Systematic Literature Reviews: A Case Study in FinTech and Automated Tool Support

Master's Thesis

Wim Spaargaren

Systematic Literature Reviews: A Case Study in FinTech and Automated Tool Support

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

Wim Spaargaren
born in Amsterdam, The Netherlands

Software Engineering Research Group
Department of Software Technology
Faculty EEMCS, Delft University of Technology
Delft, the Netherlands
www.ewi.tudelft.nl

ING Bank Personeel B.V.
Acanthus, Bijmerdreef 24
Amsterdam, the Netherlands
www.ing.nl

Systematic Literature Reviews: A Case Study in FinTech and Automated Tool Support

Author: Wim Spaargaren
Student id: 4178068
Email: w.j.spaargaren@student.tudelft.nl

Abstract

Context: Systematic literature reviews in software engineering as well as other disciplines, serve as the foundation for sound scientific research. The aim for these literature reviews is to aggregate all existing knowledge on a research problem and produce informed guidelines for practitioners. This enables practitioners to apply appropriate software engineering solutions in a specific contexts. However, one major problem exists regarding systematic literature reviews, the overall execution duration may take up as much as 24 months.
Objective & method: The first objective of this study is to provide a solid base for the AI for FinTech Research collaboration by performing a systematic literature review. This literature review is used to identify different machine learning techniques in the context of the FinTech domain. However, during this study, we found that a significant amount of time was spent on repetitive work which potentially could have been automated. Therefore, the second objective of this work is to reduce the overall workload for performing systematic literature reviews. First, a literature review is performed regarding automation solutions for different steps of systematic literature reviews. The identified solutions were used to create a tool to automate steps in both the retrieval and screening phase of systematic literature reviews.
Results & conclusions: First, this work presented the state of the art regarding machine learning applications in the FinTech domain. Afterwards, a complete overview of possible automation solutions for every step of performing literature reviews was detailed. Using this overview, a tool was created which showed that the overall workload of the retrieval and screening phase of systematic literature reviews can be significantly reduced.
Keywords: Systematic Literature Review, Information retrieval, Automation, FinTech, Machine Learning

Thesis Committee:

Chair:	Prof. Dr. A. van Deursen, Faculty EEMCS, TU Delft
University supervisor:	Prof. Dr. A. van Deursen, Faculty EEMCS, TU Delft
Company supervisor:	Dr. Hennie Huijgens, ING
Committee Member:	Dr. Georgios Gousios, Faculty EEMCS, TU Delft
	Dr. Jan Rellermeyer, Faculty EEMCS, TU Delft

Acknowledgments

This study was performed within the AI for FinTech Research(AFR) group, a research collaboration between ING and Delft University of Technology. Therefore, I would like to thank Arie van Deursen, Hennie Huijgens, Georgios Gousios, Jan Rellermeyer, Floris den Hengst, Elvan Kula, Luís Cruz and Martijn Steenbergen for their willing contributions to this study.

Wim Spaargaren
Delft, the Netherlands
July 1, 2020

Contents

List of Figures

Chapter 1

Introduction

Systematic literature reviews in software engineering as well as other disciplines, serve as the foundation for sound scientific research. The aim for these literature reviews is to aggregate all existing knowledge on a research problem and produce informed guidelines for practitioners. This enables practitioners to apply appropriate software engineering solutions in a specific contexts [45]. The execution of a literature review consists of three main phases; (1) The retrieval phase, used to retrieve relevant literature; (2) The screening phase, used to exclude any irrelevant literature found during the initial retrieval; and (3) Data synthesis, used to extract and convert data from relevant literature into common representations.

This work was carried out as part of AI for FinTech Research (AFR) [73], a scientific research collaboration between Delft University of Technology and ING Bank. The aim of this collaboration is to perform world-class research in the FinTech context with regard to Artificial Intelligence (AI), Data Analytics and Software Analytics. Financial technology (FinTech) refers to the use of technology to deliver financial solutions. The importance of research with regard to FinTech and AI is characterized by major disruptive developments in the banking industry. For example, increasing rules and regulations enforce banks to implement structural measures in the field of Customer and Customer Due Diligence to prevent money laundering and terrorist financing.

This work was carried out at the beginning of the AFR collaboration. Therefore, it was decided that this work should start by providing a solid base for the research collaboration. This was done by performing a systematic literature review, where the main focus of this review was to identify literature, describing applications of different machine learning techniques in the context of the FinTech domain.

However, during the execution of this systematic literature review one major problem was identified. It took two researchers over 3 months to complete this study, included as Chapter 2 in this Thesis. This review was set up as mapping study, which serves a the same purpose as a systematic literature review, but in general covers a wider scope. According to [101], literature surveys may even take up to as much as 24 months. Therefore, the production cost of systematic literature reviews are high and the by the time the results are produced, new literature answering the same research question might have been published. Since the systematic literature review should serve as the base of the actual research, the execution of the actual research can therefore suffer a significant delay. Because of the literature review duration, the goal of this thesis was to create a tool to reduce the overall workload needed, in order to perform a systematic literature review, such that the overall time needed to execute a systematic literature review could be reduced. In order to keep the amount of work within the limits of a nine months master thesis, the focus of the tool was set to be on both the retrieval and screening phase. These phases were chosen, because they serve as the starting point of a systematic literature review and consume a significant amount of time.

In order to create a useful tool which supports the automation of literature reviews, first the state of the art regarding automating literature reviews needs to be identified. Therefore, a short review of literature, regarding the automation of systematic literature reviews was performed, which is detailed in Chapter 3. The results of this literature review served as input for the solution design, choosing one of the identified solutions from previous literature for every step in the retrieval and screening phase of a systematic literature review. The chosen solutions resulted in a solution design for the tool which is elaborated in Chapter 4. This solution design was used in order to realize the actual implementation of the tool. The implementation details can be found in Chapter 5. In order to ensure that the tool was actually useful, Chapter 6 describes an evaluation of the created tool. Based on our evaluation, we discuss implications and future work in Chapter 7, after which, we summarize our overall conclusions in Chapter 8.

Chapter 2

Machine Learning in FinTech: A Structured Mapping Study

2.1 Introduction

The traditional banking industry has recently been characterized by two disruptive developments. Firstly, new high-tech parties face challenges in the field of financial technology, such as settlement services supported by blockchain and distributed ledgers [14]. Secondly, increasing rules and regulations put pressure on banking processes such as Know Your Customer and Customer Due Diligence, where banks are forced to implement structural measures to prevent money laundering and terrorist financing. Increasing technological possibilities—in particular in the field of artificial intelligence and machine learning, in close collaboration with innovative software engineering—are pre-eminently tools that can help to solve such complex challenges. That is why it is important for financial institutions to gain insight into the parts of their organization where the application of such new technologies lead to the greatest possible impact on business results. However, a prerequisite for such an overview—an inventory of the state of affairs in academic and industry research, regarding the use of new technologies in FinTech, preferably structured in an empirical study—is, as far as we can judge, not available. To fill this gap, we conduct a systematic mapping study in the field of machine learning (ML) and software engineering (SE) in the context of financial technology (FinTech). Systematic mapping studies or scoping studies are designed to give an overview of a research area through classification and counting contributions in relation to the categories of that classification [50]. The analysis of results focuses on frequencies of publications for categories within the scheme. Thereby, the coverage of the research field can be determined [78]. This work was carried out as part of AI for FinTech Research (AFR) [73], a scientific research collaboration between Delft University of Technology and ING Bank.

2.1.1 Challenges in FinTech

Financial technology—or FinTech—refers to the use of technology to deliver financial solutions. The term's origin can be traced to the early 1990s and referred to the "Financial

Services Technology Consortium", a project initiated by Citigroup in order to facilitate technological cooperation efforts [5]. FinTech can conceptually be defined as a new type of financial service based on IT companies' broad types of users, which is combined with IT technology and other financial services like remittance, payment, asset management and so on. FinTech includes all the technical processes from upgrading financial software, to programming a new type of financial software which can affect a whole process of finance service. Therefore, FinTech has the potential to improve the performance of financial services and spread the finance service combined with mobile environment [53].

The term FinTech now refers to a large and rapidly growing industry representing between US$12 billion and US$197 billion in investment as of 2014, depending on whether one considers start-ups (also referred to as FinTech 3.0) or traditional financial institutions (FinTech 2.0) respectively [5]. And—not surprisingly—new terminology is already popping-up; Finbrain: When finance meets AI 2.0 [112].

FinTech is increasingly characterized by disruption [110]. Old problems still occur—banks still struggle with culture and legacy systems [52]— but new ones enter the arena. New laws for payment transactions of consumers and businesses (e.g. PSD2), artificial intelligence, intellectual property, and blockchain influence banking processes to an ever greater extent [66]. Increasing rules and regulations in the field of Customer Due Diligence are causing more and more manual work in order to prevent risks such as money laundering and terrorist financing [51], leading to new approaches by making use of new technologies, such as social network analysis [18], robo-advisors adoption among customers [9], and a variety on interactive-agent applications [15, 38, 32, 67, 29].

On the other hand, new horizons open up. Fintech turns data into new business, powered by new technology [36]. Organisations are forced to "know their customers" for a number of reasons, such as increased access to financial services in emerging markets [3], and impact on consumers and regulatory responses [39]. Start-ups are playing a key role in helping the financial sector determine what ML—often in combination with SE—can do and how humans and machines can work together [111].

2.1.2 A focus on sub-symbolic AI (ML)

In this mapping study we limit our scope to sub-symbolic AI, or machine learning (ML). Most of what is considered AI today is actually sub-symbolic AI [12]. Symbolic AI is about developing intelligent systems based on symbolic reasoning about rules and knowledge, with actions that can be interpreted. Symbolic AI is sometimes called GOFAI, for "Good Old-Fashioned Artificial Intelligence" [35]. On the contrary, sub-symbolic AI—sometimes also referred to as non-symbolic AI—strives to build computational systems that are inspired by the human brain. It is about performing calculations according to principles that have demonstrated to be able to solve problems, without exactly understanding how to arrive at a solution. This is what is commonly called ML, including approaches such as genetic algorithms, neural networks and deep learning.

Although not clearly confirmed in empirical studies, the general opinion among specialists seems to be that symbolic AI has been confined to the academic world and university labs with little research coming from industry [10]. Sub-symbolic AI on the contrary seems more applied in practice, more revolutionary, futuristic and quite frankly, easier on the developers [21]. Therefore, in this mapping study we limit ourselves to the topic of sub-symbolic AI, in the remains of this study referred to as machine learning (ML).

Although an overview study with regard to ML specifically in the FinTech domain is missing, there are indeed surveys that address ML aspects themselves. However, usually these surveys relate to research into the application possibilities of a specific ML technique or algorithms, e.g. [83, 71, 54, 42, 84, 98, 72, 99, 7]. Although some recent studies do indeed describe life cycle aspects with regard to ML [6, 82], this often concerns the life cycle of the training and test data used, e.g. [80, 56].

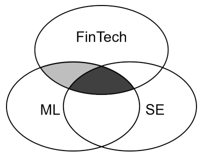
As far as we know, a large-scale study that maps the field of ML from an interest to gain overview and insight into the FinTech domain as a whole, does not exist. Two important findings from our mapping study support the importance of our study itself. Firstly, the number of scientific studies overall shows an annual doubling of published studies in the last three years—a growth that appears to reflect the massive increase in ML models in industry. Secondly, a study that focuses on the life cycle aspects of ML models—think of topics such as continuous integration, pipeline automation, maintainability, verification, deployment, and testing of ML models—does not seem to exist at the moment of writing. Given the strong growth of ML applications in the FinTech domain, we argue that a focus on research on the life cycle aspects of ML is absolutely necessary in the near future. In order to address the state of affairs in these life cycle aspects of ML in FinTech, we also examine to what extent studies overlap with specific software engineering (SE) topics.

2.1.3 Overlap of ML with SE

Although resistance to the potential benefits of new technologies occur in spite of the fact that technology is transforming the industry [26], FinTech is merging new technologies to challenge banks [85] [86]. ML plays a pivotal role in transforming business intelligence into a fully predictive probabilistic framework, enabling automation of numerous functions within companies, from pricing, budget allocation, to fraud detection and security [104].

ML and SE are often interconnected in practical applications, leading to the field of intelligent software engineering. Xie recognizes two representations of the term: *[intelligent] software engineering* and *[intelligent software] engineering*, hence ML [109]. Challenges do exist when bringing ML and SE together, such as the accuracy of systems built using ML models, since ML systems cannot guarantee 100 percent accuracy or correct answers in all cases. Furthermore, ML systems can be very difficult to test. Regrettably, a rift occurs between communities, mainly because stakeholders in the ML community focus on algorithms and their performance characteristics, whereas stakeholders in the SE community focus on implementing and deploying those algorithms [46]. We argue that combining ML and SE in a mapping study can lead to new insights and other ways of solving problems. That is why we limit our mapping study to the area where FinTech and ML overlap, and in particular where SE, ML and FinTech coincide (resp. the grey and black area in Figure 2.1).

Figure 2.1: The Scope of the Mapping Study.



2.1.4 Research Questions

Based on the above, we address the following research questions:

RQ1: Which studies address the application of ML in FinTech?

RQ2: Which countries, universities, researchers, journals and conferences are leading in ML in FinTech studies?

RQ3: What are the most frequently applied research methods, and in what study context?

RQ4: What are the most investigated topics on ML and FinTech, and how have these changed over time?

RQ5: To which extent are life cycle aspects of SE addressed in ML in FinTech studies?

The remainder of this paper is structured as follows. In Section 2.2 we outline the study design. The results of the mapping study are described in Section 2.3. We discuss the results in Section 2.4, and finally, in Section 2.5 we make conclusions and outline future work.

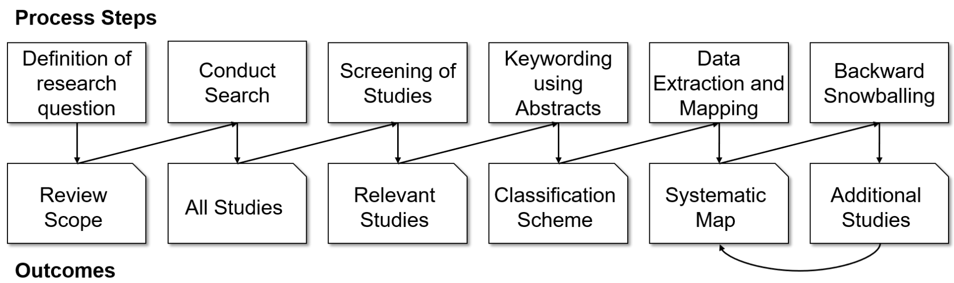
2.2 Research Design

For this study, we have adapted an applied systematic mapping, as performed earlier in a studies on software engineering [78]. We used a set of guidelines for conducting systematic mapping studies in software engineering [79] as a reference. The process that we applied for the systematic mapping study is defined by [78] and depicted in Figure 2.2. In the mapping process the following six process steps are defined: definition of research question, conduct search, screening of studies, keywording using abstracts, data extraction and mapping, backward snowballing.

2.2.1 Definition of Research Questions

The main goal of our systematic mapping study is to provide an overview of research in the topic of ML combined with SE in the area of FinTech (including banking), and to identify the quantity and type of research and its results. A secondary goal is to identify any forums (e.g. conferences and journals) in which research on this topic is published, and what trend occurs in frequency of publications over time. These goals are reflected in the research questions, as shown in Section 2.1.4. The outcome of this process step is an overview of the review scope.

Figure 2.2: The Systematic Mapping Process as defined by [78].



2.2.2 Conduct Search

Based on the review scope we designed search strings on scientific databases in order to identify primary studies. The identified keywords are on the one hand FinTech, financial technology, and banking, and on the other hand artificial intelligence, AI, machine learning, and deep learning, which are grouped into sets and their synonyms are considered to formulate the search string.

1. *Set 1*: Scoping the search for FinTech, i.e. ”FinTech”, ”financial technology” or ”banking”.

2. *Set 2*: Search terms directly related to sub symbolic AI, e.g. ”artificial intelligence”, ”AI”, ”machine learning”, or ”deep learning”.

We performed a search on five target databases, based on the above search strings: IEEE Xplore, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, and Web of Science. Google Scholar was omitted, since the query resulted in more articles then Scholar is able to show. The databases have been selected based on the experience reported by Dyba et al. [22]. We used the following search string for each database, applied on all fields:

> ((”fintech” OR ”financial technology” OR ”banking”) AND (”AI” OR ”artificial intelligence” OR ”machine learning” OR ”deep learning”))

For this purpose, we build a web scraper to be able to automatically download publications from the five target research databases. The web scraper stored relevant information about articles in a database, removed duplicates and managed the large number of references. This study has been conducted during 2019, the studies of October 2019 and before have been considered during the search.

2.2.3 Screening of studies

We applied inclusion and exclusion criteria to exclude studies that are not relevant to answer the research questions.

Table 2.1: Research Topics used as a Reference in the Keywording Process based on the definition stated in [78].

Research Topic	Description
Evaluation Research	Techniques are implemented in practice and an evaluation of the technique is performed. This means that it is demonstrated how the technology is implemented in practice (implementation of solutions) and what the consequences are of the implementation in terms of advantages and disadvantages (evaluation of the implementation). This also includes identifying problems in the industry.
Experience Papers	Experience documents explain what and how something has been done in practice. It must be the author's personal experience.
Opinion Papers	These articles reflect someone's personal opinion, whether a particular technique is good or bad, or how things should be done. They are not dependent on related work and research methods.
Philosophical Papers	These articles outline a new way of looking at existing things by structuring the field in the form of a taxonomy or conceptual framework.
Solution Proposal	A solution to a problem is proposed, the solution can be new or an important extension of an existing technique. The potential benefits and applicability of the solution are demonstrated by a small example or sound reasoning.
Validation Research	Studies that describe a process of confirming that a new or existing solution can continue or commence operation. In general, validity is an indication of how sound research is. More specifically, validity applies to both the design and the methods of a study.

1. *Inclusion*: Empirical studies that are published in the time frame 2006 to October 2019, regarding sub-symbolic artificial intelligence, Machine Learning, or Deep Learning in a FinTech or banking context. Where several papers report the same study, only the most recent is included.

2. *Exclusion*: Books and grey literature. Studies that do not report empirical findings, or literature that is only available in the form of abstracts or PowerPoint presentations. Studies presenting non-peer reviewed material, studies that are not presented in English, studies that are not accessible in full-text, or studies that are duplicates of other studies. Furthermore, we excluded studies on any symbolic AI topics.

2.2.4 Keywording using Abstracts

We applied keywording as a technique to reduce the time needed in developing a classification scheme as well as to ensure that the scheme takes existing studies into account [78, 79]. The keywording followed two steps (see Figure 2.3):

Table 2.2: ML Topics used as a Reference in the Keywording Process.

ML Topic	Description
Artificial Neural Network Algorithms	Perceptron, Multilayer Perceptrons (MLP), Back-Propagation, Stochastic Gradient Descent, Hopfield Network, Radial Basis Function Network (RBFN).
Association Rule Learning Algorithms	Apriori algorithm, Eclat algorithm.
Bayesian Algorithms	Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Averaged One-Dependence Estimators (AODE), Bayesian Belief Network (BBN), Bayesian Network (BN).
Clustering Algorithms	k-Means, k-Medians, Expectation Maximisation (EM), Hierarchical Clustering.
Deep Learning Algorithms	Convolutional Neural Network (CNN), Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), Stacked Auto-Encoders, Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN).
Decision Tree Algorithms	Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0 (different versions of a powerful approach), Chi-squared Automatic Interaction Detection (CHAID), Decision Stump, M5, Conditional Decision Trees.
Dimensionality Reduction Algorithms	Principal Component Analysis (PCA), Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), Sammon Mapping, Multidimensional Scaling (MDS), Projection Pursuit, Linear Discriminant Analysis (LDA), Mixture Discriminant Analysis (MDA), Quadratic Discriminant Analysis (QDA), Flexible Discriminant Analysis (FDA).
Ensemble Algorithms	Boosting, Bootstrapped Aggregation (Bagging), AdaBoost, Weighted Average (Blending), Stacked Generalization (Stacking), Gradient Boosting Machines (GBM), Gradient Boosted Regression Trees (GBRT), Random Forest.
Instance-based Algorithms	k-Nearest Neighbor (kNN), Learning Vector Quantization (LVQ), Self-Organizing Map (SOM), Locally Weighted Learning (LWL), Support Vector Machines (SVM).
Regression Algorithms	Linear Regression, Logistic Regression, Stepwise Regression, Multivariate Adaptive Regression Splines (MARS), Locally Estimated Scatterplot Smoothing (LOESS).
Regularization Algorithms	Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, Least-Angle Regression (LARS).
Reinforcement learning	Q-learning, State–action–reward–state–action (SARSA), Temporal difference learning (TD), Learning Automata.

Table 2.3: FinTech Topics used as a Reference in the Keywording Process.

Fintech Topic	Description (partly based on [77])
Credit Cards	Payment cards issued by a bank to cardholders, to enable the cardholder to pay a merchant for goods and services based on the cardholder's promise to the card issuer to pay them for the amounts plus the other agreed charges.
Credit Risk	The risk of default on a debt that may arise from a borrower failing to make required payments. Including credit scoring, a statistical analysis performed by financial institutions to access a customer's creditworthiness in order to help decide on whether to extend or deny credit.
Cryptocurrency	A digital asset as a medium of exchange that uses strong cryptography to secure financial transactions, control the creation of additional units, and verify the transfer of assets.
Customer Due Diligence (CDD)	The process of gathering and recording the identity of customers (or firms), and assess risks related to customers, such as money laundering and terrorist financing [18].
E-Banking	Also known as online banking or internet banking; an electronic payment system that enables customers of a financial institution to conduct financial transactions through a secure website.
Fraud Detection	Preventing money or property from being obtained through false pretenses, such as forging checks or using stolen credit cards.
Loans and Other Credit Products	Lending of money by a financial institution to customers (individuals or organizations), such as Long-Term Loans, Short-Term Loans, Lines of Credit, and Alternative Financing.
Malware Detection	The process of scanning the computer and files to detect malware.
Market Risk	Also called systematic risk. The possibility of an investor experiencing losses due to factors affecting the performance of the financial markets.
Mergers & Acquisitions (M&A)	Transactions in which the ownership of companies, or their operating units are transferred or consolidated with other entities.
Operational Risk	The risk of a change in value due to failed internal processes, people and systems, or from external events (including legal risk).
Phishing	Phishing (derived from fishing) is a form of internet fraud.
Private Banking	Financial services provided by banks to well-to-do individuals.
Stock Brokerage	Financial advisory and investment management services and execute transactions such as the purchase or sale of stocks and other investments to financial market participants.
Trade Finance	Used by importing and exporting companies and is an internationally accepted way of financing an order.

Table 2.4: SE Topics used as a Reference in the Keywording Process.

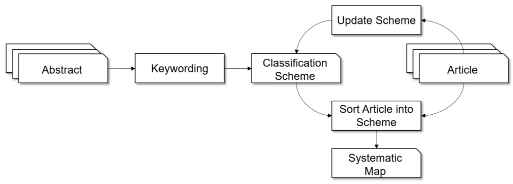
SE Topic	
Agile software development	Performance
AI and software engineering	Program / model analysis
Autonomic and (self-)adaptive systems	Program / model comprehension
Cloud computing	Program / model repair
Component-based SE	Program / model synthesis
Configuration management	Programming languages
Crowd sourced SE	Recommendation systems
Debugging	Refactoring
Dependability, safety, and reliability	Requirements engineering
Deployment	Reverse engineering
Distributed and collaborative SE	Search-based SE
Embedded software	Security, privacy, and trust
Empirical software engineering	Software architecture
End-user software engineering	Software economics and metrics
Evolution and maintenance	Software modeling and design
Fault localization	Software process
Formal methods	Software product lines
Green and sustainable technologies	Software reuse
Human and social aspects	Software services
Human-computer interaction	Software testing
Middleware, frameworks, and APIs	Specification and modeling languages
Mining repositories	Tools and environments
Mobile applications	Traceability
Model-driven engineering	Validation and verification

1. The reviewers read the abstract and searched for keywords and concepts that reflect the paper's contribution. The reviewers also identified the context of the research.

2. Once this was done, the set of keywords from different articles was combined together to develop a better understanding of the nature and contribution of the research. This helped the assessors to come up with a series of categories that are representative of the underlying population. Once a definitive set of keywords had been chosen, they were clustered and used to form the categories for a systematic map.

To support the keywording process, we compiled four overviews beforehand that we used during that process for reference:

Figure 2.3: Building the Classification Scheme as defined by [78].



1. *Research Topics*: To gain insight into the different types of research, we set up a reference overview that is derived from Wierenga et al. [108], [78] (see Table 2.1).

2. *ML Topics*: Table 2.2 provides an overview of the list of ML topics that we used as a reference for the keywording process. This overview has been prepared in collaboration with AI and ML specialists working in both organizations within the scope of this study.

3. *FinTech and Banking Topics*: An overview of applicable FinTech and banking topics has been prepared in collaboration with banking specialists within both organizations within the scope of this study (see Table 2.3).

4. *Software Engineering Topics*: We used a subset of software engineering topics as mentioned in a call for papers of the International Conference of Software Engineering (ICSE'20) [1] as a reference to examine any overlap with software engineering topics (see Table 2.4). This was done in order to see which SE methods were used in order to apply different sub-symbolic AI methods.

2.2.5 Data Extraction and Mapping of Studies

Once the classification scheme had been set up—based on the before mentioned reference tables—the papers that were in scope were classified and sorted according to this scheme. As indicated in Figure 2.3, the scheme evolved during the data classification. The analysis focused on mapping the frequencies of publications for the different categories. This way, we gained insight into which categories have been emphasized in research, and where there are gaps and therefore opportunities for future research. The results of the analysis are shown in a Systematic Map in the form of a bubble plot detailed in Figure 2.13.

2.2.6 Additional Backward Snowballing

Besides the five initial databases, we decided after performing the keywording and mapping analysis to include two additional sources to our search target. First, we included the International Joint Conference on Artificial Intelligence (IJCAI)[1] and the Journal of Artifi-

[1] https://www.ijcai.org/

Table 2.5: Number of studies per Database (incl. snowballing).

Database	Search Results	Results after Exclusion
IEEE Xplore	458	132
ACM Digital Library	1543	78
ScienceDirect (Elsevier)	110	11
SpringerLink	305	14
Web of Science	571	64
IJCAI / JAIR	4	3
Backward Snowballing	464	37
Total	3455	339

cial Intelligence Research (JAIR)[2] in our search. This conference and journal were chosen, because these are a major (A*)[3], open access conference and journal in the field of AI.

Secondly, in addition, we applied—to further optimize the subset of included studies—so-called backward snowballing, as recommended by Jalali and Wohlin [40]. We used the top-15 of the most cited studies—before backward snowballing—as a starting point for this. The result of this additional step is a top-15 of most cited papers—after backward snowballing.

Table 2.5 shows the number of search results per database, after applying inclusion and exclusion rules, and the additional backward snowballing.

2.3 Results

In this section we describe the results of the analysis of studies, the categorization of studies based on keywords and abstracts, and we plot the results of the mapping process in a bubble chart. We include overviews of analyzed data in a summarized way. A complete overview of all data used for analysis purposes in this mapping study is to be found in a publicly available analysis repository [96].

2.3.1 Screening of Papers

The initial set of papers after extraction of five libraries consisted of 2,991 studies, as shown in Figure 2.4. Application of the inclusion and exclusion rules, as described in subsection 2.2.3, resulted in a subset of 321 studies that were used as input for our mapping study. During the data extraction and mapping process itself another 19 studies where excluded, because they turned out to be out of scope for our study (e.g. they where about medical banking). Lastly, 302 studies where in scope of our study and used for mapping purposes.

As described in Section 2.2.6, we decided to extend the mapping by performing backward snowballing on the list of references from the top-15 most cited studies. Table 2.6

[2] https://www.jair.org/index.php/jair

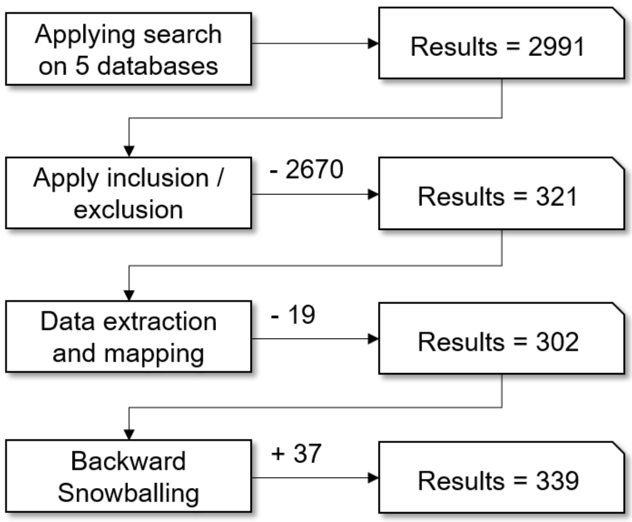
[3] flagship conference, a leading venue in a discipline area

Table 2.6: Top-15 Articles and Authors ranked on Cites before backward snowballing.

Article Title	Authors	Cites / year	Cites	Journal / Conference Name
Credit scoring with a data mining approach based on support vector machines (2007) [227]	Cheng-Lung Huang, Mu-Chen Chen, and Chieh-Jen Wang	61	729	Expert systems with applications (Elsevier Journal)
Defending against phishing attacks: taxonomy of methods, current issues and future directions (2018) [220]	B. B. Gupta, Nalin A.G. Arachchilage, and Konstantinos E. Psannis	45	45	Telecommunication Systems (Springer Journal)
Genetic algorithm-based heuristic for feature selection in credit risk assessment (2014) [323]	Stjepan Oreski and Goran Oreski	43	216	Expert systems with applications (Elsevier Journal)
A machine learning approach to Android malware detection (2012) [360]	Justin Sahs and Latifur Khan	37	261	European Intelligence and Security Informatics Conference (IEEE)
Credit card fraud detection using hidden Markov model (2008) [385]	Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun Majumdar	33	366	Transactions on Dependable and Secure Computing (IEEE Journal)
The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients (2009) [430]	I-Cheng Yeh and Che-hui Lien	33	329	Expert Systems with Applications (Elsevier Journal)
Customer churn prediction using improved balanced random forests (2009) [422]	Yaya Xie, Xiu Lia, E.W.T. Ngai, and Weiyun Ying	29	290	Expert Systems with Applications (Elsevier Journal)
Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods (2009) [155]	Melek Acar Boyacioglu, Yakup Karab Ömer, and Kaan Baykan	25	248	Expert Systems with Applications (Elsevier Journal)
Effective detection of sophisticated online banking fraud on extremely imbalanced data (2013) [415]	Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen	25	147	World Wide Web (Springer Journal)
Classifiers consensus system approach for credit scoring (2016) [121]	Maher Ala'raj and Maysam F. Abbod	24	73	Knowledge-Based Systems (Elsevier Journal)
Real-time credit card fraud detection using computational intelligence (2008) [344]	Jon T.S. Quah and M. Sriganesh	22	247	Expert Systems with Applications (Elsevier Journal)
Predicting failure in the US banking sector: An extreme gradient boosting approach (2019) [161]	Pedro Carmona, Francisco Climent, and Alexandre Momparler	16	16	International Review of Economics & Finance (Elsevier Journal)
Neural nets versus conventional techniques in credit scoring in Egyptian banking (2008) [116]	Hussein Abdou, John Pointon, and Ahmed El-Masry	16	173	Expert Systems with Applications (Elsevier Journal)
Bankruptcy prediction using a data envelopment analysis (2004) [179]	Anja Cielen, Ludo Peeters, and Koen Vanhoof	14	215	European Journal of Operational Research (Elsevier Journal)
Genetic algorithms applications in the analysis of insolvency risk (1998) [407]	Franco Varetto	14	296	Journal of Banking & Finance (Elsevier Journal)

This top-15 is based on an initial subset of 302 analyzed studies, and subsequently used as input for backward snowballing. In order to compare articles published in different years, we ranked articles by cited average per year.

Figure 2.4: Quantification of the study selection process.



provides an overview of these initial top-15 studies. During the backward snowballing we assessed 464 studies, of which we included 37 studies to our included subset, resulting in a total set of 339 studies which were included in the analysis. Figure 2.4 provides an overview of the results of the inclusion and exclusion rules after each selection step. In the remaining parts of this mapping study in all cases the total set of 339 included studies is used.

2.3.2 Classification of Studies

To provide a clear overview of the different categories within the scope of our study, the set of studies included for the analysis were categorized into the following topics: 1) country and continent, 2) conferences, journals, researchers, and universities, 3) research methods, 4) life cycle aspects from software engineering, 5) FinTech topics, and 6) ML topics. All mentioned aspects are explained in the following sections respectively.

Country and Continent

Before we start to explore the continent and country results, it must be noted that for some articles, researchers from more than one country could have collaborated. In this case, for one single article, every distinct country is counted once. Therefore, not 339 countries, but instead 405 countries were counted.

An overview of studies grouped by continent are presented in Table 2.7. As the overview clearly shows, a vast majority of studies was published by researchers at universities and companies in Asia (47%), which is followed at number two by European universities and companies, accounting for (32%). It is also worth mentioning that only 13% of the studies in the scope where from North American universities. Africa, Australia and South America conclude the list and together account for only 7% of the research.

Table 2.7: Number of studies per Continent.

Continent	Count	Percent-age
Asia	191	47%
Europe	130	32%
North America	53	13%
Africa	12	3%
Australia	13	3%
South America	6	1%

Table 2.8 inventories countries that published three or more articles in the scope of our mapping study. As shown in the table, the top three consists of China, India and the USA, covering 16%, 12% and 11% respectively.

Table 2.9: Top-10 Universities & Companies ranked by cites.

University/Company	Rank	Country	Cited Studies
University of South Carolina	173	USA	[148]
Tilburg University	173	Netherlands	[148]
University of Pennsylvania	173	USA	[148]
Case Western Reserve University	173	USA	[148]
The Pennsylvania State University	141	USA	[204, 228]
National Chiao Tung University	108	Taiwan	[417, 404, 169, 227, 419]
The Open University of Hong Kong	102	China	[402]
University of Texas, Pan American	102	USA	[402]
Institute for Development and Research in Banking Technology, Castle Hills Road	81	India	[265]
Henan University of Science & Technology	63	China	[336]

In order to compare articles published in different years, we calculated Rank for articles as a weighted rank; cited average per year.

Table 2.8: Number of studies per Country.

Country	Count	Percentage
China	64	16%
India	50	12%
United States of America	44	11%
Taiwan	22	5%
Turkey	19	5%
United Kingdom	18	4%
Spain	13	3%
Australia	13	3%
Italy	12	3%
France	11	3%
Greece	11	3%
Iran	10	2%
Canada	9	2%
Singapore	6	1%
Germany	6	1%
Morocco	6	1%
South Korea	6	1%
Pakistan	5	1%
Japan	4	1%
Poland	4	1%
Finland	4	1%
Russia	4	1%
Brazil	4	1%
Portugal	4	1%
Cyprus	3	1%
Malaysia	3	1%
United Arab Emirates	3	1%
The Netherlands	3	1%
Romania	3	1%
Thailand	3	1%
Hungary	3	1%

Countries covering less than three studies are not included in this table.

Table 2.10: Number of studies per Research Method.

Research Method	Count	Percentage
Solution Proposal	290	86%
Evaluation Research	36	11%
Philosophical Paper	7	2%
Experience Paper	3	1%
Validation Research	2	1%
Opinion Paper	1	0%

Conferences, Journals, Researchers, and Universities

A top-15 of most cited studies—after backward snowballing—is presented in Table 2.11. Instead of using the number of studies occurring within the scope of our study, we used the number of citations of each study as a reference for ranking, where we calculated a weighted average citation count per year to normalize studies for comparison.

With regard to journals and conferences, Elsevier's journal *Expert Systems with Applications* stands head and shoulders above others; three studies in the top 15 were published in this journal and it also has the highest rank.

Table 2.12: Number of studies per SE topic.

Lifecycle Aspect of SE	Count	Percentage
NA	318	94%
AI and Software Engineering	15	4%
Requirements Engineering	2	1%
Software Services	1	0%
Model-Driven Engineering	1	0%
Software Architecture	1	0%
Security, Privacy, and Trust	1	0%

Table 2.13, provides an overview of conferences and journals in which the most prevalent articles of our subset have been published. The table depicts journals and conferences which were identified for three articles or more. For every conference and journal, we inventoried the applicable H5-Index. As can be seen, the Elsevier journal *Expert Systems with Applications* was identified for 27 different papers, accounting for 8% of the studies and also has the highest H5-Index in the list of most used journals and conferences. Furthermore, it is worth mentioning that only three of the journals and conferences which are published in most occurring journals and conferences, are included in the top-15 (see Table 2.11); respectively *Expert Systems with Applications*, *European Journal of Operational Research* and *Journal Of Banking & Finance*.

Table 2.11: Top-15 Articles and Authors ranked on Cites after backward snowballing.

Article Title	Authors	Rank	Total Cites	Journal / Conference Name
How does capital affect bank performance during financial crises? (2013) [148]	A.N. Berger, C.H.S. Bouwman	173	1040	Journal of Financial Economics (Elsevier Journal)
Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review (2007) [265]	P.R. Kumar, V. Ravi	81	977	European Journal of Operational Research (Elsevier Journal)
A comprehensive survey of data mining-based fraud detection research (2007) [336]	C. Phua, V. Lee, K. Smith, R. Gayler	63	760	arXiv preprint arXiv:1009.6119
Cantina: a content-based approach to detecting phishing web sites (2007) [438]	Y. Zhang,J. I. Hong, L. F. Cranor	62	745	In Proceedings of the 16th international conference on World Wide Web
Credit scoring with a data mining approach based on support vector machines (2007) [227]	C.-L. Huang, M.-C. Chen, C.-J. Wang	61	734	Expert Systems with Applications (Elsevier Journal)
Nontraditional banking activities and bank failures during the financial crisis (2013) [197]	R. DeYoung, G. Torna	61	366	Journal of Financial Intermediation (Elsevier Journal)
The state of phishing attacks (2012) [224]	J. Hong	56	396	Communications of the ACM (ACM Journal)
Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions (2010) [375]	S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, J. Downs	51	462	In 28th international conference on human factors in computing systems
Predicting distress in European banks (2014) [150]	F. Betz, S. Oprica, T.A. Peltonen, P. Sarlin	46	232	Journal of Banking & Finance (Elsevier Journal)
Defending against phishing attacks: taxonomy of methods, current issues and future directions (2018) [220]	B. B. Gupta, Nalin A.G. Arachchilage, and Konstantinos E. Psannis	45	45	Telecommunication Systems (Springer Journal)
Genetic algorithm-based heuristic for feature selection in credit risk assessment (2014) [323]	S. Oreski and G. Oreski	43	216	Expert systems with applications (Elsevier Journal)
Phishing detection: a literature survey (2013) [252]	Khonji, M., Iraqi, Y., & Jones, A.	37	226	IEEE Communications Surveys & Tutorials (IEEE Journal)
A machine learning approach to Android malware detection (2012) [360]	J. Sahs and L. Khan	37	261	European Intelligence and Security Informatics Conference (IEEE)
Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction (2016) [448]	M. Maciej Zieba, S.K. Tomczak, J.M. Tomczak	35	107	Expert Systems with Applications (Elsevier Journal)
Credit card fraud detection using hidden Markov model(2008) [385]	A. Srivastava, A. Kundu, S. Sural, A. Majumdar	34	366	IEEE Transactions on Dependable and Secure Computing

This top-15 is based on the final subset of 339 analyzed studies, after backward snowballing. In order to compare articles published in different years, we calculated Rank for articles as a weighted rank; cited average per year.

Table 2.13: Number of studies per Conference or Journal.

Publisher	Name Conference or Journal	C/J	H5 Index	Count	Percentage
Elsevier	Expert Systems with Applications	Journal	105	27	8%
Springer	International Conference On Machine Learning And Cybernetics	Conference	30	11	3%
Elsevier	European Journal Of Operational Research	Journal	92	6	2%
Elsevier	Applied Soft Computing	Journal	83	5	1%
Springer	Computational Economics	Journal	22	5	1%
Elsevier	Journal Of Banking & Finance	Journal	70	5	1%
Elsevier	Decision Support Systems	Journal	55	4	1%
IEEE	IEEE Access	Journal	89	4	1%
Springer	International Conference On Artificial Intelligence and Computational Intelligence	Conference	7	4	1%
Springer	Neural Computing and Applications	Springer	60	4	1%
IJCAI	International Joint Conference on Artificial Intelligence	Conference	67	3	1%
IEEE	International Conference On Advanced Computing & Communication Systems	Conference	3	3	1%
IEEE	IEEE International Conference On Big Data	Conference	33	3	1%

Conferences or journals of included studies in scope occurring only 2 times or less, are not included in this Table. Rank for journals is based on H Index, derived from Google Scholar Metrics.

We assessed for the most frequently occurring universities and companies in our study subset (see Table 2.9). To accomplish this, for every paper in scope, we identified all distinctive universities and companies which contributed to an article. Afterwards, per article, the rank indicating the average amount of cites per year, was assigned to all universities and/or companies belonging to that article. Afterwards, per university and company, the sum of ranks was taken, for every paper which the company or university published in, and a top-10 was created which is presented in Table 2.9. Only two universities in the top-10 are involved in multiple articles; both *The Pennsylvania State University* and the *National Chiao Tung University* are involved in 2 and 5 articles respectively.

Research Methods

Table 2.10 depicts how the studies in scope are divided among different research methods. The first thing to notice is that the vast majority of studies have the character of a solution proposal. In most studies, a ML approach is proposed, a ML model is trained on a data source either from industry or from an open source source, and the resulting model is tested for its performance (e.g. by assessing its *Area Under the Curve* or AUC). In almost all cases however, a validation in a practical situation in industry is missing. As a result, it is not clear how a new or modified ML model is applied in an actual system or application and how it is used in a real situation in practice.

In 36 studies (11% of the set of the studies in scope) such a description of practical application in any form is found. Seven studies were characterized as philosophical studies, usually because they outline a new way of looking at existing topics or structured the field of ML research.

Life cycle aspects of Software Engineering

Table 2.12 present the mapping SE topics on to the studies in scope. Only a limited number of studies could be mapped on ML and SE or other SE-related topics. For the vast majority of the studies we found no link with SE topics or with life cycle aspects of ML in general. Therefore, our results suggest that the life cycle of models is not a topic that is high on the agenda of ML researchers, at least within the field of FinTech.

FinTech and Banking Topics

Table 2.14 summarizes the number of studies per FinTech or banking topic. In total, 86 articles focused on credit risk related topics, accounting for 25% of all papers. In most cases, a study related to credit risk, is about some form of credit scoring. It's worth mentioning that 81% of the studies relate to the 6 most counted topics, including two risk related topics namely, credit risk and operational risk, two security-related topics: fraud detection and phishing. The top 6 is concluded by stock brokerage and the category overall bank topics. The remaining 19% of research focuses on a variety of banking topics such as payment processing, cash logistics, processing handwritten bank checks, marketing approaches, or customer behavior. Furthermore, it is worth mentioning that three FinTech topics were not covered at all in the included set of studies; checking and savings accounts, credit cards and monetary policy.

The development of the number of studies over time in the final subset, is depicted in figure 2.5. The figure breaks down the FinTech topics into three categories, where each category consists of a subset of FinTech topics: 1) Risk related papers (e.g. credit risk, credit scoring, operational risk, bankruptcy prediction), 2) Security related papers (e.g. malware detection, phishing, fraud detection, customer due diligence, money laundering), and 3) Other papers (e.g. banking related topics, stock brokerage, loans and savings, cryptocurrency). Overall, Figure 2.5 shows a peak in research on credit risk related topics during the financial crisis in 2008 and 2009. The number of credit risk related studies shows a slightly downward trend after the financial crisis, with a slight upward trend in the last three years. On the contrary, studies related to security topics, such as malware detection, phishing, fraud detection, and customer due diligence, and overall banking topics show strong growing trends since the end of the financial crisis.

ML Topics

In Table 2.15 the number of studies per ML topic are listed. The first thing to notice is that fifty percent of the research has been focused on three ML topics namely, Artificial Neural Networks, Deep learning and combination of ML algorithms. In our analysis we separated Artificial Neural Networks from Deep Learning, because of the massive growth and popularity in the field of Deep Learning. For studies focusing on combinations of ML algorithms, usually a subset of machine learning techniques is applied on a modelling

Table 2.14: Number of studies per FinTech or Banking Topic.

FinTech or Banking Topic	Count	Percentage
Credit Risk	86	25%
Operational Risk	58	17%
Fraud Detection	41	12%
Overall Bank Topics	36	11%
Phishing	35	10%
Stock brokerage	21	6%
Malware	9	3%
Loans and other credit products	7	2%
Credit cards	7	2%
Customer Due Diligence	5	1%
Cryptocurrency	5	1%
Market Risk	4	1%
Money Laundering	3	1%
Mergers & Acquisitions	3	1%
NA	3	1%
Intrusion Detection	3	1%
Personal loans	2	1%
Wealth management	2	1%
Private banking	2	1%
E-Banking	2	1%
Trade finance	2	1%
Checking and savings accounts	1	0%
User trust	1	0%
Monetary policy	1	0%

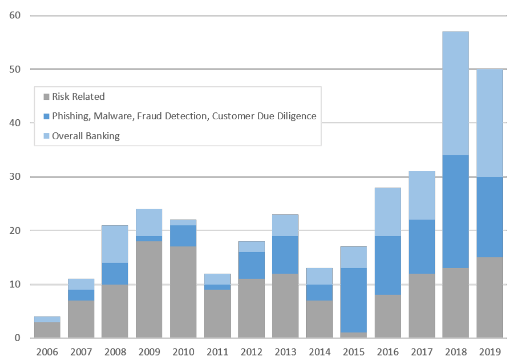
problem, where the performances of the models are compared in order to find the best fit for a specific problem. Another result worth mentioning, is the fact that only four papers focused on Reinforcement learning.

When the growth of the different ML topics is analyzed over time (see Figure 2.6), we notice a rather scattered pattern that is difficult to interpret. But splitting the growth overviews into the most occurring ML topics yields a better result, as is depicted in the Figures 2.7 to 2.12. Figure 2.7 shows that Artificial Neural Network Algorithms follows the overall pattern of ML topics in general; the technique is used for a long time and was used in studies during the financial crisis, with a clear increase in the last three years. Figure 2.8 shows that studies describing a combination of algorithms have been occurring since 2009, and that there has also been growth there over the last three years. What is immediately noticeable in Figure 2.9 is that Deep Learning is apparently a subject that was only recently broken through in FinTech studies. Here too there has been a sharp increase in studies in recent years. What is also noticeable is that Regression Algorithms, Instance-Based Algorithms, and Clustering Algorithms have been used for a long period of time in FinTech

Table 2.15: Number of studies per ML Topic.

ML Topic	Count	Percent-age
Artificial Neural Network Algorithms	71	21%
Combination of algorithms	60	18%
NA (no AI applied in the study)	36	11%
Deep Learning Algorithms	35	10%
Regression Algorithms	30	9%
Instance-based Algorithms	28	8%
Clustering Algorithms	21	6%
Ensemble Algorithms	18	5%
Decision Tree Algorithms	11	3%
Bayesian Algorithms	10	3%
Genetic programming	9	3%
Reinforcement learning	4	1%
Dimensionality Reduction Algorithms	3	1%
Association Rule Learning Algorithms	3	1%

Figure 2.5: Growth pattern of FinTech topics over time.

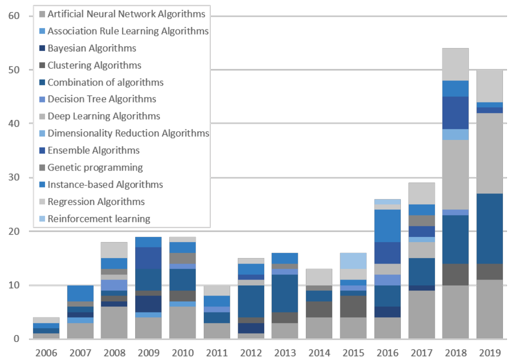


studies, but that its growth is lagging far behind the aforementioned three ML topics (see Figures 2.10, 2.11, and 2.12).

2.3.3 Results of the Mapping Process

The results of the mapping process are depicted in a bubble chart in Figure 2.13. The chart shows two matrices: 1) the left part of the chart shows ML topics on the Y-axis versus research topics on the X-axis, and 2) the right part of the chart shows ML topics on the Y-axis versus FinTech topics on the X-axis. The size of the bubbles indicates the number

Figure 2.6: Growth pattern of ML topics over time.



In the above figure 18 studies are excluded that were in scope of the mapping study but did not relate with any ML topic.

of studies that are mapped on a specific topic, small bubbles indicate little studies mapped, large bubbles indicate many studies mapped. As the part on the left side of the chart shows, most of the studies were labeled as Solution Proposal, with a relatively even spread over the ML topics. The right-hand part of the card shows an emphasis on credit risk related studies, and to a lesser extent a focus on fraud detection and operational risk. It is also clear that neural networks, a combination of algorithms, and deep learning score high in the number of studies. For reference purposes we included a detailed overview of the referenced set of all 339 included studies in our mapping study devided over two tables Table 2.16 and Table 2.17, at the end of this paper. The X-axis states the ML topics, whereas the FinTech topics are depicted on the Y-axis.

Figure 2.7: Growth pattern of Artificial Neural Network Algorithms over time.

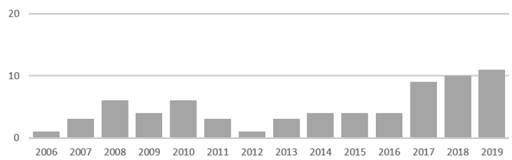


Figure 2.8: Growth pattern of Combinations of Algorithms over time.

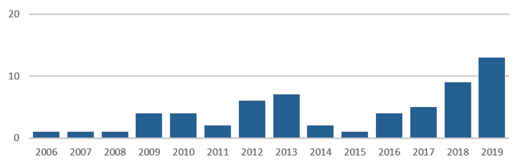


Figure 2.9: Growth pattern of Deep Learning Algorithms over time.

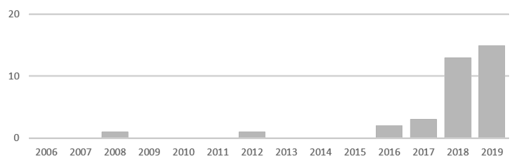


Figure 2.10: Growth pattern of Regression Algorithms over time.

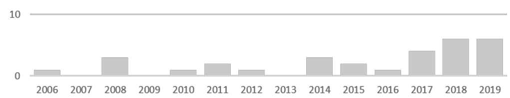


Figure 2.11: Growth pattern of Instance-Based Algorithms over time.

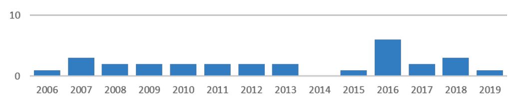


Figure 2.12: Growth pattern of Clustering Algorithms over time.

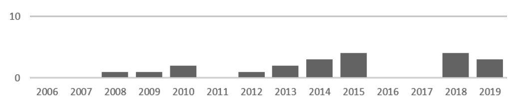
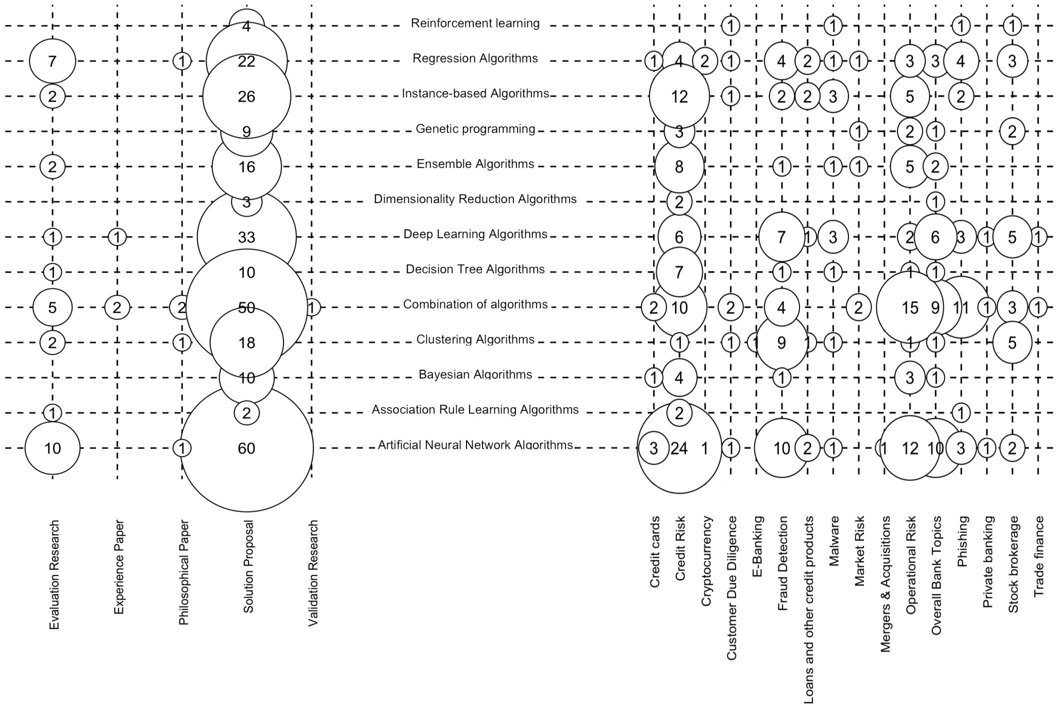


Figure 2.13: Bubble Chart of Study Results, after backward snowballing.



2.4 Discussion

In conclusion, we assess the results of our mapping study in relation to related work, we investigate the expected direction of future research and we describe the implications of our study for research, industry and education. In summary, our mapping study provides the following insights:

1. Regarding ML topics we found much focus on Artificial Neural Networks, Combinations of Algorithms, and Deep Learning, while we found only limited focus on Reinforcement Learning.

2. Regarding FinTech topics, we found that during the financial crisis (2008 - 2009) much focus was on Credit Risk (credit scoring). Since the crisis we observed a growing focus on security related topics, such as fraud detection, phishing, and malware detection. Only little research was found on Customer Due Diligence (Money laundering and terrorism financing), while these topics are of major importance to financial companies nowadays.

3. We did not find strong links with SE topics, indicating no emphasis in research on life cycle aspects of ML models, such as verification, testing, validation, maintenance, and model legacy.

4. The continents in in which ML research is carried out most are Asia (56%), Europe (26%), and North America (10%). Leading countries are China (17%), India (13%), and the USA (9%).

5. Finally, regarding research methods, almost all studies were Solution Proposals. Almost no case studies were identified. We argue that this observation relates to the lack of studies on life cycle aspects of ML models, that we mentioned above.

2.4.1 Implications

As discussed in section 2.4, we did not find strong evidence of research related to life cycle aspects of ML models. Besides that, we observed a strong preference with regard to the applied research approach in ML research for Proposal

Solutions, indicating a lack of case studies in a practical setting. Both observations can give the impression that ML research—at least within the FinTech domain—can learn from research in software engineering on the above mentioned topics, because within the SE research domain a large number of studies specifically focus on empirical studies themselves (e.g. [50, 49]), and on verification, testing, deployment, and in general the software engineering process itself (e.g. [41, 44, 102]).

A quick scan of related work in the field of ML in general, however, shows that there are indeed a number of studies in this field, such as Marijan et al. [61], Ma et al. [58][57], Wicker et al. [107], and Masuda et al. [63] on challenges of testing ML based systems. Furthermore, there is a study of Sculley et al. [88] on hidden technical debt in ML systems.

Finally, we found some recent and inspiring studies on the topic of continuous integration of ML models [82, 37, 81]. These studies indicate that especially large-scale companies such as Microsoft, Huawei, and Alibaba do recognize the need for more holistic studies in the field of ML.

Based on the observations in our mapping study, we observe that industry is well on the way to using ML on a large scale in its processes, and that recent steps are being taken towards DevOps and continuous integration of ML models. In the research world, however—and in particular within the FinTech domain—the focus is mainly on making models for solving a specific problem, and there is no targeted research into the large-scale use of AI within the business processes of organizations. That is why we believe that there is definitely room for future research in the area of life cycle aspects for ML.

Topics such as DevOps, continuous integration, pipeline automation play an important role in both SE and in ML, where it is not immediately clear in which of these research areas these topics should be put on the agenda. In fact, they do not form a clearly defined border area between SE and ML. A question that arises here is to what extent studies on such topics now belong to, an AI venue or journal, or to an SE conference or journal. A clear landing site seems to be missing at the moment. We think that there are great opportunities for cooperation between both research areas.

This also applies to the education of both SE and ML specialists. It is precisely in this area that universities can play an important role. We propose that in the portfolio of SE and AI courses particular attention is paid to aspects such as continuous integration, pipeline automation and life cycle aspects of ML.

2.4.2 Threats to Validity

For this research three main threats to validity were identified: 1) The identification of research, 2) the screening of papers, and 3) the data extraction.

Although a thorough search was done, it is impossible to guarantee to have captured all material in the area of ML in FinTech. However, the risk on missing relevant research in the identification phase was mitigated in four ways. Firstly, we've searched five databases recommended for use in systematic reviews according to [22]. Secondly, we added a major A* conference and a leading journal on the topic of AI to our mapping study. Thirdly—as recommended by [40]—we used a combined approach of database search and backward snowballing, in order to further minimize the risk of missing important studies within the scope of our mapping study. Finally, the risk of missing relevant information was reduced as much as possible by automatically gathering and analyzing the data from these sources and presenting the results in a publicly available spreadsheet [96].

The screening process for determining which papers should be used for analysis, was partially done by hand. This means human judgement was involved in the process. Because of this, a risk of bias exists which was reduced in three ways. At the start of the screening process the screening set was reduced automatically by applying a subset of the inclusion and exclusion criteria. Examples of inclusion and exclusion criteria which could automatically applied were, excluding duplicate articles, excluding books and excluding articles published before the year 2006. The remaining set of articles was screened by hand. However, to mitigate the risk of bias even further, articles for which it was unclear if they needed to be included in the set for analysis or not, were sent to the collaborating researcher for a second opinion. Finally, the analysis and mapping process was performed by two researchers.

For the data extraction on the remaining 339 articles, the article was read and for every article, the FinTech, ML, Research and SE topics were classified. Since this process depends on human judgement, for every facet a list of labels was constructed before the analysis phase started. Using these lists, the researchers were able to pick labels for every facet during the analysis phase. Furthermore, for every item on the list, it was clearly communicated what they meant, by writing down a definition for every label. After the first 10 articles, the results from both researchers were compared to check if both researchers understood the labels correctly.

2.5 Conclusions

We performed a structured mapping study on the topic of sub-symbolic AI, or ML in FinTech. For that purpose we used a combined approach of database search and backward snowballing. We assessed related work in five libraries with regard to ML topics, research topics, FinTech and banking topics, and the link with specific SE topics. The backward snowballing was performed on the top-15 studies ranked on cites per year, in order to enrich our subset for mapping analysis. In the end, we analyzed a subset of 339 studies, out of an initial set of 3,455 extracted papers.

Looking at our findings, three observations definitely stand out: 1) Life cycle aspects of ML models seem to be an undervalued topic in research. 2) We observe a strong preference with regard to the applied research approach in ML research for proposal solutions, indicating a lack of case studies in a practical setting. 3) Only limited research is performed on customer due diligence related topics, such as money laundering and terrorist financing, while this is a topic of fast growing importance in the FinTech industry.

We observe that industry anticipates attention for topics such as DevOps, continuous integration, and pipeline automation for ML models, and that the research community must make up for this. Life cycle aspects of ML models, case studies of applications of ML in practice, and Customer Due Diligence are important topics in industry, yet somewhat undervalued in research, leaving important opportunities for future studies in the field of ML and SE.

Table 2.16: Overview of Publications in Scope Part 1

Fintech Topic/ML Topic	Artificial Neural Network Algorithms	Association Rule Learning Algorithms	Bayesian Algorithms	Clustering Algorithms	Combination of algorithms	Decision Tree Algorithms	Deep Learning Algorithms
Credit cards	[177], [344], [430]		[385]		[369, 120]		
Credit Risk	[276, 226, 116, 199, 442, 294, 299, 250, 441, 137, 358, 241, 236, 195, 237, 334, 134, 152, 429, 209, 143, 243, 390, 374]	[274, 259]	[254, 127, 216, 320]	[411]	[286, 401, 449, 399, 249, 213, 165, 322, 121, 328]	[389, 304, 444, 176, 295, 163, 153]	[115, 333, 277, 412, 136, 232]
Cryptocurrency	[371]						
Customer Due Diligence	[235]			[285]	[190, 364]		
E-Banking				[395]			
Fraud Detection	[336, 184, 396, 330, 312, 139, 394, 435, 126, 192]		[166]	[238, 432, 189, 158, 398, 124, 264, 263, 130]	[415, 257, 135, 194]	[392]	[346, 114, 311, 434, 337, 154, 247]
Loans and other credit products	[223, 171]			[202]			[268]
Malware	[244]			[284]		[391]	[240, 382, 318]
Market Risk					[433, 261]		
Mergers and Acquisitions							
Operational Risk	[316, 317, 351, 155, 160, 365, 221, 145, 352, 181, 196, 361]		[278, 339, 366]	[349]	[265, 123, 440, 410, 193, 376, 321, 205, 159, 168, 353, 309, 297, 151, 378]	[179]	[355, 172]
Overall Bank Topics	[327, 255, 301, 448, 132, 142, 282, 373, 372, 271]		[305]	[292]	[356, 370, 354, 119, 183, 217, 147, 436, 211]	[206]	[293, 319, 287, 363, 386, 296]
Phishing	[141, 380, 409]	[117]			[253, 224, 239, 233, 125, 208, 439, 310, 270, 187, 331]		[303, 262, 167]
Private banking	[383]				[258]		[175]
Stock brokerage	[402, 426]			[423, 359, 186, 248, 162]	[416, 404, 164]		[191, 169, 218, 214, 400]
Trade finance					[335]		[225]
NA							

The classification topics on the Y-axis of this matrix table are FinTech-related topics derived from the classification scheme of papers in scope of this study. On the X-axis there are topics from the classification scheme related to ML topics. References in the matrix table are included in the Analysis References part at the end of this study.

Table 2.17: Overview of Publications in Scope Part 2

Fintech Topic/ML Topic	Dimensionality Reduction Algorithms	Ensemble Algorithms	Genetic programming	Instance-based Algorithms	Regression Algorithms	Reinforcement learning	NA
Credit cards					[275]		
Credit Risk	[314, 280]	[122, 288, 178, 421, 242, 267, 201, 379]	[230, 273, 323]	[445, 227, 272, 300, 428, 425, 200, 447, 446, 229, 325, 413]	[170, 298, 377, 348]		[414, 283, 384]
Cryptocurrency					[350, 408]		[424, 269]
Customer Due Diligence				[393]	[156]	[180]	[338]
E-Banking							[118]
Fraud Detection		[307]		[222, 231]	[332, 219, 403, 443]		[340, 308]
Loans and other credit products				[131, 173]	[342, 381]		[343, 266]
Malware		[329]		[302, 360, 228]	[138]	[185]	
Market Risk		[246]	[345]		[140]		
Mergers and Acquisitions							
Operational Risk		[357, 161, 422, 133, 203]	[407, 215]	[419, 437, 207, 157, 210]	[182, 420, 188]		[260, 204, 418, 279, 427, 251, 148, 197, 290]
Overall Bank Topics	[368]	[291, 417]	[281]		[289, 150, 397]		[144, 174]
Phishing				[146, 306]	[245, 149, 212, 347]	[405]	[438, 431, 375, 341, 256, 252, 128, 129, 234, 220]
Private banking							[387]
Stock brokerage			[315, 313]		[324, 406, 326]	[388]	
Trade finance							
NA				[367]	[198]		[362]

The classification topics on the Y-axis of this matrix table are FinTech-related topics derived from the classification scheme of papers in scope of this study. On the X-axis there are topics from the classification scheme related to ML topics. References in the matrix table are included in the Analysis References part at the end of this study.

Chapter 3

Automating Literature Reviews: Motivation and State of the Art

To be able to create a useful tool which supports the automation of literature reviews, first the state of the art regarding automating literature reviews needs to be identified. Therefore, we performed a short review of literature, regarding the automation of systematic literature reviews, which is detailed in this chapter. To create a structured overview, we analyzed the gathered literature as follows. First, the process of performing a systematic literature review was extracted. This resulted in 14 individual steps which need to be carried out in order to perform a successful literature review. After the extraction of individual steps, for each step the automation solutions were identified from previous literature. This chapter first details the motivation for creating a tool addressing literature review automation in Section 3.1. Afterwards, Section 3.2 describes the identified process for performing literature reviews. This is followed by a brief description per step including the identified automation solutions in Section 3.3. Chapter 4 will describe the solution design of the steps which our tool addresses. By combining the literature review steps and the identified research on automation techniques, it is possible to map the automation techniques to different steps of the systematic literature review process, resulting in a complete overview of automation potential.

3.1 Motivation

During the execution of the study ”Machine Learning in FinTech: A Structured Mapping Study”, described in Chapter 2, we found that a significant amount of time was spent on repetitive work which potentially could have been automated. In order to already try and omit some of the repetitive work, proof of concepts for automating parts of our study, such as retrieving articles from certain scholarly resource, were developed. However, these tools were not provided with a User Interface (UI), which meant that specific software development knowledge was needed in order to use them. These tools, along with the frustrating experience about time spent on repetitive work, inspired us to investigate the possibilities of creating a tool for automating the process of executing systematic literature reviews. In order to create such a tool, first the state of the art with regard to previously found automation solutions needs to be identified. The identified solutions are detailed in the next sections of this chapter.

3.2 Literature Pipeline

To learn about what has been done trying to automate systematic literature reviews, we offer a complete overview of the literature review process and the steps it contains. This Section describes the process for executing a systematic literature review by providing a high level overview. We provide a detailed description per step and the automation techniques which have been proposed by previous literature in Section 3.3.

3.2.1 Pipeline overview

Using the information provided in [101], [50], [113], [79] we created an overview of the different steps taken in the systematic literature review process. This overview is shown in Figure 3.1. As can be drawn from the Figure, the review process consists of 14 individual steps. Note that these steps and the order in which they are executed might differ per systematic literature review, as not every step is mandatory in order to complete a successful systematic review. However, the current process overview should provide a super set of steps which can be taken during a systematic literature review. The individual tasks can be classified in the following five phases, highlighted in the rightmost column: Preparation, Retrieval, Screening, Synthesis and Write up.

The preparation phase serves to create a solid base on which the literature review will be performed. This is done to ensure that duplicated work is omitted and that the review is actually useful and necessary. The retrieval, screening and synthesis phases, contain the process in which the literature is retrieved, filtered and analyzed. This includes identifying what has been done and what the outcomes of those studies were. These three phases will provide results which finally will be written down in the write up phase. Using these five phases, it is possible to think of the systematic review process as an algorithm where researchers define input parameters during the preparation phase, which afterwards are executed in the retrieval, screening and synthesis phase to provide an outcome for the final manual step which is the write up phase.

3.3 Automation solutions

In order to create a useful tool which supports the automation of literature reviews, besides the identification of the individual steps, we created an overview of previous research on automation techniques for each individual step. First, existing tools regarding systematic literature reviews will be discussed. Afterwards, this Section details the individual steps displayed in Figure 3.1 and elaborates per step, the automation techniques presented by previous literature. This way, an overview is created of potential automation techniques per systematic literature review step.

Figure 3.1: Systematic Literature Reviews - A process overview based on the description of [101].

Task	Description	Phase
1. Formulate review question(s)	Decide on the research question of the SLR	Preparation
2. Find previous SLR(s)	Search for previous SLR(s) which try to provide the answer to the same question	Preparation
3. Write the SLR protocol	Provide an objective, reproducible, sound methodology for peer review	Write-up
4. Devise search strategy	Decide on database and keywords to find all relevant research	Preparation
5. Search	Aim to find all relevant literature even if many irrelevant ones included	Retrieval
6. Remove duplicates	Remove identical literature	Retrieval
7. Screen abstracts	Remove irrelevant literature based on title and abstract	Screening
8. Obtain full text	Gather full-text versions of remaining articles	Retrieval
9. Screen full text	Exclude irrelevant literature by screening the full text	Screening
10. Snowball	Follow citations from included literature to find additional literature	Retrieval
11. Extract data	Identify literature features, methods and outcomes	Synthesis
12. Synthesize data	Convert extracted data to common representation(e.g. average or standard deviation)	Synthesis
13. Re-check literature	Repeat the search to find new literature published since the initial search	Retrieval
14. Write up review	Write down and publish the final report	Write-up

3.3.1 Existing tools

Before creating a tool, we identified an existing tool called Publish or Perish [74]. This tool also addresses systematic literature reviews, where it was found that the main focus of this tool is to identify literature and provide information with regard to indexing and citations. However, we envisage a tool which not only retrieves information, but also address additional steps of systematic literature review, such as screening and snowballing.

3.3.2 Formulate review question(s)

The first step of a systematic literature review is to formulate one or more research questions. The aim of the review, then, is to provide an answer to all of these formulated questions. Typically, the research questions in literature studies are general as they aim to discover research trends (e.g. publication trends over time, topics covered in the literature) [79]. To guide the formulation of these research questions, the population, intervention, control and outcome (PICO) are recommended elements to be used as suggested by [49]. The PICO process was originally introduced as medical guidelines for considering the effectiveness of a treatment. However, these rules can also be applied within the software engineering research domain. Here, the population can for example be software engineers, testers, but might also reference to an industry group. The intervention is the tool, technology methodology or procedure that addresses a specific issue. Then the comparison is the tool, technology methodology or procedure with which the intervention is being compared. Using these first three elements, the outcome should relate to any factor of importance for practitioners such as time reduction, reduced production cost or improved reliability.

The formulation of these research questions can be seen as a creative process and serves as the preparation or starting point of the systematic review. Therefore, no real automation tools exist. However, it might be possible to create decision support tools which help researchers define or choose between research questions. It might also be possible to detect duplicate questions from previous systematic literature reviews using evidence gap maps [91]. Furthermore, it might be possible to detect missing info or ambiguity in research questions using support tools. However, we did not find literature addressing this topic.

3.3.3 Find previous SLR(s)

Finding previous systematic literature reviews which try and answer similar or identical research questions can save a significant amount of time, since reviews, according to [101], take from 12 to 24 months for a person to complete. Therefore, if a reviewer had an automatic way of identifying previous systematic reviews and ensuring no previous reviews are missed, a considerable amount of time could be saved.

To automate this step, a system should be able to translate the research questions from the previous step 3.3.2 into keywords which find previous systematic literature reviews. In order to find an overview of previous literature regarding these keywords, a register of previous systematic reviews should be created. This is addressed for medical literature by [11]. However, we did not find such register regarding computer literature.

3.3.4 Write the SLR protocol

Once the previous step has established that no previous reviews addressing the same research question exists, and that a review on the given topic is needed, a systematic literature review protocol should be created. This protocol is the first step of actually performing the review and dictates how the review is going to be performed. This protocol details which and how every individual step in Figure 3.1 is going to be executed. Currently there is no defined method on reviewing the consistency of the review protocol. Therefore, the protocol should be peer reviewed to verify the consistency and integrity of the protocol.

Since there is no defined review method for the systematic literature review protocol, there are no real ways of automating this step. However, there are some universities and organizations which provide templates which can guide writing such a protocol [103].

3.3.5 Devise search strategy

Before the literature of a review can be retrieved, a researcher needs to define a search strategy. This search strategy ensures that a review is not biased, because some literature is more accessible than other literature. For example, some databases prioritise their literature on the number of citations, however that does not mean literature with less cite might contain useful information. This task consists of two sub steps: define a search string, commonly defined as a search query and describing the databases which will be searched.

The PICO principle, explained in Section 3.3.2, can address the automatic search strategy creation. This would involve translating the formulated review question(s), from the previous step, into a search query. Currently no tools to support this process are identified. However, automation potential can also be found in finding the right keywords regarding the topic of interest. To automate finding the right keywords, decision support tools have been proposed [60].

3.3.6 Search

Systematic literature reviews are executed in a systematic way, in order to reduce the risk of missing relevant literature. This means the search should aim for the highest recall, even if that means retrieving irrelevant literature as well. This implies that multiple databases need to be searched. For our research, the main focus is about literature related to computer science. Therefore, based on the experience reported by Dyba et al. [22], the following databases should be most relevant for the computer science field: IEEE [75], ACM [55], Google Scholar [30], Springer [97], Science Direct [20] and Web Of Science [87].

By automating this process the researchers performing the review do not have to gather all the articles by hand. This can save a considerable amount of work. It is possible to automate this process by scraping information from websites which host scientific literature databases, or by using APIs such as [76] which provide information about literature and allow users to input search queries.

3.3.7 Remove duplicates

When literature is gathered from multiple databases, it is possible that duplicates occur, since some articles may occur in multiple databases or may occur multiple times in different guises in the same database. Because of that, before the abstract screening step 3.3.8 is carried out, the workload can be reduced by removing duplicates. This way screening the same articles multiple times is avoided. However, removing duplicates is a time consuming task and challenges exist for automating this process since different databases might have variations in metadata for identical articles. For example, there might be differences in the title of an article, where different databases use different formats. The DOI or ISBN might not always be present and page numbers might differ since these are specific for the journal in which they are published.

Current solutions for automating duplication removal focus on removing duplicates using citations [23]. Most citation managers already have automatic ways of searching for duplicate records. For example, Mendeley[1], End-Note[2] and ProCite[3] have implemented forms of semi automated duplication removal using citations.

3.3.8 Screen abstract

As stated in the search step 3.3.6 the retrieval should aim for the highest recall possible. This means it is likely to find a lot of irrelevant articles. Because of that, the abstract screening step is needed to exclude the irrelevant literature which in turn reduces the workload for the next steps. During the execution of this step, usually the majority of articles which were initially found, will be discarded. The inclusion and exclusion of literature is based on the inclusion and exclusion criteria which are defined in the search strategy step 3.3.5.

The initial set of articles can contain more than a thousand individual articles. Therefore, this step may be very time consuming and error-prone, since researchers have to read the title and abstract for every article and make the inclusion/exclusion decision based on that. Therefore, it is recommended to include a second reviewer for the screening phase to make sure all articles are filtered correctly and reduce the risk of bias. However, this increases the amount of manual labour even further.

To address the automation of this step, different studies have been identified and can be categorized in three different types of (semi)automating the abstract screening step. The first way of trying to automate the screening phase is by using different types of text analysis methods and use those to partially or fully automate the selection process [90],[28],[33],[2],[48]. Other research suggests ways to semi-automate the screening of abstracts. These methods also make use of text analysis methods, but instead of using these to include or exclude articles, they are used as a decision support method to support the screening of abstracts [25],[24],[59],[47], [106],[4],[100]. Other solutions also propose decision support, but use machine learning techniques to infer exclusion and inclusion rules by observing a human screen-er [65],[70],[105],[64],[8],[17].

3.3.9 Obtain full text

Obtaining the full text is mostly manual work since PDF files for all relevant articles need to be gathered. However, since the previous step is expected to greatly reduce the total number of articles, this step is not as resource demanding as the previous one.

Most databases provide URLs to their PDFs which can be used to automatically download them. However, at the moment of writing a system which fully automates the retrieval

[1]Mendeley is a free reference manager and an academic social network tool.

[2]EndNote is a commercial reference management software package, used to manage bibliographies and references when writing essays and articles.

[3]ProCite is a commercial reference management software program.

of PDF files for articles has not been found, which might be due to the fact that articles are stored across multiple databases.

3.3.10 Screen full text

During the screening of full texts, the process of step 3.3.8 is repeated. However, instead of screening based on only the title and abstract, the full text of articles is used, which was retrieved in the previous step. As well as for obtaining the full text, only the remaining relevant articles are screened. Therefore, we found from prior experience, that since the set of articles is significantly reduced, this step usually takes less time than the screening of abstracts.

Since full text provides more information than only the abstract and title, the automation methods proposed for the abstract screening step in Section 3.3.8 could also be applied to screening the full text. In addition, automatic text summary solutions, using Natural Language Processing (NLP), could be used in order to provide summaries of literature which needs to be screened [34].

3.3.11 Snowball

Even when a sound search strategy is defined in step 3.3.5 and the search is performed thoroughly on multiple databases, it is possible that relevant literature might be missed. This can be addressed by applying snowballing. This is done by recursively pursuing relevant references from the set of articles resulting after the full text screening described in Section 3.3.10. Even though this can be very time consuming, it is a recommended step to apply, since it reduces the chance to miss any relevant literature. After the snowballing step has been completed, the newly identified literature should go through the literature review steps again, starting from duplicate removal step described in Section 3.3.7.

To be able to automate snowballing, the citation links of articles should be accessible. There are databases such as Google Scholar and Web of Science which provide this kind of information. We found Studies addressing the automation of the snowballing step, using these information source [16].

3.3.12 Extract data

The data extraction step serves to extract information in literature such as methods, outcomes or other results stated in the literature. However, this is not trivial as some articles might present relevant outcomes in the form of graphs or Figures whereas others present them in text or tables. The data extraction step is mostly performed by two researchers simultaneously such that any disagreements can be resolved and the literature review results are not biased.

To automate this process, the data extraction step can be divided into two sub steps. The first sub step is to reduce the text to process, since not all text in an article contains relevant information about the results and outcomes. The text reduction can be accomplished by algorithms which highlight sections containing information about results [48]. The second sub step is to extract the relevant information of outcomes and results for which an automation solution is proposed by [43]. Other solutions suggest the usage of context aware keyword extraction [68].

3.3.13 Synthesize data

After the results and outcomes are extracted, the data needs to by synthesised, the data needs to be converted in such a way it can be compared to all results from different articles. In cases where results are in the form of numbers this could be done using statistical methods such as continuous distribution, whereas cases of results are based on finding certain topics, the different results can be synthesized using a list of keywords.

An overview for different machine learning techniques proposed to address the automation of data synthesis in systematic literature reviews is given by [62]. This research states that the use of machine learning for addressing this step is maturing. However, it also states that systematic reviews require very high accuracy in their methods, which implies that it may be difficult to fully automate this process.

Table 3.1: Automation solutions per systematic literature review step.

Systematic review step	**Literature**
Formulate Review Questions	[91]
Find Previous SLR(s)	[11]
Device Search Strategy	[60]
Search	[76]
Remove duplicates	[23]
Screen Abstract	[90],[28],[33], [2], [48][25],[24],[59], [47], [106],[4],[100] [65],[70],[105],[64],[8], [17]
Obtain Full Text	-
Screen Full Text	Same as abstract screening
Extract Data	[48],[43],[68]
Synthesize Data	[62]
Re-Check Literature	-
Write up review	-

3.3.14 Re-check literature

Since the overall process of performing the systematic literature review may take up to 12 to 24 months [101] for a person to complete, it is good practice to repeat the search when all initial data is processed to identify new literature which has emerged. This means repeating the process starting at the search step again, which is described in Section 3.3.6.

This step can therefore use all the automation steps from the previous steps described above. To be able to identify new literature in the re-check step, the duplicate removal described in step 3.3.7 can be used.

3.3.15 Write up review

The final step of the systematic review is to write up the results as a scientific paper. Afterwards this paper is submitted to one of more journals and will be submitted to review before it will be published. Although there are templates for writing the systematic review, at the moment of writing, there are no ways identified to automate this process.

3.4 Conclusion

By identifying the state of the art regarding automating systematic literature reviews, we identified a process overview and automation solutions per step. The obtained process overview is shown in Figure 3.1 and an overview of literature which addresses automation solutions for each individual step is shown in Table 3.1. As can be seen from the automation overview, the main focus of research has been on the screening phase. For the screening phase 18 articles were identified proposing automated or semi-automated solutions. For the synthesis phase 3 articles were identified. For every step of the retrieval phase at least one article was found. The focus of research on the screening phase might be explained due to the fact that this is a very time consuming phase and consists of repeating manual labour, whereas the retrieval phase is relatively simple and the synthesis phase is not. This chapter provides current solutions, which can be used to create a tool which addresses the automation of the complete process of performing systematic literature reviews.

Chapter 4

Automating Literature Reviews: Tool Design

In chapter 3 the individual steps of a systematic literature review and possible automation solutions were identified from previous literature. This chapter details the design of a tool, called LitAutomation, addressing the automation of two phases of the systematic literature review, namely the retrieval and screening phase. These two phases consist of six individual systematic literature review steps. This chapter details the solution design for each of these steps based on solutions found in the previous chapter. Due to the limited amount of time available for this work, these six were marked as having the highest potential for automation and having significant impact on reducing the total amount of time needed to carry out a systematic literature reviews. The steps addressed in the tool are: searching articles, duplication removal, abstract screening, obtaining full texts, full text screening and snowballing. An overview of these steps is depicted in Figure 4.1.

Figure 4.1: Systematic Literature Review Automation Tool - Focus overview.

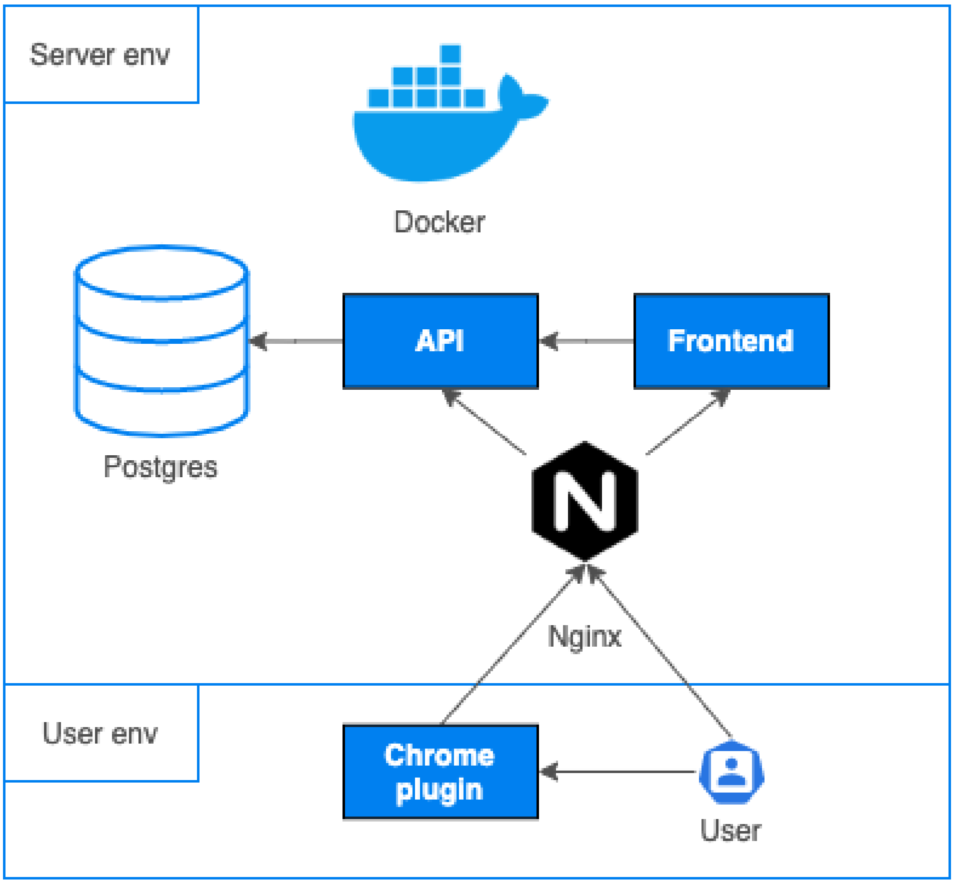
Task	Description	Phase
5. Search	Aim to find all relevant literature even if many irrelevant ones included	Retrieval
6. Remove duplicates	Remove identical literature	Retrieval
7. Screen abstracts	Remove irrelevant literature based on title and abstract	Screening
8. Obtain full text	Gather full-text versions of remaining articles	Retrieval
9. Screen full text	Exclude irrelevant literature by screening the full text	Screening
10. Snowball	Follow citations from included literature to find additional literature	Retrieval

This chapter starts by providing the overall system architecture overview and explaining the reasoning behind it in Section 4.1. This section is followed by detailing which solutions found in previous literature were chosen for each individual step. The next chapter will provide an in depth description of implementation details for each of these solutions.

4.1 System architecture

We used three main requirements for the tool design. First of all, the tool should be easily accessible for multiple users. This is important when multiple researchers are working on the same systematic literature review, such that they are able to work on the review at the same time. Secondly, the tool should be platform independent, which means that it can run on any server. It was also taken into account that the tool could be deployed on any cloud service running on Kubernetes for example. The last requirement was that the tool needs to be able to search literature from six different scientific databases as defined in Section 3.3.6.

Figure 4.2: Tool architecture overview.



These requirements resulted in the decision to built the tool as a web application. Therefore, a user is able to perform all actions using the browser and access the tool anywhere. The system architecture for the tool is depicted in Figure 4.2, which shows two environments.

The first environment is the server environment, which exists of a database, a web server and two services. The web server is an Nginx server used as reverse proxy to route incoming traffic to the services and configure SSL certificates using Let's Encrypt. The first service is the application programming interface (API) which handles the application logic and the communication with the Postgres database to store data. This API is implemented as a RESTful API implemented in Golang and will further be reffered to as the backend. The second service is the frontend, which serves as the main user interface in order for the user to execute different steps of the systematic literature review processes. The frontend is built using LitElement, such that the frontend can be built consisting of WebComponents written in TypeScript. Furthermore, docker is depicted in the server environment. This is done, to indicate that the database, API and frontend are running as docker containers on the server. This way, the tool becomes platform independent, since all individual services run in their own docker environment.

The second environment is the user environment. This environment consists of a Google Chrome plugin and a user. The use of the Google Chrome plugin requires a Google Chrome browser. The Chrome plugin is needed in order to perform the search and snowball step which will be further explained in sections 4.2 and 4.7.

4.2 Search

As stated in Section 3.3.6 the search step needs to retrieve all articles from IEEE, ACM, Google Scholar, Springer, Science Direct and Web Of Science. The easiest way to perform these searches would be by using an API, however according to [69] Google Scholar, ACM and Web Of Science do not have an API available at the moment of writing. Furthermore, for the API's of IEEE, Springer and Science direct, different types of restrictions exist such as a maximum number of articles which may be retrieved every single day. Therefore, we decided to automate the search step by using web scraping. However, this resulted in two main problems. The first problem is that website use Javascript for rendering the DOM. This means that simple http GET requests are not sufficient, and that instead some form of rendering needs to be done for full retrieval of websites. Secondly, Google Scholar has implemented a CAPTCHA which cannot be completed automatically. Therefore, it was chosen to perform the scraping using a Google Chrome plugin which is able to render DOM and halts whenever it encounters a CAPTCHA. When the user completes the CAPTCHA, the scraping process is resumed by the plugin. This way the tool is able to retrieve articles from all different databases.

4.3 Remove duplicates

Article duplication removal is handled in the backend. The de-duplication is first executed on Digital Object Identifiers(DOIs). The DOI system is a not-for-profit organization [27] which manages unique identifiers for digital objects such as scientific articles. However, not every scientific article is assigned a DOI. Therefore, after duplication removal on DOI's, duplication's are identified based on titles.

4.4 Screen abstracts

As stated in 3.3.8 three types of automation and semi-automation solutions for screening abstracts were identified: decision-support, decision-support based on machine learning and full automation. To reduce the workload of a researcher as much as possible, we used the most cited solution from the full automation category [2]. This paper combines text classification with different machine learning techniques where it concludes that Naive Bayes offered the lowest rate of mistakes in the form of False Negatives (FN) with a corresponding F1 score of 0.75. Since it is very important for literature reviews to not lose potentially relevant data, the FN rate of any classifier used to automate screening should aim to be as low as possible. Therefore, the design of the abstract screening automation is based on the screening automation process described in this work using Naive Bayes.

The described process consists of three main phases as depicted in Figure 4.3. First, documents are pre-processed. This means words in documents are labeled using Natural Language Processing (NLP). Words of insignificant meaning are removed and the remaining words are Porter stemmed.

Secondly, documents are modeled. Term Frequency-Inverse Document Frequency (TFIDF) is used, in order to perform feature extraction. This means that for an article features are extracted based on the TFIDF score of the given title and abstract. Data presented in [2] has shown that by only selecting 11 features an average F1 score of 0.75 can be achieved.

After the feature extraction has been completed, the Naive Bayes model can either be trained by providing the modeled document and the classification, or it can be used to let the model predict which class the article belongs to. For screening, documents can be classified into two categories. A document either belongs to the included set of articles or the document belongs to the excluded set of articles. When at least one included article and one excluded article are provided to the model, it is able to predict if an unseen document belongs to either the included or excluded set.

Figure 4.3: Abstract screening design.

4.5 Obtain full text

For obtaining the full text, no real automation solution is provided, because as described in Section 3.3.9 none was found at the moment of writing. Obtained articles are stored across a variety of websites, which makes it not feasible to build an automated way of retrieving all full texts. Instead, an option is provided for researchers to add the full text to articles themselves using the frontend.

4.6 Screen full text

As stated in Section 3.3.8 the screening of full texts is the replication of the abstract screening step. However, instead of using article and title only, instead the full text of an article is used. Therefore, it is possible to simply reuse the same solution as defined in Section 4.4.

4.7 Snowball

We found in Section 3.3.11 that Google Scholar indexes article references. Furthermore, Google Scholar provides the benefit of indexing articles from every website which allows indexing. Although Google Scholar does not provide an API to gather this information, it is possible to scrape the information. Therefore, to automate the snowballing step, the Chrome Plugin is used to scrape information from Google Scholar.

Chapter 5

LitAutomation: the Implementation Details

This chapter describes the implementation details of the tool called LitAutomation, for addressing the solutions stated in Chapter 4. As explained, the tool consists of three main components: the backend, frontend and the Chrome plugin. These main components will be used interchangeably in this chapter, to explain different implementation details. In order to provide a sound illustration on how the tool works, this chapter first describes the general implementation details of the tool in Section 5.1. Afterwards, for every systematic literature review step which is addressed by the tool, a section details how the solutions were implemented. This chapter concludes by detailing additional features which are added to increase the usability of the tool.

5.1 General implementation

As detailed in Section 4.1, the tool is implemented as a Web Application. The source code is published as Github organization [93] which is divided into three repositories and distributed under MIT[1]. Every repository contains a README with a detailed explanation on how to build and run for local development. The backend and frontend contain docker files for building and running as well. This enables the frontend and backend to be deployed without installing any additional software.

Because the tool is implemented as a Web Application, a user can access it on any computer using a browser. Before a user can start using the tool however, an account needs to be created. This is required, in order to make sure users can only access their own literature review projects. The account can be created through the frontend by registering with an email and a password as shown in Figure A.4.

After a user has registered an account, before using either the frontend or the plugin, the user needs to sign shown in figures A.5 and A.1.

Since the steps of the systematic literature review are sequential, the tool needs to be able to determine which articles to use for a given step. Therefore, six different article

[1] https://github.com/lit-automation/backend/blob/master/LICENSE

statuses are distinguished, which are detailed below.

- *Unproccesed*: Indicates that an article was just added to the tool an has not been processed yet.
- *Duplicate*: Is assigned by the automatic duplication removal step, indicating that an article has been marked as a duplication of another article in the set of a given project. In case duplication's are found, excessive articles will be marked as duplicate.
- *Excluded*: Indicates that the given article is not relevant for the literature review.
- *Included by Abstract*: Indicates that an article has not been excluded after screening on abstract and title.
- *Included*: Included means that an article passed all screening steps of the systematic literature review and thus, should be used.
- *Unknown*: Can be assigned by a researcher to indicate it is unsure if an article has to be included or not.

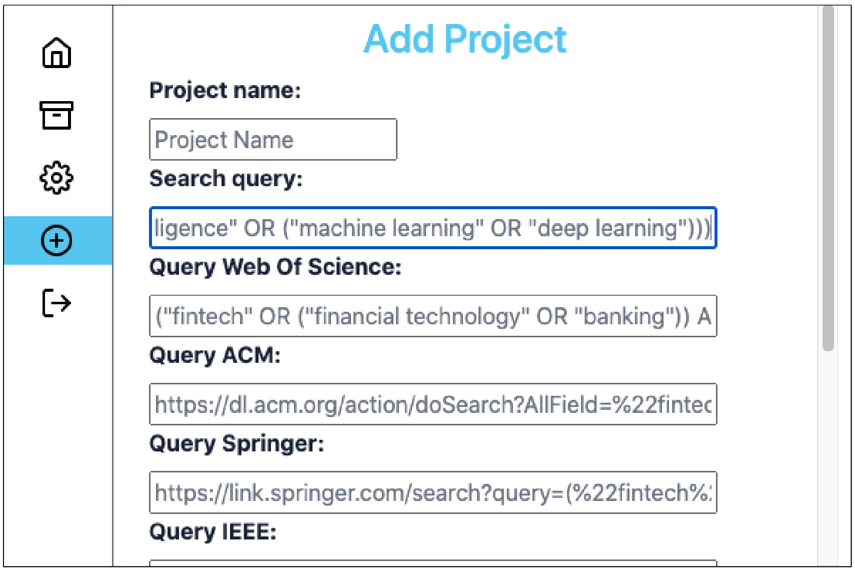
5.2 Search

After an account has been created and the user has signed in, retrieving the articles during a full search of a literature review, can be done using the Chrome Plugin. The Chrome Plugin consists of two core components.

The first component works in the browser environment. This component is able to read the DOM and control the browser. The second component is the popup. The popup contains the user interface of the Chrome Plugin and is the part which allows users to manage actions for literature reviews. After the plugin is installed, a button is added to the Chrome plugin section of the browser. A user can open the popup by clicking this button. The two components are running in separate environments in the browser and cannot access each other's DOM or local storage directly. Instead they communicate through messages. Here, the popup component sends instructions to the browser component, which in turn sends back messages about performed actions.

The first time a user signs in, the ”Home” screen of the plugin is shown 5.2. This screen shows a dummy project, to illustrate how a systematic literature review projects look like in the tool. A user can create a new project using the ”Add Project” screen 5.1. A project requires both a project name, which is used as reference for the literature review project and a search query.

Figure 5.1: Chrome plugin - Add project.

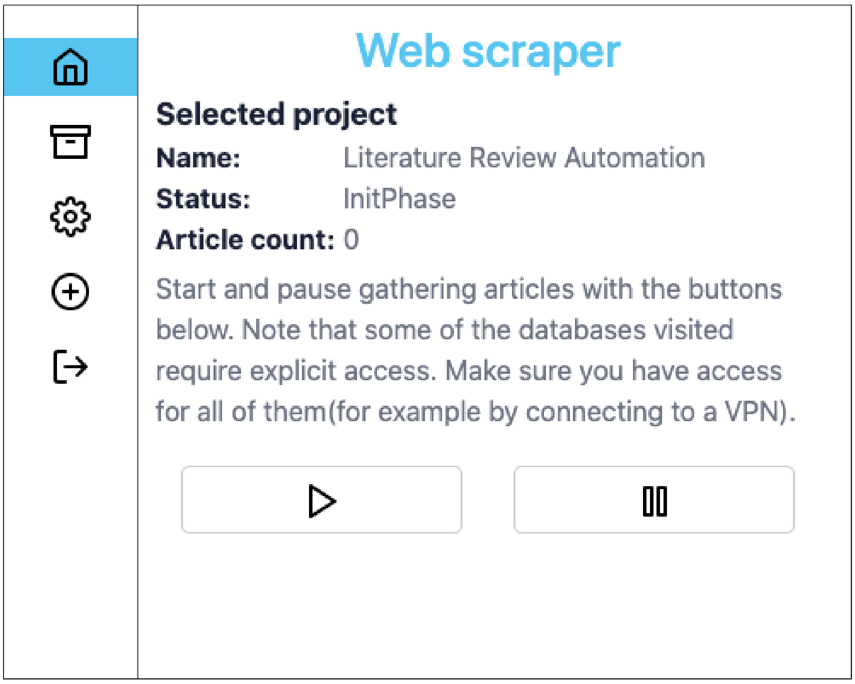


A custom parser is written to parse the search query defined by the user. This parser is needed in order to translate the query to the correct format for search queries of the digital libraries IEEE, ACM, Google Scholar, Springer, Science Direct and Web Of Science. The query language consists of five terms: opening brackets, closing brackets, the AND keyword, the OR keyword and search terms. In addition, these search terms can be concatenated by quotation marks. The parser checks for a correct number of opening and closing brackets and an even number of quotation marks.

After the parser has verified that the input query is valid, it outputs the scraping URL results in the edit fields of IEEE, ACM, Google Scholar, Springer, Science Direct and Web Of Science respectively. In these edit fields, a user can adjust the URLs in case some additional filters need to be set, such as a minimum year or publication type. In case a user wants to skip one of the digital libraries, the URL of the library can be left empty and the plugin will skip the library.

After the user has added the project, it is automatically set to be the current project, i.e. the project which is currently used. In case a user needs another project, the current project can be switched by selecting another project in the ”Project List” screen A.3. The user also has the ability to edit the search query and project name in the ”Edit Project” screen A.2. However, editing a project’s search query is only possible as long as a user has not yet started gathering articles for a given project.

Figure 5.2: Chrome plugin - Scrape articles.



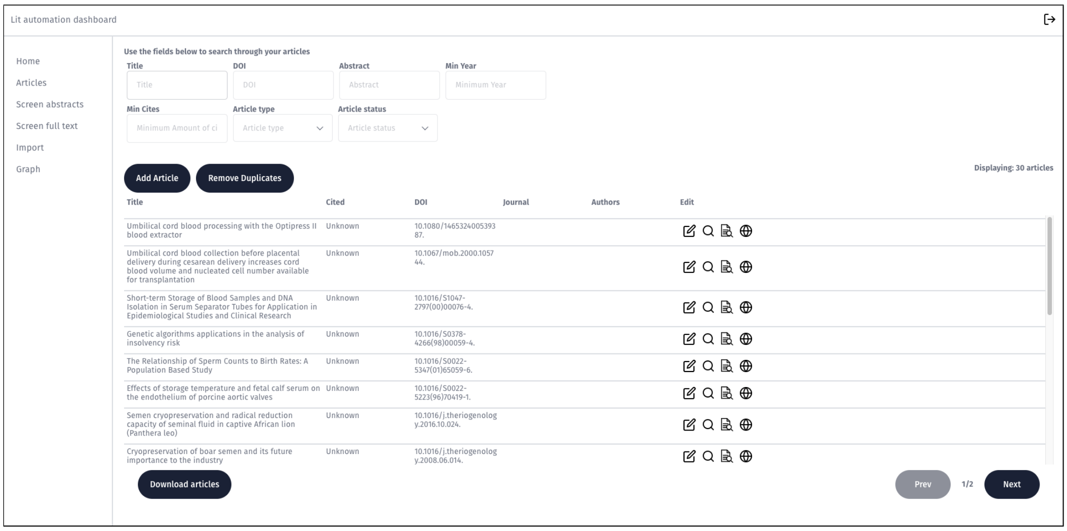
Finally, when the desired project has been selected and the search query has been entered correctly, the user can start scraping from the home screen by pressing the play button. Once the play button has been presses, the browser component of the plugin will start controlling the browser. For every digital library the popup component sends a URL containing the formatted search query to the browser component. The browser component then ”walks” through the assigned page. For every article on the page, all available information is gathered. For every article the plugin tries to find the authors, year, journal, publisher, title, URL, literature type, language, abstract, DOI, number of citations and the search result number. Once the information has been gathered it is sent to the backend along with an identifier for the current platform. The backend then stores the article information for the current project in the database.

When article information is scraped, entries on the website of a digital library may lack certain information. Therefore, when the backend receives a new article, it uses the CrossRef API [76] in order to add as much missing information as possible. In addition, the backend adds BibTex to the articles, to be able to reference found literature in a Latex document and adds an ”unprocessed” status field to the article which is used during the next steps of the literature review. While scraping articles, the user needs to keep the popup component of the plugin open in order for the popup component to communicate with the browser component. The user can also pause the plugin by pressing the ”Pause” button. When gathering literature from Google Scholar, a CAPTCHA might appear. The plugin will wait until the user completes the CAPTCHA, before continuing the scraping process.

5.3 Remove duplicates

After the scraping process has been completed, the backend stores all identified literature in the database. The user can access the list literature for a project using the frontend. After signing in the ”Home” page is shown. The ”Home” page provides an overview of all project of a user, which is shown in Figure A.6. By default, the last created project is selected for use. However, a user can switch projects using the ”Use” buttons. When the desired project is selected, the user can inspect the articles at the ”Article overview” page shown in Figure 5.3. The ”Article overview” page is the central page for executing different steps of the screening phase of the systematic literature review and is also used for managing the article set. This includes filtering, editing and browsing through articles, screening articles and downloading the complete set. These functionalities are further discussed in sections 5.4 - 5.8.

Figure 5.3: Tool - Article overview.



Duplicates can be removed by simply clicking the ”Remove Duplicates” button on the ”Article overview” page. As described in Section 5.1, every article is assigned a status. The backend retrieves all articles from the database, which are not marked with status ”duplicate” based on DOI. After these duplicates are removed, the backend repeats this process, but instead using the title to identify duplicates. The backend sends back a response to the frontend indicating the number of duplicates it has removed. The user is now able to check-out the articles indicated as duplicates, by adjusting the search filter at the top of the page to ”Article Status Duplicate”.

5.4 Screen abstracts

After duplication removal, the user is able to start screening on abstract and title using the frontend. The screening consists of three phases executed on two different pages. First the model needs initialization, which is detailed in Subsection 5.4.1. After the model is initialized, the model needs to be trained using active learning as explained in Subsection 5.4.2. Finally, if the training is done, the user can automatically screen the remaining articles as detailed in Subsection 5.4.3.

5.4.1 Initialization phase

As described in Section 4.4, the automatic screening is implemented as a Naive Bayes classifier. Therefore, the model needs to be trained with data from all different classes which the model needs to identify. In the case of screening on abstract and title, we call these classes ”included” and ”excluded”. This means that the model needs to be trained with at least one article which should be included and one article that should be excluded. Our screening evaluation, detailed in Section 6.3, has shown that the overall screening outcome is dependent on the initialization. Therefore, a detailed explanation is added to the frontend detailing how the model should be initialized. For the initialization phase, it states that a user should initialize the model by training more included articles in order to reduce recall.

The initial training can be performed at the ”Article overview” page, by pressing the magnifying glass on an article row. This will open a popup containing the title, abstract and a list of separate sentences for the article that was clicked. The model will already indicate if it thinks an article should be included or excluded and details this information by either showing the text in green color, which means an article should be included, or showing the text in red, which indicates a text should be excluded. An example of both situations is shown in figures A.10 and A.9. As can be seen, the tool also provides a confidence score. This score indicates the certainty of the screening model for assigning the include or exclude class to the given article. Using the information in the popup, the user can train the model by making the expert decision and pressing either the ”Include” or ”Exclude” button. This decision is then sent to the backend which trains the model following the steps described in Section 4.4. Besides training the screening model, the backend also updates the status field of the screened article to either, ”include on abstract” or ”excluded”, based on the expert decision.

5.4.2 Training phase

The second step is performed in order to train the model using active learning. The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer labeled training instances if it is allowed to choose the training data from which is learns [89]. This is performed at the ”Screen abstract” page shown in Figure 5.4. Since, the tool has gathered data for all articles in the search step. The model can determine which article it is most unsure about for classifying, by calculating the confidence score for every article in the current set. This way, the total number of articles which are needed for training can be reduced. The backend determines the next article which needs to be learned, which is shown in the same format as articles screened in the first step. In addition, information is provided about the percentage of articles which have been trained, as well as the distribution of articles when the current model would be used for automatic screening of the remaining ”unprocessed” articles. Since the layout is identical to the layout of the first step, the user can again use the ”Include” and ”Exclude” buttons to decide if an article should be included or not. However, after making the decision, the ”Screen abstract” page will automatically load the next article which needs to be trained.

Figure 5.4: Tool - Article screening active learning.

5.4.3 Screening phase

When the user thinks the model has been given enough training data, the remaining articles can be screened automatically by pressing the ”Auto screen abstract” button. As well as for the initialization phase screening information is included. Based on our evaluation it is suggested that a user should screen at least 30% before automatic screening is applied.

As described in Section 4.4, for screening, a document first needs to be pre-processed in order to remove words of insignificant semantic meaning and documents need to be stemmed. For every, article the backend first uses natural language processing (NLP) to tokenize every word of a provided document. This process takes approximately 300ms per article. Therefore, when new articles are added or updated in the tool, this information is already stored inside an article to speed up the automatic screening process. When the NLP data is available, the tokenization of words indicates what type of word has been found. Afterwards, stop words are removed from the text. The remaining words are then Porter stemmed in order to normalise all words in the dictionary. The remaining text is then used for term frequency–inverse document frequency (TFIDF), which returns the most important words of the provided document. The eleven most important words are then used to create a new text which is used to let the model determine if the given article should either be included or excluded. Depending on the size of the set to be screened, the automatic screening will take from a few second up to a few minutes.

5.5 Obtain full text

As stated in Section 4.5 no real automation solution is provided for retrieving full texts. However, a user can use full texts in the platform in two ways. The first is to include the full text by editing an article. An article can be edited by clicking the pencil icon in an article row on the ”Article overview” page. As shown in Figure A.8, a text field is included for the full text of an article. The second way is to include the full text by importing a list of papers as CSV. The ”full_text” CSV column header is used to identify full text imports. A detailed description of importing a project is provided in Section 5.8.1.

5.6 Screen full text

The screening process of full texts works similar to screening on title and abstract with two differences. First, instead of pressing the magnifying glass on the ”Article overview” page, the magnifying glass with a document in the background is pressed. This opens a popup containing the full text of an article and an indication if the current article should be included or excluded. Furthermore, instead of using the ”Abstract screen” for active learning, during full text screening, the ”Full text screen” page is used. When enough data has been trained, the ”Auto screen full text” button can be used, to automate the process of screening on full text.

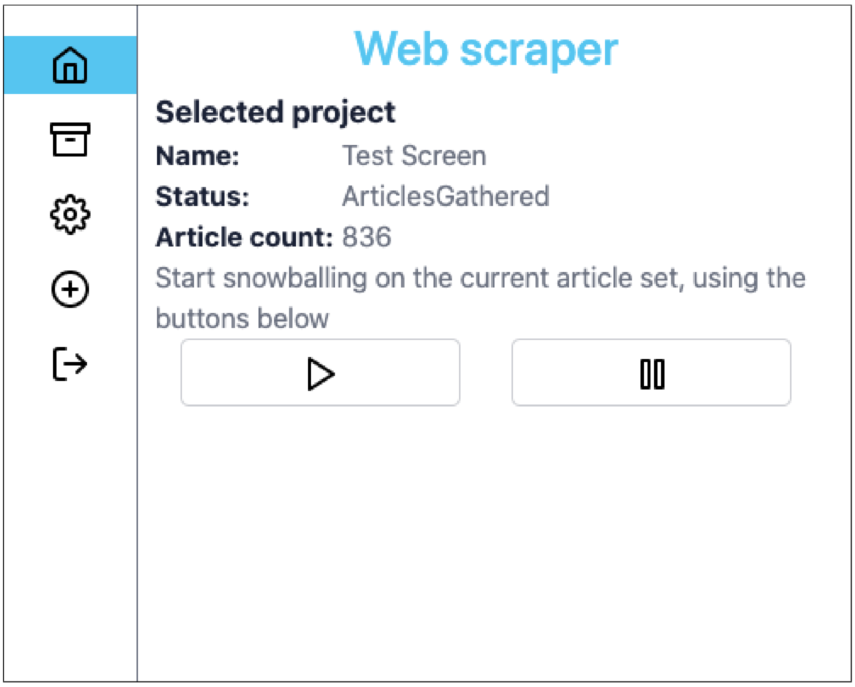
5.7 Snowballing

Snowballing of articles is performed using the Chrome plugin, as detailed in Section 4.7. When a project is selected for which the scraping step has been completed, or a project is created using the import functionality, detailed in Section 5.8.1, the home page of the plugin will provide the option to start snowballing instead of scraping. The adjusted home screen of the plugin is depicted in Figure 5.5.

The process works similar to the scraping functionality. After the user pressed the ”Play” button, the plugin sends a request to the backend, asking for an article which has not yet been snowballed. The backend retrieves a non snowballed article from the database which has a status of either ”included on abstract” or ”included”.

After receiving an article for snowballing, the popup component instructs the browser component to search a given article on Google Scholar. For this article the ”Cited by” URL is followed to add articles which reference the current article. The first five pages are scraped and the articles are sent to the backend and assigned a status ”unprocessed”. Besides adding the new article, the article which is snowballed gets updated receiving references to the newly snowballed articles using the articles’ DOI.

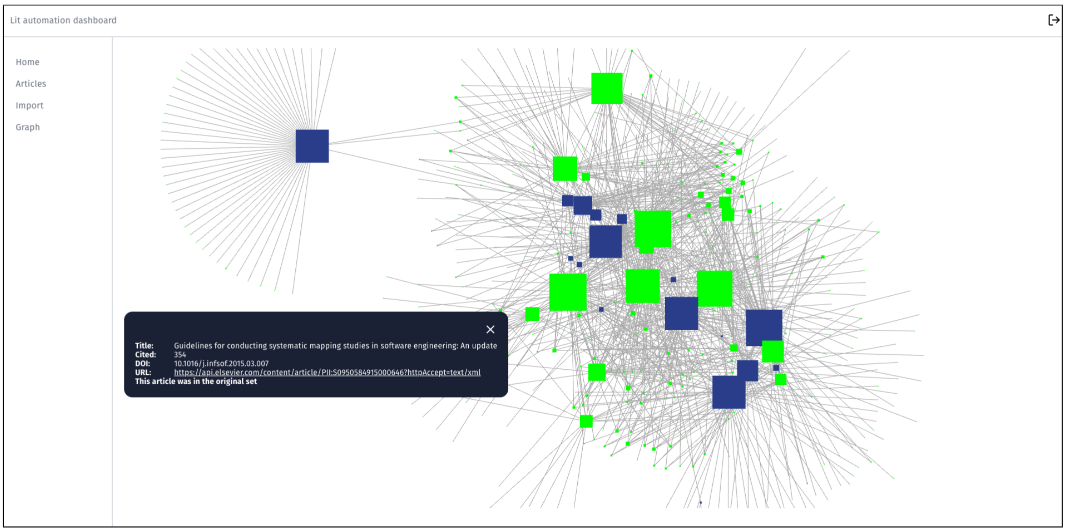
Figure 5.5: Chrome plugin - Snowball articles.



When all pages are visited, the plugin sends a request to the backend indicating that the current article has been snowballed and requesting a new one. When the snowballing has been completed, the user can start using the newly identified literature at the duplication removal step.

To provide a clear overview of the snowballed literature and how these articles are connected to the initial set, a "Graph" page is added to the tool as depicted in Figure 5.6. This graph represents articles as nodes and references are depicted as edges between those nodes. The size of the node indicates the number of references the current node has in the graph. Furthermore, the nodes are colored. Node colored green depict newly discovered articles during snowballing and blue nodes identify articles which were already identified. This graph can be used to identify snowballed literature which is highly connected to the current set of articles. The user can click on a node to show the information about the clicked article.

Figure 5.6: Tool - Snowballing graph.



5.8 Additional features

To provide more flexibility for researchers to manage the data set during the execution of a systematic literature review, additional functionality was added which not directly reflects the systematic review process. This section describes these features and details how they can be used.

5.8.1 Import projects

When a user does not want to depende on the Chrome Plugin, or already has gathered a set of articles for performing a systematic literature review, the user can import the article set into the tool using the ”Import” page, displayed in Figure A.7. The user needs to provide a name for the imported project, which is used for reference and a CSV file containing the article data. The provided CSV file should be valid RFC 4180 CSV document and may contain the headers: title, abstract, full text, DOI, URL, status and year. The CSV file and project name are sent to the backend which creates a project and adds every article included in the CSV file to the database. By default imported articles are marked with an ”unprocessed” status, however if the user provides an additional status, that status will be used instead. Because of this, a user can continue at every supported step after importing a project by CSV.

5.8.2 Article overview features

The ”Article overview” page, depicted in Figure 5.3, provides some additional functionality for managing the article set.

First of all, the page contains seven filters. These filters can be used to search through the set of articles. Implemented filters are: title, DOI, abstract, year, number of cites, the type of article and the article status. When adjusting the filter, the list of articles will automatically be reloaded in the overview page. Next to filtering, the user is able to browse through the set of articles, using the ”Prev” and ”Next” button shown on the bottom right corner of the screen.

Besides filtering and browsing functionality, the user can also add a new article using the Add Article button. This opens a popup, where the user can add the DOI and the title of the article. The backend will try to complete the article data using this information. However, when data for an article is still missing, the user has the option to edit an article using the edit button in an article row. To find the missing data, the user can press the globe button in an article row which opens the original website where the article is located in a separate tab in the browser.

Furthermore, a ”Download articles” button is added. When the download button is clicked, the tool exports the current article set as CSV, and downloadeds it to the users computer.

Chapter 6

LitAutomation: Evaluating the Tool

This chapter details the evaluation of the tool created in this work. This is described for every step addressed by the tool. To provide a sound evaluation, two questions were defined:

- In terms of time, what is the workload reduction?
- In comparison to manual execution, what is the impact on the quality of the results when using the tool?

To find answers to these questions, we defined two evaluation methods. First, for the retrieval, duplication removal and snowballing steps, we performed a qualitative comparison detailed in Sections 6.1, 6.2 and 6.5. This comparison was done by comparing the results produced by the tool, with the results of the manual execution of the AI for FinTech study. Second, for the screening of abstracts and full text, detailed in sections 6.3 and 6.6, the F1 score will be used to provide an answer to these questions. Labeled data created during the AI for FinTech study serves as input for these evaluations. All data used for this evaluation is available as technical report [95].

6.1 Retrieval evaluation

During the excution of the AI for FinTech study the following search query was used:

((”fintech” OR ”financial technology” OR ”banking”) AND (”AI” OR ”artificial intelligence” OR ”machine learning” OR ”deep learning”))

This query was entered in the Chrome Plugin which created a new project. Afterwards, the plugin was started to retrieve the articles from Web Of Science, Science Direct, Springer, IEEE and ACM. The results of executing the search using the plugin can be found in Table 6.1.

The execution of the original retrieval for the AI for FinTech study took approximately three weeks of work, with about 8 hours of work per day. This means the original retrieval took approximately 120 hours. As shown in the plugin execution results, it took the tool 20 minutes to gather articles for the given search query. This means that in terms of time, the workload was reduced by 119 hours and 40 minutes.

In order to identify the quality of the retrieved articles in comparison to the original results, we identified which articles in the original set could be found in the set of articles retrieved by the plugin. As shown in figure 2.4, the original retrieval step identified 2987 articles, whereas the search using the plugin resulted in 3096 articles. The third and fourth column of Table 6.1, show the number of articles from the original set, identified in the set retrieved by the tool. As shown, in total 1508 articles were identified and 1478 could not be found. The majority of articles which were not found, were originally extracted from ACM and Springer. The number of articles not found, might be explained due to the fact that during the execution of the original study, for these scholarly resources, the query was split up into three parts. Here, the ORs' of the left AND clause of the search query, were split up into three individual search queries. The tool is not able to execute this process, resulting in a different set of articles.

Table 6.1: Results for redoing the retrieval step.

Platform	Tool Duration	Articles Retrieved	Found	Not Found
ACM	10 min	1788	446	1096
Springer	1 min	50	102	203
IEEE	4 min	607	456	2
Web of Science	4 min	422	395	176
Science Direct	1 min	229	109	1
Total	20 min	3096	1508	1478

6.2 Duplication removal evaluation

In order to perform a comparison for duplication removal, the initial data set gathered during the AI for FinTech study was imported into the tool. Afterwards, in the articles overview page, the "Remove duplicates" button was used to identify duplicate articles.

During the original study, the duplicates were identified during the screening phase and data extraction phase. Therefore, no exact duration of removing duplicates could be given. However, we estimated this would have taken 4 hours of work. The tool was able to identify duplicates in 4 seconds, which in terms of time, resulted in a workload reduction of 3 hours and 56 minutes.

The original study identified 280 duplicate articles, whereas the tool was able to identify 252 articles as duplicates. This means that in terms of quality compared to manual execution, the tool was able to identify 90% of the duplicated articles. Table 6.2 provides an example of an article for which the tool was unable to detect duplication. As can be seen, both the DOI and Title are not equal, which causes the tool to fail in detecting the duplication.

Table 6.2: Unable to detect duplication example.

Title	**DOI**
Direct marketing campaigns in retail banking with the use of deep learning and random forests	10.1016/j.eswa.2019.05.020.
Direct marketing campaigns in retail banking with the use of deep	10.1016/j.eswa.2019.05.020

6.3 Abstract screening evaluation

To evaluate the abstract screening, the data set of the AI for FinTech study could be used. Since the data set was already labeled, it was possible to determine the accuracy of the model. Common measures for determining the performance of a classifier are precision and recall. Recall indicates the percentage of the included articles which were actually identified as being included by the classifier, whereas precision indicates the percentage of documents which were assigned a certain class were actually correct. These measurements can be expressed by:

$$recall = \frac{|TP|}{|TP| + |FN|} \tag{6.1}$$

$$precision = \frac{|TP|}{|TP| + |FP|} \tag{6.2}$$

Here, TP indicates the True Positives, which contains the set of documents for which both the screening model as well as the expert decision indicated that the document should be included. On the contrary, FN indicates the False Negatives, which are documents for which the expert decision indicated that the document should be included, but the screening model labeled the article excluded. Furthermore, False Positives are indicated by FP. This contains all documents for which the expert decision indicated that the document should be excluded, but the model indicated the document should be included. A full overview of the different screening result classification can be found in Table 6.3

Table 6.3: Different screening result classifications.

Expert decision	**Model prediction**	**Result**
Include	Include	True Positive
Include	Exclude	False Negative
Exclude	Include	False Positive
Exclude	Exclude	True Negative

To indicate the overall accuracy of the model, the F1 score is a common measure. This measure contains both the precision and recall and can be expressed by:

$$F1 = \frac{2 * precision * recall}{precision + recall} = \frac{2|TP|}{2|TP| + |FP| + |FN|} \tag{6.3}$$

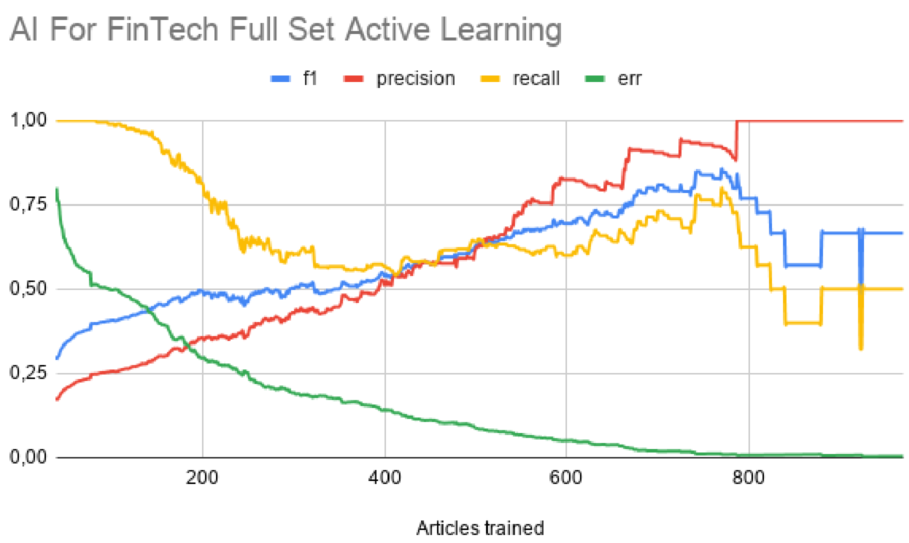
To indicate the inaccuracy of the model, the error was measured. The error can be expressed by:

$$err = \frac{|FP| + |FN|}{|TP| + |FP| + |TN| + |FN|} \tag{6.4}$$

We used the original data set from the AI for FinTech study, a subset of articles was created consisting of all articles for which both a title and abstract were present in the original set. This resulted in a data set consisting of 1165 articles for which 942 articles were labeled as excluded and 223 were labeled as included. Since a literature review should aim to identify as much relevant literature as possible, the model should aim for the highest recall possible, while preserving a high precision. While training on this data set, we found that the model had a hard time reducing the number of false negatives, thus having a low recall.

To address this issue, we initialized the model by training it with 25 included articles and 1 excluded article, before applying active learning. The training was stopped at 990 articles, which accounts for 85% of the total set. This was chosen because if more training is needed, we argue that the total workload reduction is not sufficient. The evaluation results are shown in Figure 6.1. The figure displays the F1, precision, recall and error. The y-axis indicates the score, whereas the x-axis indicates the number of articles used for training.

Figure 6.1: Abstract screening performance on full article set of literature review Machine Learning in FinTech: A Structured Mapping Study.



To provide a better understanding of Figure 6.1, data about the FN, FP, TN and TP is shown in Table 6.4 for four different data points.

Table 6.4: Data points from Figure 6.1

Articles Trained	FN	FP	TN	TP
83	0	556	342	184
500	28	32	556	50
740	6	1	404	14
850	3	0	310	2

Since active learning is applied, the model identifies the article for which the model is most unsure in which class it belongs. As shown, while using active learning, for training the first 83 articles, the recall remained 1.0. This means no false negatives were identified during this time as detailed in Table 6.4. Furthermore, the model identified 342 true negatives and 184 true positives, which means at this point the model would have saved the screening of about 45% of the articles. Originally the screening of articles took two researchers 5 days, with 8 hours of work per day, accounting for a total duration of 80 hours. This means that in terms of time, the tool was able to reduce the abstract screening workload by 36 hours.

In order to determine the quality of the screening result produced by the tool, we need to take a look at the FN and FP. As shown in Table 6.4, after training 500 articles, 28 relevant articles would have been discarded, whereas our included set would have contained 32 irrelevant articles.

It is also worth mentioning that the F1 score is gradually increasing until 740 articles were trained. The reason for the degrading F1 score after training over 740 articles, can be explained due to the fact that only 5 articles labeled included remained after 850 articles were trained. For these 5 articles, 3 articles were identified as False Negatives and 2 as true positives. Note that even though the F1 score degrades after 740 articles, the error rate is going closer to 0 as the number of articles trained increases. A more detailed evaluation of the screening model performance is elaborated in Section 6.6, where the model is tested against four different data sets.

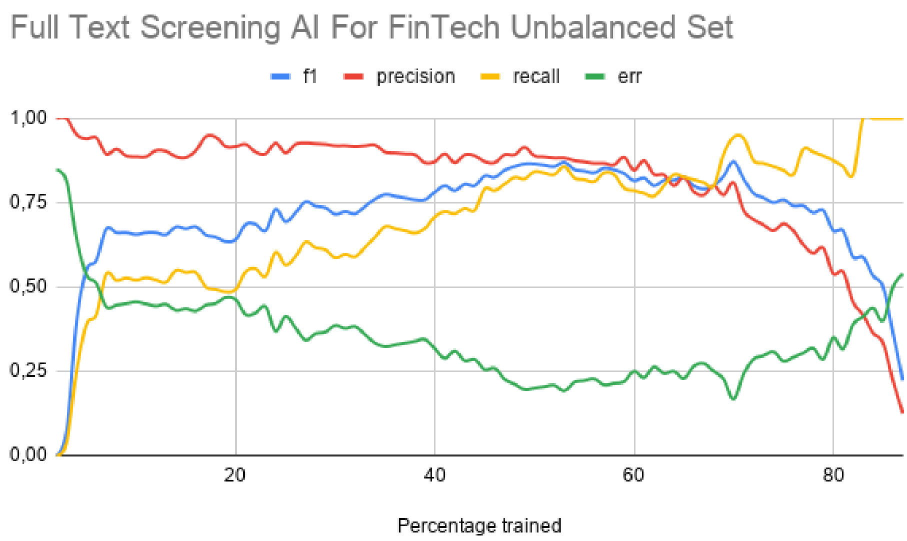
6.4 Screening full text

To evaluate the full text screening implemented in the tool, a data set was created containing 17 articles which were excluded and 83 articles which were included during the AI for FinTech study. The set was screened by the tool on full text using active learning. The result from screening the data set is shown in Figure 6.2.

Table 6.5: Data points from Figure 6.2

Articles Trained	FN	FP	TN	TP
2	83	0	15	0
20	34	3	10	33
40	14	5	7	34
50	6	4	8	32
60	6	4	8	22
80	1	6	6	7

Figure 6.2: Full text screening performance of included set after abstract screening of literature review Machine Learning in FinTech: A Structured Mapping Study.



To provide a better understanding of Figure 6.2, data about the FN, FP, TN and TP is shown in Table 6.5 for six different data points.

First of all, it is worth mentioning that in contradiction to screening on abstract, the recall starts at 0. This is resolves from the fact that the model is not initialized with multiple included articles. As shown in the figure, between training 40% and 60% of the articles the F1, Precision and Recall are greater than 0.75. This means in order for the model to be relevant, training on full text should be in this interval.

During the AI for FinTech study, in order to reduce time, the screening of full texts was combined with the synthesis phase. Therefore, we could not determine the exact time used for screening full texts, but we estimate it would have taken about 24 hours for a single researcher. As shown in table 6.5, in case the tool was used and we would have trained 50% 32 True Positives and 8 False Negatives would have been identified, accounting for a total of 40% of the full text screening workload. In terms of time, this means the tool could have reduced the workload by 9 hour and 36 minutes.

To determine the quality of the results produced by the full screening step, the FN and FP need to be identified. As shown in Table 6.5, in case 50% of the articles would have been trained, 6% of the relevant articles would have been discarded and the included set would have contained 4% irrelevant articles.

6.5 Snowballing

To evaluate the snowballing step, a CSV file was created containing the top 15 articles, detailed in Section 2.2.6. These articles were imported into the project with a status ”included”, after which the plugin was used for snowballing the articles from Google Scholar. The snowballing set contained 1399 articles. After merging the new articles with the initial set of articles, the tool was able to identify 937 duplicates, resulting in 479 new articles.

During the AI for FinTech study, this process was partially automated. For every article a script was run to scrape the first five pages of Google Scholar. This process took 4 hours. The tool took 9 minutes in order to retrieve all articles and remove duplicates. This means that in terms of time, the workload was reduced by 3 hours and 51 minutes.

Since the tool uses the same retrieval process as applied during the AI for FinTech study, the quality of the articles did not change, since the same articles were identified.

After the duplicate articles were removed, the remaining 479 articles were screened by the model created by screening the initial set on abstract and title. This resulted in 63 articles which were included and 416 articles which were excluded.

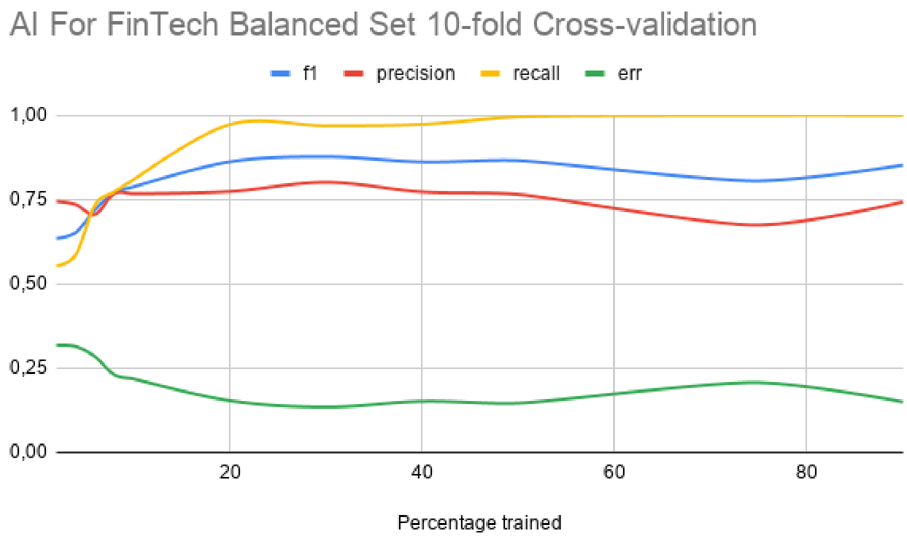
6.6 Additional screening evaluation

This section provides a detailed evaluation on the screening model as described in section 4.4. First Sections 6.6.1 - 6.6.3 are used to argue why active learning works better than normal supervised learning. Afterwards Section 6.6.4 is used in order to evaluate the model on four data sets resolving from four different systematic literature reviews. A complete overview of all data used for evaluation purposes can be found in a publicly available analysis repository [94].

6.6.1 Balanced data set

First we evaluate the model performance on a balanced data set, created as a subset from the AI for FinTech mapping study consisting of 50 articles which were labeled excluded an 50 articles were labeled included. Figure 6.3 shows the result of using normal supervised learning in combination with 10-fold cross-validation on this balanced data set. As shown, after training 10% of the articles, the F1 score is greater than 0.75. Furthermore, it is worth mentioning, that the recall is greater than 0.9 after training 20% of the articles and gradually increases towards 1.0. The F1 scores created by the model are similar results as found by [2].

Figure 6.3: Screening model performance on balanced data subset from Machine Learning in FinTech: A Structured Mapping Study using 10-fold cross-validation.



6.6.2 Issues with unbalanced data set

However, by analyzing data sets of different literature reviews, it becomes clear that data sets for abstract screening tend to be imbalanced. This can be further substantiated by the fact that the initial search query for literature reviews aim for a high recall in order to not miss out on any relevant literature. This means that the number of included and excluded articles are not evenly distributed. Therefore, a trend was discovered where the number of excluded articles are usually greater than the the number of included articles.

Figure 6.4 details the result of using 10-fold cross-validation on an unbalanced data set, extracted from the AI for FinTech study. This data set contained 20 articles labeled as included and 80 articles labeled as excluded. As shown, the overall error tends to behave the same as for the balanced set from subsection 6.6.1, however in contrast to the results of a balanced set, the recall tends to be lower than the precision. This is expected, since the model calculates the change for both the included and excluded class, whereas it will be biased towards the excluded class in case it learns a lot of excluded articles. Furthermore, the F1 score tends to be lower than the F1 score observed for the balanced data set. Instead of reaching 0.75 after 10%, the model needs to train around 88% of articles before reaching a F1 score of 0.75.

Figure 6.4: Screening model performance on unbalanced data subset from Machine Learning in FinTech: A Structured Mapping Study using 10-fold cross-validation.

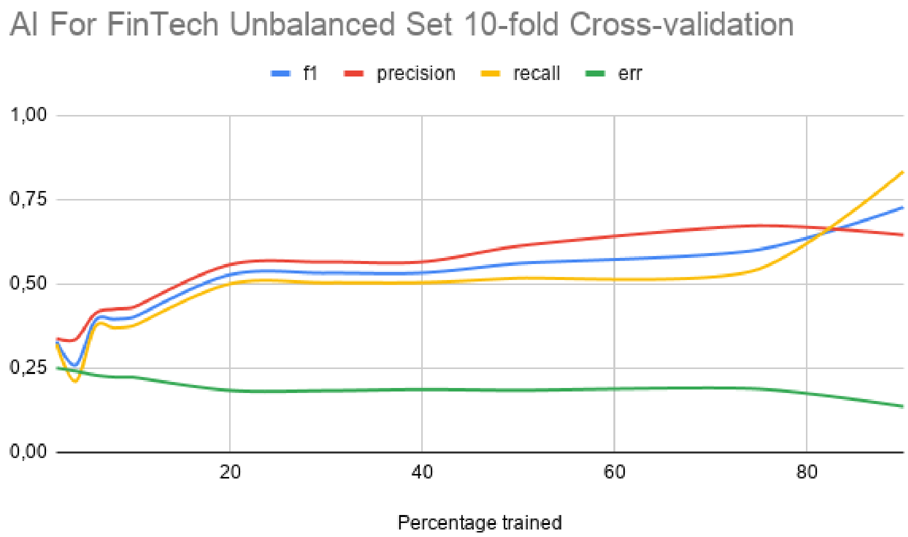
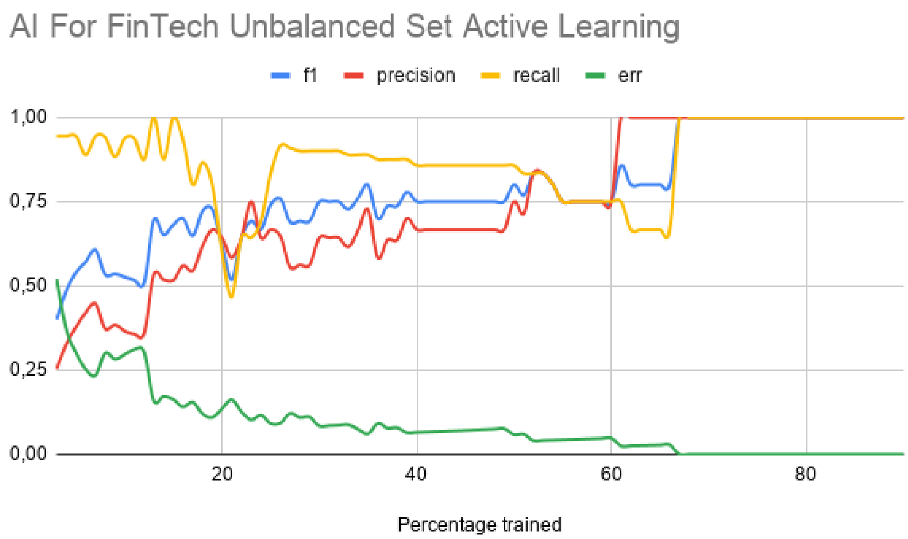


Figure 6.5: Screening model performance on unbalanced data subset from Machine Learning in FinTech: A Structured Mapping Study using active learning.



6.6.3 Active learning

To cope with the issue of unbalanced data, as stated in section 4.4, we introduced active learning for training the model. The results for applying active learning on the same data set used in sub Section 6.6.2 is shown in Figure 6.5. We can see major improvements compared to the result of normal supervised learning. As shown, the F1 score of 0.75 is now reached after training 25% of the articles instead of 88%. Furthermore, it is worth mentioning that the recall tend to be a lot higher than the precision.

6.6.4 Evaluation on four different data sets

To evaluate whether the model does not just work for one data set, a total of four data sets were retrieved from different literature reviews to evaluate the model. The list of literature reviews consisted of:

- Machine Learning in FinTech: A Structured Mapping Study [2]
- Contemporary Software Monitoring: A Systematic Literature Review [13]
- Self-Adaptation in Mobile Apps: a Systematic Literature Study [31]
- Reinforcement learning for personalization: A systematic literature review [19]

For every literature review a random subset of 100 articles was extracted. Afterwards, the model was used to classify articles and trained using active learning.

Figure 6.6 compares the F1 scores for the different data sets. First of all, it is worth mentioning that the F1 scores for [19] and [2] behave similarly. The F1 scores tend to be a little worse on the data set from [31], however the overall trend follows the scores of [19] and [2]. On the other hand, the F1 scores for the data set from [13] are significantly lower than the other scores.

Figure 6.6: F1 score comparison on 4 different data sets.

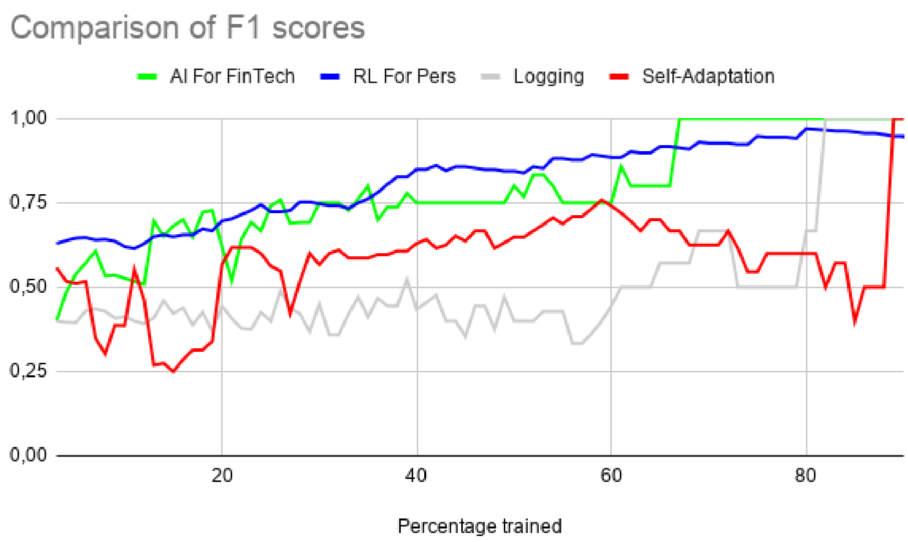
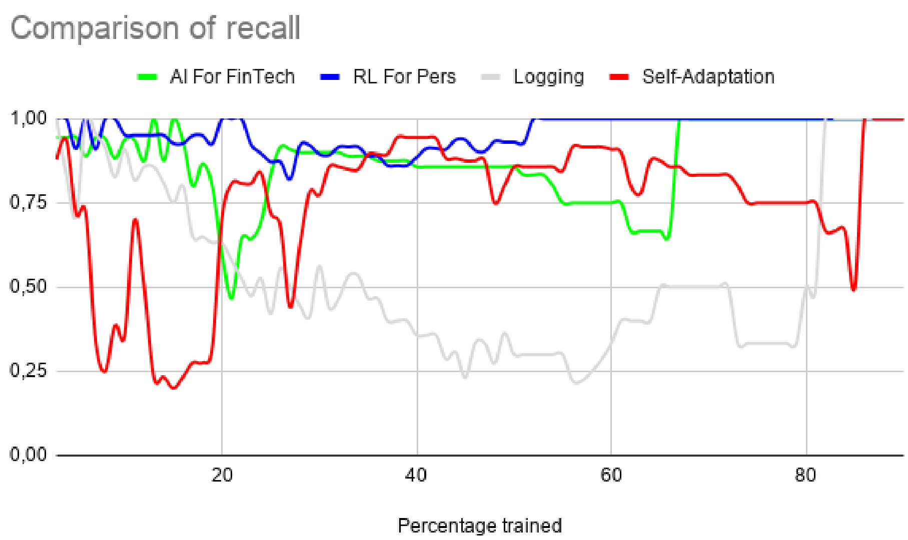


Figure 6.7 compares the recall scores for the different data sets. As well as for the F1 scores, it is worth mentioning that the recall scores for [19], [31] and [2] behave similarly. However the recall for the data set [13] is significantly lower.

Figure 6.7: Recall score comparison on 4 different data sets.



To explain the lower scores for the data from paper [13], the paper was analyzed. We found that the inclusion/exclusion criteria contained venue rank and the type of paper. The implemented screening model uses the title and abstract from an article which does not contain information about the criteria defined in this paper. This could very well explain the lower scores found for this data set. A further discussion about these results can be found in Chapter 7. Furthermore, it is worth mentioning that aside from the data regarding [13], the recall for the other three data sets was over 0.85 after training 30% of the articles.

Chapter 7

Discussion

The goal for this research is to partially automate the screening and retrieval phase for systematic literature reviews. The first section discusses the meaning of our results found in Chapter 6. This is followed by Section 7.2, discussing the limitations of this work. Afterwards, this chapter is concluded, by providing insights in future research directions regarding systematic literature review automation in Section 7.3.

7.1 Interpretation of findings

This section elaborates the meaning of the results found in Chapter 6. This is done in the order in which systematic literature review steps are described in Figure 4.1.

Search Our work has shown that a tool can be created to automatically gather over 3000 articles in about 20 minutes. This is arguably faster than searching these articles by hand. Furthermore, it requires less work than manually gathering all article information, whereas all gathered data is automatically stored in a database.

Remove duplicates The duplication removal of our work has proved to be effective. The automatic duplication removal was able to reduce the work of manually removing duplicates by 90%. The unidentified duplicates might be explained due to the fact that for some articles the DOI was missing. This means that in case the data retrieval for articles can be further improved, the automatic duplication removal might also further improve. This is further discussed in Section 7.2.

Screen abstracts The abstract screening has to be proven effective when using it for separating included from excluded articles based on the information present in abstract and title. Our evaluation data shows as that if one is willing to omit 15% of relevant literature the screening workload can be reduced by 70% with an average F1 score of 0.75. This means the amount of time needed to be spent on screening can be significantly reduced.

Screen full text The screening of full texts works identical to the screening of abstract. However, screening on full text is usually done on data sets which are more similar than during abstract screening. Therefore it was found that, to screen articles on full text after abstract screening, 50% should be screened to ensure the recall is sufficient. This means, for full text screening, the workload can be reduced by 50% as well.

Snowball This work proved that the tool produced is able to correctly perform snowballing using Google Scholar. Furthermore, the newly identified literature can be further processed starting at the duplication removal step. Therefore, this work showed that it is possible to create a tool for partially automating the screening and retrieval phase of systematic literature reviews and thus reducing the overall workload of performing a literature review.

7.2 Limitations

We have identified three main threats to the validity of this work: (1) The available data for scraping, (2) Screening model: user input, (3) screening model: external validity.

Even though the tool tries to gather as much data as possible, during the search step of the systematic literature review, the tool is dependent on the information present on the given scholarly resource. That is to say, some resources do not have all information about literature present, or might have incomplete data for some literature. This is a threat, since the follow up steps of the systematic literature review are dependent on the quality of the gathered data. For example, the abstract screening will be influenced by the completeness of abstract texts of articles. However, by performing an additional search for missing information, using the Crossref API, for every article which is added to the tool, this risk should have been omitted as much as possible.

During the screening of both abstract and full text, a researcher is able to determine how many and which articles to train before starting to train the screening model using active learning. In case the model is initialized with a lot of negative articles, the recall will be lower compared to when the model is initialized using more included articles. It is also possible to not even use active learning, but instead use automatic screening directly after initialization. Furthermore, inclusion and exclusion criteria can include properties such as number of citations and year. Since the screening model uses the abstract and title, it is not able to identify these properties. Therefore, users should first remove articles by any of these additional properties before starting to screen using the model. To omit this risk as much as possible, a detailed description is provided in the tool, for users on how to properly use the screening automation.

The model used for screening was designed using the data resulting from the literature review performed in this work. Therefore, the question arises if this model also works for different data sets. In order to reduce this risk as much as possible, a total of four different data sets were used in order to evaluate the model. The scores for different data sets were compared and evaluated.

7.3 Future directions

By performing this research an overview of automation techniques has been created for every step of a literature review. The information provided by this research can be used to further improve the current automation solutions, or extend the current work by automating steps which are not addressed by the created tool. Furthermore, by creating the tool in this work, not only has it been proven that it is possible to automate the retrieval and screening phase of systematic literature reviews. The tool can actually be used since it is deployed at [92]. Therefore, this work realized that the overall workload of performing a systematic literature review can be significantly reduced.

This section discusses points and improvements, which were out of scope of this work, but could be used to inspire future research regarding the automation of literature reviews.

First of all, the tool can be extended by adding automation solutions for systematic literature review steps which were out of scope of this research. For example, the data extraction and data synthesis could be addressed, using the results detailed in Chapter 3. Adding these steps to the tool would have made it possible to fully automate the execution of the AI for FinTech study, detailed in chapter 2. In addition, it would be interesting to see if this extracted data can automatically be used to create graphs of interest to use in systematic literature reviews.

Furthermore, the model used for screening automation could be extended. This could be done by not only using the title and abstract as input data for screening, but include additional parameters such as, keywords, journal, year, number of citations etc. This way, the screening model would be able to classify articles when researchers exclude them on other properties than the title and abstract alone.

It would also be interesting to compare the results of a support vector machine classifier (SVM) classifier to the results of this work. As argued by [2], together with Naive Bayes, an SVM classifier seems very promising when it is used for the screening on abstracts.

As argued in Section 7.2, the quality of literature reviews are dependent on the completeness of the input data. However, every scholarly resource maintains its own API and its own standards on how data is formatted. Therefore, we argue that it would be very interesting to see if an open standard could be created for how literature data should be formatted. On top of this standard, an API combining all scholarly resources could be build, to provide one single entry point for retrieving literature data. The Crossref API [76] tries to accomplish this. However, during this research it was found that for the majority of articles not all data is present in this API.

Chapter 8

Conclusions

Systematic literature reviews are essential for producing sound scientific research. However, the overall execution duration of these reviews may consume a significant amount of time. Therefore, it is important to perform research on reducing of the overall workload of performing such reviews.

First a structured mapping study was performed on the topic of machine learning in FinTech. This mapping study resulted in three main findings. First, life cycle aspects of machine learning seem to be an undervalued topic. Second, a lack of case studies were observed. Furthermore, it was found that only limited research is performed on the topic of customer due diligence.

Even though this study resulted in new insights with regard to machine learning in FinTech, it was also found that producing these results was not trivial and consumed a significant amount of time.

Because of that, this work has provided a complete overview of possible automation solutions for every step of performing literature reviews. Using this overview a tool was created, which showed that the overall workload of the retrieval and screening phase of systematic literature reviews can be significantly reduced. This tool was not only created, verified and published as source code, in addition a live version has been published [92] which can be used by anyone.

During this research, it was found that executing a search, resulting in over 3000 articles, can be performed in 20 minutes. It was also argued that the tool is able to automatically identify 90% of duplicated articles. Furthermore, it was shown, that the workload of abstract screening can be reduced by 70% and the workload of full text screening can be reduced by 50%. Lastly, it was shown that the tool is able to perform snowballing on a given article set.

We argued that future research regarding literature review automation should be focused on extending the created tool by addressing the data extraction and synthesis phase of the systematic literature review. Furthermore, the screening model created in this work could be further developed by adding additional article metadata such as year, journal number of citations and keywords. Another possibility would be to compare the results of this study with a screening model using an SVM classifier, since [2] argued that this classifier also has potential for supporting the screening automation. Furthermore, it was argued that the field of automation with regard to literature reviews would benefit from standardization on literature data.

We estimate that in case this tool would have been available when the initial mapping study was performed, in terms of time, the overall workload could have been reduced by approximately 173 hours.

Bibliography

[1] ICSE 2020. *42nd International Conference on Software Engineering*, 2019. URL https://conf.researchr.org/track/icse-2020/icse-2020-papers.

[2] J.J. García Adeva, J.M. Pikatza Atxa, M. Ubeda Carrillo, and E. Ansuategi Zengotitabengoa. Automatic text classification to support systematic reviews in medicine. *Expert Systems with Applications*, 41(4):1498–1508, mar 2014. doi: 10.1016/j.eswa.2013.08.047. URL https://doi.org/10.1016/j.eswa.2013.08.047.

[3] Manuela Adl and William Haworth. How a know-your-customer utility could increase access to financial services in emerging markets. *EMCompass*, 2018.

[4] Sophia Ananiadou, Brian Rea, Naoaki Okazaki, Rob Procter, and James Thomas. Supporting systematic reviews using text mining. *Social Science Computer Review*, 27(4):509–523, 2009.

[5] Douglas W Arner, Janos Barberis, and Ross P Buckley. The evolution of fintech: A new post-crisis paradigm. *Geo. J. Int'l L.*, 47:1271, 2015.

[6] Rob Ashmore, Radu Calinescu, and Colin Paterson. Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *arXiv preprint arXiv:1905.04223*, 2019.

[7] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18(153), 2018.

[8] Tanja Bekhuis and Dina Demner-Fushman. Towards automating the initial screening phase of a systematic review. In *MedInfo*, pages 146–150, 2010.

[9] Daniel Belanche, Luis V Casaló, and Carlos Flavián. Artificial intelligence in fintech: understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*, 2019.

[10] Richa Bhatia. *Understanding the difference between symbolic AI & non symbolic AI*, 2017. URL `https://analyticsindiamag.com/`.

[11] Alison Booth, Mike Clarke, Davina Ghersi, David Moher, Mark Petticrew, and Lesley Stewart. An international registry of systematic-review protocols. *Lancet*, 377 (9760), 2011.

[12] Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] Jeanderson Candido, Maurício Aniche, and Arie van Deursen. Contemporary software monitoring: A systematic literature review. *arXiv preprint arXiv:1912.05878*, 2019.

[14] Mirka Snyder Caron. The transformative effect of ai on the banking industry. *Banking & Finance Law Review*, 34(2):169–214, 2019.

[15] Nader Chmait, David L Dowe, David G Green, and Yuan-Fang Li. Agent coordination and potential risks: Meaningful environments for evaluating multiagent systems. In *Evaluating General-Purpose AI, IJCAI Workshop*, 2017.

[16] Miew Keen Choong, Filippo Galgani, Adam G Dunn, and Guy Tsafnat. Automatic evidence retrieval for systematic reviews. *Journal of medical Internet research*, 16 (10):e223, 2014.

[17] Aaron M Cohen, Kyle Ambert, and Marian McDonagh. Studying the potential impact of automated document classification on scheduling a systematic review update. *BMC medical informatics and decision making*, 12(1):33, 2012.

[18] Andrea Fronzetti Colladon and Elisa Remondi. Using social network analysis to prevent money laundering. *Expert Systems with Applications*, 67:49–58, 2017.

[19] Floris den Hengst, Eoin Martino Grua, Ali el Hassouni, and Mark Hoogendoorn. Reinforcement learning for personalization: A systematic literature review. *Data Science*, pages 1–41, 2020.

[20] Science Direct. *Science Direct*, 2020. URL `https://www.sciencedirect.com/`.

[21] Rhett D'souza. *Symbolic AI versus Non-Symbolic AI, and everything in between?*, 2018. URL `https://medium.com/datadriveninvestor/`.

[22] Tore Dyba, Torgeir Dingsoyr, and Geir K Hanssen. Applying systematic reviews to diverse study types: An experience report. In *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, pages 225–234. IEEE, 2007.

[23] Ahmed K Elmagarmid, Panagiotis G Ipeirotis, and Vassilios S Verykios. Duplicate record detection: A survey. *IEEE Transactions on knowledge and data engineering*, 19(1):1–16, 2006.

[24] Katia R. Felizardo, Gabriel F. Andery, Fernando V. Paulovich, Rosane Minghim, and José C. Maldonado. A visual analysis approach to validate the selection review of primary studies in systematic reviews. *Information and Software Technology*, 54(10): 1079–1091, oct 2012. doi: 10.1016/j.infsof.2012.04.003. URL `https://doi.org/10.1016/j.infsof.2012.04.003`.

[25] Katia Romero Felizardo, Simone R.S. Souza, and Jose Carlos Maldonado. The use of visual text mining to support the study selection activity in systematic literature reviews: A replication study. In *2013 3rd International Workshop on Replication in Empirical Software Engineering Research*. IEEE, oct 2013. doi: 10.1109/reser.2013.9. URL `https://doi.org/10.1109/reser.2013.9`.

[26] Mark Fenwick and Erik PM Vermeulen. How to respond to artificial intelligence in fintech. *Japan SPOTLIGHT*, 2017.

[27] DOI Foundation. *DOI*, 2020 (accessed June 6, 2020). URL `https://www.doi.org/`.

[28] Oana Frunza, Diana Inkpen, Stan Matwin, William Klement, and Peter O'Blenis. Exploiting the systematic review protocol for classification of medical abstracts. *Artificial Intelligence in Medicine*, 51(1):17–25, jan 2011. doi: 10.1016/j.artmed.2010.10.005. URL `https://doi.org/10.1016/j.artmed.2010.10.005`.

[29] Shijia Gao and Dongming Xu. Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering. *Expert Systems with Applications*, 36(2):1493–1504, 2009.

[30] Google. *Google Scholar*, 2020. URL `https://scholar.google.com/`.

[31] Eoin Martino Grua, Ivano Malavolta, and Patricia Lago. Self-adaptation in mobile apps: a systematic literature study. In *2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, pages 51–62. IEEE, 2019.

[32] Igli Hakrama and Neki Frashëri. Agent-based modeling and simulation of an artificial economy with repast. *International Journal on Information Technologies & Security*, 4(2), 2018.

[33] Marie J Hansen, Nana Ø Rasmussen, and Grace Chung. A method of extracting the number of trial participants from abstracts describing randomized controlled trials. *Journal of Telemedicine and Telecare*, 14(7):354–358, 2008.

[34] Md Majharul Haque, Suraiya Pervin, and Zerina Begum. Literature review of automatic single document text summarization using nlp. *International Journal of Innovation and Applied Studies*, 3(3):857–865, 2013.

[35] John Haugeland. *Artificial intelligence: The very idea*. MIT press, 1989.

[36] Dan Headrick and MaryAnne M Gobble. Ai-powered fintech turns data into new business. *Research-Technology Management*, 62(1):5, 2019.

[37] Frances Ann Hubis, Wentao Wu, and Ce Zhang. Ease. ml/meter: Quantitative overfitting management for human-in-the-loop ml application development. *arXiv preprint arXiv:1906.00299*, 2019.

[38] Mirjana Ivanović, Zoran Budimac, Miloš Radovanović, Vladimir Kurbalija, Weihui Dai, Costin Bădică, Mihaela Colhon, Sran Ninković, and Dejan Mitrović. Emotional agents–state of the art and applications. *Computer Science and Information Systems*, 12(4):1121–1148, 2015.

[39] Julapa Jagtiani and Kose John. Fintech: The impact on consumers and regulatory responses, 2018.

[40] Samireh Jalali and Claes Wohlin. Systematic literature studies: database searches vs. backward snowballing. In *Proceedings of the 2012 ACM-IEEE international symposium on empirical software engineering and measurement*, pages 29–38. IEEE, 2012.

[41] Muhammad Abid Jamil, Muhammad Arif, Normi Sham Awang Abubakar, and Akhlaq Ahmad. Software testing techniques: A literature review. In *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, pages 177–182. IEEE, 2016.

[42] Tommi Sakari Jauhiainen, Marco Lui, Marcos Zampieri, Timothy Baldwin, and Krister Lindén. Automatic language identification in texts: A survey. *Journal of Artificial Intelligence Research*, 65:675–782, 2019.

[43] Siddhartha R Jonnalagadda, Pawan Goyal, and Mark D Huffman. Automating data extraction in systematic reviews: a systematic review. *Systematic reviews*, 4(1):78, 2015.

[44] Upulee Kanewala and James M Bieman. Testing scientific software: A systematic literature review. *arXiv preprint arXiv:1804.01954*, 2018.

[45] Staffs Keele et al. Guidelines for performing systematic literature reviews in software engineering. Technical report, Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.

[46] Foutse Khomh, Bram Adams, Jinghui Cheng, Marios Fokaefs, and Giuliano Antoniol. Software engineering for machine-learning applications: The road ahead. *IEEE Software*, 35(5):81–84, 2018.

[47] Su Nam Kim, David Martinez, Lawrence Cavedon, and Lars Yencken. Automatic classification of sentences to support evidence based medicine. In *BMC bioinformatics*, volume 12, page S5. BioMed Central, 2011.

[48] Svetlana Kiritchenko, Berry De Bruijn, Simona Carini, Joel Martin, and Ida Sim. Exact: automatic extraction of clinical trial characteristics from journal publications. *BMC medical informatics and decision making*, 10(1):56, 2010.

[49] B. Kitchenham and S Charters. Guidelines for performing systematic literature reviews in software engineering, 2007.

[50] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1):7–15, 2009.

[51] Rick Kuhlman, Jeff Ziesman, Carolyn Browne, and Jason Kempf. Is your anti-money laundering program ready for fincen's customer due diligence rule? *Journal of Investment Compliance*, 19(2):42–44, 2018.

[52] Karry Lai. Fintech: banks still struggle with culture and legacy systems. *International Financial Law Review*, 2019.

[53] Tae-heon Lee and Hee-Woong Kim. An exploratory study on fintech industry in korea: crowdfunding case. In *2nd International conference on innovative engineering technologies (ICIET'2015). Bangkok*, 2015.

[54] Artuur Leeuwenberg and Marie-Francine Moens. A survey on temporal reasoning for temporal information extraction from text. *Journal of Artificial Intelligence Research*, 66:341–380, 2019.

[55] ACM Digital Library. *ACM Digital Library*, 2020. URL `https://dl.acm.org/`.

[56] Qiang Liu, Pan Li, Wentao Zhao, Wei Cai, Shui Yu, and Victor CM Leung. A survey on security threats and defensive techniques of machine learning: A data driven view. *IEEE access*, 6:12103–12117, 2018.

[57] Lei Ma, Felix Juefei-Xu, Fuyuan Zhang, Jiyuan Sun, Minhui Xue, Bo Li, Chunyang Chen, Ting Su, Li Li, Yang Liu, et al. Deepgauge: Multi-granularity testing criteria for deep learning systems. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, pages 120–131. ACM, 2018.

[58] Lei Ma, Felix Juefei-Xu, Minhui Xue, Bo Li, Li Li, Yang Liu, and Jianjun Zhao. Deepct: Tomographic combinatorial testing for deep learning systems. In *2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 614–618. IEEE, 2019.

[59] Viviane Malheiros, Erika Hohn, Roberto Pinho, Manoel Mendonca, and Jose Carlos Maldonado. A visual text mining approach for systematic reviews. In *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*. IEEE, sep 2007. doi: 10.1109/esem.2007.21. URL `https://doi.org/10.1109/esem.2007.21`.

[60] Samuel Marcos-Pablos and Francisco José García-Peñalvo. Decision support tools for slr search string construction. In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 660–667, 2018.

[61] Dusica Marijan, Arnaud Gotlieb, and Mohit Kumar Ahuja. Challenges of testing machine learning based systems. In *2019 IEEE International Conference On Artificial Intelligence Testing (AITest)*, pages 101–102. IEEE, 2019.

[62] Iain J Marshall and Byron C Wallace. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic reviews*, 8(1):163, 2019.

[63] Satoshi Masuda, Kohichi Ono, Toshiaki Yasue, and Nobuhiro Hosokawa. A survey of software quality for machine learning applications. In *2018 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, pages 279–284. IEEE, 2018.

[64] Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, and Peter OBlenis. A new algorithm for reducing the workload of experts in performing systematic reviews. *Journal of the American Medical Informatics Association*, 17(4):446–453, jul 2010. doi: 10.1136/jamia.2010.004325. URL `https://doi.org/10.1136/jamia.2010.004325`.

[65] Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, and Peter OBlenis. Performance of SVM and bayesian classifiers on the systematic review classification task. *Journal of the American Medical Informatics Association*, 18(1): 104.2–105, jan 2011. doi: 10.1136/jamia.2010.009555. URL `https://doi.org/10.1136/jamia.2010.009555`.

[66] Lizzie Meager and Jimmie Franklin. Fintech europe 2019: key takeaways. *International Financial Law Review*, 2019.

[67] Sara Mehryar, RV Sliuzas, N Schwarz, Ali Sharifi, and MFAM van Maarseveen. From aggregated knowledge to interactive agents: an agent based approach to support policy making in social-ecological systems. *Journal of Environmental Management*, 2018.

[68] Sepideh Mesbah, Kyriakos Fragkeskos, Christoph Lofi, Alessandro Bozzon, and Geert-Jan Houben. Facet embeddings for explorative analytics in digital libraries. In *International conference on theory and practice of digital libraries*, pages 86–99. Springer, 2017.

[69] MIT. *APIS FOR SCHOLARLY RESOURCES*, 2020 (accessed June 6, 2020). URL `https://libraries.mit.edu/scholarly/publishing/apis-for-scholarly-resources/`.

[70] Makoto Miwa, James Thomas, Alison O'Mara-Eves, and Sophia Ananiadou. Reducing systematic review workload through certainty-based screening. *Journal of Biomedical Informatics*, 51:242–253, oct 2014. doi: 10.1016/j.jbi.2014.06.005. URL `https://doi.org/10.1016/j.jbi.2014.06.005`.

[71] Raphaël Mourad, Christine Sinoquet, Nevin Lianwen Zhang, Tengfei Liu, and Philippe Leray. A survey on latent tree models and applications. *Journal of Artificial Intelligence Research*, 47:157–203, 2013.

[72] Thuy TT Nguyen and Grenville Armitage. A survey of techniques for internet traffic classification using machine learning. *IEEE communications surveys & tutorials*, 10 (4):56–76, 2008.

[73] Delft University of Technology. *AI for Fintech Research*, 2020. URL `https://se.ewi.tudelft.nl/ai4fintech/`.

[74] Publish or Perish. *Publish or Perish*, 2020. URL `https://harzing.com/resources/publish-or-perish`.

[75] IEEE ORG. *Advancing Technology for Humanity*, 2020. URL `https://ieeexplore.ieee.org/`.

[76] Crossref Organization. *The Crossref Curriculum*, 2020 (accessed May 7, 2020). URL `https://www.crossref.org/education/retrieve-metadata/rest-api/`.

[77] Arthur O'Sullivan and Steven M. Sheffrin. *Economics: Principles in Action*. Google Books, 2003. ISBN 0-13-063085-3.

[78] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. In *Ease*, volume 8, pages 68–77, 2008.

[79] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18, 2015.

[80] Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. Data lifecycle challenges in production machine learning: a survey. *ACM SIGMOD Record*, 47(2):17–28, 2018.

[81] Cedric Renggli, Frances Ann Hubis, Bojan Karlaš, Kevin Schawinski, Wentao Wu, and Ce Zhang. Ease.ml/ci and ease.ml/meter in action: Towards data management for statistical generalization. *Proc. VLDB Endow.*, 12(12):1962–1965, August 2019. ISSN 2150-8097. doi: 10.14778/3352063.3352110. URL `https://doi.org/10.14778/3352063.3352110`.

[82] Cedric Renggli, Bojan Karlaš, Bolin Ding, Feng Liu, Kevin Schawinski, Wentao Wu, and Ce Zhang. Continuous integration of machine learning models with ease. ml/ci: Towards a rigorous yet practical treatment. *arXiv preprint arXiv:1903.00278*, 2019.

[83] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

[84] Sebastian Ruder, Ivan Vulić, and Anders Søgaard. A survey of cross-lingual word embedding models. *arXiv preprint arXiv:1706.04902*, 2017.

[85] Paul Schulte and Gavin Liu. Fintech is merging with iot and ai to challenge banks: How entrenched interests can prepare. *The Journal of alternative investments*, 20(3): 41–57, 2017.

[86] Paul Schulte, David Kuo Chuen Lee, et al. Case study: Us vs prc in the banking lizard brain shift to ai fintech machines. *World Scientific Book Chapters*, pages 27–72, 2019.

[87] Web Of Science. *Web Of Science*, 2020. URL `https://www.webofknowledge.com/`.

[88] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In *Advances in neural information processing systems*, pages 2503–2511, 2015.

[89] Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

[90] Ian Shemilt, Antonia Simon, Gareth J. Hollands, Theresa M. Marteau, David Ogilvie, Alison OMara-Eves, Michael P. Kelly, and James Thomas. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods*, 5(1):31–49, aug 2013. doi: 10.1002/jrsm.1093. URL `https://doi.org/10.1002/jrsm.1093`.

[91] Birte Snilstveit, Martina Vojtkova, Ami Bhavsar, and Marie Gaarder. *Evidence gap maps—a tool for promoting evidence-informed policy and prioritizing future research*. The world bank, 2013.

[92] Wim Spaargaren. *Literature Automation Tool*, 2020. URL `https://lit.wimsp.nl/`.

[93] Wim Spaargaren. *The Literature Survey Automation Tool*, 2020 (accessed June 10, 2020). URL `https://github.com/lit-automation`.

[94] Wim Spaargaren. *Screening model data*, 2020 (accessed June 20, 2020). URL `https://docs.google.com/spreadsheets/d/1hSq1NqO1WCV6vCODvzKF7fpkd4lPkZf_OSCk5xugRrs/edit?usp=sharing`.

[95] Wim Spaargaren. *Evaluation data*, 2020 (accessed June 29, 2020). URL https://docs.google.com/spreadsheets/d/1l7FTPOwTQpAAr0839b2j7sAvSYaSN7B2s7_L9CKqFnY/edit?usp=sharing.

[96] Wim Spaargaren, Hennie Huijgens, and Arie van Deursen. *Analysis Repository - ML in FinTech: A Systematic Mapping Study*, 2019. URL https://docs.google.com/spreadsheets/d/1YTrzZotWqW0q0-See4iDD3fv0YcWiKrmSrJPTyRM0Dc/edit?usp=sharing.

[97] Springer. *Springer*, 2020. URL https://www.springer.com/.

[98] Peter Stone and Manuela Veloso. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3):345–383, 2000.

[99] Shiliang Sun. A survey of multi-view machine learning. *Neural computing and applications*, 23(7-8):2031–2038, 2013.

[100] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, and M. Morisio. Linked data approach for selection process automation in systematic reviews. In *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*. IET, 2011. doi: 10.1049/ic.2011.0004. URL https://doi.org/10.1049/ic.2011.0004.

[101] Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani, and Enrico Coiera. Systematic review automation technologies. *Systematic Reviews*, 3 (1), jul 2014. doi: 10.1186/2046-4053-3-74. URL https://doi.org/10.1186/2046-4053-3-74.

[102] Ömer Uludag, Martin Kleehaus, Christoph Caprano, and Florian Matthes. Identifying and structuring challenges in large-scale agile development based on a structured literature review. In *2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, pages 191–197. IEEE, 2018.

[103] Dalhousie University. Libguides: Systematic reviews: A how-to guide: Review protocol development, 2020. URL http://dal.ca.libguides.com/systematicreviews/Protocol.

[104] Armando Vieira and Attul Sehgal. How banks can better serve their customers through artificial techniques. In *Digital Marketplaces Unleashed*, pages 311–326. Springer, 2018.

[105] Byron C Wallace, Thomas A Trikalinos, Joseph Lau, Carla Brodley, and Christopher H Schmid. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*, 11(1), jan 2010. doi: 10.1186/1471-2105-11-55. URL https://doi.org/10.1186/1471-2105-11-55.

[106] Byron C Wallace, Kevin Small, Carla E Brodley, Joseph Lau, and Thomas A Trikalinos. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In *proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 819–824, 2012.

[107] Matthew Wicker, Xiaowei Huang, and Marta Kwiatkowska. Feature-guided black-box safety testing of deep neural networks. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 408–426. Springer, 2018.

[108] Roel Wieringa, Neil Maiden, Nancy Mead, and Colette Rolland. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, 11(1):102–107, 2006.

[109] Tao Xie. Intelligent software engineering: Synergy between ai and software engineering. In *International Symposium on Dependable Software Engineering: Theories, Tools, and Applications*, pages 3–7. Springer, 2018.

[110] Liudmila Zavolokina, Mateusz Dolata, and Gerhard Schwabe. Fintech–what's in a name? *Thirty Seventh International Conference on Information Systems*, 2016.

[111] Xiao Ping Steven Zhang and David Kedmey. A budding romance: Finance and ai. *IEEE MultiMedia*, 25(4):79–83, 2018.

[112] Xiao-lin Zheng, Meng-ying Zhu, Qi-bing Li, Chao-chao Chen, and Yan-chao Tan. Finbrain: When finance meets ai 2.0. *Frontiers of Information Technology & Electronic Engineering*, 20(7):914–924, 2019.

[113] Naomie Salim Zuhal Hamad. Systematic literature review (slr) automation: A systematic literature review. *Journal of Theoretical and Applied Information Technology*, 59(3), jan 2014. URL `http://www.jatit.org/volumes/Vol59No3/15Vol59No3.pdf`.

Appendix A

Tool screenshots

A.1 Chrome Plugin

Figure A.1: Chrome plugin - Signin screen.

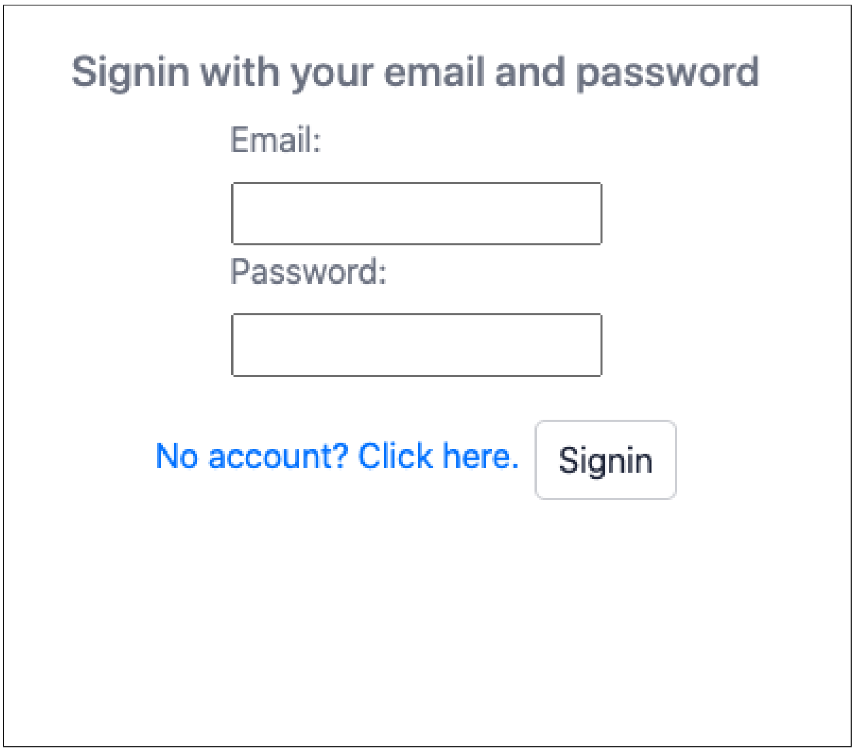


Figure A.2: Chrome plugin - Edit project.

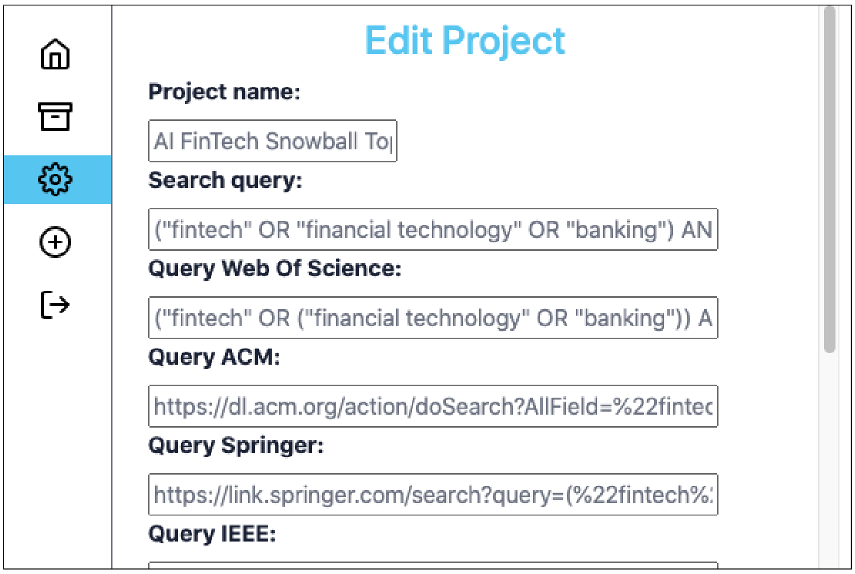
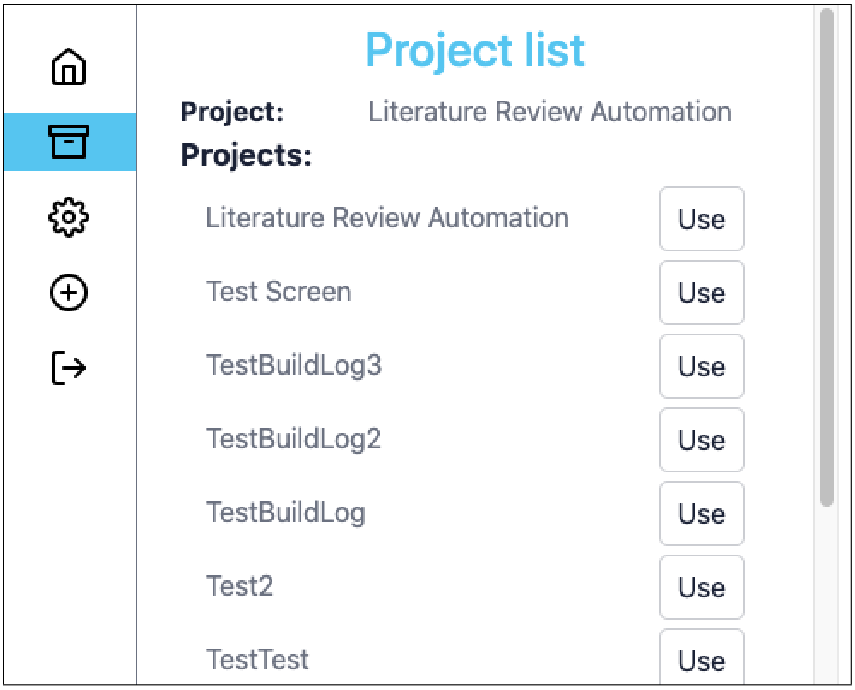


Figure A.3: Chrome plugin - Project overview.



A.2 Dashboard

Figure A.4: Tool - Create account.

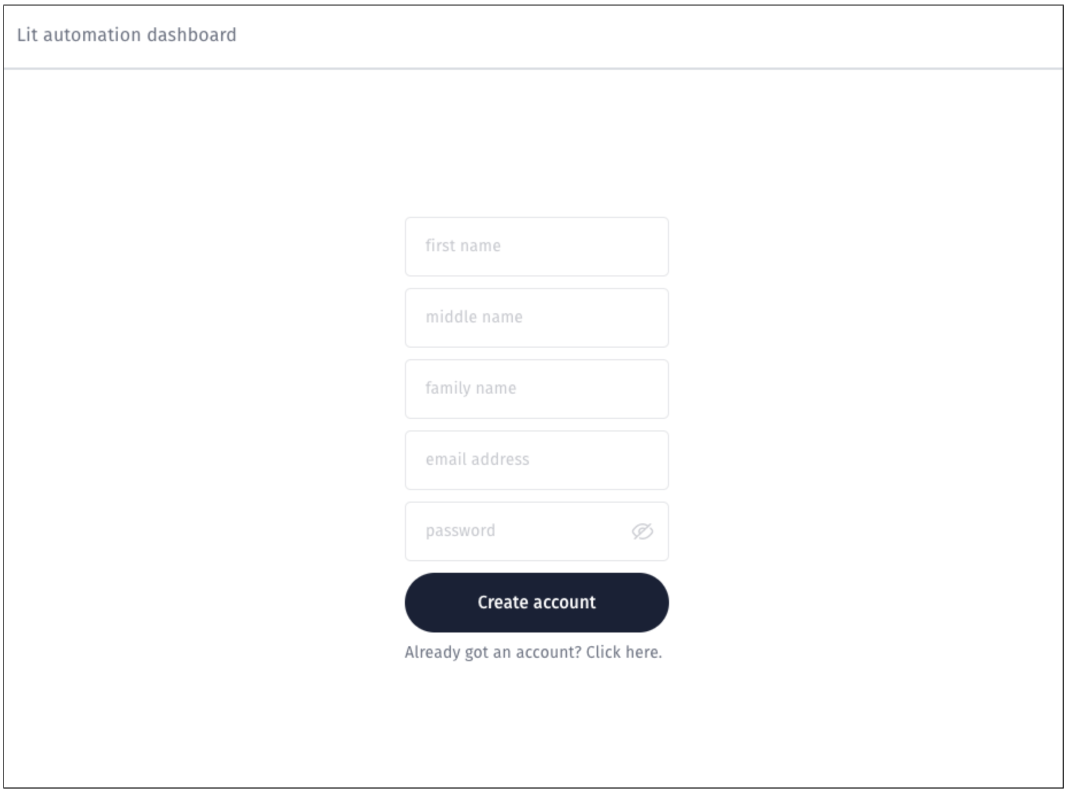


Figure A.5: Tool - Signin.

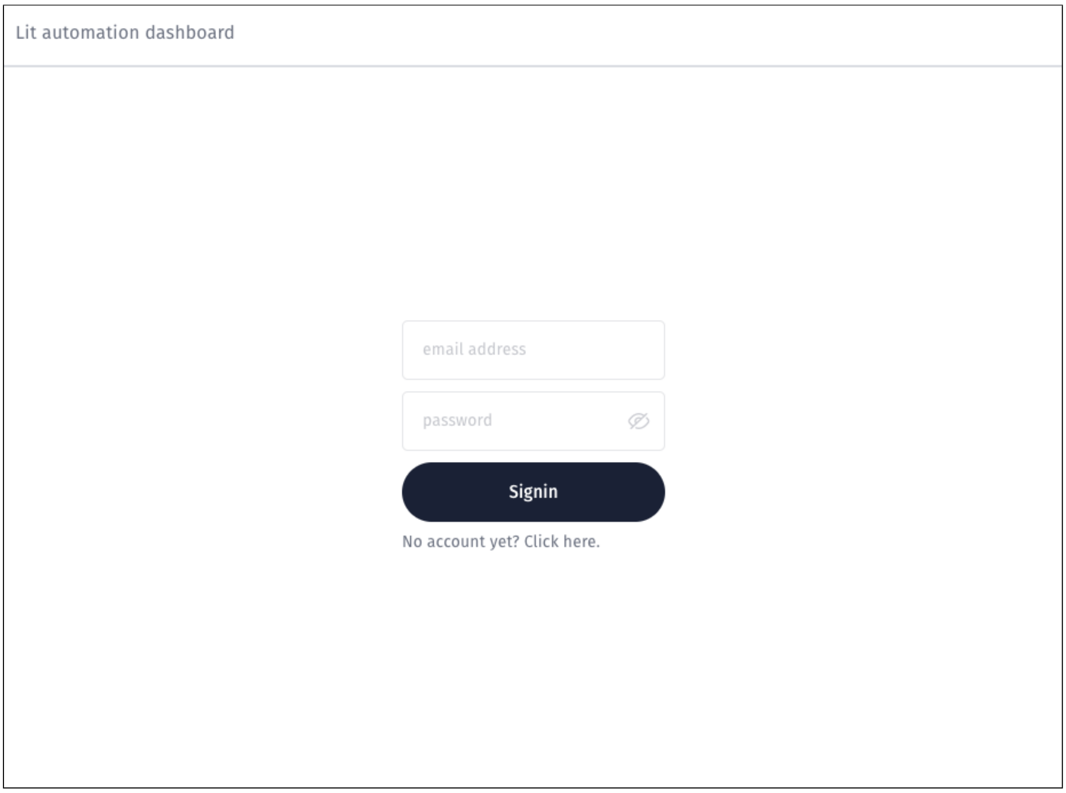


Figure A.6: Tool - Project overview.

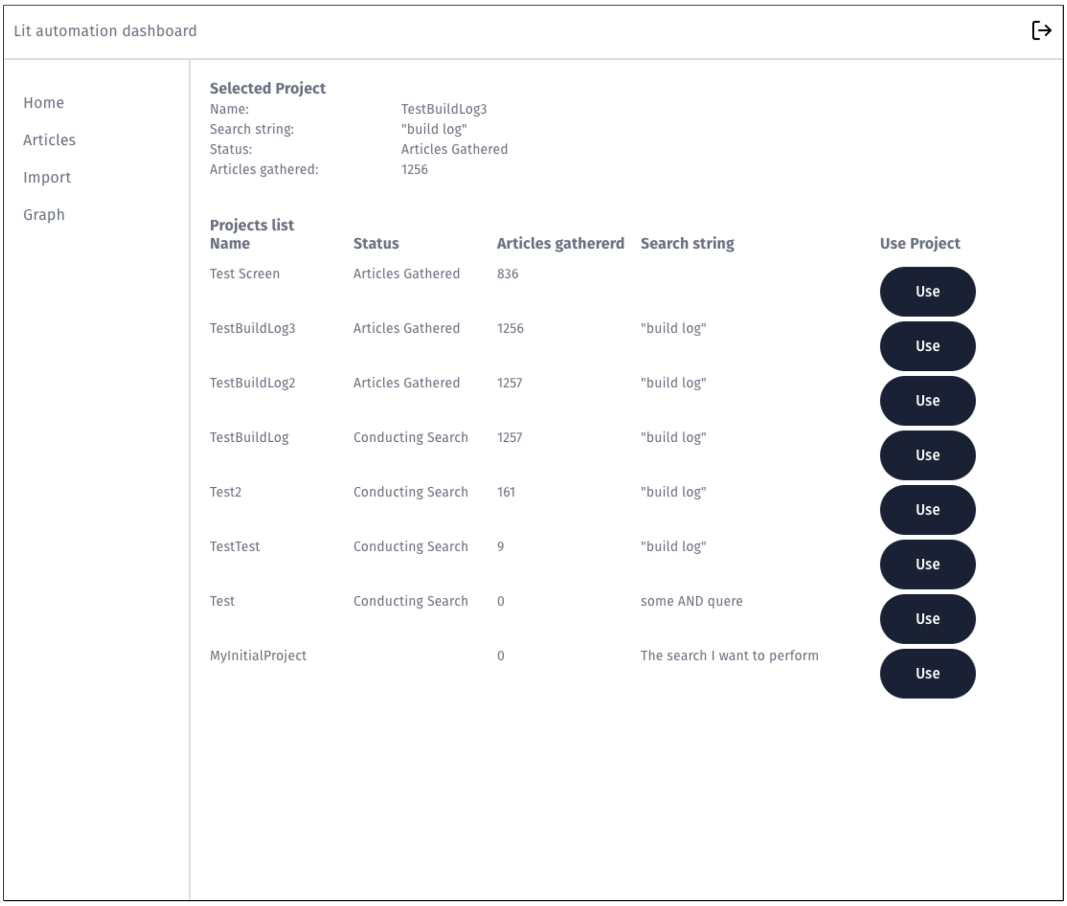


Figure A.7: Tool - Import project.

Figure A.8: Tool - Article edit.

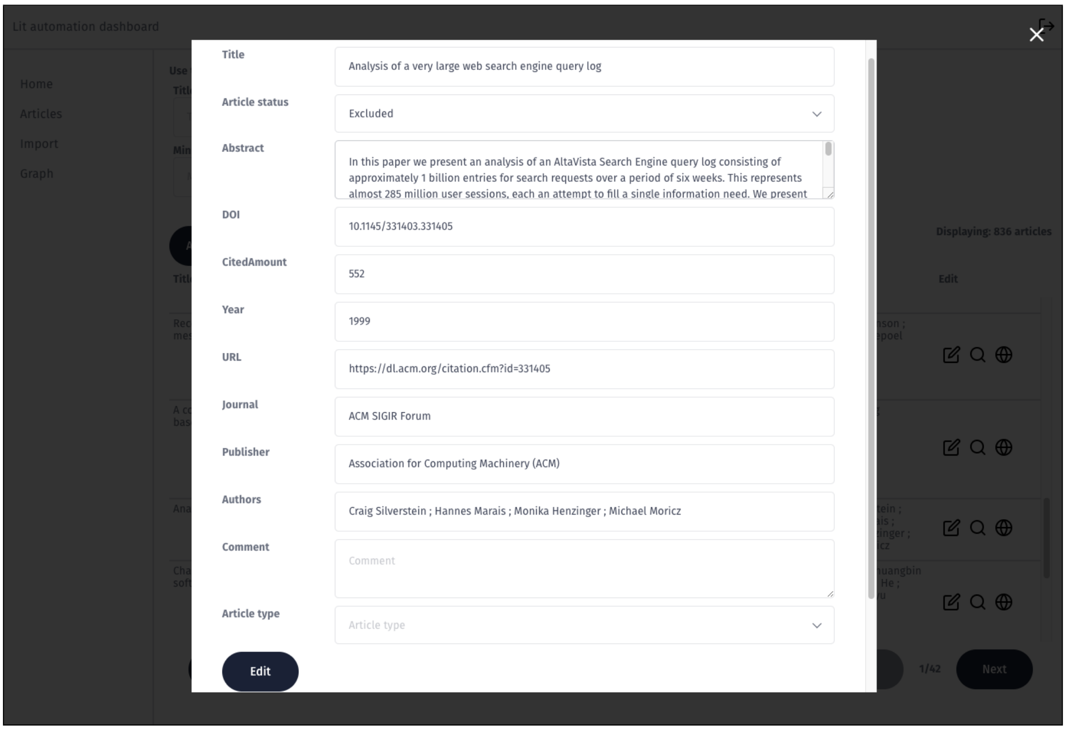


Figure A.9: Tool - Article screening include example.

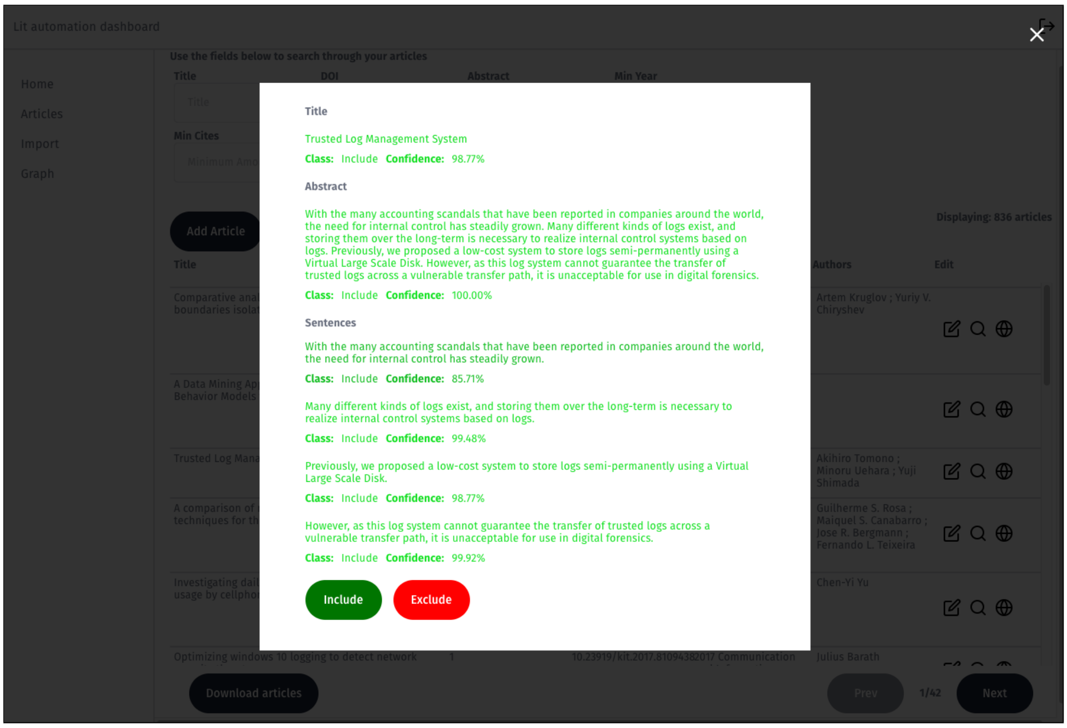
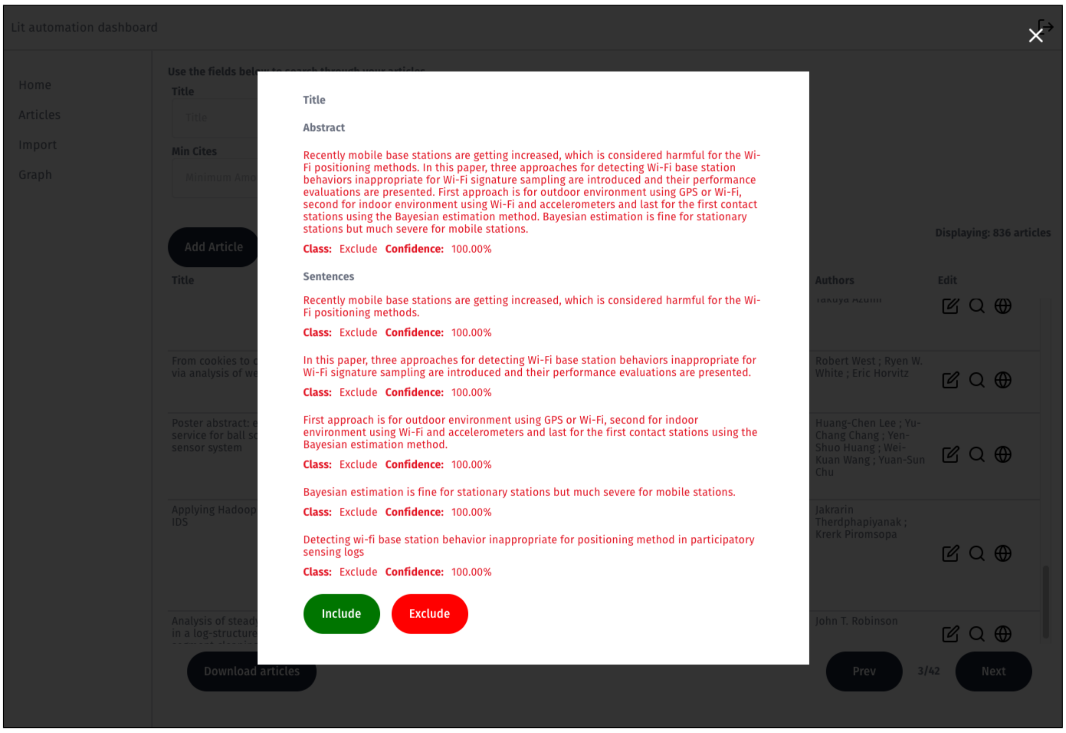


Figure A.10: Tool - Article screening exclude example.



B ML in FinTech References

[114] Youness Abakarim. An Efficient Real Time Model For Credit Card Fraud Detection Based On Deep Learning. *International Conference on Intelligent Systems: Theories and Applications (SITA)*.

[115] Youness Abakarim, Mohamed Lahby, and Abdelbaki Attioui. Towards An Efficient Real-time Approach To Loan Credit Approval Using Deep Learning. *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)*, pages 306–313.

[116] Hussein Abdou, John Pointon, and Ahmed El-Masry. Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking. *Expert Syst. Appl.*, 2008. doi: 10.1016/j.eswa.2007.08.030.

[117] Maher Aburrous and M A Hossain. Associative Classification Techniques for predicting e-Banking Phishing Websites Fadi Thabt ah. *Multimedia Computing and Information Technology (MCIT)*, pages 9–12, 2010. doi: 10.1109/MCIT.2010.5444840.

[118] Maher Ragheb Aburrous, Alamgir Hossain, Fadi Thabatah, and Keshav Dahal. Intelligent Quality Performance Assessment for E-Banking Security Using Fuzzy Logic. *International Conference on Information Technology: New Generations (ITNG)*, 2008. doi: 10.1109/ITNG.2008.154.

[119] Bogdan Adamyk and V Fin. Analysis of Trust in Ukrainian banks based on Machine Learning Algorithms. *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)*, pages 234–239, 2019.

[120] S. Akkoç. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222 (1) (2012), pp. 168-178*, 2012.

[121] Maher Ala'raj and Maysam F. Abbod. Classifiers Consensus System Approach for Credit Scoring. *Know.-Based Syst.*, 2016. doi: 10.1016/j.knosys.2016.04.013.

[122] Lucia Alessi and Carsten Detken. Identifying excessive credit growth and leverage. *Journal of Financial Stability*, 35:215–225, 2018. ISSN 1572-3089. doi: 10.1016/j.jfs.2017.06.005. URL `http://dx.doi.org/10.1016/j.jfs.2017.06.005`.

[123] E. Alfaro, N. García, M. Gámez, and D. Elizondo. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems, 45 (2008), pp. 110-122*, 2008.

[124] S S Alkhasov, A N Tselykh, and A A Tselykh. Application of Cluster Analysis for the Assessment of the Share of Fraud Victims among Bank Card Holders. *SIN '15 Proceedings of the 8th International Conference on Security of Information and Networks*.

[125] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani. A survey of phishing email filtering techniques. *IEEE Communications Surveys & Tutorials, 15(4), 2070–2090*, 2013.

[126] Jon Ander Gomez, Juan Arevalo, Roberto Paredes, and Jordi Nin. End-to-end neural network architecture for fraud scoring in card payments. *Pattern Recognition Letters*, 2018. doi: 10.1016/j.patrec.2017.08.024.

[127] M. E. Antonakis, A. C. Sfakianakis. Assessing naive Bayes as a method for screening credit applicants. *Journal of Applied Statistics*, 2009. doi: 10.1080/02664760802554263.

[128] N. A. G. Arachchilage and S. Love. A game design framework for avoiding phishing attacks. *Computers in Human Behavior, 29(3), 706–714*, 2013.

[129] N. A. G. Arachchilage, S. Love, and K. Beznosov. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior, 60, 185–197*, 2016.

[130] Abdelkbir Armel. Fraud Detection Using Apache Spark. *2019 5th International Conference on Optimization and Applications (ICOA)*, pages 1–6.

[131] G Arutjothi. Prediction of Loan Status in Commercial Bank using Machine Learning Classifier. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, (Iciss):416–419, 2017.

[132] Justice Asare-frempong. Predicting Customer Response to Bank Direct Telemarketing Campaign. *International Conference on Engineering Technology and Technopreneurship (ICE2T)*, pages 3–6, 2017.

[133] Yashodhan Athavale, Pouyan Hosseinizadeh, Sridhar Krishnan, and Aziz Guergachi. Identifying the potential for Failure of Businesses in the Technology, Pharmaceutical and Banking Sectors using Kernel-based Machine Learning Methods. *2009 IEEE International Conference on Systems*, 2009. doi: 10.1109/ICSMC.2009.5345982.

[134] Girija V. Attigeri, M. M. Pai, and Radhika M. Manohara Pai. Credit Risk Assessment Using Machine Learning Algorithms. *Advanced Science Letters*, 2017. doi: 10.1166/asl.2017.9018.

[135] Suresh Yaram Author. Machine Learning Algorithms for Document Clustering and Fraud Detection. *2016 International Conference on Data Science and Engineering (ICDSE)*, pages 1–6, 2016. doi: 10.1109/ICDSE.2016.7823950.

[136] Dmitrii Babaev, Maxim Savchenko, Alexander Tuzhilin, and Dmitrii Umerenkov. ET-RNN: Applying Deep Learning to Credit Loan Applications. *SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019. doi: 10.1145/3292500.3330693.

[137] Arash Bahrammirzaee, Ali Ghatari, Parviz Rajabzadeh Ahmadi, and Kurosh Madani. Hybrid credit ranking intelligent system using expert system and artificial neural networks. *Applied Intelligence*, 2011. doi: 10.1007/s10489-009-0177-8.

[138] Serif Bahtiyar, Mehmet Baris Yaman, and Can Yilmaz Altinigne. A multi-dimensional machine learning approach to predict advanced malware. *Computer Networks*, 2019. doi: 10.1016/j.comnet.2019.06.015.

[139] N Balasupramanian, Ben George Ephrem, and Imad Salim Al-barwani. User Pattern Based Online Fraud Detection and Prevention using Big Data Analytics and Self Organizing Maps. *2017 International Conference on Intelligent Computing*, pages 691–694, 2017.

[140] Anasse Bari, Pantea Peidaee, and Hongting Chen. Predicting Financial Markets Using The Wisdom of Crowds. *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, c:334–340, 2019.

[141] P A Barraclough. Intelligent phishing detection parameter framework for E-banking transactions based on Neuro-fuzzy. *2014 Science and Information Conference*, pages 545–555, 2014. doi: 10.1109/SAI.2014.6918240.

[142] Renzo Barrueta-meza, Jean Paul Castillo-villarreal, and Jimmy Armas-aguirre. Predictive model to determine customer desertion in Peruvian banking entities. *2018 Congreso Internacional de Innovacion y Tendencias en Ingenier'ia (CONIITI)*, pages 1–5, 2018.

[143] Mustafa Bayraktar, Mehmet S Akta, and Oya Kal. Credit Risk Analysis with Classification Restricted Boltzmann Machine. *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pages 3–6. doi: 10.1109/SIU.2018.8404397.

[144] Oleg A Bayuk, Dmitry V Berzin, and Bogdan A Timov. Analysis of the Russian Banking System Stability. *2018 Eleventh International Conference "Management of large-scale system development" (MLSD*, pages 1–4.

[145] Vahid Behbood, Jie Lu, and Guangquan Zhang. Long Term Bank Failure Prediction using Fuzzy Refinement-based Transductive Transfer Learning. *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, 1:2676–2683, 2011. doi: 10.1109/FUZZY.2011.6007633.

[146] A. Belabed, E. Aimeur, and A. Chikh. A personalized whitelist approach for phishing webpage detection. *International Conference on Availability*, 2012. doi: 10.1109/ARES.2012.54.

[147] Daniel Belanche, Luis V. Casalo, and Carlos Flavian. Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*, 2019. doi: 10.1108/IMDS-08-2018-0368.

[148] A.N. Berger and C.H.S. Bouwman. How does capital affect bank performance during financial crises? *Journal of Financial Economics, 109 (2013), pp. 146-176*, 2013.

[149] A. Bergholz, G. Paaß, F. Reichartz, S. Strobel, and J. H. Chang. Improved phishing detection using model based features. *In Proceedings on conference on email and anti-spam (CEAS)*, 2008.

[150] F. Betz, S. Oprica, T.A. Peltonen, and P. Sarlin. Predicting distress in European banks. *Journal of Banking & Finance, 45 (2014), pp. 225-241*, 2014.

[151] Johannes Beutel, Sophia List, and Gregor Von Schweinitz. Does Machine Learning Help us Predict Banking Crises? *Journal of Financial Stability*, page 100693, 2019. ISSN 1572-3089. doi: 10.1016/j.jfs.2019.100693. URL `https://doi.org/10.1016/j.jfs.2019.100693`.

[152] Biplab Bhattacharjee, Amulyashree Sridhar, and Muhammad Shafi. An artificial neural network-based ensemble model for credit risk assessment and deployment as a graphical user interface. *International Journal of Data Mining Modelling And Management*, 2017. doi: 10.1504/IJDMMM.2017.10006638.

[153] Shiivong Birla, Kashish Kohli, and Akash Dutta. Machine Learning on Imbalanced Data in Credit Risk. *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 1–6, 2016. doi: 10.1109/IEMCON.2016.7746326.

[154] Abdelali El Bouchti, Ahmed Chakroun, Hassan Abbar, and Chafik Okar. Fraud Detection in Banking Using Deep Reinforcement Learning. *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, (Intech), 2017.

[155] Melek Acar Boyacioglu, Yakup Kara, and Omer Kaan Baykan. Predicting Bank Financial Failures Using Neural Networks, Support Vector Machines and Multivariate Statistical Methods: A Comparative Analysis in the Sample of Savings Deposit Insurance Fund (SDIF) Transferred Banks in Turkey. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2008.01.003.

[156] K. Buza and D. Neubrandt. How you type is who you are. *11th IEEE International Symposium on Applied Computational Intelligence and Informatics*, 2016. doi: 10.1109/SACI.2016.7507419.

[157] B. Gutiérrez-Nieto C. Serrano-Cinca. Partial least square discriminant analysis for bankruptcy prediction. *Decision Support Systems, 54 (2013), pp. 1245-1255*, 2013.

[158] Guenael Cabanes, Younes Bennani, and Nistor Grozavu. Unsupervised Learning for Analyzing the Dynamic Behavior of Online Banking Fraud. *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 513–520, 2013. doi: 10.1109/ICDMW.2013.109.

[159] Francesco Camastra, Angelo Ciaramella, and Antonino Staiano. Machine learning and soft computing for ICT security: an overview of current trends. *Journal of Ambient Intelligence And Humanized Computing*, 2013. doi: 10.1007/s12652-011-0073-z.

[160] Li-jie Cao, Li-jun Liang, and Zhi-xiang Li. The research on the early-warning system model of operational risk for commercial banks based on bp neural network analysis. *2009 International Conference on Machine Learning and Cybernetics*, 5(July):2739–2744, 2009. doi: 10.1109/ICMLC.2009.5212096.

[161] Pedro Carmona, Francisco Climent, and Alexandre Momparler. Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 2019. doi: 10.1016/j.iref.2018.03.008.

[162] Otavio A S Carpinteiro. Forecasting financial series using clustering methods and support vector regression. *Artificial Intelligence Review*, 52(2):743–773, 2019. ISSN 1573-7462. doi: 10.1007/s10462-018-9663-x. URL `https://doi.org/10.1007/s10462-018-9663-x`.

[163] M. Carr, V. Ravi, G. Reddy, and D. Sridharan Veranna. Machine Learning Techniques Applied to Profile Mobile Banking Users in India. *International Journal of Information Systems in The Service Sector*, 2013. doi: 10.4018/jisss.2013010105.

[164] Jun Chang and Wenting Tu. A Stock-movement Aware Approach for Discovering Investors' Personalized Preferences in Stock Markets. *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 275–280, 2018. doi: 10.1109/ICTAI.2018.00051.

[165] Shiow-Yun Chang and Tsung-Yuan Yeh. An Artificial Immune Classifier for Credit Scoring Analysis. *Applied Soft Computing*, 2012. doi: 10.1016/j.asoc.2011.11.002.

[166] Anusorn Charleonnan. Credit Card Fraud Detection Using RUS and MRN Algorithms. *2016 Management and Innovation Technology International Conference (MITicon)*, pages MIT–73–MIT–76, 2016. doi: 10.1109/MITICON.2016.8025244.

[167] Moitrayee Chatterjee. Detecting Phishing Websites through Deep Reinforcement Learning. *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, 2(1):227–232, 2019. doi: 10.1109/COMPSAC.2019.10211.

[168] Ching-Hsue Chen, You-Shyang Cheng. Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. *Knowledge-based Systems*, 2013. doi: 10.1016/j.knosys.2012.11.004.

[169] Jou-fan Chen, Wei-lun Chen, and Chun-ping Huang. Financial Time-series Data Analysis using Deep Convolutional Neural Networks. *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, pages 87–92, 2016. doi: 10.1109/CCBD.2016.027.

[170] Meixuan Chen. Domain Adaptation Approach for Credit Risk Analysis Preprocessing Data Mining Result Validation. *International Conference on Software Engineering and Information Management*, (315):104–107, 2018.

[171] Ya-qi Chen, Jianjun Zhang, and Wing W Y Ng. Loan default prediction using diversified sensitivity undersampling. *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, 1:240–245.

[172] Yuzhou Chen, Yulia R. Gel, Vyacheslav Lyubchich, and Todd Winship. Deep Ensemble Classifiers and Peer Effects Analysis for Churn Forecasting in Retail Banking. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2018. doi: 10.1007/978-3-319-93034-3_30.

[173] Dawei Cheng, Zhibin Niu, Yi Tu, and Liqing Zhang. Prediction Defaults for Networked-guarantee Loans. *International Conference on Pattern Recognition (ICPR)*, 2018.

[174] Frank Cheng and Michael P Wellman. Accounting for Strategic Response in an Agent-Based Model of Financial Regulation. *Conference on Economics and Computation*, pages 187–203, 2017.

[175] Tun-kung Cheng. AI Robo-Advisor with Big Data Analytics for Financial Services. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1027–1031, 2018. doi: 10.1109/ASONAM.2018.8508854.

[176] Jiun-Yao Chiua, Yan Yan, Gao Xuedongb, and Rung-Ching Chen. A New Method for Estimating Bank Credit Risk. *International Conference on Technologies and Applications of Artificial Intelligence*, 2010. doi: 10.1109/TAAI.2010.85.

[177] Tsung-nan Chou. A Novel Prediction Model for Credit Card Risk Management. *Second International Conference on Innovative Computing*, pages 0–3, 2007.

[178] Mariola Chrzanowska, Esteban Alfaro, and Dorota Witkowska. The Individual Borrowers Recognition: Single and Ensemble Trees. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2008.07.048.

[179] Anja Cielen, Ludo Peeters, and Koen Vanhoof. Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, 154:526–532, 2004. doi: 10.1016/S0377-2217(03)00186-3.

[180] claudio Alexandre and Joao Balsa. A Multiagent Based Approach to Money Laundering Detection and Prevention. *International Conference on Agents and Artificial Intelligence*, 2015. doi: 10.5220/0005281102300235.

[181] S. Cleary and G. Hebb. An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance, 64 (2016), pp. 101-111*, 2016.

[182] Francisco Climent, Alexandre Momparler, and Pedro Carmona. Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 2019. doi: 10.1016/j.jbusres.2018.11.015.

[183] Francesco Corea. How AI Is Transforming Financial Services. *Applied Artificial Intelligence: Where AI Can Be Used in Business*, 2019. doi: 10.1007/978-3-319-77252-3_3.

[184] Ivo Correia and Inna Skarbovsky. Industry Paper : The Uncertain Case of Credit Card Fraud Detection. *The 9th ACM International Conference on Distributed Event-Based Systems (DEBS15)*, pages 181–192, 2015.

[185] Kelton A P Costa, Luis A Silva, Guilherme B Martins, Gustavo H Rosa, Clayton R Pereira, and P Papa. Malware Detection in Android-based Mobile Environments using Optimum-Path Forest. *International Conference on Machine Learning and Applications (ICMLA)*, 2015. doi: 10.1109/ICMLA.2015.72.

[186] Adrian Costea. Applying fuzzy logic and machine learning techniques in financial performance predictions. *International Conference on Applied Statistics (ICAS)*, 2014. doi: 10.1016/S2212-5671(14)00271-8.

[187] Alfredo Cuzzocrea, Fabio Martinelli, and Francesco Mercaldo. Applying Machine Learning Techniques to Detect and Analyze Web Phishing Attacks. *International Conference on Information Integration and Web-based Applications & Services (iiWAS)*.

[188] Ansar Daghouri, Khalifa Mansouri, and Mohammed Qbadou. Towards a decision support system, based on the systemic and multi-agent approaches for organizational performance evaluation of a risk management unit: Banks case. *Intelligent Systems and Computer Vision (ISCV)*, 2017.

[189] Marc Damez, Marie-jeanne Lesot, Adrien Revault, Marc Damez, Marie-jeanne Lesot, Adrien Revault, Dynamic Credit-card Fraud, Marc Damez, Marie-jeanne Lesot, and Adrien Revault. Dynamic Credit-Card Fraud Profiling To cite this version : HAL Id : hal-01282301 Dynamic Credit-Card Fraud Profiling. *International conference on Modeling Decisions for Artificial Intelligence*, 2017.

[190] Emad Abd Elaziz Dawood, Essamedean Elfakhrany, and Fahima A. Maghraby. Improve Profiling Bank Customer's Behavior Using Machine Learning. *IEEE ACCESS*, 2019. doi: 10.1109/ACCESS.2019.2934644.

[191] M. Day and C. Lee. Deep learning for financial sentiment analysis on finance news providers. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016. doi: 10.1109/ASONAM.2016.7752381.

[192] Simon Delecourt. Building a robust mobile payment fraud detection system with adversarial examples. *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 103–106, 2019. doi: 10.1109/AIKE.2019.00026.

[193] Y. Demyanyk and I. Hasan. Financial crises and bank failures: A review of prediction methods. *Omega, 38 (2010), pp. 315-324*, 2010.

[194] Dhanashree Deshpande, Shrinivas Deshpande, and Vilas Thakare. Analysis of Online Suspicious Behavior Patterns. *Ambient Communications And Computer Systems*, 2018. doi: 10.1007/978-981-10-7386-1_42.

[195] C. Lakshmi Devasena. Adeptness Evaluation of Memory Based Classifiers for Credit Risk Analysis. *2014 International Conference on Intelligent Computing Applications (ICICA 2014)*, 2014. doi: 10.1109/ICICA.2014.39.

[196] K Bavithra Devi. Deep Learn Helmets-Enhancing Security at ATMs. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 1111–1116, 2019.

[197] R. DeYoung and G. Torna. Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation, 22 (3) (2013), pp. 397-421*, 2013.

[198] Cees Diks, Cars Hommes, Valentyn Panchenko, and Roy Weide. EF Chaos: A User Friendly Software Package for Nonlinear Economic Dynamics. *Computational Economics*, 2008. doi: 10.1007/s10614-008-9130-x.

[199] Alina Mihaela Dima and Simona Vasilache. ANN Model for Corporate Credit Risk Assessment. *2009 International Conference on Information and Financial Engineering*, pages 94–98, 2009. doi: 10.1109/ICIFE.2009.33.

[200] Constantin Doumpos, Michael Zopounidis. Monotonic support vector machines for credit risk rating. *New Mathematics And Natural Computation*, 2009. doi: 10.1142/S1793005709001520.

[201] C R Durga. A Relative Evaluation of the Performance of Ensemble Learning in Credit Scoring. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pages 161–165, 2016. doi: 10.1109/ICACA.2016.7887943.

[202] Sara Sadat Ebadati, Omid Mahdi E. Babaie. Implementation of Two Stages k-Means Algorithm to Apply a Payment System Provider Framework in Banking Systems. *Artificial Intelligence Perspectives And Applications (CSOC2015)*, 2015. doi: 10.1007/978-3-319-18476-0_21.

[203] Aykut Ekinci and Halil Ibrahim Erdal. Forecasting Bank Failure: Base Learners, Ensembles and Hybrid Ensembles. *Comput. Econ.*, 2017. doi: 10.1007/s10614-016-9623-y.

[204] William Enck, Damien Octeau, Patrick McDaniel, , and Swarat Chaudhuri. A study of android application security. *In Proceedings of the 20th USENIX Security Symposium, 2011*, 2011.

[205] Halil Ibrahim Erdal and Aykut Ekinci. A Comparison of Various Artificial Intelligence Methods in the Prediction of Bank Failures. *Computational Economics*, 2013. doi: 10.1007/s10614-012-9332-0.

[206] Ilhami Erdal, Hamit Karahanoglu. Bagging ensemble models for bank profitability: An emprical research on Turkish development and investment banks. *Applied Soft Computing*, 2016. doi: 10.1016/j.asoc.2016.09.010.

[207] Birsen Eygi Erdogan. Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *Journal of Statistical Computation And Simulation*, 2013. doi: 10.1080/00949655.2012.666550.

[208] B. B. Gupta et al. Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, 2016.

[209] Qi Fan. A Denoising Autoencoder Approach for Credit Risk Analysis. *International Conference on Computing and Artificial Intelligence (ICCAI)*, 2018.

[210] Shuoshuo Fan, Guohua Liu, and Zhao Chen. Anomaly detection methods for bankruptcy prediction. *2017 4th International Conference on Systems and Informatics (ICSAI)*, (17):1456–1460, 2017.

[211] Mark Fenwick, Wulf A. Kaal, and Erik P. M. Vermeulen. Regulation Tomorrow: Strategies for Regulating New Technologies. *Transnational Commercial And Consumer Law: Current Trends in International Business Law*, 2018. doi: 10.1007/978-981-13-1080-5_6.

[212] Mohammed Nazim Feroz and Susan Mengel. Examination of Data , Rule Generation and Detection of Phishing URLs using Online Logistic Regression. *2014 IEEE International Conference on Big Data (Big Data)*, pages 241–250, 2014. doi: 10.1109/BigData.2014.7004239.

[213] S. Finlay. Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research, 210 (2) (2011), pp. 368-378*, 2011.

[214] Pier Giuseppe Fioribello, Simone Giribone. Design of an Artificial Neural Network battery for an optimal recognition of patterns in financial time series. *International Journal of Financial Engineering*, 2018. doi: 10.1142/S2424786318500317.

[215] Alma Lilia Garc. Understanding bank failure : A close examination of rules created by Genetic Programming. *Electronics*, 2010. doi: 10.1109/CERMA.2010.14.

[216] Gulefsan Bozkurt Gonen, Mehmet Gonen, and Fikret Gurgen. Probabilistic and Discriminative Group-wise Feature Selection Methods for Credit Risk Analysis. *Expert Syst. Appl.*, 2012. doi: 10.1016/j.eswa.2012.04.050.

[217] Israel Luis Gonzalez-Carrasco, Jose Luis Jimenez-Marquez, Jose Lopez-Cuadrado, and Belen Ruiz-Mezcua. Automatic detection of relationships between banking operations using machine learning. *Information Sciences*, 2019. doi: 10.1016/j.ins.2019.02.030.

[218] Hakan Gunduz. Finansal Haberler Kullanılarak Derin grenme ile Borsa Tahmini Stock Market Prediction with Deep Learning Using Financial News. *Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2018. doi: 10.1109/SIU.2018.8404616.

[219] Chaonian Guo, Hao Wang, Hong-Ning Dai, Shuhan Cheng, and Tongsen Wang. Fraud Risk Monitoring System for E-Banking Transactions. *Intl Conf on Dependable*, 2018. doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00030.

[220] B. B. Gupta, Nalin A. G. Arachchilage, and Kostas E. Psannis. Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems*, 2018. doi: 10.1007/s11235-017-0334-z.

[221] P. A. Gutierrez, M. J. Segovia-Vargas, S. Salcedo-Sanz, C. Hervas-Martinez, A. Sanchis, J. A. Portilla-Figueras, and F. Fernandez-Navarro. Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises. *Omega-international Journal of Management Science*, 2010. doi: 10.1016/j.omega.2009.11.001.

[222] Nana Kwame Gyamfi. Bank Fraud Detection Using Support Vector Machine. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 37–41, 2018.

[223] Firoozaeh Hajialiakbari and Mohamad H Gholami. Assessment of the effect on technical efficiency of bad loans in banking industry: a principal component analysis and neuro-fuzzy system. *Neural Computing and Applications*, 23, 2013. doi: 10.1007/s00521-013-1413-z.

[224] J. Hong. The state of phishing attacks. *Communications of the ACM, 55(1), 74–81*, 2012.

[225] Pei-ying Hsu and An-pin Chen. A Market Making Quotation Strategy Based on Dual Deep Learning Agents for Option Pricing and Bid-Ask Spread Estimation. *2018 IEEE International Conference on Agents (ICA)*, pages 99–104, 2018. doi: 10.1109/AGENTS.2018.8460084.

[226] Xin-yue Hu and Yong-li Tang. Ann-based credit risk identificaion and control for. *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, (August):13–16, 2006.

[227] Cheng-Lung Huang, Mu-Chen Chen, and Chieh-Jen Wang. Credit Scoring with a Data Mining Approach Based on Support Vector Machines. *Expert Syst. Appl.*, 2007. doi: 10.1016/j.eswa.2006.07.007.

[228] H. Huang, C. Zheng, J. Zeng, W. Zhou, S. Zhu, P. Liu, S. Chari, and C. Zhang. Android malware development on public malware scanning platforms: A large-scale data-driven study. *2016 IEEE International Conference on Big Data (Big Data)*, 2016. doi: 10.1109/BigData.2016.7840712.

[229] Shian-chang Huang. Credit Quality Assessments Using Manifold Based Semi-supervised Discriminant Analysis and Support Vector Machines. *2011 Seventh International Conference on Natural Computation*, 4:2037–2041, 2011. doi: 10.1109/ICNC.2011.6022386.

[230] Ming hui Jiang and Xu chuan Yuan. Personal Credit Scoring Model of Non-linear Combining Forecast Based on GP. *Third International Conference on Natural Computation*, 2007. doi: 10.1109/ICNC.2007.551.

[231] Walid Hussein, Mostafa A. Salama, and Osman Ibrahim. Image Processing Based Signature Verification Technique to Reduce Fraud in Financial Institutions. *International Conference on Circuits*, 2016. doi: 10.1051/matecconf/20167605004.

[232] Zsolt Ilonczai. Echo State Network-based Credit Rating System. *2012 4th IEEE International Symposium on Logistics and Industrial Informatics*, 2012. doi: 10.1109/LINDI.2012.6319485.

[233] Jemal Islam, Rafiqul Abawajy. A multi-tier phishing detection and filtering approach. *Journal of Network And Computer Applications*, 2013. doi: 10.1016/j.jnca.2012.05.009.

[234] A. K. Jain and B. B. Gupta. A novel approach to protect against phishing attacks at client side using auto-updated white-list. *EURASIP Journal on Information Security*, 2016.

[235] Mohammad Behdad Jamshidi. A Novel Multiobjective Approach for Detecting Money Laundering with a Neuro-Fuzzy Technique. *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, pages 454–458, 2019.

[236] Ming-hui Jiang and Jian-hua Hu. Combining multiple classifiers based on Dempster-Shafer theory for personal credit scoring. *2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings*, 1(1):167–172, 2014. doi: 10.1109/ICMSE.2014.6930225.

[237] S. Jones, D. Johnstone, and R. Wilson. An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance, 56 (2015), pp. 72-85*, 2015.

[238] Piotr Juszczak, Niall M. Adams, David J. Hand, Christopher Whitrow, and David J. Weston. Off-the-peg and Bespoke Classifiers for Fraud Detection. *Computational Statistics & Data Analysis*, 2008. doi: 10.1016/j.csda.2008.03.014.

[239] Archana S. Kadam and S. S. Pawar. Comparison of association rule mining with pruning and adaptive technique for classification of phishing. *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*.

[240] Rutvik Kakadiya. AI Based Automatic Robbery / Theft Detection using Smart Surveillance in Banks. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 201–204, 2019.

[241] Ehsan Kamalloo and Mohammad Saniee Abadeh. Credit Risk Prediction Using Fuzzy Immune Learning. *Adv. Fuzzy Sys.*, 2014, 2014.

[242] Elias Kamos, Foteini Matthaiou, and Sotiris Kotsiantis. Credit Rating Using a Hybrid Voting Ensemble. *Hellenic conference on Artificial Intelligence*, pages 165–173, 2012.

[243] Y. Kang, R. Cui, J. Deng, and N. Jia. A novel credit scoring framework for auto loan using an imbalanced-learning-based reject inference. *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 2019. doi: 10.1109/CIFEr.2019.8759110.

[244] T. Prem Kanimozhi, V Jacob. Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. *ICT EXPRESS*, 2019. doi: 10.1016/j.icte.2019.03.003.

[245] Murat Karabatak and Twana Mustafa. Performance Comparison of Classifiers on Reduced Phishing Website Dataset. *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, pages 1–5, 2018. doi: 10.1109/ISDFS.2018.8355357.

[246] Mehmet Fatih Kartal, Cem Bayramoglu. What Are Relations Between the Domestic Macroeconomic Variables and the Convertible Exchange Rates? *Global Approaches in Financial Economics*, 2018. doi: 10.1007/978-3-319-78494-6_22.

[247] Zahra Kazemi and Houman Zarrabi. Using deep networks for fraud detection in the credit card transactions. *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pages 630–633, 2017.

[248] Vivek Kedia. Portfolio Generation for Indian Stock Markets using Unsupervised Machine Learning. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–5.

[249] Amir E. Khandani, Adlar J. Kim, and Andrew W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 2010. doi: 10.1016/j.jbankfin.2010.06.001.

[250] A. Khashman. Neural networks for credit risk evaluation: investigation of different neural models and learning schemes. *Expert Systems with Applications, 37 (2010), pp. 6233-6239*, 2010.

[251] Adel Khelifi. Enhancing Protection Techniques of E-Banking Security Services Using Open Source Cryptographic Algorithms. *2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pages 89–95, 2013. doi: 10.1109/SNPD.2013.47.

[252] M. Khonji, Y. Iraqi, and A. Jones. Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials, 15(4), 2091–2121*, 2013.

[253] Mahmoud Khonji and Andrew Jones. A Study of Feature Subset Evaluators and Feature Subset Searching Methods for Phishing Classification. *Proceedings of the 8th Annual Collaboration*.

[254] Jinhwa Kim, Kook Jae Hwang, and Jae Kwon Bae. Prediction of Personal Credit Rates with Incomplete Data Sets Using Cognitive Mapping. *International Conference on Convergence Information Technology*, 2007. doi: 10.1109/ICCIT.2007.310.

[255] Kee-hoon Kim, Chang-seok Lee, and Sang-muk Jo. Predicting the Success of Bank Telemarketing using Deep Convolutional Neural Network. *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pages 314–317, 2015. doi: 10.1109/SOCPAR.2015.7492828.

[256] I. Kirlappos and M. A. Sasse. Security education against phishing: A modest proposal for a major rethink. *IEEE Security and Privacy Magazine, 10(2), 24–32*, 2012.

[257] Timo Klerx, Maik Anderka, and Hans Kleine Buning. On the Usage of Behavior Models to Detect ATM Fraud. *ECAI'14 Proceedings of the Twenty-first European Conference on Artificial Intelligence*, 2014. doi: 10.3233/978-1-61499-419-0-1045.

[258] Bay Kliman, Russ Arinze. Cognitive Computing: Impacts on Financial Advice in Wealth Management. *Aligning Business Strategies and Analytics*, 2019. doi: 10.1007/978-3-319-93299-6_2.

[259] S. B. Kotsiantis, D. Kanellopoulos, V. Karioti, and V. Tampakas. An ontology-based portal for credit risk analysis. *2009 2ND IEEE International Conference on Computer Science And Information Technology*, 2009. doi: 10.1109/ICCSIT.2009.5234452.

[260] Sotiris Kotsiantis, Dimitris Kanellopoulos, Vasiliki Karioti, and Vasilis Tampakas. On Implementing an Ontology-Based Portal for Intelligent Bankruptcy Prediction. *International Conference on Information and Financial Engineering*, 2009. doi: 10.1109/ICIFE.2009.21.

[261] Gang Kou, Xiangrui Chao, Yi Peng, Fawaz E. Alsaadi, and Enrique Herrera-Viedma. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 2019. doi: 10.3846/tede.2019.8740.

[262] Sai Sreewathsa Kovalluri. LSTM Based Self-Defending AI Chatbot Providing Anti-Phishing. *RESEC '18 Proceedings of the First Workshop on Radical and Experiential Security Pages 49-56*, pages 49–56, 2018.

[263] Mehmet Ufuk Kultur, Yigit Caglayan. A Novel Cardholder Behavior Model for Detecting Credit Card Fraud. *Intelligent Automation & Soft Computing*, 2018. doi: 10.1080/10798587.2017.1342415.

[264] Yigit Kultur and Mehmet Ufuk Caglayan. A Novel Cardholder Behavior Model for Detecting Credit Card Fraud. *2015 9th International Conference on Application of Information and Communication Technologies (AICT)*, pages 148–152. doi: 10.1109/ICAICT.2015.7338535.

[265] P.R. Kumar and V. Ravi. Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review. *European Journal of Operational Research, 180 (1) (2007), pp. 1-28*, 2007.

[266] Sangeet Kumar. Fuzzy Logic Based Decision Support System for Loan Risk Assessment. *ACAI '11 Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*, pages 179–182, 1920.

[267] Vinod Kumar L, S Natarajan, S Keerthana, K M Chinmayi, and N Lakshmi. Credit Risk Analysis in Peer-to-Peer Lending System. *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)*, pages 193–196, 2016. doi: 10.1109/ICKEA.2016.7803017.

[268] Piotr Ladyzynski, Kamil Zbikowski, and Piotr Gawrysiak. Direct marketing campaigns in retail banking with the use of deep learning and random forests. *Expert Syst. Appl.*, 2019. doi: 10.1016/j.eswa.2019.05.020.

[269] Marek Laskowski and Henry M Kim. Rapid Prototyping of a Text Mining Application for Cryptocurrency Market Intelligence. *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 448–453, 2016. doi: 10.1109/IRI.2016.66.

[270] Rana M Amir Latif, Muhammad Umer, Tayyaba Tariq, Muhammad Farhan, Osama Rizwan, and Ghazanfar Ali. A Smart Methodology for Analyzing Secure E- Banking and E-Commerce Websites. *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pages 589–596, 2019.

[271] Raymond S T Lee. Chaotic Interval Type-2 Fuzzy Neuro-oscillatory Network (CIT2- FNON) for Worldwide 129 Financial Products Prediction. *International Journal of Fuzzy Systems*, 21(7):2223–2244, 2019. ISSN 2199-3211. doi: 10.1007/s40815-019-00688-w. URL `https://doi.org/10.1007/s40815-019-00688-w`.

[272] Y.C. Lee. Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications, 33 (2007), pp. 67-741*, 2007.

[273] K M Leung. Discrete versus Continuous Parametrization of Bank Credit Rating Systems Optimization Using Differential Evolution. *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 265–272, 2010.

[274] Kevin Leung, France Cheong, and Christopher Cheong. Consumer Credit Scoring using an Artificial Immune System Algorithm. *2007 IEEE Congress on Evolutionary Computation*, pages 3377–3384, 2007.

[275] Aihua Li, Yong Shi, Meihong Zhu, and Jingran Dai. A Data Mining Approach to Classify Credit Cardholders' Behavior. *Sixth IEEE International Conference on Data Mining - Workshops*, 2006. doi: 10.1109/ICDMW.2006.4.

[276] Rong-zhou Li and Su-lin Pang. Neural network credit-risk evaluation model based on back-propagation algorithm. *International Conference on Machine Learning and Cybernetics*, (November):4–5, 2002.

[277] Wei Li, Shuai Ding, Yi Chen, Hao Wang, and Shanlin Yang. Transfer Learning-based Default Prediction Model for Consumer Credit in China. *J. Supercomput.*, 2019. doi: 10.1007/s11227-018-2619-8.

[278] Li-jun Liang, Li-jie Cao, and Zhi-xiang Li. Research on operational risk measurement for commercial bank based on credal network. *2009 International Conference on Machine Learning and Cybernetics*, 5(July):2634–2640, 2009. doi: 10.1109/ICMLC.2009.5212128.

[279] Li-Jun Liang, Fan-Chen Meng, and Li-Jie Cao. Research on affecting factors of operational risk management for commercial bank based on structural equation model. *2012 International Conference on Machine Learning and Cybernetics*, 2012. doi: 10.1109/ICMLC.2012.6359007.

[280] Caterina Liberati, Furio Camillo, and Gilbert Saporta. Advances in credit scoring: combining performance and interpretation in kernel discriminant analysis. *Advances in Data Analysis and Classification*, 2017. doi: 10.1007/s11634-015-0213-y.

[281] F. Liebana-Cabanillas, R. Nogueras, L. J. Herrera, and A. Guillen. Analysing user trust in electronic banking using data mining methods. *Expert Syst. Appl.*, 2013. doi: 10.1016/j.eswa.2013.03.010.

[282] J. Lin, Z. Zhang, J. Zhou, X. Li, J. Fang, Y. Fang, Q. Yu, and Y. Qi. NetDP: An Industrial-Scale Distributed Network Representation Framework for Default Prediction in Ant Credit Pay. *2018 IEEE International Conference on Big Data (Big Data)*, 2018. doi: 10.1109/BigData.2018.8622169.

[283] Shu-ling Lin, Shun-jyh Wu, Hsiu-lan Ma, and Der-bang Wu. Development of credit risk model in banking industry based on gra. *2009 International Conference on Machine Learning and Cybernetics*, 5(July):2903–2909, 2009. doi: 10.1109/ICMLC.2009.5212585.

[284] Jiayong Liu, Zhiyi Tian, Rongfeng Zheng, and Liang Liu. A Distance-Based Method for Building an Encrypted Malware Traffic Identification Framework. *IEEE ACCESS*, 2019. doi: 10.1109/ACCESS.2019.2930717.

[285] Xuan Liu and Pengzhu Zhang. A Scan Statistics Based Suspicious Transactions Detection Model for Anti-money Laundering (AML) in Financial Institutions. *2010 International Conference on Multimedia Communications*, 2010. doi: 10.1109/MEDIACOM.2010.37.

[286] Y. Liu, A. Ghandar, and G. Theodoropoulos. A Metaheuristic Strategy for Feature Selection Problems: Application to Credit Risk Evaluation in Emerging Markets. *Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 2019. doi: 10.1109/CIFEr.2019.8759117.

[287] Zining Liu, Chong Long, Xiaolu Lu, Zehong Hu, J I E Zhang, and Yafang Wang. Which Channel to Ask My Question ?: Personalized Customer Service Request Stream Routing Using Deep Reinforcement Learning. *IEEE ACCESS*, 7, 2019.

[288] Ioannis E. Livieris, Niki Kiriakidou, Andreas Kanavos, Vassilis Tampakas, and Panagiotis Pintelas. On Ensemble SSL Algorithms for Credit Scoring Problem. *Informatics-basel*, 2018. doi: 10.3390/informatics5040040.

[289] Jorge Lopez Lazaro, Alvaro Barbero Jimenez, and Akiko Takeda. Improving cash logistics in bank branches by coupling machine learning and robust optimization. *Expert Syst. Appl.*, 2018. doi: 10.1016/j.eswa.2017.09.043.

[290] W. Lu and D.A. Whidbee. Bank structure and failure during the financial crisis. *Journal of Financial Economic Policy, 5 (3) (2013), pp. 281-299*, 2013.

[291] Xiao-Yong Lu, Xiao-Qiang Chu, Meng-Hui Chen, Pei-Chann Chang, and Shih-Hsin Chen. Artificial Immune Network with Feature Selection for Bank Term Deposit Recommendation. *Journal of Intelligent Information Systems*, 2016. doi: 10.1007/s10844-016-0399-2.

[292] Ling Luo, Xiang Ao, Feiyang Pan, Jin Wang, Tong Zhao, Ningzi Yu, and Qing He. Beyond polarity: Interpretable financial sentiment analysis with hierarchical query-driven attention. In *IJCAI*, pages 4244–4250, 2018.

[293] Thai-le Luong, Minh-son Cao, Duc-thang Le, and Xuan-hieu Phan. Intent Extraction from Social Media Texts Using Sequential Segmentation and Deep Learning Models. *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, (978):215–220, 2017.

[294] Haiying Ma. Credit risk evaluation based on artificial intelligence technology. *2010 International Conference on Artificial Intelligence and Computational Intelligence*, 1:200–203, 2010. doi: 10.1109/AICI.2010.48.

[295] Haiying Ma. Customer segmentation for B2C e-commerce websites based on the Generalized association rules and decision tree. *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pages 4600–4603, 2011. doi: 10.1109/AIMSEC.2011.6010255.

[296] Shenglan Ma, Lingling Yang, Hao Wang, Hong Xiao, Hong-Ning Dai, Shuhan Cheng, and Tongsen Wang. MHDT: A Deep-Learning-Based Text Detection Algorithm for Unstructured Data in Banking. *ICMLC 2019: 2019 11TH International Conference on Machine Learning Andcomputing*, 2019. doi: 10.1145/3318299.3318327.

[297] Georgios Manthoulis, Michalis Doumpos, Constantin Zopounidis, and Emilios Galariotis. An ordinal classification framework for bank failure prediction : Methodology and empirical evidence for US banks. *European Journal of Operational Research*, (xxxx), 2019. doi: 10.1016/j.ejor.2019.09.040.

[298] P Marikkannu. Classification of customer credit data for intelligent credit scoring system using fuzzy set and mc2 – domain driven approach. *2011 3rd International Conference on Electronics Computer Technology*, 3:410–414. doi: 10.1109/ICECTECH.2011.5941782.

[299] Amparo Marin-de-la barcena, Alexis Marcano-cedeo, Juan Jimenez-trillo, Juan A Piuela, and Diego Andina. Artificial Metaplasticity : an Approximation to Credit Scoring modeling. *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pages 2817–2822, 2010. doi: 10.1109/IECON.2010.5675097.

[300] Yannis Marinakis, Magdalene Marinaki, Michael Doumpos, Nikolaos Matsatsinis, and Constantin Zopounidis. Optimization of Nearest Neighbor Classifiers via Metaheuristic Algorithms for Credit Risk Assessment. *Journal of Global Optimization*, 2008. doi: 10.1007/s10898-007-9242-1.

[301] Gautier Marti, Sébastien Andler, Frank Nielsen, and Philippe Donnat. Clustering financial time series: How long is enough? *arXiv preprint arXiv:1603.04017*, 2016.

[302] Bisyron Wahyudi Masduki and Kalamullah Ramli. Study on Implementation of Machine Learning Methods Combination for Improving Attacks Detection Accuracy on Intrusion Detection System (IDS). *2015 International Conference on Quality in Research (QiR)*, pages 56–64, 2015. doi: 10.1109/QiR.2015.7374895.

[303] E. Medvet, E. Kirda, and C. Kruegel. Visual-similarity-based phishing detection. *In Proceedings of the 4th international conference on Security and privacy in communication networks, SecureComm '08, Article no 2, pp*, 2008.

[304] Ying Mei and Liangsheng Zhu. High Efficiency Association Rules Mining Algorithm for Bank Cost Analysis. *International Symposium on Electronic Commerce and Security*, 2008. doi: 10.1109/ISECS.2008.47.

[305] Vahab S Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Dynamic auctions with bank accounts. In *IJCAI*, pages 387–393, 2016.

[306] M. Moghimi and A. Y. Varjani. New rule-based phishing detection method. *Expert Systems with Application, 53, 231–242*, 2016.

[307] Patrick Monamo. A Multifaceted Approach to Bitcoin Fraud Detection : Global and Local Outliers. *International Conference on Machine Learning and Applications (ICMLA)*, 2016. doi: 10.1109/ICMLA.2016.0039.

[308] Woo Young Moon and Soo Dong Kim. Adaptive Fraud Detection Framework for FinTech Based on Machine Learning. *Advanced Science Letters*, 2017. doi: 10.1166/asl.2017.10412.

[309] Roberto Mourao, Ricardo Carvalho, Rommel Carvalho, and Guilherme Ramos. Predicting Waiting Time Overflow on Bank Teller Queues. *2017 16TH IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017. doi: 10.1109/ICMLA.2017.00-51.

[310] Youness Mourtaji and Pr Mohammed Bouhorma. Perception of a new framework for detecting phishing web pages. *SCAMS '17 Proceedings of the Mediterranean Symposium on Smart City Application*.

[311] A. M. Mubalaike and E. Adali. Deep Learning Approach for Intelligent Financial Fraud Detection System. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 2018.

[312] A. M. Mubarek and E. Adalı. Multilayer perceptron neural network technique for fraud detection. *2017 International Conference on Computer Science and Engineering (UBMK).*

[313] Tamil Nadu. Forecasting stock price using soft. *International Conference on Science Technology Engineering & Management (ICONSTEM)*, pages 1–4, 2017.

[314] Anahita Namvar and Mohsen Naderpour. Handling Uncertainty in Social Lending Credit Risk Prediction with a Choquet Fuzzy Integral Model. *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, 2018.

[315] C. Neelima and S. S. V. N. Sharma. Comparative study and analysis of innovative approaches in financial assets identical allocation and genetic algorithm implications. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, pages 61–66, 2017.

[316] G. S. Ng, C. Quek, and H. Jiang. Fcmac-ews: a bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network. *Expert Syst. Appl.*, 2008. doi: 10.1016/j.eswa.2006.10.027.

[317] M. N. Nguyen, D. Shi, and C. Quek. A Nature Inspired Ying-Yang Approach for Intelligent Decision Support in Bank Solvency Analysis. *Expert Syst. Appl.*, 2008. doi: 10.1016/j.eswa.2007.04.020.

[318] Umara Noor, Zahid Anwar, Tehmina Amjad, and Kim-Kwang Raymond Choo. A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems-the International Journal of Escience*, 2019. doi: 10.1016/j.future.2019.02.013.

[319] K. Oh, H. Choi, S. Kwon, and S. Park. Question Understanding Based on Sentence Embedding on Dialog Systems for Banking Service. *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2019. doi: 10.1109/BIGCOMP.2019.8679121.

[320] Olatunji J Okesola, Samuel N John, and Kennedy O Okokpujie. An improved Bank Credit Scoring Model A Naive Bayesian Approach. *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 228–233, 2017. doi: 10.1109/CSCI.2017.36.

[321] D. Olson, D. Denle, and Y. Meng. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems, 52 (2012), pp. 464-473*, 2012.

[322] S. Oreski, D. Oreski, and G. Oreski. Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications, 39 (2012), pp. 12605-12617*, 2012.

[323] Stjepan Oreski and Goran Oreski. Genetic Algorithm-based Heuristic for Feature Selection in Credit Risk Assessment. *Expert Syst. Appl.*, 2014. doi: 10.1016/j.eswa.2013.09.004.

[324] Ralf Ostermark. Incorporating asset growth potential and bear market safety switches in international portfolio decisions. *Applied Soft Computing*, 2012. doi: 10.1016/j.asoc.2012.03.052.

[325] Huseyin Ozturk, Ersin Namli, and Halil Ibrahim Erdal. Reducing Overreliance on Sovereign Credit Ratings: Which Model Serves Better? *Computational Economics*, 2016. doi: 10.1007/s10614-015-9534-3.

[326] Kunal Pahwa. Stock Market Analysis using Supervised Machine Learning. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 197–200, 2019.

[327] Rafael Palacios and Amar Gupta. A System for Processing Handwritten Bank Checks Automatically. *Image Vision Comput.*, 2008. doi: 10.1016/j.imavis.2006.04.012.

[328] Trilok Nath Pandey. Credit Risk Analysis using Machine Learning Classifiers. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 1850–1854, 2017.

[329] Harris Papadopoulos, Nestoras Georgiou, Charalambos Eliades, and Andreas Konstantinidis. Android malware detection with unbiased confidence guarantees. *Neurocomputing*, 2018. doi: 10.1016/j.neucom.2017.08.072.

[330] Priyanka S Patil and Nagaraj V Dharwadkar. Analysis of banking data using machine learning. *International Conference on I-SMAC (IoT in Social*, pages 876–881, 2017.

[331] Srushti Patil and Sudhir Dhage. A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 588–593, 2019.

[332] Suraj Patil, Varsha Nemade, and Piyushkumar Soni. ScienceDirect Predictive Modelling For Credit Card Fraud Detection Using Data Analytics. *Procedia Computer Science*, 00, 2018. doi: 10.1016/j.procs.2018.05.199.

[333] Pawe Pawiak, Moloud Abdar, and U Rajendra Acharya. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied Soft Computing Journal*, 84:105740, 2019. ISSN 1568-4946. doi: 10.1016/j.asoc.2019.105740. URL `https://doi.org/10.1016/j.asoc.2019.105740`.

[334] Anastasios Petropoulos, Sotirios P. Chatzis, and Stylianos Xanthopoulos. A Novel Corporate Credit Rating System Based on Student'S-t Hidden Markov Models. *Expert Syst. Appl.*, 2016. doi: 10.1016/j.eswa.2016.01.015.

[335] Anastasios Petropoulos, Sotirios P. Chatzis, Vasilis Siakoulis, and Nikos Vlachogiannakis. A Stacked Generalization System for Automated FOREX Portfolio Trading. *Expert Syst. Appl.*, 2017. doi: 10.1016/j.eswa.2017.08.011.

[336] C. Phua, V. Lee, K. Smith, and R. Gayler. A Comprehensive Survey of Data Mining-Based Fraud Detection Research. 2007.

[337] Thulasyammal Ramiah Pillai. Credit Card Fraud Detection Using Deep Learning Technique. *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, pages 1–6, 2018.

[338] Vassilis Plachouras. Information Extraction of Regulatory Enforcement Actions : From Anti-Money Laundering Compliance to Countering Terrorism Finance. *International Conference on Advances in Social Networks Analysis and Mining*, pages 2013–2016, 2015.

[339] Danae Politou and Paolo Giudici. Modelling Operational Risk Losses with Graphical Models and Copula Functions. *Methodology and Computing in Applied Probability*, pages 65–93, 2009. doi: 10.1007/s11009-008-9083-5.

[340] Rimpal R Popat and Jayesh Chaudhary. A Survey on Credit Card Fraud Detection using Machine Learning. *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, (Icoei):1120–1125, 2018.

[341] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta. PhishNet: Predictive blacklisting to detect phishing attacks. *In Proceedings of the INFOCOM-2010 IEEE, San Diego, pp*, 2010.

[342] Loan Pricing, Model For, Smes Based, O N Credit, and Risk Adjustment. Loan pricing model for smes based on credit risk adjustment 2. *2011 International Conference on Machine Learning and Cybernetics*, pages 10–13, 2011. doi: 10.1109/ICMLC.2011.6016806.

[343] Hong Qiao and Xue-Chen Dong. Research on the risk evaluation in loan projects of commercial bank in financial crisis. *2009 International Conference on Machine Learning and Cybernetics*, (July):12–15, 2009. doi: 10.1109/ICMLC.2009.5212440.

[344] Jon T. S. Quah and M. Sriganesh. Real-time Credit Card Fraud Detection Using Computational Intelligence. *Expert Syst. Appl.*, 2008. doi: 10.1016/j.eswa.2007.08.093.

[345] Marco Raberto, Andrea Teglio, and Silvano Cincotti. Integrating Real and Financial Markets in an Agent-Based Economic Model: An Application to Monetary Policy Design. *Comput. Econ.*, 2008. doi: 10.1007/s10614-008-9138-2.

[346] Mizanur Rahman, Nestor Hernandez, and Georgia Tech. Search Rank Fraud De-Anonymization in Online Systems. *The Eleventh International Conference on Web Search and Data Mining. Los Angeles*, pages 174–182, 2018.

[347] Majed Rajab. Visualisation Model Based on Phishing Features. *Journal of Information & Knowledge Management*, 2019. doi: 10.1142/S0219649219500102.

[348] Ranjith Ramesh. Predictive Analytics for Banking User Data using AWS Machine Learning Cloud Service. *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, (May 2008):210–215, 2017. doi: 10.1109/ICCCT2.2017.7972282.

[349] G Ramos and B Dias Sofia. The aftermath of the subprime crisis : a clustering analysis of world banking sector. *Review of Quantitative Finance and Accounting*, pages 293–308, 2014. doi: 10.1007/s11156-013-0342-3.

[350] Prachi Vivek Rane and Sudhir N Dhage. Systematic Erudition of Bitcoin Price Prediction using Machine Learning Techniques. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, (January 2009):594–598, 2019.

[351] V. Ravi and C. Pramodh. Threshold Accepting Trained Principal Component Neural Network and Feature Subset Selection: Application to Bankruptcy Prediction in Banks. *Applied Soft Computing*, 2008. doi: 10.1016/j.asoc.2007.12.003.

[352] V. Ravi, N. Naveen, and Manideepto-Das. Hybrid classifier based on particle swarm optimization trained auto associative neural networks as non-linear principal component analyzer: Application to banking. *International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 77–82, 2012. doi: 10.1109/ISDA.2012.6416516.

[353] Vadlamani Ravi. Auto-Associative Extreme Learning Factory as a Single Class Classifier. *2014 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–6, 2014. doi: 10.1109/ICCIC.2014.7238402.

[354] Mario Romao, Joao Costa, and Carlos J Costa. Robotic Process Automation : A case study in the Banking Industry. *Iberian Conference on Information Systems and Technologies (CISTI)*, (June):19–22, 2019.

[355] Alessandro Rossi, Antonio Rizzo, and Francesco Montefoschi. ATM Protection Using Embedded Deep Learning Solutions. *Artificial Neural Networks in Pattern Recognition*, 2018. doi: 10.1007/978-3-319-99978-4_29.

[356] Pumitara Ruangthong and Saichon Jaiyen. Bank Direct Marketing Analysis of Asymmetric Information Based on Machine Learning. *2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 93–96, 2015. doi: 10.1109/JCSSE.2015.7219777.

[357] Zuherman Rustam and Glori Stephani Saragih. Predicting Bank Financial Failures using Random Forest. *2018 International Workshop on Big Data and Information Security (IWBIS)*, pages 81–86, 2018. doi: 10.1109/IWBIS.2018.8471718.

[358] Morteza Saberi, Monireh Sadat Mirtalaie, Farookh Khadeer Hussain, Ali Azadeh, Omar Khadeer Hussain, and Behzad Ashjari. A Granular Computing-based Approach to Credit Scoring Modeling. *Neurocomput.*, 2013. doi: 10.1016/j.neucom.2013.05.020.

[359] Chinmayee Sadhu, Rituja Rane, Nandini Waghmare, and A. B. Patki. Android application for stock market prediction by fuzzy logic. *International Conference on Advanced Communications*, (978):1681–1685, 2014. doi: 10.1109/ICACCCT.2014.7019395.

[360] Justin Sahs and Latifur Khan. A Machine Learning Approach to Android Malware Detection. *2012 European Intelligence and Security Informatics Conference*, pages 141–147, 2012. doi: 10.1109/EISIC.2012.34.

[361] Ahmed Salih. Towards a Type-2 Fuzzy Logic Based System for Decision Support to Minimize Financial Default in Banking Sector. *2018 10th Computer Science and Electronic Engineering (CEEC)*, pages 46–49, 2018. doi: 10.1109/CEEC.2018.8674212.

[362] C. Salinesi, E. Ivankina, and W. Angole. Using the RITA Threats Ontology to Guide Requirements Elicitation: an Empirical Experiment in the Banking Sector. *International Workshop on Managing Requirements Knowledge*, 2008.

[363] A. P. Sam, B. Singh, and A. S. Das. A Robust Methodology for Building an Artificial Intelligent (AI) Virtual Assistant for Payment Processing. *IEEE Technology & Engineering Management Conference (TEMSCON)*, 2019. doi: 10.1109/TEMSCON.2019.8813584.

[364] Fatima Samea, Muhammad Anwar, Waseem, Farooque Azam, Mehreen Khan, and Muhammad Fahad Shinwari. An Introduction to UMLPDSV for Real-Time Dynamic Signature Verification. *Information And Software Technologies*, 2018. doi: 10.1007/978-3-319-99972-2_32.

[365] Nasredine Sammar, Hassane Essafi, Ali Al-jaoua, Jihad Al Jaam, and Helmi Hammami. Financial Events Detection by Conceptual News Categorization. *2010 10th International Conference on Intelligent Systems Design and Applications*, pages 1101–1106, 2010. doi: 10.1109/ISDA.2010.5687040.

[366] A. D. Sanford and I. A. Moosa. A Bayesian network structure for operational risk modelling in structured finance operations. *Journal of The Operational Research Society*, 2012. doi: 10.1057/jors.2011.7.

[367] Parinya Sanguansat. Paragraph2Vec-Based Sentiment Analysis on Social Media for Business in Thailand. *2016 8th International Conference on Knowledge and Smart Technology (KST)*, pages 175–178. doi: 10.1109/KST.2016.7440526.

[368] Arun Sasidharan. Performance of pattern recognition algorithms in. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1463–1467, 2018.

[369] Yashna Sayjadah. Credit Card Default Prediction using Machine Learning Techniques. *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, pages 1–4, 2018.

[370] Will Serrano. ScienceDirect ScienceDirect ScienceDirect Systems Network with Genetic Fintech Model : The Random Neural Algorithm Fintech Model : The Random Neural Network with Genetic Algorithm. *Procedia Computer Science*, 126:537–546, 2018. ISSN 1877-0509. doi: 10.1016/j.procs.2018.07.288. URL `https://doi.org/10.1016/j.procs.2018.07.288`.

[371] Will Serrano. Genetic and deep learning clusters based on neural networks for management decision structures. *UK Workshop on Computational Intelligence*, 2019. doi: 10.1007/978-3-319-97982-3_11.

[372] Will Serrano. Genetic and deep learning clusters based on neural networks for management decision structures. *Neural Computing and Applications*, 8, 2019. ISSN 1433-3058. doi: 10.1007/s00521-019-04231-8. URL `https://doi.org/10.1007/s00521-019-04231-8`.

[373] Akber Aman Shah, Desheng Dash Wu, Vladimir Korotkov, and Gul Jabeen. Do Commercial Banks Benefited From the Belt and Road Initiative ? A Three-Stage DEA-Tobit-NN Analysis. *IEEE Access*, 7:37936–37949, 2019. doi: 10.1109/ACCESS.2019.2897137.

[374] Kao-Yi Shen, Hioshi Sakai, and Gwo-Hshiung Tzeng. Comparing Two Novel Hybrid MRDM Approaches to Consumer Credit Scoring Under Uncertainty and Fuzzy Judgments. *International Journal of Fuzzy Systems*, 2019. doi: 10.1007/s40815-018-0525-0.

[375] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *In 28th international conference on human factors in computing systems, 10–15 April, 2010, Atlanta, GA*, 2010.

[376] Fu Shuen, Shie Mu-yen Chen, Fannie Mae, Freddie Mac, and Washington Mutual. Prediction of corporate financial distress : an application of the America banking industry. *Neural Computing and Applications*, pages 1687–1696, 2012. doi: 10.1007/s00521-011-0765-5.

[377] Mohammad Siami, Mohammad Reza Gholamian, and Javad Basiri. An application of locally linear model tree algorithm with combination of feature selection in credit scoring. *International Journal of Systems Science*, 2014. doi: 10.1080/00207721.2013.767395.

[378] B Singh and Emil Richard. Risk analysis in Electronic payments and settlement system using Dimensionality reduction techniques. *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 14–19, 2018. doi: 10.1109/CONFLUENCE.2018.8442666.

[379] Pradeep Singh. Comparative Study of Individual and Ensemble Methods of Classification for Credit Scoring. *International Conference on Inventive Computing and Informatics (ICICI)*, (Icici):968–972, 2017.

[380] Priyanka Singh, Yogendra P S Maravi, and Sanjeev Sharma. Phishing Websites Detection through Supervised Learning Networks. *2015 International Conference on Computing and Communications Technologies (ICCCT)*, pages 61–65, 2015. doi: 10.1109/ICCCT2.2015.7292720.

[381] Atish P. Sinha and Huimin Zhao. Incorporating Domain Knowledge into Data Mining Classifiers: An Application in Indirect Lending. *Decision Support Systems*, 2008. doi: 10.1016/j.dss.2008.06.013.

[382] P Sirisha, B Kamala Priya, K Aditya Kunal, and T Anuradha. Detection of Permission Driven Malware in Android Using Deep Learning Techniques. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 941–945, 2019.

[383] Ion Smeureanu, Gheorghe Ruxanda, and Laura Maria Badea. Customer segmentation in private banking sector using machine learning techniques. *Journal of Business Economics And Management*, 2013. doi: 10.3846/16111699.2012.749807.

[384] Min Song, Zhehua Wang, Chao Ma, and Huiying Li. Evolutionary Game Analysis of Bank Credit Retreat. *2010 International Conference on Artificial Intelligence and Education (ICAIE)*, pages 499–504. doi: 10.1109/ICAIE.2010.5640965.

[385] Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun Majumdar. Credit Card Fraud Detection Using Hidden Markov Model. *IEEE Transactions on Dependable and Secure Computing*, 2008. doi: 10.1109/TDSC.2007.70228.

[386] Shriansh Srivastava, J. Priyadarshini, Sachin Gopal, Sanchay Gupta, and Har Shobhit Dayal. Optical Character Recognition on Bank Cheques Using 2D Convolution Neural Network. *Applications of Artificial Intelligence Techniques in Engineering*, 2019. doi: 10.1007/978-981-13-1822-1_55.

[387] Rosalvo Ermes Streit and Denis Borenstein. Structuring and Modeling Data for Representing the Behavior of Agents in the Governance of the Brazilian Financial System. *Appl. Artif. Intell.*, 2009. doi: 10.1080/08839510902804796.

[388] Zheyuan Su and Mirsad Hadzikadic. An Agent-based System for Issuing Stock Trading Signals. *International Conference on Simulation and Modeling Methodologies*, 2015. doi: 10.5220/0005508203520358.

[389] Pang Sulin, Liu Yongqing, and Li Rongzhou. The optimal design on the decision mechanism of credit risk for banks with imperfect information. *World Congress on Intelligent Control and Automation (WCICA)*, 2000. doi: 10.1109/WCICA.2000.862805.

[390] Yunchuan Sun, Xiaoping Zeng, Xuegang Cui, Guangzhi Zhang, and Rongfang Bie. An active and dynamic credit reporting system for SMEs in China. *Personal and Ubiquitous Computing*, 2019.

[391] G Ganesh Sundarkumar, Vadlamani Ravi, Ifeoma Nwogu, and Venu Govindaraju. Malware Detection via API calls , Topic Models and Machine Learning. *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1212–1217, 2015. doi: 10.1109/CoASE.2015.7294263.

[392] Gian Antonio Susto, Matteo Terzi, Chiara Masiero, Simone Pampuri, and Andrea Schirru. A Fraud Detection Decision Support System via Human On-line Behavior Characterization and Machine Learning. *2018 First IEEE International Conference on Artificial Intelligence For Industries (Ai4i 2018)*, 2018. doi: 10.1109/ai4i.2018.00011.

[393] Chih-hua Tai. Identifying Money Laundering Accounts. *2019 International Conference on System Science and Engineering (ICSSE)*, pages 379–382, 2019.

[394] Sumit Baburao Tamgale. Application of Deep Convolutional Neural Network to Prevent ATM Fraud by Facial Disguise Identification. *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5, 2017.

[395] Lili Tao, Xiaodong Liu, and Yan Chen. Online banking performance evaluation using data envelopment analysis and axiomatic fuzzy set clustering. *Quality & Quantity*, pages 1259–1273, 2013. doi: 10.1007/s11135-012-9767-3.

[396] Surya Susan Thomas. Implementation of Data Mining Techniques Monetary Domains. *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, (Icimia):730–734, 2017. doi: 10.1109/ICIMIA.2017.7975561.

[397] S. Tian and Y. Yu. Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance, 51 (2017), pp. 510-526*, 2017.

[398] Tian Tian, Jun Zhu, Fen Xia, Xin Zhuang, and Tong Zhang. Crowd Fraud Detection in Internet Advertising. *International World Wide Web Conference Committee (IW3C2)*, (10):1100–1110.

[399] Chih-Fong Tsai and Ming-Lun Chen. Credit Rating by Hybrid Machine Learning Techniques. *Appl. Soft Comput.*, 2010. doi: 10.1016/j.asoc.2009.08.003.

[400] Yu-Che Tsai, Chih-Yao Chen, Shao-Lun Ma, Pei-Chi Wang, You-Jia Chen, Yu-Chieh Chang, and Cheng-Te Li. FineNet: A Joint Convolutional and Recurrent Neural Network Model to Forecast and Recommend Anomalous Financial Items. *RecSys '19 Proceedings of the 13th ACM Conference on Recommender Systems*, 2019. doi: 10.1145/3298689.3346968.

[401] A. Tsakonas, N. Ampazis, and G. Dounias. Towards a comprehensible and accurate credit management model: Application of four computational intelligence methodologies. *2006 International Symposium on Evolving Fuzzy Systems*, 2006. doi: 10.1109/ISEFS.2006.251142.

[402] Philip M Tsang, Paul Kwok, S O Choy, Reggie Kwan, S C Ng, Jacky Mak, Jonathan Tsang, Kai Koong, and Tak-lam Wong. Design and implementation of NN5 for Hong Kong stock price forecasting. *Engineering Applications of Artificial Intelligence*, 20: 453–461, 2007. doi: 10.1016/j.engappai.2006.10.002.

[403] Alexey Tselykh and Dmitry Petukhov. Web Service for Detecting Credit Card Fraud in Near Real-Time. *International Conference on Security of Information and Networks*, pages 2–5, 2015.

[404] H. Tung, C. Cheng, Y. Chen, Y. Chen, S. Huang, and A. Chen. Binary Classification and Data Analysis for Modeling Calendar Anomalies in Financial Markets. *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 2016. doi: 10.1109/CCBD.2016.032.

[405] M Amaad Ul, Haq Tahir, Sohail Asghar, Ayesha Zafar, and Saira Gillani. A Hybrid Model to Detect Phishing-Sites Using Supervised Learning Algorithms. *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1126–1133, 2016. doi: 10.1109/CSCI.2016.0214.

[406] K S Umadevi, Abhijitsingh Gaonka, and Ritwik Kulkarni. Analysis of Stock Market using Streaming data Framework. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1388–1390, 2018.

[407] Franco Varetto. Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance*, 22, 1998.

[408] Guandong Vo, Nhi N. Y. Xu. The volatility of Bitcoin returns and its correlation to financial markets. *International Conference on Behavioral*, 2017. doi: 10.1109/BESC.2017.8256365.

[409] Grega Vrbani. Swarm Intelligence Approaches for Parameter Setting of Deep Learning Neural Network : Case Study on Phishing Websites Classification. *International Conference on Web Intelligence*, 2000.

[410] Guoxun Wang, Liang Liu, Yi Peng, Guangli Nie, Gang Kou, and Yong Shi. Predicting Credit Card Holder Churn in Banks of China Using Data Mining and MCDM. *International Conference on Web Intelligence and Intelligent Agent Technology*, 2010. doi: 10.1109/WI-IAT.2010.237.

[411] Haomin Wang. Cost-sensitive Classifiers in Credit Rating A Comparative Study on P2P Lending. *2018 7th International Conference on Computers Communications and Control (ICCCC)*, (Icccc):210–213, 2018. doi: 10.1109/ICCCC.2018.8390460.

[412] Maoguang Wang. Research on Financial Network Loan Risk Control Model based on Prior Rule and Machine Learning Algorithm. *ICMAI 2019 Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence*, pages 76–79, 2019.

[413] Liwei Wei. Credit Risk Evaluation Using : Least Squares Support Vector Machine with Mixture. *2016 International Conference on Network and Information Systems for Computers (ICNISC)*, pages 237–241, 2016. doi: 10.1109/ICNISC.2016.059.

[414] Ran Wei. Development of Credit Risk Model Based on Fuzzy Theory and Its Application for Credit Risk Management of Commercial Banks in China. *2008 4th International Conference on Wireless Communications*, (072102350034):1–4, 2008.

[415] Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen. Effective Detection of Sophisticated Online Banking Fraud on Extremely Imbalanced Data. *World Wide Web*, 2013. doi: 10.1007/s11280-012-0178-0.

[416] A. Weissensteiner. A Q -Learning Approach to Derive Optimal Consumption and Investment Strategies. *IEEE Transactions on Neural Networks*, 2009. doi: 10.1109/TNN.2009.2020850.

[417] Yu-ting Wen, Pei-wen Yeh, Tzu-hao Tsai, Wen-chih Peng, and Hong-han Shuai. Customer Purchase Behavior Prediction from Payment Datasets. *WSDM '18 Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 628–636, 2018.

[418] Charles A Worrell, Shaun M Brady, and Jerzy W Bala. Comparison of data classification methods for predictive ranking of banks exposed to risk of failure. *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 1–6, 1802. doi: 10.1109/CIFEr.2012.6327823.

[419] C.H. Wu, G.H. Tzeng, Y.J. Goo, and W.C. Fang. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications, 32 (2007), pp. 397-408*, 2007.

[420] Der-bang Wu and Shu-ling Lin. Predicting Risk Parameters Using Intellegent Fuzzy / Logistic Regression. *2010 International Conference on Artificial Intelligence and Computational Intelligence*, 3:397–401, 2010. doi: 10.1109/AICI.2010.320.

[421] Quan-wu Xiao and Lei Shi. Gradient Learning Approach for Variable Selection in Credit Scoring. *2009 International Conference on Business Intelligence and Financial Engineering*, pages 219–222, 2009. doi: 10.1109/BIFE.2009.59.

[422] Yaya Xie, Xiu Li, E. W. T. Ngai, and Weiyun Ying. Customer Churn Prediction Using Improved Balanced Random Forests. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2008.06.121.

[423] J U N Xu, Qing-cai Chen, Xiao-long Wang, and Zhong-yu Wei. One-class classification models for financial indus try information recommendation. *2010 International Conference on Machine Learning and Cybernetics*, (July):11–14, 2010. doi: 10.1109/ICMLC.2010.5580675.

[424] Ronghua Xu, Xuheng Lin, Qi Dong, and Yu Chen. Constructing Trustworthy and Safe Communities on a Blockchain-Enabled Social Credits System. *International Conference on Mobile and Ubiquitous Systems: Computing*, 2018. doi: 10.1145/3286978.3287022.

[425] Xiujuan Xu, Chunguang Zhou, and Zhe Wang. Credit Scoring Algorithm Based on Link Analysis Ranking with Support Vector Machine. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2008.01.024.

[426] Jingming Xue, SiHang Zhou, Qiang Liu, Xinwang Liu, and Jianping Yin. Financial Time Series Prediction Using 2,1RF-ELM. *Neurocomputing*, 2018. doi: 10.1016/j.neucom.2017.04.076.

[427] Jiaqi Yan and Leon Zhao. An Ontology-based Approach for Bank Stress Testing. *2013 46th Hawaii International Conference on System Sciences*, pages 3407–3415, 2013. doi: 10.1109/HICSS.2013.91.

[428] Chen-guang Yang and Xiao-bo Duan. Credit risk assessment in commercial banks based on svm. *Machine Learning and Cybernetics*, (July):12–15, 2008.

[429] M. H. Yazdani. Developing a model for validation and prediction of bank customer credit using information technology (case study of dey bank). *Journal of Fundamental and Applied Sciences*, 2017. doi: 10.4314/jfas.v9i1s.693.

[430] I-Cheng Yeh and Che hui Lien. The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2007.12.020.

[431] W. D. Yu, S. Nargundkar, and N. Tiruthani. A phishing vulnerability analysis of web based systems. *In Proceedings of the 13th IEEE symposium on computers and communications (ISCC 2008), IEEE*, 2008.

[432] Wen-Fang Yu and Na Wang. Research on Credit Card Fraud Detection Model Based on Distance Sum. *2009 International Joint Conference on Artificial Intelligence*, 2009. doi: 10.1109/JCAI.2009.146.

[433] Xiaojian Yu. A Comparison Study on Interest Rate Models of SHIBOR Based on MCMC Method. *International Conference on Artificial Intelligence and Computational Intelligence*, 2009. doi: 10.1109/AICI.2009.283.

[434] Mohamad Zamini. Credit Card Fraud Detection using autoencoder based clustering. *2018 9th International Symposium on Telecommunications (IST)*, pages 486–491, 2018.

[435] Qing Zhan and Hang Yin. A Loan Application Fraud Detection Method Based on Knowledge Graph and Neural Network. *International Conference on Innovation in Artificial Intelligence (ICIAI)*, pages 111–115, 2018.

[436] David Zhang, Xiao-Ping (Steven) Kedmey. A Budding Romance: Finance and AI. *IEEE MULTIMEDIA*, 2018. doi: 10.1109/MMUL.2018.2875858.

[437] Jisong Zhang and Xingfen Wang. Breaking Internet Banking CAPTCHA Based on Instance Learning. *2010 International Symposium on Computational Intelligence and Design*, 1:39–43, 2010. doi: 10.1109/ISCID.2010.18.

[438] Y. Zhang, J. I. Hong, and L. F. Cranor. Cantina: A content-based approach to detecting phishing web sites. *In Proceedings of the 16th international conference on World Wide Web*, 2007.

[439] Zhao Zhang, Qinggang He, and Bailing Wang. A Novel Multi-Layer Heuristic Model for Anti-Phishing. *ICIE '17 Proceedings of the 6th International Conference on Information Engineering*.

[440] Huimin Zhao, Atish P. Sinha, and Wei Ge. Effects of Feature Construction on Classification Performance: An Empirical Study in Bank Failure Prediction. *Expert Syst. Appl.*, 2009. doi: 10.1016/j.eswa.2008.01.053.

[441] Zhenyu Zhao. National Student Loans Credit Risk Assessment Based on GABP Algorithm of Neural Network. *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, 2011:2196–2199, 2011. doi: 10.1109/AIMSEC.2011.6010910.

[442] Yu Zhaoji, Mao Qiang, and Wang Wenjuan. The Application of WN Based on PSO in Bank Credit Risk Assessment. *2010 International Conference on Artificial Intelligence and Computational Intelligence*, 3:444–448, 2010. doi: 10.1109/AICI.2010.331.

[443] Maria Zhdanova, Juergen Repp, Roland Rieke, Chrystel Gaber, and Baptiste Hemery. No Smurfs: Revealing Fraud Chains in Mobile Money Transfers. *International Conference on Availability*, 2015. doi: 10.1109/ARES.2014.10.

[444] Jian-guo Zhou and T A O Bai. Customer classification in commercial bank based on. *International Conference on Machine Learning and Cybernetics*, (July):12–15, 2008.

[445] Jianguo Zhou, Zhaoming Wu, Chenguang Yang, and Qi Zhao. The Integrated Methodology of Rough Set Theory and Support Vector Machine for Credit Risk Assessment. *Sixth International Conference on Intelligent Systems Design and Applications*, 2006. doi: 10.1109/ISDA.2006.267.

[446] Xiaofei Zhou, Wenhan Jiang, Yong Shi, and Yingjie Tian. Credit Risk Evaluation with Kernel-based Affine Subspace Nearest Points Learning Method. *Expert Syst. Appl.*, 2011. doi: 10.1016/j.eswa.2010.09.095.

[447] Chunsheng Zhu, Yuanrui Zhan, and Shijun Jia. Credit Risk Identification of Bank Client Basing on Supporting Vector Machines. *2010 Third International Conference on Business Intelligence and Financial Engineering*, pages 62–66, 2010. doi: 10.1109/BIFE.2010.25.

[448] M. Maciej Zieba, S.K. Tomczak, and J.M. Tomczak. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications, 58 (2016), pp. 93-101*, 2016.

[449] M. Šušteršic, D. Mramor, and J. Zupan. Consumer credit scoring models with limited data. *Expert Systems with Applications, 36 (2009), pp. 4736-4744*, 2009.