APPLICABILITY STUDY OF ARTIFICIAL INTELLIGENCE TO FORECAST NEW INFRASTRUCTURE PROJECT INTRODUCTION BASED ON THE DECISION-MAKING DURATION BY THE GOVERNMENT

CASE STUDY : PUBLIC ROAD INFRASTRUCTURE IN THE NETHERLANDS

YOSEP PANDJI HARIO WICAKSONO

CONSTRUCTION MANAGEMENT & ENGINEERING

TU DELFT IN COLLABORATION WITH BAM INFRA BV

bam TUDelft

COLOPHON

Report:	
Type:	Graduation Thesis
Title:	*Applicability Study of Artificial Intelligence to Forecast New Infrastructure Project Introduction Based on The Decision-Making Duration by The Government (Case Study: Public Road Infrastructure in the Netherlands)*
Place:	Delft
Date:	5 August 2019
Authors:	
Name:	Yosep Pandji Hario Wicaksono
Student Number:	4739280
Email:	yosepandji@gmail.com
Graduation Company:	BAM Infra BV
Study Program:	
University:	Delft University of Technology
Faculty:	Technology, Policy and Management, Civil Engineering and Geosciences, Architecture and the Built Environment
Master:	Construction Management & Engineering
Course:	CME2000-Master Thesis
Graduation Committee:	
Chairman:	Prof.dr. G.P. (Bert) van Wee Faculty of Technology, Policy and Management
1st Supervisor:	Dr. Jan Anne Annema Faculty of Technology, Policy and Management
2nd Supervisor:	Dr. J.S. Rellermeyer Faculty of Electrical Engineering, Mathematics & Computer Science
Company Supervisor:	Jeroen Nuijten BAM Infra BV

Company:
BAM Infra BV
H.J. Nederhorststraat 1
2801 SC Gouda

University:
Delft University of Technology
Mekelweg 2
2628 CD Delft

This page is left blank intentionally

Foreword

This report is the result of my graduation research to obtain a master's degree in Construction Management and Engineering from Delft University of Technology. The research aim is to explore the applicability of Artificial Intelligence to forecast a new infrastructure project introduction to the market based on the decision-making duration by the government regarding the project's preferred design alternative. By forecasting the project introduction to the market, the user of the model is expected to be able to prepare better for the forecasted project by better allocating its internal resources, establishing a partnership with strategic partners, and initiating research and development for the special requirements of the new project.

This research is a collaborative effort between Delft University of Technology and BAM Infra BV. It all started when I was offered an opportunity to work together with BAM Infra BV on this idea to explore the potential of AI in forecasting future project. This idea is also echoed with the university's interest who consider this research to be interesting, innovative, and could provide good insight into the academic field. When the kick-off meeting started, my journey in artificial intelligence and data analysis world was officially started. To be honest, I challenged myself to do this research because the risk is quite high due to my lack of knowledge in AI and programming. But, at the end of the research, I realized that I had learned a lot through the processes, and I cannot be thankful enough for this opportunity which has been given to me.

As a sign of my gratitude, I would like to mention special people in my life in this section. First of all, I want to say thank you to my Jesus Christ and Mother Mary for always be here by my side no matter what the circumstances are. Secondly, I want to say thank you to my lovely family (Mom, Dad, Mbak Linda, Yohan, and Ateja's Family) for always believing in me and provided me with an opportunity to be in this wonderful country and learning my craft at one of the best universities in the world.

Thirdly, I cannot express my gratitude enough to my graduation committee, who has supported me since the start of the research. To Professor Bert van Wee, thank you for your willingness and kindness to take me as one of your graduate students, Jan Anne Annema for your full support since day one we met, Jan Rellermeyer for your patience and time, and Jeroen Nuijten for believing in me and giving me this amazing opportunity to work together with an amazing company such as BAM Infra BV. I also want to say thank you to Sebastiaan Geenen who supported me during my time at BAM Infra BV.

In addition to that, I want to say thank you to my best friends in Indonesia and Delft, who always cheered me up and gave me the much-needed pushes during my study here in the Netherlands. There are too many of you to be written here, but you must have known that I already wrote it down in another space.

And finally, I want to say thank you to Maria Margaretha Kania Anggaraputri and her wonderful family for always believing in me and supporting me in every step of the way. You are one of the reasons why I am here. Thank you once again.

Without support from all of you, I would not be here and completed this research. God bless you all.
I hope you enjoy reading this report and please do not hesitate to contact me if you have any question.

Yosep Pandji Hario Wicaksono

Delft, 5 August 2019

This page is left blank intentionally

Executive Summary

Research Objective and Scope

This research aims to provide an insight about the applicability of Artificial Intelligence (AI) in making a forecast about decision-making duration of a new infrastructure project by the government during infrastructure planning procedure prior to the market introduction. The decision-making duration in this research is the duration which the government requires to decide on the preferred alternative for a certain infrastructure problem. By identifying the likely timeframe of a new infrastructure project introduction from the government, the construction firms can prepare itself better by allocating its resources accordingly to the profile and requirement of the upcoming projects.

This duration forecasting is not a straight forward task due to some influence from various variables during the decision-making procedure, which makes its environment dynamic. This research focuses on the variables which originating from the drivers of the infrastructure demand that underlying the initiation of a new infrastructure project by the government and the involvement of various interested parties during the decision-making procedure itself.

This research decided to explore further about the applicability of AI to forecast the duration within this context based on the potentials of AI that shown by previous researches. From these researches, it is proven that the implementation of AI provides a better prediction accuracy in comparison to the conventional method / tool such as linear regression. AI also provides an ability to capture the non-linearity between the variables in a simpler way than the conventional models. For a model deployment in the early stage of a process, AI shows advantages over conventional methods which usually requires the knowledge and experience of the experts; because it needs only basic information which normally available during the early stage of project development.

With regard to the scope of the research, the Netherlands construction industry is chosen with the perspective of a construction firm, which is BAM Infra BV. With regard to the legal framework of the industry, Dutch Infrastructure Planning Act (*Tracewet*) is the main reference of the research, which acts as the foundation of the proposed forecasting model. Lastly, the research focuses on the road infrastructure project from the Netherlands government, which falls under *Tracewet*.

Based on the aforementioned objective and scope, the following research question is formulated:

"To what extent can an Artificial Intelligence (AI) technology be applied to forecast new infrastructure project introduction based on the decision-making duration by the Dutch Government?"

Research Method

This research applies a combination of three methods to fulfill the aforementioned objective, which are literature study, experts interview, and model simulation. This research is divided into four main steps to answer the main research question: *[1] AI Theory and System Design for Forecasting, [2] Data and Variable Exploratory Study, [3] AI Forecasting Model Implementation, and [4] Result Discussion*. The first step is started with explaining why AI, theoretically, could be a useful tool in the forecasting field and designing the AI system to be used for tender forecasting. Then, it is followed by an exploratory study on both the data availability and relevant variables to the AI system. After these two steps are done, the third step is where the AI model is implemented to solve the problem at hand. After the model is implemented and the result is known, a discussion is done about the influential factor behind the model's result and a comparison is made with the conventional statistic models' result.

Research Results

The first research step shows that AI could potentially be a tool for the prediction task of decision-making duration on new road infrastructure with the advantages that it has in comparison to the conventional methods; which are better accuracy, ability to handle imprecise data, and good non-linear approach. The analysis on *System's Operating Environment* and its relationship with the proposed AI system provides knowledge about the type of AI for the forecasting problem; which in this case is Artificial Neural Network (ANN). Based on its operating environment, the purpose of the ANN model is to be implemented in the phase before the new road infrastructure is announced in TenderNed with a focus on duration between the release of *startnotitie* (the announcement of the infrastructure problem) and the release of draft-track decision (when the Minister announces her/his preferred alternative).

The second step shows that a total of ninety-five road infrastructure projects are available as the data entries for the dataset. The data is gathered by collecting *startnotitie* which released by the Dutch government. This step also provides the input features for the model based on the identified independent variables; which are Road Category, Type of Network Intervention, Gross Domestic Product (GDP), three years average of GDP growth, Regional Domestic Product (RDP), Population size, Geographical profile of the project area, *Ecologische Hoofd Structuur* (EHS) Intersection, Car to road area ratio (car/km2), Dominant Political Party Change, Dominant Political Ideology, and Number of Provinces. After the data is gathered along with its input features, the features are visualized. From the pair-wise visualization of the categorical data within the dataset, four data are considered as outliers and removed from the dataset. This removal which makes the final number of data entries is ninety-one road infrastructure projects.

The third step shows that the prediction made by the models are not sufficiently reliable. There are two models implemented in this research, which are Regression ANN Model and Classification ANN Model. The first model produced a prediction with Root Mean Squared Error (RMSE) of 2.565 years; while the second model produced a prediction with an accuracy of 55%. Based on an interview done with the commercial manager and innovation of BAM, these values are not acceptable to be implemented to the company's business processes due to the level of uncertainty that present. To address this result, an optimization effort has been done through a reduction in the number of input variables involved in the forecasting model based on the curse of dimensionality (Bellman, 1961). The result of optimization shows that no significant improvement occurred for both the regression model and the classification model. Although the RMSEs are higher for the regression model, the difference with the original model, which embedded the initial input features set is relatively low. A similar outcome also happened in the optimization of the classification model, which indicated by the accuracy values that are not far apart from the original model accuracy. The result of the proposed AI forecasting model implementation and improvement could mean that the combination of the following factors influences the model performance: [1] the number of data entries is too low to make an adequate generalization; [2] the identified variables do not have enough influence on the decision-making duration of new road infrastructure project in regard to the preferred design alternative; and [3] the model is unable to properly represent the "world" which influences the decision-making duration for new road infrastructure projects.

In the last step of the research, which is the fourth step, the discussion on the result shows that the combination of the three factors is indeed affecting the model performance. For the first factor, it can be considered that the research has used a decent number of data entries for the proposed ANN forecasting model in comparison with the existing researches which implemented ANN for forecasting in civil engineering domain. However, the variables which used for the model are different with those researches. Hence, the number of data is deemed to be insufficient due to the high number of incorporated input variables in this research. For the second factor, the non-existent of a substantial correlation between the numeric input variable and the target value is assumed to be the potential cause to the poor model performance. It is indicated by the fact that besides the GDP and 3 years average GDP growth variable, the other numeric input variables have a small influence on the target value despite being well-founded in theory. For the third factor, it is found out that there are other policies outside Infrastructure Planning Act which influence the Infrastructure Planning Act and eventually affected the decision-making procedure duration; and were not incorporated to the proposed model due to the scope limitation of the research. The policies which discussed in this research in relation to the decision-making duration are the modifications of Infrastructure Planning Act, mobility plans from the government, and spoedwet Wegverbreding 2003. In regard to the comparison of ANN models with Multiple Linear Regression (MLR) and Logistic Regression (LR), the results show that the ANN models do not perform superior to their counterparts. This is indicated by the RMSE value of MLR on training set which lower than the ANN Regression Model (i.e., 2.175 years compares to ANN's 2.565 years) and model accuracy of LR on training set which higher than the ANN Classification Model (i.e., 66.67% compares to ANN's 55%).

Conclusion

The AI technology, specifically ANN, is not applicable to be used as a forecasting model to predict a new infrastructure project introduction based on the duration of decision-making on the preferred design alternative by the Dutch Government. Two approaches have been explored in this research, namely the ANN regression model and ANN classification model. It is found that neither models produce a reliable prediction, which indicated by high RMSE and low model accuracy on the dataset. This reliability is evaluated by interviewing BAM's commercial manager about the acceptable range of error for the prediction made.

The optimization effort has been done to address these results by iterating several different variables combinations into the models. These optimization results revealed that some factors influenced the performance of the models. A further discussion has been done on these factors; namely number of data entries, input variables influence, and representation of the world by the model. It is found out that the combination of these factors has an impact to a certain degree on the model performance. Besides that, a comparison with other conventional methods (i.e., Multiple Linear Regression and Logistic Regression) has been done. The result shows that the ANN models do not perform better than the conventional methods being compared in term of model prediction accuracy, which transcends into its ability to handle imprecise data and non-linear approach. This comparison result indicates the importance of dataset quality over the decision of a forecasting model to be used.

Based on the research result, a recommendation for future research which aims to forecast the introduction of new infrastructure project has been proposed with the model focus shifts from: [1] Dynamic decision-making procedure towards a more stable physical state of the existing road network, and [2] Large infrastructures project towards small projects (e.g. replacement project / maintenance project). By realigning the focus of the future AI model as mentioned above, the three factors which affected the model performance can be addressed. Firstly, a more stable system environment to be interpreted by the AI model might reduce the difficulty in identifying the relevant independent variables as the predictor for a future project introduction period. Secondly, the focus on smaller projects means a higher project number is available for the database. This addition means the future AI model has a chance to be more reliable with more input data available to be processed

This page is left blank intentionally

Table of Content

Foreword ... iii

Executive Summary ... v

Table of Content ... ix

Chapter 1 Introduction ... 1

1.1 Problem Statement ... 1

1.2 Research Objective ... 2

1.3 Research Questions ... 2

1.4 Research Outline ... 3

Chapter 2 Research Method ... 5

2.1 Research Scope ... 5

2.2 Research Approach ... 5

2.2.1 Step 1: AI Theory and System Design for Tender Forecasting ... 6

2.2.2 Step 2: Data and Variables Exploratory Study ... 8

2.2.3 Step 3: AI Forecasting Model Implementation ... 9

2.2.4 Step 4: Result Discussion ... 10

Chapter 3 Artificial Intelligence Theory and System Design ... 12

3.1 Artificial Intelligence Theory ... 12

3.1.1 Definition of Artificial Intelligence ... 12

3.1.2 AI in forecasting field ... 13

3.2 AI System Design for Project Introduction Forecasting ... 14

3.2.1 Operating Environment ... 14

3.2.2 Artificial Intelligence for Infrastructure Tender Forecasting ... 16

3.2.3 Artificial Neural Network (ANN) ... 18

3.3 Conclusion ... 20

Chapter 4 Data and Variables Exploratory Study ... 22

4.1 Data Availability Study ... 22

4.2 Variables Study ... 22

4.2.1 Variables Identification ... 22

4.2.2 Variables/Features' Data Type Illustration ... 26

4.2.3 Features Visualisation ... 27

4.3 Conclusion ... 33

Chapter 5 ANN Forecasting Model Implementation ... 35

5.1 Dataset Pre-processing ... 35

5.2 Model Implementation ... 37

5.2.1 ANN Regression Model ... 37
5.2.2 ANN Classification Model ... 41
5.3 Model Optimization ... 45
5.3.1 ANN Regression Model Optimization ... 45
5.3.2 ANN Classification Model Optimization ... 46
5.4 Conclusion ... 47
Chapter 6 Result Discussion ... 50
6.1 Influential Factor Discussion ... 50
6.1.1 Number of Data Entries ... 50
6.1.2 Input Variables Influence ... 51
6.1.3 Representation of the "World" ... 52
6.2 Result Comparison to other statistical methods ... 56
6.2.1 Multiple Linear Regression Model ... 56
6.2.2 Logistic Regression Model ... 56
6.3 Conclusion ... 57
Chapter 7 Research Conclusion ... 60
Chapter 8 Future Recommendation ... 63
References ... 66
Appendix 1 Infrastructure Planning Act (IPA) ... 69
Appendix 2 Interviews with BAM's Expert ... 72
Pre-qualification Tender department (BAM_1) ... 72
Tender Strategy Department (BAM_2) ... 72
Commercial Manager (BAM_3) ... 73
Commercial Manager & Innovation (BAM_4) ... 74
Appendix 3 Models' Dataset ... 75
Raw Dataset ... 75
ABT Dataset for Regression Models ... 75
ABT Dataset for Classification Model ... 76

List of Figures

Figure 1 Research Outline 3
Figure 2 Research Approach Illustration 5
Figure 3 Step One Illustration 6
Figure 4 Adopted Analytical Perspective of Ai System Operating System Environment Architecture (Charles S. Watson, 2016) 7
Figure 5 Variables Exploratory Study Method Illustration 9
Figure 6 BAM's Tender Procedure 15
Figure 7 A simple neuron mathematical model proposed by McCulloch and Pitts (1943) 18
Figure 8 Multilayer ANN Structure Illustration 19
Figure 9 ANN for Decision-making duration on New Road Infrastructure Illustration 20
Figure 10 Dataset Quantity Representation 22
Figure 11 Independent Variables Overview, and it's Physical Environment Domain Representation 23
Figure 12 Road Category Illustration 23
Figure 13 RDP Percentage of the Netherlands Provinces in 1995 (CBS) 24
Figure 14 Categorical Data Visualisation 28
Figure 15 Box Plot Structure Illustration 29
Figure 16 Road Category Box Plot 29
Figure 17 Network Intervention Box Plot 30
Figure 18 Geographical Profile Box Plot 30
Figure 19 EHS Intersection Box Plot 31
Figure 20 Political Party Change Box Plot 31
Figure 21 Dominant Political Ideology Box Plot 31
Figure 22 GDP and Decision-Making Duration 32
Figure 23 Pre-Standardization and Post-Standardization Categorical Features Example 36
Figure 24 5-Fold Cross Validation Illustration 37
Figure 25 Relu Activation Function Illustration 38
Figure 26 Road Infrastructure ANN Regressor Structure 39
Figure 27 Model Loss on Training Data 40
Figure 28 Regression Model Five-Fold Cross-Validation Result 40
Figure 29 The Difference between Prediction and Actual Value 41
Figure 30 Decision-Making Duration Yearly and Global Average 42
Figure 31 Road Infrastructure Projects ANN Classifier Structure 42
Figure 32 Sigmoid Function Illustration (Karlik and Olgac, 2010) 43
Figure 33 Model Accuracy and Loss on Training Data 44
Figure 34 Classification Model Five-Fold Cross-Validation Result 44
Figure 35 Timeline of IPA and other policies related to road infrastructure 53
Figure 36 Decision Making Duration Average for each IPA (years) 54
Figure 37 Decision Making Duration Average for each mobility plan (years) 55
Figure 38 Modification of Operating Environment for future research 64

List of Tables

Table 1 Four Types of Agent Structure (Norvig and Russell, 2010) 12
Table 2 AI implementation benefit in forecasting field 13
Table 3 AI System Attributes based on the interaction with its Operating Environment (Norvig and Russell, 2010). 17
Table 4 Data Types (Kelleher et al., 2015) 26
Table 5 Data Type of the Variables 26
Table 6 Individual Correlation of Input Variables with the decision-making duration.................... 33
Table 7 ABT Structure adopted from J. D. Kelleher, B. Mac Namee, and A. D'Arcy (2015).................... 36
Table 8 Raw Dataset (Partial).................... 36
Table 9 Regression Model Standardized ABT Dataset (Partial) 38
Table 10 Regression Model Prediction on Test Data 40
Table 11 Classification Model Standardized ABT Dataset (Partial).................... 42
Table 12 Testing Result on New Data Entries 44
Table 13 Variables Combination and its corresponding RMSE.................... 45
Table 14 Variables Combination and its corresponding Accuracy.................... 46
Table 15 Data Entries Number Comparison with existing research 50
Table 16 Full Procedure and Shortened Procedure (Hobma and De Jong, 2016) 70

List of Formulas

4.1 Correlation between Feature a and b. 32
4.2 Covariance of Feature a and b 32
5.1 Standard Score 35
5.2 Rectified Linear Units 38
5.3 Mean Squared Error 39
5.4 Sigmoid Function 43
5.5 Cross-entropy Function 43
6.1 Multiple Linear Regression Function 56
6.2 Logit Transformation Function 57

This page is left blank intentionally

Chapter 1 Introduction

The high number of construction firms involved and the low amount of new large infrastructure project from the government makes the market competitive. This situation leads to a need of competitive advantage to elevate the owner from the rest of its competitors. One of the strategies that are commonly used by the company to achieve this is forecasting. Forecasting is a mean to accurately predict the future as far as possible, based on the available information; which includes both historical data and any knowledge of events in the future that might have an impact on the forecasts (Hyndman and Athanasopoulos, 2018). By forecasting the likely timeframe of a new project from the government in a particular time and particular location, the firm can prepare better by allocating its resources accordingly; e.g. forming a concession with the relevant specialized companies and investing in R&D for the project's work packages. With a better information retrieval and resources allocation, the firm can compete better during tendering because the offer from the company is the result of the firm's better problem understanding, innovative material and efficient process in designing, constructing, and maintaining its projects.

1.1 Problem Statement

Forecasting the introduction of a new infrastructure project to the market is not a straightforward task as an infrastructure demand forecasting due to several reasons. The first reason is the uncertainty from the government to initiate an infrastructure project to supply the presenting demand. This uncertainty is indicated by the commonly known infrastructure gap, which is the difference between infrastructure needs and the resources invested by the governments (Deloitte Research Study, 2006). The existence of this gap is proven by the infrastructure needs as high as trillions of dollars for the European Union, where $1.2 trillion over the next 20 years is required for the energy sector only (European Commission, 2006). Yearly infrastructure investment of approximately $90 billion is required in Germany alone (Jack Welch, 2006).

The second reason is when the government has decided to start building an infrastructure project to supply the demand, they need to undergo a procedure called infrastructure planning procedure to inform and seek acceptance from the public on the planned project; which obligatory in various countries such as the Netherlands, Germany, United Kingdom, and Belgium (Hobma and Jong, 2016). During this planning procedure, the demand is not the sole influential factor, for example in Infrastructure Planning Act 2012 from the Netherlands, it mentions that input from various parties would need to be taken into account throughout the procedure. The involvement of more parties would have an unexpected influence on the project and could delay the duration of decision making on new infrastructure by the government.

Based on these reasons, forecasting upcoming infrastructure projects can be considered difficult to be done due to the strong influence and interrelation of the independent variables to be considered. The context of independent variables here is the variables that influence the infrastructure demand, which drives the government to initiate the infrastructure project in the first place; and the variables that influence the decision-making process by the government on a certain infrastructure project during the infrastructure planning procedure. The variety of the variables and its interrelations produce an additional difficulty to forecasting with conventional methodology or tool. The lack of the previous study about upcoming new infrastructure project forecasting also increases the difficulty to implement the proposed forecasting approach.

Artificial Intelligence (AI) might provide an alternative to model the dynamic and complex nature of the infrastructure planning procedure by mapping-out the inter-relations of the relevant features and produce reliable duration forecasting. This is possible because AI possesses the following abilities; [1] deal with uncertainty, [2] working with incomplete data, and [3] judge new cases based on previous experiences from similar cases [Elfaki et al., 2014]. Artificial Intelligence for forecasting or prediction in the construction industry has been studied several times before. Sonmez and Ontepeli (2009), Cheng et al (2009), Elsawy et al. (2010), Wang et al. (2010), Arafa and alqedra (2011), Petroutsatou et al. (2011), Alqahtani and Whyte (2013), Lyne and Maximinio (2014), Wang et al. (2017), Zhou (2018) studied AI implementation in cost prediction task; while other researchers such as Yahia et al. (2011) and Maghrebi et al. (2014) explored the implementation of AI in duration forecasting.

In their respective researches, they prove that the implementation of AI provides a better prediction accuracy in comparison to conventional method / tool such as linear regression. AI also provides an ability to capture the non-linearity between the variables in a simpler way than the conventional models. For a model deployment in the early stage of a process, AI shows advantages over conventional methods which usually requires the knowledge and experience of the experts; because it needs only basic information which normally available during the early stage of project development. These previous studies result about AI implementation provides an argument that artificial intelligence might be able to provide an ability to its user to forecast new road infrastructure within a complex environment; In a performance level which better than the conventional methods.

However, to the author's knowledge, there is no research has been carried on which explores the application of AI in the field of forecasting task related to public infrastructure project during early development process prior to the introduction to the tender market.

1.2 Research Objective

Based on the aforementioned problem statement, this research aims to explore the applicability of AI technology to forecast the decision-making duration by the Dutch government on the publicly known infrastructure project in term of preferred design alternative.

1.3 Research Questions

Based on the research objective, the following main research question is formulated:

To what extent can an Artificial Intelligence (AI) technology be applied to forecast new infrastructure project introduction based on the decision-making duration by the Dutch Government?

In order to find an answer to the proposed main research question, the following sub-research questions are formulated:

1. What is the state of AI implementation in the field of forecasting?
2. Which type of AI is relevant for a forecasting model of an upcoming infrastructure project?
3. What are the relevant independent variables to forecast the decision-making duration of a new infrastructure project by the Dutch government?
4. How an AI forecasting model to predict decision-making duration of an upcoming infrastructure project by the Dutch government might look like?
5. How the proposed model performs in forecasting the decision-making duration of an upcoming infrastructure project by the Dutch government?
6. Does the proposed AI model fit for purpose?
7. Does the proposed AI model produce a superior result in comparison to conventional statistical methods?

1.4 Research Outline

This research is structured in eight parts, and the order is designed in a way that a proper assessment of the research can be done.

The report is started with the first part about an introduction to the research where the background of the research, the problem to be researched on, the research objective, and the research questions are elaborated.

The second part of the research is about the research methodology. In this part, the research scope and the approach to the problem are elaborated. For the research approach, an elaboration about the steps to be taken, the underlying motivation, and the respective research methodologies for each step are given.

The third part consists of the theory about artificial intelligence and its application in the field of forecasting. Based on this, the AI system which suitable to forecast the decision-making duration of a new road infrastructure project from the Dutch government is designed with a system design approach. At the end of this part, the type of AI to be used for the forecasting task is decided.

The fourth part of the research is about an exploratory study on both the data availability and the variables which might have an influence on the target value of the forecasting AI model. The end result of this part is the raw database to be processed further by the model.

The fifth part of the research consists of the AI model implementation for forecasting the decision-making duration by the government on the new road infrastructure project regarding the preferred design alternative. Prior to the implementation, the dataset pre-procession is also elaborated in this part. A possible result optimization is discussed after the model implementation's result is validated and evaluated.

The sixth part of the research discusses the result of the model and the potential factors which influence it. In this part, a comparison between conventional statistic tools and artificial intelligence model is also done.

The research conclusion, which is an answer for the main research question, is elaborated in the seventh part of the research by providing answers for each one of sub-research questions.

For the eight and the last part of the research, the future recommendation is elaborated. The following Figure 1 is presented to give a better overview of the research outline.

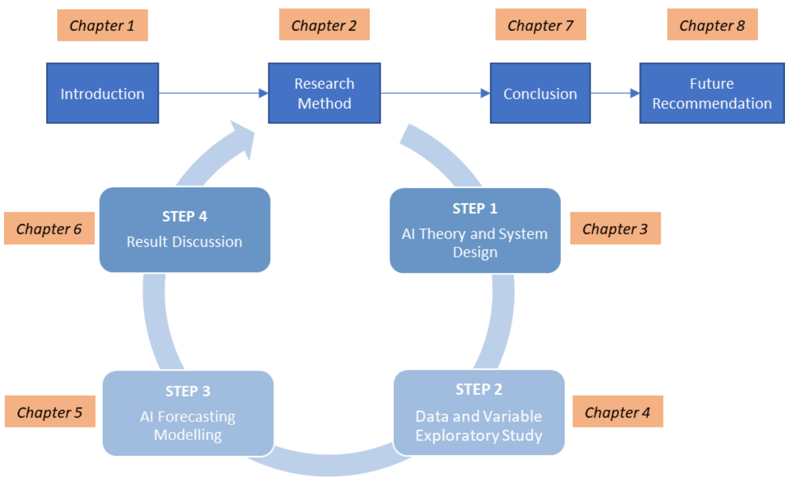


Figure 1 Research Outline

This page is left blank intentionally

Chapter 2 Research Method

Artificial Intelligence in forecasting new road infrastructure project is an innovative application which requires further research regarding the applicability of the technology. Thus, analysis regarding the state of, the potential improvement by the AI implementation, the type of technology to be implemented, the data availability, and the conceptual architecture of the system is required. The research uses several methods to achieve the desired result, which are literature study, interview, and modeling. A more detailed elaboration is explained in the subsequent sections.

2.1 Research Scope

For this research, a certain limitation on the scope has been set for the research in order to achieve the intended result within a reasonable amount of time. The research boundaries and the underlying reasons are elaborated as follows:

1. **Research Perspective**
   The Dutch construction industry is chosen with the perspective of a construction firm, ***BAM Infra BV***. With this scope limitation, a better understanding of the industry can be achieved.
2. **Legal Framework**
   Dutch Infrastructure Planning Act (*Tracewet*) is the main reference of the research to base the proposed forecasting model.
3. **Project Type**
   The research focuses on road infrastructure projects from the Netherlands government, which fall under Tracewet. This decision is taken due to the majority of infrastructure projects from the government is road infrastructure (MIRT, 2008-2019). In addition to that, the proposed model could represent the "world" better by focusing only on one type of infrastructure.

2.2 Research Approach

There are four steps to be done to answer the research question to fulfill the research objective. The first step is started with explaining why AI, theoretically, could be a useful tool in the forecasting field and designing the AI system to be used for tender forecasting. Then, it is followed by an exploratory study on both the data availability and relevant variables to the AI system. After these two steps are done, the third step is where the AI model is implemented to solve the problem at hand. After the model is implemented and the result is known, a discussion is done about the influential factor behind the model's result and the comparison of the result with the conventional statistic model's result. The illustration of the proposed research approach can be seen in Figure 2.

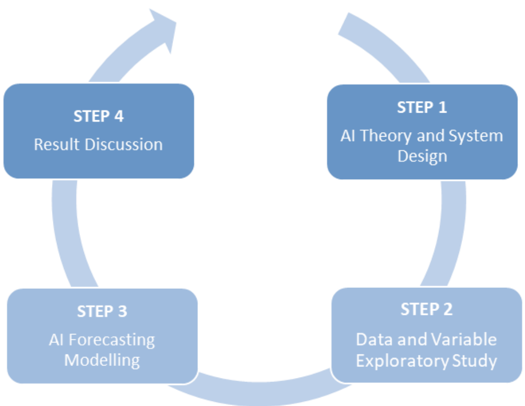


Figure 2 Research Approach Illustration

2.2.1 Step 1: AI Theory and System Design

In the first step, the research aims to provide answers to **sub-research questions number [1]** and **[2]** by conducting a literature study and experts interview. The following Figure 3 illustrates the method used to answer the sub-research questions and the corresponding topics which covered by each method.

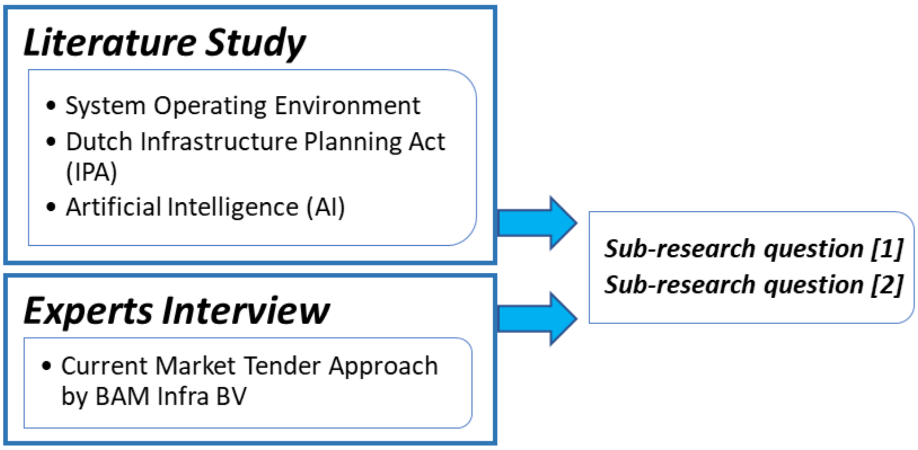


Figure 3 Step One Illustration

Artificial Intelligence Theory

This topic is studied to give an overview of AI implementations in forecasting field and how it could be useful. This overview is intended to provide a background to support the decision to use AI as the forecasting tool for a new road infrastructure project in this research. In order to grasp the concept of Artificial Intelligence and its implementation in the forecasting field, the following sources of information are used:

- A book of *Artificial Intelligence: A Modern Approach Third Edition* from Russel and Norvig (2010) is used.
- Relevant scholar papers from academic websites, such as Scopus.com, Scholar.google.com, and ScienceDirect.com. The keywords which are used as follows: “Artificial Intelligence for prediction”, ”machine learning”, and ”artificial intelligence forecasting”.

Artificial Intelligence System Design

Prior to designing an AI system, the understanding of the System Architectural Frameworks and its Operating Environment is important. As mentioned by Norvig and Russell (2010), the environment where the AI exists is seen as the “problems” where the AI is seen as the “solution”.

In order to grasp the concept of the operating environment for the proposed AI system, a system engineering approach (Watson, 2016) is used to identify and map out the system elements which included within the operating environment. The operating environment itself comprises of Higher-Order System Domain (HOSD) and Physical Environment Domain (PED). The overview of the adopted analytical perspective for AI system design of this research is illustrated in Figure 4.

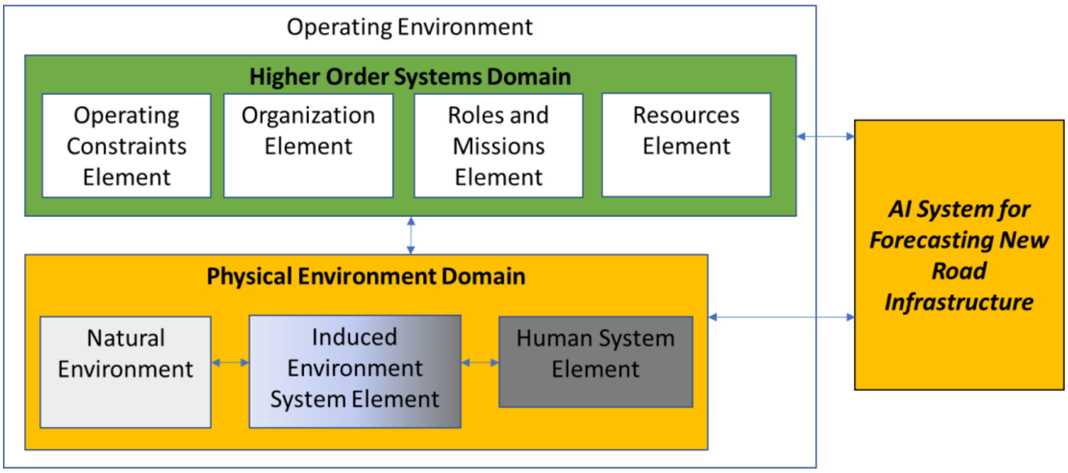


Figure 4 Adopted Analytical Perspective of Ai System Operating System Environment Architecture (Charles S. Watson, 2016)

The HOSD's principle is every system exists and performs its' purpose under the authority of HOSD. It means HOSD determines the operational boundaries of a system in design; in this research is the AI system. HOSD is comprised of four analytical elements of system: [1] Organization Element, [2] Roles, and Missions Element, [3] Operating Constraints Element, and (4) Resources Element. The definition of each is elaborated further as follows, according to Watson (2016).

a. *Organization Element* is the organization who have influence, authority, and responsibility of the system.

b. *Roles and Missions Element* is the roles allocated to the system elements within the higher-order system domain and the system utilization objective

c. *Resources Element* is the investments, raw materials, time, and money that are allocated to Physical Environment Doman and the System of Interest (SOI), in this case, is the AI system for forecasting.

d. *Operating Constraints Element* is statutory, regulatory, policy, and procedures in international, regional, and local level, which governs the SOI behavior and action.

The topics discussed in this section are chosen to determine the system elements of the HOSD, excluding *Organization Element*. The exclusion is based on the decision of the research scope in the previous section, which only focuses on the perspective of BAM Infra BV. This decision means the *Organization Element* is pre-determined as **BAM Infra BV**.

With regard to PED, an extensive analysis would be done in **Step 2**; while only a brief elaboration for each system element is done in this step based on Watson (2016) definition of each system element within the PED.

The two chosen topics to be analyzed further in this step are Dutch Infrastructure Planning Procedure and Current Market Tender Approach by BAM Infra BV. These topics are addressed by conducting an extensive literature study and experts interview.

A further elaboration about the method and data source for each topic is as follows:

I. Dutch Infrastructure Planning Procedure

A literature study is done on this topic to determine the *Operating Constraint Element, Roles and Mission Element, and Resources Element* of the HOSD from the legal perspective. The sources of information used for the literature study are listed below:

- Relevant documents such as MIRT, NCMA, Mobiliteitsbeeld, startnotitie/ startbeslissing, structure vision, trajectnota, and tracebesluit are extensively studied. All these documents are accessed from the following websites:
    - http://Google.com
    - http://publicaties.minienm.nl
    - https://www.commissiemer.nl
    - http://www.infrasite.nl
    - https://www.platformparticipatie.nl

- Relevant scholar papers from academic websites, such as Scopus.com, Scholar.google.com, and ScienceDirect.com, are extensively studied. The keywords used are as follows: *"Infrastructure Planning", "Infrastructure Planning", "Infrastructure Planning Act", "Decision Making in Infrastructure Planning".* In addition to that, the Planning and Development Law in the Netherlands book by F.A.M. Hobma & P. Jong (2016) is also studied for this topic.

II. Current Market Tender Approach

The interviews with BAM's experts are conducted to address the procedure of the company in regard to tender market approach; which determine the *Operating Constraint Element, Roles and Mission Element, and Resources Element* of the HOSD from the organization perspective. The focuses of the interviews are: **[1]** the current practice of the company's tender procedure; and **[2]** the current tender forecasting practice by the company.

There are three persons who are interviewed for this research from three departments within BAM Infra BV; ***commercial manager, pre-qualification tender department***, and ***tender strategy department***. The underlying reason is these three departments are responsible for engaging with the tender market and possess the relevant knowledge about the tender process.

2.2.2 Step 2: Data and Variables Exploratory Study

In the second step, the research aims to explore the availability of data and identifying the type of variable to be considered as the input features for the model. The study of these variables would provide an answer to the **sub-research question [3]**. The method used for both studies is an extensive literature study on the official documents from the govenment with an addition of infrastructure development studies for the variables exploratory study.

Data Availability Study

It is important to have enough historical data to be able to formulate good statistical data. The method of this step is literature study on official government documents to compile the necessary data and formulate a project database for the AI forecasting model. The official government documents which studied are as follows:

- MIRT/MIT
  The list of a public infrastructure project from the government is obtained from MIRT's section of a finished project. This document is the starting point before exploring the *startnotitie* and *ontwerp-tracebesluit* documents. MIRT itself stands for *Meerjarenprogramma Infrastructuur, Ruimte en Transpor*t. The MIRT consists of projects and programs that take place within the

physical and spatial domain. These documents allow the central and local government bodies to work together to improve competitiveness, accessibility, and quality of life in the Netherlands (MIRT, 2017).

- *Startnotitie / Startbeslissing*.
  After the name of the project is obtained from MIRT, the project's *startnotitie* is browsed and downloaded from the aforementioned websites in Step 1, specifically in the topic of Dutch Infrastructure Planning Procedure. Startnotitie is an official government document which lists the following aspects: [1] the description of the exploration area; [2] the description of the problem nature which being explored and the description of the spatial developments within the area; [3] the procedure how the interested parties would be involved during the exploration; and [4] the term for the exploration (Hobma and De Jong, 2016).

Variables Exploratory Study

1. The features or relevant independent variables, which have an influence on the decision-making duration on new road infrastructure project by the government, are identified by conducting literature study on infrastructure development studies and official government documents; which illustrated in Figure 5. In addition to that, the independent variables to be chosen are assumed to represent the *Physical Environment Domain* which has an influence on the *Higher-Order System Element.*

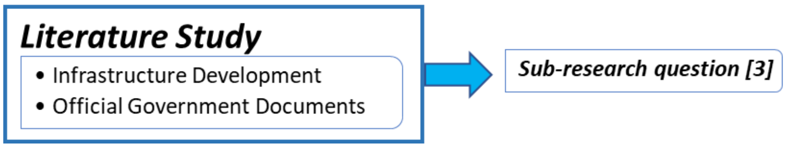


Figure 5 Variables Exploratory Study Method Illustration

After the identification, the data type of independent variables is also illustrated. This is done to give a better overview of the data type of features within the database, which are to be further processed during the AI forecasting model implementation.

2.2.3 Step 3: AI Forecasting Model Implementation

In this step, the main focuses are deploying, validating, and evaluating the artificial intelligence forecasting model. A computer software called Python 3, with the addition of Scikit-learn and Keras libraries, is used as the tool for this step.

With regard to the model deployment, the components of the model can be depicted into three, which are the Input Data, the AI Algorithm, and the Output Data.

1. Input Data: The result of Step 2, which are the project database and the relevant variables, form this model component.
2. AI Algorithm: The structure of the AI which dependent on the result of Step 1.
3. Output Data: The expected output/dependent variables from the AI algorithm are comprised under this component.

After the model is deployed, the performance of the model and its' result are validated and evaluated. The validation of the model is done by utilizing the model parameter which embedded in the Scikit-learn library which called K-Fold cross-validation; while the evaluation is done by interviewing BAM's expert about the acceptable accuracy of the model to be implemented into the company's business process. By conducting this step, the **sub-research question [4] and [5]** are answered.

2.2.4 Step 4: Result Discussion

After the result of the model is known, a discussion is done to unearth the underlying factor behind it. In addition to that, a comparison is done between the proposed artificial intelligence forecasting model with a conventional statistical model. The aim of this comparison is to prove whether the proposed AI model outperforms the conventional model in any way and justify the decision to use AI as the forecasting model. By conducting this step, the answer to the **sub-research question [6]** and **[7]** can be provided.

For the influential factor discussion, a literature study on official government documents and other researches which implemented AI for forecasting / prediction in the construction industry are done. As for the result comparison discussion, a Microsoft Excel software with XLSTAT add-on is used to simulate the conventional statistical model.

This page is left blank intentionally

Chapter 3 Artificial Intelligence Theory and System Design

Forecasting model for an introduction of new infrastructure project has not been done before to the author's knowledge and can be considered as an innovative approach. In order to design a new system as a forecasting model, an understanding of the system and its operating environment is necessary. Based on this necessity, this chapter elaborates the theory of the system of interest (SOI), which is artificial intelligence, along with its application in the field of forecasting. In order to decide the type of AI system to be used as the forecasting model, a system design approach based on the operating environment of the SOI is also done in this chapter.

3.1 Artificial Intelligence Theory

3.1.1 Definition of Artificial Intelligence

The definition of artificial intelligence in this research uses the definition by Norvig and Russell (2010), or in their term, it is called a rational agent. The definition of the **rational agent** is for each possible **percept sequence**; a rational agent should select an action that is expected to maximize its **performance measure**, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has. The action or behavior of the agent is described by the **agent function,** which maps out any given percept sequence to an action. The implementation of an agent function for an artificial agent is done by an **agent program.** The definition of the highlighted terms are as follows:

- ***Percept sequence***: History of everything that the agent has perceived.
- ***Performance measure***: The desirability of the consequences from the agent's behavior, which evaluates any given sequence of environment states.
- ***Agent function***: An abstract mathematical description
- ***Agent program***: A concrete implementation

The main concern in the AI job is the design of the agent program, which implements the agent function. There are four types of agent program proposed, but all of them use the same framework where the agent takes the current percept as input through its sensors and returns the action to the actuators, the output. There are four basic types of agent program that encompass the basic principles of an intelligent system; which are illustrated in Table 1. The process to choose which kind of agent structure is suitable for forecasting decision-making duration on new road infrastructure is elaborated further in section 3.2.

Table 1 Four Types of Agent Structure (Norvig and Russell, 2010)

Structure	Description
Simple reflex agents	This kind of agent only considers the current percept and ignores the percept history in taking actions. It is only applicable to fully observable task environments.
Model-based reflex agents	This kind of agent suits to handle the partial observable environment. It uses a **model** of the world to provide knowledge about "the way the world works" to the agent. The agent maintains an **internal state** that is dependent on the percept history, which reflects some of the unobserved aspects from the current state.
Goal-based agents	Go**al** information is required by the agent to make a decision in addition to the knowledge about the current state of the environment. The **goal** itself describes the desirable situation. Thus, the combination of goal and the model, the agent can choose actions to achieve its goal.

Structure	Description
Utility-based agents	This structure is the expansion of goal-based agents. The concept of a goal is expanded to a more general performance measure which allows the comparison of several different world states; which the agent can refer to decide how much **utility** it can have from them. This kind of agent chooses the action in order to maximize the **expected utility** of the output.

3.1.2 AI in forecasting field

Implementation of artificial intelligence in the field of forecasting, specifically in the construction industry, has been done several times before. Kim et al. (2004) compared the ability of CBR, regression analysis, and Artificial Neural Network in making a cost prediction of Korea residential construction projects; with a database which consisted of 530 historical cost data. Sodikov (2005) studied the implementation of ANN for cost estimation in highway projects. Sonmez and Ontepeli (2009) developed parametric models to forecast construction cost for urban railway system with regression analysis and ANN. Vahdani et al. (2012) presented an efficient model to improve the accuracy of conceptual cost estimation in the early phase of the project lifecycle, which called the support vector machine (SVM). Petruseva et al. (2013) proposed SVM as an algorithm to forecast the construction duration with a database of 75 projects. El-Sawah and Moselhi (2014) employed three neural networks types, Probabilistic Neural Network (PNN), Generalized Regression Neural Network (GRNN), and Back Propagation Neural Network (BP-NN). The neural networks models then compared to regression analysis models in their capability to make a cost prediction of low-rise steel structure buildings. Lastly is the research by (Zhou, 2018) which implemented Adaptive Network-based Fuzzy Inference System (ANFIS) to forecast the cost of road pavement projects. Majority of the aforementioned researches compared their innovative AI implementation with conventional methods. The results are AI implementation provides additional benefit to the forecasting task being researched. The comparison of those researched is presented in Table 2.

Table 2 AI implementation benefit in forecasting field

Research	AI Type	Comparison	AI Benefit
Kim et al. (2004)	ANN	CBR Regression Analysis	Better Accuracy
Sodikov (2005)	ANN	-	Ability to handle imprecise data Good non-linear approach Better Accuracy
Sonmez and Ontepeli (2009)	ANN	Regression Model	Better Accuracy
Vahdani et al. (2012)	SVM BPNN	Non-linear regression	Better Accuracy
Petruseva et al. (2013)	SVM	Regression Analysis	Better Accuracy
El-Sawah and Moselhi (2014)	NNs	Regression Model	Better Accuracy
(Zhou, 2018)	ANFIS ANN Random Forest SVM	Linear regression	Modeling non-linear relationship Better Accuracy

Based on the aforementioned researches, AI implementation has shown superior performance level if compared to the conventional method. AI also provides a better model capability to capture non-linearity within the dataset and to handle imprecise data which usually available during the early phase of project development. These benefit that the AI implementation can bring to the forecasting field give a solid foundation to explore the applicability of AI in forecasting decision-making duration of new road infrastructure project; which exists within a complex environment where various variables with varying degree of influence present and data which available is imprecise during the early stage of infrastructure planning procedure.

3.2 AI System Design for Project Introduction Forecasting

In order to decide on the type of AI to be used as the forecasting tool / system for future road infrastructure projects, the operating environment of the system needs to be identified first. An agent can be considered as perceiving its environment through sensors and acting upon it through actuators. Task environments are considered as the "problems" where, on the other hand, the rational agents are seen as the "solutions". The task environment mentioned above is considered as operating environment in this research; Thus, to define this term as mentioned in the Research Approach, an analytical perspective of a system operating environment by Charles S. Watson (2016) is used.

3.2.1 Operating Environment

3.2.1.1 Higher-Order System Domain

For this research, the *Organization Element* is BAM Infra BV as the company where this research is conducted. For the other three elements, integrated analysis is done from two perspectives, which are Legal and Organization perspective.

Legal Perspective

The operating constraint element from a legal perspective for the SOI is the Netherlands' Infrastructure Planning Act (IPA). IPA's effect is limited to only national infrastructure, which comprises of motorways, railways, and waterways at the national level; and relevant not only for new infrastructure but also for modification of existing infrastructure. A further elaboration about the legal framework, definition, number of steps in the procedure, and other aspects of IPA can be found in the Appendix 1 Infrastructure Planning Act (IPA). Based on the literature study on this topic, the following *Roles and Mission Element, and Resources Element* are identified for the research:

a. For *Roles and Mission Element,* the IPA has a role as a legislative system and a mission to establish societal compliance guidance and constraints of the construction industry; which eventually establishes the constraints of the forecasting AI system being developed.
b. In regard to *Resources Element*, the relevant period within the procedure for construction company such as BAM Infra BV to fulfill its objective, which is to forecast future road project, is between **Decision to Start** and **Draft Track Decision**. The decision to Start is chosen as the start of the forecasting period because it describes the project area and its nature of problem early in the process before the government involves the market in the project development. On the other end, Draft Route Decision is chosen as the end period of forecasting because, in this stage, the design is well defined and sometimes the construction market is also invited by the government to develop the design and prepare the environmental study for the **Track Decision** (BAM_3, 2019). Based on this period, the official documents such as the *Startnotitie* (released at Decision to Start) and Draft-Track decision (*ontwerp-tracebesluit)* are the main data sources for SOI. But for some projects, *Staatscourant van het Koninkrijk der Nederlanden* about the decision made by the Minister is utilized as the data source of SOI.

Organization Procedure

Based on the conducted interview with the company's experts from three different departments, namely Commercial Manager, Pre-Qualification Department, and Tender Strategy Department; the current tender practice and forecasting tender practice, which acts as the <u>*Operating Constraints Elements*</u> for the SOI are identified and elaborated further in the following sections.

a. Current Tender Practice

There is one formal tender procedure which complemented by the tender evaluation process implemented by BAM Infra BV (BAM_2, 2019). The overview of the company's tender procedure is presented in Figure 6 In relevance to the research scope, the explanation given in the subsequent paragraph would only cover the procedure until Stage Gate 2, Validation to Tender. It is because, in Stage Gate 2, the company already have all the required information to formulate a strategy to win the tender based on the department's assessment.

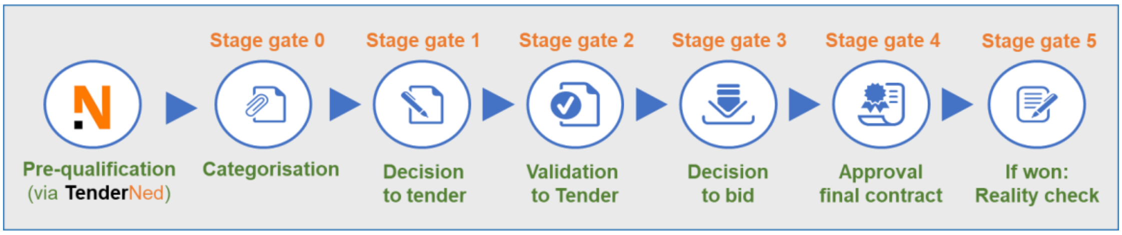


Figure 6 BAM's Tender Procedure

II. Pre-qualification
The tender procedure within the company starts when an invitation to tender is received by Pre-Qualification Department from TenderNed (BAM_1, 2019). TenderNed is part of the PIANOo Procurement Expertise Centre of the Ministry of Economic Affairs and Climate which lists all the public projects in the Netherlands ("Organisatie | TenderNed," n.d.). After the invitation is received, the documents which come in a bundle with the invitation would be distributed to the Tender Strategy Department to be further reviewed.

III. Categorization
In this stage, the project is reviewed and categorized into four different categories based on their project value and risk profile (BAM_2, 2019). Within this stage, the Tender Strategy department develops Policy to Win (P2W), identifies the opportunity of the project, and seeks strategic Partner Alignment if necessary.

IV. Decision to Tender
After the categorization stage, the Tender Strategy department reviews both the P2W and the opportunity; and Partnering Joint Venture & Supply Chain partners if necessary. In addition, the department also formulates Indicative Costing based on Key Figures if required.

V. Validation to Tender
The department conducts a kick-off meeting with the main reference to the Strategy to Win (S2W).

b. **Tender Forecasting Practice**

In term of forecasting upcoming infrastructure project tender before the announcement in TenderNed (a type of project which the government is obligated to announce to the public via Startnotitie), one of the Commercial Managers from BAM explains that there is no formal procedure within the company to monitor that kind of projects in order to make a prediction based on it (BAM_3, 2019). The non-existent of a formal procedure to assess those projects might eliminate the potential benefit that the company could gain from it.

From the procedures mentioned above, it is apparent that the main approach of the company to the market is a passive approach by waiting for the invitation to tender from the client through TenderNed before doing any business process to engage with the market, that includes forecasting practice. Based on these procedures, the *Roles and Mission Element, and Resources Element* of the SOI are as followed:

a. For the *Roles and Mission Element*, the role of the organization tendering practice is a resource system with a mission to provide the organization with projects to be executed by winning a tender.
b. In regard to *Resources Element*, the proposed forecasting model is suited to be used by the Pre-qualification Department in prospect management; which is the current business process of BAM to map out the potential project in the future, which announced in TenderNed, based on various aspects and assign its resources accordingly.

3.2.1.2 Physical Environment Domain

The PED comprises of three system elements; which are *Natural Environment, Human System Element, and Induced Environment System Element*. The definition of each of this system element, according to Charles S. Watson (2016) and the instances in relation to the SOI are as follows.

a. *Natural Environment* is the atmospheric, living, geophysical, and aquatic entities that comprise the Earth; for instance, the existing habitats for flora and fauna in the projected area of the new infrastructure project, air, the lake, river, sea, etc.
b. *Human-System Element* is systems created by humans that have interaction with the SOI. The examples for the SOI at hand are the road infrastructures, the cities, bridges, tunnels, government system, election, businesses, companies, cars, etc.
c. *Induced Environment System Element* is the situation which resulted from the interaction of *Natural Environment* and *Human System Element;* such as air pollution, sound pollution, traffic jam, disturbance on the existing habitat due to new construction, etc.

3.2.2 Artificial Intelligence for Infrastructure Tender Forecasting

Based on the interaction of road infrastructure tender forecasting's operating environment with the AI system; the attributes of the system can be identified. These attributes are the foundation for the **agent program** decision to be used for the AI system. The overview of the system attributes based on its operating environment is presented in Table 3

Table 3 AI System Attributes based on the interaction with its Operating Environment (Norvig and Russell, 2010).

System Attributes	Interaction with Operating Environment
Observability	The environment can be considered as partial observable, the reason being that the agent cannot access the environment all the time. The Resources Element from the Higher-Order Domain designated to use an official document from the government which not available all the time. The government releases the documents when there is a new project, and sometimes the data are removed at one point in time.
Number of agents	Single-agent is designated to interact with the environment. In this case, is the number of AI systems in design.
Agent process impact on the environment state	The impact of the AI executing the algorithm is uncertain because, within the operating environment, the organizational element and the organization procedure might be affected. However, the rest of the system elements within the operating environment will not be affected.
Agent Experience	The agent experience in this operating environment is Episodic; because the agent's experience is separated into several atomic episodes. The episode is when the agent receives a percept. The next episode does not depend on the actions taken in previous episodes. The experience, in this case, is when the agent percept the operating environment.
Change in the environment	The environment in which the AI system exists is continuous. It is because the environment does change while the agent is deliberating. This is in line with the dynamicity of the Physical Environment Domain. For example, the features identified as relevant for the decision-making process (4.2.1), such as the traffic jam and air pollution, will always keep changing over time. The political party composition also changes in every election.
Knowledge of outcomes	There is a necessity for the agent to learn before it decides on a certain outcome because the outcomes are unknown. This is expected due to the nature of forecasting or predicting the future.

By considering the aforementioned AI system attributes, the ***Model-based reflex agent*** is deemed to be suitable for the AI system; because this kind of agent program is capable of handling the partial observable environments. It uses a model of the world to provide knowledge about "the way the world works" to the agent. The agent maintains an internal state that is dependent on the percept history, which reflects some of the unobserved aspects from the current state. In order to perform better, the agent program usually designed with the capability to **learn**. The advantage of learning that the agent is able to operate in unknown environments and is more competent in comparison to its initial knowledge.

In addition to that, the way the agent program represents its environment is also important to consider the type of AI system for the forecasting problem. To better address, the uncertainty aspects of the forecasting environment **factored representation** is better suited for it. Factored representation splits up each state into a fixed set of attributes or variables, each with its own value (Norvig and Russell, 2010).

In conclusion, the type of AI which has the attributes of a model-based reflex agent, with a capability to learn, and has a factored representation of its environment is **machine learning (ML)**. Since ML itself has various branches, a decision has to be made in regard to a certain type of ML for this research. There are several types of ML which suit a forecasting or prediction task: *Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Functions (RBF), Generalized*

Regression Neural Networks (GRNN), K-Nearest Neighbour regression (KNN), CART regression trees (CART), Support Vector Regression (SVR), and Gaussian Processes (GP) (Ahmed, 2010). Amongst these ML types, there are three types under the branch of artificial neural networks (ANN); namely MLP, BNN, and GRNN. In addition to that, from the conducted literature study on existing researches about the AI implementation in the field of forecasting, it shows that majority of those researches implemented ANN as the tool in their respective forecasting research. Thus, ANN is chosen as the type of AI to be implemented in this research.

3.2.3 Artificial Neural Network (ANN)

This branch of ML is inspired by a hypothesis which says electrochemical activity in networks of brain cells called neurons is the driver of mental activity. The following Figure 7 illustrates a simple neuron mathematical model proposed by McCulloch and Pitts (1943). The main concept of this model is the neuron "fires" when a linear combination of its input surpasses a certain threshold which implements a linear classifier.

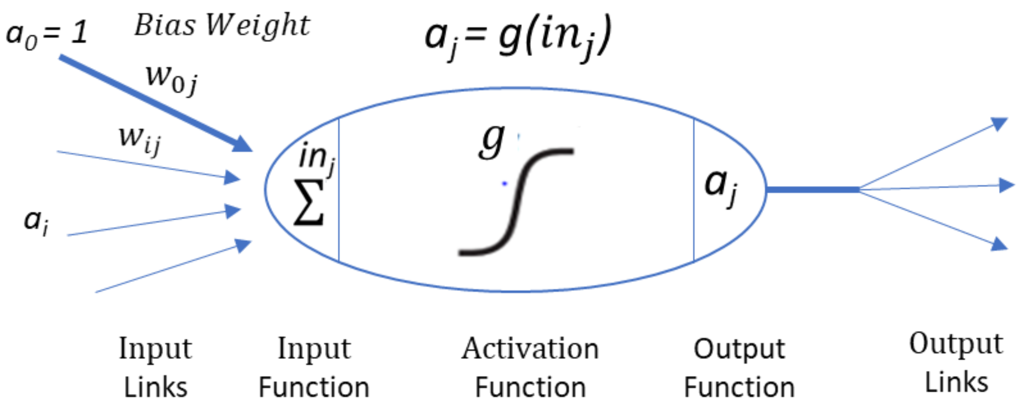


Figure 7 A simple neuron mathematical model proposed by McCulloch and Pitts (1943)

By referring to Figure 7**,** the neuron output activation function is as follows:

$$a_j = g\left(\sum_{i=0}^{n} w_{i,j} a_i\right) \tag{1}$$

Where:

a_j = Neuron's output activation function

a_i = Output activation of neuron *i*

w_{ij} = The weight of the link from the previous neuron to this neuron

A neural network is a collection of the "neurons" which connected together; which makes both the topology and properties of the neurons determines the network properties.

ANN structures

As mentioned above, ANN consists of interconnected neurons. The connection of the neurons within ANN are denoted by w_{ij}, which determines both the sign and strength of the connection. In regard to the networks, ANN usually constructed in several layers which each neuron receives input only from the neurons in the immediate previous layer. The following Figure 8 illustrates the typical structure of ANN with three layers. The first layer, which is the input layer, represents the input variables listed in the dataset. The second layer is the hidden layer, which represents the relationship between the independent and dependent variables. The number of the hidden layers and its neurons are usually defined with trial and error. The dependent variables or the output of the algorithm is represented in the third layer, the output layer. The combination of these neurons forms a model which to be used to forecast or predict the duration of the decision making on a newly announced road infrastructure project.

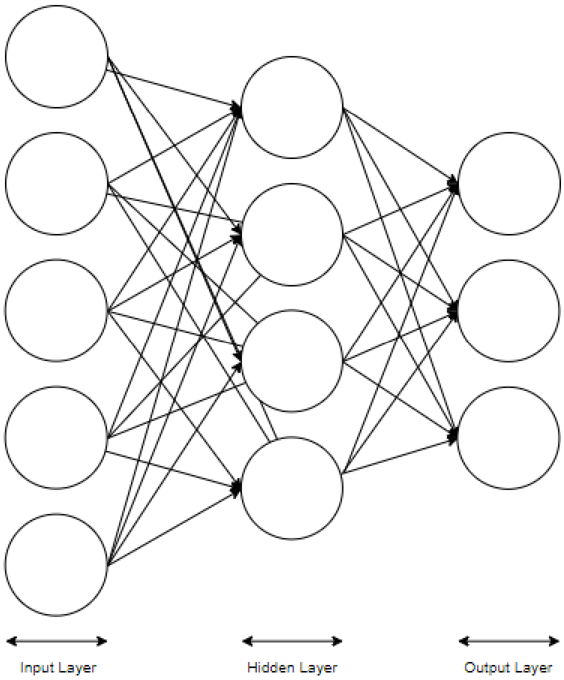


Figure 8 Multilayer ANN Structure Illustration

Learning in Multi-layers ANN

As mentioned before, the connection of the neurons within ANN are denoted by weight w_{ij}. The objective of ANN is to find the weight for each connection by mapping out the input features set with the output features. It is difficult to realize this objective due to the high number of connection between neurons present in the network. Thus, an algorithm is used to solve this search or optimization problem and identify the best possible set of weights which can provide a good prediction. This search of best possible weights set is what known as learning for an ANN model.

Normally, an ANN is trained by using an optimization algorithm called stochastic gradient descent, and the weights are updated by using an algorithm called error back-propagation. The “gradient” term in “stochastic gradient descent” stands for a gradient of error.

Firstly, an ANN model with certain weights set is used to make predictions on the dataset and then the error, which is the difference between the predicted value and the actual value within the dataset, is calculated. The gradient descent algorithm then seeks to modify the weights so that for the next run, the error would be reduced. Which means, the optimization algorithm is steering down the error slope or gradient. This process is repeated as many as the set iteration number in the model parameter. With regard to the optimization algorithm, it is normally referred as cost function or a loss function; with the calculated value from the function is simply called “loss” (Goodfellow et al., 2016).

3.3 Conclusion

Theory shows that AI could potentially be a tool for a prediction task of decision-making duration on new road infrastructure with the advantages that it has in comparison to the conventional methods. The previous studies about an AI implementation for prediction task in construction industry prove that AI has advantages over the conventional method in the form of better accuracy, ability to handle imprecise data, and good non-linear approach

The analysis on *System's Operating Environment* and its relationship with the proposed AI system provides knowledge about the type of AI for the forecasting problem; which is an Artificial Neural Network (ANN). The following Figure 9 is presented to give an overview of the *Operating Environment* and how artificial neural network interacts with/represent its operating environment to fulfill its purpose.

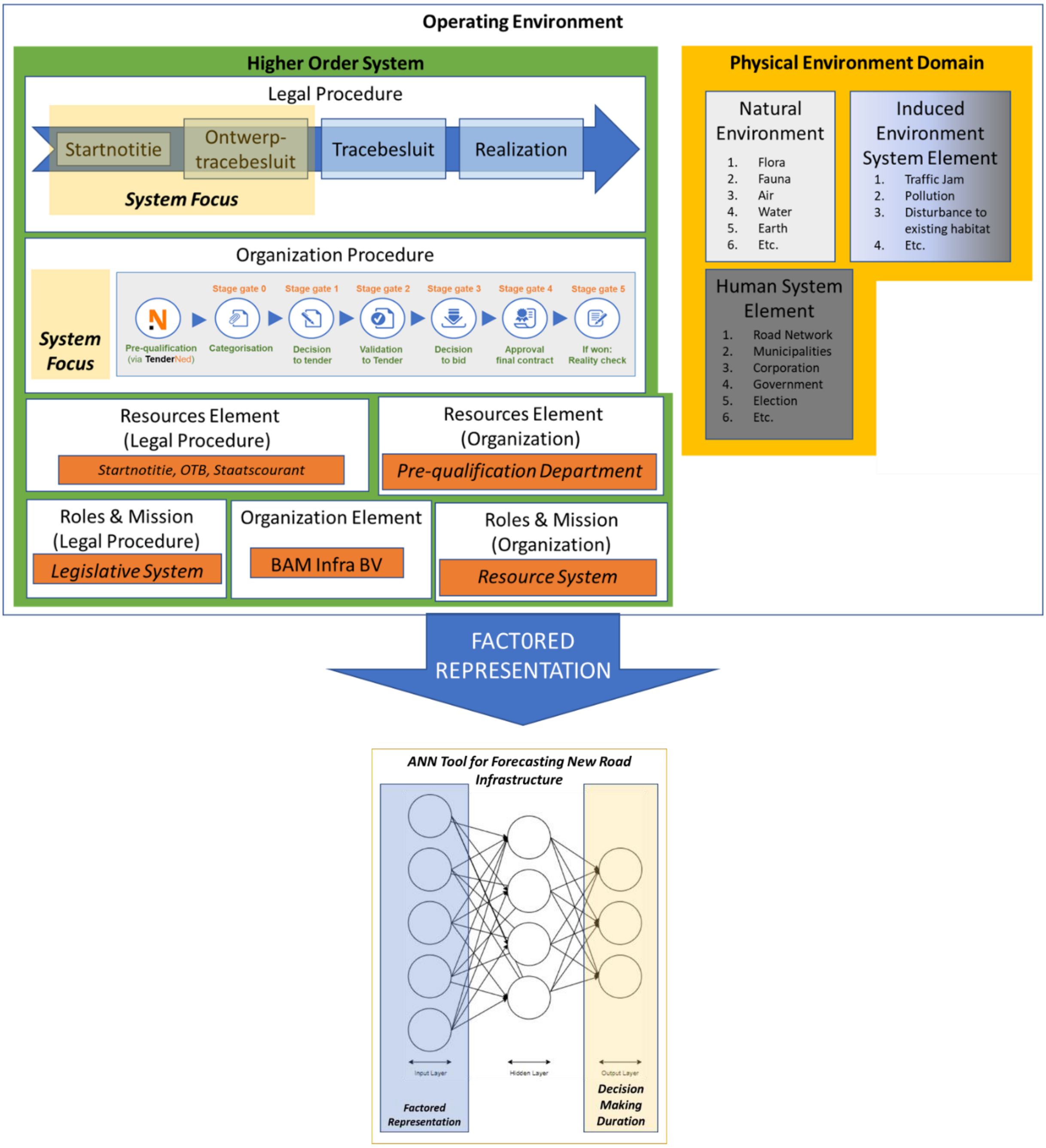


Figure 9 ANN for Decision-making duration on New Road Infrastructure Illustration

This page is left blank intentionally

Chapter 4 Data and Variables Exploratory Study

This chapter elaborates the process and result of data availability study and the study of the variables. For the data availability study, a closer look at the official documents released by the Dutch government is done to learn about the number of road infrastructure available. For the other study, which is variables study, an elaboration of the variables or features identification, and the visualization of the features are presented.

4.1 Data Availability Study

In the previous chapter, it is clear that *Resources System Element* is important for the SOI. For this research, as mentioned above, the main *Resources System Element* for the AI is the official documents from the Dutch government. Based on this decision, a data availability study is done to find the relevant documents for the AI system.

The result of this study is a database with a total of 95 national infrastructure projects, which covers road projects from the Dutch government with the *startnotitie* released ranging from the year of 1990-2015. The data is gathered from the government's official website, as explained before, and the number of project for each year is represented in the following Figure 10. It is shown that the number of the project released for each year is varied and, in some years, the government did not release a new road infrastructure project through *startnotitie* (i.e., 2000, 2001,and 2012).

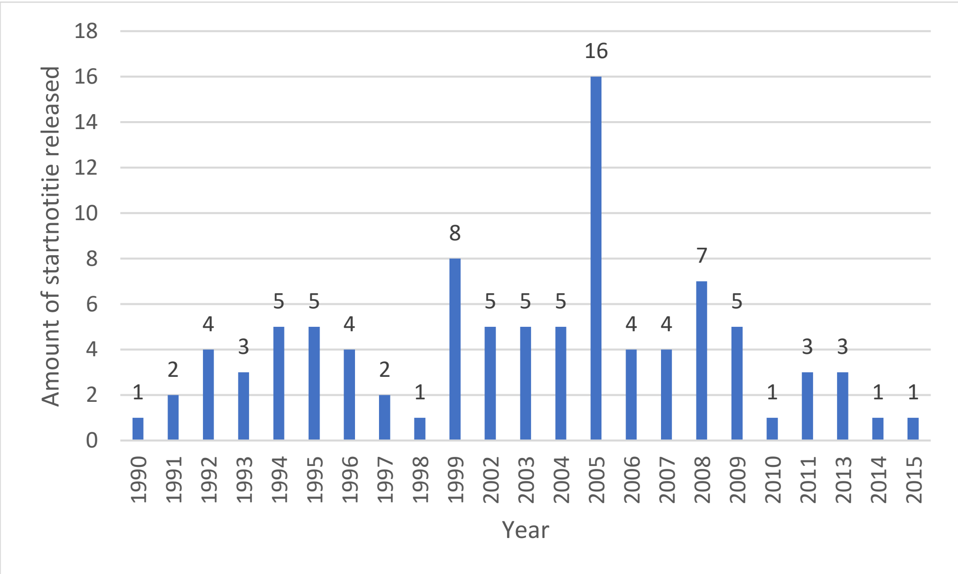


Figure 10 Dataset Quantity Representation

4.2 Variables Study

4.2.1 Variables Identification

As mentioned before in Research Approach, there are two methods for the identification of the variables, which are the literature study on infrastructure development and the government official document (i.e. *startnotitie*). All these variables are used as the data 'features' for the input data component in Chapter 5 AI Forecasting Model.

Independent Variables

These independent variables are *location-based or* representing the attributes of a project spesific location/city/province/countries. For this research, the scope of these variables is adjusted to three different levels; which is country, province, and project specific location. Based on the aforementioned studies and assumption, the following list of independent variables is decided to be explored further (Figure 11).

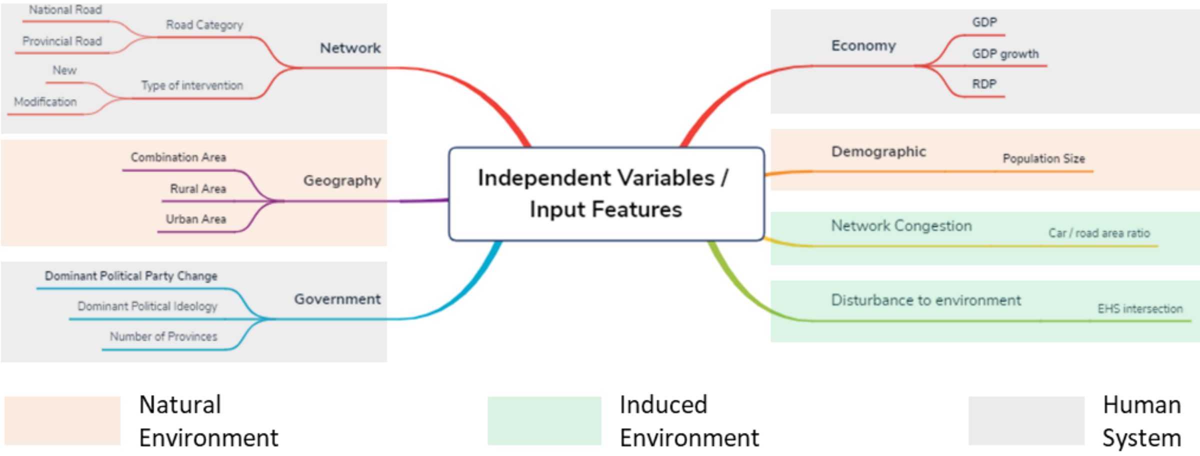


Figure 11 Independent Variables Overview, and it's Physical Environment Domain Representation

1. *Road Category*
   From the description of startnotitie, there are two categories of the road which were studied; Motorways (which denoted by A on the road name) and Provincial Roads (which denoted by N on the road name). This variable can be seen as an indicator of the sense of urgency for the decision-making of the respective project. The following figure illustrates the two categories based on the information listed in the startnotitie A1 Eemnes-Barneveld (1999) and startnotitie N11 Zoeterwoude-Alphen aan den Rijn (2002).

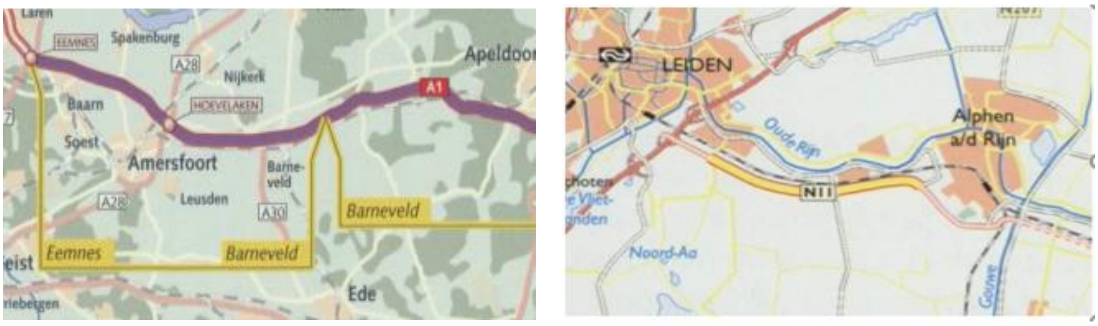


Figure 12 Road Category Illustration

2. *Type of Network Intervention*
   In every startnotitie, it is explained what kind of problem is being explored and in which part of the road network. There are two types of intervention which considered in this research, which are the new road line and modification of the existing road line. This variable is assumed to be relevant because the exploration study for a completely new route and the expansion of an existing route might differ in term of decision-making duration due to the requirement to release the required land for the new infrastructure. For a new route project, the land-use of a certain area might be required to be changed, and the former inhabitants need to be

expropriated first (Hobma and Jong, 2016); which potentially increase the number of interested parties in the project plan that lead to longer duration of decision-making process on the preferred design alternative.

3. *Gross Domestic Product (GDP)*
   This variable can be considered as an indicator of the country's infrastructure demand and the government's ability to fund the project. With a better economy, the demands for different types of infrastructure would be increased because of the increase in production and main resources usage. Also, the same can be said with the government's ability to supply the demand (Rothman et al., 2014).
4. *Three years of GDP growth*
   In addition to the *GDP* variable, the GDP growth of the Netherlands for the last three years is also considered. This variable is chosen to give an indication of the government's performance in regard to GDP over the period of the last three years and investigate the influence it has on the decision-making duration.
5. *Regional Domestic Product (RDP)*
   The RDB and RDB/capita that are listed in the project database is the % of a certain province RDP with the GDB of the Netherlands as a country. The underlying reason to choose this variable is similar to the GDP variable where an increase in the economy would also increase the demand for various types of infrastructure, but in a smaller scope. This variable can be an indicator of the financial ability of the province to fund the project and the importance of the province to the central government based on the economy level of the respective province.

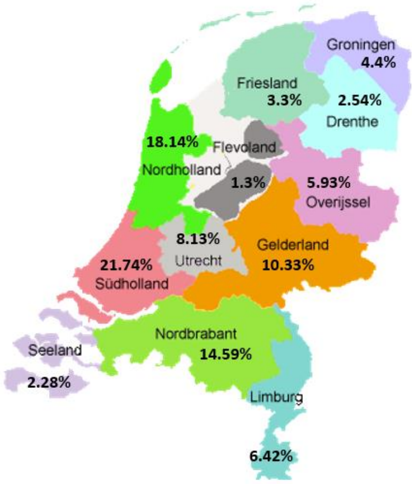


Figure 13 RDP Percentage of the Netherlands Provinces in 1995 (CBS)

6. *Population size*
   The population size of a province is assumed to be relevant because not only it produces the infrastructure demand at the provincial level (D.S. Rothman et al., 2014) behind the initiation of *startnotitie*; but also determines the number of affected people, both negative and positive, because of the new potential infrastructure project.
7. *The geographical profile of the project area*
   The project location's geography would influence the type and scale of the planned project impact. There are three categories for the geographical profile of the project area which considered in this research: **urban, rural,** and **combination**. In relation to the project impact (e.g., environment, social, and economy), the acceptance behavior of the interested parties is assumed to differ for each geographical profile. This assumption is based on the fact that several large infrastructure projects were delayed due to protest from the interested parties

due to the project area's geographical profile; which led to a delay or modification during scope definition of the respective projects; e.g., Betuweroute and HSL-South projects (Hertogh and Westerveld, 2009). Both of these examples are railway project, but the infrastructure planning procedure is the same as road infrastructure project, which makes it comparable.

8. *Ecologische Hoofd Structuur (EHS) Intersection*
   EHS is a coherent system of nature development areas, connecting zones and core areas that are prioritized within the landscape and nature policy of the Netherlands government (*Startnotitie A27 Plusstrook,* 2003*)*. This variable is decided to be included in the model to represent the induced environment system element between the road infrastructure project with the flora and fauna which live in EHS. This variable is chosen also based on the fact that the interested parties, including the environmental groups, have legal rights to bring forward their opinion on infrastructure projects being discussed by the Minister (IPA, 2012). This could lead to longer duration of decision-making on project scope and design.
9. *Car to road area ratio (car/km2)*
   This ratio is decided as one of the independent variables because it also represents the induced environment system element of traffic jam. The higher the ratio between the number of cars with the road area, the higher the probability of occurrence of traffic jam and potential environmental deterioration in the surrounding area (Fan and Yan, 2009).
10. *Dominant Political Party Change*
    As mentioned in the system attributes, the operating environment is dynamic or continuously changing. The government is seen as a human system element within the physical environment domain, which comprises of various political parties with different ideology and goal. The dynamicity of this composition can be represented with the change during the election. This variable is chosen to represent this dynamicity. The influence of election on the public spending, which this research assumes to have an influence on the decision-making duration on new projects, is acknowledged by Sakurai and Gremaud (2007), Salvato et al. (2007), and Sakurai (2009).
11. *Dominant Political Ideology*
    As mentioned above, the government is comprised of political parties with diverse ideology. This variable is chosen because based on a research done by Hiromoto (2012), in regard to the ideology of the political parties affecting governmental action in term of public spending; when centrist and right-wing parties are in power, they have a tendency to spend lesser in comparison to left-wing parties. The public spending itself includes infrastructure spending, which then assumed to have a relation with duration of decision-making on a new road infrastructure project.
12. *Number of Provinces*
    This variable is considered to be influential because the number of provinces involved in the project is always mentioned in the *startnotitie.* The number of involvement is assumed to be an indication that there is more interest to be considered and potential for conflict is higher as the number of province increases; which eventually affects the duration of decision-making on a new road infrastructure project. It is because, in IPA, it is stipulated that lower bodies of government have been legally ensured to influence the decision-making process of new infrastructure.

Dependent Variables

The dependent variable or the output of the forecasting model is the ***decision-making duration*** on the project scope / design alternatives which normally known as 'preferred alternative' and stipulated in *ontwerp-tracebesluit* after the release of *Startnotitie.* For the project which decided to be delayed or canceled by the government, the specific decision date is found in the MIRT document, Staatscourant, or information from https://www.commissiemer.nl/.

4.2.2 Variables/Features' Data Type Illustration

Twelve independent variables/features are identified to be the potential decision-making duration drivers, namely*: Road Category**,** Type of Network Intervention**,** Country GDP, Country GDP Growth average of the last 3 years, Province RDP, Population Size, Geographical Area, EHS Intersection, Dominant Political Party Change,* and *Dominant Political Ideology*. All these variables have a different type of data which have a big influence on how the designated algorithm performs. According to Kelleher et al. (2015), there are six types of data that can be prepared prior to a model building which illustrated in the following Table 4.

Table 4 Data Types (Kelleher et al., 2015)

Type	Description
Numeric	Arithmetic operation is possible
Interval	Ordering and subtraction are possible
Ordinal	Ordering is allowed, but arithmetic is permitted
Categorical	Ordering and arithmetic are not possible
Binary	Just two values
Textual	Free-form and short, text data

To provide a better understanding of the variables to be used as the input features and the output of the forecasting model data set; an illustration is provided in the following Table 5.

Table 5 Data Type of the Variables

	Features	Categories	Data Type	Unit	Data Source
Input Features	Road Category **[RC]**	I. National Road (A) II. Provincial Road (N)	Categorical	N/A	*Statnotitie*
	Network Intervention Type **[NI]**	I. New Route II. Expansion	Categorical	N/A	*Statnotitie*
	Country GDP **[GDP]**		Numeric	Million euros	CBS
	Country GDP Growth of the last 3 years **[GGDP]**		Numeric	%	CBS
	Province RDP **[RDP]**		Numeric	%	CBS
	Population Size **[P]**		Numeric	%	CBS
	Geographical Area **[GA]**	I. Urban II. Rural III. Combination	Categorical	N/A	*Startnotitie*

	Features	Categories	Data Type	Unit	Data Source
	EHS Intersection **[EI]**	I. Intersects with the EHS area II. No intersection	Binary	N/A	*Startnotitie*
	Car / road area ratio **[CR]**		Numeric	Cars/ km^2	CBS
	Dominant Political Party Change **[DPC]**	I. Change of political party from the last election II. No change / stable III. Election Year	Categorical	N/A	*Kiesraad.nl*
	Dominant Political Party's Ideology **[DPI]**	I. Right-Centre ii. Centre iii. Left-Centre iv. Election year	Categorical	N/A	(Keman, 2008) (Colomer, 2008) (Bremmer,2012)
	Number of Provinces **[PV]**		Numeric	N/A	*Startnotitie*
Output	Decision-making duration		Numeric	Year	*Ontwerp-tracebesluit Staatscourant*

4.2.3 Features Visualisation

It is clear that there are three data types within the compiled database, which are categorical, binary, and numeric. The features are visualized to give an overview of the frequency of each category within each feature (*Individual Visualisation*) and the relationship between the features (*Pair-wise Visualisation*).

Individual Visualisation

The first feature is Road Category. Most of the projects which initiated via *startnotitie* are a national road with a frequency of 79% in comparison to the provincial road with 21% frequency from the total data. This proportion is reasonable because the national road has a high economic impact due to its utilization, both personal and commercial. Most people use the national road to travel to their workplace in major economic cities such as Amsterdam, Rotterdam, and Utrecht, from their home in neighbor cities or cities which located quite far due to the relatively high living price within those major economic cities. Most companies also distributed their goods mainly through national roads in comparison to provincial roads.

For the second feature, Type of Network Intervention, majority of the data entry is a modification project on existing road line with a frequency of 86% in comparison to 14% of new road line project. This distribution is also reasonable due to relatively higher cost of new road line project in term of feasibility study, design, tendering, environmental impact, construction, and compensation in comparison to the modification project on existing road line.

For Geographical Profile feature, Rural area dominates the data set with a frequency of 57%, followed by Combination between the rural and urban area with 24%, and Urban area with 19% of data entries. This distribution shows that the government has a preference to initiate an infrastructure project in a rural area. This is reasonable due to the relatively lower money required in term of compensation for land acquisition and mitigation measures if necessary.

The fourth category is the EHS intersection. From the collected startnotitie, the EHS intersection is mentioned in 67% of the dataset. It is reasonable because as mentioned before, most of the road infrastructure project, which was initiated by the government is within the rural areas and EHS is normally located in that type of area.

In relation to politic influence on the projects, Change of Dominant Party and Dominant Political Ideology are visualized. For Change of Dominant Party during the national election, most of the road infrastructure project is published when the political environment is relatively stable with no change in regard to the dominant party from previous regime. It is shown by 62% of the dataset being dominated by the project which released when there is no change of political party, and then followed by 22% and 18% of projects released on an election year and when there is a change of political party respectively.

Last but not least, the dominant political ideology is visualized. Most of the road infrastructure projects are released when the dominant political ideology is Centre, which shown by its 48% share within the dataset. Then it is followed by uncertain (during an election), center-left, and center-right with a percentage of 30%, 13%, and 9% respectively. To give an overview of each pair-wise visualization, Figure 14 is presented.

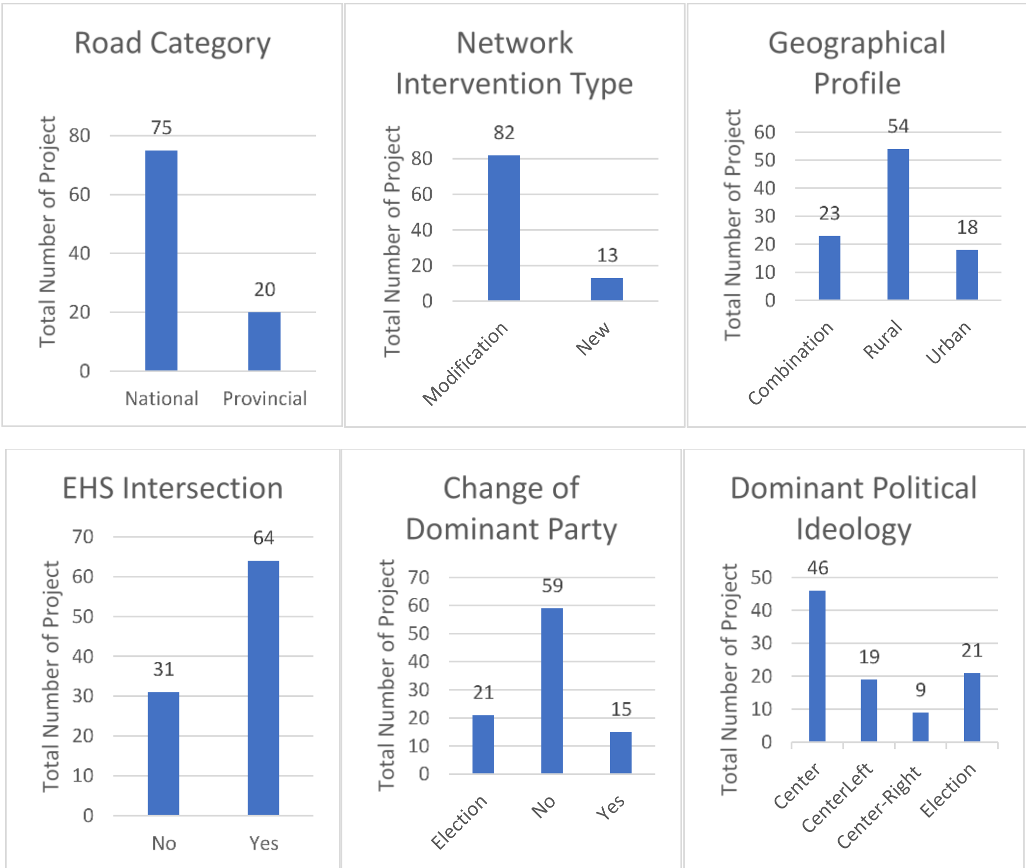


Figure 14 Categorical Data Visualisation

Pair-wise Visualisation

A box plot graph is used to visualize input features and duration of decision making on a new infrastructure project. This approach is used to identify the outliers of the categorical features and eliminate it to improve the performance of the model. Outliers itself are values that plotted far from the central tendency of a certain feature (Kellehe et al., 2015). To give an overview of the box plot's structure, the following Figure 15 is presented. Based on this structure, the outliers which located above the 3rd quartile plus 1.5 times inter-quartile range (IQR) and below the 1st quartile minus 1.5 times IQR will be removed.

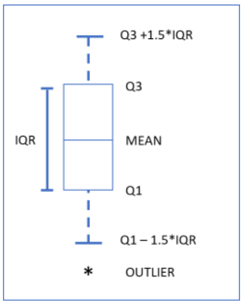


Figure 15 Box Plot Structure Illustration

The first box plot (Figure 16) shows the Road Category feature and its related decision-making duration. There are three national road projects which considered as outliers and thus removed from the dataset for the ANN modeling. The box plot also shows that the provincial road has a lower duration average compares to the national road.

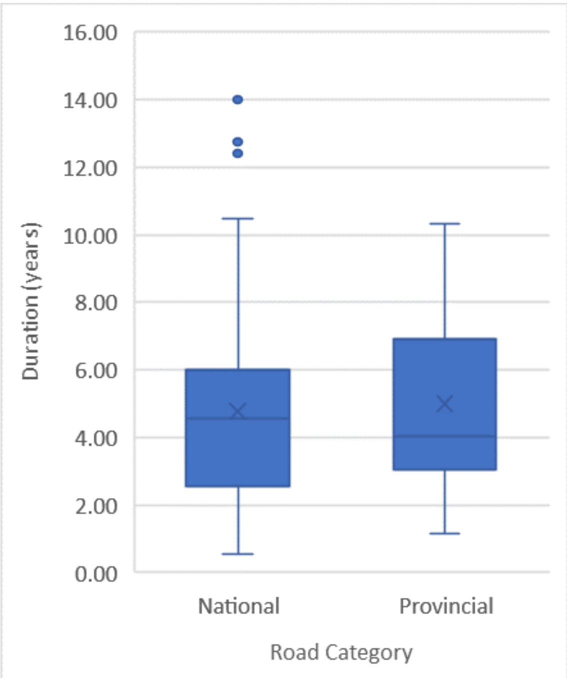


Figure 16 Road Category Box Plot

The second box plot (Figure 17) illustrates network intervention and its related decision-making duration. The box plot also shows there are three outliers, which are the same as what the previous road category box plot shows. The average duration of road modification project is slightly lower in comparison to the average of new road modification project.

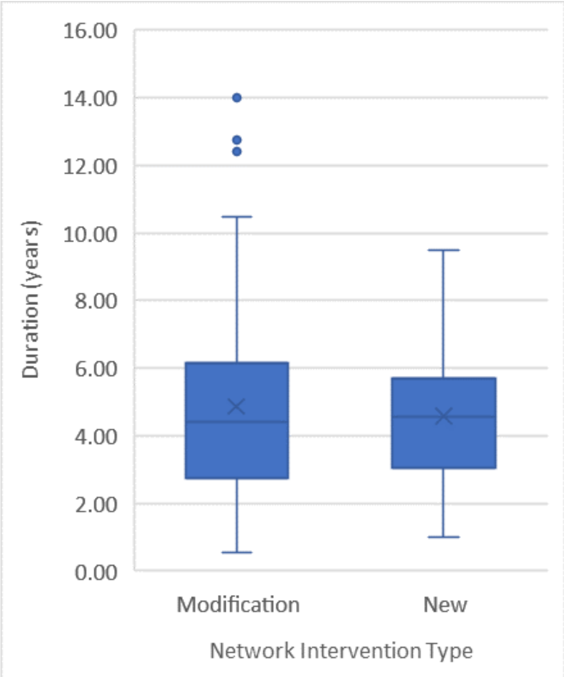


Figure 17 Network Intervention Box Plot

The third box (Figure 18) plots the geographical profile feature with its corresponding decision-making duration. From the box plot, three outliers which similar to the previous two box plot and one additional outlier are identified. The box plot also shows that the duration average of the rural profile is slightly lower compares to urban profile, while the combination profile has the lowest average.

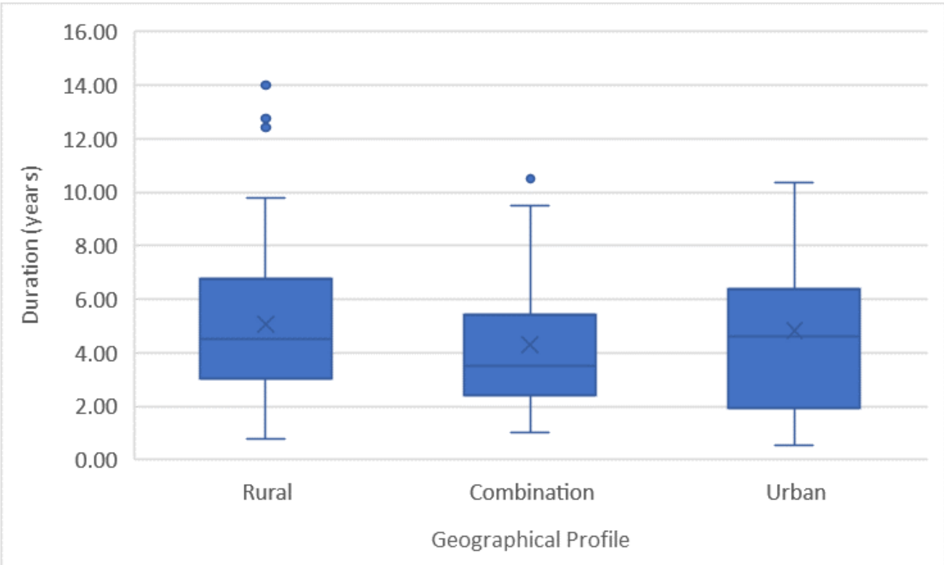


Figure 18 Geographical Profile Box Plot

The fourth box plot (Figure 19) illustrates the EHS intersection with its related decision-making duration. One outlier which similar to the previously mentioned outliers is identified from this box plot. The box plot shows that the projects which initiated in area where no intersection with EHS exists have a lower average compared to where EHS intersection is present.

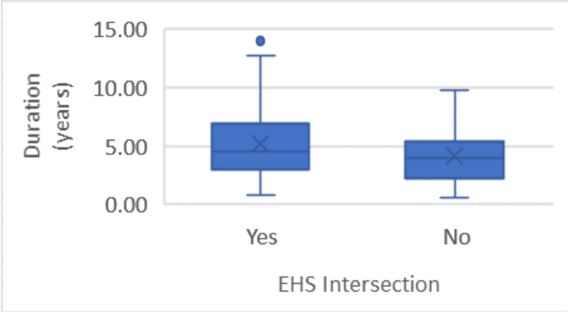


Figure 19 EHS Intersection Box Plot

The fifth box (Figure 20) plots the change of dominant party in the government feature with its related duration of decision-making on a new road infrastructure project. There is one outlier identified from the box plot with the same duration as the previously identified outliers. The average of decision-making duration of projects which initiated at a particular year where there is a dominant political party swift during an election is the highest amongst the three categories. The second highest average-duration is the projects which announced when no political party swift present and followed by the projects which announced during an election year.

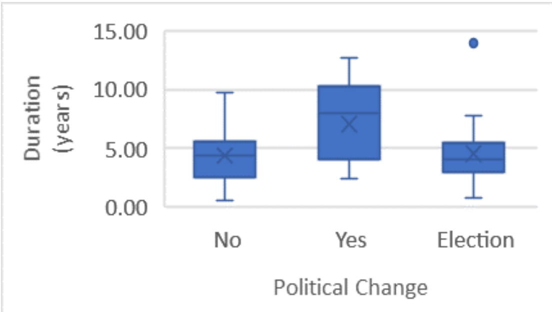


Figure 20 Political Party Change Box Plot

The sixth and the final box plot (Figure 21) shows the dominant political party ideology with its corresponding duration. One outlier with the same duration is identified from the box plot, and the overview of the average decision-making duration is shown. The project which initiated by the government with an ideology of CenterLeft has the highest average duration of decision-making followed by projects initiated by Center, uncertain (political year), and Center-Right minded government respectively.

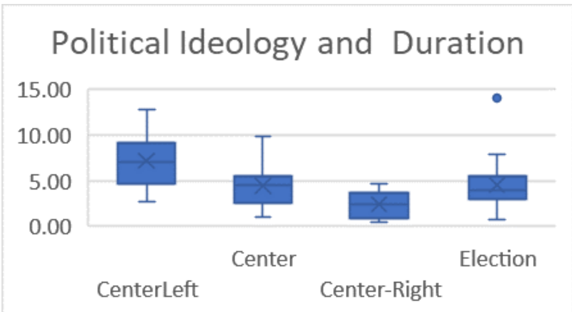


Figure 21 Dominant Political Ideology Box Plot

For the continuous features, the correlation between the input feature and output is used as the indication of their relationship. Correlation between two variables (*a* and *b*) is calculated with the following formulas:

$$corr(a,b) = \frac{cov(a,b)}{sd(a)x\ sd(b)} \tag{4.1}$$

$$cov(a,b) = \frac{1}{n-1}\sum_{i=1}^{n}((a_i - \bar{a})\ x\ (b_i - \bar{b}) \tag{4.2}$$

Where:
corr(a, b) = Correlation between feature *a* and *b.*
a_i and b_i = The i^{th} instance v*alues of features a* and *b* within the dataset
sd(a) and *sd(b)* = Standard deviation of feature *a* and *b*
cov(a,b) = Covariance of feature *a* and *b*
$\bar{a}$ and $\bar{b}$ = Mean values of features *a* and *b*

The correlation values fall within the range of [-1,1], where values close to -1 indicates a very strong negative correlation and an indication of a very strong positive correlation on the other end. As for the current dataset, the highest correlation with the decision-making duration is the GDP of the Netherlands with the value of -0.44. This is a negative value which shows a negative correlation and an indication that the higher the country GDP value is, the lower the decision-making duration on a road infrastructure project from the government would be. The following Figure 22 illustrates the correlation between the aforementioned two values.

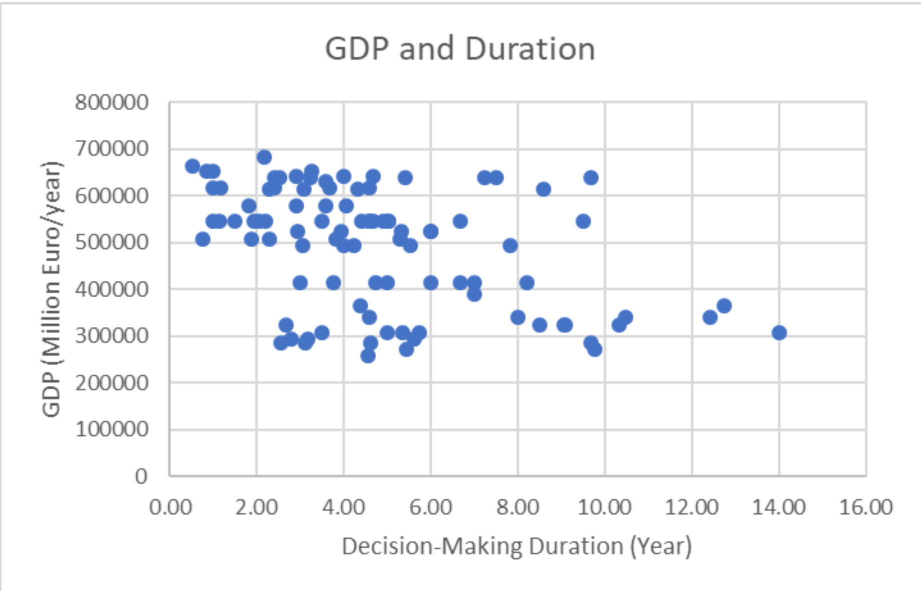


Figure 22 GDP and Decision-Making Duration

Although the correlation value for some other independent variables are low, they are still considered for the model because in a multiple regression model such as ANN, although an independent variable is not related to the dependent variable, it can be useful because it helps whittle away otherwise unexplained variance in one or more of the remaining independent variables. This kind of variable is referred to as a suppressor variable (Thompson and Levine, 1997). The following Table 6 presents the correlation value for other independent variables to be considered in the model.

Table 6 Individual Correlation of Input Variables with the decision-making duration

	GDP	3 years average GDP Growth	Province RDP	Car/Road Ratio	Population	Number of Province
correlation	-0.44	0.37	-0.09	-0.11	0.01	0.097

4.3 Conclusion

There are ninety-five national road infrastructure projects identified by conducting data availability study, which covers road projects from the Dutch government with the *startnotitie* released ranging from the year of 1990-2015. With regard to the variables study, there are twelve independent variables identified as the input features for the model; which are Road Category, Type of Network Intervention, Gross Domestic Product (GDP), GDP growth of the last 3 years, Regional Domestic Product (RDP), Population size, the geographical profile of the project area, Ecologische Hoofd Structuur (EHS) Intersection, Car to road area ratio (car/km2), Dominant Political Party Change, Dominant Political Ideology, and Number of Provinces. All these variables are visualized, and the pair-wise visualization identified four outliers, which eventually were removed from the database. This removal makes the final database for the model consists of ninety-one road infrastructure projects from the Dutch government.

This page is left blank intentionally

Chapter 5 ANN Forecasting Model Implementation

This chapter explains the processes of ANN model implementation starting from the dataset pre-processing, the ANN model simulation, and the optimization of the ANN model's result. All these processes are done in Python 3.0 along with its libraries (i.e., Scikit-learn and Keras).

5.1 Dataset Pre-processing

The total data entry within the compiled database prior to the features visualisation was ninety-five entries. After visualising the variables, four data entries were removed based on their nature as outliers; which is the result of pair-wise visualization of the categorical data where the values are located in the box-plot above the 3rd quartile plus 1.5 times inter-quartile range (IQR) and below the 1st quartile minus 1.5 times IQR. This removal makes the total data entries for the dataset to be feed into the model is ninety-one road infrastructure projects. The last two procedures to be done before the dataset is ready to be fed into the model are *Standardizing the Features*, both the continuous and categorical features, and structuring the dataset into an *Analytics Base Table (ABT)* after the features are standardized.

Standardizing Continuous Features

Dataset standardization is a common requirement for various machine learning algorithms. The reason is the algorithms might behave poorly if each of the features does not fall within or have features of a standard normally distributed data with 0 mean and standard deviation of 1 (*Pedregosa et al., 2011).* The following formula is used to find the standard score (*z)* of a continuous feature (*x)*:

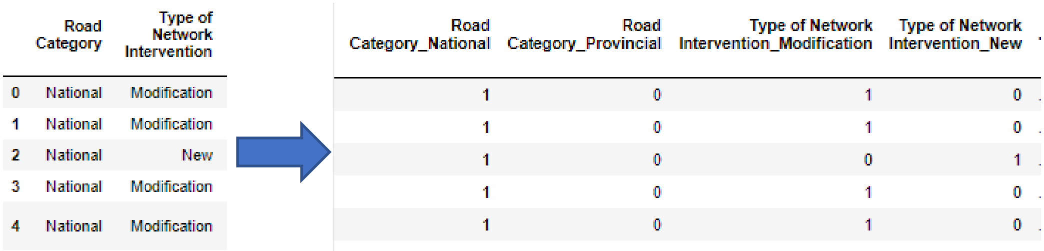
$$z = \frac{(x - u)}{s} \qquad 5.1$$

Where:
u = Mean of the training samples within the corresponding feature set.
s = Standard deviation of the training samples within the corresponding feature set.

Standardizing Categorical Features

In order to handle the categorical features prior to the feeding into the machine learning algorithm, the most common approach is to transform a single categorical feature into several continuous features according to the number of category present within the corresponding feature set. This is done to encode the levels of the categorical feature. In order to do this, each category value is converted into a new column and assign 0 or 1 (False/True) to that column. This method is called One Hot Encoding (Pathak, 2018). For this research, *.get_dummies()* method from the Pandas library is used. The function is called this way because it produces dummy or indicator variables with a value of either 1 or 0.

To give an overview of what had been done to the dataset, the following example is presented. Two categorical feature sets of the database, which are *Road Category,* and *Type of Network*, are transformed into four different columns which each column contains a continuous value of either 1 or 0. This transformation process is illustrated in Figure 23. The number of new columns generated is dependent on the number of category within each categorical feature. *Road Category* feature contains two categories, which are National and Provincial Road. Thus, after *.get_dummies()* function is implemented, two new columns are generated which represent the two categories present. To give an overview of the continuous value within the columns; the column of *Road_Category_National* contains the value of 1 and *Road_Category_Provincial* has the value of 0 when the corresponding data entry has *National* in the Road Category column in the pre-standardized categorical feature.



	Road Category	Type of Network Intervention
0	National	Modification
1	National	Modification
2	National	New
3	National	Modification
4	National	Modification

Road Category_National	Road Category_Provincial	Type of Network Intervention_Modification	Type of Network Intervention_New
1	0	1	0
1	0	1	0
1	0	0	1
1	0	1	0
1	0	1	0

Figure 23 Pre-Standardization and Post-Standardization Categorical Features Example

Analytics Base Table (ABT)

An Analytics Base Table is a flat and simple tabular structure of data which made of columns and rows. ABT is the format which the dataset structured for the machine learning model to process (J. D. Kelleher, B. Mac Namee, and A. D'Arcy; 2015). A typical structure of ABT is illustrated by the following Table 7.

Table 7 ABT Structure adopted from J. D. Kelleher, B. Mac Namee, and A. D'Arcy (2015)

Input Features				Target Feature

After conducting the aforementioned processes, the raw dataset, as shown in Table 8, is standardized, transformed, and structured into an ABT, which dependent on the type of ANN to be used. The final ABT for the ANN models is shown in the Model Implementation section.

Table 8 Raw Dataset (Partial)

Road Category	Type of Network Intervention	Number of Province	Geographical Profile	EHS/GHS	GDP	3 years Average GDP Growth	RDP	Population	Ideoogy	Car/Road Ratio	Duration (days)
National	Modification	1	Urban	No	663008	0.010316531	0.1827	2741369	Center-Right	13175.35509	196
National	Modification	1	Rural	Yes	506671	0.041956742	0.2151	3439982	Election	12764.14374	276
National	Modification	1	Rural	No	652748	0.011103421	0.2103	3563935	Center-Right	14501.52862	308
National	New	1	Combination	Yes	617540	0.022397292	0.0594	1122604	Center	6948.008541	365
National	Modification	1	Rural	Yes	652748	0.011103421	0.15	2471011	Center-Right	7244.206161	365
National	Modification	1	Urban	No	545609	0.033350577	0.1857	2599103	Center	11495.00193	365
Provincial	New	1	Rural	No	545609	0.033350577	0.0164	365859	Center	4756.167	416
National	Modification	1	Urban	No	617540	0.022397292	0.2117	3481558	Center	13513.1658	430

Training and Validating Dataset

The dataset is split into three separate sets, training set, validation set, and test set. The training set is utilized to fit the parameters of the input features on the model. After the training set is fitted to the model, the validation set is used to measure the generalizability of the post-trained model. Finally, the test set can be used on the model to provide independent performance measure of the model both during and post-training. For this model, 10% of the dataset is held out as the test set.

With regard to validation, the objective of a machine learning model is to produce a sufficient l generalization to any data from the problem domain according to the training data. The result of the generalization allows the model to make predictions on new data that the machine learning model never seen before. There are two common problems regarding this model generalization, which are underfitting and overfitting. A model is deemed to be underfitting when it is too simple to represent the relationship between the descriptive feature (input feature) and the target feature (output feature) within the dataset (J. D. Kelleher, B. Mac Namee, and A. D'Arcy; 2015). On the other hand, overfitting happens when the model is too complex and fits the dataset too well; which make the model too sensitive to noise within the data.

A validation method called K-Fold Cross Validation is used to overcome this problem. This method splits the training dataset into k-consecutive folds. Each fold is used once as a validation set, while the remaining k-1 folds are treated as the training set (*Pedregosa et al., 2011).* For this research, 5-fold cross-validation is applied. The *K* selection is not an exact science because it is difficult to estimate how good is the chosen fold to represent the whole dataset. The 5-fold cross-validation is commonly used because the 20% split generally produce a pretty accurate result.

To give an overview of the data split and K-Fold Cross Validation method, the following Figure 24 is provided. For the training data, 90% from the dataset is used for training data with a 5-fold cross-validation method while the other 10% is used to the test data.

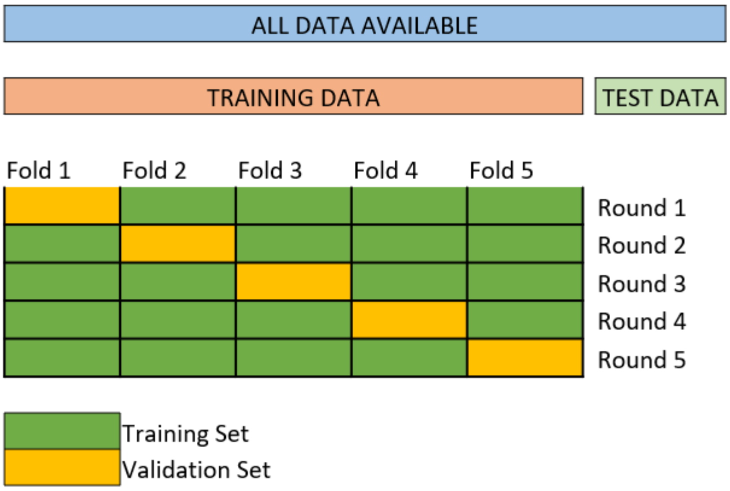


Figure 24 5-Fold Cross Validation Illustration

5.2 Model Implementation

Artificial Neural Network (ANN) or Machine Learning, in general, can be utilized to analyze various kind of problems. This research treats the forecasting task as a regression problem and classification problem. Thus, in this section, two model would be proposed to address these two problems. The first model is a regression model which proposed to predict the precise duration of the decision-making process. The second model is a classification model which try to provide a prediction in a higher level; which is the timeframe of the project introduction whether the project would take longer or shorter in comparison to its average duration.

5.2.1 ANN Regression Model

The target feature in this model is the value of decision-making duration on new road infrastructure projects. The standardized data set in the form of ABT for the model to process is shown in Table 9. Python 3 is used with additional libraries from Scikit-learn and Keras to pre-processing, implementing, and evaluating the model.

Table 9 Regression Model Standardized ABT Dataset (Partial)

	Input Features																			Target Feature
	Number of Province	GDP	3 years Average GDP Growth	RDP	Population	Car/Road Ratio	Road Category_National	Road Category_Provincial	Type of Network Intervention_Modification	Type of Network Intervention_New	Geographical Profile_Combination	Geographical Profile_Rural	Geographical Profile_Urban	EHS/GHS_No	EHS/GHS_Yes	Ideoogy_Center	Ideoogy_Center-Right	Ideoogy_CenterLeft	Ideoogy_Election	Duration (days)
0	-0.37798	1.478076	-1.61188	0.76594	0.697088	0.986575	1	0	1	0	0	0	1	1	0	0	1	0	0	196
1	-0.37798	0.24027	0.327056	1.128225	1.242077	28417.69	1	0	1	0	0	1	0	0	1	0	0	0	1	276
2	-0.37798	1.396842	-1.56371	1.074553	1.338772	32286.11	1	0	1	0	0	1	0	1	0	0	1	0	0	308
3	-0.37798	1.118081	-0.8716	-0.61275	-0.56571	15467.64	1	0	0	1	1	0	0	0	1	1	0	0	0	365
4	-0.37798	1.396842	-1.56371	0.400301	0.486182	16127.15	1	0	1	0	0	1	0	0	1	0	1	0	0	365

ANN Structure

The best performing ANN structure for the regression problem at hand is an ANN with a structure of 22-11-11-1, as shown in Figure 26. The "(+12)" in the input layer represent the number of additional nodes that are present in the model but not drawn in the figure. It also applies for "(+1) on both hidden layers. The activation function for nodes within the input layer and the hidden layer is "Rectified Linear Units (Relu)."

Relu is introduced by Richard HR Hahnloser et al. (2000) and regarded as the most used function for ANN due to its simplicity. This function is defined as follows:

$$f(x) = \max(0, a) = \max(0, \sum_{i=1}^{i=n} w_i\, x_i + b) \qquad 5.2$$

Where:
w = weight of the neuron connection *i.*
x = value of input *i*.
b = bias.

With this function, the weighted sum of inputs above 0 would be returned as the number itself; while negative-sum would be returned as zero. This process is illustrated in Figure 25.

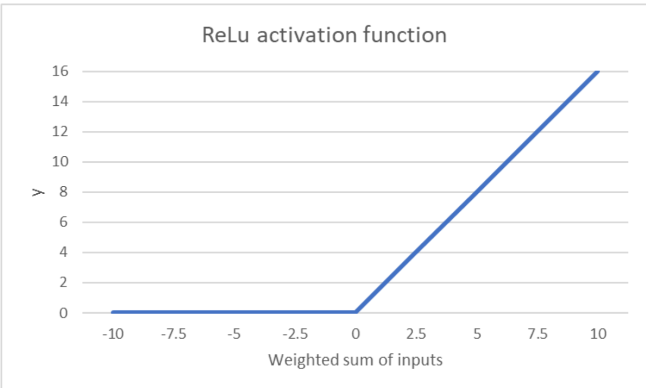


Figure 25 Relu Activation Function Illustration

As mentioned in 3.2.3, the ANN model "learns" from the training data by modifying its weights to reduce the loss function gradually. For the loss function of this ANN regression model, "mean_squared_error" is used. Mean Squared Error (MSE) measures the average squared difference of an observation's predicted values with the actual value. The output from the function represents the score associated with the current weights set. The MSE function is defined as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^{i=n} (y_i - (mx_i + b))^2 \quad 5.3$$

Where:
N = the number of total observation
$\frac{1}{N} \sum_{i=1}^{i=n} (..)$ = mean value
y_i = the actual value of the observation
$mx_i + b$ = the predicted value

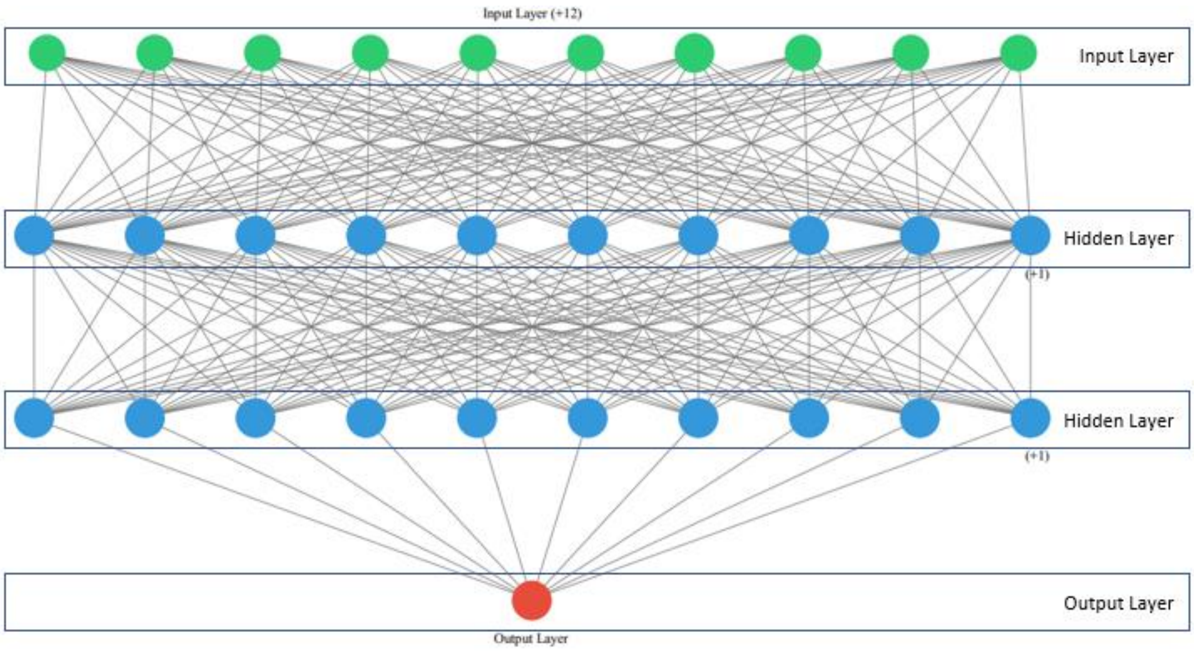


Figure 26 Road Infrastructure ANN Regressor Structure

Training and Validation

The input features are fed into the model with a batch size of 15 and Epoch of 500. Batch size and epoch are two model hyperparameters which define how the ANN model performs its training process. The batch size is a hyperparameter which defines the number of samples to be processed before the model updates its internal model parameters during the training process. On the other hand, an epoch is a hyperparameter which defines the frequency of the model to working through the training dataset during the training phase to minimize its loss value.

Figure 27 below shows that with an increase in the epoch the loss of the model decreases and after the epoch of 300, the model loss started to remain constant which indicates the model was not learning much with additional epochs. However, the model loss on training data cannot represent the actual performance of the model because the model might overfit and be too sensitive to new data. Thus, the five-fold cross-validation was used to validate the proposed ANN regression model.

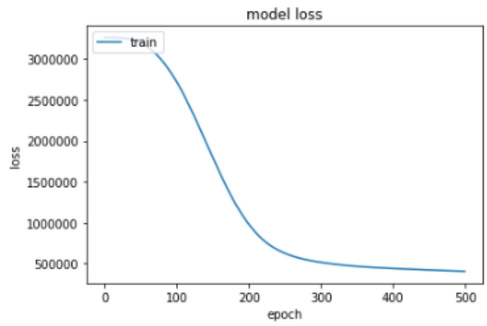


Figure 27 Model Loss on Training Data

The five-fold cross-validation (Figure 28) on the training data shows the mean squared value for each fold. The average of those five values are as follows:

a. Mean squared error (MSE) = 876160.71
b. Root mean squared error (RMSE) = 936.03 days / 2.565 years

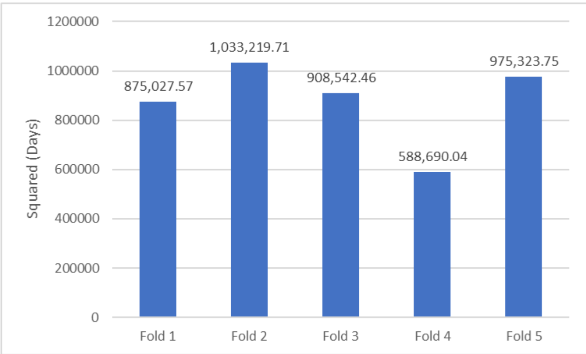


Figure 28 Regression Model Five-Fold Cross-Validation Result

Testing on Unseen Data

After the model is trained and validated, the test data were fed into the model to see its performance on unseen data in the future. The unseen data, in this case, is 10% of the original data which not included in the model during the training and validation process. The result from the testing is the model predicted the test data with:

a. Mean squared error score = 1054666.482.
b. Root mean squared error = 1026.97 days / 2.81 years

Based on Table 10, the best prediction from the model is on data **ID 3** with the error of only 0,5 years. On the other hand, the worst prediction is data **ID 6** with an error of 5,03 years. Figure 29 illustrates the prediction error / differences.

Table 10 Regression Model Prediction on Test Data

ID	Prediction	Actual	Difference	
	[days]	**[days]**	**[days]**	**[years]**
1	1343.3629	1825	481.6371	1.319554
2	1863.5	3532	1668.5	4.571233
3	1695.7	1479	-216.7	-0.5937
4	24.3	1704	1679.7	4.601918
5	2034.7	2434	399.3	1.093973

ID	Prediction	Actual	Difference	
6	1631.085	3468	1836.915	5.032644
7	1360	1840	480	1.315068
8	1956	1677	-279	-0.76438
9	2501.6	1600	-901.6	-2.47014
10	1487.33	1400	-87.33	-0.23926

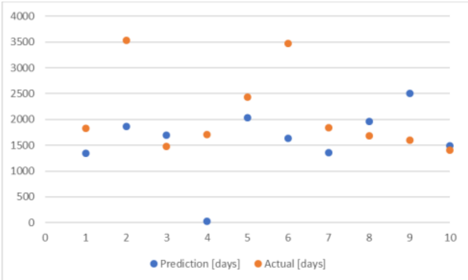


Figure 29 The Difference between Prediction and Actual Value

From the value above, it is clear that the model performs slightly worse on the future dataset in comparison to the training dataset which indicated by the higher RMSE value on the test data. The general acceptance is a lower RMSE value indicates a better model performance; However, there is no theoretical sound method to assess whether an RMSE value is acceptable or not. Thus, an interview with BAM expert was done to assess their acceptance of the model's prediction performance.

The acceptable minimum RMSE range of a prediction model from the company's perspective is 0.75 year – 1 year (BAM_4, 2019). The underlying motivation is the prediction with an RMSE value above that limit would not provide additional value for the company's portfolio management in term of reducing the uncertainty of a newly announced project. Because if the project decision is delayed too far behind the predicted timeframe and the company already hired additional human resource or re-allocating its resources to formulate a strategy to engage with the project, then it would be an inefficient business procedure.

5.2.2 ANN Classification Model

In this model, the target feature, which is the decision-making duration, is clustered into two categories: ***Below Average*** and ***Above Average***. The mean value of the target feature is 4,57 years, as illustrated in Figure 30. It means the data entry with the target feature value below 4,57 years clustered into the first category, Below Average; while the data entry with feature value above 4,57 years clustered into Above Average category. This approach transforms the target feature data type from numeric into categorical data; which makes the data set for the pre-processing different from the regression model. The ABT for this model is shown in Table 11 below.

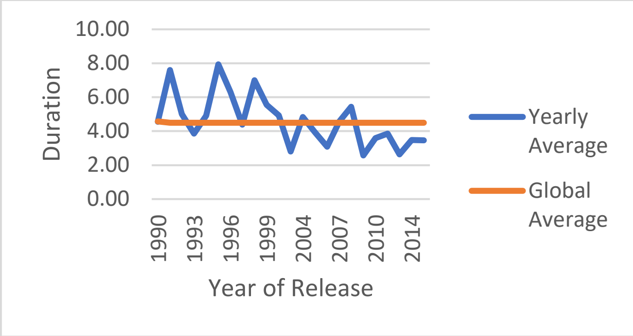


Figure 30 Decision-Making Duration Yearly and Global Average

Table 11 Classification Model Standardized ABT Dataset (Partial)

	Input Features																	Target Features
	Number of Province	GDP	3 years Average GDP Growth	RDP	Population	Car/Road Ratio	Road Category_National	Road Category_Provincial	Type of Network Intervention_Modification	Type of Network Intervention_New	...	EHS/GHS_No	EHS/GHS_Yes	Ideoogy_Center	Ideology_Center-Right	Ideoogy_CenterLeft	Ideoogy_Election	Above Average?
0	-0.37798	1.478076	-1.61188	0.76594	0.697088	0.986575	1	0	1	0	...	1	0	0	1	0	0	No
1	-0.37798	0.24027	0.327056	1.128225	1.242077	28417.69	1	0	1	0	...	0	1	0	0	0	1	No
2	-0.37798	1.396842	-1.56371	1.074553	1.338772	32286.11	1	0	1	0	...	1	0	0	1	0	0	No
3	-0.37798	1.118081	-0.8716	-0.61275	-0.56571	15467.64	1	0	0	1	...	0	1	1	0	0	0	No
4	-0.37798	1.396842	-1.56371	0.400301	0.486182	16127.15	1	0	1	0	...	0	1	0	1	0	0	No

ANN Structure

As already mentioned, Python 3.0 is used to conduct processes concerned with the classification model building. Additional libraries from Scikit-learn and Keras are used to pre-processing the dataset, build and deploy the ANN, and evaluate the result. The best performing ANN structure for classification problem at hand is three layers with one hidden layer consists of 9 nodes. The structure of the model is shown in the following Figure 31 Road Infrastructure Projects ANN Classifier Structure Figure 31. The "(+12)" in the input layer represent the number of additional nodes that are present in the model but not drawn in the figure.

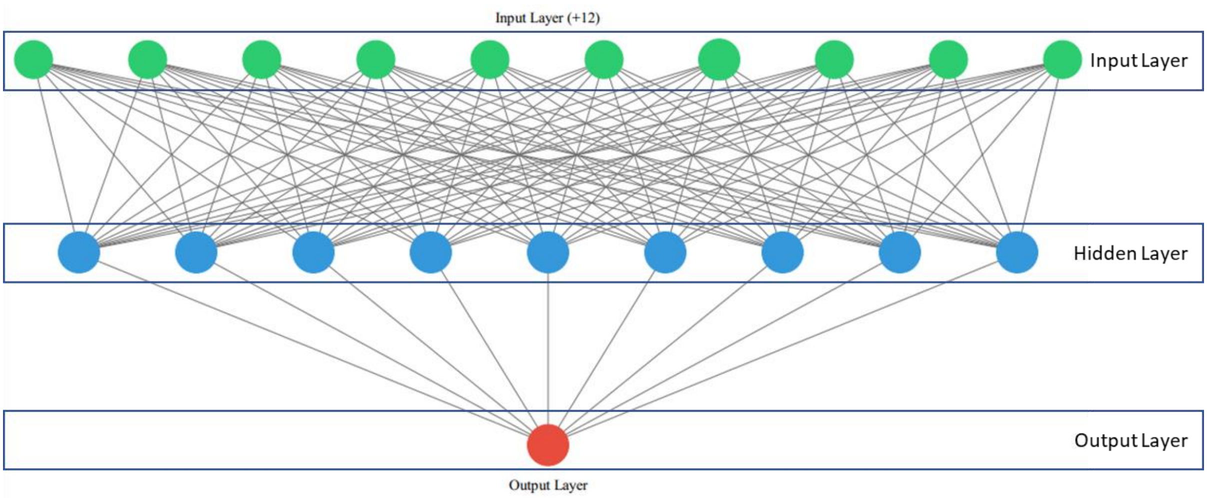


Figure 31 Road Infrastructure Projects ANN Classifier Structure

The activation function for nodes within the input layer and the hidden layer is "Relu" while for a node in the output layer is "Sigmoid." Sigmoid function squashes the output signal's permissible amplitude range from (-∞.∞) to (0,1) (Karlik and Olgac, 2010). The function is identified as follows:

$$g(x) = \frac{1}{1 + e^{-x}} \qquad 5.4$$

The illustration of this function is shown in Figure 31

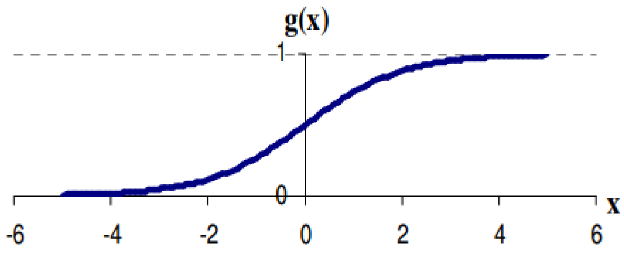


Figure 32 Sigmoid Function Illustration (Karlik and Olgac, 2010)

For the loss function in the node of the output layer, "binary cross-entropy" is used. Cross-entropy measures classification model performance, which the output of the model is a probability in the range (0,1). Cross-entropy loss would increase when the predicted probability is further away from the actual label. The function is identified as follows:

$$H_{(p,q)} = -\sum_{i=1}^{i=n} p_i log(q_i) \qquad 5.5$$

Where:
$H_{(p,q)}$ = Cross-entropy loss
p_i = the true probability values
q_i = the predicted probability values

In the case of binary cross-entropy, the cross-entropy loss would only be $-\sum_{i=1}^{i=n} log(q_i)$ (Géron, 2017).

Training and Validation

The input features are feed into the model with a batch size of 15 and Epoch of 150. These values are chosen because, with larger batch size, the speed of the training is faster; besides that, the epoch of 15 is chosen because the model did not learn much above 140 epoch. Figure 33 below shows that with an increase in epoch, the accuracy of the model increases while the loss of the model decreases. However, the accuracy of the model on training data alone cannot represent the actual performance of the model. Thus, five-fold cross-validation was used to validate the proposed classifier ANN model.

The five-fold cross-validation on the training data shows that the average accuracy of the model is 55% (Figure 34). Accuracy, in this case, means that 55% of the prediction made by the model on the validation set was correct.

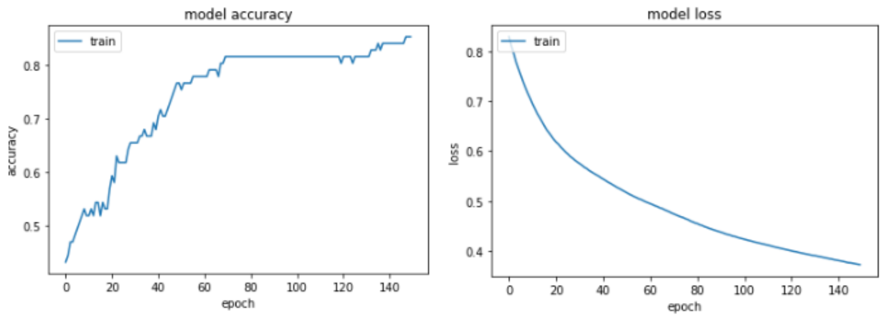


Figure 33 Model Accuracy and Loss on Training Data

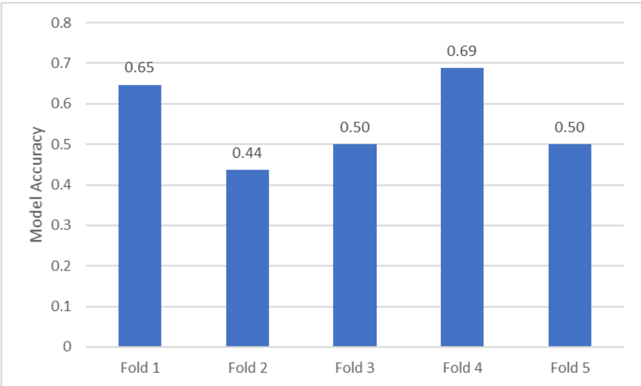


Figure 34 Classification Model Five-Fold Cross-Validation Result

Testing on unseen data

After the model is trained and validated, the test data were fed into the model to see its performance on unseen data in the future. The unseen data, in this case, is 10% of the original data which not included in the model during the training and validation process. The result from the testing is the model predicted 70% of the test data correctly.

Table 12 Testing Result on New Data Entries

Prediction	Actual	Status
Yes	Yes	**Correct**
Yes	Yes	**Correct**
No	No	**Correct**
No	Yes	**Wrong**
Yes	Yes	**Correct**
Yes	Yes	**Correct**
No	Yes	**Wrong**
Yes	Yes	**Correct**
Yes	No	**Wrong**
No	No	**Correct**

Based on the model accuracy on the training data and test data with a binary target feature, it can be concluded that the model is not sufficiently reliable to be applied. An optimization for the classification model is done to address the result above which elaborated in the next section.

5.3 Model Optimization

One of the main difficulties in building an ANN model is the selection of input features / variables for the model due to its impact on the specification of the model. With the chosen input variables, the model might be under-specified or over-specified with the inclusion of redundant uninformative variables (May et al., 2011). One of the main concerns for a multi-layer perceptron (MLP) type such as ANN is the curse of dimensionality (Bellman, 1961), which is, as the model dimensionality increases linearly, the modeling problem domain volume increases exponentially. Thus, to map a given function to the model parameter space with a sufficient confidence level, a higher number of samples is required (Scott, 1992). ANN architecture with an MLP is susceptible to this curse because the high number of additional connection weights for each additional input variable inclusion in the model.

Thus, for the proposed models, an optimization effort is done by reducing the number of input variables included in the models. The minimum threshold of the amount of variables inclusion in each of the three system elements within the *Physical Environment Domain* needs to be represented by the variables in the model. It means the combination of three variables as the input variables is the minimum threshold for the optimization effort because there are three system elements present; *Natural Environment, Induced Environment System Element*, and *Human System Element*.

One important remark for the model optimization is that the structure of ANN is also modified to find the best performance ANN with the given input features based on the training set.

5.3.1 ANN Regression Model Optimization

In regard to the regression model, several variables combinations have been tried in order to reduce the dimensionality of the regression model. The result of this approach is the model performed slightly worse, which indicated by the increase of RMSE on the training set with lesser variables embedded to the model. The summary of the results is illustrated in Table 13; which shows that the average RMSE of the model based on 5-fold cross-validation falls within the range of *2.78-3.19* years. The lowest RMSE, with a value of 2.78 years occurred when *Variables Combination 1 and 2* were embedded. On the other end, the highest RMSE with a value of 3.19 years occurred when *Variables Combination 8* were embedded.

Table 13 Variables Combination and its corresponding RMSE

Variables Combination	Number of Variables	Natural Environment	Induced Environment	Human-System	Structure	**5 folds average** RMSE (years)
1	11	GA P	CR EHS	*GDP* *GGDP* *RDP* *RC* *NI* *DPC* *DPI*	21--11--11-1	2.78
2	10	GA P	CR EHS	*GDP* *GGDP* *RDP* *RC* *NI* *DPC*	22--11--11--1	3.05

Variables Combination	Number of Variables	Natural Environment	Induced Environment	Human-System	Structure	**5 folds average** RMSE (years)
3	9	GA P	CR EHS	*GDP* *RDP* *RC* *NI* *DPC*	16--10--10--1	3.12
4	8	GA P	CR EHS	*GDP* *RC* *NI* *DPC*	15--8--8--1	2.92
5	7	GA P	CR EHS	*GDP* *RC* *DPC*	15--05--01	2.78
6	6	GA P	CR	*GDP* *RC* *DPC*	11--05--01	2.9
7	5	GA P	CR	*GDP* *DPC*	9--4--4--1	2.867
8	4	GA P	CR	*GDP*	6--4--1	3.19
9	3	GA	CR	*GDP*	3--2--1	2.97
10	3	P	CR	*GDP*	5--3--1	3.04

5.3.2 ANN Classification Model Optimization

The same optimization approach is taken for the classification model. The result of the optimization is presented in Table 14; which shows that the performance of the model remains at relatively the same level of accuracy, which falls within the range of 54 – 62%, even with the reduction in the number of input variables involved. The highest accuracy on training set based on 5-fold cross-validation occurred when *Variables Combination 10* was embedded to the model with sixty-two percent of model accuracy; while the lowest accuracy occurred when *Variables Combination 3* was used with a model accuracy of fifty-four percent.

Table 14 Variables Combination and its corresponding Accuracy

Variables Combination	Number of Variables	Natural Environment	Induced Environment	Human-System	Structure	**5 folds average** *Accuracy* (years)
1	11	GA P	CR EHS	*GDP* *GGDP* *RDP* *RC* *NI* *DPC* *DPI*	21--10--10-1	59%

Variables Combination	Number of Variables	Natural Environment	Induced Environment	Human-System	Structure	**5 folds average** *Accuracy* (years)
2	10	GA P	CR EHS	*GDP* *GGDP* *RDP* *RC* *NI* *DPC*	22--8--8--1	57%
3	9	GA P	CR EHS	*GDP* *RDP* *RC* *NI* *DPC*	16--10--10--1	54%
4	8	GA P	CR EHS	*GDP* *RC* *NI* *DPC*	15--7--1	57%
5	7	GA P	CR EHS	*GDP* *RC* *DPC*	15--05--01	57%
6	6	GA P	CR	*GDP* *RC* *DPC*	11--05--01	57%
7	5	GA P	CR	*GDP* *DPC*	9--4--1	61%
8	4	GA P	CR	*GDP*	6--2--1	57%
9	3	GA	CR	*GDP*	3--2--1	57%
10	3	P	CR	*GDP*	5--3--1	62%

5.4 Conclusion

The result of forecasting models with the original set of variables shows that the prediction made by the model is not sufficiently reliable based on training and test data; with RMSE of 2.565 years for the regression model and an accuracy of 55% for the classification model. To address these result, an optimization effort has been done through a reduction in the number of input variables involved in the forecasting model.

The result of optimization shows that no significant improvement occurred for both the regression model and the classification model. Although the RMSEs are higher for the regression model, the difference with the original model, which embedded the initial input features set, is relatively low. A similar outcome also happened in the optimization of the classification model, which indicated by the accuracy values that are not far apart from the original model accuracy.

The result of the proposed AI forecasting model implementation and improvement could be explained with a further discussion about the potential individual and collective effect of the following factors: [1] the number of data entries is too low to make an adequate generalization; [2] the identified variables simply do not have enough influence on the decision-making duration of new road infrastructure project in regard to the project scope; or [3] the model is unable to represent the "world" which influences the decision-making duration for new road infrastructure projects. These three potential causes are discussed further in the next chapter.

This page is left blank intentionally

Chapter 6 Result Discussion

This chapter discusses the potential causes of the proposed AI forecasting model result. For this discussion section, a reflection on the potential causes mentioned in 5.4 is done with the support of a literature study on the relevant topics. In this chapter, a comparison of the ANN model's result with conventional statistical methods is also done. The objective of this comparison is to explore whether the benefit that the artificial intelligence possesses in comparison to the conventional statistical method is proven in this research.

6.1 Influential Factor Discussion

There are three potential causes of why the result of the forecasting model is not reliable enough as mentioned in 5.4; which are [1] the number of data entries, [2] the chosen independent variables' influence on the variable being forecasted; and [3] the model capability to represent the "world". Each of these causes is discussed and elaborated further in the following sections.

6.1.1 Number of Data Entries

The total number of projects in the final dataset for the AI forecasting model is ninety-one projects which ranging from 1990 until 2015. In this section, the number of data entries used for this research is compared with other researches which implement the same type of AI, Artificial Neural Network, in the civil engineering field. Although the problem addressed by those researches are different from this research problem, this comparison is still relevant because the goal of an ANN model is to produce a good generalization to any data from the domain of a certain problem based on the available training data. The comparison is presented in Table 15.

Based on the comparison with the existing researches which implemented ANN in civil engineering domain, it can be considered that the research has used a decent number of data entries for the proposed ANN forecasting model; in order to produce a well generalization of the decision-making duration on new road infrastructure projects. Although the number of data entries in other researches presented in this discussion are relatively smaller or similar; the distinctive factor might be the influence of the chosen independent variables with the dependent variables. This factor is explored and elaborated in the next section.

Table 15 Data Entries Number Comparison with existing research

Research	Problem	Data entries number
Sonmez and Ontepeli (2009)	cost estimation	13
Wang et al. (2010)	cost estimation	16
Alqahtani and Whyte (2013)	cost estimation	20
Cheng et al. (2009)	cost estimation	28
Lyne and Maximinio (2014)	cost estimation	30
Wang et al. (2017)	cost estimation	46
Elsawy et al. (2011)	cost estimation	52
Yahia et al. (2011)	duration prediction	52
Zhou (2018)	cost estimation	71
Arafa and alqedra (2011)	cost estimation	71
Yu and Skibniewski (2009)	cost estimation	110
Petroutsatou et al. (2011)	cost estimation	149
Maghrebi et al. (2014)	duration prediction	200
This Research	**duration prediction**	**91**

6.1.2 Input Variables Influence

The independent variables which were chosen for the input features are based on the variables that drive the development of infrastructure and variables which were mentioned within *startnotitie* documents. The initial approach was by trying to incorporate all selected variables into the model to better represent the "world." After the result from this approach is presented, an optimization effort had been done by reducing the number of variables which incorporated into the model.

One key insight from the model optimization result is the selected input variables do not have a strong influence, both individually or in combination with other variables, on the decision-making duration of a new road infrastructure project. The influence of the numeric input variables is indicated by the optimization result, which not far apart from the initial approach result for both regression and classification models. The absent of strong influence can be represented by a low co-relation value of an individual input variable with the output variable, which is the decision-making duration on a new road infrastructure project. The correlation value for each independent variables with numeric data type is presented in Table 6.

Besides the GDP and 3 years average of the GDP growth variable, the other numeric input variables have a small influence on the target value despite being well-founded in theory. To discuss the underlying reason for these values, a closer look at each of the variable is given as follows:

1. *GDP*
   This correlation value shows that the higher the GDP, the lower the duration needed to decide on a new infrastructure project by the Dutch government. This value is in conform to the previous research, which states that the level of GDP represents the infrastructure demand and the ability by the government to fund the project (Rothman et al., 2014). Thus, the higher the GDP value of the Netherlands, both the infrastructure demand and the government's ability to fund the project are also improved and lead to a faster decision-making procedure.
2. *Three years average of GDP growth*
   This variable is chosen as an indicator of the direction where the country is going in term of GDP and complimenting the GDP variable due to the consideration of only its annual value. The correlation value shows a positive relationship, which means the decision-making duration on new infrastructure increases with the increase in the GDP growth of the last three years. This value indicates that better growth in the Netherlands' economy does not necessarily mean the government is willing to allocate more fund for infrastructure investment. This claim is supported by Briceno-Garmendia, Estache, and Shafik (2004) and Estache (2010). In their respective research, they suggest that public spending on infrastructure tends to decline when GPD rises. This situation eventually might lead to a longer duration to decide on new infrastructure project by the government. The difference in "direction" of the correlation value in comparison to the *GDP* variable might be caused by the span of the growth which considered in this research. The number of year for the analysis period, which is three years, is chosen arbitrary, and this decision might cost the model's prediction performance. The fact that an infrastructure project is a long-term investment justifies that a higher number of analysis period for the GDP growth should have been chosen for this research.
3. *Province RDP*
   The low correlation value indicates that the economic performance of a province where a certain infrastructure project is initiated might not have an influence on the decision-making duration by the government. This might be influenced by the fact that only in recent years, specifically in 2012, that the Dutch government try to stimulate the province to take the

initiative in finding the fund for the infrastructure development in their respective area (IPA, 2012). Thus, in previous years, the main government played an important role in deciding the project scope and funding its cost, which explains the low correlation value.

4. *Car/Road ratio*
   The correlation value of this variable is also low. The underlying reason might be the level of data collection. This variable is considered from the province level, whereas each project mentioned in *startnotitie* is route-specific. The decision to choose a province level for the data collection is based on the data available on CBS, which lists the number of cars and the road area in each province. Thus, the different level between data collection and the project itself might explain the low correlation value and the low model prediction performance.
5. *Population*
   There is almost no correlation between the population of a province where a certain project is initiated and the duration of decision-making by the government for the respective project. This value can be influenced by the data collection method, which only focuses on the province level. Meanwhile, in reality, the user of the infrastructure project in plan is not only from that particular province but also from other provinces which were neglected in the model. Due to the lack of available data in the *starnotitie* during the early stage of the planning procedure, it is difficult to gather the actual number of people who would be impacted both positively and negatively with the initiation of the new project for the model.
6. *Number of provinces*
   The underlying reason for choosing this variable is the involvement of lower government bodies during the decision-making procedure. With an increase in the number of provinces involved in the decision-making process, the chance that the government takes longer to decides on a certain design alternative is also increased. The low correlation value might be influenced by the high-level data collection method that is taken in this research. Besides the province where a certain project is initiated, the municipalities in the vicinity of the project area are also involved. In this research, the number of municipalities is not considered because it is not always mentioned in every *startnotitie,* and the number might increase or decrease during the decision-making process. Thus, only the number of the province involved is considered in this research to maintain consistency during the data collection process.

From this discussion about the correlation values of the numeric variables, it is reasonable to consider that the non-existent of a substantial correlation between the numeric input variable and the target value is a potential cause to the poor model's prediction performance.

6.1.3 Representation of the "World"

The third potential cause is the capability of the model to represent the "world"; which is the SOI architecture and its operating environment that interacts with the SOI. The focus of discussion in this section is the operating environment, especially the *Higher-Order System*. It is because the *Physical Environment Domain,* which represented by the independent variables, has been discussed in the previous section; and SOI architecture is designed to be fixed, which is an ANN model structure.

As illustrated in Figure 9, the operating environment which interacts with SOI comprises of *Higher-Order System* and *Physical Environment Domain.* Operating environment is represented to the SOI with the so-called factored representation; where the "world" is split up into a fixed set of attributes or variables for each episode to be experienced by the SOI. *Physical Environment Domain* is represented to the SOI through the selected independent variables; while the *Higher-Order System* is represented to the SOI through system elements from two perspectives: Legal and Organization perspective.

In Research Method, it is mentioned that a certain limitation is chosen to determine the scope of the research. This limitation had become the foundation for system elements from both the Legal and Organization perspectives within the HOSD. From the legal perspective, the Infrastructure Planning Act (*Tracewet)* is chosen as the sole legal framework, which acts as the foundation for the forecasting model and considered as the *Operating Constraint Element*. As from the organization perspective, BAM Infra BV is chosen as the organization element where the system would be implemented. The discussion in this section focuses on the operating constraint element from legal perspective instead of from organization perspective because, during the Data Availability Study, it is found that there are other policies which influence the IPA which were not incorporated to the model due to the scope limitation of the research.

The release year of ninety-five projects which compiled in the dataset are ranging from 1990 until 2015, and the IPA itself was established in 1994. Prior to 1994, the startnotitie which released by the government was based on the mobility plan of the Netherlands called *Tweede Structuurschema Verkeer en Vervoer (*SVV-2)*.* Even after the IPA was established, each *startnotitie* released was still based its motivation on SVV-2. Beside the SVV-2, there were other policies which also influenced the projects further down the years. The IPA itself was amended several times to suit the infrastructural need in that particular period. To give a representation of these dynamic relationships surrounding the *Operating Constraint Element* from a legal perspective, which is IPA, the following Figure 35 is presented. Further elaborations for each influential policy are given in the following sub-sections to give a better overview of the impact of these dynamic relationships on the model performance.

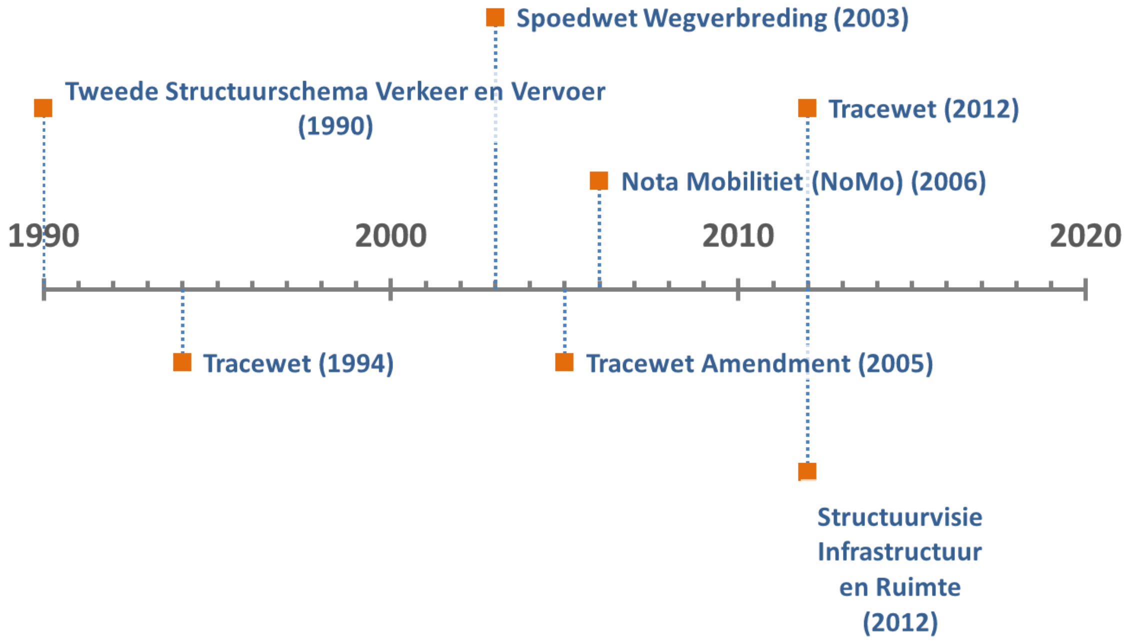


Figure 35 Timeline of IPA and other policies related to road infrastructure

Infrastructure Planning Act (IPA) potential impact on the model performance

With regard to the Infrastructure Planning Act, there were two modifications since the IPA was released. The first modification or amendment was in 2005, followed by the release of IPA 2012. In the first amendment, the major change was the introduction of a new procedure which called shortened procedure (Gierveld, 2016). With the 2005 amendment, there are two separate procedures for the planning of new infrastructure based on its scale. The procedure which still in line with the existing requirements was called extended procedure; while the new procedure where some requirements could be abandoned was called shortened procedure. The motivation behind this amendment was the fact that at the beginning of the 21^{st}, in regard to large infrastructure projects, adaptations to the main road were increasingly possible due to the relatively small measures required; such as the construction of additional permanent lanes and traffic lanes. Thus, the necessity to follow the existing procedure (where EIA / Trajectnota mandatory for every infrastructure project) was questioned. The introduction of this new procedure affected the decision-making duration on new road infrastructure projects in a positive way, as shown in Figure 36.

For the second modification, IPA 2012, the main modification is the statutory limitation for the parties who have the right to oppose the government's decision on new infrastructure. But this change can be considered as less influential to the decision-making duration on new road infrastructure because the modification is on the process after the government made a decision; not during the decision-making process on a certain infrastructure project. Thus, the reduction during its respective period, which shown in Figure 36, can be considered as the result of other aspects / interventions.

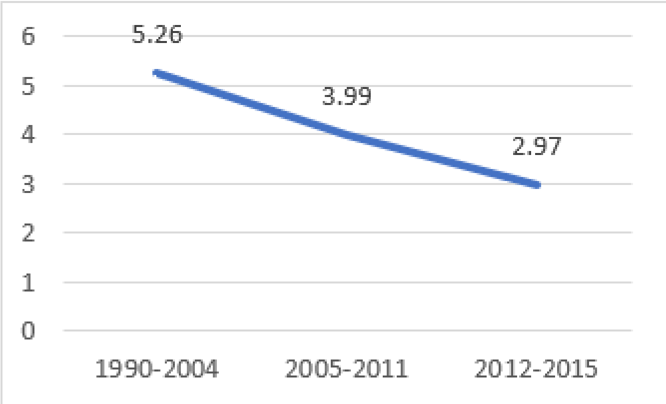


Figure 36 Decision Making Duration Average for each IPA (years)

Mobility Plan potential impact on the model performance

There are three mobility plans released by the Dutch government within the period of 1990-2015. As illustrated by Figure 35, the mobility plans are Tweede Structuurschema Verkeer en *Vervoer (1990)*, *Nota Mobilitiet (NoMo) (2006),* and *Structuurvisie Infrastructuur en Ruimte (2012).*

SVV-II is a mobility plan of the Netherlands which adopted in the period of 1988-1991 as a successor of the first Structuurschema Verkeer en Vervoer (SVV-1). The main focus of this mobility plan is to shift the focus of the government towards the development of public transportation. Thus, limiting the expansion possibility of the existing road network (Ministerie van Verkeer en Waterstaat, 1988). This limitation impact is apparent in Figure 37, where the average of the decision-making duration on projects within the period of 1990-2006 ranked the highest in comparison to the other periods where other mobility plans were implemented.

In 2006, *Nota Mobilitiet (NoMo)* was implemented. This mobility plan was meant as the replacement of SVV-II to abandon the limitation on road infrastructure expansion. In *NoMo,* the government acknowledged that traffic and transport growth had to be facilitated because this growth was seen as the consequence of the ever-changing economic, demographic, spatial, and international developments (Ministerie van Verkeer en Waterstaat, 2004). The change of attitude from the government on the expansion of the existing road network through the implementation of *NoMo* affected the decision-making duration on a new road infrastructure project, which showed in Figure 37. The duration was lower in comparison to the previous period where SVV-II was implemented.

The latest mobility plan which released by the Dutch government is *Structuurvisie Infrastructuur en Ruimte* in 2012 which replaced NoMo. This is the central government's structural vision which addresses future ambitions and policies in regard to spatial issues within the Netherlands; which includes road, railway, shipping route, main ecological structure, and other spatial issues. One of the main takeaways from this vision is the emphasize on the decentralization of governance to the local governments. The reasoning behind this is the central government found that implementation of the macro policy is not as effective as micro policy or area-oriented policy (Ministerie van Infrastructuur en Milieu, 2012). Thus, through this structural vision, the local government bodies are stimulated to develop their own area along with their corresponding infrastructure and buildings. This kind of policy approach might have an impact on the decision-making duration on new road infrastructure project because of a less bureaucracy required with lesser main government involvement. One important note is the government still emphasizes the necessity for integration despite this governance decentralization during the infrastructure development procedure via MIRT. The result of the latest mobility plan implementation is shown in Figure 37, where the average of decision-making duration is ranked the lowest.

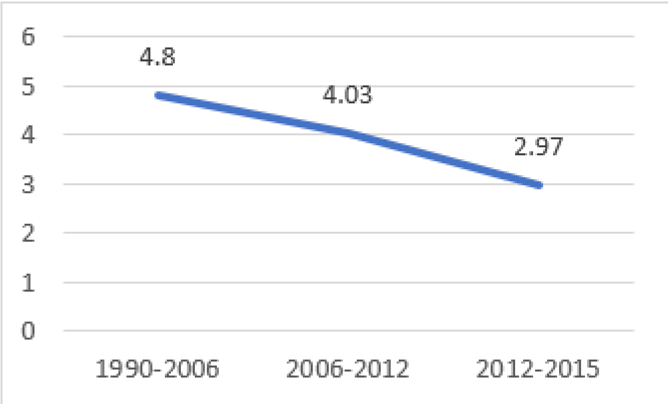


Figure 37 Decision Making Duration Average for each mobility plan (years)

Other Policy potential impact on the model

Spoedwet Wegverbreding (2003) is a policy which the government introduced to accelerate the decision-making of thirty capacity expansion projects on the main road infrastructure networks. This policy allows the government to revoke the applicability of IPA on those projects; Thus, allow a special decision-making procedure for those projects to accelerate the required duration (Henk Gierveld, 2016). From this policy, it is clear the government has the power to make this kind of intervention to tackle the occurring problem at that time; which makes it difficult to develop a forecasting model which incorporates this kind of intervention since there is no trend or pattern from the previous data related to this. Thus, it is reasonable to consider *Spoedwet Wegverbreding* has an influence at a certain degree on why the model does not perform well.

6.2 Result Comparison to other statistical methods

In this section, two comparisons are done for the proposed ANN models. Multivariate linear regression is done to compare its result with the ANN regression model's result. On the other hand, a logistic regression model is done to compare its result with the ANN classification model's result.

6.2.1 Multiple Linear Regression Model

Linear regression is one of the most commonly used statistical methods for modeling. There are two categories for this method which governed by the number of the explanatory variable; which are the simple regression with one explanatory variable and the multiple regression with several explanatory variables. The principle of this method is to model a dependent variable Y through a linear combination of *n* explanatory variables: x1, x2, …., x_n (Stulp and Sigaud, 2015).The variable Y is written for observation *i* as follows:

$$Y_i = a_1 x_{1i} + a_2 x_{2i} + .. + a_n x_{ni} + e_i \tag{6.1}$$

Where:
Y_i = The dependent variable value being observed for observation *i*
x_{ni} = The value taken by variable n for observation *i*
a_n = coefficient estimated by the regression model
e_i = model error

The model is concluded based on the least-squares method, as explained before in 5.2.1. The model implementation, in this case, is done by utilizing the XLSTAT add-on for Microsoft Excel.

The model implementation result on training dataset is the model predicted the duration with the following values:

1. Mean squared error (MSE) of 4.732 year
2. Root mean squared error (RMSE) of 2.175 year

On the other hand, the model performance on the test set is provided below:

1. Mean squared error (MSE) of 7.375 year
2. Root mean squared error (RMSE) of 2.716 year

Based on the aforementioned values, it is proven that the ANN regression model does not perform superior in comparison to the conventional statistical method, specifically multiple linear regression.

6.2.2 Logistic Regression Model

Logistic regression is a statistical method to analyze a dataset which comprises of one or more explanatory variables which determine a dependent variable. In this case, the dependent variable is measured with a binary variable where there are only two outcomes. The aim of this model is to find the most suitable structure to describe the relationship between the dependent variable and the explanatory variables (Hosmer et al., 2013).

Logistic regression model generates the coefficients of a formula to predict *a logit transformation* of the characteristic of interest's probability. The structure of the formula is looked like the following:

$$logit(p) = b_0 + \ b_1 x_1 + b_2 x_2 + b_3 x_3 + .. + b_k x_k \quad 6.2$$

Where:
p = the characteristic of interest's probability of presence
x_k = The value taken by variable k
b_k = coefficient estimated by the regression model
b_0 = estimated constant

The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}$$

and

$$logit(p) = \ln\left(\frac{p}{1-p}\right)$$

In comparison to the previous regression model, which minimize the sum of squared error, a logistic regression chooses its parameters that could maximize the probability of observing the characteristic of interest.

The implementation for the logit regression model is done with XLSTAT add-on in Microsoft Excel software. The model implementation results are the model predicted the classification with an accuracy of 66.67% on the training dataset and 50% on the test dataset. Based on these two values, it is proven that the proposed ANN model for classification model also does not yield superior performance in comparison to logistic regression model which considered as a more straightforward and simple statistical method.

6.3 Conclusion

With regard to the influential factor behind the model's performance discussion, the combination of the three factors can explain the performance of the proposed ANN models. For the first factor, the *number of data entries* used in this research ranked in a respectable position compares to the other researches which used artificial intelligence prediction model within the construction industry. This *number of data entries* cannot be observed individually to formulate a conclusion but instead should be done collectively with the other factors. In that regard, the *input variables influence* discussion shows that the chosen variables for the ANN models in this research are insufficient as the predictor for the decision-making duration by the government on the preferred design alternative of new road infrastructure; both individually and collectively. The focus of discussion was the numeric input variables, and the correlation value of these variables are proven to be low in general with only GDP and GDP growth have a noticeable correlation value with the decision-making duration. The third factor, representation of the "world", proven to have the strongest influence on the decision-making duration based on the comparison of its average values. The discussion on this factor shows that there is a reduction of decision-making duration in the period where the government intervened the industry environment with a new policy outside the main legal procedure which the model fixated on. These interventions make the environment of the system to be dynamic and difficult to grasp by the ANN models and other conventional models.

A comparison of the ANN models' result with other conventional models was also done in this chapter. The comparison of ANN Regression model with Multiple Linear Regression (MLR) model shows that the former does not perform better with the RMSE value produced by MLR has a lower value for both training and test dataset; with 2.175 years and 2.716 years of RMSE in comparison to ANN's 2.565 years and 2.81 years. For the ANN Classification model, a comparison is done with Logistic Regression Model (LRM), and it shows that the former does not perform superior either with accuracy produced by LRM higher on the training set with an accuracy of 66.67% compares to ANN's 55% accuracy.

This page is left blank intentionally

Chapter 7 Research Conclusion

The main objective of this research is to explore the applicability of AI technology in forecasting the decision-making duration by the Dutch government on the publicly known infrastructure project in term of preferred design alternative by answering the following main research question:

To what extent can an Artificial Intelligence (AI) technology be applied to forecast new infrastructure project introduction based on the decision-making duration by the Dutch Government?

Several sub-research questions were proposed to answer the main research question above, and the development of an AI forecasting model has provided an answer to each of them as follows:

1. ***What is the state of AI implementation in the field of forecasting?***
   This research answers this question through literature study on the previous studies which implements AI to forecast a certain value, e.g., cost, duration, with comparison to the conventional methods. The results show that AI is capable of providing additional benefit to the forecasting field in the form of better accuracy, better capability in dealing with the non-linear relationship of the variables, and ability to process imprecise data.

2. ***Which type of AI is relevant for the forecasting model of an upcoming infrastructure project?***
   The Artificial Neural Network (ANN) is chosen as the AI type for the problem at hand. The decision is supported with a system design approach which considered the operating environment of where the AI forecasting system exists and the way the system represents it. Based on these two, the suitable agent structure is chosen which eventually lead to the type of AI chosen. In addition to that, a literature study on previous researches which implement AI to make a prediction or forecasting in the construction industry strengthen the AI type decision.

3. ***What are the relevant independent variables to forecast the decision-making duration of a new infrastructure project by the Dutch government?***
   The independent variables are identified based on the literature study on the drivers behind infrastructure development, official government documents, and the assumption that the chosen variables represent the *Physical Environment Domain* which has an influence on the Higher-Order System Element. There are twelve independent variables identified as the input features for the model; which are Road Category, Type of Network Intervention, Gross Domestic Product (GDP), GDP growth of the last 3 years, Regional Domestic Product (RDP), Population size, the geographical profile of the project area, *Ecologische Hoofd Structuur* (EHS) Intersection, Car to road area ratio (car/km2), Dominant Political Party Change, Dominant Political Ideology, and Number of Provinces.

4. ***How an AI forecasting model to predict decision-making duration of an upcoming infrastructure project by the Dutch government might look like?***
   There are two ANN models proposed in this research, Regression ANN Model and Classification ANN Model. The best performing ANN structure for the regression model is ANN with a structure of 22-11-11-1. On the other hand, the best performing ANN structure for the classification model is ANN with a structure of 22-10-1.

5. ***How the proposed model performs in forecasting the decision-making duration of an upcoming infrastructure project by the Dutch government?***

   The ANN regression model predicts the decision-making duration with RMSE of 2.565 years on the training set and 2.81 years on unseen data. On the other hand, the ANN classification model predicts the decision-making duration with an accuracy of 55% on the training set and 70% on unseen data.

6. ***Does the proposed AI model fit for purpose?***

   The proposed forecasting AI system is designated to be used prior to the tender announcement in TenderNed; as an identification tool for project's likelihood to be introduced to the market. This knowledge could help the company to plan both its internal source allocation and to form a collaboration with other companies.

   For the proposed models, the initial result indicates that the models are not able to provide a sufficient generalization for making a prediction on the decision-making duration based on the available road infrastructure projects data.

   An optimization effort has been done to address these results by iterating several different variables combinations onto the model. The optimization results revealed that some factors influence the performance of the models; namely number of data entries, input variables influence, and representation of the world by the model. A combination of these three factors, which are: a small number of data entries, insufficient independent variables' influence on the dependent variable, and the dynamic system environment, affects the model prediction performance. With regard to the third factor, the model unable to represent the *Operating Environment* of the decision-making by the Dutch government due to the high dynamicity of the environment itself. This dynamicity is imposed by the introduction of other policies related to the decision-making process by the Dutch government outside the legal policy which the proposed model based on.

   Thus, it can be concluded that these AI models do not fit for its purpose, which is to be applied to forecast the introduction of new infrastructure project which based on the decision-making duration by the Dutch government on the preferred design alternative of the project.

7. ***Does the proposed AI model produce a superior result in comparison to conventional statistical methods?***

   The comparison of ANN models with Multiple Linear Regression model and Logistic Regression model shows that the AI model produced inferior result than its counterparts in term of model accuracy. While for the other benefits mentioned in 3.1.2, the inferior accuracy compares to the conventional statistical methods means the AI models were unable to handle the imprecise data and capture the non-linearity present in the dataset. This result proves the importance of data quality for any model. Because in the end, AI is a tool which will produce a bad output if feed in with bad input.

This page is left blank intentionally

Chapter 8 Future Recommendation

The finding of this research disincentives the future researches to forecast new road infrastructure introduction based on the decision-making duration by the Dutch government with Artificial intelligence. The low performance of AI forecasting model shows that the selected infrastructure development drivers are not enough as the predictor of new road project introduction due the strong influence of the policy intervention made by the government compares to the influence of those drivers on the decision-making duration. This is proven in the result discussion that the correlation value of the selected drivers with the decision-making duration is low while the influence from the policy interventions are more noticeable with a reduction in the duration average within the period where the policy was implemented.

It is true that those policy interventions are motivated by the drivers, but it is relatively difficult to forecast beforehand whether the government is going to introduce new policy at some point in the future based on the projected value of a certain set of variables within the forecasting period. For example, SVV-II was introduced with the main focus in the development of public transportation and the limitation in the expansion possibility of the existing road network (Ministerie van Verkeer en Waterstaat, 1988). This mobility plan was meant to be implemented until the year of 2010 based on the future prediction of infrastructure development drivers made by the Dutch government. However, during its implementation period, specifically in 2006, another mobility plan was introduced, which is known as Nota Mobilitiet (NoMo) to replace the SVV-II. In NoMo, the government abandoned the plan to restrict the existing road network expansion and be more open to facilitate traffic and transport growth (Ministerie van Verkeer en Waterstaat, 2004).

Thus, for the future researches which aim to forecast the introduction of a new road infrastructure project to the market with artificial intelligence in the Netherlands, a model focus realignment is proposed. It is recommended to change the model focus from a highly dynamic political environment towards a relatively more stable physical environment of an infrastructure asset in order to provide a reliable prediction. By shifting this focus, the following adjustments to the AI model and its *Operating Environment* are needed:

1. For the *AI Model Mission/Objective,* the AI model aims to predict a replacement project from the government on a certain component of an existing road infrastructure asset; such as a pavement component replacement or a bridge replacement. It means the focus of the forecasting model is no longer the large infrastructure projects but instead the smaller ones which do not require an exhaustive political infrastructure planning procedure.
2. For the *Operating Constraint Element*, the system focus is the management procedure during the Maintenance and Operation stage of the whole physical asset's life cycle (Hastings, 2014).
3. In regard to *Roles and Missions Element*, the role of Asset Management procedure during the maintain and operate stage is public utility system with a mission to provide its user with road service with a certain level of quality
4. For the *Resource Element or* the type of information and its source for the database of the AI forecasting model in the future, there are two aspects to be considered; which are functional lifetime and technical lifetime. Functional lifetime is the total time in which an asset is able to fulfill its functional performance requirements (Jaspers and Havelaar, 2017); while the Technical lifetime is the total time in which an asset is able to operate technically (UNFCCC/CCNUC, 2009).

a. *Functional Lifetime*
With regard to functional lifetime, a road infrastructure main function is to provide a service to its user with a certain set of performance requirements such as the road availability and the congestion level within a certain route.

b. *Technical Lifetime*
This aspect is directly connected with the life expectancy of the structural components, which comprise a road infrastructure asset. This life expectancy is important to be considered by the AI model because of maintenance or replacement decision for an infrastructure asset dependent on it (Hastings, 2014).

Thus, the potential type of information to be explored and its source for future research based on the functional and technical lifetime is as follows:

Type of information	**Data source**
List of the infrastructure Projects	www.tenderned.nl Company's Internal Database
Design Capacity	Project Technical Description
Usage Capacity	http://research.cbs.nl/verkeerslus/
Asset Component Life Expectancy	Component Technical Description

The following Figure 38 is presented to give a better overview of the *Operating Environment* adjustment for future research.

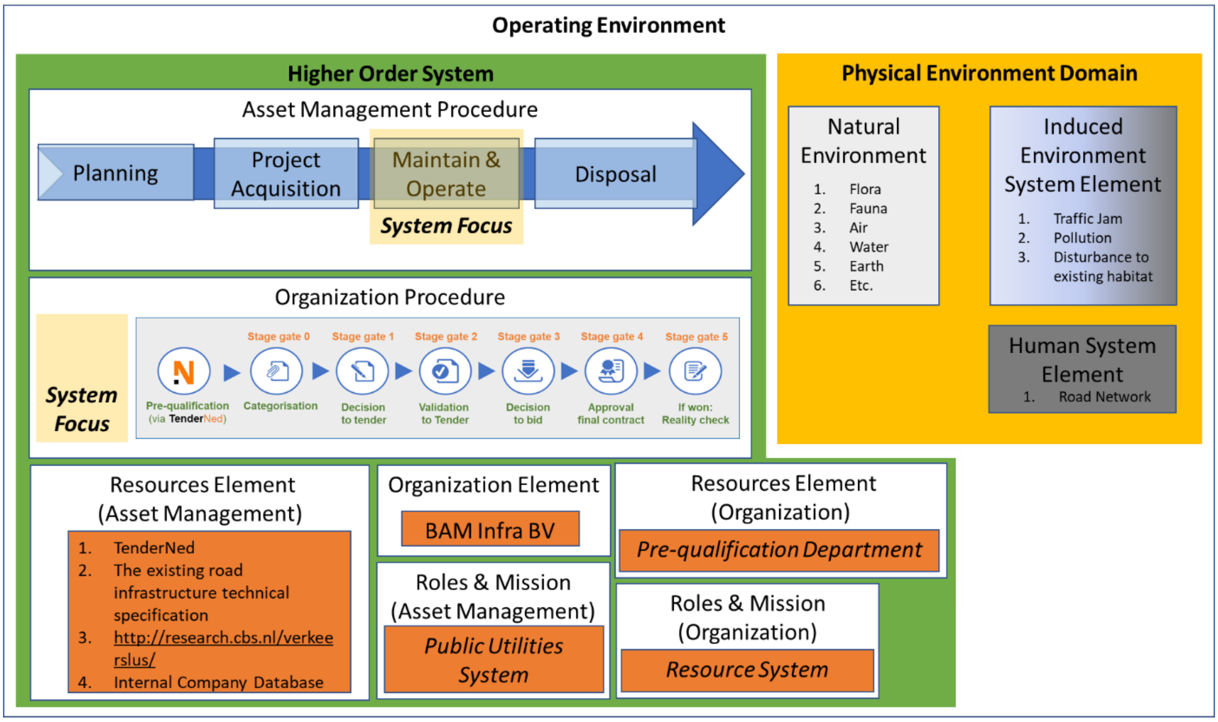


Figure 38 Modification of Operating Environment for future research

By realigning the focus of the future AI model as mentioned above, the following improvement on the factors which influence the performance of the proposed AI model in this research can be achieved:

1. In term of *number of data entries,* the focus on the smaller projects means a higher number of project is available for the database. This addition means the future AI model potentially be more reliable with more input data available to be considered.
2. In term of *Input Variables Influence* and *Representation of the world,* a more stable environment to be interpreted by the AI model might reduce the difficulty in identifying the relevant independent variables as the predictor for a future project introduction. This reduced selection difficulty might yield a better correlation value between the (to be) selected independent variables and the introduction of new maintenance or replacement project from the Dutch government; which eventually would produce a better prediction performance by the AI model.

References

al., P. e. (2011). Scikit-learn: Machine Learning in Python. *JMLR 12*, 2825-2830.

Alqahtani, A., & Whyte, A. (2013). Artificial Neural Networks incorporating cost significant items towards enhancing estimation for (life-cycle) costing of construction projects. *Construction Economics and Building, 13(3)*, 51-64.

Arafa, M., & Alqedra, M. (2009). Early stage cost estimation of buildings construction projects using artificial neural networks. *Journal of Computing in Civil Engineering 24, 35* .

Bremmer, I. (2012). Going Dutch: The Netherlands' election results roll in. *Foreign Policy*.

César CeCalderón, & Servén, L. (2010). Infrastructure in Latin America. *WB Policy Research Working Paper no. 5317*.

Cheng, M., Tsai, H., & Hsieh, W. S. (2009). Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model. *Automation in Construction 18*, 164.

Colomer, J. M. (2008). *Comparative European Politics.* Taylor & Francis.

De Jong, Gerard, & van de Riet, O. (2004). Drivers of Demand for Passenger Transport Worldwide. *European Transport Conference.* Strasbourg, France.

Elfaki, A. O., Alatawi, S., & Abushandi, E. (2014). Using intelligent techniques in construction project cost setimation: 10-year survey. *Advances in Civil Engineering* .

Estache, Antonio, & Goicoechea, A. (2005). A 'Research' Database on Infrastructure Economic Performance. *WB Policy Research Working Paper no. 3643*.

European Commision. (2006). *The Green Paper Energy.* Directorate General for Energy and Transport. Retrieved from http:// europa.eu.int/comm/energy/green-paper-energy/index_en.html

European Union. (2014). Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014. *Official Journal of the European Union*.

Fan, J., & Yan, G.-l. (2009). The System Dynamics Research on the Private Cars' Amount in Beijing. *Complex Scinces: FIrst International Conference, Complex 2009*.

Fay, M. (2001). Financing the Future: Infrastructure Needs in Latin America, 2000–05. *WB Policy Research Working Paper no. 2545*.

Géron, A. (2017). *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media.

Gierveld, H. (2016). *Tracéwet Tekst & Commentaar.* Utrecht.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* London: The MIT Press.

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M., & Seung, H. (2000). Digital Selection and Analogue Amplification coexist in a cortex-inspired silicon circuit. *Nature 405, 6789*, 947.

Hastings, N. A. (2014). *Physical Asset Management With an introduction to ISO55000 .* Melbourne: Springer International Publishing AG Switzerland .

Hertogh, M., & Westerveld, E. (2009). *Playing with Complexity: Management and Organisation of Large Infrastructure Projects.* Amsterdam: THIEME AMSTERDAM BV.

Hiromoto, M. H. (2012). Uma análise do efeito dos gastos dos governos sobre a pobreza no Brazil 1987 a 2009. *ANPEC*.

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression. Third Edition.* New Jersey: John Wiley & Sons.

Hyndman, R., & Athanasopoulos, G. (2019). *Forecasting: Principles and Practice.* Retrieved from https://otexts.com/fpp2/index.html

Jaspers, W., & Havelaar, M. (2017). *Long-term planning of large interventions within complex and dynamic infrastructure systems.* Thesis, University of Technology Delft, Delft.

Jong, F. (2016). *Planning and Development Law in the Netherlands. An Introduction.* The Hague: Instituut vor Bouwrecht.

Karlik, B., & Olgac, A. (2010). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks . *International Journal of Artificial Intelligence And Expert Systems (IJAE), Volume (1):*.

Kelleher, J., Mac Namee, B., & D'Arcy, A. (2015). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. *MIT Press*.

Keman, H. (2008). The Low Countries: Confrontation and Coalition in Segmented Societies. *Comparative European Politics*, p. 221.

Ko, C., & Cheng, M. (2007). Dynamic Prediction of Project Success Using Artificial Intelligence. *Journal of Construction Engineering and Management, 133(4)*, 316-324. doi:https://doi.org/10.1061/(asce)0733-9364(2007)133:4(316)

Lyne, R., & Maximinio, O. (2014). An artificial neural networks approach to structural cost estimation of building projects in Palestine. *DLSU Research Congress.*

Maghrebi, M., Sammut, C., & Waller, T. (2014). Predicting the duration of concrete operations via artificial neural networks and by focusing on supply chain parameters. *Building Research Journal, 61*, 1-14.

Marcio A. Salvato, e. (2007). Ciclos politicos: um estudo sobre a relação entre flutuações econômicas e calendário eleitoral no Brasil, 1985-2006. *Texto para Discussão, n. 42*.

Marianne, F., & Yepes, T. (2003). Investing in Infrastructure: What Is Needed from 2000 to 2010? *WB Policy Research Working Paper no. 3102.*

May, R., Dandy, G., & Maier, H. (2011). Review of Input Variable Selection Methods for Artificial Neural Networks. *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, 19-44.

Milieu, M. v. (2012). *Structuurvisie Infrastructuur en Ruimte.* Den Haag: rijksoverheid.

Ministerie van Infrastructuur en Milieu. (2017). *MIRT Overzicht.* Amsterdam: Must Stedebouw.

Ministerie van Verkeer en Waterstaat. (1988). *Tweede Struktuurschema Verkeer n Vervoer.*

Ministerie van Verkeer en Waterstaat, Diretoraat-General Rijkswaterstaat, Directie Utrecht. (2003). *Startnotitie A27 Plusstrook.* Apeldoorn/Amsterdam: Tik Design bv.

Pathak, M. (2018, May 22). Retrieved from https://www.datacamp.com/community/tutorials/categorical-data

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., & Pantouvakis, J. (2011). Early cost estimating of road tunnel construction using neural networks. *Journal of construction engineering and management 138*, 679.

Rothman, D. (2014). *Building Global Infrastructure: Forecasting the Next 50 Years.* Colorado, United States: Paradigm Publishers.

Russel, S., Norvig, P., & Davis, E. (2010). *Artificial Intelligence: A Modern Approach (3rd ed.).* Amsterdam: Prentice Hall.

Sakurai, S. N. (2009). Ciclos políticos nas funções orçamentárias dos municípios brasileiros: uma análise para o período 1990-2005 via dados em painel. *Estudos Econômicos, São Paulo, v. 39*, p. 39-58.

Sakurai, S. N., & Gremaud, A. (2007). Political business cycles: evidências empíricas para os municípios paulistas (1989-2001). *Estudos Econômicos, São Paulo, v. 11, n. 1*, p. 27-54.

Sawy, I. E., Hosny, H., & Razek, M. (2011). A neural network model for construction projects site overhead cost estimating in egypt. *International Journal of Computers Science, 8(3)*, 273–283.

Stulp , F., & Sigaud, O. (2015). Many regression algorithms, one unified model: A review. *Neural Networks 69*, 60.

Verschuren, P., & Doorewaard, H. (2007). *Het ontwerpen van een onderzoek (3rd ed.).* The Hague, The Netherlands: LEMMA.

Wang, W., Bilozerov, T., Dzeng, R., Hsiao, F., & Wang, K. (2017). Conceptual cost estimations using neuro-fuzzy and multi-factor evaluation methods for building projects. *Journal of Civil Engineering and Management 23*.

Wang, X., Duan, X., & Liu, J. (2010). Application of neural network in the cost estimation of highway engineering. *Journal of Computers, 5(11)*, 1762–1766.

Waterstaat, M. v. (2004). *Nota Mobiliteit Naar een betrouwbare en voorspelbare bereikbaarheid.*

Welch, J. (2006). Public-Private Partnerships in Germany: German Budget Deficits Present Investment Opportunities. *Infrastructure Journal*. Retrieved from http://www.ijonline.com/pdf/2006springIJ/ 10PPPsinGermany.pdf

Yahia, H., Hosny, H., & Razik, M. (2011). Time contingency assessment in construction projects in Egypt using artificial neural networks model. *International Journal of Computer Science, 8(4)*, 523–531.

Yu, W. d., & Skibniewski, M. (2011). Integrating neurofuzzy system with conceptual cost estimation to discover cost-related knowledge from residential construction projects. *Journal of Computing in Civil Engineering 24*, 35.

Zhou, L. (. (2018). *A Machine Learning Approach for Conceptual Cost Prediction of Road Pavement Project.*

Appendix 1 Infrastructure Planning Act (IPA)

In the Netherlands, there is almost continuous and long debate about the decision-making process for national infrastructure projects. This long debate finally led to a new act in 1994, which called The Infrastructure Planning Act 1994 (Dutch: *Tracewet*). The implementation of this act was still considered insufficient, and the need for a more efficient and faster decision-making procedure persisted. Because of this, the Government appointed the commission-Elverding in 2008 to assist them. Based on their analysis, one of the main causes which can be addressed with the legal measure is the number of the party who has the right to oppose through court against infrastructure decision. With the commission's assistance, a revised Infrastructure Planning Act 2012 was published and activated (Hobma and De Jong, 2016).

Infrastructure Planning Act Scope

IPA's effect is limited to only national infrastructure*,* which comprises of motorways, railways, and waterways at the national level; and relevant not only for new infrastructure but also for modification of existing infrastructure. Through this act, not only the Minister could have a legal platform to direct the infrastructure planning, but also the lower government bodies and interested parties could give their opinion on the infrastructure being planned legally.

Within IPA, relevant aspects from both the Spatial Planning Act (SPA) (Dutch: *Wet ruimtelijke ordening)* and Environmental Management Act (EMA) (Dutch: *Wet milliebeheer)* are considered. It means when planning new infrastructure, the government needs to consider the requirements from the spatial planning law and environmental law. Although there is a necessity to follow EMA, which mentions the requirement to draw Environmental Impact Statement (EIS); The release of the EIS is not explicitly mentioned in the procedures of IPA. It is because not every infrastructure project needs EIS even though in practice, most of the infrastructure projects need one.

Infrastructure Planning Act Procedure

There are two main procedures within the Infrastructure Planning Act, which are the full procedure and the shortened procedure. These two procedures would be illustrated in Table 16. The main difference between the two is the requirement to prepare a Structure Vision for the project. The following paragraph would give a clear definition of each term and how does it relate with each other. Due to the scope of the research, which is until the project is put to the tender market; then the explanation would only be given until step procedure number ten, *Appeal with the Council of State (*in relation to the *Track decision)*.

1. **Decision to start**
   Ministry of Infrastructure and the Environment decides to start an explorative study on an infrastructural problem; which can be an existing infrastructural problem or a future expected problem (art. 2 IPA). The contents of every decision to start are: [1] the description of the exploration area; [2] the description of the problem nature which being explored and the description of the spatial developments within the area; [3] the procedure how the interested parties would be involved during the exploration; and [4] the term for the exploration. The decision to start is both political and administrative decision to start a procedure, which makes it not possible to be appealed.
2. **Exploratory phase**
   In this phase, the minister will compile the knowledge required and insights about the problem's nature (art. 3 IPA). The main purpose of this phase is to study both the benefits and necessity of the proposed solution(s) for the infrastructure problem being studied.

Table 16 Full Procedure and Shortened Procedure (Hobma and De Jong, 2016)

Full Procedure	Shortened Procedure
1. Decision to start	1. Decision to start
2. Exploratory phase	2. Exploratory phase
3. Draft Structure vision	
4. View regarding Draft structure vision	
5. Structure vision	
6. Preferred decision	
7. Draft track decision	7. Draft track decision
8. Views regarding Draft track decision	8. Views regarding Draft track decision
9. Track Decision	9. Track Decision
10. Appeal with the Council of State	10. Appeal with the Council of State
11. Coordinated permit procedure	11. Coordinated permit procedure
12. Appeal with the Council of State	12. Appeal with the Council of State
13. Construction and operation	13. Construction and operation
14. Evaluation test	14. Evaluation test

3. **Draft structure vision**
   The ‘decision to start’ would include the ‘structure vision’ preparation under certain conditions. The motivation behind this additional requirement is the big impact of new national infrastructure or major intervention, which essential to be coordinated with the other spatial development within the area. Although the decision to prepare a structure vision is made, it does not necessarily mean that the proposed project or intervention would be realized. Only after the structure vision is finished, the minister would be clearer with his ‘preferred decision’. A major intervention in this context is the new national project or major modification to existing national infrastructure (higher than two lanes for motorway or tracks for railway). IF NO ‘structure vision’ is required, then the ‘shortened procedure’ would apply.
4. **View regarding Draft structure vision**
   Everybody can have the opportunity to express their opinion on the draft structure vision when it is ready (art. 6 IPA). Thus, the minister could incorporate their view on the final version of the structure vision.
5. **Structure vision**
   The possible solutions and each alternative’s environmental effects are reported in ‘structure vision’; It means EIS is also included in this vision due to its mandatory stature according to European Environmental Law. The final version would be sent to the Second Chamber of Parliament, the government bodies involved, and the national railway's manager (art. 7 IPA). Within the final version, the ‘preferred decision’ of the minister is also be included. There is a possibility for the final version to be discussed not only in the parliament but also in the concerned provincial and municipal council. This situation puts the ‘structure vision to the domain of the politics. During the discussion on these different level councils, the interested civilians and group of people with special interest would try to influence the politicians involved to either support or reject, at least weaken the minister’s preferred decision.
6. **Preferred Decision**
   The ‘structure vision’ is not completely neutral, because it contains the preferred decision of the minister to solve the problem being studied. One important note is, the preferred decision

does not necessarily mean that it will be realized; because there are more steps to come in the procedure.

7. **Draft track decision**

   At some point, which solution to be realized to solve the infrastructural problem being studied would be clear. The solution could be the minister's preferred solution or the result from the discussion with the interested parties, especially the Second Chamber of Parliament. The solution to be realized would be further developed in a draft-track decision (*ontwerp-tracebesluit).* Province, municipalities, and water boards are involved in the development (art. 11, para. 2, IPA). Most of the time, the draft track decision includes the Environmental Impact Statement. There are many specifications within the draft track decision, such as:

   - The number of lanes for a motorway or number of railway track.
   - Infrastructural measures to be taken.
   - Measures to be taken to compensate for damage incurred to nature.
   - Safety measures
   - Maps in detail
   - The noise production ceilings

8. **Views regarding Draft track decision**

   The municipalities concerned, civilians, organizations, and a group of special interest can express their opinions about the draft track decision. It is specified in IPA that 'everyone' can put forward their opinion (art. 11, para. 1).

9. **Track Decision**

   The minister would decide on the solution, which is to be realized in detail after reviewing the opinion from the parties on the draft track decision. This decision is called 'track decision' (*tracebesluit)* with the same specifications as the draft track decision. One of the most important traits from track decision is its's special planning status; The track decision counts as a decision to deviate from a land-use plan (art. 13, para. 4, IPA). It means the concerned municipality can no longer object and block the realization of the infrastructure based on their current land-use plan.

10. **Appeal with the Council of State**

    There is still an opportunity for the interested parties to lodge an appeal against the released 'track decision' as long as they had expressed their view in the previous stages with the exclusion of municipality. This exclusion is stipulated in the Crisis and Recovery Act (art. 1.4)(*Crisis en herstelwet)* with the purpose of swift decision-making. Prior to the Crisis and Recovery Act, it is very common for local government and the central government to fight each other in the court. In Crisis and Recovery Act article 1.6, para. 4, it is also stipulated that the Council of State has to come to a ruling within six months.

Appendix 2 Interviews with BAM's Expert

Pre-qualification Tender department (BAM_1)

Interview Date : 18 February 2019
Person : XXX
Subject : Current Market Tender Approach

Q1: What is the general procedure of this department in term of addressing the upcoming infrastructure project from the government?

Every region branch of this company has its own pre-qualification department which responsible for receiving an invitation to tender from the client via TenderNed and distribute it to Tender Strategy Department.

Q2: Based on the procedure of this department, is it safe to conclude that the company approaches the market in a passive manner by waiting for an invitation to tender via TenderNed?

Yes, that is correct. The company is waiting for a project to be available on the market before making any engagement with the market.

Q3: In addition to the current passive approach, is there any procedure to forecast the upcoming infrastructure project based on the external factors (e.g., demand drivers of the project)?

Currently, there is none. But, as a company, there is a necessity to have a knowledge about the upcoming project prior the market announcement in order for the company to prepare better for the tender in term of acquiring the potential partners / subcontractors for the project, researching on new innovative materials to be applied, and forecasting the potential volume of a certain material for the future project (e.g. asphalt, concrete).

Tender Strategy Department (BAM_2)

Interview Date : 01 April 2019
Person : XXX
Subject : Current Market Tender Approach

Q1: What is the general procedure of this department to process a new project after receiving the invitation and its documents from the Pre-qualification Department?

In the company, there is a procedure which addresses the tender procedure called Stage-Gate Procedure that the whole BAM Group, including BAM Infra BV, refers to it. In regard to tender strategy department, our responsibility is to oversee those gates starting from GATE 2 in the form of analyzing the project characteristics and the risk profile, presenting to the board about the project prospect prior and post tender related to the company interest (potential profit, expertise suitability with the company, ability to minimize the risk), formulate the strategy to win the tender, and finally undergoing the tender process.

Q2: What is the relevancy of the official documents and information announced by the government for BAM Infra BV to identify an upcoming infrastructure project?

The company does not utilize the documents due to the uncertain nature of the information it provides. Although they do inform the company about the potential timeframe of a new project introduction, it is still uncertain when the project would be introduced to the market.

Q3: How does the early information provided by the official documents could help the tender Strategy team to prepare better prior to submitting a tender bid?

The early information could help the team to have a better approximation about the potential criteria to be asked by the client regarding the new project during the tender procedure. It Is helpful because identifying the important criteria of a project takes a lot of our time in formulating a bid.

Q4: What is your view on the benefit of this early knowledge about a new project introduction to the company's processes in general?

The early information about a new project introduction could provide benefit in the earlier phases prior to preparing a bid. For example, the early information about the project could help the company in identifying the type of material to be researched on based on the project's requirements, reviewing the suitability of internal expertise with the potential project's requirements, early communication with potential partners to fill the resource gap.

Commercial Manager (BAM_3)

Interview Date : 11 April 2019
Person : XXX
Subject : Current Market Tender Approach

Q1: What is your main responsibility as a Commercial Manager?

Responsibility of commercial manager is dependent on the project valuation. There are three project categories, which are Small Projects, Medium Projects, and Large Projects. My responsibility is to identify the small project which does not require an extensive planning procedure within my region.

Q2: This research focus is the large road infrastructure projects and its related planning procedure. Is it possible for us to discuss this matter considering your responsibility in the Small Project category?

Yes, my study is related to spatial planning. Thus, I have the knowledge about the planning procedure related to the large road infrastructure projects.

Q3: Based on my previous interviews, there are several potential benefits for BAM Infra BV by utilizing the official government documents (e.g., Startnotitie, MIRT, Ontwerp-Tracebesluit, Tracebesluit) about an upcoming infrastructure project. What is your view on this matter?

All the processes related to those documents are mostly engineering works (i.e., design, impact assessment). That kind of work is the main interest of consultant engineering companies instead of a construction company like BAM Infra BV. Thus, if the government announce an engineering consultancy work about a certain project, there is already an indication that the project is going to be introduced to the market within several years. This kind of announcement normally happens after the OTB is ready, and the government wants to make a more detailed design. In regard to design development after OTB, the government is starting to involve the construction company more in recent years. This development indicates that in the future, the construction companies, such as BAM Infra BV, might already be involved as early as the project design development. Thus, information after OTB would not be a useful addition to the company.

Q4: Is there any formal procedure within BAM Infra BV to manage these official governments and utilize it to forecast the upcoming project?

We do not have a formal procedure and resources which allocated to manage these documents at the moment.

Commercial Manager & Innovation (BAM_4)

Interview Date : 29 July 2019
Person : XXX
Subject : ANN Models Validation

Q1: What is your main responsibility within BAM Infra BV?

My main responsibility is the development of digital services for asset management. I tried to explore how can BAM utilize technology, such as artificial intelligence, BIM, data centralization, to improve the current asset management processes. In the past, I was involved a lot in the tender as the tender manager for infrastructure project throughout the whole process from the announcement, prospect, to bidding.

Q2: After I explained this research objective, method, and its outcome, what do you think about the result? Are the model results acceptable by BAM as the future user of the model? If not, what is the range of error that you expect from the model to be deemed as applicable for the company?

From the strategic portfolio management point of view, the range that you want to consider is 0,5 years – 1,5 years in the future. Because of this kind of management concerns about two things, the workload management, and strategic partnership. With a standard deviation of 2,5 years as the model suggested; then we are looking at a five-year time-window. A time-window with this value still possesses a high uncertainty and not applicable yet from our perspective. The range that we expect is a deviation of (0,75 – 1 year) or (1,5 – 2 year) time-window.

Appendix 3 Models' Dataset

Raw Dataset

N	Startnotitie	Road Catego	Type of Network Interventio	Number of Provin	Geographical Profile	EHS/GH	GDP (million euro / year)	3 years Average GDP Growth (%)	RDP (%)	Province(s) RD	Populatio	Change of Powe	Ideology	c/km2 P1	Duration (years
1	**A2 Holendrecht-Oudenrijn**	National	Modification	2	Rural	Yes	414838	0.053703539	0.2715	112644.2005	3601880	No	CenterLeft	20357.33251	6.67
2	**A2 Maasbracht-Geleen**	National	Modification	1	Rural	No	545609	0.033350577	0.0612	33414.1104	1136695	No	Center	6868.199366	5.04
3	**A10 Tweede Coentunnel / A5 Westrandweg**	National	New	1	Combination	Yes	523939	0.032014268	0.1864	97684.9526	2587265	No	Center	11389.30057	2.95
4	**A12 Utrecht-Veenendaal**	National	Modification	1	Rural	Yes	364961	0.057197969	0.0826	30159.8804	1079428	Yes	CenterLeft	9363.410326	4.38
5	A28 Utrecht-Amersfoort	National	Modification	1	Combination	Yes	639163	0.054203372	0.0844	53950.5218	1201350	Yes	Center-Right	10836.177	2.42
6	N50 Ramspol-Ens	Provincial	Modification	2	Rural	Yes	494501	0.060448747	0.0738	36483.7641	1435753	Election	Election	9276.929775	5.54
7	**A2 Everdingen - Deil**	National	Modification	2	Rural	Yes	308909	0.04235228	0.3242	100151.6723	5164595	Election	Election	13370.8098	14.00
8	A2 Holendrecht-Maarssen (A2 Holendrecht-Oudenrijn)	National	Modification	2	Rural	Yes	579212	0.045676375	0.2675	154944.8068	3786623	Election	Election	22096.36228	3.59
9	**A2 Rondweg Den Bosch**	National	Modification	1	Rural	Yes	293610	0.044177598	0.1494	43859.4610	2243546	No	Center	4256.673762	2.79
10	A2 St.Joost-Urmond	National	Modification	1	Rural	No	545609	0.033350577	0.0612	33414.1104	1136695	No	Center	6868.199366	1.50
11	**A2 Tangenten Eindhoven**	National	Modification	1	Urban	No	308909	0.04235228	0.1494	46143.6017	2259779	Election	Election	4348.547697	5.75
12	**A7 Rondweg Sneek**	National	Modification	1	Urban	No	325341	0.045099165	0.0330	10751.5864	609579	Yes	CenterLeft	2347.527638	9.08
13	**A9 Holendrecht-Diemen**	National	Modification	1	Urban	Yes	545609	0.033350577	0.1857	101330.5587	2599103	No	Center	11495.00193	1.95
14	**A12 Woerden-Gouda**	National	Modification	2	Rural	Yes	545609	0.033350577	0.3031	165388.2628	4629672	No	Center	23705.57927	4.92
15	**A1 Diemen-Muiderberg & A1/A6 Muiderberg-Almere Stad West**	National	Modification	2	Combination	Yes	545609	0.033350577	0.2021	110294.6202	2964962	No	Center	16251.16893	2.06
16	**A2 Leenderheide-Valkenswaard**	National	Modification	1	Rural	Yes	613280	0.053921893	0.1467	89960.5616	2419042	No	Center	6724.632478	2.30
17	**A2/A27 Everdingen-Lunetten**	National	Modification	1	Rural	Yes	545609	0.033350577	0.0859	46863.9222	1171291	No	Center	10691.14551	5.04
18	**A9 Raasdorp-Badhoevedorp & A9 Velsen-Raasdorp**	National	Modification	1	Rural	Yes	545609	0.033350577	0.1857	101330.5587	2599103	No	Center	11495.00193	4.58
19	A12 Woerden-Oudenrijn	National	Modification	1	Rural	Yes	545609	0.033350577	0.0859	46863.9222	1171291	No	Center	10691.14551	4.42
20	A28 Zwolle - Meppel	National	Modification	2	Combination	Yes	545609	0.033350577	0.0811	44231.5593	1592801	No	Center	8253.494209	3.51
21	A50 Valburg-Grijsoord	National	Modification	1	Combination	Yes	340964	0.05110686	0.1029	35100.3487	1876299	Yes	CenterLeft	4768.926625	10.49
22	A58 Eindhoven-Oirschot	National	Modification	1	Rural	Yes	579212	0.045676375	0.1474	85371.7333	2415946	Election	Election	6594.468956	2.92
23	N34 Omleiding Ommen	Provincial	Modification	1	Urban	Yes	325341	0.045099165	0.0593	19307.6664	1050389	Yes	CenterLeft	3553.652505	10.34
24	A9 Alkmaar-Uitgeest	National	Modification	1	Rural	Yes	613280	0.053921893	0.1810	110978.0311	2613070	No	Center	11666.43374	3.08
25	**A12 Zoetermeer-Zoetermeer centrum (noordbaan)**	National	Modification	1	Urban	No	617540	0.022397292	0.2117	130734.4574	3481558	No	Center	13513.1658	1.18
26	A74 Venlo	National	New	1	Rural	No	325341	0.045099165	0.0642	20888.2178	1130050	Yes	CenterLeft	4648.614773	8.51
27	N9 Koedijk-De Stolpen	Provincial	Modification	1	Rural	Yes	325341	0.05110686	0.1814	59018.6208	2463611	Yes	CenterLeft	8829.972099	9.08
28	N200 Halfweg	Provincial	Modification	1	Urban	No	293610	0.04235228	0.1830	53719.1750	2440165	No	Center	8480.177376	5.62
29	A4 Burgerveen - Leiden	National	Modification	1	Combination	Yes	308909	0.04235228	0.2204	68074.2722	3313193	Election	Election	9668.005803	3.51
30	A4 Dinteloord-Bergen op Zoom	National	New	1	Rural	Yes	285059	0.053703539	0.1494	42582.7967	2225331	No	Center	4064.157142	4.63
31	N11 Zoeterwoude-Alphen aan den Rijn	Provincial	Modification	1	Rural	Yes	494501	0.060448747	0.2145	106068.9732	3423780	Election	Election	12523.82684	7.84
32	N31 Leeuwarden (Haak)	Provincial	New	1	Combination	Yes	494501	0.060448747	0.0316	15639.2396	636184	Election	Election	3235.22839	4.25
33	N33 Assen-Zuidbroek	Provincial	Modification	2	Rural	Yes	579212	0.045676375	0.0687	39783.4593	1058523	Election	Election	7387.079503	4.05
34	N61 Hoek-Schoondijke	Provincial	Modification	1	Rural	Yes	389315	0.06755634	0.0207	8054.5190	369949	Election	Election	2003.009586	7.00
35	Kanaalkruising Sluiskil	Provincial	New	1	Rural	Yes	506671	0.041956742	0.0212	10733.7403	378348	Election	Election	2222.7945	3.84
36	A28/A1 Knooppunt Hoevelaken	National	Modification	1	Rural	Yes	639163	0.054203372	0.0844	53950.5218	1201350	No	Center	10836.177	9.67
37	A7/A8 Purmerend - Zaandam - Coenplein	National	Modification	1	Rural	Yes	494501	0.060448747	0.1856	91787.1256	2559477	Election	Election	11476.36461	4.00
38	A12 Ede - Grijsoord	National	Modification	1	Rural	Yes	364961	0.061710222	0.1035	37766.5005	1886109	Yes	CenterLeft	4885.546229	12.76
39	A15 Maasvlakte - Vaanplein	National	Modification	1	Urban	No	340964	0.05110686	0.2166	73856.0728	3332860	Yes	CenterLeft	10794.92708	4.59
40	A4 Delft-Schiedam	National	New	1	Combination	No	523939	0.032014268	0.2151	112716.6889	3451942	No	Center	12929.6424	6.00
41	A2 passage Maastricht	National	Modification	1	Urban	Yes	523939	0.032014268	0.0628	32924.3240	1139335	No	Center	6852.056972	6.00
42	A9 omlegging Badhoevedorp	National	Modification	1	Urban	Yes	613280	0.053921893	0.1810	110978.0311	2613070	No	Center	11666.43374	4.33
43	A50 Ewijk - Valburg	National	Modification	1	Rural	Yes	340964	0.05110686	0.1029	35100.3487	1876299	Yes	CenterLeft	4768.926625	12.42
44	N50 Ens - Emmeloord	Provincial	Modification	1	Rural	Yes	642929	0.002291302	0.0172	11045.9200	391967	Yes	Center-Right	5683.5215	2.92
45	A1 Apeldoorn-Zuid - Beekbergen	National	Modification	1	Combination	No	617540	0.022397292	0.1029	63552.5347	1991062	No	Center	6039.115205	4.58
46	N18 Varsseveld-Enschede	Provincial	Modification	2	Combination	Yes	545609	0.033350577	0.2783	151864.5852	3081442	No	Center	10857.72655	6.67
47	N31 Harlingen (traverse Harlingen)	Provincial	Modification	1	Rural	No	617540	0.022397292	0.0314	19404.1370	644811	No	Center	3645.96837	2.42
48	N35 Zwolle - Wijthmen	Provincial	Modification	1	Combination	No	631512	0.010333096	0.0596	37649.2816	1130345	Election	Election	5305.70102	3.59
49	**A27/A1 Utrecht-Noord - knooppunt Eemnes -aansluiting Bunschoten-Spakenburg**	National	Modification	1	Rural	Yes	639163	0.054203372	0.0844	53950.5218	1201350	No	Center	10836.177	2.53
50	A27/A12 Ring Utrecht	National	Modification	1	Combination	Yes	639163	0.054203372	0.0844	53950.5218	1201350	No	Center	10836.177	7.25
51	**A1/A6/A9 Schiphol-Amsterdam-Almere**	National	Modification	2	Combination	Yes	523939	0.032014268	0.2367	124023.7841	2947169	No	Center	16046.2402	5.32
52	A12/A15 Ressen - Oudbroeken (ViA15)	National	Modification	1	Urban	Yes	639163	0.054203372	0.1013	64722.5635	1983869	No	Center	5925.330589	7.51
53	Rijnlandroute (A4 and A44)	National	Modification	1	Combination	Yes	639163	0.054203372	0.2133	136302.5062	3461435	No	Center	13229.72191	5.42
54	**A4 Vlietland-N14**	National	Modification	1	Rural	No	652748	0.011103421	0.2103	137285.9651	3563935	No	Center-Right	14501.52862	0.84
55	A15/A24 Blankenburgverbinding	National	New	1	Urban	No	642929	0.002291302	0.2090	134357.3254	3528324	Yes	Center-Right	14113.296	4.67
56	A76 Ansluiting Nuth	National	New	1	Combination	Yes	617540	0.022397292	0.0594	36687.0912	1122604	No	Center	6948.008541	1.00
57	Zuidasdok	National	Modification	1	Urban	No	663008	0.010316531	0.1827	121134.3100	2741369	No	Center-Right	13175.35509	0.54
58	A7 Zuidelijke Ringweg Groningen Fase 1	National	Modification	1	Urban	No	285059	0.044177598	0.0366	10436.4447	555226	No	Center	2833.285618	9.68
59	A7 Zuidelijke Ringweg Groningen Fase 2	National	Modification	1	Urban	No	617540	0.022397292	0.0457	28228.2146	574092	No	Center	4456.51078	3.67
60	A27 Lunetten-Hooipolder	National	Modification	2	Rural	Yes	613280	0.053921893	0.0855	52423.0070	3609646	No	Center	17688.11098	8.59
61	A2 Grathem - Urmond	National	Modification	1	Rural	Yes	414838	0.06755634	0.0651	26989.6647	1139302	No	CenterLeft	6406.509797	7.00
62	A4 De Hoek - Prins Clausplein route	National	Modification	2	Rural	Yes	414838	0.06755634	0.4374	181452.8568	5881928	No	CenterLeft	22044.72319	6.00
63	A2 Oudenrijn-Deil	National	Modification	2	Rural	Yes	414838	0.06755634	0.3234	134150.5852	3005500	No	CenterLeft	15051.75706	4.73
64	A2: Everdingen junction Deil junction (future) Zaltbommel junction Empel junction	National	Modification	3	Rural	Yes	308909	0.045099165	0.4736	146295.2740	7424374	Election	Election	17719.3575	5.00
65	A2: route 's-Hertogenbosch - Eindhoven	National	Modification	1	Rural	Yes	414838	0.06755634	0.1495	62001.2516	2337709	No	CenterLeft	6006.828424	3.75
66	A1: Eemnes - Barneveld	National	Modification	2	Combination	Yes	414838	0.06755634	0.1909	79190.1641	3005500	No	CenterLeft	15051.75706	3.00
67	A1 Barneveld - Deventer	National	Modification	2	Rural	No	414838	0.06755634	0.1618	67132.5749	2977222	No	CenterLeft	9701.79829	5.00
68	A1 Apeldoorn - Azelo	National	Modification	2	Rural	No	652748	0.011103421	0.1603	104652.4080	3155141	No	Center-Right	11929.64657	3.27
69	N31 Zurich - Harlingen	Provincial	Modification	1	Combination	No	494501	0.060448747	0.0316	15639.2396	636184	Election	Election	3235.22839	3.05
70	A2 Meerenakkerweg	National	Modification	1	Urban	No	579212	0.045676375	0.1474	85371.7333	2415946	Election	Election	6594.468956	1.81
71	**A59 Brug over het Drongelens Kanaal**	National	Modification	1	Rural	Yes	652748	0.011103421	0.1500	97934.4674	2471011	No	Center-Right	7244.206161	1.00
72	N35 Nijverdal-Wierden	Provincial	Modification	1	Rural	No	642929	0.002291302	0.0596	38320.4990	1134465	Yes	Center-Right	5371.34449	4.01
73	A2 Leenderheide - Budel	National	Modification	1	Rural	Yes	414838	0.06755634	0.1495	62001.2516	2337709	No	CenterLeft	6006.828424	8.22
74	A1 t' Gooi	National	Modification	1	Urban	Yes	545609	0.033350577	0.1857	101330.5587	2599103	No	Center	11495.00193	4.59
75	A27 Gorinchem - Noordeloos	National	Modification	1	Rural	Yes	506671	0.041956742	0.2151	108968.0735	3439982	Election	Election	12764.14374	0.76
76	A12 Veenendaal-Ede	National	Modification	2	Rural	Yes	340964	0.057197969	0.1858	63349.5341	2946884	Yes	CenterLeft	13881.34779	8.00
77	A4 Badhoevedorp-Nieuwe Meer	National	Modification	1	Rural	No	545609	0.033350577	0.1857	101330.5587	2599103	No	Center	11495.00193	4.67
78	A1 Watergraafsmeer - Diemen	National	Modification	1	Urban	No	545609	0.033350577	0.1857	101330.5587	2599103	No	Center	11495.00193	1.00
79	A15 Papendrecht - Sliedrecht	National	Modification	1	Combination	No	683457	0.019438693	0.2147	146710.7635	3577032	No	Center-Right	14775.69732	2.17
80	A15 Papendrecht - Hardinxveld-Giessendam - Gorinchem	National	Modification	1	Combination	Yes	545609	0.033350577	0.2172	118524.3407	3458381	No	Center	13014.43376	2.20
81	A59 Rosmalen-Geffen	National	Modification	1	Rural	Yes	285059	0.053703539	0.1494	42582.7967	2225331	No	Center	4064.157142	2.57
82	A16 Rotterdam	National	New	1	Combination	Yes	545609	0.033350577	0.2172	118524.3407	3458381	No	Center	13014.43376	9.50
83	A15 knooppunt Ressen-rijksweg 12	National	New	1	Rural	No	272796	0.053703539	0.1045	28505.8244	1816935	No	Center	3371.531791	5.43
84	RIJKSWEG 14 'VERLENGDE LANDSCHEIDINGSWEG'	National	New	1	Urban	No	257943	0.053703539	0.2020	52107.6996	3219839	No	Center	8790.671556	4.55
85	N23 Lelystad-Dronten	Provincial	New	1	Rural	No	545609	0.033350577	0.0164	8964.0615	365859	No	Center	4756.167	1.14
86	N33, Spijk - Eemshaven	Provincial	Modification	1	Combination	No	325341	0.045099165	0.0440	14329.4091	557995	Yes	CenterLeft	2885.631286	2.67
87	Rijksweg 69, Eindhoven - Belgische	National	Modification	2	Rural	Yes	308909	0.04235228	0.1038	32077.4001	3384966	Election	Election	8900.440561	5.35
88	A2: route 's-Hertogenbosch - Eindhoven	National	Modification	1	Rural	Yes	639163	0.054203372	0.1460	93342.2837	2424827	No	Center	6792.527807	3.22
89	RIJKSWEG 37 EN RIJKSWEG 34 (GEDEELTE HOLSLOOT-EMMEN-ZUID)	Provincial	Modification	1	Rural	No	272796	0.053703539	0.0262	7150.6812	443510	No	Center	1789.511279	9.76
90	Rijksweg 73-Zuid	Provincial	New	1	Rural	No	285059	0.053703539	0.0675	19232.4332	1115485	No	Center	4300.814389	3.13
91	rijksweg 31, gedeelte Leeuwarden–Drachten	Provincial	Modification	1	Rural	Yes	293610	0.044177598	0.0342	10042.0792	603998	No	Center	2255.354234	3.17
92	A1 Hoevelaken - Barneveld	National	Modification	1	Rural	Yes	523939	0.032014268	0.1011	52970.5503	1966929	No	Center	5752.593719	3.94
93	A28 Leusden Zuid - A1 Hoevelaken	National	Modification	2	Combination	Yes	506671	0.041956742	0.1886	95539.2125	3112640	Election	Election	16049.81478	5.30
94	A12 Woerden - Gouwe	National	Modification	2	Rural	Yes	506671	0.041956742	0.3015	152755.0202	4592200	Election	Election	23134.11	2.30
95	A27 Plusstrook Utrecht richting Eemnes	National	Modification	1	Combination	Yes	506671	0.041956742	0.0864	43786.9467	1152218	Election	Election	10369.962	1.87

ABT Dataset for Regression Models

Startnotitie	National Road	Provincial Road	Modification Project	New Project	Number of Provin	Rural	Urban	Combination	EHS YE	EHS NO	GDP	3 years Average GDP Growth	RDP (%)	Populatio	Yes Change of Power	No Change of Power	Election Change of Power	CenterLeft Ideolog	CenterRight Ideolo	Center Ideology	Election	c/km2 P1	Duration (year
A2 Holendrecht-Oudenrijn	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	0.712	1.433	1.060	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	2.124	6.668
A2 Maasbracht-Geleen	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-0.932	-0.874	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.529	5.041
A10 Tweede Coentunnel / A5 Westrandweg	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.235	-0.625	0.476	0.264	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.360	2.951
A12 Utrecht-Veenendaal	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.031	0.927	-0.691	-0.919	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.039	4.384
A28 Utrecht-Amersfoort	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	-0.671	-0.823	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	0.251	2.416
N50 Ramspol-Ens	-1.884	1.884	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	-0.791	-0.639	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.056	5.537
A2 Holendrecht-Maarssen (A2 Holendrecht-Oudenrijn)	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	1.387	1.205	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	2.466	3.589
A2 Rondweg Den Bosch	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.599	0.125	0.059	-0.006	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.043	2.789
A2 St.Joost-Urmond	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-0.932	-0.874	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.529	1.504
A2 Tangenten Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.477	0.012	0.059	0.007	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.025	5.745
A7 Rondweg Sneek	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.346	0.181	-1.249	-1.288	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.419	9.079
A9 Holendrecht-Diemen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	1.945
A12 Woerden-Gouda	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	1.788	1.866	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	2.783	4.921
A1 Diemen-Muiderberg & A1/A6 Muiderberg-Almere Stad West	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.653	0.560	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.316	2.063
A2 Leenderheide-Valkenswaard	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	0.029	0.132	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.558	2.299
A2/A27 Everdingen-Lunetten	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	-0.655	-0.847	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.223	5.044
A9 Raasdorp-Badhoevedorp & A9 Velsen-Raasdorp	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	4.584
A12 Woerden-Oudenrijn	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	-0.655	-0.847	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.223	4.416
A28 Zwolle - Meppel	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	-0.709	-0.516	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.257	3.507
A58 Eindhoven-Oirschot	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	0.037	0.129	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.583	2.915
N34 Omleiding Ommen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	-1.346	0.181	-0.953	-0.942	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.181	10.340
A9 Alkmaar-Uitgeest	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	0.414	0.284	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.415	3.082
A12 Zoetermeer-Zoetermeer centrum (noordbaan)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.980	-1.218	0.760	0.965	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.778	1.178
A74 Venlo	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.346	0.181	-0.898	-0.879	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.966	8.510
N9 Koedijk-De Stolpen	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.346	0.552	0.419	0.167	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.143	9.085
N200 Halfweg	-1.884	1.884	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.599	0.012	0.437	0.148	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.212	5.622
A4 Burgerveen - Leiden	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	-1.477	0.012	0.857	0.833	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.021	3.512
A4 Dinteloord-Bergen op Zoom	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.667	0.712	0.059	-0.020	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.081	4.630
N11 Zoeterwoude-Alphen aan den Rijn	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	0.791	0.920	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.583	7.838
N31 Leeuwarden (Haak)	-1.884	1.884	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.000	1.127	-1.265	-1.267	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.244	4.247
N33 Assen-Zuidbroek	-1.884	1.884	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	-0.848	-0.935	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.427	4.052
N61 Hoek-Schoondijke	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.837	1.565	-1.388	-1.476	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.486	7.000
Kanaalkruising Sluiskil	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	-1.382	-1.469	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.443	3.836
A28/A1 Knooppunt Hoevelaken	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	9.671
A7/A8 Purmerend - Zaandam - Coenplein	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	0.467	0.242	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.377	4.003
A15 Maasvlakte - Vaanplein	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.222	0.552	0.815	0.849	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	0.243	4.595
A4 Delft-Schiedam	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.235	-0.625	0.798	0.942	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.663	6.003
A2 passage Maastricht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.235	-0.625	-0.914	-0.872	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.533	6.003
A9 omlegging Badhoevedorp	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.946	0.725	0.414	0.284	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.415	4.332
N50 Ens - Emmeloord	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.182	-2.456	-1.427	-1.458	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	-0.762	2.918
A1 Apeldoorn-Zuid - Beekbergen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.980	-1.218	-0.463	-0.204	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.692	4.584
N18 Varsseveld-Enschede	-1.884	1.884	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	1.509	0.651	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.255	6.674
N31 Harlingen (traverse Harlingen)	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.980	-1.218	-1.267	-1.260	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.163	2.419
N35 Zwolle - Wijthmen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	1.091	-1.961	-0.950	-0.879	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.837	3.589
A27/A1 Utrecht-Noord - knooppunt Eemnes -aansluiting Bunschoten-Spakenburg	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	2.534
A27/A12 Ring Utrecht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	7.252
A1/A6/A9 Schiphol-Amsterdam-Almere	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.235	-0.625	1.041	0.546	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.276	5.318
A12/A15 Ressen - Oudbroeken (ViA15)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	1.152	0.742	-0.482	-0.210	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.715	7.507
Rijnlandroute (A4 and A44)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	0.777	0.950	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.722	5.416
A4 Vlietland-N14	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	1.260	-1.913	0.744	1.030	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.972	0.844
A15/A24 Blankenburgverbinding	0.531	-0.531	-2.449	2.449	-0.517	-1.129	2.014	-0.565	-1.391	1.391	1.182	-2.456	0.729	1.002	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	0.896	4.668
A76 Ansluiting Nuth	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.980	-1.218	-0.952	-0.885	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.514	1.000
Zuidasdok	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	1.342	-1.962	0.434	0.385	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.711	0.537
A7 Zuidelijke Ringweg Groningen Fase 1	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.667	0.125	-1.209	-1.330	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.323	9.677
A7 Zuidelijke Ringweg Groningen Fase 2	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.980	-1.218	-1.106	-1.315	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.004	3.668
A27 Lunetten-Hooipolder	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	-0.659	1.066	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.599	8.589
A2 Grathem - Urmond	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	-0.889	-0.872	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.620	7.000
A4 De Hoek - Prins Clausplein route	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	3.298	2.848	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	2.456	6.000
A2 Oudenrijn-Deil	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	2.016	0.592	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	1.080	4.729
A2: Everdingen junction Deil junction (future) Zaltbommel junction Empel junction	0.531	-0.531	0.408	-0.408	3.961	0.886	-0.497	-0.565	0.719	-0.719	-1.477	0.181	3.704	4.058	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	1.605	5.000
A2: route 's-Hertogenbosch - Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	0.060	0.068	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.699	3.753
A1: Eemnes - Barneveld	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	-0.634	1.565	0.526	0.592	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	1.080	3.000
A1 Barneveld - Deventer	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	-1.391	1.391	-0.634	1.565	0.199	0.570	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	0.028	5.000
A1 Apeldoorn - Azelo	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	-1.391	1.391	1.260	-1.913	0.182	0.709	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.466	3.268
N31 Zurich - Harlingen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.000	1.127	-1.265	-1.267	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.244	3.049
A2 Meerenakkerweg	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.675	0.217	0.037	0.129	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.583	1.811
A59 Brug over het Drongelens Kanaal	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.260	-1.913	0.067	0.173	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	-0.455	1.000
N35 Nijverdal-Wierden	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	1.182	-2.456	-0.950	-0.876	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	-0.824	4.005
A2 Leenderheide - Budel	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	0.060	0.068	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.699	8.216
A1 t' Gooi	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	4.589
A27 Gorinchem - Noordeloos	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	0.798	0.933	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.630	0.756
A12 Veenendaal-Ede	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-1.222	0.927	0.469	0.546	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	0.850	8.000
A4 Badhoevedorp-Nieuwe Meer	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	4.666
A1 Watergraafsmeer - Diemen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	1.000
A15 Papendrecht - Sliedrecht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	1.505	-1.400	0.793	1.040	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	1.026	2.170
A15 Papendrecht - Hardinxveld-Giessendam - Gorinchem	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.822	0.947	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.680	2.203
A59 Rosmalen-Geffen	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.667	0.712	0.059	-0.020	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.081	2.567
A16 Rotterdam	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.822	0.947	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.680	9.501
A15 knooppunt Ressen-rijksweg 12	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.765	0.712	-0.445	-0.340	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.217	5.433
RIJKSWEG 14 'VERLENGDE LANDSCHEIDINGSWEG'	0.531	-0.531	-2.449	2.449	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.883	0.712	0.651	0.760	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.151	4.553
N23 Lelystad-Dronten	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-1.436	-1.479	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.945	1.140
N33, Spijk - Eemshaven	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	-1.346	0.181	-1.125	-1.328	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.313	2.668
Rijksweg 69, Eindhoven - Belgische	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-1.477	0.012	-0.453	0.890	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.130	5.351
A2: route 's-Hertogenbosch - Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	0.022	0.136	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.544	3.222
RIJKSWEG 37 EN RIJKSWEG 34 (GEDEELTE HOLSLOOT-EMMEN-ZUID)	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.765	0.712	-1.326	-1.418	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.528	9.764
Rijksweg 73-Zuid	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.667	0.712	-0.862	-0.891	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.034	3.129
rijksweg 31, gedeelte Leeuwarden–Drachten	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.599	0.125	-1.236	-1.292	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.437	3.167
A1 Hoevelaken - Barneveld	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.235	-0.625	-0.484	-0.223	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.749	3.942
A28 Leusden Zuid - A1 Hoevelaken	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.097	-0.012	0.500	0.676	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	1.277	5.296
A12 Woerden - Gouwe	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	1.769	1.837	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	2.670	2.296
A27 Plusstrook Utrecht richting Eemnes	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.097	-0.012	-0.649	-0.862	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.159	1.868

ABT Dataset for Classification Model

Startnotitie	National Road	Provincial Road	Modification Project	New Project	Number of Provin	Rural	Urban	Combination	EHS YE	EHS NO	GDP	3 years Average GDP Growth	RDP (%)	Populatio	Yes Change of Power	No Change of Power	Election Change of Power	CenterLeft Ideolog	CenterRight Ideolo	Center Ideology	Election	c/km2 P1	Below Average
A2 Holendrecht-Oudenrijn	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	0.712	1.433	1.060	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	2.124	0
A2 Maasbracht-Geleen	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-0.932	-0.874	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.529	0
A10 Tweede Coentunnel / A5 Westrandweg	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.235	-0.625	0.476	0.264	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.360	1
A12 Utrecht-Veenendaal	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.031	0.927	-0.691	-0.919	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.039	1
A28 Utrecht-Amersfoort	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	-0.671	-0.823	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	0.251	1
N50 Ramspol-Ens	-1.884	1.884	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	-0.791	-0.639	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.056	0
A2 Holendrecht-Maarssen (A2 Holendrecht-Oudenrijn)	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	1.387	1.205	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	2.466	1
A2 Rondweg Den Bosch	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.599	0.125	0.059	-0.006	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.043	1
A2 St.Joost-Urmond	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-0.932	-0.874	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.529	1
A2 Tangenten Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.477	0.012	0.059	0.007	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.025	0
A7 Rondweg Sneek	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.346	0.181	-1.249	-1.288	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.419	0
A9 Holendrecht-Diemen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	1
A12 Woerden-Gouda	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	1.788	1.866	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	2.783	0
A1 Diemen-Muiderberg & A1/A6 Muiderberg-Almere Stad West	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.653	0.560	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.316	1
A2 Leenderheide-Valkenswaard	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	0.029	0.132	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.558	1
A2/A27 Everdingen-Lunetten	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	-0.655	-0.847	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.223	0
A9 Raasdorp-Badhoevedorp & A9 Velsen-Raasdorp	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	0
A12 Woerden-Oudenrijn	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.407	-0.543	-0.655	-0.847	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.223	1
A28 Zwolle - Meppel	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	-0.709	-0.516	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.257	1
A58 Eindhoven-Oirschot	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	0.037	0.129	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.583	1
N34 Omleiding Ommen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	-1.346	0.181	-0.953	-0.942	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.181	0
A9 Alkmaar-Uitgeest	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	0.414	0.284	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.415	1
A12 Zoetermeer-Zoetermeer centrum (noordbaan)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.980	-1.218	0.760	0.965	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.778	1
A74 Venlo	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.346	0.181	-0.898	-0.879	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.966	0
N9 Koedijk-De Stolpen	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.346	0.552	0.419	0.167	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-0.143	0
N200 Halfweg	-1.884	1.884	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.599	0.012	0.437	0.148	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.212	0
A4 Burgerveen - Leiden	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	-1.477	0.012	0.857	0.833	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.021	1
A4 Dinteloord-Bergen op Zoom	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.667	0.712	0.059	-0.020	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.081	0
N11 Zoeterwoude-Alphen aan den Rijn	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	0.791	0.920	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.583	0
N31 Leeuwarden (Haak)	-1.884	1.884	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.000	1.127	-1.265	-1.267	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.244	1
N33 Assen-Zuidbroek	-1.884	1.884	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.675	0.217	-0.848	-0.935	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.427	1
N61 Hoek-Schoondijke	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.837	1.565	-1.388	-1.476	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.486	0
Kanaalkruising Sluiskil	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	-1.382	-1.469	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.443	1
A28/A1 Knooppunt Hoevelaken	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	0
A7/A8 Purmerend - Zaandam - Coenplein	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.000	1.127	0.467	0.242	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.377	1
A15 Maasvlakte - Vaanplein	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.222	0.552	0.815	0.849	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	0.243	0
A4 Delft-Schiedam	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.235	-0.625	0.798	0.942	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.663	0
A2 passage Maastricht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.235	-0.625	-0.914	-0.872	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.533	0
A9 omlegging Badhoevedorp	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.946	0.725	0.414	0.284	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.415	1
N50 Ens - Emmeloord	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.182	-2.456	-1.427	-1.458	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	-0.762	1
A1 Apeldoorn-Zuid - Beekbergen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.980	-1.218	-0.463	-0.204	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.692	0
N18 Varsseveld-Enschede	-1.884	1.884	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	1.509	0.651	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.255	0
N31 Harlingen (traverse Harlingen)	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.980	-1.218	-1.267	-1.260	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.163	1
N35 Zwolle - Wijthmen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	1.091	-1.961	-0.950	-0.879	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.837	1
A27/A1 Utrecht-Noord - knooppunt Eemnes -aansluiting Bunschoten-Spakenburg	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	1
A27/A12 Ring Utrecht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	-0.671	-0.823	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.251	0
A1/A6/A9 Schiphol-Amsterdam-Almere	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.235	-0.625	1.041	0.546	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.276	0
A12/A15 Ressen - Oudbroeken (ViA15)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	1.152	0.742	-0.482	-0.210	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.715	0
Rijnlandroute (A4 and A44)	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	1.152	0.742	0.777	0.950	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.722	0
A4 Vlietland-N14	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	1.260	-1.913	0.744	1.030	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.972	1
A15/A24 Blankenburgverbinding	0.531	-0.531	-2.449	2.449	-0.517	-1.129	2.014	-0.565	-1.391	1.391	1.182	-2.456	0.729	1.002	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	0.896	0
A76 Ansluiting Nuth	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.980	-1.218	-0.952	-0.885	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.514	1
Zuidasdok	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	1.342	-1.962	0.434	0.385	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.711	1
A7 Zuidelijke Ringweg Groningen Fase 1	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.667	0.125	-1.209	-1.330	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.323	0
A7 Zuidelijke Ringweg Groningen Fase 2	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.980	-1.218	-1.106	-1.315	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.004	1
A27 Lunetten-Hooipolder	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.946	0.725	-0.659	1.066	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	1.599	0
A2 Grathem - Urmond	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	-0.889	-0.872	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.620	0
A4 De Hoek - Prins Clausplein route	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	3.298	2.848	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	2.456	0
A2 Oudenrijn-Deil	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	2.016	0.592	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	1.080	0
A2: Everdingen junction Deil junction (future) Zaltbommel junction Empel junction	0.531	-0.531	0.408	-0.408	3.961	0.886	-0.497	-0.565	0.719	-0.719	-1.477	0.181	3.704	4.058	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	1.605	0
A2: route 's-Hertogenbosch - Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	0.060	0.068	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.699	1
A1: Eemnes - Barneveld	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	-0.634	1.565	0.526	0.592	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	1.080	1
A1 Barneveld - Deventer	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	-1.391	1.391	-0.634	1.565	0.199	0.570	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	0.028	0
A1 Apeldoorn - Azelo	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	-1.391	1.391	1.260	-1.913	0.182	0.709	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	0.466	1
N31 Zurich - Harlingen	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	0.000	1.127	-1.265	-1.267	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-1.244	1
A2 Meerenakkerweg	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.675	0.217	0.037	0.129	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.583	1
A59 Brug over het Drongelens Kanaal	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.260	-1.913	0.067	0.173	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	-0.455	1
N35 Nijverdal-Wierden	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	1.182	-2.456	-0.950	-0.876	2.566	-1.358	-0.531	-0.462	3.018	-1.011	-0.531	-0.824	1
A2 Leenderheide - Budel	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-0.634	1.565	0.060	0.068	-0.390	0.736	-0.531	2.165	-0.331	-1.011	-0.531	-0.699	0
A1 t' Gooi	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	0.719	-0.719	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	0
A27 Gorinchem - Noordeloos	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	0.798	0.933	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.630	1
A12 Veenendaal-Ede	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-1.222	0.927	0.469	0.546	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	0.850	0
A4 Badhoevedorp-Nieuwe Meer	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	0
A1 Watergraafsmeer - Diemen	0.531	-0.531	0.408	-0.408	-0.517	-1.129	2.014	-0.565	-1.391	1.391	0.407	-0.543	0.468	0.273	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.381	1
A15 Papendrecht - Sliedrecht	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	1.505	-1.400	0.793	1.040	-0.390	0.736	-0.531	-0.462	3.018	-1.011	-0.531	1.026	1
A15 Papendrecht - Hardinxveld-Giessendam - Gorinchem	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.822	0.947	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.680	1
A59 Rosmalen-Geffen	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.667	0.712	0.059	-0.020	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.081	1
A16 Rotterdam	0.531	-0.531	-2.449	2.449	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.407	-0.543	0.822	0.947	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	0.680	0
A15 knooppunt Ressen-rijksweg 12	0.531	-0.531	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.765	0.712	-0.445	-0.340	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.217	0
RIJKSWEG 14 'VERLENGDE LANDSCHEIDINGSWEG'	0.531	-0.531	-2.449	2.449	-0.517	-1.129	2.014	-0.565	-1.391	1.391	-1.883	0.712	0.651	0.760	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.151	0
N23 Lelystad-Dronten	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	0.407	-0.543	-1.436	-1.479	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.945	1
N33, Spijk - Eemshaven	-1.884	1.884	0.408	-0.408	-0.517	-1.129	-0.497	1.771	-1.391	1.391	-1.346	0.181	-1.125	-1.328	2.566	-1.358	-0.531	2.165	-0.331	-1.011	-0.531	-1.313	1
Rijksweg 69, Eindhoven - Belgische	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	-1.477	0.012	-0.453	0.890	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	-0.130	0
A2: route 's-Hertogenbosch - Eindhoven	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	1.152	0.742	0.022	0.136	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.544	1
RIJKSWEG 37 EN RIJKSWEG 34 (GEDEELTE HOLSLOOT-EMMEN-ZUID)	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.765	0.712	-1.326	-1.418	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.528	0
Rijksweg 73-Zuid	-1.884	1.884	-2.449	2.449	-0.517	0.886	-0.497	-0.565	-1.391	1.391	-1.667	0.712	-0.862	-0.891	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.034	1
rijksweg 31, gedeelte Leeuwarden–Drachten	-1.884	1.884	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	-1.599	0.125	-1.236	-1.292	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-1.437	1
A1 Hoevelaken - Barneveld	0.531	-0.531	0.408	-0.408	-0.517	0.886	-0.497	-0.565	0.719	-0.719	0.235	-0.625	-0.484	-0.223	-0.390	0.736	-0.531	-0.462	-0.331	0.989	-0.531	-0.749	1
A28 Leusden Zuid - A1 Hoevelaken	0.531	-0.531	0.408	-0.408	1.722	-1.129	-0.497	1.771	0.719	-0.719	0.097	-0.012	0.500	0.676	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	1.277	0
A12 Woerden - Gouwe	0.531	-0.531	0.408	-0.408	1.722	0.886	-0.497	-0.565	0.719	-0.719	0.097	-0.012	1.769	1.837	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	2.670	1
A27 Plusstrook Utrecht richting Eemnes	0.531	-0.531	0.408	-0.408	-0.517	-1.129	-0.497	1.771	0.719	-0.719	0.097	-0.012	-0.649	-0.862	-0.390	-1.358	1.884	-0.462	-0.331	-1.011	1.884	0.159	1