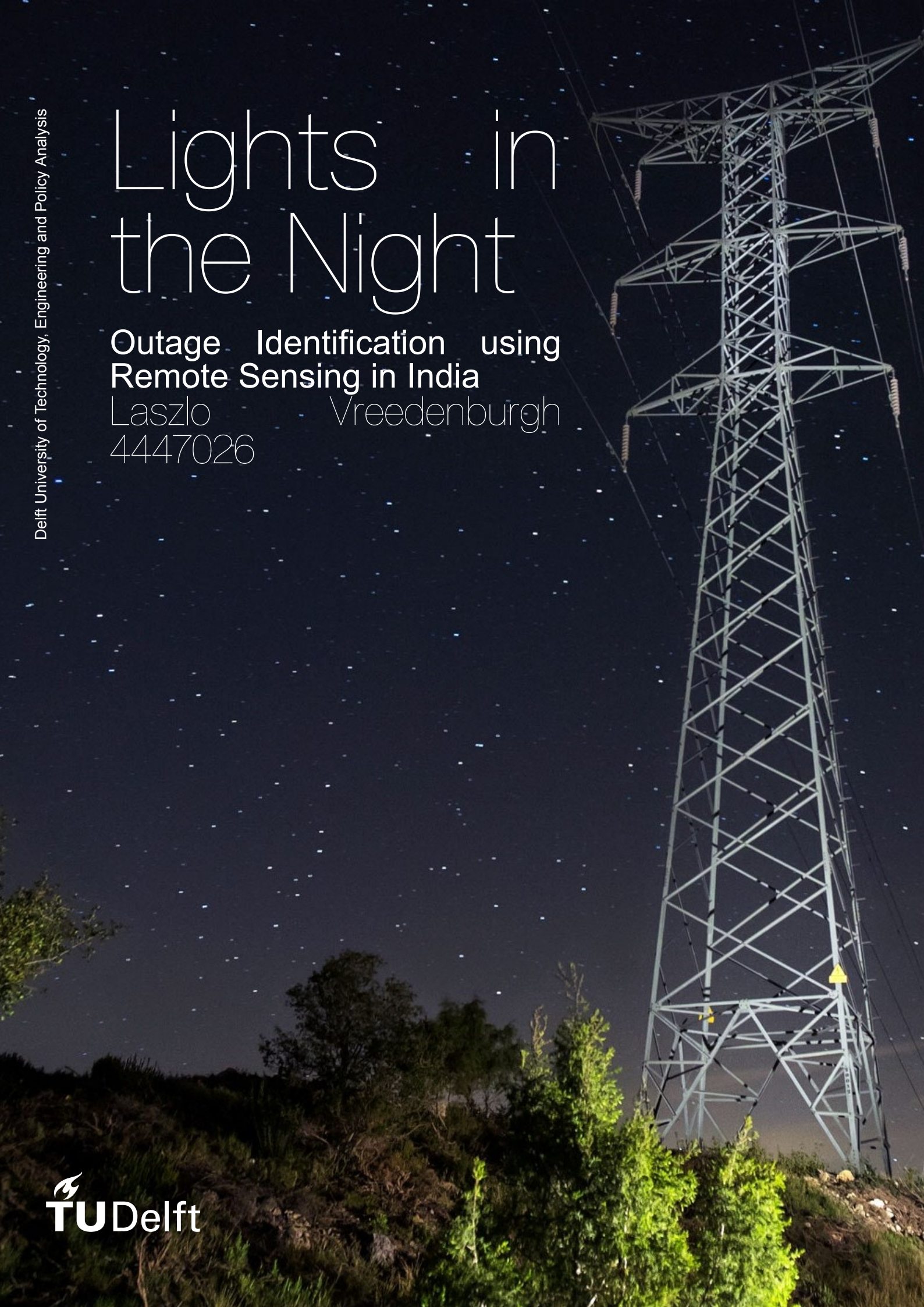


Lights in the Night

Outage Identification using Remote Sensing in India

Laszlo Vreedenburgh
4447026

Delft University of Technology, Engineering and Policy Analysis

TU Delft

Lights in the Night

Outage Identification using Remote Sensing in India

by

Laszlo Vreedenburgh 4447026

to obtain the degree of Master of Science Engineering and Policy Analysis
at the Delft University of Technology,

Student number:	4447026	
Project duration:	March 3, 2022 – June 23, 2022	
Thesis committee:	Dr. N. Goyal,	TU Delft, Supervisor
	Dr. ir. T. Verma,	TU Delft, Supervisor
	Prof. dr. M. E. Warnier,	TU Delft, Chair

TU Delft

Abstract

Electricity is essential in the modern world. Although India is near reaching 100% electrification, for many of its citizens, reliable access to electricity is still a major issue. It is not clear how, when, and where the reliability issues arise. Previous studies have not shed light on electricity outages at a granular level. Furthermore, identifying outages is difficult without proper metering.

This research aims to acquire insight into how electrical outages vary during the nighttime in the Uttar Pradesh, one of the states experiencing the most outages in India. Remote sensing data in the form of nighttime radiance, wind, precipitation, temperature, air quality, land cover, and population are used to identify outages, as they have been identified in the literature to be influencing or correlated with outages. Using a Random Forest (RF) classification algorithm, outages are identified in India for the year 2018. The RF is trained on Electricity Supply Monitoring Initiative (ESMI) real-time electricity household senors. RF is cross-validated using multiple strategies to accurate measure the performance and to ensure no data leakage. Using the resampling technique Synthetic Minority Oversampling technique (SMOTE), the performance of the RF is increased for outage classification. RF is used to classify three classes: Never Access, Normal Access, and Outage. The distribution of these three classes is highly imbalanced, with the Outage class being in the minority.

To further validate the methodology, the spatial and temporal sampling was done during cross-validation. Using this method, 91% of Never Access samples were categorized as such, with a precision of 84%. 69% of Normal Access samples were categorized as such, but with 84% precision. The performance of the outage classification is the worst. Nonetheless, 55% of the Outage samples were categorized as such, although with a precision of just 39%. A map is created for Uttar Pradesh showing the results of the rates for each of the classes in 2018 between 00:00 and 02:00 with a spatial granularity of 0.1° x 0.1°. The map indicate that the electricity usage during the nighttime in Uttar Pradesh is in large parts of the state non-existent. The map be used as a precursor for future field work and help policymaker and researchers identify reliability and fairness issues.

Preface

This thesis was written for the Master of Science Engineering and Policy Analysis program at Delft University of Technology. I've met several great individuals over this project, and I would like to thank them for their contributions to my work. First and foremost, I would want to thank my first supervisor, Dr. Nihit Goyal, for exposing me to this topic and allowing me to complete the thesis in the manner that best suits me. I've learned a lot from you over this process. Thank you very much. I would also like to thank Dr. ir. T. Verma and Prof. dr. M. E. Warnier for their feedback on all the major milestones. I enjoyed the casual and approachable way of the meetings. Furthermore, I would like to thank everyone from the CUSP and ETL thesis circles. The feedback and discussions from these meetings have been very helpful.

Laszlo Vreedenburgh 4447026
Delft, June 2021

Contents

List of Figures

List of Tables

1 Introduction

The problem at the center of this research proposal is introduced in this section. The problem's scientific and societal importance is also addressed, as well as the structure of the proposal.

1.1. Problem Situation

Access to electricity is well recognized within the scope of human rights (Tully, 2006). While the western world is currently focusing on transitioning to renewable energy, India still struggles to provide reliable electricity access to all its citizens (Kennedy et al., 2019, 2020; Phadke et al., 2019). India has made commendable progress in the last decade in the electrification of the country, increasing connectivity to electricity from 76.3% to 97.8% (World Bank, 2019). However, measuring electricity access entails considerably more than merely documenting whether or not houses are connected to the grid. To reach the sustainable development goal of access to affordable, reliable, sustainable, and modern energy for all, efforts to assess progress in energy access are crucial (IEA et al., 2021). For example, Mukherji (2021) reports in The Times of India that only 63% of households are satisfied with the reliability of the electricity access. In rural regions, 50% of the households face daily power outages of 8 hours or more. Agricultural users in most states only get 7-8 hours of electricity. It is important to understand that the average number of hours of power available on a normal day has a significant impact on home satisfaction. Aklin et al. (2016) found that increasing the number of hours per day by 4.8 increases satisfaction almost as much as electrifying a non-electrified home. These results highlight the necessity of moving away from counting electrical connections and instead focusing on improving the quality of the power supply.

Even though India has recently developed a power surplus (Parray & Tongia, 2019), the frequency of outages has not improved at the same rate as the development of electricity generation. As a result, while new capacity installations increased the supply of electricity closer to demand, the necessary expenditures in sufficient distribution infrastructure fell short (Athawale, 2021).

Recently, the Electricity Right of Power Consumers Rules was released by the Ministry of Power with the goal of ensuring a minimum quality of service for the delivery of electricity to end-users (Ministry of Power, 2020). This regulation mandates that state electricity regulatory commissions set clear rules for distribution companies to follow in order to maintain a reliable distribution network 24 hours a day, seven days a week.

1.2. Scientific and Societal Relevance

Determining reliability can be difficult, common methods generally fall short in India. The government does not collect and publish statistics on local electrical delivery, and the data available from utility companies might be deceptive. While household surveys are valuable for assessing household connections, they only provide limited information on service reliability (Aklin et al., 2016). Cross-sectional surveys represent the current situation but often fall short of providing an accurate picture of service reliability over seasons and over longer periods of time (Gibson & Olivia, 2010). Technical reliability measurement, on the other hand, is costly and difficult in remote rural locations. Few governments in the developing world collect and publish statistics on power delivery at the local level, and the

data provided by utilities might be deceptive (Kennedy et al., 2019). Crowdsourcing has proven to be a reasonably effective way for tracking access reliability over time, although it is far from ideal (Kennedy et al., 2020).

For this research, a different method is proposed than in the previous paragraph described methods. Data from remote sensing has the potential to transform social science (Dugoua et al., 2018). Combined with a socioeconomic survey, remote sensing data can be used for a predictive machine learning algorithm applied to energy poverty (Wang et al., 2021). Machine learning has been used in a number of studies to predict power outages during extreme weather (Eskandarpour & Khodaei, 2017; Nateghi et al., 2013; Yang et al., 2020). As well as to predict the risk of failures for components and systems during standard situations (He et al., 2018; Rudin et al., 2012; Wu et al., 2011). Machine learning can also be used to assist in outage planning for maintenance activities (Toubeau et al., 2022). Machine learning presents opportunities to help gain insights into the reliability of the electricity access in India. Which could lessen the need to use untrustworthy surveys or invest in expensive technical measurement equipment.

The Engineering and Policy Analysis program deals with problems that are related to the Sustainable Development Goals (SDGs) that are defined by the United Nations (United Nations, 2021). Ensuring reliable electricity access in India is a goal is in line with the following SDGs:

- SDG 1: No Poverty
- SDG 7: Affordable and Clean Energy
- SDG 9: Industry, Innovation, and Infrastructure
- SDG 10: Reduced Inequalities

This demonstrates the societal importance of the study on India's electricity access. Furthermore, by employing machine learning methods to acquire insight into India's electrical reliability, this project seeks to make a significant scientific contribution to the field.

1.3. Research Objectives

The following research question will be studied: *How do the number of electricity outages vary across one of India's less developed states?* This question is formulated this way, because while the outage problem is a nationwide problem, for this research it is too big of a scope. The focus will primarily lie on a state that less developed, as the need for intervention and improvement is urgent. To help answer this question, a predictive algorithm might be able to provide insight. An algorithm that can estimate the electricity outages in India decreases the can compliment the lacking metering, and in doing so, can help identify reliability and fairness issues.

The goal of this study is to gain an understanding of the relationship between specific characteristics and electricity reliability in India. It will explore in-depth how the use of machine learning algorithms can improve the information gathering of electricity usage in India. To achieve this goal, I aim to approximate the electricity supply on the grid in India. For this, I will conduct a descriptive, empirical (based primarily on real-world observation), quantitative analysis, to provide the machine learning algorithms with relevant and comprehensive data. To help answer the main research question, the following sub-questions are proposed:

1. What is the status of electricity access in India? *Why: It is important to have a clear understanding of the history and current situation surrounding electricity access in India.*
2. How have existing literature used remote sensing and machine learning to measure aspects of the energy system? *Why: To comprehend contemporary remote sensing and machine learning solutions, scientific literature applied to energy systems will provide useful insights.*
3. What are the spatial characteristics that influence (or are correlated with) electricity reliability in India? *Why: To identify which characteristics are important predictors of electricity outages.*

4. To what extent can historical data identify electricity outages in India? *Why: To verify and analyze the results found for the previous research question.*

1.4. Research Flow

In figure 1.1, the research flow of the thesis is presented. The research flow summarizes the thesis into the major phases of research. The main tasks, outputs, and methods are presented in the diagram for each phase and the respective questions. For reference the main question and research questions are:

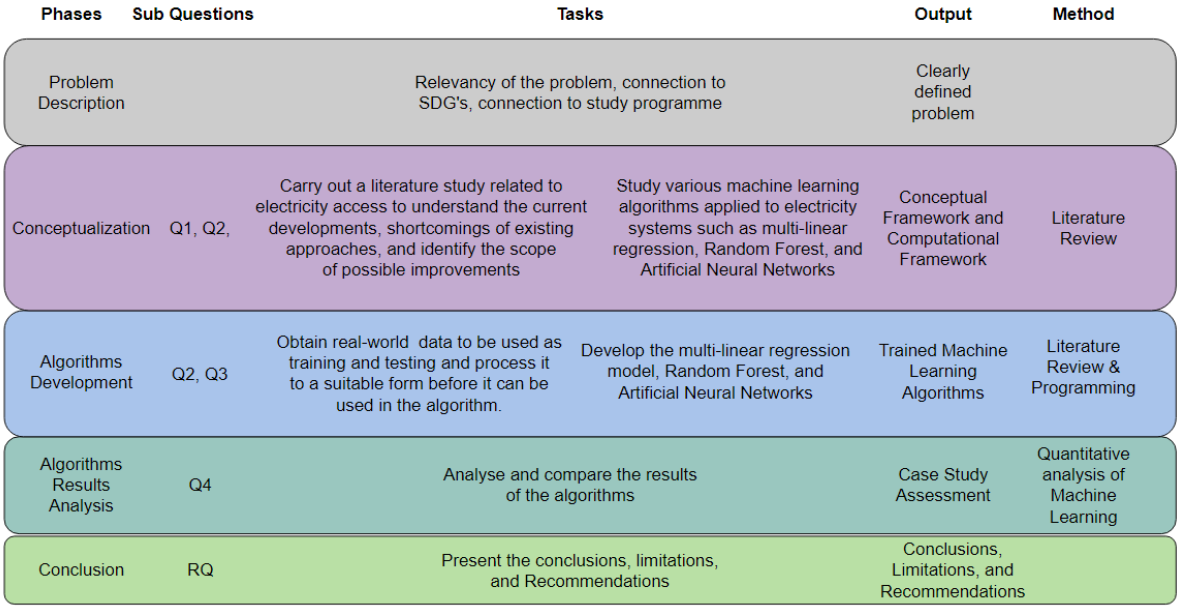


Figure 1.1: The Research Flow Diagram shows how the research is divided in tasks, what their output will be, and how they will be performed.

1.5. Thesis Outline

A literature review defines and describes the important core concepts of the identified issue in chapter 2. As a result, a knowledge gap is identified, and the main research question for the master thesis is proposed. The overall research approach, the sub questions, and the research methods for each sub question are described in chapter 3. Chapter 4 describes the methods and data used in this research. In chapter 4, the results of the classifications are examined, and the strengths and weaknesses of each classification are highlighted. Chapter 5 discusses the implications of the results and limitations of the research, and chapter 6 closes with an overview of the algorithm and the usefulness of the results, a suggestion for the best algorithm to adopt, and ideas for additional study and improvement.

2
Literature Review

Previously, Indian policy and literature on electricity access centered on the electrification of households. (Pelz, Aklin, et al., 2021) However, having access to the grid just through physical connections is inadequate. Only when the electricity supply is reliable and sufficient is it beneficial. People who do not have reliable access to electricity are forced to rely on polluting resources such as kerosene to meet their daily energy needs. This not only results in the depletion of scarce resources but also in the generation of greenhouse gases that threaten our environment (Rao, 2013; Sedai, Nepal, et al., 2021).

In general, there is a strong case to be made for more research on the effects and moderators of reliable, inexpensive, and modern electricity (Bayer et al., 2020; Jeuland et al., 2021; Peters & Sievert, 2016; Riva et al., 2018). However, lack of electricity access is a part of the bigger problem of energy poverty. So, we first need to broadly explore what energy exactly is.

2.1. Energy Poverty

There are many various definitions and perspectives of energy poverty, but they all pertain to an insufficient amount of energy use to fulfill certain fundamental necessities. A comprehensive definition for energy poverty defined by A. K. Reddy et al. (2000) as “the absence of sufficient choice in accessing adequate, affordable, reliable, high-quality, safe and environmentally benign energy services to support economic and human development”. This definition is used by many authors, as it contains a variety of intriguing components and intricacies (de Groot et al., 2017; González-Eguino, 2015; Khanna et al., 2019; Pelz et al., 2018; Samarakoon, 2019).

First, the definition relates energy poverty to the absence of choice. It is not so much a question of attaining a specific amount of wealth or energy per capita as it is of not being barred from those possibilities that allow people to select and get welfare in its fullest meaning (Acemoglu & Robinson, 2013; Sen, 2014). Without access to energy, people may be deprived not only of basic services such as lighting, house heating, and cooking, but also of other aspects essential for individual and communal development such as access to health, education, information, and political involvement (Pelz et al., 2018). As can be observed, a lack of competence or choice can impact aspects that are needed for involvement in and management of institutions. When they do not serve the common interest, true development is unlikely.

Second, the term emphasizes addressing demand for energy services. Although it may appear apparent, it is worth noting that the aim is not so much energy consumption as it is the supply of energy services from diverse energy sources (Gupta et al., 2020; Pelz & Urpelainen, 2020). Primary energy sources (coal, oil, gas, biomass, and so on) are processed, and energy is stored and distributed through various energy ”vectors” (heat, electricity, and solid, liquid, or gaseous fuel) to provide the various energy services that are truly required: cooking, heating, cooling, lighting, transportation, work, and access to information and communication technologies (González-Eguino, 2015). The composition of major energy sources and energy vectors utilized may vary greatly depending on geographical factors and the energy policy enacted in a specific nation, but the energy services required are generally similar everywhere.

Third, the definition defines certain desired features of the energy sources used to obtain energy services. These sources must be adequate, that is, they must be appropriate for the geographical features, knowledge base, and culture of the region (A. K. Reddy et al., 2000). It is generally recognized that development assistance initiatives might fail if they merely try to replicate the deployment of the same technology in numerous regions without taking into consideration the unique characteristics of each region or community (Derks & Romijn, 2019; Lu et al., 2020; Ulsrud et al., 2018).

Energy sources must also be affordable, that is, as inexpensive as feasible in comparison to the available alternatives. Normally, as average family income levels improve, fuel sources such as biomass are gradually replaced by sources such as kerosene, oil, and, eventually, electricity, which is the cleanest and most adaptable energy vector of all. This is described by the theory of the Energy Ladder, as income rises, low-quality fuels are substituted by higher-quality, more adaptable fuels (van der Kroon et al., 2013). One key limitation to Energy Ladder theory is that low-quality energy sources are not necessarily the cheapest as commonly assumed, all too frequently, they are just the only alternative, which again relates to the absence of sufficient choice.

Next, energy must also be reliable. Because of the absence of reliability in energy services, which limits productive activities, reliability might influence an individual's choice of fuel to assure energy security. Because modern energy sources are unreliable in many developing countries, many households employ a range of fuels to maintain their energy security (B. S. Reddy, 2015). Renewable energy advocates must be alert to concerns of energy reliability, which is a key concern in developing countries. This is especially significant because it pushes policymakers to properly distinguish preferences for renewable energy sources and their associated influence on local vs global pollution from preferences for greater reliability (Bakkensen & Schuler, 2020).

Finally, the definition describes that the energy sources should be safe, implying that the energy sources should be unlikely to threaten human health, and environmentally benign, that is, they should not jeopardize future generations (de Groot et al., 2017; Samarakoon, 2019). It is critical that technical solutions to eliminate energy poverty include the effects of climate change and the environment in order to sustain growth in the future. Furthermore, as stated in the definition, the objective of energy consumption is "to support economic and human development," therefore the sheer existence of energy resources and the economic activity linked with their extraction does not ensure general or energy development (González-Eguino, 2015).

2.2. The Importance of Electricity Access

In table 2.1 the definitions of basic energy services are given for the three most used multidimensional energy poverty measurement frameworks are given (Pelz et al., 2018). The definitions reflect widespread agreement that energy poverty can result in a variety of ways, they also highlight the historical discrepancy on a prescriptive set of basic energy services, which has had consequences for the development of a universal SDG7 measurement. However, most of the basic energy services across the different frameworks need electricity or electricity can be used as an energy source. Reliable electricity access is clearly instrumental in fighting energy poverty.

The effects of a lack of access to electricity in India are felt throughout many aspects of life. Electricity access is a key driver of economic progress. Access and supply of electricity have a significant impact on household revenue (Rao, 2013). According to Chakravorty et al. (2014), a grid connection raises rural households' non-agricultural income by around 9% over the study period. A grid connection combined with fewer outages and more hours per day of electricity, on the other hand, increase non-agricultural incomes by roughly 28.6%.

The effects of a shortage of electricity are not only economical, shortages are linked to a considerable and significant reduction in the number of treatments, as well as the quality of the treatments. Where women's healthcare access is disproportionally worse than men's (Shastry & Rai, 2021). Dependable electricity increases women's standing in comparison to males overall, specifically by increas-

Table 2.1: Basic Energy Service Definitions in Frameworks

Year	Source	Basic Energy Service Definition
2012	(Practical Action, 2012)	Cooking and water heating, lighting, information and communications, space heating and cooling, and earning a living
2012	(Nussbaumer et al., 2012)	Cooking, lighting, services provided by means of household appliances, communication and entertainment
2015	(Bhatia & Angelou, 2015)	Cooking, lighting, entertainment and communications, space cooling and heating, refrigeration, mechanical loads, and product heating

ing work options and reducing time spent at home (Sedai, Vasudevan, et al., 2021).

Having more electricity hours increases the chance of having bathroom facilities, lowers interior pollution, provides piped water, and provides artificial lighting. For both men and women, longer electricity hours mean more agricultural income, longer labor hours, more businesses, and less fuel collection. In rural regions, reliable electricity enhances the possibility of borrowing for productive purposes and lowers the risk of poverty (Richmond et al., 2020; Sedai, Nepal, et al., 2021). Electricity availability has been proven to be a highly important predictor of powering appliances for longer periods of time in all categories, but notably in the lighting and cooling categories, where items are more easily attainable (Richmond et al., 2020).

2.3. Electricity Access Defined

The multidimensionality of electricity access is now generally acknowledged, and current measures reflect increasingly comprehensive characteristics such as availability, reliability, affordability, quality, and safety, in addition to the continued predominance of energy consumption. Despite this, there is currently no agreement on which dimensions and thresholds are required for measuring electricity access and monitoring intended development results.

2.3.1. Energy Supply Index

The Energy Supply Index (ESI) was the first ordinal framework to provide energy supply parameters, mostly on a qualitative basis (Practical Action, 2012). The ESI is divided into three categories: home fuels (cooking), electricity, and mechanical power, each with six quality levels (0–5). Table 2.2 shows the residential electricity supply measurement component. Although Practical Action (2012) mentions the aspects of reliability, affordability, and adequacy, they are not expressly defined, necessitating subjective interpretation throughout data collection and analysis. While the lack of clarity in the definition of chosen parameters allows for some contextual flexibility, it makes it difficult to compare electricity access across time and between nations, and it has thus far not gained widespread use.

2.3.2. Multidimensional Energy Poverty Index

The Multidimensional Energy Poverty Index (MEPI) defines basic household energy services as ”cooking, lighting, services provided by means of household appliances, communication and entertainment” (Nussbaumer et al., 2012). It uses proxy indicators derived from widely accessible Demographic and Health Survey (DHS) data (The DHS Program, n.d.) to assess access to these resources, shown in table 2.3. It is made up of five dimensions that reflect basic energy services and six indicators. The use of proxy measures decreases data collecting costs greatly, but also limits the MEPI to solely measuring particular appliance ownership and physical access to specific power sources, as measured by the DHS surveys. As a result, as the authors point out, these proxy measures may not always be useful

Table 2.2: ESI household electricity supply index (Practical Action, 2012)

Level	Quality of supply
0	No access to electricity at all
1	Access to third party battery charging only
2	Access to stand-alone electrical appliance
3	Own limited power access for multiple home applications
4	Poor quality and/or intermittent AC connection
5	Reliable AC connection for all uses

in a particular setting. This is due to the fact that customer preferences and cultural norms may have a substantial influence on energy service demands, equipment utilized to meet those needs, and, as a result, energy consumption and delivery costs (Nussbaumer et al., 2012). For example, the proxy measure for lighting suggests that a household's connection to an electric power carrier is adequate to guarantee that the lighting services required in that context are reliable, affordable, and available. The use of the incidence of appliance ownership as a proxy measure suggests that ownership entails the capacity to operate the device reliably and affordably, hence meeting desired energy service demands. The MEPI has a distinct benefit in that it makes use of existing statistics to assess access to a variety of energy services. Its failure to capture service features such as availability, reliability, and affordability, on the other hand, is a flaw.

Table 2.3: Dimensions and cut-offs for individual variables, including relative weights (in parentheses) (Nussbaumer et al., 2012)

Dimension	Indicator (weight)	Variable	Deprivation cut-off (poor if...)
Cooking	Modern cooking fuel (0.2)	Type of cooking fuel	Use any fuel beside electricity, LPG, kerosene, natural gas, or biogas
	Indoor polluion	Food cooked on stove or open fire (no hood/ chimney) if using any fuel beside electricity, LPG, natural gas, or biogas	True
Lighting	Electricity access (0.2)	Has electricity access	False
Services provided by means of household appliances	Household appliance ownership (0.13)	Has a fridge	False
Entertainment/ education	Entertainment/education appliance ownership (0.13)	Has a radio OR television	False
Communication	Telecommunication means (0.13)	Has a phone land line OR a mobile phone	False

2.3.3. Multi-tier Framework for Measuring Energy Access

The World Bank's Multi-tier Framework for Measuring Energy Access (MTF) connects the definition of energy services with a selection of common appliances and ultimately specified dimensions and thresholds organized into ordinal measurement matrices (Bhatia & Angelou, 2015). Through its incorporation in Sustainable Energy for All (SE4ALL) Country Action Agendas, the MTF has already been utilized in different national electrification strategies (Pelz & Urpelainen, 2020). The MTF provides three electricity access measurement sub-frameworks: access to electricity supply, access to electricity-powered services, and access to electricity consumption. The framework for measuring electricity supply is almost

the only one that is presently used for measurement and policy development. The criteria for seven dimensions are defined across six levels (Tier 0–5) in every one of the matrices for energy supply or access to residential energy service. Table 2.4 shows selected parameters for residential electricity supply as specified by the MTF. For the goal of defining and assessing electricity access, eight key attributes have been chosen: capacity, affordability, availability, reliability, quality, health and safety, legality, and convenience. The definitions of these attributes are described in Appendix A.

Table 2.4: Multi-tier Matrix for Measuring Access to Household Electricity Supply (Bhatia & Angelou, 2015)

Multi-tier Matrix for Measuring Access to Household Electricity Supply

ATTRIBUTES			TIER 0	TIER 1	TIER 2	TIER 3	TIER 4	TIER 5
	1. Peak Capacity	Power capacity ratings[28] (in W or daily Wh)		Min 3 W	Min 50 W	Min 200 W	Min 800 W	Min 2 kW
				Min 12 Wh	Min 200 Wh	Min 1.0 kWh	Min 3.4 kWh	Min 8.2 kWh
		OR Services		Lighting of 1,000 lmhr/ day	Electrical lighting, air circulation, television, and phone charging are possible			
	2. Availability (Duration)	Hours per day		Min 4 hrs	Min 4 hrs	Min 8 hrs	Min 16 hrs	Min 23 hrs
		Hours per evening		Min 1 hr	Min 2 hrs	Min 3 hrs	Min 4 hrs	Min 4 hrs
	3. Reliability						Max 14 disruptions per week	Max 3 disruptions per week of total duration <2 hrs
	4. Quality						Voltage problems do not affect the use of desired appliances	
	5. Afford-ability					Cost of a standard consumption package of 365 kWh/year < 5% of household income		
	6. Legality						Bill is paid to the utility, pre-paid card seller, or authorized representative	
	7. Health & Safety						Absence of past accidents and perception of high risk in the future	

Respondent surveys have struggled to capture normative criteria for reliability, legality, and quality characteristics, and have struggled to achieve consensus due to the broad range of politically and socially acceptable degrees of supply disruption in diverse geographical settings (Aklin et al., 2016; Chaurey et al., 2004; Onyeji et al., 2012). Health and safety are also difficult to adequately capture, as it is often dependent on subjective respondent recollection of earlier incidents, which may not fully reflect the true nature of the current danger to health and safety owing to energy supply, particularly in the case of fuel stacking (Tait, 2017).

Finally, despite the fact that affordability is a crucial decision consideration in most energy access activities, it is not examined below Tier 3 and is only rudimentary described above. Asset or wealth

indices may be a better baseline for gauging affordability in many impoverished agricultural areas than income or expenditures, according to (Pelz et al., 2018). The dimension in its current form does not reflect the true cost of energy services (including appliance acquisition) at higher tiers of energy supply, which can have a serious influence on appliance purchase and thus access to modern energy services, nor does it capture the price of energy services at lower tiers of energy supply in various environmental, cultural, and economic contexts.

Despite this general critique, it must be emphasized that the MTF's present dimensions and thresholds have various degrees of functioning. A normative set of characteristics and thresholds is required at the highest global monitoring level to facilitate meaningful analysis and comparison of the global development of SDG 7 across time. An examination of individual dimensions of the electricity supply, on the other hand, quickly reveals that the MTF's complexity of dimensions may pose methodological challenges throughout their application, such as measuring reliability, legality, and quality characteristics. Efforts to better understand the elements that influence satisfaction with electricity supply, on the other hand, have highlighted the contextual nature of the perceived significance of household electricity supply dimensions evaluated by the MTF (Aklin et al., 2016). Therefore, a thorough examination of relevant dimensions and thresholds is required for use in global tracking as well as national or regional energy planning.

2.3.4. MTF CEEW

Although a multidimensional approach to measuring electricity access has not been formally accepted into India's national electrification policy, the announcement of complete household electrification and shift in focus towards supply challenges point towards this as a logical next step. A modification of the MTF developed by the Council on Energy, Environment and Water (CEEW) has been trialed in a representative household panel survey targeting rural areas in six contiguous states across India (Jain et al., 2015; Jain et al., 2018). The CEEW MTF follows the same principles of the global MTF framework for measuring electricity supply, describing access to electrical energy services ranked by appliance power ratings in the Indian context as shown in table 2.5, with several modifications to suit the local context. For instance, the reliability measure differs from the MTF. The reason CEEW MTF uses complete black-out days as a measure for reliability instead of any outage, because unplanned outages are widespread in rural India.

Table 2.5: MTF adaption by CEEW (Jain et al., 2015)

Tier / Dimension	Tier 0	Tier 1	Tier 2	Tier 3
Capacity	No electricity	Lighting + Basic entertainment / communication (Radio/ Mobile) (~1-50W)	Lighting + Air circulation + entertainment / communication (TV/ Computer) (~50-500W)	Tier 2 services + Medium to Heavy loads (>500W)
Duration	<4hrs	>=4hrs and <8hrs	>=8hrs and <20hrs	>=20hrs
Reliability (Black-out Days)	5 or more days	2-4 days	1 day	0
Quality*	$N_H > 3$; $N_L > 6$	N_H = 0-3; N_L = 0-6	N_H = 0-1; N_L = 0-3	$N_H + N_L = 0$
Affordability	Unaffordable		Affordable	
Legality	Illegal		Legal	

*N_H is number of high voltage days in a month causing appliance damage; N_L is number of low voltage days in a month limiting appliance usage.

NOTE: *For dimensions where the categories span multiple tiers, only the higher tier values apply. For example, affordability can only be categorised as Tier 1 or Tier 3. The same is the case for legality.*

2.4. Electricity Access in India

In 1995, judicial activism resulted in the 'right to electricity' being interpreted as a basic right under the 'right to life and liberty' as provided in Article 21 of the Indian Constitution (Palit & Bandyopadhyay, 2017). This was continued by a Chief Ministers' Conference in 1996 and the final creation of the Common Minimum National Action Plan for Power. Reforms to enhance the power distribution sector developed at the beginning of the 2000s, and rural electrification began to be recognized as a major driver of rural development. So, what changes have these decisions brought on for electricity access in India the following years?

2.4.1. Electricity access in India 2000-2010

Banerjee et al. (2015) compare National Sample Survey (NSS) data from 2000 with 2010. They found that India has made significant progress in providing all of its residents with a connection to electricity. Between 2000 and 2010, approximately 283 million people were newly connected, far exceeding the natural population increase. By 2010, the country's total electrified population reached 881 million. Despite this great accomplishment, 311 million people remain without electricity, primarily in impoverished rural homes. By the year 2000, more than 90% of homes in Goa, Delhi, Himachal Pradesh, and Punjab had access to electricity. Uttarakhand and Jharkhand, the two states with the greatest rates of electricity adoption during the decade, began with very low electrification percentages in 2000. Some states with adoption rates below 90% in 2000 did badly during the decade, notably Chhattisgarh, Mizoram, and Uttar Pradesh, with rate increases ranging from 4 to 10%, see figure 2.1. Increase in electricity adoption vary considerably between states during the decade, particularly when population growth was taken into account. Access expansion has lagged behind population growth in the major states of Uttar Pradesh and Bihar. Uttar Pradesh made the most progress in the beginning of the decade, however fell behind in the second half. Andhra Pradesh and West Bengal saw the greatest absolute gains in population with connection to electricity. For the whole of India, there was a rise in electricity access from 59% to 74%. However, using Census of India 2011 data rather than Nation Sample Survey data shows a lower access percentage of 67% (Das & Mistri, 2013).

In addition, there is a data gap in accurately estimating the number of electrified houses. According to field research, there are a substantial number of families who rely on privately run solar systems/micro-grids, utilize diesel generators, or have an illegal connection, and many of these households do not have an electricity access from the grid. It is unclear if these households, who have access to electricity in certain manner in their houses, are termed unelectrified or electrified. In other words, the real number of electrified households may be greater than what the Census or NSS numbers show (Palit & Bandyopadhyay, 2017).

Unreliable power supply jeopardizes India's huge investments in rural electrification, as well as the considerable advantages it provides to both individual homes and the country as a whole. Based on India Human Development Survey 2005 (Desai et al., 2005), Banerjee et al. (2015) show that the issue of electricity reliability in rural India is not restricted to a few areas. Only approximately 7% of rural families with access to electricity report no power disruptions. One-fifth of families with electricity have outages lasting up to four hours per day, and the same amount has inconsistent electricity supply for the most of the day. Bihar and Uttar Pradesh, the two states with the lowest village coverage and household adoption, have the highest average daily outages, see figure 2.2.

2.4.2. Current status of electricity access in India

In 2020 CEEW published their report on electricity access based on their India Residential Energy Survey (IRES) (Agrawal et al., 2020). Their research reveals that, while not ubiquitous, residential electricity in India is approaching saturation levels. Around 96.7% of Indian homes are grid-connected, with 0.33% relying only on off-grid options such as solar home systems, solar mini-grids, battery storage, and diesel generators. Overall, 6.8% of homes have non-grid solutions, with 95% using them as a backup source alongside grid energy and the rest relying solely on them. Eight states have reached 100% electrification, shown in 2.3. Still, 4 percent or more of homes in Chhattisgarh, Haryana, Rajasthan, and Uttar Pradesh do not have an electricity connection. More than a third of the unconnected

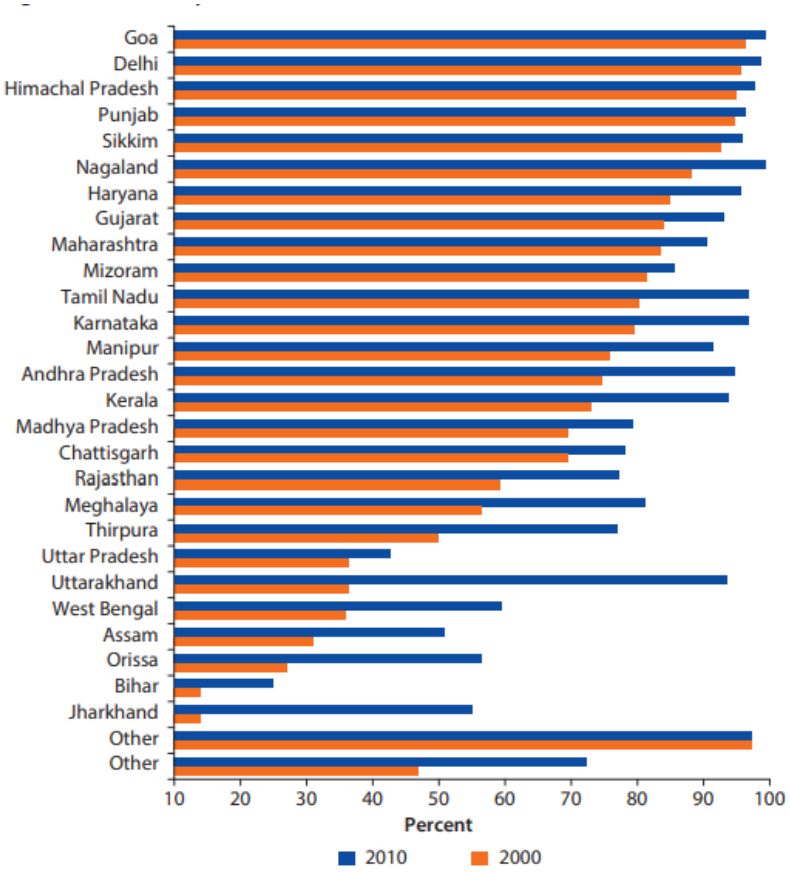


Figure 2.1: Electricity connectivity percentages across Indian states in 2000 and 2010 (Banerjee et al., 2015)

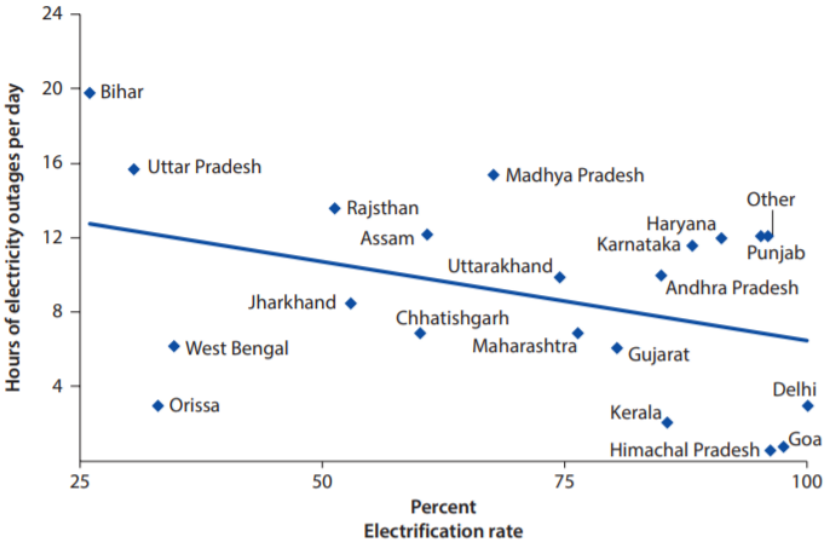


Figure 2.2: Electricity outages for a few selected states in 2005 (Banerjee et al., 2015)

homes are in Uttar Pradesh alone. Future attempts to close the electricity gap must concentrate on these states.

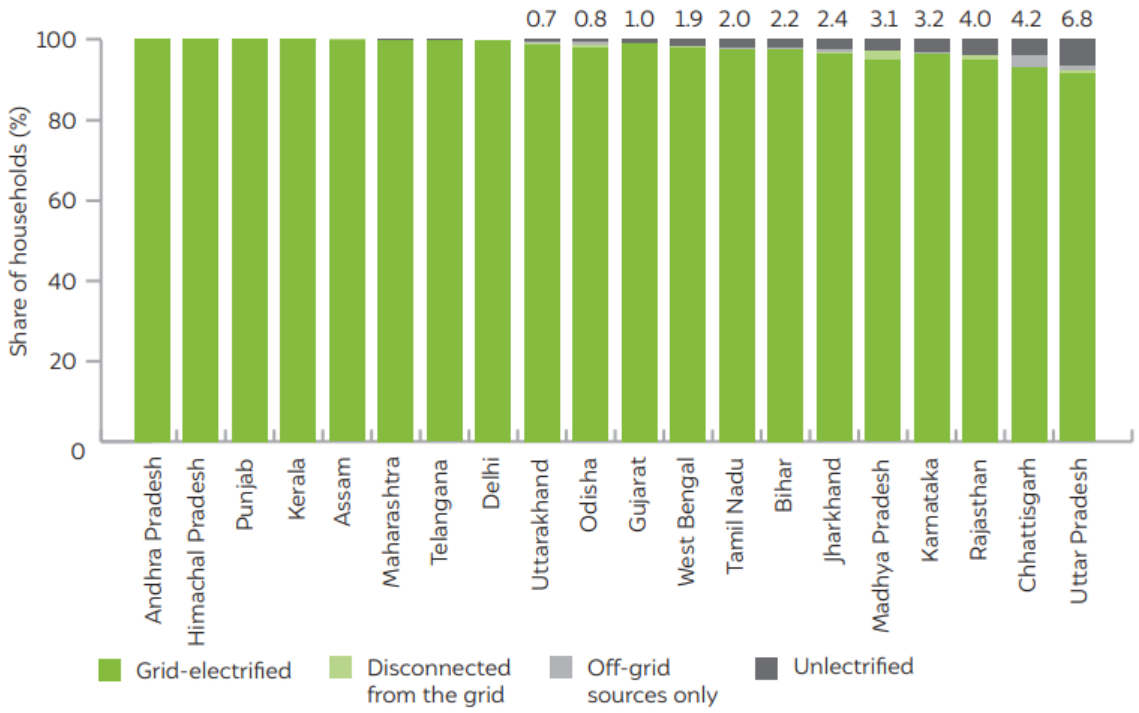


Figure 2.3: Electricity connectivity percentages across Indian states in 2020 (Agrawal et al., 2020)

According to the same study, Indian homes receive an average of 20.6 hours of power every day (Agrawal et al., 2020). Because the study was performed during the winter months, when power consumption is at its lowest, the estimates of supply length and quality are likely to be optimistic. Supply-related difficulties are generally more obvious during the summer, when residential power consumption is at its greatest. Still, the study find that rural homes receive supplies for just 19.9 hours per day, whereas urban households receive supply for 22.3 hours per day. However, national averages conceal the considerable variance in supply offered throughout India's states. In figure 2.4 the averages of availability of electricity are shown for rural and urban areas across the states. Delhi, Kerala, Gujarat, and Tamil Nadu are the best performing states in terms of supply availability, with both urban and rural households receiving power for 23 hours or more each day on average. Homes in Uttar Pradesh, Jharkhand, Haryana, Assam, and Bihar, on the other hand, have the longest electricity disruptions, with rural homes in these states experiencing six or more hours of daily power outages on average. Power outages are highly unpredictable in most villages and towns, and residents must accept many disruptions on a daily basis. In India, about two-thirds of rural homes and two-fifths of urban homes experience power outages at least once each day. According to the survey, just 9% of rural grid users and 23% of urban grid users get near-constant supply. Rural families are particularly impacted by power outages, which frequently occur in the nighttime hours, with half of them experiencing daily outages between 18:00 and midnight. However, most urban families get an uninterrupted supply throughout the late hours.

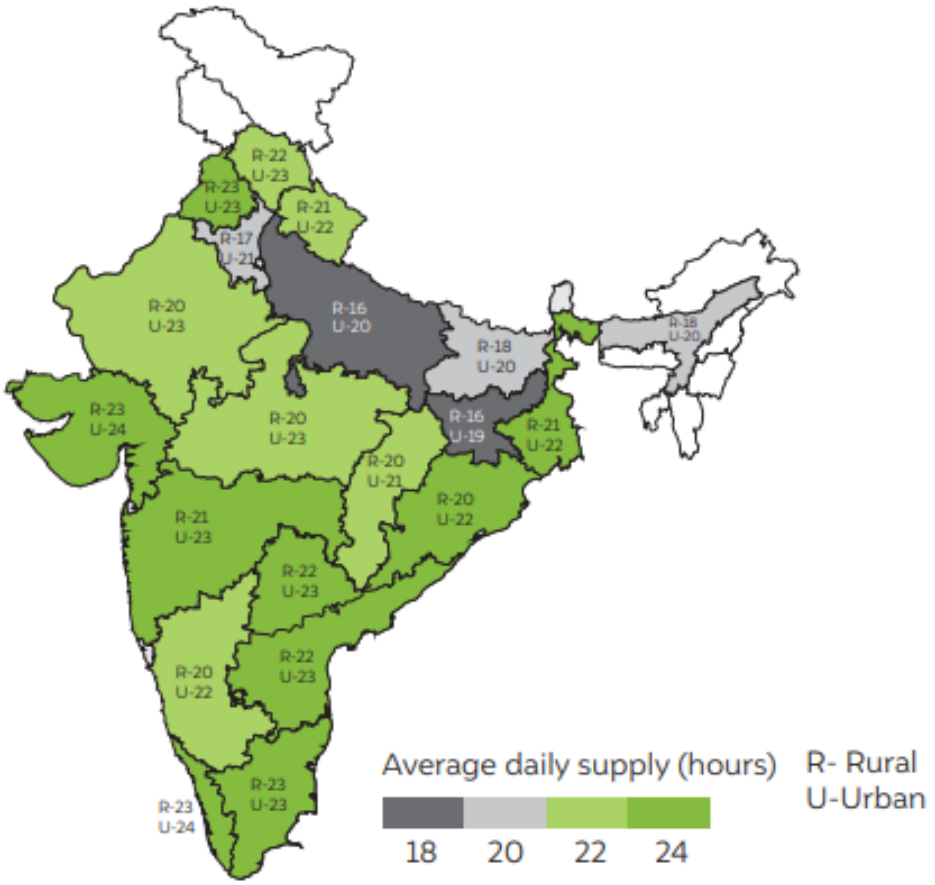


Figure 2.4: Average daily supply across Indian states in 2020 (Agrawal et al., 2020)

2.5. Electricity Access Barriers Identified in Literature

As India approaches 100% home electrification, it is important to grasp the fundamental challenges encountered by the actual implementors (World Bank, 2019). Several barriers to electricity consumption in India have been identified in the literature. Although the articles identify many barriers to electricity reliability, there are three core concepts that may be applied to the barriers identified:

- Technical: Barriers with a technical nature, such as faulty connections or metering.
- Pricing: Barriers that concern the price of the electricity.
- Social: Barriers that involve social aspects of Indian society

In table 2.6 [1], the studies are categorized.

2.5.1. Technical

It is well documented that the quality of the electricity is an important factor for electricity usage and customer satisfaction. The frequent outages discourage households from using electricity (Aklin et al., 2021; Kemmler, 2007). Aklin et al. (2016) discovered a substantial and almost linear relationship between hours of supply and satisfaction, this variable should be highlighted in research of electricity access. Customers are dissatisfied owing to the erratic power supply and long-duration power outages, leading to insufficiency, and the prioritization of industrial districts for energy delivery, resulting

[1] ACCESS: Access to Clean Cooking Energy and Electricity – Survey of States, CEA: Central Electricity Authority, NSSO: National Sample Survey Organisation, CSO: Central Statistical Organization, TERI: The Energy and Resources Institute, SPI-ISEP: Smart Power India and the Initiative for Sustainable Energy Policy

Table 2.6: Studies that identified barriers

Study	Scope	Barrier	Data source
(Aklin et al., 2016)	Northern, central and eastern India	Technical	ACCESS
(Aklin et al., 2021)	Northern, central and eastern India	Technical	ACCESS
(Dugoua et al., 2017)	Northern and eastern India	Pricing, Social	Census of India, ACCESS
(Joseph, 2010)	India	Technical, Pricing	CEA
(Kemmler, 2007)	India	Technical, Pricing, Social	NSSO
(Khan et al., 2014)	India	Technical	Ministry of Power
(Kulkarni & Kulkarni, 2020)	India (rural)	Technical, Pricing	SPI-ISEP
(A. Kumar, 2018)	Bihar, India	Pricing, Social	Home Tours, Interviews, Group Discussions
(Pelz, Aklin, et al., 2021)	North India (rural)	Pricing	SPI-ISEP REDI dataset
(Pelz, Chindarkar, et al., 2021)	Uttar Pradesh, Bihar, West Bengal, Jharkhand, Odisha and Madhya Pradesh	Social	ACCESS
(Rajkumari, 2020)	Karnataka	Technical, pricing	CSO, CEA TERI
(Shrimali & Sen, 2020)	India	Technical, Pricing, Social	Semi-structured interviews
(Urpelainen, 2016)	Uttar Pradesh, India	Pricing, Social	Household Surveys
(Wong et al., 2021)	Uttar Pradesh, India	Social	Household Surveys

in load shedding in rural areas (Kulkarni & Kulkarni, 2020). However, industrial consumers have also had issues with outages. Joseph (2010) and Shrimali and Sen (2020) found Industrial users' choice to leave the state-run system is a direct result of the low-quality electricity they received from the state utilities. The significant aggregate technical and commercial losses, as well as the disparity between the average cost of supply and average revenue collected, are examples of issues faced. These findings suggest that novel methods for not only operations and maintenance, but also metering, pricing, and connecting, may be necessary.

The current metering, billing, and collection systems have progressed over time, but they still lack visibility, automated analysis, and are prone to human mistake (Khan et al., 2014; Shrimali & Sen, 2020). In contrast to the findings in section 2.2, Rajkumari (2020) found that there is no causal relationship between electricity consumption and economic development in Karnataka for total, agricultural, and industrial consumption. However, Rajkumari claims that this is due to erroneous agricultural consumption figures, a greater reliance on captive power, and insufficient grid supply.

2.5.2. Pricing

Every home has to pay a set cost due to unmetered connections. As a result, poorer households are forced to pay for more electricity than they really consume (Kulkarni & Kulkarni, 2020). Lack of metering, billing inefficiency and inconsistency, high connection prices, and the availability of cheaper

alternatives all contribute to the poor's inability to adopt electricity. Without a subsidy or other assistance mechanism, it appears that the poorest households will be unable to pay even basic electricity services (Urpelainen, 2016). Even though India's electricity is extensively subsidized, some sections of the country are experiencing affordability concerns. Joseph (2010) emphasized that authorities must guarantee that producers do not participate in cherry-picking, in which only the best customers, those who are most likely to pay for power on a constant basis, receive electricity. If this is the case, not only would price rises affect agricultural customers who are accustomed to getting discounted power, but reforms may also leave rural consumers worse off than they were before.

To compensate for the unreliability of power supply, almost 70% of grid-connected households utilize kerosene as a backup lighting source (Banerjee et al., 2015). In most cases, regardless of the quantity of electricity utilized, these families must pay a mandated minimum payment for electricity service. Furthermore, they spend around Rs. 26 ($0.6) per month on kerosene for lighting, which is comparable to purchasing 10 kilowatt-hours of electricity at an indicative cost of Rs. 2.5 ($0.06) per kilowatt-hour. The much higher amounts that these families pay to cover their monthly lighting demands result in revenue loss for the state electricity corporations. The amount of money people have to spend on alternative fuels for residential lighting is closely tied to the reliability of power supply. In essence, these households must buy two electricity sources in order to light their dwellings.

The claimed power use appears to be unrelated to supply quality, according to Pelz, Aklin, et al. (2021). Their findings suggest that consumption patterns are unlikely to be linked to grid supply reliability issues. The distribution of business types in their sample hints at the larger issue, since manufacturing businesses, which are normally linked with productive power consumption, are in the minority compared to retail businesses. This clearly implies that additional hurdles to productive energy usage in rural parts of the states in issue, aside from supply reliability, demand significantly more attention.

2.5.3. Social

Theft is a key barrier to securing safe, proper access to electricity in the electrical industry. Social norms concerning thieving activities may be an understudied aspect that stifles cooperative conduct, influencing the efficacy of suggested policy solutions directly (Wong et al., 2021). The seeming unwillingness to prosecute thieves may reflect the vicious loop of mistrust that exists between utility companies and citizens. In their research, Urpelainen (2016) discovered that respondents did not trust local enterprises and favored government leadership in rural electrification.

Traditional social structures and hierarchies continue to impact rural development processes in India, particularly among India's marginalized caste-based and tribal populations, who are constitutionally recognized as Scheduled Castes and Scheduled Tribes (SC&ST) (Kemmler, 2007; A. Kumar, 2018). In rural areas, Dugoua et al. (2017) discovered a negative association between the proportion of SC&ST families and the chance of village electricity. Furthermore, in terms of daily supply hours and monthly outage days, SC&ST electricity supply improved less (Pelz, Chindarkar, et al., 2021).

2.6. Spatial analysis

Implementing initiatives or policies to alleviate electricity outages necessitates accurate forecasting of where the electricity outages exist. Dubois (2012) plainly states that effective policy measures rely on accurate targeting and identification of program participants, which presents a variety of challenges and expenses. Among other things, there is a strong information barrier in identifying the specific homes that may benefit from the initiatives. Given the complex nature of electricity access, relying on a single indicator to correctly reflect electricity access conditions is impossible. To identify energy-poor homes for policy interventions, information on a variety of parameters would be required, as was also identified in section 2.5. Unfortunately, such a large collection of information is frequently unavailable due to privacy concerns and coordination issues among the different organizations that possess the data. Identifying minor risk zones is an alternate method, however, it is confined to aggregate identification and is imprecise. Because homes in the same region frequently share socioeconomic and material features, spatial analysis is particularly well suited to this problem.

This is well explained by the spatial dimension of the factors that influence electricity access, which is systematically discussed in Bouzarovski and Simcock (2017). End-use energy injustice, like other types of inequality, is also a fundamentally geographical phenomenon. It is unequally distributed and experienced in various locations, and it is formed through many complex spatialities of distribution and perception. The main implication of this reasoning in terms of vulnerability to electricity access is that where someone lives appear to be at least as important as the socioeconomic group that they are a part of. Nevertheless, in much of the current literature and policy discourse, inequalities and vulnerability tend to be defined in terms of the latter.

The considerable expense and difficulties of getting extensive survey data are highlighted in Kennedy et al. (2019), Gibson and Olivia (2010), and other studies, support the use of geo-referenced datasets in conjunction with widely accessible socio-economic indicators to help predict electricity access or usage. However, such systems demand access to a diverse set of indicators with a low resolution. Granular socioeconomic data is difficult to acquire and sometimes unavailable in developing nations such as India. As an alternative, this study recommends investigating if and how these socioeconomic indicators might be substituted by remote sensing data, which is inexpensive and generally available.

2.7. Use of Remote Sensing Data

One of the first studies to use nighttime light intensity and population counts yielded poverty indicators that were highly linked with sub-national indices of income poverty. This highlighted the feasibility of employing remote sensing data to assist in the monitoring of socioeconomic aspects on the ground (C. D. Elvidge et al., 2009).

Remote sensing data can also assist forecast rural well-being and poverty. For example, utilizing data, a vegetation index along with rainfall has been used to forecast rural poverty through agricultural productivity and food consumption (Imran et al., 2014; Morikawa, 2014). In one research, poverty status was predicted on a very granular scale using two dimensions of vegetation land cover and travel time to marketplaces, which were evaluated using road maps and awareness of impediments such as water bodies. Travel time to market towns, the proportion of a village covered in woods, and the percentage of a village covered in winter crop were all found to be strongly associated to wellbeing (Watmough et al., 2016). The promise of remote sensing in terms of poverty resides in its ability to monitor SDGs and anticipate rural poverty levels where field surveys are difficult and expensive.

Remote sensing data has also found use in the energy sector. They've been used to track changes in electricity access and socio-economic welfare by utilizing long-term nighttime lights (Falchetta & Noussan, 2019; Proville et al., 2017). Dugoua et al. (2018) compares the nighttime lights data to specific rural electrification statistics from India's 2011 Census. Even at the village level, the findings imply that many nighttime illumination measurements generated from satellite data are remarkably accurate for quantifying rural electrification, and with simple statistical approaches. Other studies report similar findings as well when the consistency of electrification rates are compared to survey-based findings (Falchetta et al., 2019; B. Min & Gaba, 2014; B. Min et al., 2013; Xie et al., 2016). The prediction of electricity access using remote sensing data has largely been handled in these researches using pictures of nocturnal lights combined with population counts.

However, nighttime light evidence, informative as it may be, is limited and does not account for the multidimensional aspect of electricity access, as highlighted in section 2.5. In Wang et al. (2021), a broader variety of remote sensing data was employed to address the multidimensional aspect of electricity access and estimate electricity access prevalence more efficiently and reliably.

However, few studies examine the actual outages of electricity using remote sensing. Cole et al. (2017) used nighttime satellite imagery to estimate power outages following Hurricane Sandy in 2012 in the US and showed promising results. Mann et al. (2016) show first results for developing estimates of electrical reliability for western India at a granular scale using satellite data of nighttime lights. The study investigated quite a short time frame, nonetheless, results were promising for further investigation.

Machine learning methods are inherently suited to the nature of remote sensing data, as demonstrated by the use of classification trees and random forests (Mann et al., 2016; Wang et al., 2021; Watmough et al., 2019). In other experiments, neural networks were employed to map electricity usage and poverty mapping via satellite photos, with results that matched those of field surveys (Cole et al., 2017; Xie et al., 2016). These strategies are especially beneficial in underdeveloped nations because satellite imagery is often lacking or unstructured.

2.8. Key Findings

Electricity reliability should be identified in order to improve electricity access and sustainable development. However, no study sheds light on electricity reliability at a granular level for all of India. It was shown in the previous sections, that there is a clear issue with the reliability of the electricity supply, but because of the lack of metering, it is not clear how, when, and where this issue arises. As a result, households have to pay more than they actually use. For the vulnerable in India, this is unaffordable. There is also a strong mistrust and prejudice between the electricity companies and the vulnerable, due to this unfairness. New methods should be used to help advance the prediction of the electricity supply in India, and in doing so compliment information provided by metering or surveys. This requirement highlights the potential for remote sensing and machine learning to give important insights. This may bridge the gap between data normally accessible at a low granular level and the need for high granular analysis, as well as aiding in better tailoring policy actions to local realities.

3
Methods

This chapter provides an examination of the technique created to solve the research questions of this graduation project based on the research objectives specified in Chapter 1: Introduction and the associated scientific research and knowledge examined in Chapter 2: Literature Review. A special emphasis is placed on conceptually defining the design of the mathematical equations and algorithms that have been employed. The next sections offer an overview of the created approach as well as information regarding the modelling framework that was used. Finally, the data processing methods employed, and the evaluation metrics produced for the categorization findings are presented briefly.

3.1. Overview

This section provides an overview of the modelling approach that was developed and used to train and test our machine learning models. The methodology is broken down into three steps, as seen in figure 3.1, the gathering and understanding of the data sources (marked in green), the processing of the data alignment, training and evaluating of the algorithms (marked in yellow), and, finally, applying the algorithms to generate output maps (marked in blue). These steps are explained in detail in the following sections.

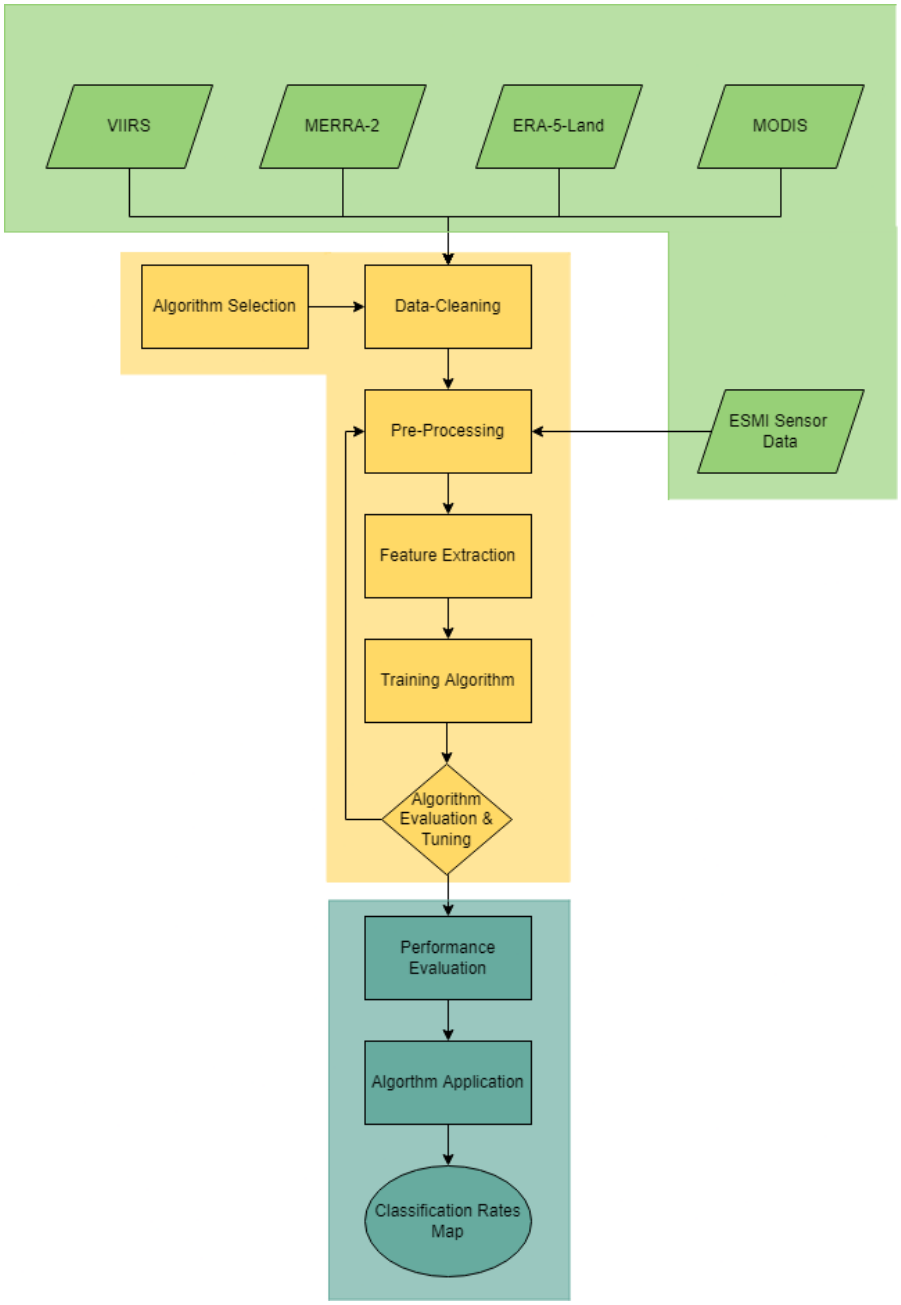


Figure 3.1: Workflow of processes, with marked in green; the gathering and understanding of the data sources, marked in yellow; the processing of the data alignment, training and evaluating of the algorithms, and marked in blue; applying the algorithm to generate output maps.

3.2. Data Sources

The algorithms will be trained with the use of remote sensing and real-time electricity data. Wang et al. (2021) showed that the use of multiple remote sensing and simulation data can accurately predict the risk of energy poverty. Thus, this study will use multiple remote sensing and simulation data as well to identify the outages of the electricity supply, shown in table 3.1. The availability of electrical supply is frequently more important during the nighttime hours, particularly for lighting requirements (Bhatia & Angelou, 2015). Nighttime lights are essential to capture the reliability of the electricity supply of India. Next to nighttime lights, wind, precipitation, temperature, fine particulate matter. Wang et al. (2021) used similar data sources for identification of energy poverty. As reliability issues are more prevalent in areas where energy poverty is found, these data will also be used as predictors in this study. In the

following sections, the data sources will be described and explored.

Table 3.1: Electricity reliability predictors

Predictors	Source	Spatial Resolution	Temporal Resolution
Nighttime Lights	SNPP VIIRS DNB NASA	15 arc-seconds	Daily (overpass 00:00-02:00)
Wind	Copernicus ERA5-Land	0.1° x 0.1°	Hourly
Precipitation	Copernicus ERA5-Land	0.1° x 0.1°	Hourly
Temperature	Copernicus ERA5-Land	0.1° x 0.1°	Hourly
Air Quality	NASA MERRA-2 OR NASA MODIS	0.5° x 0.625° OR 0.1° x 0.1°	Hourly OR Daily (overpass 05:00-10:00)
Land Cover	Copernicus Global Land Operations	100 m x 100 m	Yearly
Population	NASA SEDAC a	1 km x 1 km	10-year

3.2.1. ESMI

Voltage sensors data supplied by the Electricity Supply Monitoring Initiative (ESMI) developed by the Indian NGO Prayas in late 2014 were among the target data used to train the machine learning algorithm, (Canares et al., 2017). The ESMI data are gathered through battery-powered voltage monitoring devices installed into power outlets in a few hundred households and businesses across India. Voltage readings are taken every one minute and sent to a central database via the local cellular network.

The dataset on Harvard Dateverse contains data from 2014 to 2019 (Prayas, 2019). The dataset contains 528 locations. In figure 3.2 the distribution of the sensor locations is presented per state. It is obvious that by far the most sensors are placed in the state of Uttar Pradesh. As was shown in the previous chapter, this state is one of the worst performing states considering electricity connections and reliability. With this in mind, and computational power, it is chosen to scope down to Uttar Pradesh, in order to get meaningful results in time.

During the exploration of the sensor locations in Uttar Pradesh, it was discovered that most sensor locations came online in 2017 or the beginning of 2018, shown in figure 3.3a. The sensor location data seems to be lacking in regard to the year 2019, as the data shows that all sensor locations go offline at the end of 2018, shown in figure 3.3b. Even tough, the dataset contains sensor data from 2019, the data on the sensor locations does not contain the information for that year. The decision was made to focus on the year 2018, as this year would contain the most online sensor locations for a full year. This resulted in 67 sensor locations for further analysis.

The locations of the sensors could only be partly found on the ESMI website. Furthermore, the locations shown on the ESMI website are actually the locations of the GPRS tower nearest to the sensor. The actual location of the sensor could be a few meters to a few hundred meters away from the GPRS tower (Personal communication source here). Moreover, only 15 of the 67 sensor locations from 2018 in Uttar Pradesh could be found on the ESMI website. ESMI could not provide the location of the others (personal communication source again).

To overcome this issue, the remaining sensor locations had to be estimated based on their names with the help of Google Maps. Sensor locations were only kept if it was certain that the name pointed to one area on Google Maps. Thankfully, only one sensor location had to be dropped from the dataset, however, this manual estimation often pointed to areas which spanned 1 km x 1 km, thus, this introduces extra uncertainty about the exact location of the sensors, and this has to be taken into account in the rest of the research. In figure 3.4 the sensor locations are shown as red dots. As can be seen, the locations are quite concentrated in certain areas and completely absent in others. This is important to note, as this may mean that the sensor data might not be representative of the truth for the whole of Uttar Pradesh

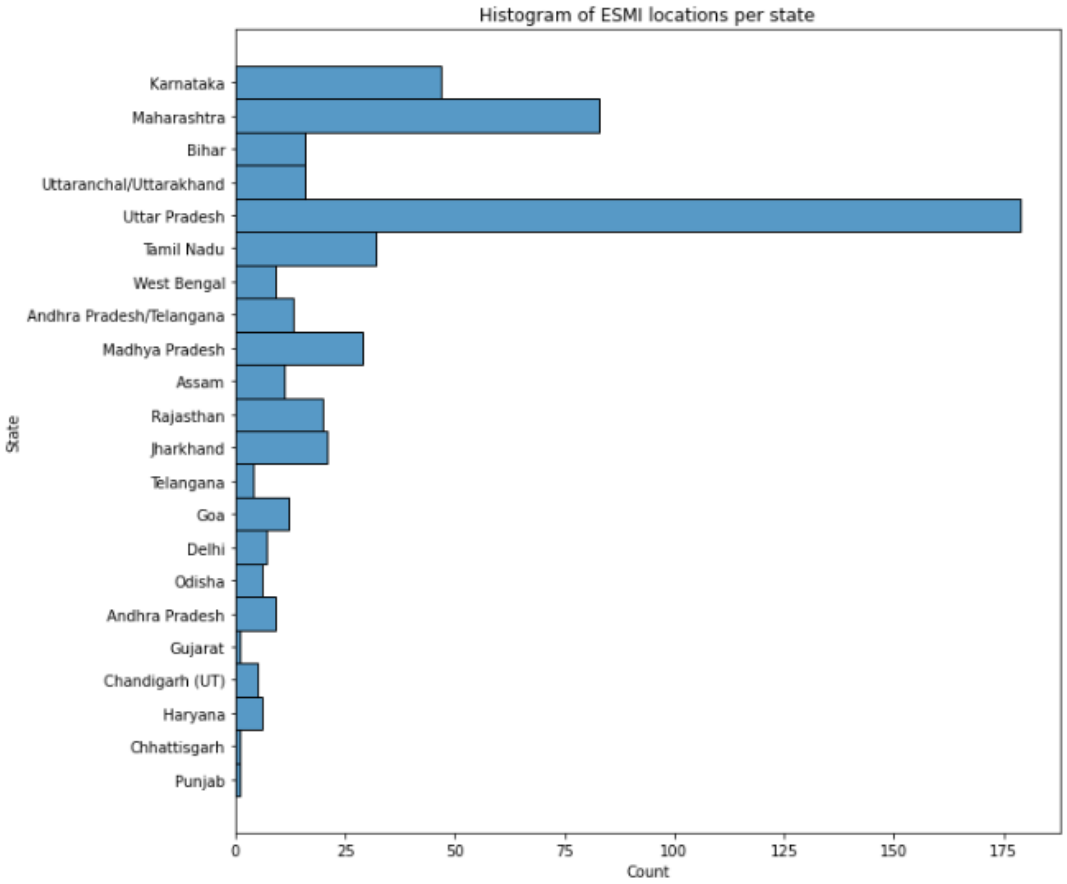
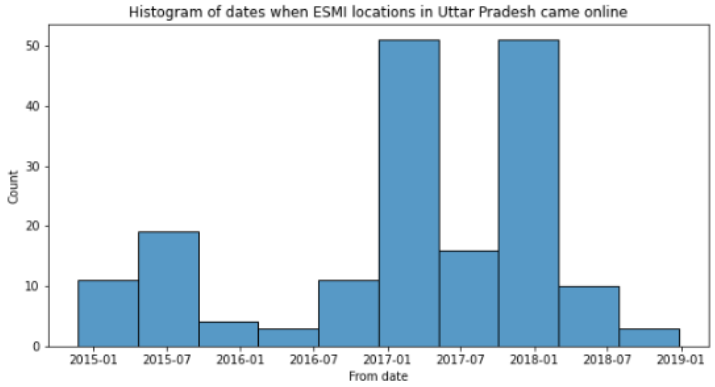
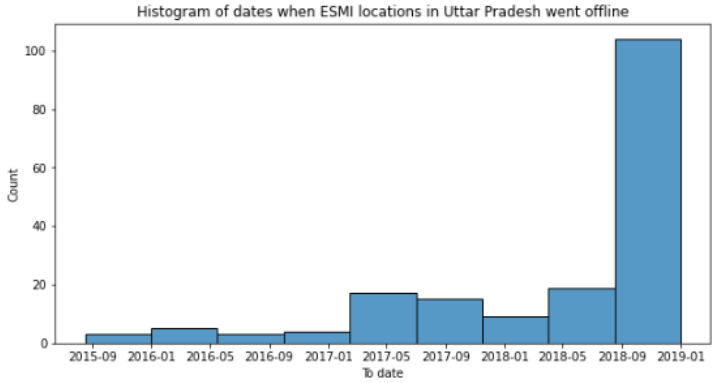


Figure 3.2: ESMI sensor locations per state, Uttar Pradesh has by far the most sensors, Maharashtra, which has the second most sensors, has less than half when compared to Uttar Pradesh.



(a) Distribution of when ESMI sensors in Uttar Pradesh came online



(b) Distribution of when ESMI sensors in Uttar Pradesh went offline

Figure 3.3: Histograms of first online and offline dates of the ESMI sensors in Uttar Pradesh, there are two big impulses of new sensors in early 2017 and 2018. The sensor location data stops at 2019, although not all sensors went offline, the information which did not is not present.

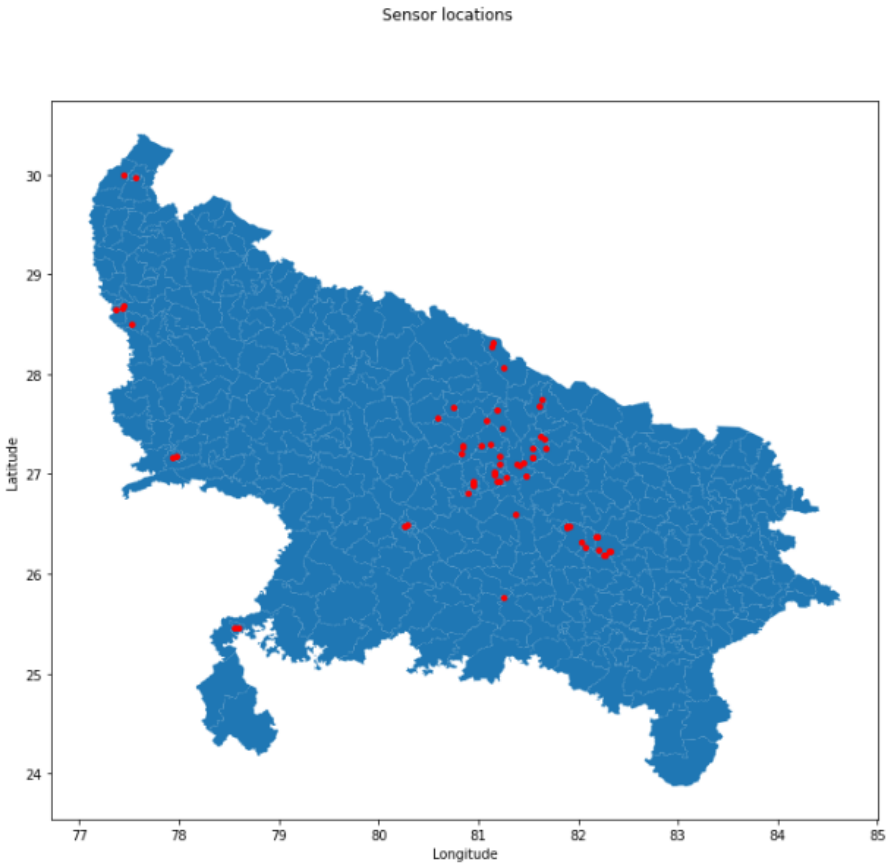


Figure 3.4: ESMI sensor locations in Uttar Pradesh. As can be seen, the locations are quite concentrated in certain areas and completely absent in others. This is important to note, as this may mean that the sensor data might not be representative of the truth for the whole of Uttar Pradesh

3.2.2. Nighttime Lights

The Visible Infrared Imaging Radiometer Suite (VIIRS) instruments aboard NASA/Suomi NOAA's National Polar-orbiting Partnership (Suomi NPP) and NOAA-20 satellites deliver global daily measurements of nighttime visible and near-infrared light suitable for Earth system science and applications research (Román et al., 2018). The spatial granularity of VIIRS is 15 arc-seconds, with a daily temporal resolution with an overpass time between 00:00 and 02:00. Due to this, it was decided that the research should focus on this time range, as VIIRS is more likely to identify outages that occur during its overpass time. However, this comes with a data loss of 22 hours. Is it still possible to get generalizable results for the whole day, based on these two hours?

Mann et al. (2016) had access to feeder-line data from Maharashtra, India, and looked at the correlation of outages during the day, the evening, and the night. Figure 3.5 depicts the proportion of zero feeder-line voltage readings during daytime hours (6 a.m.–6 p.m.), VIIRS overpass periods (midnight–2 a.m.), and Defense Meteorological Satellite Program, (DMSP) overpass times (7–10 p.m.). It should be noted that these measurements are primarily generated from MahaDiscom feeder-line voltage data during time periods that correspond with satellite flyover. They compared these readings to zero voltage readings produced from feeder-line voltage data for daylight hours. As can be observed, there is a strong relationship between the frequency of power outages throughout the day and the frequency of outages at satellite overpass periods. This applies to both the VIIRS and DMSP.

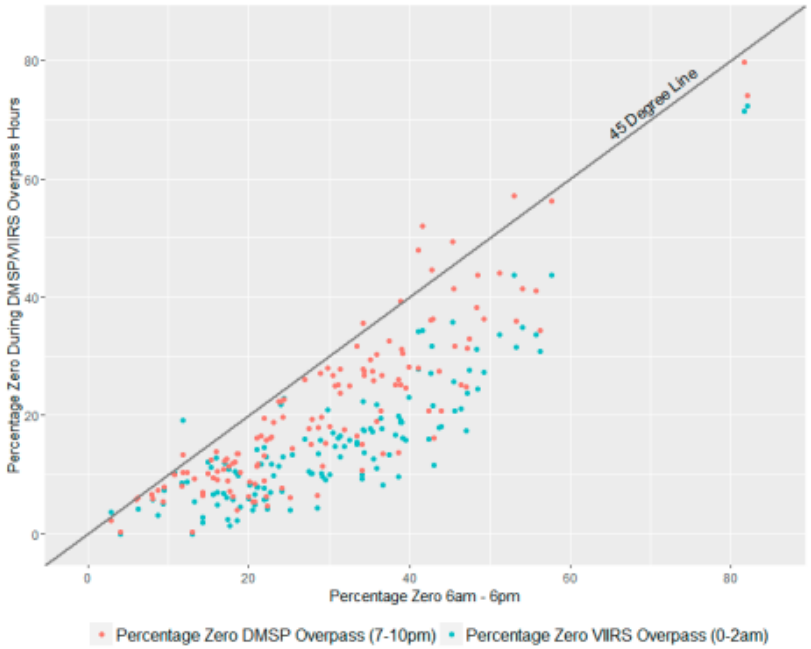


Figure 3.5: The correlation of outages during the day, the evening, and the night on feeder lines in 137 Maharashtra divisions from 2007 to 2013. With a 45-degree line, observations of day and nighttime outage rates for the DMSP overpass time (red) and the VIIRS sensor overpass time (blue) are given (black line) (Mann et al., 2016)

NASA produced a set of products known as the Black Marble to facilitate the utilization of Nighttime Lights data for scientific purposes. NASA's Black Marble product package for the Suomi NPP (VNP46) enhances Nighttime Lights data in a variety of ways (NASA, 2021):

- Atmospheric correction: The VNP46 method corrects the effects of aerosols, water vapor, and ozone on evening light radiances.
- Bidirectional Reflectance Distribution Function: Based on surface BRDF/albedo, the VNP46 algorithm calculates and eliminates moonlight influence, which was a major concern for Mann et al. (2016).
- Angular Correction: By performing consistency tests as a function of satellite view and lighting geometry, the VNP46 method adjusts for differences in artificial light sources.

Thus, VNP46 is well suited for nighttime lights analysis. Specifically, product VNP46A2 which is the daily moonlight-adjusted nighttime lights product will be used for this research. The data is supplied in the standard land hierarchical data format - Earth observing system (HDF-EOS) format. This product includes seven Science Data Sets (SDS), which are as follows:

- DNB BRDF-Corrected NTL
- Gap-Filled DNB BRDF-Corrected NTL
- DNB Lunar Irradiance
- Latest High-Quality Retrieval
- Mandatory Quality Flag
- Cloud Mask Quality Flag
- Snow Flag

However, for this research only DNB BRDF-Corrected NTL and Mandatory Quality Flag will be used. This DNB BRDF-Corrected NTL is computed from the raw sensor data and with an algorithm designed by NASA. But the algorithm is not perfect. It is advised to mask data for DNB BRDF-Corrected NTL based on where the Mandatory Quality Flag reports low-quality. The SDS are transformed to Geotiffs for further analysis.

The VNP46A2 product is gridded in a linear latitude/longitude format. To capture the whole of Uttar Pradesh, three tiles have to be used; h25v05, h25v06, h26v06. For the analysis, these have to be concatenated to each other. Furthermore, as mentioned earlier, the spatial granularity of VIIRS is 15 arc-seconds, which is about 500 m, depending where on earth the measurement is taken. However, during the location retrieval of the ESMI it was found that the actual location of the sensor could be a few meters to a few hundred meters away from the GPRS tower if the sensor could be found on the ESMI website. If the sensor could not be found, locations had to be estimated based on their names with the help of Google Maps. As the estimated locations often pointed to areas which spanned 1 km x 1 km, the spatial granularity of the nighttime lights will be down sampled from 15-arcsecond (about 500 m) to 30-seconds (about 1 km) using bilinear interpolation. The final product for 01-01-2018 is shown in figure 3.6.

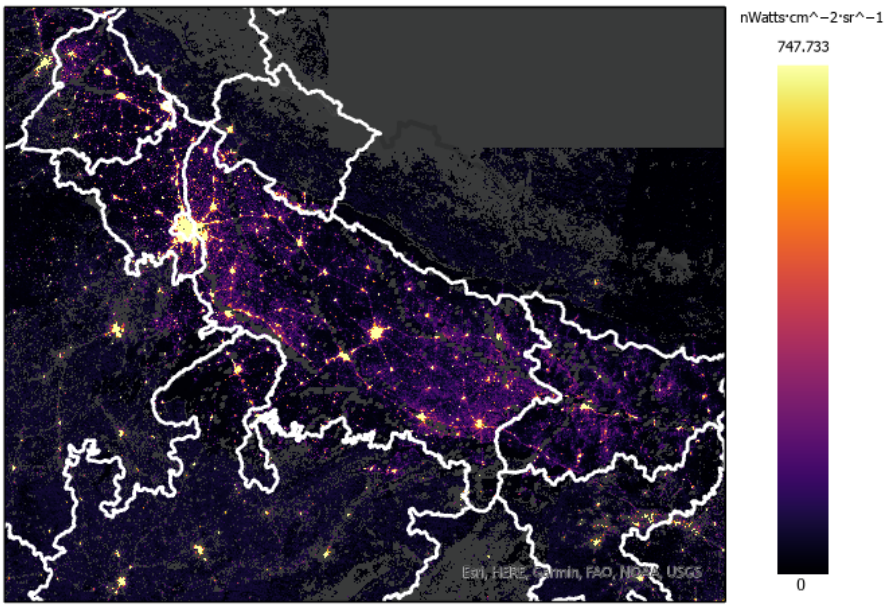


Figure 3.6: VIIRS VNP46A2 input 01-01-2018 for the algorithms (dark grey is the color of the basemap, thus a null value for VIIRS).

3.2.3. Weather Data

Kenward, Raja, et al. (2014) studied power outage data over a 28-year period in the United States. According to the analysis, weather was responsible for 80% of all disruptions between 2003 and 2012. The authors' data indicates a definite trend of weather-related accidents, but the authors also note the fact that physical and cyberattacks on the power infrastructure have escalated and should be recorded. Abi-Samra et al. (2014) highlight the damage to the electrical grid caused by seasonal storms, rain, and high winds in North America and in South-East Asia. These weather occurrences can cause trees to fall on local distribution and transmission lines to break, resulting in power outages. High temperatures also bring risk for the electricity reliability. Wildfires can cause large outages, as was seen in Australia (der Linde, 2019). The heat can also cause overheated and overloaded subterranean cables and overhead wires to melt (Parashar, 2022).

For this research, ERA5-Land data was used for weather data. ERA5 is the fifth generation Euro-

pean Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis of global climate and weather during the last four to seven decades (Muñoz-Sabater et al., 2021). ERA5-Land is a reanalysis dataset with higher resolution than ERA5 that provides a consistent perspective of the evolution of land characteristics across multiple decades. The land portion of ECMWF ERA5 climate reanalysis was replayed to create ERA5-Land. Using physical laws, reanalysis integrates model data using observations from across the world to create a globally comprehensive and consistent dataset. Reanalysis generates data that extends many decades back in time, offering an accurate depiction of previous climate. The data includes hourly readings of atmospheric variables with a spatial resolution of 0.1° x 0.1°. This study will use the wind, precipitation, and temperature variables of ERA5-Land from 2018 between 00:00 - 02:00.

Wind
The ERA5-Land contains two variables to compute the wind speed at for a cell, u- and v-component of wind, both components are measured at 10 m height (Copernicus, 2019-07-12) . The u-component is the wind's eastward component. It denotes the horizontal speed of air traveling eastward in meters per second. The v-component is the wind's northward component. It denotes the horizontal speed of air traveling northward in meters per second. To calculate the wind speed from u and v components of the wind, the magnitude formula is used (ECMWF, 2021), shown in equation 3.1.

$$| \vec{V} | = \sqrt{u^2 + v^2} \tag{3.1}$$

After the wind speeds are calculated, the mean of the two hours is calculated. In figure 3.7 the result of this computation is shown for 01-01-2018 00:00 - 02:00.

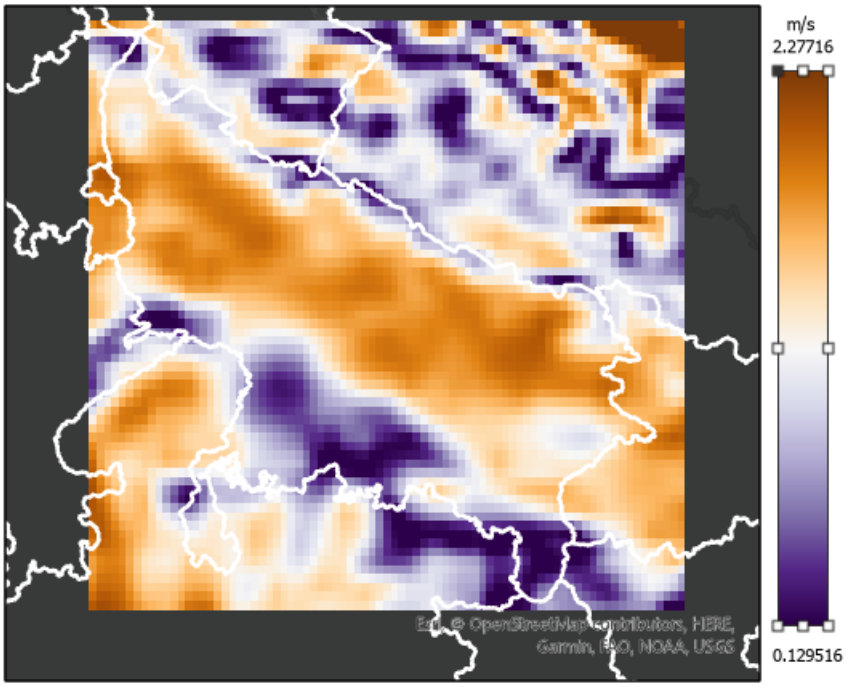


Figure 3.7: ERA5-Land wind speed input 01-01-2018 for the algorithms

Precipitation The ERA5-Land contains a variable for total precipitation. This is defined as: "*Accumulated liquid and frozen water, including rain and snow, that falls to the Earth's surface*" (Copernicus, 2019-07-12). Precipitation is measured in meters of depth, however, for this research millimeters are used. To get the total precipitation for the time range 00:00 - 02:00, the data are summed up. In figure 3.8 the result of this computation is shown for 01-01-2018 00:00 - 02:00.

Temperature The ERA5-Land contains multiple temperature variables, however, only one is used for this research, as they would all be very heavily correlated (Copernicus, 2019-07-12). For this research, the 2 m temperature variable is used. This is the temperature of the air at 2 meters above the

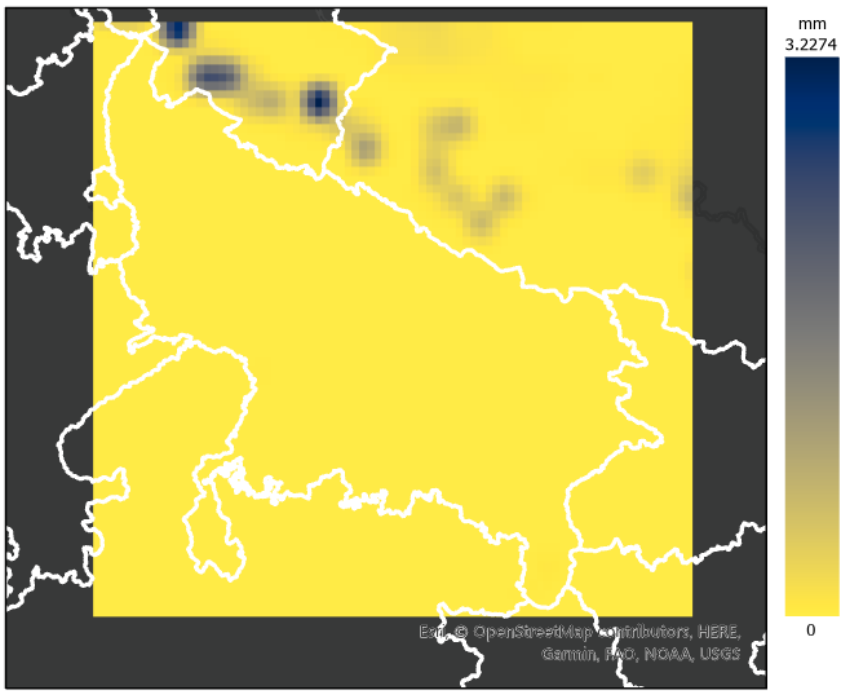


Figure 3.8: ERA5-Land total precipitation input 01-01-2018 for the algorithms

surface of land, sea, or in-land waters, measured in kelvin. Little processing is needed. Only the mean between the two hours is calculated. In figure 3.9 the result of this computation is shown for 01-01-2018 00:00 - 02:00.

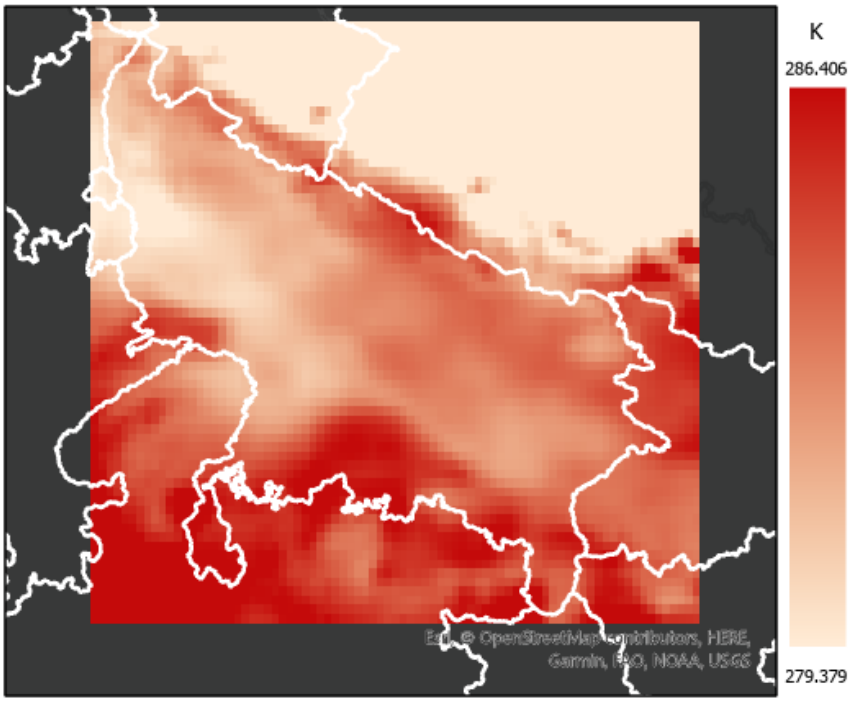


Figure 3.9: ERA5-Land 2 m temperature input 01-01-2018 for the algorithms

3.2.4. Air Quality

Households experiencing energy poverty are more likely to utilize wood or other biomass fuels instead of high-cost electricity, potentially causing air pollution. High PM2.5 concentrations are related with energy poverty in this scenario Wang et al. (2021). In India, many firms and households rely on their own generators for electricity (Szakonyi & Urpelainen, 2013). For this research, it is theorized that when there is an electricity outage, people will resort back to more polluting sources for their energy.

However, measuring air quality is quite difficult at a high granular level itself. Projects similar to ESMI measure the air quality in India with sensors, such as Prana Air (https://www.pranaair.com/about-us) and Atmos (http://urbansciences.in/). Another way of measuring the air quality is to measure the aerosol optical depth (AOD) or thickness (AOT) (used interchangeably).

AOD is the measure of aerosols (e.g., urban haze, smoke particles, desert dust, sea salt) spread inside a column of air from the sensor (Earth's surface) to the top of the atmosphere (Allen, 2005). It is a unitless scale that ranges from 0 to 5. A score of less than 0.1 denotes a crystal-clear sky having maximum visibility, whereas a score of 4 denotes the presence of thick particles that would make it hard to see the midday sun.

There are two datasets used in this research for the air quality. The first dataset measures AOT at 550 nm at an hourly time-averaged rate with spatial resolution of 0.5° x 0.625° using Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2). This collection consists of aerosol diagnostics (Global Modeling and Assimilation Office (GMAO), 2015). The spatial resolution is quite low, however, the temporal resolution allows for precise measurements that fall right in the time range from 00:00 - 02:00. Again, the mean of the two hours is calculated. In figure 3.10 the result of this computation is shown for 01-01-2018 00:00 - 02:00.

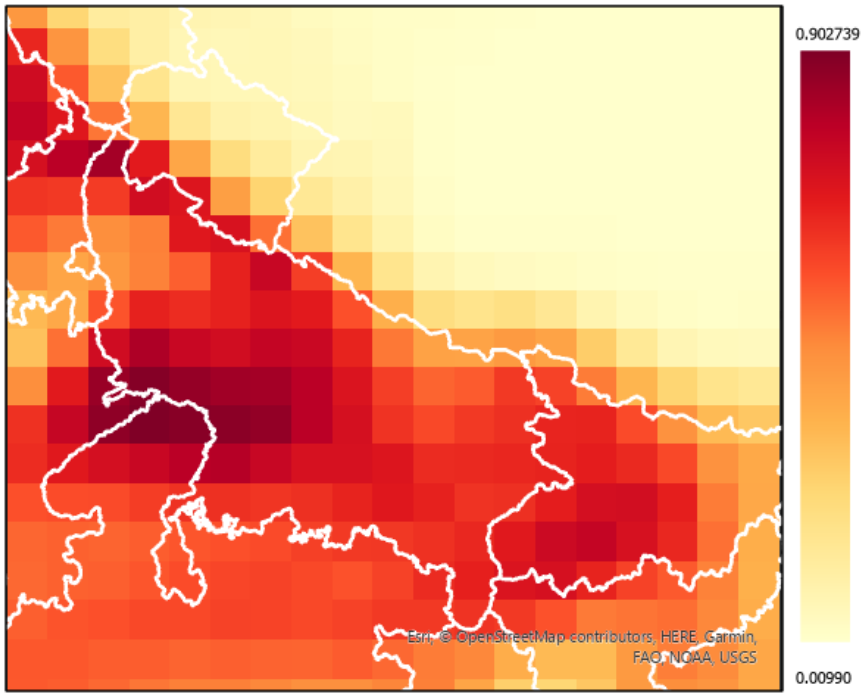


Figure 3.10: MERRA-2 AOT at 550 nm input 01-01-2018 for the algorithms

The second dataset measures AOD at 550 nm at a daily rate with spatial resolution of 0.1° x 0.1° using Moderate Resolution Imaging Spectroradiometer (MODIS) (Lyapustin & Wang, 2015). MODIS is a satellite sensor that is aboard on the Terra (formerly EOS AM-1) and Aqua (formerly EOS PM-1) satellites. Terra's orbit around the Earth is scheduled such that it crosses the equator from north to south in the morning, whereas Aqua crosses the equator from south to north in the afternoon. The overpass

times are not constant, resulting in 3 to 5 individual measurements every day. This has the benefit, that a better measurement for the day can be captured. Clouds can obstruct the view during one overpass, but might be gone for the next. However, it makes it more difficult to get precise measurement for the same time of day for every day. Still, the fine spatial resolution and multiple measurements per day can be of great value. The mean is computed for every day. In figure 3.11 the result of this computation is shown for 01-01-2018 00:00 - 02:00. It can be seen that there are quite a lot of data gaps. During further exploration of the data, it was discovered that more than 30% of the year extreme gaps were present in the data. Thus, it was decided to use only AOT MERRA-2 dataset for the air quality feature in this research.

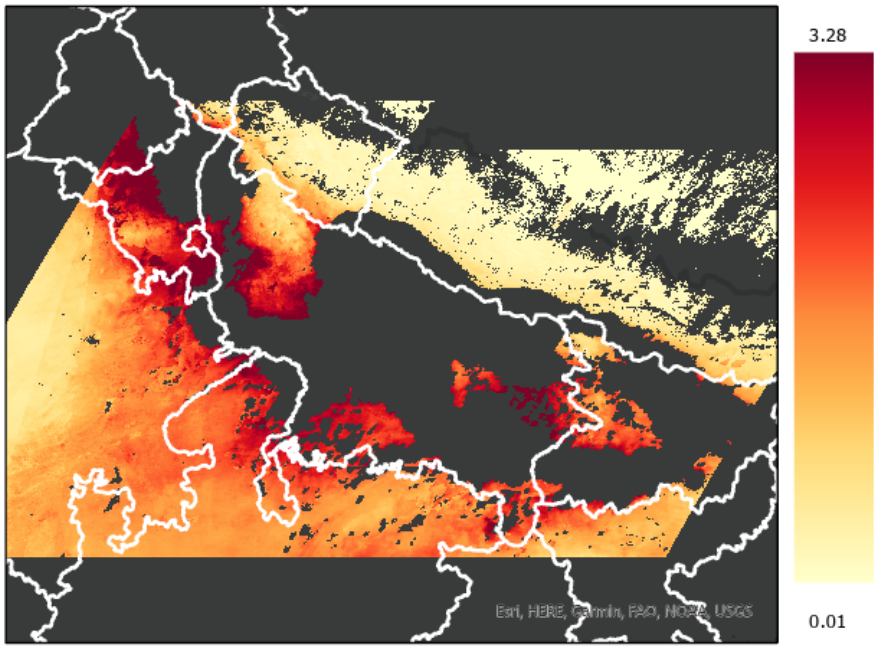


Figure 3.11: MODIS AOD at 550 nm input 01-01-2018 for the algorithms. It can be seen that there are quite a lot of data gaps, which makes the data less usable for the research.

3.2.5. Land Cover

The Copernicus Global Land Service (CGLS) is designated as a part of the Land service to run ”a multi-purpose service component” that delivers a variety of bio-geophysical products on the state and change of the global land surface (Buchhorn et al., 2020). The CGLS provides an annual dynamic global Land Cover product with a spatial resolution of 100 m. Land cover has a significant impact on the surface climate and biogeochemistry. The CGLS Land Cover package offers a main land cover scheme with three classification levels and class descriptions based on the Land Cover Classification System (LCCS). The Land Cover map is generated from the PROBA satellite’s Vegetation sensor. Figure 3.12 shows the classifications for Uttar Pradesh in 2018. This dataset was chosen because it is not only important to recognize outages when they happen, but also area’s that do not have any electricity access at all, like forests or deserts, to not misclassify an area having outages the whole year through when the algorithms are presented with unseen data.

3.2.6. Population

Finally, as described in section 2.7, one of the first studies to use nighttime light intensity and population counts yielded poverty indicators that were highly linked with sub-national indices of income poverty (C. D. Elvidge et al., 2009). Furthermore, the prediction of electricity access using remote sensing data has largely been handled in these researches using pictures of nocturnal lights combined with population counts (Falchetta & Noussan, 2019; B. Min & Gaba, 2014; B. Min et al., 2013; Proville et al., 2017; Xie et al., 2016). Spatial population data from the 2011 Census at a resolution of 1 km by 1 km is used for this research (Balk et al., 2020). While the population of Uttar Pradesh has grown during the

time between 2011 and 2018, the population data from 2011 still gives a clear image of the distribution of the population in Uttar Pradesh.

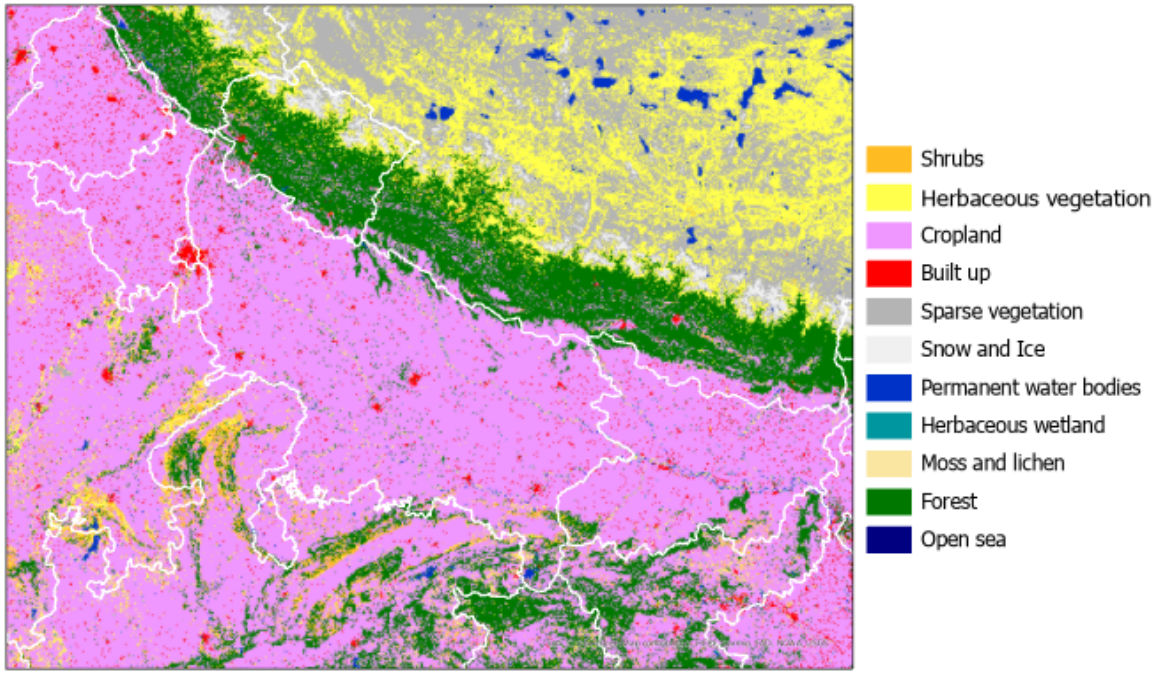


Figure 3.12: Land cover input 2018 for the algorithms. As can be seen, most of the land is used for cropland. Still, other land uses can be identified which might help with the classifications.

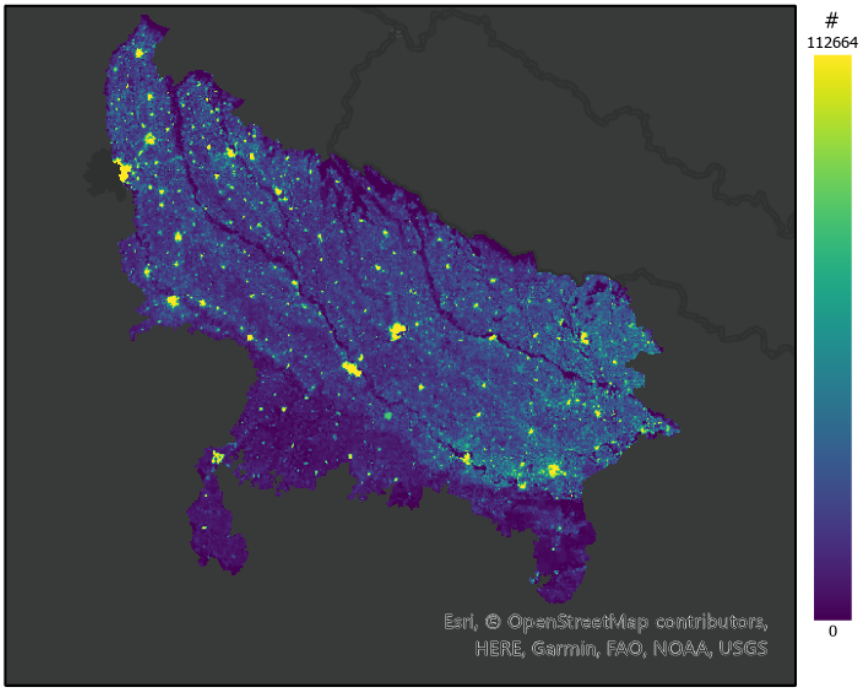


Figure 3.13: Population spread in Uttar Pradesh, 2011. It can be seen that the spread of population seems very similar to he spread of nighttime radiance in figure 3.6

3.3. Machine Learning Algorithm

Machine learning algorithms are computational algorithms that are meant to learn from their surroundings in order to mimic human intelligence (El Naqa & Murphy, 2015; Jordan & Mitchell, 2015). It is at the interface of computer science and statistics, as well as at the heart of artificial intelligence and data science, and is one of today's fastest expanding technological topics. The development of novel learning algorithms and theory, as well as the continual expansion in the availability of digital data and low-cost computation, have fueled recent advances in machine learning. Data-intensive machine learning approaches are being used throughout science, technology, and commerce, resulting in greater evidence-based decision-making in a variety of fields such as health care, manufacturing, education, modeling, policing, and marketing. The processes' degrees of complexity can vary, and they may entail numerous phases of complicated human-machine interactions and decision-making, which would naturally attract the application of machine learning algorithms in optimizing and automating processes. I will be using the open-source programming language Python for programming the algorithms (Python, 2021). Python has a lot of libraries that can be used for machine learning, such as Numpy, Scipy. Scikit-learn, Theano, TensorFlow, Keras, PyTorch, Pandas, Matplotlib, and Seaborn. The last question will be answered by analyzing the results of the algorithms.

The majority of machine learning algorithms feature variables known as hyperparameters that influence their behavior (Jordan & Mitchell, 2015; O'Shea & Nash, 2015). The hyperparameters are established before the learning process and can have a significant impact on their performance. Furthermore, the effectiveness of machine learning algorithms is strongly reliant on having input features that are informative, discriminating, and independent, as well as the correctness of the labelled data used to train them.

In general, machine learning algorithms can be categorized as supervised, unsupervised, semi-supervised, or reinforcement learning algorithms (Louridas & Ebert, 2016). Supervised learning algorithms are given labeled datasets to use to learn from, and they are used to solving problems such as classification and regression. Unsupervised learning methods employ unlabeled datasets to discover beneficial features of underlying structures and trends. Typically, these techniques are employed to solve clustering and dimensionality reduction problems. Semi-supervised learning techniques are a combination of two categories, since they train on both labeled and unlabeled data. They are utilized when the majority of the observations lack labels and make use of a little quantity of tagged data to increase learning accuracy. However, semi-supervised learning still often suffers from performance degradation caused by the introduction of unlabeled data (Van Engelen & Hoos, 2020). Finally, reinforcement learning is a strategy that rewards positive actions while penalizing undesirable ones. A reinforcement learning agent, in general, is capable of seeing and interpreting its surroundings, taking actions, and learning via trial and error (François-Lavet et al., 2018).

Machine learning methods are inherently suited to the nature of remote sensing data, as demonstrated by the use of random forests (Mann et al., 2016; Wang et al., 2021; Watmough et al., 2019). In other experiments, convolutional neural networks were employed to map electricity usage via satellite photos, with results that matched those of field surveys (Cole et al., 2017; Xie et al., 2016). These strategies are especially beneficial in underdeveloped nations because satellite imagery is often lacking or unstructured. As mainly random forests and convolutional neural networks are used in similar studies, only these two algorithms will be further explored.

3.3.1. Random Forest

Random Forest (RF) algorithms are built up from Decision Trees. Decision Trees are a non-parametric supervised classification/regression approach that learns basic decision rules from input data characteristics and estimates the value of a target based on these rules. Figure 3.14 depicts how a decision tree creates a forecast and its essential components. It may be described as a set of nodes and edges arranged in a hierarchical framework (Criminisi et al., 2012). Each internal (split) node has a function that tests the supplied data and makes a decision. Every branch relates to the result of that test. Terminal (leaf) nodes have a class label that is the forecast of the target. The root node is the tree's highest node. The decision tree may be thought of as a hierarchical piecewise model that divides complex

problems into easier ones.

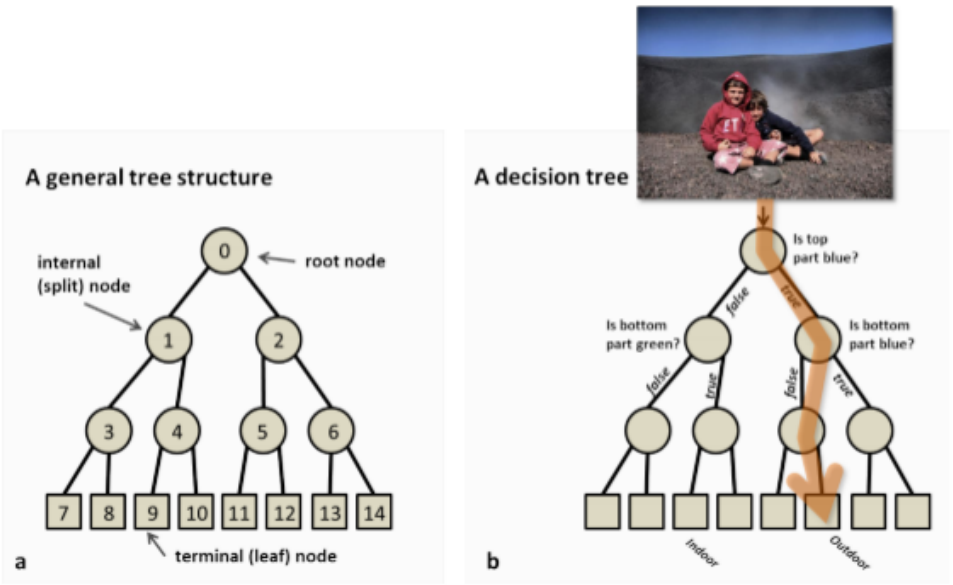


Figure 3.14: (a) The primary components of a decision tree. (b) An example of a decision tree that attempts to determine if a photograph depicts an indoor or outdoor environment (Criminisi et al., 2012).

The RF is a straightforward and simple machine learning algorithm for tackling classification and regression problems. It is an ensemble approach that consists of numerous smaller decision trees that are trained on a randomized subset of the training data and make their own estimates, which are then aggregated to get a more precise estimate (Breiman, 2001), shown in 3.15. As a result, they are less prone to overfitting than typical decision tree classifiers and can offer estimates of which variables are crucial in the classification or regression. Random Forests can also manage big data with high dimensionality and varied feature types. When comparing the Random Forest classifier to other machine learning and deep learning algorithms, it delivers an excellent balance of accuracy and efficiency (Fernández-Delgado et al., 2014).

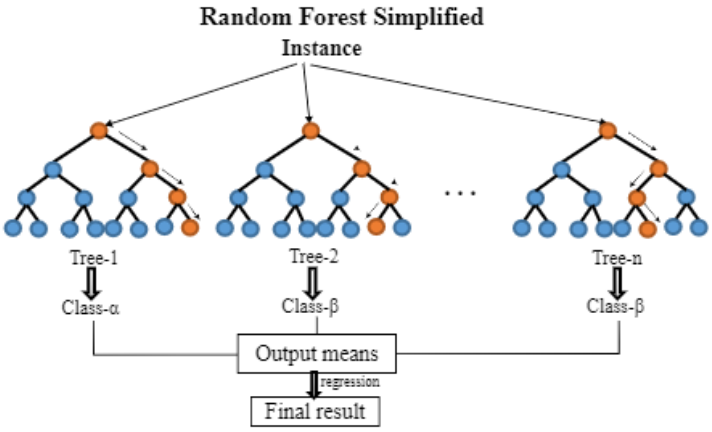


Figure 3.15: Different trees have different estimations for the outcome, via majority vote an estimation is made (Gao et al., 2021).

3.3.2. Convolutional Neural Networks

A Deep Learning algorithm is a Convolutional Neural Network (CNN). Deep Learning is a class of machine learning, and early machine learning techniques heavily impacted its development. Deep Learning algorithms, in particular, enable a system to autonomously find the representations required for feature identification or classification using raw data (François-Lavet et al., 2018). CNNs have been used for regressions as well (Fischer et al., 2015), however, they are much more commonly used for classification. They concentrate on learning successive layers of more relevant data representations using artificial neural networks in a hierarchical learning process. Thus, deep learning is a multiphase method of learning data representations using appropriately scaled basic methods. A neural network consists of an input layer, hidden layer(s), and an output layer, illustrated in 3.16. Hidden layers are using store weights that parameterize the change they apply to the incoming data. These algorithms strive to identify the optimal set weights of all layers in the complete network through an iterative training process in order to enhance their performance. This is accomplished with the use of a loss function, which computes a similarity rating between their predictions and the real labels of the training data. This score is then utilized as a feedback signal by an optimizer that employs a back propagation technique to alter the weights.

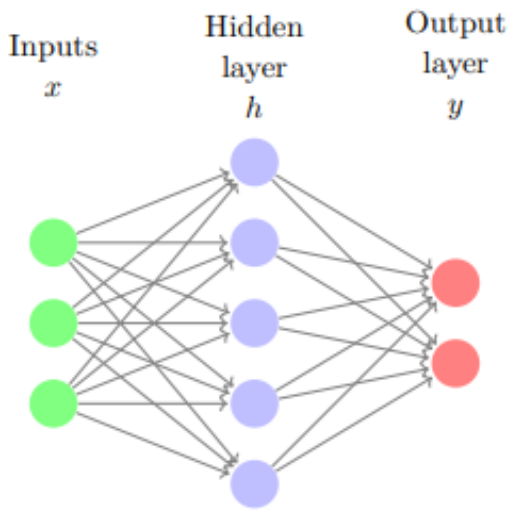


Figure 3.16: A neural network with one hidden layer is illustrated (François-Lavet et al., 2018)

A CNN can take an image as input, assign weights to distinct characteristics of the image, and distinguish them. When compared to other neural network techniques, the amount of pre-processing required by a CNN is significantly less (Albawi et al., 2017). While filters in primitive approaches are hand-crafted, CNNs can learn these filters/characteristics with adequate training. The core functionality of a CNN may be divided into four distinct phases (O'Shea & Nash, 2015), shown in 3.17.

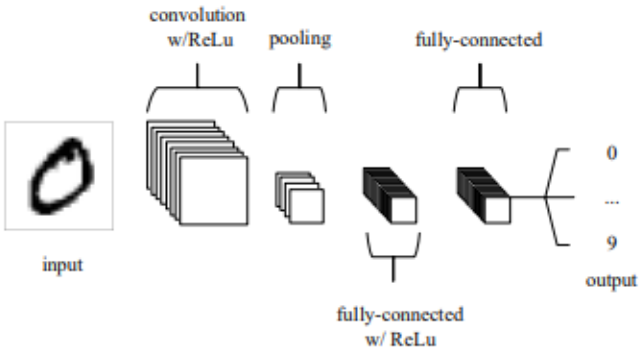


Figure 3.17: Core functionality of a CNN (O'Shea & Nash, 2015).

1. The input layer, like in other kinds of ANN, will store the data values.

2. The output of neurons connected to local portions of the input will be determined by the convolutional layer by calculating the scalar product of its weights and the region related to the input volume. The rectified linear unit (ReLu) applies an activation function to the specific regions to the output of the preceding layer's activation.

3. The pooling layer will then simply downsample along the regions of the given input with a specific precision, reducing the amount of variables within that activation even further.

4. The fully-connected layers will then execute the same functions as typical neural networks, attempting to generate output values from the activations for classification.

3.3.3. Random Forest vs Convolutional Neural Networks

RF is less complex than CNN. As a result, RF is less computationally expensive (Lan et al., 2020). Lan et al. found that the random forest algorithm was the simpler and had an easier structure to understand, as well as the faster interpretation time. Although, in terms of classification outcomes, the CNN method was the more effective. "For the sake of performance, the CNN will reduce the comprehensibility of the features to the extent that the algorithm becomes a black box. CNN require defining an appropriate architecture for specific problems. In this research, classification will be attempted. It is therefore to an advantage to easily switch between these objectives, without too much effort, like creating another architecture. CNNs also have more hyperparameters than RFs, and they are much more sensitive to them (Roßbach, 2018)

CNN will require more data to be effective, than RF (Xu et al., 2021). RF thrives at tabular and structured data using small sample sizes (less than 10,000 samples), but deep nets excel at structured data with greater sample sizes.

RF is employed as the classification algorithm in this research, using the Python library Scikit-Learn (sklearn) (Pedregosa et al., 2011). Robustness, computational performance, feature significance estimation, and comprehensibility are its benefits over the CNN algorithm (Breiman, 2001; Roßbach, 2018). It is an algorithm designed to cope with multi-label classification and regression out of the box, and it handles multicollinearity of features and data dimensionality efficiently (Belgiu & Drăguţ, 2016; Criminisi et al., 2012).

3.4. Machine Learning Set-Up

The sections that follow go through the algorithm set-up that was chosen and the approach that was utilized to validate the algorithm.

3.4.1. Target Data

As described in section 3.2.1, ESMI data is real-time data, however, the predictor data are not. So, the ESMI data has to be adapted to fit the research problem. For the final application of the algorithm, it needs to predict normal access and outages correctly, but also locations where there is no access at all. So, in addition to the 66 places where there is access and electricity is monitored, 30 additional spatially diverse locations in Uttar Pradesh with no electricity access were included in the target data set. These include rivers, lakes, forests, and other naturally occurring vegetation. The locations were manually selected by using Google Earth and Earth at Night. It is believed that the 30 chosen locations properly capture the diverse landscapes of Uttar Pradesh, although there is no literature backing this up. The sensor locations and additional locations are presented in figure 3.18, with sensor locations in red, no access locations in black. This brings the total up to 96 locations in the target data, with 29407 samples. During the preparation of the target data, it turned out that not all ESMI locations were consistently online during the whole year. However, this does not affect the RF algorithm, so it was chosen to continue with this set-up of the target data.

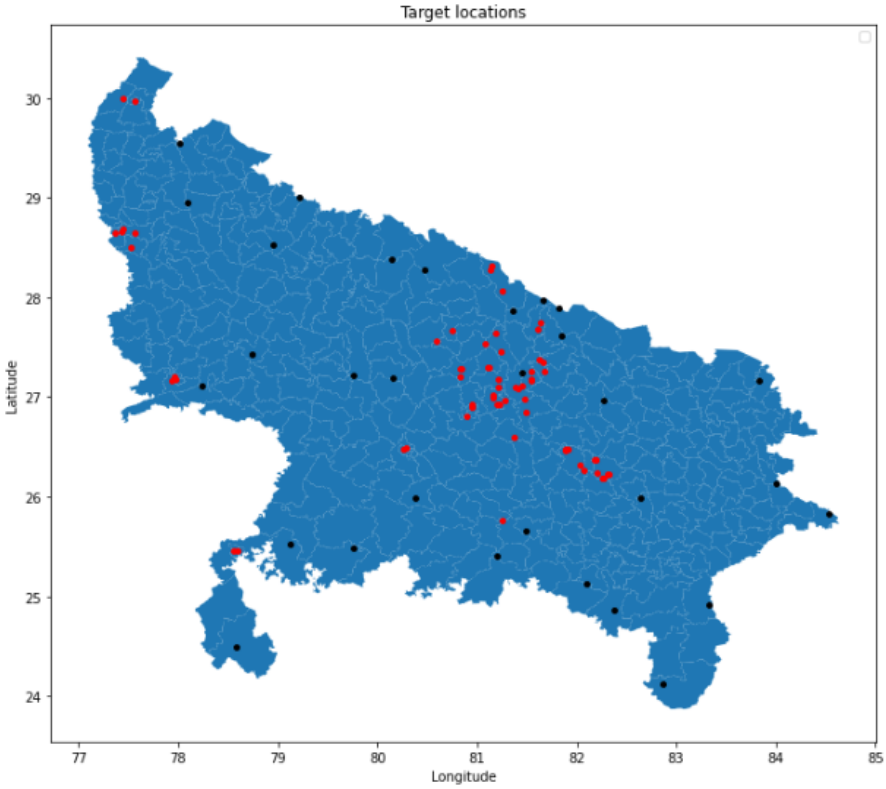
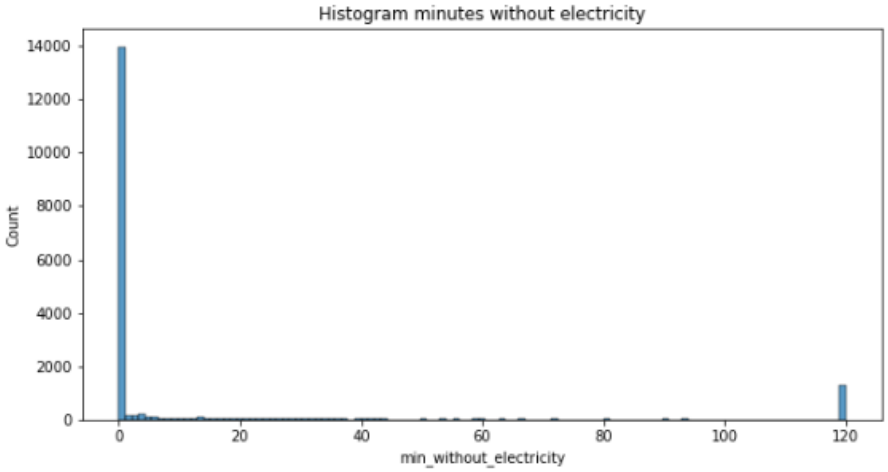


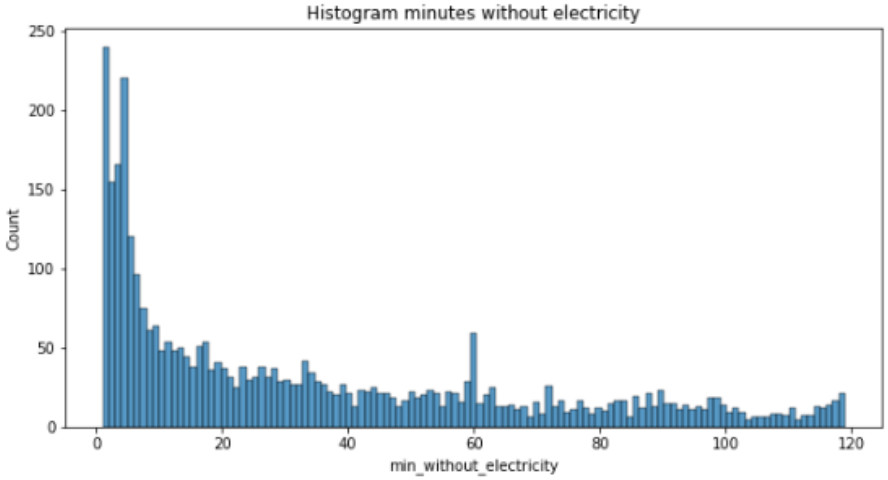
Figure 3.18: Target locations in Uttar Pradesh, sensor locations in red, no access locations in black, the sensor locations are quite concentrated in certain areas, while the no access locations are more spread out through the state.

Every minute between 00:00 and 02:00 that there was an outage reported are summed up per day and sensor for a new variable containing the minutes without electricity variable per location. So, this variable ranges from 0 to 120. With no access locations always showing 120 minutes without electricity. In figure 3.19a, the histogram of the new variable is shown for only the ESMI locations. Only the values for the ESMI locations are shown, because for all the non access locations the values are 120 (10950 samples). Clearly, zero minutes without an outage occurs much more often than the rest. What should be noted, is that 120 minutes without electricity is the second most common. At face value, this looks like it indicates that outages longer than two hours are more common during the night than shorter ones, but the cutoff point for this dataset is 120 minutes. Thus, all the outages longer than 120 minutes are grouped in 120 minutes. In total, there are 4752 samples with more than zero minutes without electricity, of which 1264 two hours or longer. This is still quite a large part of all the outages, these long outages might not be accidental, planned outages are quite common in Uttar Pradesh and

they often last longer than two hours (Kennedy et al., 2020; B. K. Min et al., 2017; Parashar, 2022). In figure 3.19b the histogram of minutes without electricity is shown for samples with more minutes than 0 and less than 120 for only the ESMI locations. Here, it is clear that shorter outages are more common, but there are some aberrations. For instance, at around 16 and 17 an increase is seen, and again at around 33. However, at minute 60 is a much more pronounced jump is seen.



(a) Histogram of minutes without electricity of ESMI locations



(b) Histogram of minutes without electricity (0 < minutes < 120) of ESMI locations

Figure 3.19: Histograms of minutes without electricity of the ESMI sensors in Uttar Pradesh. The distribution is split into two figures, as the visibility of the distribution of shorter minutes without electricity is lost when 0 minutes and the aggregate of 120 minutes without electricity are present.

As the predictor data sources do not capture information every minute of the 00:00 - 02:00 time range, it expected that a classification algorithm used for identifying outages would perform better than a regression algorithm identifying the actual minutes without electricity. Looking at the distribution of minutes without electricity, there is a great disparity in frequency for the individual minutes without electricity, forming highly imbalanced data. The larger the imbalance, the stronger the model's bias towards the majority class (Krawczyk, 2016). To reduce this imbalance, grouping the individual minutes without electricity over zero minutes into a class will be done. Although, regression still might lead to interesting results, based on the high imbalance and time constraints on the research, only classification is performed. In order to do that, the target data needs to be classified. The classes are Never Access, Normal Access, and Outage. Non access locations, like forests, should be classified as Never Access. An important decision is when to classify a day to Outage or not. There is no clear definition on how many minutes should pass for it to be called an outage. Strictly speaking, one minute without electricity

is an outage. This can of course be catastrophic during a surgery, but for lighting it is not as important. An outage of one minute is also probably much hard to identify an outage of two hours. So, based on the distributions of minutes without electricity shown in figure 3.19, six classification variables were created; over 0 minutes without electricity, over 5 minutes without electricity, over 15 minutes without electricity, over 30 minutes without electricity, over 60 minutes without electricity, and over 120 minutes without electricity. In figure 3.20 the distributions of the six classes in the target data are presented, this time for all the locations. While the samples with Never Access is constant, it is important to show this is not a binary classification problem. There is very clearly still a class imbalance in all the variables. Thus, this imbalance should be handled with care.

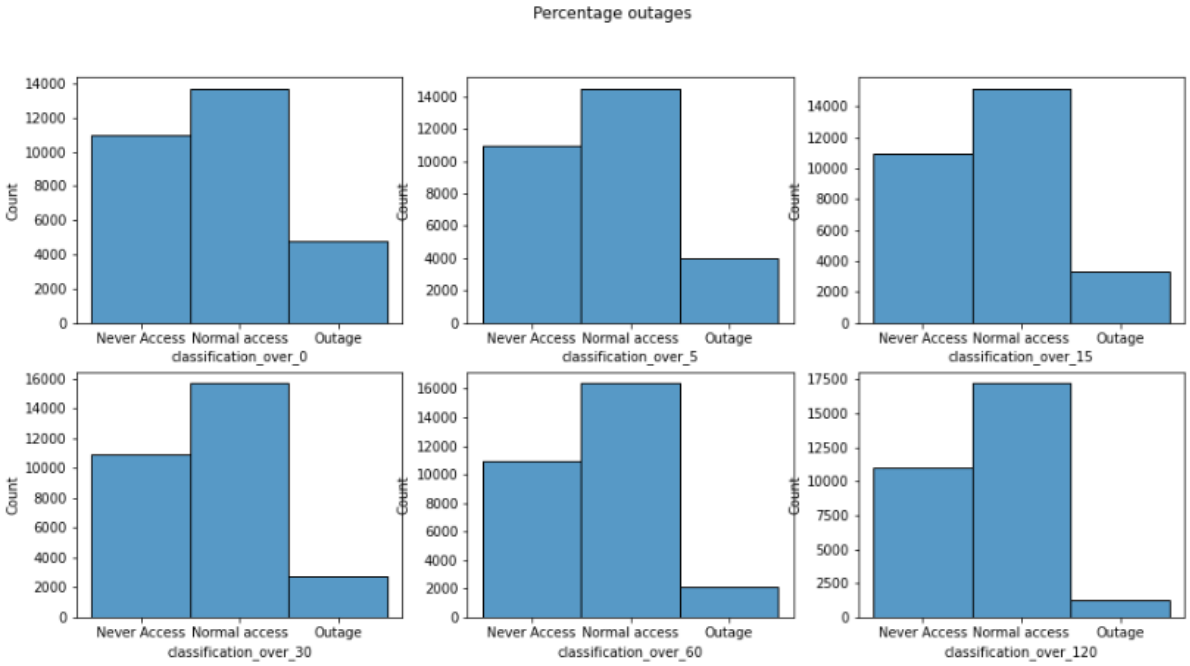


Figure 3.20: Distributions of the six classification variables in the target data. These distributions show a clear class imbalance present. The longer an outage has to last to be classified an outage, the bigger the class imbalance.

3.4.2. Validation

An RF can be validated in a variety of ways. A classic approach is to take the dataset and to divide it into a training set and a test set, then train the model on the training set and estimate the model error on the test set (Hastie et al., 2009, p. 222). Another approach is to divide the training set once more into a training set and a validation set. The validation set is utilized for model optimization and model selection procedures, while the test set is used to estimate the model's ultimate error rate.

K-fold cross-validation is another validation strategy (Hastie et al., 2009, pp. 241–249). The data is divided into K equal-sized subsets that are mutually exclusive and comprehensive. These subsets are referred to as folds. The model is trained and evaluated on the union of K-1 folds. For each fold, the procedure is repeated K times, and the model error rate is the mean of all repetitions. The leave-one-out validation method is a variation of K-fold cross-validation in which one sample of data is considered a fold, leading to a number of folds equivalent to the total of samples. This is a more computationally expensive strategy that is intended to produce more accurate findings. However, according to research, the improvement in accuracy over 10-fold cross validation is not usually substantial (Kohavi, 1995). Thus, in this research, 10-fold cross validation will be used.

3.4.3. Evaluation Metrics

Accuracy is defined as the proportion of correct predictions made to total forecasts made (Powers, 2020). It is calculated in most benchmarking studies, however it can be a deceptive and inaccurate statistic for unbalanced datasets, such as the one used in this study. The accuracy worsens as a result of this issue, which might obscure facts about a model's performance. For example, when a high accuracy score is obtained for an unbalanced dataset, it is not clear if this is because all classifications are predicted equally well or because certain classes are ignored by the model. A high accuracy score can be attained by constantly guessing the most common class value. Still, it gives a good first impression if the model is performing well. The accuracy is shown in equation 3.2, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. For multi-class problems, the TP, TN, FP, and FN are determined for each class separately and aggregated.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}$$

Precision is calculated by dividing the number of accurately classified members of a class by the number of times the algorithm predicted that class, shown in equation 3.3 (Powers, 2020). The precision score for Outages would be the number of properly detected Outages divided by the total number of instances the algorithm predicted Outages, correctly or incorrectly.

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

Recall is the number of samples of a class properly recognized by the algorithm divided by the total number of samples in that class, shown in equation 3.4 (Powers, 2020). In the case of Outages, this is the number of real Outages successfully recognized by the algorithm.

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

The F1 score is less apparent since it combines accuracy and recall into a single score, shown in equation 3.5 (Powers, 2020). F1 will be high if precision and recall both are high. F1 will be low if both are low. F1 will also be low if precision is high and recall is low, and vice versa. F1 is a fast technique to determine if the algorithm is truly competent at recognizing samples of a class, or whether it is merely looking for workarounds. This is why F1 is considered the most important metric in this study, as any sort of misclassification is unwanted.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{3.5}$$

The Cohen's Kappa coefficient (Cohen, 1960) is another measure of accuracy that reveals how well the algorithm performs as compared to random behavior (Landis & Koch, 1977). The Cohen's Kappa formula is shown in equation 3.6. Where p_0 is the proportion of actual agreement and p_c is the fraction of predicted agreements (Sim & Wright, 2005). Typically, the range is between 0 and 1. Perfect agreement is indicated by a 1, indicating that the raters agree on every case's classification. Zero represents no agreement, suggesting agreement that is no better than would be predicted by chance or random guessing. Cohen's kappa might be negative as well. The total performance of the algorithm would thus be even worse than what a random guess might have gotten. Landis and Koch (1977) present a method for characterizing values. According to their scheme, a value of 0 indicates no agreement, a value of 0–0.20 indicates slight agreement, a value of 0.21–0.40 indicates fair agreement, a value of 0.41–0.60 indicates moderate agreement, a value of 0.61–0.80 indicates substantial agreement, and a value of 0.81–1 indicates almost perfect agreement.

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}, \tag{3.6}$$

3.4.4. Random Forest Hyperparameters

As a reminder, RF is a straightforward and simple machine learning algorithm for tackling classification and regression problems. It is an ensemble approach that consists of numerous smaller decision trees that are trained on a randomized subset of the training data and make their own estimates, which are then aggregated to get a more precise estimate (Breiman, 2001).

RF is a non-parametric classification and regression algorithm, meaning that the set of parameters is not defined before training and the tree structures can closely match training data (Criminisi et al., 2012). To avoid overfitting and underfitting, the technique employs hyperparameters to modify the structure of the trees and forest, a process known as regularization (Probst et al., 2019). The regularization hyperparameters consist of bootstrapping, the number of decision trees, the maximum depth of a decision tree, the maximum number of features evaluated for splitting a node, the minimum amount of samples permitted in a leaf node, and the number of samples per tree. It is imperative to prevent both overfitting and underfitting.

Bootstrapping

The main concept of bootstrapping is to generate datasets using replacement from the training set at random, with each sample being the same size as the initial training set. This is repeated B times, yielding B bootstrap datasets. The model is then fit to everyone of the bootstrap datasets, and the behavior of the fits is examined throughout the B replications (Hastie et al., 2009, p. 249). In figure 3.21 an example of bootstrapping is given. Bootstrap sampling in action. Because objects are subsampled using replacement, some classes may be over-represented (yellow in bootstrap samples 1 and 2), while others may be under-represented (red in bootstrap samples 1 and 2) or even absent (green in bootstrap sample 3) (Galdi & Tagliaferri, 2018).

Figure 3.21: A bootstrap example. Because items are subsampled via replacement, some classes may be overrepresented (yellow circles in bootstrap samples 1 and 2), while others may be underrepresented (red circles in bootstrap samples 1 and 2), or even absent (green circles in bootstrap sample 3). (Galdi & Tagliaferri, 2018)

Maximum depth of a tree

The maximum depth of a tree refers to the maximum number of layers that a tree in the forest can have. The greater the value, the deeper the tree and the greater the number of splits (Criminisi et al., 2012). As the depth of the tree increases, so does the overall prediction accuracy. A high value for tree depth, on the other hand, tends to result in overfitting, which means that the trees become sensitive to noisy training data. The maximum tree depth parameter, in reality, regulates the amount of overfitting. Similarly, excessively shallow trees result in underfitting. As a result, while employing several trees reduces the overfitting issue of individual trees, it does not totally eliminate it. In practice, one must exercise extreme caution while determining the most appropriate value of depth, as the best value is determined by the problem's complexity.

Number of decision trees

Based on the results of their study on 29 datasets, Oshiro et al. (2012) reveal that the best number of

trees to utilize inside RF ranges between 64 and 128. Increasing the number of trees above that did not substantially increase effectiveness and primarily raised the computational cost. However, in order to discover the appropriate number of trees and achieve optimal performance, the dataset's properties must also be taken into account. Typically, the best results are obtained with the first 100 trees (Probst & Boulesteix, 2017). Wright and Ziegler (2017) also showed that the computing time of RF grows linearly with the number of trees.

Maximum number of features evaluated for splitting a node
Bernard et al. (2009) describe that this number is an integer with values ranging from 1 to M, where M represents the number of features in the input features vector. The lower the value of K, the more randomness there is in the feature selection process. If K is equal to one, just one feature is chosen at random when the tree structure is split. If, on the other hand, K = M, all accessible features are employed to create each tree, and no randomization is added. The value of the K parameter affects the reliability and accuracy of the decision trees as well. Low K values result in less correlated trees, which are more robust when aggregating and utilize less significant characteristics more effectively (Probst et al., 2019). Geurts et al. (2006) have demonstrated via testing that placing the K value equal to $sqrtM$ is usually close to the ideal value. However, Bernard et al. (2009) presented another option for the K parameter, $log2(M) + 1$, in their experimental research, which was then utilized in further alongside $sqrtM$ and were both proven to yield accurate outcomes and sometimes perfectly match to the optimal parametrization of K. However, these trials also demonstrated that the global relevance of the characteristics should be considered when determining the value of the K parameter. Specifically, if there are too many irrelevant and unimportant characteristics, randomization will rapidly reduce tree accuracy, but having too many features with significant information enables overfitting of trees and diminishes the randomization benefits in split selection. Finally, another study demonstrated that decreasing the K value reduces the computing time of the RF algorithm practically linearly (Wright & Ziegler, 2017).

Minimum amount of samples permitted in a leaf node
The node size hyperparameter influences tree complexity by providing the minimal number of samples stored in leaf nodes for predictions (Criminisi et al., 2012; Probst et al., 2019). The default value for this hyperparameter in sklearn, 1, often produces satisfactory results (Buitinck et al., 2013). However, Segal (2004) demonstrated that if the data predictors are noisy and higher number of features examined for splitting a node work well, then the ideal node size to maximize performance would be larger. Furthermore, increasing the node size reduces the tree complexity, which generally results in a significant reduction in computation time.

Number of samples per tree
The number of samples needed to train the trees is determined by the sample size hyperparameter. Low values improve the diversity of the decision trees, but they also diminish their predictive performance. The appropriate sample size is determined by achieving the best trade-off between tree stability and accuracy (Criminisi et al., 2012).

3.4.5. Sampling Design

Commonly, data for RF is split into a training and testing set. The data in the training set is meant for the algorithm to learn from, and the test set is meant to test how well the algorithm performs on unseen data. RF is a sampling-sensitive algorithm that needs specific requirements to be met in order to get accurate performance (Belgiu & Drăguţ, 2016). The classes should be balanced in terms of the number of samples in each class. As shown in the previous section, there are unbalanced classes in this problem. It is important that data is independently and identically distributed (IID), and a helpful assumption for statistical reasoning. A typical train-test split can be used to predict the risk of diabetes in a random group of individuals. However, there are some datasets where this assumption of IID does not hold true. Normal train-test split can result in data leakage.

Stratified sampling is a strategy that assures increased success rates in issues with numerous classes and an imbalanced class distribution in the input data (Dekanovsky, 2021). It involves assigning samples from a class to the training and testing set in the same distribution as in the original

dataset. This solves the imbalance of the target data. However, the predictor classes may as well be imbalanced.

Under- and Oversampling

Stratified sampling might not be enough to get accurate results, though. Resampling data might be an option to increase model performance for the minority classes. However, it is important that resampling is best conducted within the cross-validation process (Santos et al., 2018).

A common technique is to use oversampling of the minority class. The most common method is Synthetic Minority Oversampling technique (SMOTE) (Bowyer et al., 2011; Rok & Lusa, 2013). SMOTE has shown to be effective in a wide range of applications across several fields (Fernández et al., 2018). SMOTE generates synthetic samples from a minority class using the K-nearest-neighbor technique, shown in figure 3.22. As a result, it solves the overfitting problem caused by random oversampling. SMOTE generates instances in feature space that are near to each other by employing interpolation between positive examples that are close to each other. It chooses a minority class instance at random and discovers its K closest neighbors. Then it generates synthetic models by selecting one of the neighbors at random and forming a line segment in the feature space, SMOTE can be implemented by using the Python library Imbalanced-Learn (imb-learn) (Lemaître et al., 2017).

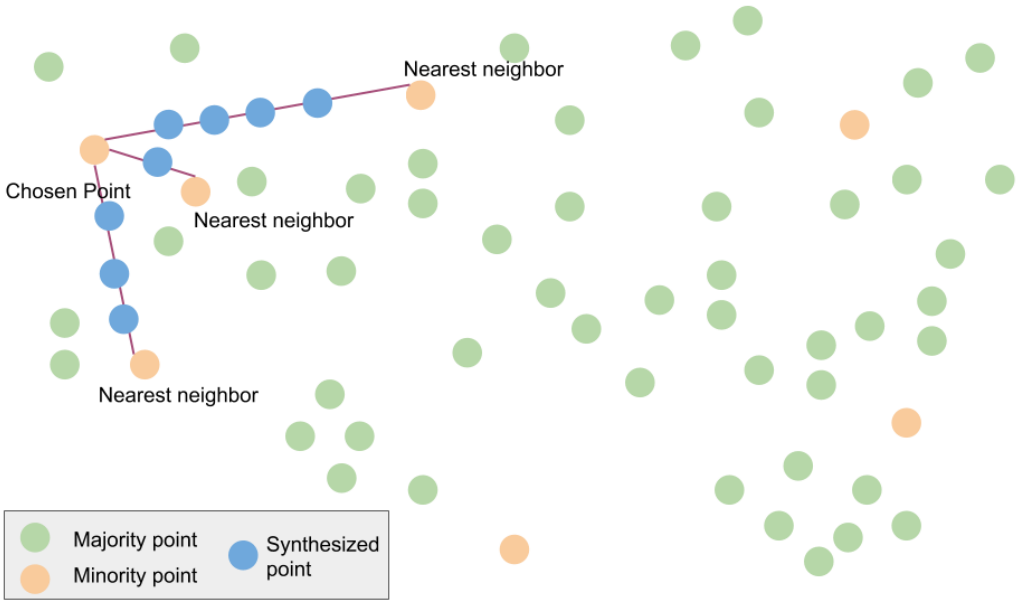


Figure 3.22: Example of SMOTE algorithm, in this example $k = 3$. The three nearest neighbors of the minority class are chosen and between the chosen point and neighbors new samples are synthesized (Huh, 2021).

Tomek undersapmling is a technique for under-sampling. The technique discovers what are known as Tomek's links (Tomek, 1976). If two samples of different classes are nearest neighbors to one another, a Tomek's link exists. A Tomek's link between two samples of different class x and y is defined such that for any sample z, shown in equation 3.7. A Tomek's link is depicted in the figure 3.23, highlighted in green. Once identified, the minority sample, majority sample, or both can be removed. In this research, the removal of the majority sample will be tested, because the minority samples should be kept in the analysis as much as possible.

$$d(x,y) < d(x,z) \text{ and } d(x,y) < d(y,z)" > \tag{3.7}$$

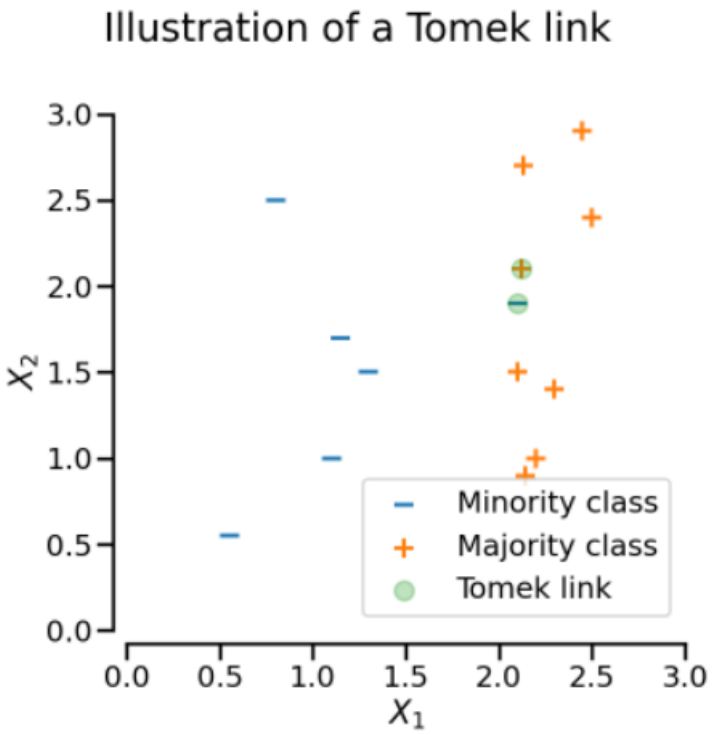


Figure 3.23: Illustration of Tomek link, a Tomek's link exists when two samples of different classes are nearest neighbors. The samples of interest are highlighted in green to represent a Tomek's link. (Lemaître et al., 2017)

SMOTE was previously presented, however, by interpolating additional points between marginal outliers and inliers, this approach can yield noisy samples (Lemaître et al., 2017). This problem can be resolved by clearing up the over-sampled area. In this sense, Tomek's link may be utilized as a cleaning approach that can be added to the pipeline after SMOTE over-sampling is used to get a cleaner region. This can be easily done using imb-learn's SMOTETomek function.

Spatial and Temporal Sampling

An important class feature to take into account is the geography of the samples. As mentioned in section 2.6, end-use energy injustice, like other types of inequality, is also a fundamentally geographical phenomenon (Bouzarovski & Simcock, 2017). When data is autocorrelated, extra caution is required when it comes to overfitting. The IID assumption is breached if random samples for train-test splits are utilized in this situation, as such samples are not statistically independent (Hengl et al., 2018). For example, Location X might be in the training set, but location Y in the test set, which might only be a few kilometers distant and has very comparable properties to location X. Because it encountered a highly comparable case in the training set, the algorithm would make a more accurate prediction for location Y.

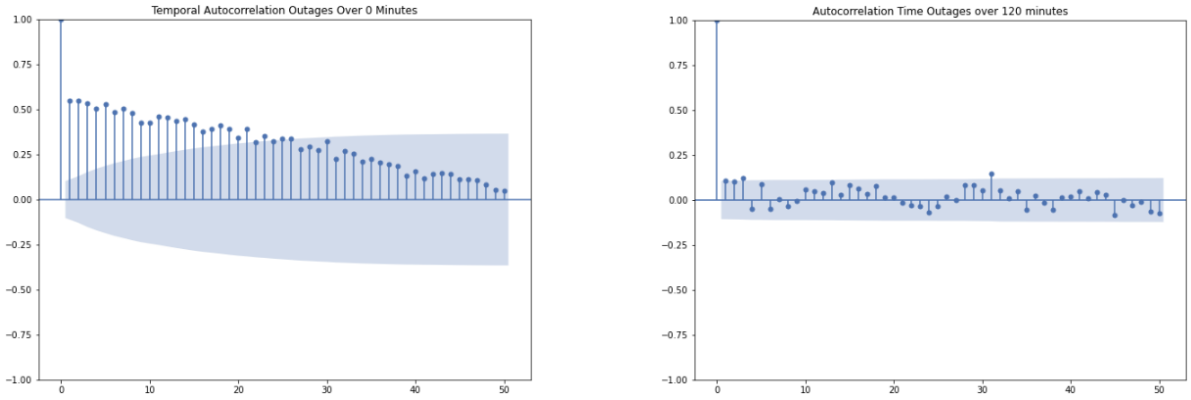
An important issue to address here is whether it is good to always avoid overfitting. It depends, overfitting may even be useful if it matches the use case. Assume there is a nationwide wealth survey with a random sample. There are wealth values for a dispersed group of families around the nation, and the need to infer wealth levels for unsurveyed locations in order to obtain comprehensive wealth data for the whole country. The purpose here is just to fill up spatial voids (Hengl et al., 2018). Training using data from nearby places would undoubtedly help fill in the gaps more correctly. However, it is a different matter if when attempting to create a generalizable algorithm, like in this research. The algorithms might be applied to a different state or country entirely. In this scenario, taking advantage of the spatial autocorrelation trait during training will most certainly increase the accuracy of a possibly weak algorithm. This is especially troubling if we apply this supposedly good algorithm to an area in which there is no ground truth to check (Meyer et al., 2018).

From figures 3.4 and 3.18 it becomes clear that the ESMI sensor locations are not spatially diverse.

In a lot of states, only non-access locations are sampled. The distribution of Never Access, Normal Access, and Outage should be somewhat preserved in the sampling, otherwise the algorithm will get biased during the learning phase. To do this, spatial sampling will be performed based on the individual locations of the samples and manually checked.

Another important thing to account for is the autocorrelation between dates. Time series problems are also sensitive to the sequence of the samples. It is typically far easier to forecast the past based on knowledge further in time than it is to anticipate the future. As a result, older data is often put into time series models for forecasting newer data. However, does this also hold true for the ESMI data?

To explore this, the temporal autocorrelation of the amount of outages is calculated. This is done with the help of the Python library Statsmodels (Seabold & Perktold, 2010). To do this the autocorrelation the correlation coefficient is calculated. Correlation of a time series is discovered by comparing it with a lagged version of itself, rather than between two characteristics. In figure 3.24 the temporal autocorrelations of outages over 0 and 120 minutes are presented. This image needs some explaining. The blue shaded area represents the confidence interval, which has a default value of 0.05. Anything in this range reflects a value that has no substantial relationship with the most recent amount of outages value. The vertical lines with marks at their tops are the ”lags,” which signify a specified number of prior dates (50 in this example). These indicate the correlation coefficient (shown on the y-axis) and decrease at a constant rate as the distance from the date rises. In figure 3.24a it shows that there is a significant correlation until day 26 approximately. However, no such correlation can be found for outages over 120 minutes, shown in figure 3.24b, except for one outlier at around 31 days. In appendix B, the temporal autocorrelations of all the classification variables are presented. The temporal autocorrelation seems to be more pronounced for shorter outages than for longer ones.



(a) Temporal autocorrelation outages over 0 minutes

(b) Temporal autocorrelation outages over 120 minutes

Figure 3.24: Temporal autocorrelation outages of over 0 and 120 minutes. The blue shaded area represents the confidence interval, which has a default value of 0.05. Anything in this range reflects a value that has no substantial relationship with the most recent amount of outages value. The vertical lines with marks at their tops are the ”lags,” which signify a specified number of prior dates (50 in this example). These indicate the correlation coefficient (shown on the y-axis) and decrease at a constant rate as the distance from the date rises.

Figure 3.25 shows a diagram with different validation strategies using 3-fold sampling; random cross-validation (CV), Leave-Location-Out (LLO), and Leave-Time-Out (LTO) (Meyer et al., 2018). The initial data in this example are on three data loggers, as shown by color, each measured at three points in time (t), as represented by shape.

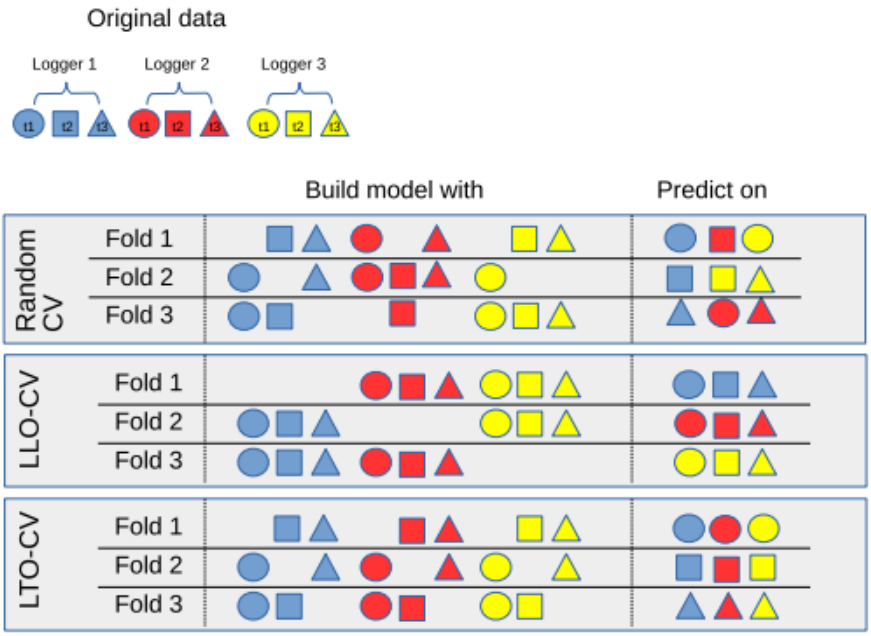


Figure 3.25: Schematic overview of the validation 3-fold strategies; random cross-validation (CV), Leave-Location-Out (LLO), and Leave-Time-Out (LTO) (Meyer et al., 2018).

3.4.6. Feature Engineering and Selection

Feature engineering is either selecting a subset of the original features (feature selection), combining original features to create new ones (feature extraction), or gathering more data for new feature calculation (Heaton, 2016).

Feature extraction

For all the daily features, two extra features are calculated; the median DNB value for the specific pixel, and the difference between the daily measurement and the median value for that pixel. This is done based on work by Mann et al. (2016). This gives the algorithm extra information to detect abrupt changes for a specific pixel. In table 3.2 all the features are presented. In total 17 features will be used.

Table 3.2: Features for RF

Features	Median	Difference daily measure and median
nighttime_radiance	median_radiance	diff_radiance_median
temperature	median_temperature	diff_temperature_median
windspeed	median_windspeed	diff_windspeed_median
precipitation	median_precipitation	diff_precipitation_median
AOT_MERRA	median_AOT_MERRA	diff_AOT_MERRA_median
land_cover	x	x
population	x	x

Feature selection

The process of scanning the feature space for the optimal subset for data classification is known as feature selection. Because the number of possible subsets grows exponentially with the size of the set of features, exhaustive search is impossible, and multiple search methods with varying levels of complexity, such as forward selection and backward elimination, best-first search, beam search, and genetic algorithms search, have been proposed (V. Kumar & Minz, 2014).

In this research, two sklearn methods will be used; recursive feature elimination and recursive feature elimination with cross-validation (Pedregosa et al., 2011). The purpose of recursive feature

elimination (RFE), a form of backward eleimination, is to pick features by iteratively examining smaller and smaller sets of features given an external estimator that gives weights to features. To begin, RF is trained on the original set of features, and the significance of each feature is determined using feature importance. The least significant features are then removed from the present list of features. This technique is continued recursively on the trimmed set until the appropriate set of features to pick is attained. RFE in a cross-validation loop to determine the best amount of features is called RFECV.

Consequences in practice, SMOTE can be troublesome for classifiers that presume sample independence (Rok & Lusa, 2013). After employing SMOTE, it is important to perform feature selection with caution because most feature selection methods presume that the samples are independent. Thus, in this research, the feature selection will be performed before the different sampling strategies.

3.5. Outcomes

The most immediate result is a map that displays at which rate the classes are assigned to each pixel across Uttar Pradesh. This outcome is possible because for every date in 2018 classifications are made. The map does not provide direct indication of how confident the algorithm is in the classifications.

3.6. Conclusion Methods

This chapter described the classification procedure, beginning with data preparation for classification. The Random Forest classification method was then introduced. Because the class distribution is imbalance, resampling approaches such as SMOTE, Tomek, and SMOTETomek were developed to compensate the imbalance. Because this is a geographical study that spans a whole year, spatial and temporal cross-validation was used to provide an accurate performance indication. The extraction and selection of features were then addressed. Finally, the classification method will be utilized to generate an outage variation map for Uttar Pradesh that will aid in the assessment of electricity reliability. The results will be presented in the next chapter.

4
Results

The outcomes of the study's experiments are presented in Chapter 4. More precisely, the collected findings from the experiments are provided and examined in the following sections to understand how different characteristics, hyperparameters, data sizes, and data sampling might impact the performance of the algorithms. We talk about the limitations we encountered and which techniques and parameters we think are the best. Finally, comparisons between the results and the survey results are made, and an application is provided.

4.1. Hyperparameter Tuning

For the tests with the RF hyperparameters, the default hyperparameters given by sklearn were utilized (Pedregosa et al., 2011). The settings for one of the hyperparameters, described in section 3.4.4, were adjusted each time, controlling for the others, to see what influence they had on the behavior of the model. The default settings for the hyperparameters that we examined are shown in table 4.1 and the tuning settings are shown in table 4.2. The metrics, described in section 3.4.3, were utilized as the assessment tool to compare all categorization outcomes.

Table 4.1: Default hyperparameter settings

Hyperparameter	**Value**
Number of trees (n_estimators)	100
Max depth	None
Min samples split	2
Min samples leaf	1
Max features	sqrt
Bootstrap	True

Table 4.2: Hyperparameter tuning settings

Hyperparameter	**Value**
Number of trees (n_estimators)	50, 100, 150, 200, 250, 300, 350, 400
Max depth	5, 10, 15, 20, 25, 30, 35, 40, 45, 50
Min samples split	2, 4, 8, 10, 20, 40, 80, 160, 240, 320
Min samples leaf	1, 2, 4, 8, 10, 15, 20, 40, 60, 80, 100, 120
Max features	1, 2, 4, 8, 10, 12, 14, 16
Bootstrap	True, False

The hyperparameter settings used for the next sections of the research are presented in table 4.3. How these values were chosen is described in appendix C.

4.2. 10-Fold vs. Stratified 10-Fold

The confusion matrixes in figures 4.1a and 4.1b capture the RF's capacity to detect individual, household-level electrical outages for the known validation data set. The performance is nearly the same for both validation strategies. It seems that normal 10-fold cross-validation captures the distribution of the targets the same as the stratified sampling does. This can be more easily deduced from tables 4.5, 4.5,

Table 4.3: Final hyperparameter settings

Hyperparameter	Value
Number of trees (n_estimators)	100
Max depth	10
Min samples split	8
Min samples leaf	4
Max features	sqrt
Bootstrap	True

and 4.6. The scores are almost exactly the same for both the cross-validation strategies.

Furthermore, RF seems to be able to classify Never Access perfectly. It has a bit more trouble with Normal Access and Outage, though. The algorithm has a tendency to classify more Outages as Normal access than actual Outages. While in 69% of the cases when RF classifies a pixel as an Outage it is correct, it actually misclassifies more than half of the Outages as Normal Access, only 44% of the Outages are correctly classified by the algorithm. Looking at the Cohen's kappa scores, the overall performance of the algorithm is quite high, with a score of 0.79, it is in the absolute upper range of substantial agreement, nearing the range of almost perfect agreement. However, the minority class of Outages has a smaller impact on this overall score than Never Access and Normal Access. That is why it is important to look at the individual scores as well. Still, a score of 0.43 indicates moderate agreement.

The performance for the other target variables, introduced in section 3.4.1, are shown in appendix D. These results show that the performance for Outage recognition decreases with the higher classifications of outages. Also, the scores are almost exactly the same for both the cross-validation strategies for each of the target variables. Because of the decrease in performance for the other target variables, it was decided to report only on the target variable *classification_over_0* for the remainder of the research. Unless, unexpected results are found for the others.

Table 4.4: Scores 10-Fold cross-validation

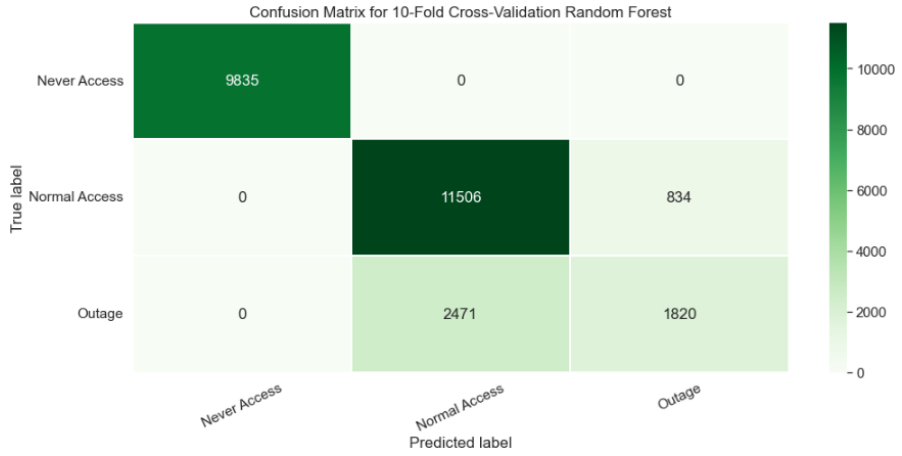
Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.94	0.88
Outage	0.69	0.44	0.54

Table 4.5: Scores stratified 10-fold cross-validation

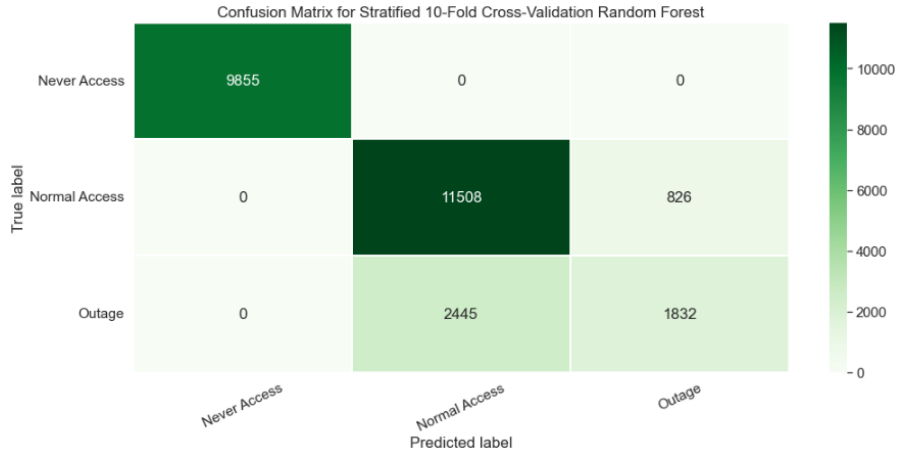
Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.93	0.87
Outage	0.69	0.44	0.53

Table 4.6: Cohen's kappa scores 10-fold and stratified 10-fold cross-validations

Class	10-Fold	Stratified 10-Fold
Overall	0.79	0.79
Never Access	1.00	1.00
Normal Access	0.91	0.91
Outage	0.43	0.43



(a) Confusion matrix 10-fold cross-validation



(b) Confusion matrix stratified 10-fold cross-validation

Figure 4.1: Confusion matrixes 10-fold and stratified 10-fold cross-validations

4.3. Feature Importance

In figure 4.2 the importance of the individual features and their standard deviations are plotted for the 10-fold cross-validation. It is very obvious that feature *median_radiance* is the most important feature, followed by population. It is also obvious that some features have little to no importance for the algorithm, especially the daily features seem to be of little value to the algorithm. However, since the dataset is imbalanced, some features might be important for the minority class, but get overshadowed by others when it comes to the overall importance. To methodically choose the features that will be used for further experimentation, RFE and RFECV, as explained in section 3.4.6, will be used.

RFE

When the RFE was applied, only 9 features were deemed important enough; *temperature, nighttime_radiance, land_cover, population, median_radiance, median_temperature, median_windspeed, median_precipitation, median_AOT_MERRA*. Once these features were selected, another round of hyperparameter tuning was performed, shown in appendix E, in order to get the best performance out of the new algorithm. The values chosen for the RFE features algorithm are shown in table 4.7

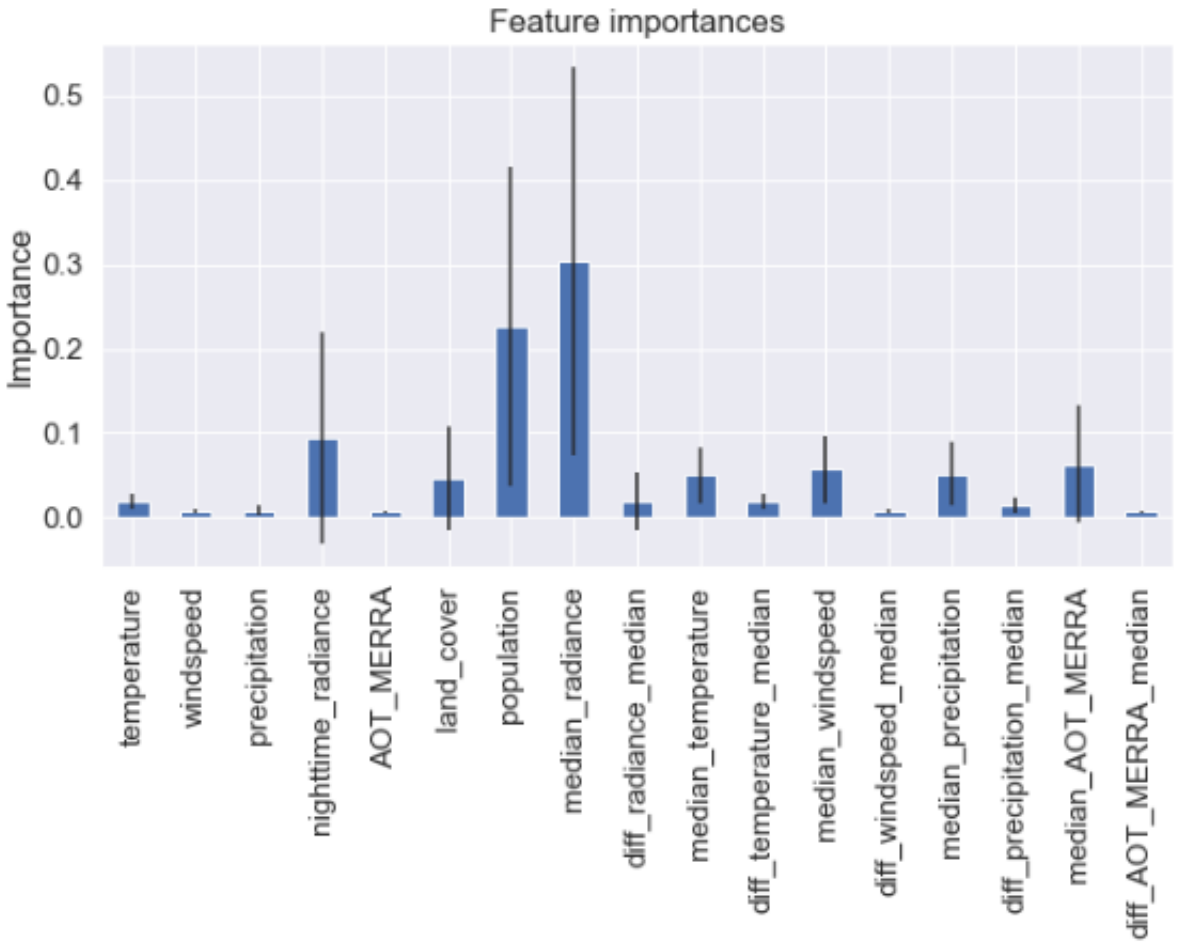


Figure 4.2: Feature importance for 10-fold cross-validation with 19 features

Table 4.7: Hyperparameter settings RFE features

Hyperparameter	**Value**
Number of trees (n_estimators)	100
Max depth	8
Min samples split	160
Min samples leaf	20
Max features	sqrt
Bootstrap	True

RFECV

When the RFECV was applied, 16 features were deemed important enough; *temperature, windspeed, nighttime_radiance, land_cover, population, median_radiance, diff_radiance_median, median_temperature, diff_temperature_median, median_windspeed, diff_windspeed_median, median_precipitation, diff_precipitation_median, median_AOT_MERRA, diff_AOT_MERRA_median*. It is curious that the RFECV method includes more features than RFE. Even features like *diff_windspeed_median and diff_AOT_MERRA_median* are still included, even though, the feature importance graph shows almost no importance for these features. Again, another round of hyperparameter tuning was performed, shown in appendix E, in order to get the best performance out of this algorithm as well. The values chosen for the RFECV features algorithm are shown in table 4.8.

Table 4.8: Hyperparameter settings RFECV features

Hyperparameter	Value
Number of trees (n_estimators)	100
Max depth	8
Min samples split	200
Min samples leaf	60
Max features	sqrt
Bootstrap	True

Tables 4.9 and 4.10 show the scores for the cross-validation with the RFE and RFECV features only. In table 4.11 the Cohen's kappa scores are presented. The performance of both RFE and RFECV are very similar to the performance of all features. Never Access still gets perfectly classified in both subsets of features. RFECV also has the same scores for Normal Access classification, while the other scores for the other classes are not exactly the same as when all features are used, RFE and RFECV do not seem to underperform. Looking at the Cohen's kappa scores for the Normal Access and Outage, it is also apparent that the performance of the algorithms stayed on par for this pair.

Table 4.9: Scores 10-fold cross-validation RFE features

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.83	0.92	0.87
Outage	0.68	0.45	0.54

Table 4.10: Scores 10-fold cross-validation RFECV features

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.94	0.88
Outage	0.69	0.42	0.52

Table 4.11: Cohen's kappa scores RFE and RFECV

Class	RFE	RFECV
Overall	0.79	0.79
Never Access	1.00	1.00
Normal Access	0.90	0.92
Outage	0.44	0.41

4.4. Sampling Results

4.4.1. Under- and Oversampling

Tables 4.12, 4.13, and 4.14 show the scores for the cross-validation with the SMOTE, Tomek, and SMOTETomek resampling strategies using all available features. In table 4.15 the Cohen's kappa scores are presented. Just like all previous results, Never Access is perfectly classified. However, the results for show a clear increase in the F1 score for Outage classification compared to the results from section 4.2 where no resampling was performed. Specifically, the Recall score increases significantly. This does come at a cost of Precision, but the difference is limited compared to the increase in Recall. Tomek resampling shows a smaller difference on its own compared to the results from section 4.2. Interestingly enough, the combination of SMOTE and Tomek does not result in new results, as SMOTE and SMOTETomek perform practically the same. Following this, it can be concluded that SMOTETomek superfluous to perform, as it is more computationally complex and the results are practically the same as SMOTE. Taking a closer look at the Cohen's kappa scores, it seems that the resampling does not increase overall performance. This is probably due to the algorithm misclassifying Normal Access as Outage a lot more than previously, so the overall performance against random performance did not increase, however, it did reach a better balance between Normal Access and Outage classifications. Appendix F presents the resampling results of the features selected in the previous section. The results show the same pattern as for the resampling techniques as using all the features; higher recall, but lower precision.

Table 4.12: Scores SMOTE resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.88	0.76	0.82
Outage	0.51	0.71	0.60

Table 4.13: Scores Tomek resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.84	0.91	0.87
Outage	0.65	0.51	0.57

Table 4.14: Scores SMOTETomek resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.88	0.86	0.81
Outage	0.50	0.71	0.59

Table 4.15: Cohen's kappa scores SMOTE, Tomek, SMOTETomek

Class	SMOTE	Tomek	SMOTETomek
Overall	0.75	0.80	0.75
Never Access	1.00	1.00	1.00
Normal Access	0.71	0.88	0.70
Outage	0.70	0.50	0.70

4.4.2. Spatial and Temporal Sampling

Tables 4.16 and 4.17 show the scores for the cross-validation with spatial and temporal sampling strategies using all available features. In table 4.18.

First, taking a closer look at the results for spatial sampling, Never Access is not classified perfectly, unlike in all the previous results. It appears that it is much more difficult for the algorithm to recognize a Never Access pixel if it has never seen that location during the training. Also, Outage recognition performance has diminished quite considerably. The F1 scores in the previous results for Outage never went below 0.50, however, using spatial sampling the score only reaches 0.19. This can be seen in the Cohen's kappa score as well, where it only has a score of 0.11, meaning that it only slightly performs better than random guessing. The overall Cohen's kappa score has dropped to 0.60, which is quite a fall from 0.79. Figure 4.3a shows the confusion matrix for the spatial sampling cross-validation to show how the algorithm misclassifies Never Access and Outage more often as Normal Access than in the previous results. This clearly shows that the algorithm actually performs quite poorly, unlike the results from the normal 10-fold and stratified 10-fold cross-validation would seem.

Second, looking at the results for temporal sampling, the difference in performance compared to the normal 10-fold cross-validation is not as staggering as for the spatial sampling. However, while Never Access locations are classified perfectly, there is a significant drop in performance for Outage recognition as well, while the performance for Normal Access only decreases slightly. Because Normal Access is so much more prominent in the data set, the effects of misclassifying the minority class Outage as Normal Access has little effect on the scores for Normal Access. The effect of temporal sampling is also clearly shown in the Cohen's kappa score for Outage, dropping from 0.43 for normal 10-fold cross-validation to 0.33 for temporal 10-fold cross-validation.

In the previous section, it was shown that the resampling techniques can lead to higher performance for Outage recognition when all features are used, and Tomek performed better for the subset of features using RFE and RFECV. Thus, these resampling techniques will also be employed for the spatial and temporal sampling strategies. In appendix G the results of this are shown. The results of resampling show the same pattern as before. The performance for Outage increases drastically for SMOTE and SMOTETomek, but does come at a cost of the performance for Normal Access. While

Tomek does increase the performance for Outage, without decreasing the performance for Normal Access, although only limited. An interesting thing to notice is that the subsets of features selected by RFE and RFECV perform better for Outage recognition than when all features are used for no resampling for both spatial and temporal cross-validation. RFE and RFECV perform very closely to all the features using SMOTE and SMOTETomek for spatial cross-validation, while they actually perform as good or better for Outage recognition using Tomek for both spatial and temporal cross-validation. Still, the overall Cohen's kappa score is around the 0.60, indicating moderate agreement of the algorithms with the actual measurements.

Table 4.16: Scores spatial cross-validation

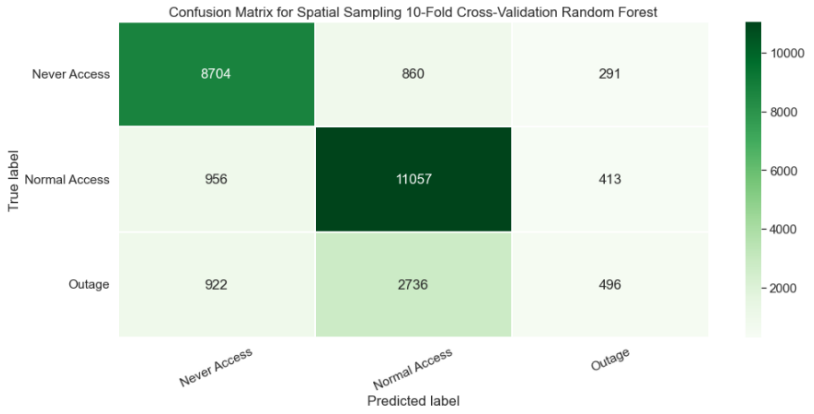
Class	Precision	Recall	F1
Never Access	0.82	0.88	0.85
Normal Access	0.75	0.89	0.82
Outage	0.41	0.12	0.19

Table 4.17: Scores temporal cross-validation

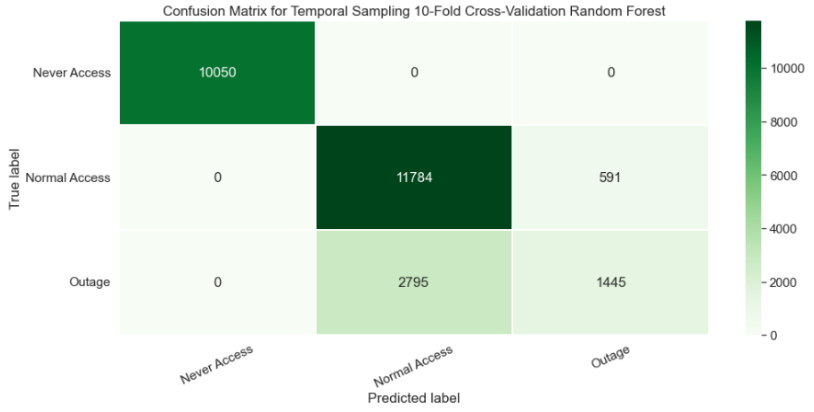
Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.81	0.95	0.87
Outage	0.71	0.34	0.46

Table 4.18: Cohen's kappa scores spatial and temporal sampling

Class	Spatial	Temporal
Overall	0.60	0.79
Never Access	0.86	1.00
Normal Access	0.85	0.94
Outage	0.11	0.33



(a) Confusion matrix for spatial sampling 10-fold cross-validation



(b) Confusion matrix for temporal sampling 10-fold cross-validation

Figure 4.3: Confusion matrixes spatial and temporal 10-fold cross-validations

4.4.3. Original Data

As an extra experiment, the spatial and temporal sampling strategies were once more employed, but now for only the input directly from the data sources; *nighttime_radiance, temperature, windspeed, precipitation, AOT_MERRA, land_cover, population*. The results of this are presented in appendix H. The results show that without the median variables, RF cannot identify Never Access samples perfectly anymore for the normal 10-fold and temporal 10-fold cross-validation, although it is still quite high. What is interesting to notice is that RF performs much better on the normal 10-fold cross-validation than spatial and temporal cross-validation, indicating that even without the easy identifying tags, in form of the median values, it is still very important to perform spatial and temporal cross-validation to get an accurate idea of the performance of the algorithm. Another interesting thing to notice is that the performance of RF using only original data for spatial cross-validation with the resampling techniques applied is nearly the same as when all features are used. However, comparing the results to the results found using RFE, the performance is slightly worse. Thus, this mean that while not all features created as described in 3.4.6 are useful, the median values are.

4.5. Unmonitored Areas

Finally, the RF algorithm with the RFE selected features is used to present a more complete view of outage variation over the year 2018 across the whole state of Uttar Pradesh, beyond locations where ground truth data is available. The results of the classifications for 2018 between 00:00 and 02:00 are shown with a spatial granularity of 0.1° x 0.1°. Finer granularity would be better, but due to computational constraints this was not possible.

Where three-part compositions are provided by geographical location or other ordered pairs of variables, the issue of displaying ternary compositions on a surface, such as the Earth's surface, arises. The ternary balancing scheme is a color scale that is appropriate for that purpose (Schöley, 2021). The approach encodes the respective rates of three classifications as a three-color combination. This figure is made using the R package Tricolore (Schöley, 2018). In figure 4.4, the variation of outages across Uttar Pradesh for the year 2018 are presented. Normal Access is mapped to yellow, Outage to cyan, and Never Access to magenta. The more pronounced the yellow in a region, the higher the rate of Normal Access. The same reasoning applies to the other two groups. The more gray a place appears, the more balanced the three ratios are, with dark gray representing an equal rate of all three groups. A ternary diagram is utilized as a color key, see top right. The diagram grid is made up of lines that run parallel to the triangle edges. A parallel to a triangle side is the constant set of points in the component located at the vertex opposite the side (Stover, 2021). Each class is 100% in a triangle corner and 0% on the opposing edge, decreasing linearly with increasing distance (perpendicular to the opposite edge) from the corner. Fine divisions can be produced for simple estimate by drawing parallel lines at regular intervals between the zero line and the corner, in this case intervals of 20%.

The pixels that are classified as Never Access at a rate of 80% or more of the year are most likely to be actual locations without access. Looking at the figure, it would seem that large parts of Uttar Pradesh never have access to electricity, specifically the south. Adding to this, large parts of the state are colored green, indicating between outages happen between 40% to 60% of the days in 2018. Furthermore, the algorithm classified some pixels as Outage 100% of the year, specifically in the north-east. Looking at the figure, it seems that there are only a few pixels that have 80% or more days without outages. The yellow pixels primarily in or near big cities in Uttar Pradesh. Furthermore, it can be seen that Normal Access is mainly recognized on the pixels that are never classified as Never Access, as there are only a few pixels with a brownish color. There are a few pixels north of Bareilly where the algorithm cannot decide between Normal Access with Never Access.

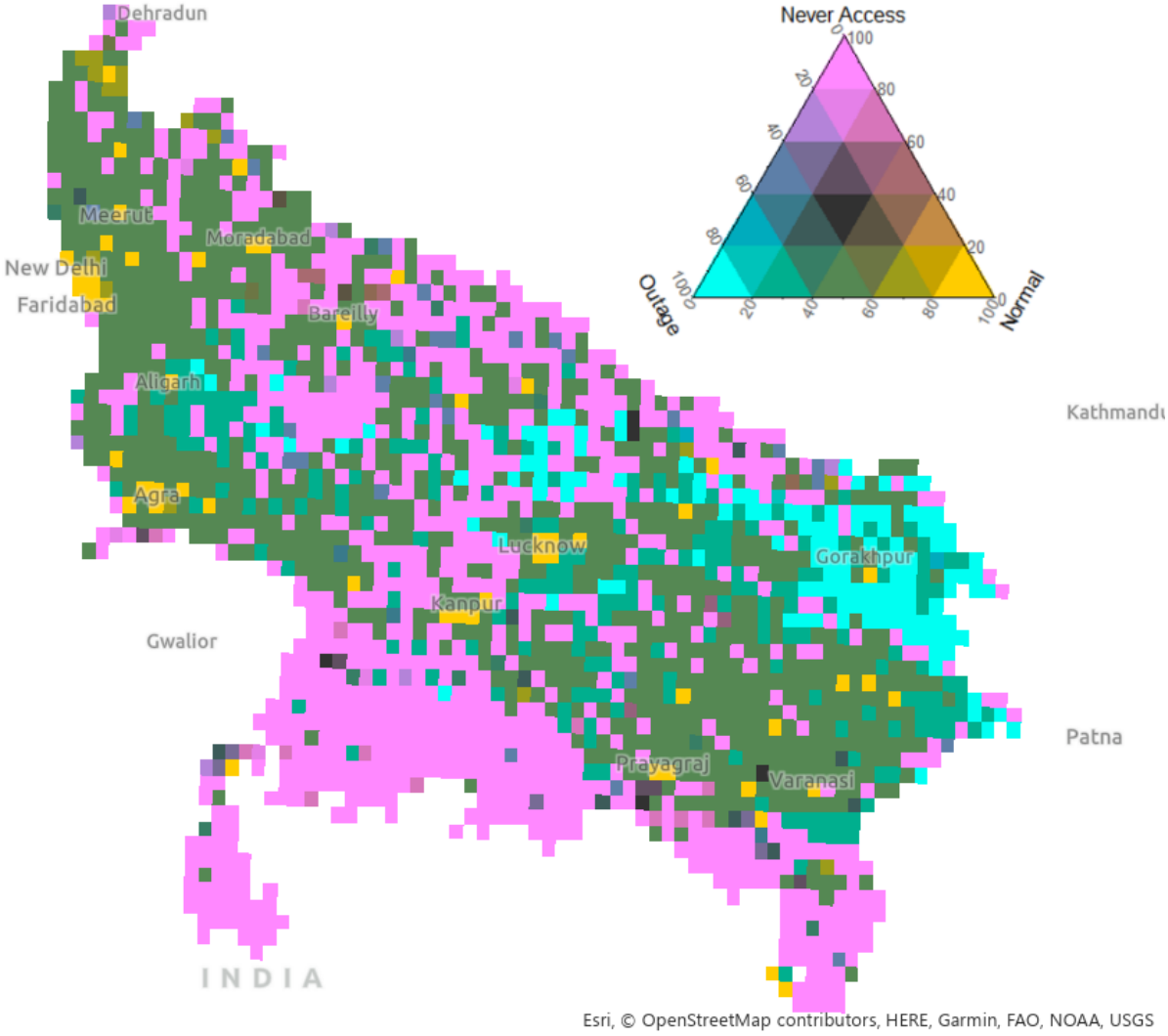


Figure 4.4: Outage variation in Uttar Pradesh for 2018. A ternary diagram is utilized as a color key, see top right. The diagram grid is made up of lines that run parallel to the triangle edges. A parallel to a triangle side is the constant set of points in the component located at the vertex opposite the side (Stover, 2021). Each class is 100% in a triangle corner and 0% on the opposing edge, decreasing linearly with increasing distance (perpendicular to the opposite edge) from the corner.

5
Discussion

This chapter looks at data availability, methodological challenges, and outcomes in a broader context. It begins with a description of what data was utilized and what factors impact data selection. The approach is then outlined, as well as what must be addressed before utilizing specific algorithms and how to train. Then there is a discussion on how the findings are shown and the challenges with accuracy.

5.1. Algorithm results

In this section, the performance of the RF will be examined and discussed.

5.1.1. Faulty Optimism

When examining the results of the normal and stratified 10-fold cross-validations, it is very peculiar to see the algorithm identifying Never Access samples perfectly. RF had a bit more difficulty with distinguishing Normal Access and Outage samples. Still, Normal Access samples were classified with high precision and recall. However, because Outage is a minority class it is easy for an algorithm to just classify every Outage as Normal Access and the scores would still be quite high. In this research, the algorithm was mostly judged on how it performed on the classification of Outage samples, because it is thought that accurately identifying Outage samples is more important than correctly identifying Normal Access or Never Access samples. A misclassification of an outage is a costly error for both users who may suffer from an unreliable electricity connection and researchers or NGOs who are unable to discover proper patterns regarding the network's weaknesses. RF was able to identify almost half of the Outage samples as such, with a precision of 69%. On one hand, this could be seen as a bad results if the goal is to identify all the outages, on the other hand, without any extra costs, almost half of household outages can be identified using readily available data. The Cohen's kappa gives an indication of how well the algorithm performs as compared to random behavior (Landis & Koch, 1977). The overall kappa indicates substantial agreement, nearing the range of almost perfect agreement. However, the minority class of Outage has a smaller impact on this total score than Never Access and Normal Access. That is why it is important to look at the individual scores as well. Still, the score indicates moderate agreement. Thus, the algorithm is performing quite well. However, perfect identification using machine learning is almost always a sign of data leakage. It is important to take every precaution to prevent this.

5.1.2. Importance of Sampling Design for Cross-Validation

The premise of doing this research is that spatial analysis in terms of vulnerability to electricity access is that where someone lives appear to be at least as important as the socioeconomic group that they are a part of, as discussed in section 2.6. In section 3.4.5 it was also shown that short outages are autocorrelated. Thus, if the performance of the algorithm needs to be accurately validated, both spatial and temporal sampling during the validation is necessary. The results showed that the performance of RF steeply declined when spatial and temporal sampling was used during the cross-validation. Especially when spatial sampling was employed. When the same location is in the training and validation set (or testing set). Because of the median values calculated for each location, RF will recognize the median values as tags for a location and will be able to check if that tag was coupled to a Never Access location. However, in the real world, RF will have to perform on unseen locations and will not be able to easily distinct Never Access locations from the other classes, as was proven by the spatial cross-validation. It

was expected that the median values for radiance would off set the Never Access samples much more against the Normal Access and Outage samples, however, during closer examination of these values, it was found that some sensor locations actually have a lower median radiance value than some none access locations, this can be seen in figure 5.1. To a lesser extent, RF also performed worse when temporal sampling was employed during the cross-validation, as it might have expected more or less Outage samples than there actually were. However, as the same locations were present in the training and validation set, RF still was able to easily determine Never Access samples.

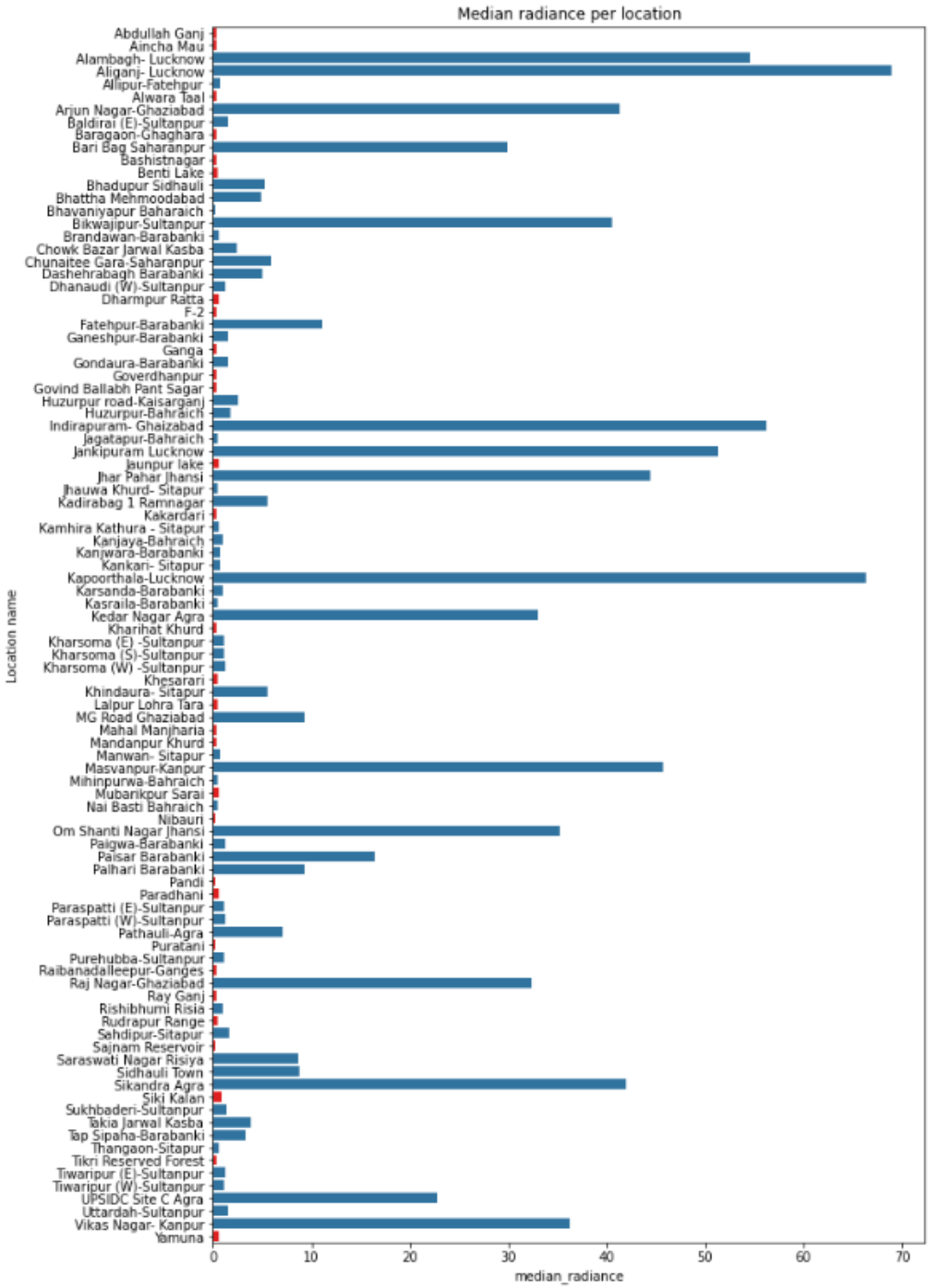


Figure 5.1: Median radiance per location (ESMI locations in blue, none access locations in red)

5.1.3. Features

The feature selection strategies, RFE and RFECV, selected different features. Where RFE selected only 9 of the 19 features, RFECV selected 16. When looking at the feature importance in figure 4.2, it seems that RFE chose features that had at least 0.04 importance score, whereas RFECV kept features with even smaller importance scores. It may seem odd to keep these features in, however, may seem not important for all the classes and get a lower importance score, but are important for the classification of the minority class. The performance of both these subsets is worse than when all features were used for 10-fold cross-validation, however, not by much. Interestingly, when spatial sampling is used, to ensure no data leakage, in combination with the resampling techniques, the performance of RFE is nearly as good for Outage samples as when all features are used, and RFECV actually performs better, shown in appendix G. Indicating, that not all features should be used when trying to identify outages.

5.2. Similar Studies

This research took inspiration from Mann et al. (2016), using remote sensing to identify three classes, in this research Never Access, Normal Access, Outage. Although, they only used nighttime radiance data as the predictor. The feature extraction, described in section 3.4.6, were also inspired by their research. Comparing the results found in this research to the results found by Mann et al. (2016) similar performance is found. Perfect identification of Never Access samples and much lower accuracy for Outage. However, they used different metrics, so, calculating the metrics used in this research based on their confusion matrix, the scores of their research are presented in tables 5.1 and 5.2. It becomes clear that the algorithm in their study actually struggles a lot to identify Outage samples. Still, they discovered that the technique can reasonably forecast electricity reliability rates for specific locations, with the predicted vs real regression producing an R^2 greater than 0.5. However, perfect identification using machine learning is almost always a sign of data leakage. It is important to take every precaution to prevent this.

Table 5.1: Scores Mann et al. (2016)

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.97	0.95	0.96
Outage	0.39	0.27	0.33

Table 5.2: Cohen's kappa scores Mann et al. (2016)

Class	SMOTE
Overall	0.93
Never Access	1.00
Normal Access	0.96
Outage	0.27

In this research, it was chosen to use a classification method to identify outages instead of a regression identifying the actual minutes without electricity, as it was expected to perform better, since the original data is highly imbalanced. The larger the imbalance, the stronger the model's bias towards the majority class (Krawczyk, 2016). To reduce this imbalance, grouping the individual minutes without electricity over zero minutes into a class has been done. Two very recent studies also study the reliability of electricity networks with the use of nighttime lights and make use of regression. The first study performed a regression to estimate the hours the amount of hours with normal voltage in Uttar Pradesh, India, using monthly nighttime luminosity, village characteristics, and ESMI sensors (Dugoua et al., 2022). Which lead to promising results. However, while the authors perform an extensive cross-validation, no spatial or temporal sampling is performed. As shown in this research, the lack of such sampling can lead to overconfidence in the performance of the model.

The other study by Shah et al. (2022) examines the electricity reliability in the city of Accra, Ghana, using a dataset representing the largest collection of utility-independent electricity reliability measurements on the African continent. Using a variety of machine learning algorithms, random forest and

logistic regression, to predict possible outages from nightlight images, they obtain high scores for recall, between 0.79 and 0.90, however, very low scores for precision, between 0.03 and 0.21. The balance, or imbalance, of these scores were the result of the authors purposely only focussing on the recall. The performance of the algorithms holds up during the spatial and temporal validation performed in the study. While, vulnerable parts of the cities can be mapped using their technique, the low precision make their produced map less actionable. However, both studies have shown that regression may as well serve as a valid technique for examining the reliability of the electricity network.

5.3. Possibilities for Usage

It was quickly determined that the performance of RF for Outage classification degraded when the Outage variable considered longer outages only, shown in appendix D. Thus, for the remainder of the research, only the Outage classification for any outages over 0 minutes was used. But it is important to reflect on this choice. In the literature review, multiple frameworks for electricity access were introduced. Electricity reliability is specifically mentioned in two of the frameworks; MTF and CEEW MTF. The identification of outages using remote sensing could be a help in determining electricity reliability when researcher or policymakers would apply one of those frameworks. The two frameworks differ on how the reliability attribute is defined, though. The MTF has no specific duration of outages defined for the reliability attribute in tier 4, only a maximum of 14 disruptions. For tier 5 the total duration of the outages should be less than two hours with a maximum of three disruptions. The CEEW MTF uses black-out days to measure the reliability attribute. The results from this research can thus only be applied when using the MTF. However, the CEEW MTF was specifically developed for India. The reason CEEW MTF uses complete black-out days as a measure for reliability instead of any outage, because unplanned outages are widespread in rural India (Jain et al., 2015). Jain et al. argue that using MTF will exaggerate the absence of electrical access in India. In India, short outages are frequently utilized to regulate demand or balance the grid. Furthermore, minor outages (less than 5 minutes) are not even considered in the reliability measures used by utilities in India. As a result, they calculate the reliability of the electricity supply by counting the number of days in a month when there was no power, which is generally due to factors other than purposeful demand control by load dispatch centers. While this is a clear argument to use black-out days to measure reliability, it does ignore the impact even small outages can have on a society. A streetlight that loses power for a minute is not a big deal. However, in a hospital it can mean life or death and there are many more situations constant connection to electricity is essential.

The results of this research cannot be directly applied to either of these frameworks, however. The algorithm can try to identify if there was an outage between 00:00 and 02:00, but it cannot estimate how many outages there were in this timespan or the duration of the outage, although this was the thought process behind creating multiple target variables. Still, it is interesting how the algorithm performed on completely unknown data. In figure 4.4 the results of the classifications for 2018 between 00:00 and 02:00 are shown with a spatial granularity of 0.1° x 0.1° using RF with the RFE selected features, as this combination of feature performed the best. Finer granularity would be better, but due to computational constraints this was not possible.

Looking at the figure, it seems that there are only a few pixels that have 80% or more days without outages. These yellow pixels are primarily in or near big, densely populated cities in Uttar Pradesh. This is consistent with utility companies prioritizing high-demand regions in order to recuperate the high fixed costs of infrastructure expansion. As was shown in figure 5.1, some locations of sensors that have electricity access, actually have a lower median radiance than non access locations. With the electrification of nearly all villages in India (World Bank, 2019), the actual usage of electricity, specifically nightlights, seem to be nonexistent in many parts of Uttar Pradesh. The algorithm does not classify all the classes as Outage perfectly, but it does classify areas as Never Access quite accurately. If these areas were to be compared to maps containing information on where the supposed electricity connections should exist, villages can be identified where policy should focus on implementation of electricity infrastructure and promotion of electricity usage. Also, the maps for Normal Access and Outage can also still be used to determine where more work needs to be done for electricity reliability. Again, the algorithm does not perform perfectly for these classes, but it does point out locations where more out-

ages are happening than others. Households in these locations are unlikely to benefit significantly from the grid connection they acquired. The map in figure 4.4 can be used as a precursor for future fieldwork and help policymaker and researchers identify reliability and fairness issues across Uttar Pradesh.

Future research will most certainly benefit from the use of a variety of methods and data, such as government statistics, surveys, and satellite photography. This study adds by reporting the performance of one such instrument utilizing freely accessible data and an open analytic technique. Especially, the importance of spatial and temporal validation for the algorithms used has been highlighted. As a result, despite the difficulties in assessing outages, it gives some optimism that NGOs and scholars may build monitoring tools to inform policymakers and hold government and utilities responsible at a relatively cheap cost in terms of resources and knowledge. This type of evaluation is a first step toward incentivizing higher reliability, especially in low-income areas, and fulfilling the full promise of national electrification.

5.4. Data Quality

Everything is influenced by the quality of the acquired data. Thus, it is important to highlight the limitations of the data used in this research.

5.4.1. ESMI Data

As discussed in section 3.2.1, the Prayas cannot provide a precise location of the sensors. The few sensors found on their website still have an uncertainty range of a few hundred meetings, per Prayas' own admission. The sensors that were not found on their site had to be manually located using Google Maps based on the name of the sensor. This introduces even more uncertainty about the locations of the sensor that could not be found on the Prayas website. This could of course impact the performance of RF if a different pixel is associated with a sensor than in reality is the case. However, since the pixel should likely be a neighboring pixel, similar outage frequency is to be expected for both pixels.

Another issue is that it is unknown in what kind of socio-economic households the sensors are placed. The Prayas website does not offer any information on this. As these sensors are expensive, it is probable that they are placed in more wealthy households. In section 3.2.1, it was shown that the ESMI sensors are quite densely placed in certain areas and completely absent in others. This does not offer a complete picture of the whole state and might mean that results from the study are too positive.

In section 3.4.1, another limitation of the ESMI data was briefly presented. Not all ESMI locations were consistently online during the whole year of 2018. While this does not impact the ability of RF to use the ESMI data, it does raise a question if the distribution of outages in the dataset is representative of the truth. The sensors are battery-backed-up, but nowhere on the Prays website is it mentioned how long they can operate without electricity. It is possible that big outages stop the sensor from working and therefore miss outages.

5.4.2. VIIRS

The SNPP satellite's after-midnight flyby period calls the VIIRS data's appropriateness for estimating energy reliability into doubt. Although lighting peaks in the evening before 10 p.m., new studies employing VIIRS DNB data show that there is still a significant quantity of lighting present beyond midnight (C. Elvidge et al., 2013; Kyba et al., 2015; Mann et al., 2016). This is not unexpected considering that satellite photos typically capture streetlights and other types of outdoor illumination, which are likely to be visible long after midnight.

In section 3.2.2, a graph from Mann et al. (2016) was presented in figure 3.5 that showed seven years of hourly feeder-line voltage data for Maharashtra, to confirm the potential of early-morning VIIRS data to offer fair estimates of electricity reliability. These findings show a 0.85 correlation between the frequency of outages throughout the day (6 a.m.–6 p.m.) and the frequency during VIIRS overhead times (midnight–2 a.m.). Although this significant link between electricity reliability during the daytime

and during VIIRS overpass periods, it may not be universal. However, the fact that it exists in one of India's largest states, Maharashtra, is promising. Although, the downward bias of observations implies that the overnight outage rates reported here would understate the frequency of outages during the daylight hour.

The quality of the images is partly determined by the satellite technology, such as the images sensors utilized. However, these raw photographs will be corrected to enhance the impacts generated by capturing these images. These modifications can improve the image's usability, but some of them can potentially delete information acquired in the original image. For both VIIRS and MODIS data, clouds and bad quality flags were masked, leaving big gaps for some days in Uttar Pradesh. Specifically, with the importance of the radiance found in the results, this can have a significant impact on the performance of the algorithm.

Another issue that might come from using VIIRS to identify household outages, is that in India, many firms and households rely on their own generators for electricity (Szakonyi & Urpelainen, 2013). Which means that an outage which affects streetlights does not necessarily mean that the outage is experienced indoors as well. VIIRS mostly captures streetlights, thus RF might see a drop in nighttime radiance, but no Outage classification in the target data, as the household still has access to electricity. It is not clear if the ESMI sensors are placed in buildings with or without backup generators, so this again adds to the uncertainty of the algorithm.

6
Conclusion

In this last chapter, we analyze the degree to which the intended methodology and research questions of this master thesis have been answered and draw conclusions. Furthermore, recommendations are made for future studies to improve the classification of electricity outages. Conclusions are offered as responses to the primary research question as well as the sub-questions.

6.1. Research Overview

The following research problem was investigated: *How do the number of electricity outages vary across one of India's less developed states?* The Random Forest (RF) method was introduced in this paper to detect power outages. If RF can accurately predict electrical outages in India, it may assist reduce reliance on metering while also identifying reliability and fairness concerns. The state studied in this research is Uttar Pradesh.

The purpose of this research is to learn more about the association between particular features and electricity outages in India. It will investigate in depth how the use of RF might enhance data collection on electricity outages in India. The sub-questions will be answered in the following parts to assist answer the main research question.

What is the status of electricity access in India?

India has made considerable strides toward providing electricity to all of its citizens. Between 2000 and 2010, over 283 million people were added to the internet, significantly outnumbering natural population growth. By 2010, the total electrified population in the country had reached 881 million. Despite this achievement, 311 million people remained without power, the majority of whom live in poor rural areas. Only around 7% of rural families that have access to electricity report no outages. 20% of homes with electricity have outages lasting up to four hours per day, and the same proportion gets intermittent power supply for the most of the day. Bihar and Uttar Pradesh have the greatest average daily outages, while having the lowest village coverage and household adoption rates. Residential electricity in India is expected to reach saturation levels by 2020. Approximately 96.7 percent of Indian houses are grid-connected. Nonetheless, 4% or more of residences in Chhattisgarh, Haryana, Rajasthan, and Uttar Pradesh do not have an electricity connection. More than a third of the disconnected households are in Uttar Pradesh alone. Every day, Indian households receive an average of 20.6 hours of power.Rural homes receive supplies for just 19.9 hours per day, whereas urban households receive supply for 22.3 hours per day. However, national averages obscure the significant variation in supplies available throughout India's states. On average, homes in Uttar Pradesh, Jharkhand, Haryana, Assam, and Bihar experience six or more hours of daily power disruptions.Rural families are disproportionately affected by power outages, which usually occur at night, with 50% having daily outages between 18:00 and midnight. However, most metropolitan homes have an uninterrupted supply until late at night.

How have existing literature used remote sensing and machine learning to measure aspects of the energy system?

By deploying long-term evening illumination, remote sensing data has been utilized to measure changes in energy access and socioeconomic wellbeing. However, few studies have used remote sensing to investigate actual power outages. Nighttime satellite imagery was used to predict power disruptions in the aftermath of Hurricane Sandy in the United States in 2012, with promising results. Another research

used satellite data of nighttime lights to create detailed estimates of electricity outages in western India. Despite the fact that the study was conducted over a very short period of time, the results were encouraging and warranted more examination.

As illustrated by the usage of RF, machine learning approaches are intrinsically suited to the nature of remote sensing data. Other studies used neural networks to map power consumption and poverty using satellite images, with findings that matched field surveys. These solutions are especially useful in developing countries because satellite imagery is frequently inadequate or unorganized. RF was used for this study due to its computational restrictions and convenience of usage.

What are the spatial characteristics that influence (or are correlated with) electricity reliability in India?
Nighttime lights are the primary predictor used for electricity outages in this research. Lighting is one of the first services electricity will be used for. As light can be observed from satellites during the night, it can be used as a proxy for electricity usage during the overpass time of the satellite.

Weather data in the form of temperature, wind speed, and precipitation are used as well. Damage to the electrical grid caused by seasonal weather is common in South-East Asia.

Energy poverty stricken households are more inclined to use wood or other biomass fuels instead of high-cost electricity, thereby generating air pollution. It is hypothesized in this study that when there is an electricity outage, individuals would turn to more polluting sources of energy.

Land cover is employed in this study as well. Land cover data is not expected to directly identify outages, it is important to not only recognize outages when they occur, but also areas that do not have any electricity access at all, such as forests or deserts, to not misclassify an area having outages all year when the algorithms are presented with unseen data.

Finally, spatial population data based on the Census of India 2011 is used. Population counts yielded poverty indicators that were highly linked with sub-national indices of income poverty. Furthermore, the prediction of electricity access using remote sensing data has largely been handled in these researches using pictures of nightlights combined with population counts. Utility companies are likely to cherry-pick areas with large populations for investments, leaving rural consumers in the dark.

To what extent can historical data identify electricity outages in India?
Because the predictor data sources do not collect information every minute between 00:00 and 02:00 and the original data is highly imbalanced. It was chosen to use a classification method to identify outages instead of a regression identifying the actual minutes without electricity, as it was expected to perform better. Since, the larger the imbalance, the stronger the model's bias towards the majority class. To do this, the target data must be classified. Never Access, Normal Access, and Outage are the three classifications used. Never Access places, like as forests, should be identified as such. When to identify a pixel on a specific day as an Outage sample or not is a critical choice. There is no clear definition of how many minutes must go before an outage is declared. One minute without electricity is strictly speaking considered an outage. This can be disastrous during surgery, but it is not as critical for illumination. Longer outages were also predicted to be simpler to identify than shorter ones. Six outage classification variables were created to investigate this: over 0 minutes without electricity, over 5 minutes without electricity, over 15 minutes without electricity, over 30 minutes without electricity, over 60 minutes without electricity, and over 120 minutes without electricity.

When evaluating the results of the normal and stratified 10-fold cross-validations, the algorithm perfectly identifies Never Access samples. For the target variable of over 0 minutes without electricity, 94% of Normal Access samples were identified as such, with a total precision of 82%, which is highly accurate. With a precision of 69%, RF was able to classify 44% of the Outage samples as such. Thus, the method is working pretty well, but the performance of the other target variables decreased as the duration of the outage definition increased. Perfect machine learning identification is nearly always a sign of data leakage. It is critical to take every precaution to avoid this.

The premise of this research is that where someone lives appears to be at least as relevant as the socioeconomic class that they are a part of in terms of vulnerability to electricity access. Short outages have also been proven to be autocorrelated. As a result, if the algorithm's performance should be properly tested, both spatial and temporal sampling are required throughout the validation. The results indicated that when spatial and temporal sampling were applied during cross-validation, RF performance dropped dramatically. Particularly when spatial sampling was used. So, using a standard 10-fold cross-validation actually result in data leakage.

As a result, the scores obtained from spatial sampling in conjunction with 10-fold cross-validation are more likely to be an accurate estimate of RF performance. 88% of Never Access samples were correctly identified, with a total accuracy of 82%, which is not perfect but still quite high. Normal Access samples were categorized as such with an even higher rate of 89%, albeit with a lower precision of 75%. Outage classification is where performance suffers the most. Only 12% of the Outage samples were correctly identified, with a precision of only 41%. RF also performed poorly when temporal sampling was used during cross-validation, but to a smaller degree. However, as the same locations were present in the training and validation set, RF still was able to easily determine Never Access samples.

Different resampling techniques for the training data were investigated. Resampling has the potential to boost performance of the algorithm. The sampling techniques used in this research are Synthetic Minority Oversampling technique (SMOTE), Tomek undersapmling, and a combination of the two SMOTETomek. Using the resampling techniques, the performance for Outage increases drastically for SMOTE and SMOTETomek, but does come at a cost of the performance for Normal Access. While Tomek does increase the performance for Outage, without decreasing the performance for Normal Access, although only limited.

In this study, two feature selection procedures were used: recursive feature elimination (RFE) and recursive feature elimination with cross-validation (RFECV). Whereas RFE selected just 9 of the 19 traits, RFECV selected 16. Both of these subsets perform similar as when all features were utilized for the standard 10-fold cross-validation. Surprisingly, when spatial sampling is utilized in conjunction with resampling techniques to ensure minimal data leakage, the performance of RFECV is almost as good for Outage samples as when all features are used, and RFE actually performs better. This indicates that not all features should be used to identify outages. Using the RFE subset of features, 92% of Never Access samples were categorized as such, with a total accuracy of 87%, which isn't ideal but is still fairly high. 69% of Normal Access samples were classified as such, but with 84% precision. The performance of the outage classification is the worst. Nonetheless, 55% of the Outage samples were categorized as such, although with an accuracy of just 39%. Using the RFE subset, with SMOTE resampling, the map in figure 4.4 shows the results of the classifications for 2018 between 00:00 and 02:00 are shown with a spatial granularity of 0.1° x 0.1°. Looking at the figure, it seems that there are only a few pixels that have 80% or more days without outages. The outage free days are primarily experienced in or near big, densely populated cities in Uttar Pradesh. This is consistent with utility companies prioritizing high-demand regions in order to recuperate the high fixed costs of infrastructure expansion. For most parts of Uttar Pradesh, outages happen 40% to 60% of the days, with some parts in the north-east experiencing outages more than 80% of the days.

6.2. Policy Recommendations

One interesting discovery is that Normal Access samples are mostly identified on pixels that are never labeled as Never Access. However, as was shown in figure 5.1, certain sensor locations with electricity access have a lower median radiance than non-access locations. Despite the electrification of virtually all villages in India, actual electricity usage, particularly nightlights, appears to be nonexistent in many regions of Uttar Pradesh. The method does not correctly classify all Outage samples as Outage, but it is highly accurate for Never Access areas. When these locations are compared to maps that show where the intended electricity connections should be, communities may be identified where policy should focus on the development of electrical infrastructure and the promotion of electricity usage. Additionally, the maps for Normal Access and Outage may still be utilized to indicate where more work is needed

for electrical reliability. Again, the algorithm does not function flawlessly for these classes, but it does identify areas with more outages than others.

6.3. Scientific Contribution

Scientific research can benefit from a range of approaches and data, such as government statistics, surveys, and satellite imagery. This thesis contributes by reporting on the performance of one such instrument using publicly available data and an open analytic approach. Especially, the importance of spatial and temporal validation for the algorithms used has been highlighted. As a result, despite the difficulty in analyzing outages, there is some hope that NGOs and researchers may be able to develop monitoring tools to inform policymakers and hold government and utility companies accountable at a low cost in terms of resources and expertise. This form of assessment is a first step toward promoting greater reliability, particularly in low-income regions, and delivering the full promise of national electrification.

6.4. Further Research

The first thing that can be expanded on, is to use a longer time range for the training and validating of RF. With more samples to learn from, it is likely that performance will increase. However, it should be noted that this also contains a risk of overtraining on the few specific locations that are used in this research. More locations with sensors could be a solution to counter this risk.

To gain more confidence in the algorithm, an interesting next step would be to research why Normal Access and Outage samples are confused with each other. Does it happen under specific conditions? If under some conditions, the algorithm is less accurate, these samples can be removed from the training and test samples. These conditions can then be accounted for in the algorithm and specific advice can be provided for the use of the algorithm. Other data sources may be able to provide extra context for these conditions and increase the accuracy of the algorithm, such as a map with fine spatial granularity for literacy or scheduled caste and tribe counts are probably able to further fine tune the algorithm, at the time of writing these could not readily be found for spatial analysis.

Extending the experiment to other nations, might assist validate or invalidate the technique in a larger variety of circumstances. Because other developing countries are quickly expanding their electricity infrastructure and experiencing the same problems with monitoring outages, satellite data might be a critical component in monitoring access.

In this research it was opted to use RF for classification, but as mentioned in 3.3 and section 2.7, other researchers using nighttime lights have used CNNs and regression techniques to predict various other targets. Another machine learning technique that might be worth exploring is the use of semi-supervised algorithms. Semi-supervised learning techniques are a combination of supervised and unsupervised learning, since they train on both labeled and unlabeled data. They are utilized when the majority of the observations lack labels and make use of a little quantity of tagged data to increase learning accuracy. However, semi-supervised learning still often suffers from performance degradation caused by the introduction of unlabeled data. Still, the potential of using unlabeled data is worth exploring for an issue like this where getting labeled data is a constraint.

Bibliography

Abi-Samra, N., McConnach, J., Mukhopadhyay, S., & Wojszczyk, B. (2014). When the bough breaks: Managing extreme weather events affecting electrical power grids. *IEEE Power and Energy Magazine*, *12*(5), 61–65. https://doi.org/10.1109/MPE.2014.2331899

Acemoglu, D., & Robinson, J. (2013). *Why Nations Fail : The Origins of Power, Prosperity and Poverty* (Main). Adfo Books.

Agrawal, S., Mani, S., Jain, A., & Ganesan, K. (2020). State of electricity access in india: Insights from the india residential energy survey. *New Delhi: Council on Energy, Environment and Water*.

Aklin, M., Cheng, C.-y., Urpelainen, J., Ganesan, K., & Jain, A. (2016). Factors affecting household satisfaction with electricity supply in rural India. *Nature Energy*, *1*(11). https://doi.org/10.1038/nenergy.2016.170

Aklin, M., Chindarkar, N., Urpelainen, J., Jain, A., & Ganesan, K. (2021). The hedonic treadmill: Electricity access in india has increased, but so have expectations. *Energy Policy*, *156*, 112391. https://doi.org/https://doi.org/10.1016/j.enpol.2021.112391

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Allen, J. (2005). NASA - Enhancing Research and Education Through Partnerships. https://www.nasa.gov/audience/foreducators/k-4/features/F_Enhancing_Research.html

Athawale, T. (2021). India's electric grid reliability and its importance in the clean energy transition. https://www.raponline.org/blog/indias-electric-grid-reliability-its-importance-in-clean-energy-transition/

Bakkensen, L., & Schuler, P. (2020). A preference for power: Willingness to pay for energy reliability versus fuel type in vietnam. *Energy Policy*, *144*, 111696. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111696

Balk, D., Montgomery, M. R., Engin, H., Lin, N., Major, E., & Jones, B. (2020). Spatial Data from the 2011 India Census, v1: India Data Collection | SEDAC. https://doi.org/10.3390/data4010035

Banerjee, S. G., Barnes, D., Singh, B., Mayer, K., & Samad, H. (2015). *Power for all: Electricity access challenge in india*. World Bank Publications. https://openknowledge.worldbank.org/handle/10986/20525

Bayer, P., Kennedy, R., Yang, J., & Urpelainen, J. (2020). The need for impact evaluation in electricity access research. *Energy Policy*, *137*, 111099. https://doi.org/https://doi.org/10.1016/j.enpol.2019.111099

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bernard, S., Heutte, L., & Adam, S. (2009). Influence of hyperparameters on random forest accuracy. *International workshop on multiple classifier systems*, 171–180. https://doi.org/https://doi.org/10.1007/978-3-642-02326-2_18

Bhatia, M., & Angelou, N. (2015). Beyond connections. *Energy Sector Management Assistance Program (ESMAP)*. http://documents.worldbank.org/curated/en/650971468180259602/Beyond-connections-energy-access-redefined-technical-report

Bouzarovski, S., & Simcock, N. (2017). Spatializing energy justice. *Energy Policy*, *107*, 640–648. https://doi.org/https://doi.org/10.1016/j.enpol.2017.03.064

Bowyer, K. W., Chawla, N. V., Hall, L. O., & Kegelmeyer, W. P. (2011). SMOTE: synthetic minority over-sampling technique. *CoRR*, *abs/1106.1813*. http://arxiv.org/abs/1106.1813

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N. E., Herold, M., & Fritz, S. (2020). Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2018: Globe. https://doi.org/10.5281/zenodo.3518038

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). Api design for machine learning software: Experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*. https://doi.org/https://doi.org/10.48550/arXiv.1309.0238

Canares, S., Dinesh, A., Verhulst, S., & Young, A. (2017). *Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How*. African Minds.

Chakravorty, U., Pelli, M., & Ural Marchand, B. (2014). Does the quality of electricity matter? evidence from rural india. *Journal of Economic Behavior & Organization*, *107*, 228–247. https://doi.org/https://doi.org/10.1016/j.jebo.2014.04.011

Chaurey, A., Ranganathan, M., & Mohanty, P. (2004). Electricity access for geographically disadvantaged rural communities—technology and policy insights. *Energy Policy*, *32*(15), 1693–1705. https://doi.org/https://doi.org/10.1016/S0301-4215(03)00160-5

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37–46.

Cole, T. A., Wanik, D. W., Molthan, A. L., Román, M. O., & Griffin, R. E. (2017). Synergistic use of nighttime satellite data, electric utility infrastructure, and ambient population to improve power outage detections in urban areas. *Remote Sensing*, *9*(3). https://doi.org/10.3390/rs9030286

Copernicus. (2019-07-12). Copernicus Climate Data Store | Copernicus Climate Data Store. https://doi.org/10.24381/cds.e2161bac

Criminisi, A., Shotton, J., & Konukoglu, E. (2012). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, *7*(2–3), 81–227. https://doi.org/10.1561/0600000035

Das, B., & Mistri, A. (2013). Household quality of living in indian states: Analysis of 2011 census. *Environment and Urbanization ASIA*, *4*(1), 151–171. https://doi.org/10.1177/0975425313477759

de Groot, J., Mohlakoana, N., Knox, A., & Bressers, H. (2017). Fuelling women's empowerment? an exploration of the linkages between gender, entrepreneurship and access to energy in the informal food sector. *Energy Research & Social Science*, *28*, 86–97. https://doi.org/https://doi.org/10.1016/j.erss.2017.04.004

Dekanovsky, V. (2021). Complete guide to Python's cross-validation with examples. https://towardsdatascience.com/complete-guide-to-pythons-cross-validation-with-examples-a9676b5cac12

Derks, M., & Romijn, H. (2019). Sustainable performance challenges of rural microgrids: Analysis of incentives and policy framework in indonesia. *Energy for Sustainable Development*, *53*, 57–70. https://doi.org/https://doi.org/10.1016/j.esd.2019.08.003

der Linde, M. v. (2019). Fire risk mitigation in the overhead electricity distribution network. *2019 29th Australasian Universities Power Engineering Conference (AUPEC)*, 1–6. https://doi.org/10.1109/AUPEC48547.2019.211806

Desai, S., Vanneman, R., & National Council of Applied Economic Research, New Delhi. (2005). India Human Development Survey (IHDS), 2005. https://doi.org/10.3886/ICPSR22626.v12

Dubois, U. (2012). From targeting to implementation: The role of identification of fuel poor households [Special Section: Fuel Poverty Comes of Age: Commemorating 21 Years of Research and Policy]. *Energy Policy*, *49*, 107–115. https://doi.org/https://doi.org/10.1016/j.enpol.2011.11.087

Dugoua, E., Kennedy, R., Shiran, M., & Urpelainen, J. (2022). Assessing reliability of electricity grid services from space: The case of uttar pradesh, india. *Energy for Sustainable Development*, *68*, 441–448. https://doi.org/https://doi.org/10.1016/j.esd.2022.04.004

Dugoua, E., Kennedy, R., & Urpelainen, J. (2018). Satellite data for the social sciences: Measuring rural electrification with night-time lights. *International Journal of Remote Sensing*, *39*(9), 2690–2701. https://doi.org/10.1080/01431161.2017.1420936

Dugoua, E., Liu, R., & Urpelainen, J. (2017). Geographic and socio-economic barriers to rural electrification: New evidence from indian villages. *Energy Policy*, *106*, 278–287. https://doi.org/https://doi.org/10.1016/j.enpol.2017.03.048

ECMWF. (2021). ERA5: How to calculate wind speed and wind direction from u and v components of the wind? - Copernicus Knowledge Base - ECMWF Confluence Wiki. https://confluence.ecmwf.int/pages/viewpage.action?pageId=133262398

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li, & M. J. Murphy (Eds.), *Machine learning in radiation oncology: Theory and applications* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1

Elvidge, C., Baugh, K., Zhizhin, M., & Hsu, F.-C. (2013). Why viirs data are superior to dmsp for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*, 62–69. https://doi.org/10.7125/APAN.35.7

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *35*(8), 1652–1660. https://doi.org/https://doi.org/10.1016/j.cageo.2009.01.009

Eskandarpour, R., & Khodaei, A. (2017). Machine learning based power grid outage prediction in response to extreme events. *IEEE Transactions on Power Systems*, *32*(4), 3315–3316. https://doi.org/10.1109/TPWRS.2016.2631895

Falchetta, G., & Noussan, M. (2019). Interannual variation in night-time light radiance predicts changes in national electricity consumption conditional on income-level and region. *Energies*, *12*(3). https://doi.org/10.3390/en12030456

Falchetta, G., Pachauri, S., Parkinson, S., & Byers, E. (2019). A high-resolution gridded dataset to assess electrification in sub-saharan africa. *Scientific data*, *6*(1), 1–9.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863–905.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, *15*(1), 3133–3181.

Fischer, P., Dosovitskiy, A., & Brox, T. (2015). Image orientation estimation with convolutional networks. *German conference on pattern recognition*, 368–378.

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, *11*(3–4), 219–354. https://doi.org/10.1561/2200000071

Galdi, P., & Tagliaferri, R. (2018). Data mining: Accuracy and error measures for classification and prediction. https://doi.org/10.1016/B978-0-12-809633-8.20474-3

Gao, G., Wang, M., Huang, H., & Tang, W. (2021). Agricultural irrigation area prediction based on improved random forest model. https://doi.org/10.21203/rs.3.rs-156767/v1

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, *63*(1), 3–42. https://doi.org/https://doi.org/10.1007/s10994-006-6226-1

Gibson, J., & Olivia, S. (2010). The effect of infrastructure access and quality on non-farm enterprises in rural indonesia. *World Development*, *38*(5), 717–726. https://doi.org/https://doi.org/10.1016/j.worlddev.2009.11.010

Global Modeling and Assimilation Office (GMAO). (2015). *MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics*. https://doi.org/10.5067/KLICLTZ8EM9D

González-Eguino, M. (2015). Energy poverty: An overview. *Renewable and Sustainable Energy Reviews*, *47*, 377–385. https://doi.org/https://doi.org/10.1016/j.rser.2015.03.013

Gupta, S., Gupta, E., & Sarangi, G. K. (2020). Household energy poverty index for india: An analysis of inter-state differences. *Energy Policy*, *144*, 111592. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111592

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

He, J., Cheng, M. X., Fang, Y., & Crow, M. L. (2018). A machine learning approach for line outage identification in power systems. *International Conference on Machine Learning, Optimization, and Data Science*, 482–493.

Heaton, J. (2016). An empirical analysis of feature engineering for predictive modeling. *SoutheastCon 2016*, 1–6. https://doi.org/10.1109/SECON.2016.7506650

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/https://doi.org/10.7717/peerj.5518

Huh, K. (2021). Surviving in a Random Forest with Imbalanced Datasets. https://medium.com/sfu-cspmp/surviving-in-a-random-forest-with-imbalanced-datasets-b98b963d52eb

IEA, IRENA, UNSD, World Bank, & WHO. (2021). *Tracking SDG 7* (tech. rep.). https://trackingsdg7.esmap.org/data/files/download-documents/2021_tracking_sdg7_report.pdf

Imran, M., Stein, A., & Zurita-Milla, R. (2014). Investigating rural poverty and marginality in burkina faso using remote sensing-based products. *International Journal of Applied Earth Observation and Geoinformation*, *26*, 322–334. https://doi.org/https://doi.org/10.1016/j.jag.2013.08.012

Jain, A., Ray, S., Ganesan, K., Aklin, M., Cheng, C., & Urpelainen, J. (2015). Access to Clean Cooking Energy and Electricity: Survey of States. *Council on Energy, Environment and Water*. https://www.ceew.in/publications/access-clean-cooking-energy-and-electricity-survey-states

Jain, A., Tripathi, S., Mani, S., Patnaik, S., Shahidi, T., & Ganesan, K. (2018). Access to Clean Cooking Energy and Electricity: Survey of States 2018. *Council on Energy, Environment and Water*. https://www.ceew.in/publications/access-clean-cooking-energy-and-electricity

Jeuland, M., Fetter, T. R., Li, Y., Pattanayak, S. K., Usmani, F., Bluffstone, R. A., Chávez, C., Girardeau, H., Hassen, S., Jagger, P., Jaime, M. M., Karumba, M., Köhlin, G., Lenz, L., Litzow, E. L., Masatsugu, L., Naranjo, M. A., Peters, J., Qin, P., … Toman, M. (2021). Is energy the golden thread? a systematic review of the impacts of modern and traditional energy use in low- and middle-income countries. *Renewable and Sustainable Energy Reviews*, *135*, 110406. https://doi.org/https://doi.org/10.1016/j.rser.2020.110406

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Joseph, K. L. (2010). The politics of power: Electricity reform in india. *Energy Policy*, *38*(1), 503–511. https://doi.org/https://doi.org/10.1016/j.enpol.2009.09.041

Kemmler, A. (2007). Factors influencing household access to electricity in india. *Energy for Sustainable Development*, *11*(4), 13–20. https://doi.org/https://doi.org/10.1016/S0973-0826(08)60405-6

Kennedy, R., Mahajan, A., & Urpelainen, J. (2019). Quality of service predicts willingness to pay for household electricity connections in rural india. *Energy Policy*, *129*, 319–326. https://doi.org/https://doi.org/10.1016/j.enpol.2019.01.034

Kennedy, R., Mahajan, A., & Urpelainen, J. (2020). Crowdsourcing data on the reliability of electricity service: Evidence from a telephone survey in uttar pradesh, india. *Energy Policy*, *145*, 111746. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111746

Kenward, A., Raja, U. et al. (2014). Blackout: Extreme weather climate change and power outages. *Climate central*, *10*, 1–23.

Khan, M. F., Jain, A., Arunachalam, V., & Paventhan, A. (2014). Roadmap for smart metering deployment for indian smart grid. *2014 IEEE PES General Meeting | Conference Exposition*, 1–5. https://doi.org/10.1109/PESGM.2014.6939919

Khanna, R. A., Li, Y., Mhaisalkar, S., Kumar, M., & Liang, L. J. (2019). Comprehensive energy poverty index: Measuring energy poverty and identifying micro-level solutions in south and southeast asia. *Energy Policy*, *132*, 379–391. https://doi.org/https://doi.org/10.1016/j.enpol.2019.05.034

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1137–1143.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/https://doi.org/10.1007/s13748-016-0094-0

Kulkarni, V., & Kulkarni, K. (2020). A blockchain-based smart grid model for rural electrification in india. *2020 8th International Conference on Smart Grid (icSmartGrid)*, 133–139. https://doi.org/10.1109/icSmartGrid49881.2020.9144898

Kumar, A. (2018). Justice and politics in energy access for education, livelihoods and health: How socio-cultural processes mediate the winners and losers. *Energy Research & Social Science*, *40*, 3–13. https://doi.org/https://doi.org/10.1016/j.erss.2017.11.029

Kumar, V., & Minz, S. (2014). Feature selection: A literature review. *SmartCR*, *4*(3), 211–229.

Kyba, C. C. M., Garz, S., Kuechly, H., De Miguel, A. S., Zamorano, J., Fischer, J., & Hölker, F. (2015). High-resolution imagery of earth at night: New sources, opportunities and challenges. *Remote Sensing*, *7*(1), 1–23. https://doi.org/10.3390/rs70100001

Lan, T., Hu, H., Jiang, C., Yang, G., & Zhao, Z. (2020). A comparative study of decision tree, random forest, and convolutional neural network for spread-f identification. *Advances in Space Research*, *65*(8), 2052–2061. https://doi.org/https://doi.org/10.1016/j.asr.2020.01.036

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*(17), 1–5. http://jmlr.org/papers/v18/16-365.html

Louridas, P., & Ebert, C. (2016). Machine learning. *IEEE Software*, *33*(5), 110–115. https://doi.org/10.1109/MS.2016.114

Lu, Y., Khan, Z. A., Alvarez-Alvarado, M. S., Zhang, Y., Huang, Z., & Imran, M. (2020). A critical review of sustainable energy policies for the promotion of renewable energy sources. *Sustainability*, *12*(12). https://doi.org/10.3390/su12125078

Lyapustin, A., & Wang, Y. (2015). MCD19A2 MODIS/Terra+Aqua Aerosol Optical Thickness Daily L2G Global 1km SIN Grid. https://doi.org/10.5067/MODIS/MCD19A2.006

Mann, M. L., Melaas, E. K., & Malik, A. (2016). Using viirs day/night band to measure electricity supply reliability: Preliminary results from maharashtra, india. *Remote Sensing*, *8*(9). https://doi.org/10.3390/rs8090711

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*, 1–9. https://doi.org/https://doi.org/10.1016/j.envsoft.2017.12.001

Min, B., & Gaba, K. M. (2014). Tracking electrification in vietnam using nighttime lights. *Remote Sensing*, *6*(10), 9511–9529. https://doi.org/10.3390/rs6109511

Min, B., Gaba, K. M., Sarr, O. F., & Agalassou, A. (2013). Detection of rural electrification in africa using dmsp-ols night lights imagery. *International Journal of Remote Sensing*, *34*(22), 8118–8141. https://doi.org/10.1080/01431161.2013.833358

Min, B. K., O'Keeffe, Z., & Zhang, F. (2017). Whose power gets cut? using high-frequency satellite images to measure power supply irregularity. *Using High-Frequency Satellite Images to Measure Power Supply Irregularity (June 29, 2017). World Bank Policy Research Working Paper*, (8131). https://ssrn.com/abstract=3006205

Ministry of Power. (2020). Electricity (Rights of Consumers) Rules, 2020. https://pib.gov.in/PressReleasePage.aspx?PRID=1682384

Morikawa, R. (2014). Remote sensing tools for evaluating poverty alleviation projects: A case study in tanzania [Humanitarian Technology: Science, Systems and Global Impact 2014, HumTech2014]. *Procedia Engineering*, *78*, 178–187. https://doi.org/https://doi.org/10.1016/j.proeng.2014.07.055

Mukherji, J. (2021). Beyond access: The future of electricity in India. https://timesofindia.indiatimes.com/blogs/voices/beyond-access-the-future-of-electricity-in-india/

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383.

NASA. (2021). Nighttime Lights | Earthdata. https://earthdata.nasa.gov/learn/backgrounders/nighttime-lights

Nateghi, R., Guikema, S., & Quiring, S. M. (2013). Power outage estimation for tropical cyclones: Improved accuracy with simpler models. *Risk Analysis*, *34*(6), 1069–1078. https://doi.org/https://doi.org/10.1111/risa.12131

Nussbaumer, P., Bazilian, M., & Modi, V. (2012). Measuring energy poverty: Focusing on what matters. *Renewable and Sustainable Energy Reviews*, *16*(1), 231–243. https://doi.org/https://doi.org/10.1016/j.rser.2011.07.150

Onyeji, I., Bazilian, M., & Nussbaumer, P. (2012). Contextualizing electricity access in sub-saharan africa. *Energy for Sustainable Development*, *16*(4), 520–527. https://doi.org/https://doi.org/10.1016/j.esd.2012.08.007

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. https://doi.org/https://doi.org/10.48550/arXiv.1511.08458

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? *International workshop on machine learning and data mining in pattern recognition*, 154–168. https://doi.org/https://doi.org/10.1007/978-3-642-31537-4_13

Palit, D., & Bandyopadhyay, K. R. (2017). Rural electricity access in india in retrospect: A critical rumination. *Energy Policy*, *109*, 109–120. https://doi.org/https://doi.org/10.1016/j.enpol.2017.06.025

Parashar, B. (2022). Heat wave conditions aggravate power crisis in Uttar Pradesh. https://www.hindustantimes.com/cities/lucknow-news/heat-wave-conditions-aggravate-power-crisis-in-uttar-pradesh-101650818270180.html

Parray, M. T., & Tongia, R. (2019). *Understanding India's Power Capacity: Surplus or not, and for how long?* (Tech. rep.). Brookings India. https://www.brookings.edu/research/understanding-indias-power-capacity-surplus-or-not-and-for-how-long/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelz, S., Aklin, M., & Urpelainen, J. (2021). Electrification and productive use among micro- and small-enterprises in rural north india. *Energy Policy*, *156*, 112401. https://doi.org/https://doi.org/10.1016/j.enpol.2021.112401

Pelz, S., Chindarkar, N., & Urpelainen, J. (2021). Energy access for marginalized communities: Evidence from rural north india, 2015–2018. *World Development*, *137*, 105204. https://doi.org/https://doi.org/10.1016/j.worlddev.2020.105204

Pelz, S., Pachauri, S., & Groh, S. (2018). A critical review of modern approaches for multidimensional energy poverty measurement. *WIREs Energy and Environment*, *7*(6), e304. https://doi.org/https://doi.org/10.1002/wene.304

Pelz, S., & Urpelainen, J. (2020). Measuring and explaining household access to electrical energy services: Evidence from rural northern india. *Energy Policy*, *145*, 111782. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111782

Peters, J., & Sievert, M. (2016). Impacts of rural electrification revisited – the african context. *Journal of Development Effectiveness*, *8*(3), 327–345. https://doi.org/10.1080/19439342.2016.1178320

Phadke, A., Park, W. Y., & Abhyankar, N. (2019). Providing reliable and financially sustainable electricity access in india using super-efficient appliances. *Energy Policy*, *132*, 1163–1175. https://doi.org/https://doi.org/10.1016/j.enpol.2019.06.015

Powers, D. M. W. (2020). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *CoRR*, *abs/2010.16061*. https://arxiv.org/abs/2010.16061

Practical Action. (2012). *Poor people's energy outlook 2012: Energy for earning a living*. Practical Action Publishing.

Prayas, E. G. (2019). *Minute-wise voltage data collected in Electricity Supply Monitoring Initiative*. https://doi.org/10.7910/DVN/CLLZZM

Probst, P., & Boulesteix, A.-L. (2017). To tune or not to tune the number of trees in random forest. *J. Mach. Learn. Res.*, *18*(1), 6673–6690. https://doi.org/https://doi.org/10.48550/arXiv.1705.05654

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/https://doi.org/10.1002/widm.1301

Proville, J., Zavala-Araiza, D., & Wagner, G. (2017). Night-time lights: A global, long term look at links to socio-economic trends. *PloS one*, *12*(3), e0174610. https://doi.org/https://doi.org/10.1371/journal.pone.0174610

Python. (2021). Welcome to Python. https://www.python.org/

Rajkumari, L. (2020). Relation between electricity consumption and economic growth in karnataka, india: An aggregate and sector-wise analysis. *The Electricity Journal*, *33*(5), 106768. https://doi.org/https://doi.org/10.1016/j.tej.2020.106768

Rao, N. D. (2013). Does (better) electricity supply increase household enterprise income in india? *Energy Policy*, *57*, 532–541. https://doi.org/https://doi.org/10.1016/j.enpol.2013.02.025

Reddy, A. K., Annecke, W., Blok, K., Bloom, D., Boardman, B., Eberhard, A., & Ramakrishna, J. (2000). Energy and social issues. *World energy assessment*, 39–60. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.4978&rep=rep1&type=pdf

Reddy, B. S. (2015). Access to modern energy services: An economic and policy framework. *Renewable and Sustainable Energy Reviews*, *47*, 198–212. https://doi.org/https://doi.org/10.1016/j.rser.2015.03.058

Richmond, J., Agrawal, S., & Urpelainen, J. (2020). Drivers of household appliance usage: Evidence from rural india. *Energy for Sustainable Development*, *57*, 69–80. https://doi.org/https://doi.org/10.1016/j.esd.2020.05.004

Riva, F., Ahlborg, H., Hartvigsson, E., Pachauri, S., & Colombo, E. (2018). Electricity access and rural development: Review of complex socio-economic dynamics and causal diagrams for more appropriate energy modelling. *Energy for Sustainable Development*, *43*, 203–223. https://doi.org/https://doi.org/10.1016/j.esd.2018.02.003

Rok, B., & Lusa, L. (2013). Smote for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1), 106–121. https://doi.org/10.1186/1471-2105-14-106

Román, M., Wang, Z., Sun, Q., Kalb, V., Miller, S., Molthan, A., Schultz, L., Bell, J., Stokes, E., Pandey, B., & Seto, K. (2018). *NASA's Black Marble nighttime lights product suite*. https://doi.org/10.1016/j.rse.2018.03.017

Roßbach, P. (2018). Neural networks vs. random forests–does it always have to be deep learning. *Germany: Frankfurt School of Finance and Management*.

Rudin, C., Waltz, D., Anderson, R. N., Boulanger, A., Salleb-Aouissi, A., Chow, M., Dutta, H., Gross, P. N., Huang, B., Ierome, S., Isaac, D. F., Kressner, A., Passonneau, R. J., Radeva, A., & Wu, L. (2012). Machine learning for the new york city power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(2), 328–345. https://doi.org/10.1109/TPAMI.2011.108

Samarakoon, S. (2019). A justice and wellbeing centered framework for analysing energy poverty in the global south. *Ecological Economics*, *165*, 106385. https://doi.org/https://doi.org/10.1016/j.ecolecon.2019.106385

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Santos, J. (2018). Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. *IEEE Computational Intelligence Magazine*, *13*(4), 59–76. https://doi.org/10.1109/MCI.2018.2866730

Schöley, J. (2018). Tricolore. a flexible color scale for ternary compositions.

Schöley, J. (2021). The centered ternary balance scheme: A technique to visualize surfaces of unbalanced three-part compositions. *Demographic Research*, *S29*(19), 443–458. https://doi.org/10.4054/DemRes.2021.44.19

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Sedai, A. K., Nepal, R., & Jamasb, T. (2021). Flickering lifelines: Electrification and household welfare in india. *Energy Economics*, *94*, 104975. https://doi.org/https://doi.org/10.1016/j.eneco.2020.104975

Sedai, A. K., Vasudevan, R., Pena, A. A., & Miller, R. (2021). Does reliable electrification reduce gender differences? evidence from india. *Journal of Economic Behavior & Organization*, *185*, 580–601. https://doi.org/https://doi.org/10.1016/j.jebo.2021.03.015

Segal, M. R. (2004). Machine learning benchmarks and random forest regression. https://escholarship.org/uc/item/35x3v9t4

Sen, A. (2014). Development as freedom (1999). *The globalization and development reader: Perspectives on development and global change*, *525*.

Shah, Z., Klugman, N., Cadamuro, G., Hsu, F.-C., Elvidge, C. D., & Taneja, J. (2022). The electricity scene from above: Exploring power grid inconsistencies using satellite data in accra, ghana. *Applied Energy*, *319*, 119237. https://doi.org/https://doi.org/10.1016/j.apenergy.2022.119237

Shastry, V., & Rai, V. (2021). Reduced health services at under-electrified primary healthcare facilities: Evidence from India. *PLOS ONE*, *16*(6), e0252705. https://doi.org/10.1371/journal.pone.0252705

Shrimali, G., & Sen, V. (2020). Scaling reliable electricity access in india: A public-private partnership model. *Energy for Sustainable Development*, *55*, 69–81. https://doi.org/https://doi.org/10.1016/j.esd.2020.01.006

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical therapy*, *85*(3), 257–268. https://doi.org/https://doi.org/10.1093/ptj/85.3.257

Stover, C. (2021). Ternary Diagram. https://mathworld.wolfram.com/TernaryDiagram.html

Szakonyi, D., & Urpelainen, J. (2013). Electricity sector reform and generators as a source of backup power: The case of india. *Energy for Sustainable Development*, *17*(5), 477–481. https://doi.org/https://doi.org/10.1016/j.esd.2013.05.006

Tait, L. (2017). Towards a multidimensional framework for measuring household energy access: Application to south africa. *Energy for Sustainable Development*, *38*, 1–9. https://doi.org/https://doi.org/10.1016/j.esd.2017.01.007

The DHS Program. (n.d.). The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs. https://dhsprogram.com/

Tomek, I. (1976). Two modifications of cnn. *IEEE Trans. Systems, Man and Cybernetics*, *6*, 769–772.

Toubeau, J.-F., Pardoen, L., Hubert, L., Marenne, N., Sprooten, J., De Grève, Z., & Vallée, F. (2022). Machine learning-assisted outage planning for maintenance activities in power systems with renewables. *Energy*, *238*, 121993. https://doi.org/https://doi.org/10.1016/j.energy.2021.121993

Tully, S. (2006). The human right to access electricity. *The Electricity Journal*, *19*(3), 30–39. https://doi.org/https://doi.org/10.1016/j.tej.2006.02.003

Ulsrud, K., Rohracher, H., Winther, T., Muchunku, C., & Palit, D. (2018). Pathways to electricity for all: What makes village-scale solar power successful? *Energy Research & Social Science*, *44*, 32–40. https://doi.org/https://doi.org/10.1016/j.erss.2018.04.027

United Nations. (2021). *The sustainable development goals report 2021*. United Nations Publications. https://unstats.un.org/sdgs/report/2021/The-Sustainable-Development-Goals-Report-2021.pdf

Urpelainen, J. (2016). Energy poverty and perceptions of solar power in marginalized communities: Survey evidence from uttar pradesh, india. *Renewable Energy*, *85*, 534–539. https://doi.org/https://doi.org/10.1016/j.renene.2015.07.001

van der Kroon, B., Brouwer, R., & van Beukering, P. J. (2013). The energy ladder: Theoretical myth or empirical truth? results from a meta-analysis. *Renewable and Sustainable Energy Reviews*, *20*, 504–513. https://doi.org/https://doi.org/10.1016/j.rser.2012.11.045

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/https://doi.org/10.1007/s10994-019-05855-6

Wang, H., Maruejols, L., & Yu, X. (2021). Predicting energy poverty with combinations of remote-sensing and socioeconomic survey data in india: Evidence from machine learning. *Energy Economics*, *102*, 105510. https://doi.org/https://doi.org/10.1016/j.eneco.2021.105510

Watmough, G. R., Atkinson, P. M., Saikia, A., & Hutton, C. W. (2016). Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: An example from assam, india. *World Development*, *78*, 188–203. https://doi.org/https://doi.org/10.1016/j.worlddev.2015.10.031

Watmough, G. R., Marcinko, C. L. J., Sullivan, C., Tschirhart, K., Mutuo, P. K., Palm, C. A., & Svenning, J.-C. (2019). Socioecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences*, *116*(4), 1213–1218. https://doi.org/10.1073/pnas.1812969116

Wong, J. C. Y., Blankenship, B., Urpelainen, J., Ganesan, K., Bharadwaj, K., & Balani, K. (2021). Perceptions and acceptability of electricity theft: Towards better public service provision. *World Development*, *140*, 105301. https://doi.org/https://doi.org/10.1016/j.worlddev.2020.105301

World Bank. (2019). Access to electricity. https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=IN

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, *77*(1). https://doi.org/10.18637/jss.v077.i01

Wu, L. L., Kaiser, G. E., Rudin, C., Waltz, D. L., Anderson, R. N., Boulanger, A. G., Salleb-Aouissi, A., Dutta, H., & Pooleery, M. (2011). *Evaluating Machine Learning for Improving Power Grid Reliability* (tech. rep.). Department of Computer Science, Columbia University. https://academiccommons.columbia.edu/doi/10.7916/D8SB4F1Z

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016). *Transfer learning from deep features for remote sensing and poverty mapping*. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12196/12181

Xu, H., Kinfu, K. A., LeVine, W., Panda, S., Dey, J., Ainsworth, M., Peng, Y.-C., Kusmanov, M., Engert, F., White, C. M., Vogelstein, J. T., & Priebe, C. E. (2021). When are deep networks really better

than decision forests at small sample sizes, and how? https://doi.org/10.48550/ARXIV.2108.13637

Yang, F., Wanik, D. W., Cerrai, D., Bhuiyan, M. A. E., & Anagnostou, E. N. (2020). Quantifying uncertainty in machine learning-based power outage prediction model training: A tool for sustainable storm restoration. *Sustainability*, *12*(4). https://doi.org/10.3390/su12041525

A

Access Attributes Defined

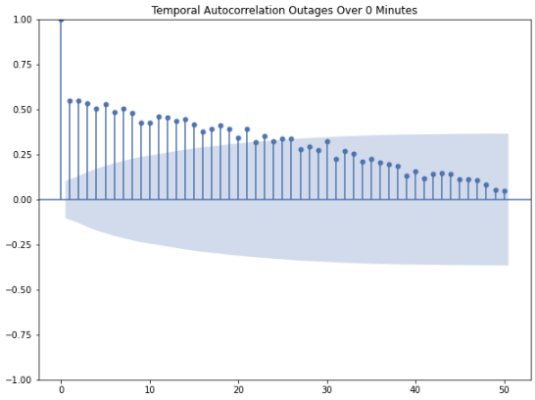
Bhatia and Angelou (2015) present clear definitions for the attributes of access. These are described in table A.1 for an easy overview.

Table A.1: Electricity Access Attributes Bhatia and Angelou (2015)

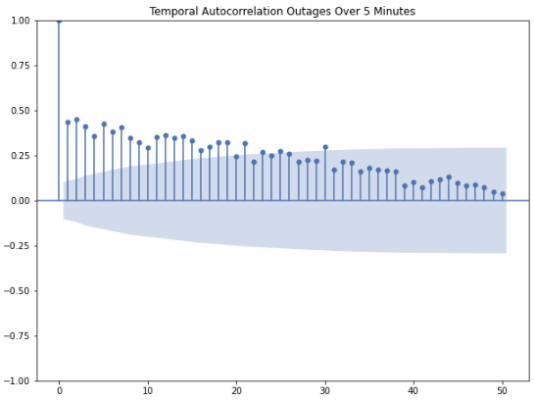
Attribute	**Definition**
Capacity	Refers to the amount of electricity made accessible to the user. It is defined as the sum of all energy available over a certain time period and the greatest power (rate of electricity delivery) that may be utilized.
Availability	Implies the ability to extract electricity as needed for usage. The period throughout the day (and night) when electricity is available, or the total number of hours when electricity is available each day, can be used to calculate availability.
Reliability	Implies the lack of unpredictability in electricity supply. It is determined by the frequency and duration of unexpected outages.
Quality	It denotes the right amount and stability of voltage and frequency.
Affordability	End-user ability to pay for electricity required for a given package of electricity use. Affordability includes one-time connection fees, electricity fees, capacity fees, maintenance fees, and replacement fees. Affordability is determined by the defined package, the price of electricity (including all of the aforementioned fees), and the user's income level. When the cost of electricity for a specific package of electricity use does not exceed a normative proportion of household income, energy supply is deemed affordable.
Legality	Implies that the end-user is not engaging in any illegal conduct by using the electricity supply
Health and Safety	Concerns the possibility of negative health implications from the usage of electricity.

B

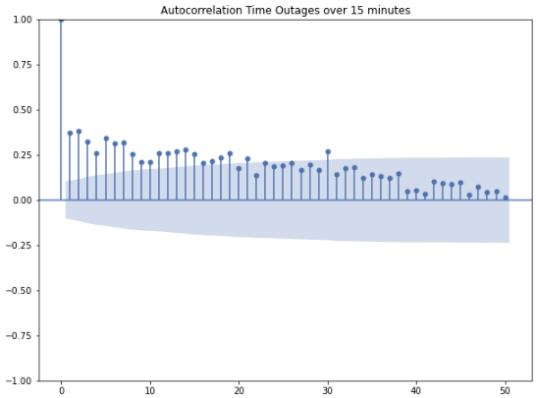
Temporal Autocorrelations



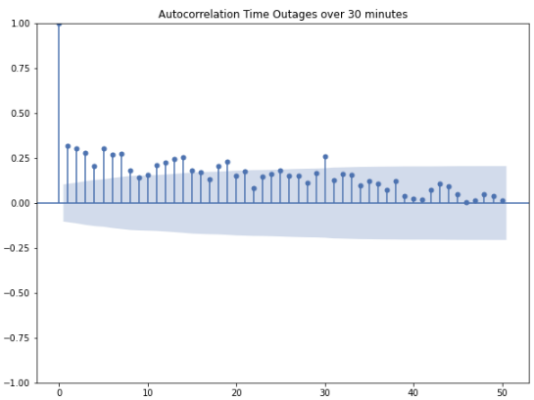
(a) Temporal autocorrelation outages over 0 minutes



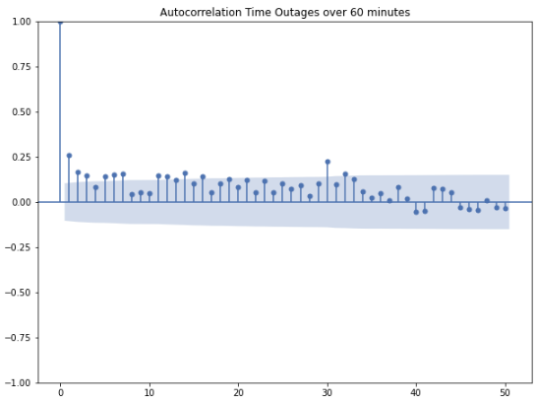
(b) Temporal autocorrelation outages over 5 minutes



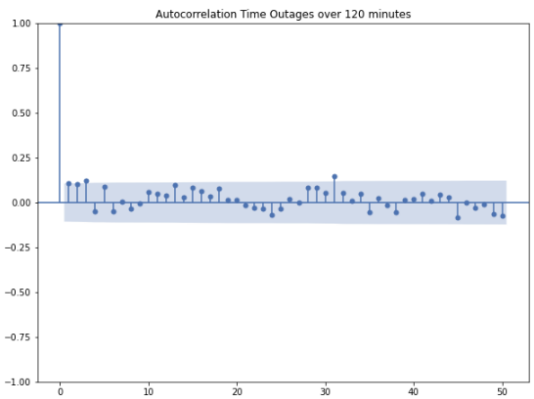
(c) Temporal autocorrelation outages over 15 minutes



(d) Temporal autocorrelation outages over 30 minutes



(e) Temporal autocorrelation outages over 60 minutes



(f) Temporal autocorrelation outages over 120 minutes

Figure B.1: Temporal autocorrelation outages of over 0 and 120 minutes

C

Hyperparameter Tuning

C.1. Target Over 0 Minutes Outage

C.1.1. Number of trees

The number of trees in the forest is denoted by the hyperparameter n_estimators. Generally, the more trees there are, the better RF fits and learns from the data (Oshiro et al., 2012). A relatively small number may result in overfitting, while adding numerous trees severely slows down the training process and does not appreciably increase performance. As a result, a compromise must be struck between accuracy and efficiency. We evaluated 50, 100, 150, 200, 250, 300, 350, and 400 trees for this parameter. Figure C.1 and appendix C.2.1 depict the F1 score values for algorithms trained and cross-validated with various numbers of trees. The tests seem to show that the number of trees do not have a big influence on the algorithm's performance. Therefore, the default setting of 100 trees will be used.

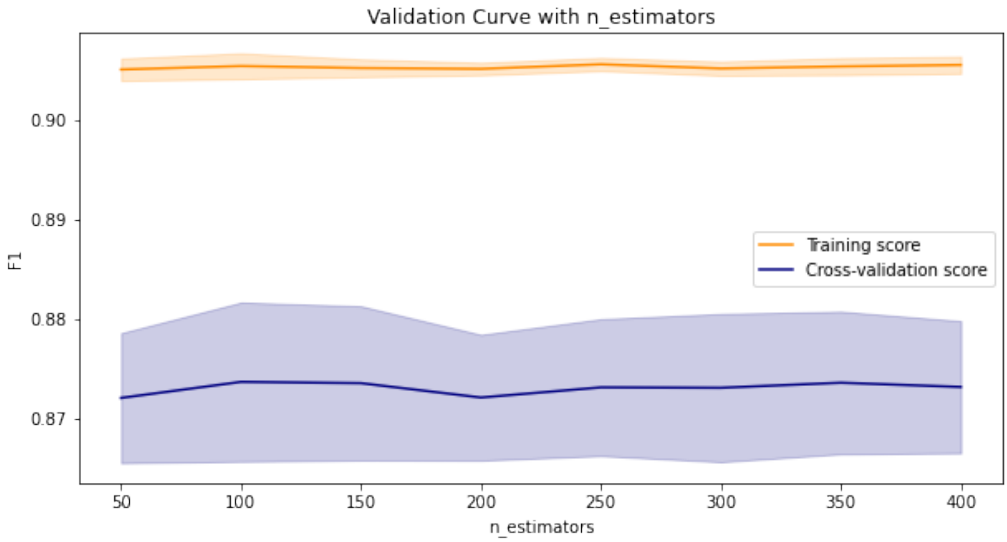


Figure C.1: F1 score for the number of trees

C.1.2. Max depth

The maximum depth hyperparameter pertains to the depth of a tree in RF. Deeper trees have a higher number of splits. The settings evaluated for this hyperparameter ranged from 5 to 50. Figure C.2 and appendix C.2.2 depict the F1 score values for algorithms trained and cross-validated with varying tree depths. The best results are obtained when the depth of each tree in the forest is set to 10. The train and cross-validation curves follow distinct patterns. With an increase in depth, the F1 score for the training dataset increases as well. At a depth of 20, the training F1 score becomes 1. However, beyond a value of 10, the cross-validation score no longer increases. This occurs when the model overfits the training data when the depth of each tree is quite big. Thus, a maximum depth of 20 will be used in this study.

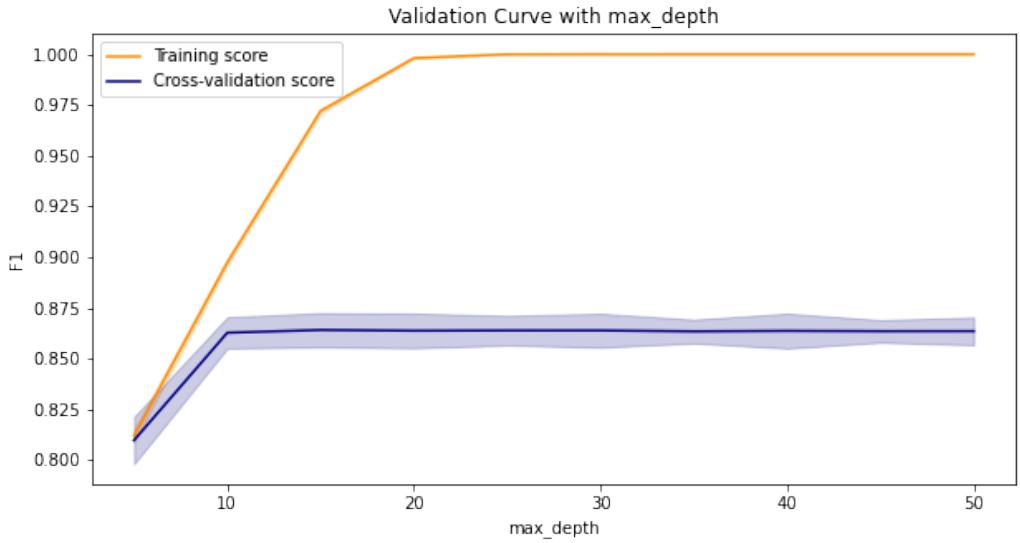


Figure C.2: F1 score for the maximum depth of trees

C.1.3. Min samples split

The min_samples_split hyperparameter specifies the minimal number of samples from the dataset are necessary to separate an internal node. Its value can range from 1 to the total number of samples at each node. The impact on the model for sample sizes ranging from 1 to 320 is examined. The F1 score values for the models trained and evaluated with varied values for the min samples split hyperparameter are shown in figure C.3 and appendix C.2.3. The findings show that when the minimum number of samples necessary to divide an internal node is 8, the best performance is attained out of all the samples evaluated. After 8, the F1 score value gradually decreases. The smaller the value of min samples split, the better the algorithm's performance, but the bigger the risk of overfitting. At a value of 8, cross-validation performance does not degrade, but there is less risk of overfitting.

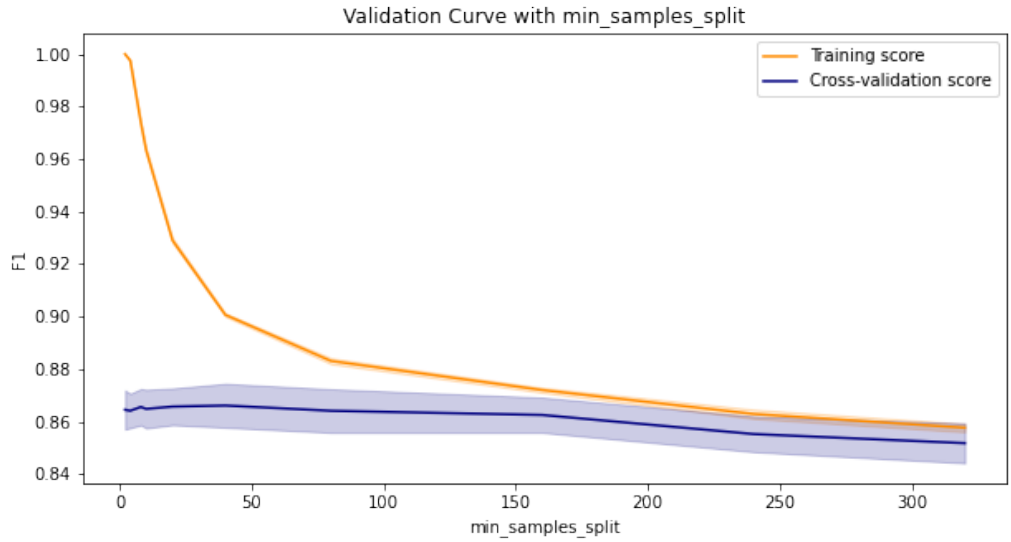


Figure C.3: F1 score for the minimum samples in a split

C.1.4. Min samples leaf

The hyperparameter min_samples_leaf specifies the number of samples from the dataset that must be present at a leaf node. It is analogous to the minimum sample split hyperparameter. However, it is about the leaves. This hyperparameter's value can vary from 1 to all the samples at each node. The impacts on the algorithm for sample values ranging from 1 to 120 are examined. The F1 score values for the algorithms trained and cross-validated with varied values for the min_samples_leaf hyperparameter are shown in figure C.4 and appendix C.2.4. The findings show that the highest performance is attained when the minimum number of samples at a leaf node is 4. After 4, the F1 score value begins to steadily fall, following a declining trend as the number of samples utilized increases. The lower the amount of min samples leaf, the better the algorithm's performance, although when smaller values are utilized, the method tends to overfit. At 4, there appears to be a decent balance between the F1 training and cross-validation scores.

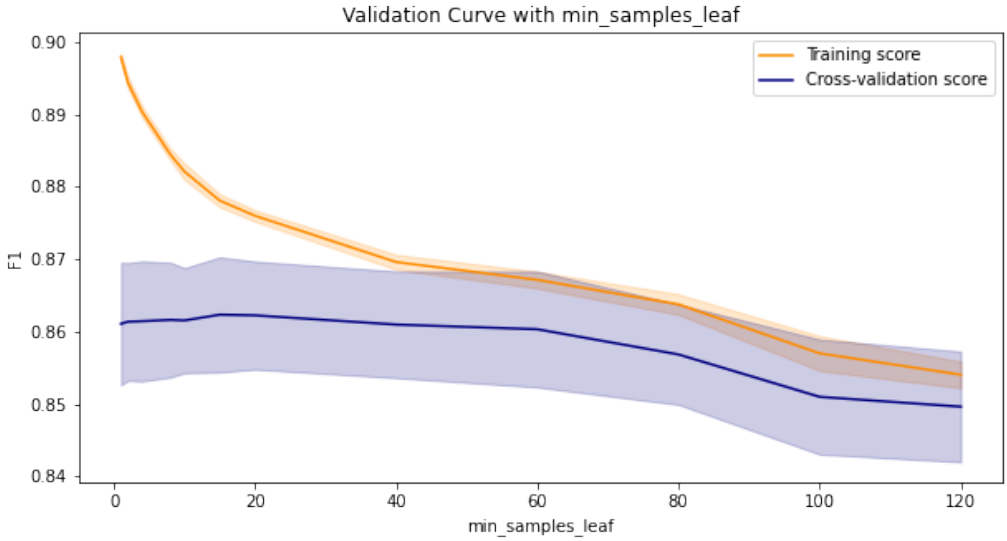


Figure C.4: F1 score for the minimum samples in a leaf

C.1.5. Max features

The max features hyperparameter specifies the amount of features to evaluate while calculating the optimal split. The values examined for this hyperparameter vary from 1 to a maximum of 18 characteristics. Figure C.5 and appendix C.2.5 depict the F1 score values for the algorithms trained and cross-validated with different hyperparameter max values. The experiments show that this hyperparameter has little influence on the performance. Therefore, the default value of sklearn's RF will be used, which is close to 4.

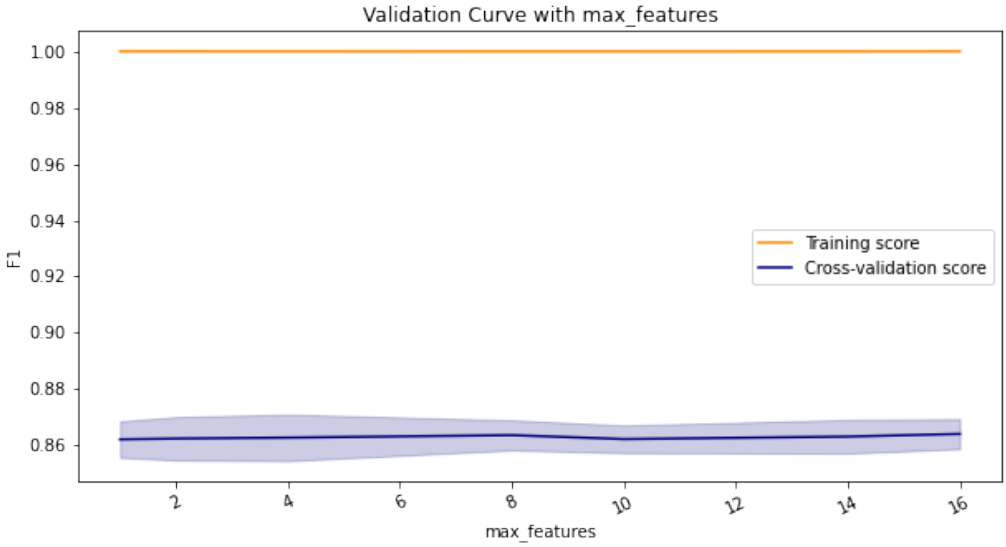


Figure C.5: F1 score for the maximum features

C.1.6. Bootstrap

The bootstrap hyperparameter specifies whether bootstrap samples are utilized for generating trees, and its value can be either False or True. Figure C.6 and appendix C.2.6 show the results of the two tests. There is barely any difference between the two settings. Thus, the default setting will be used.

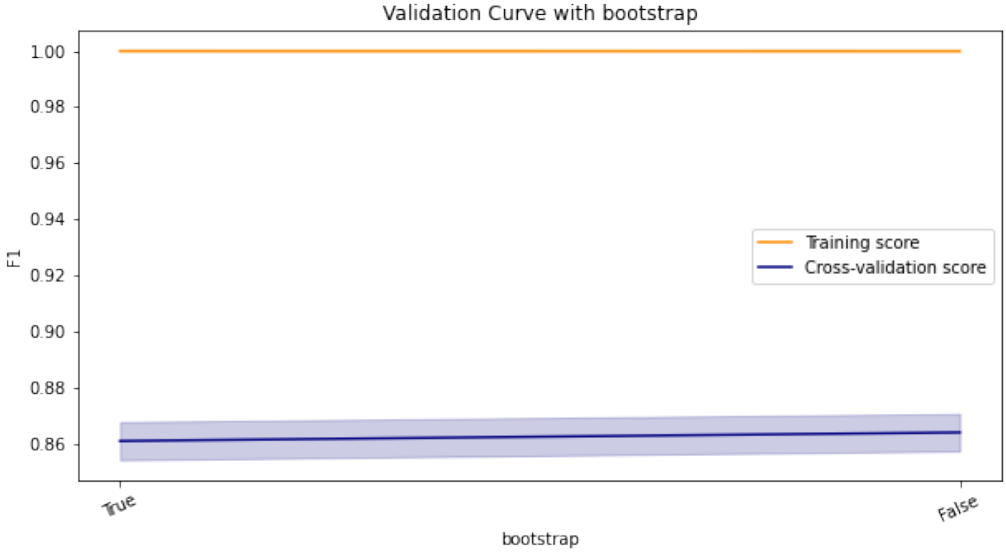
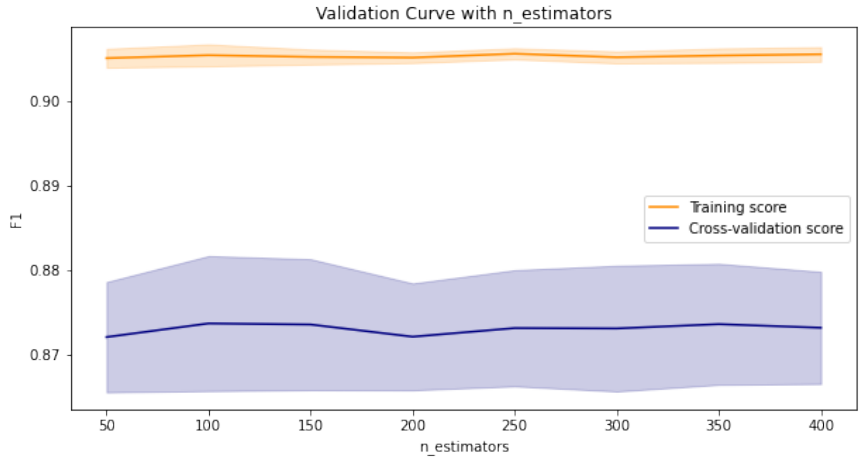


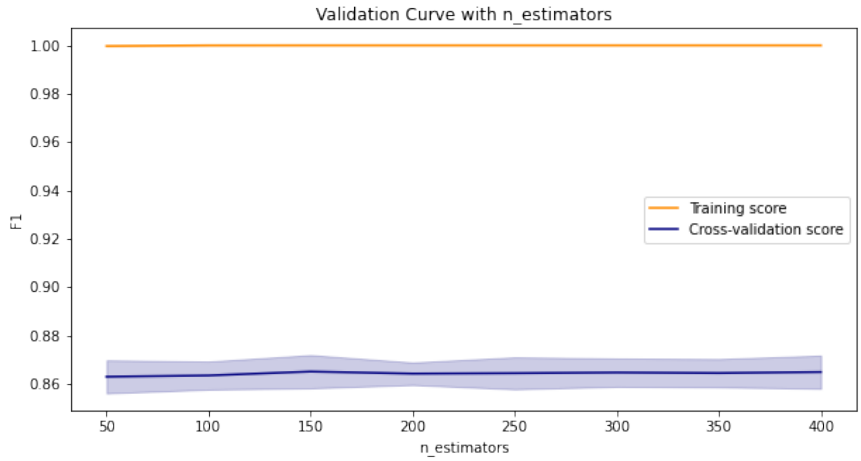
Figure C.6: F1 score for bootstrap

C.2. All Target Variables

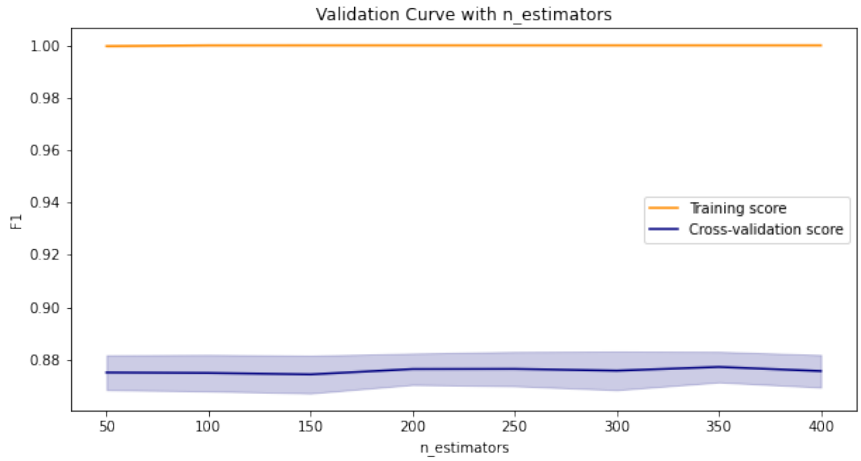
C.2.1. N Estimators



(a) F1 score for the number of trees outages over 0 minutes

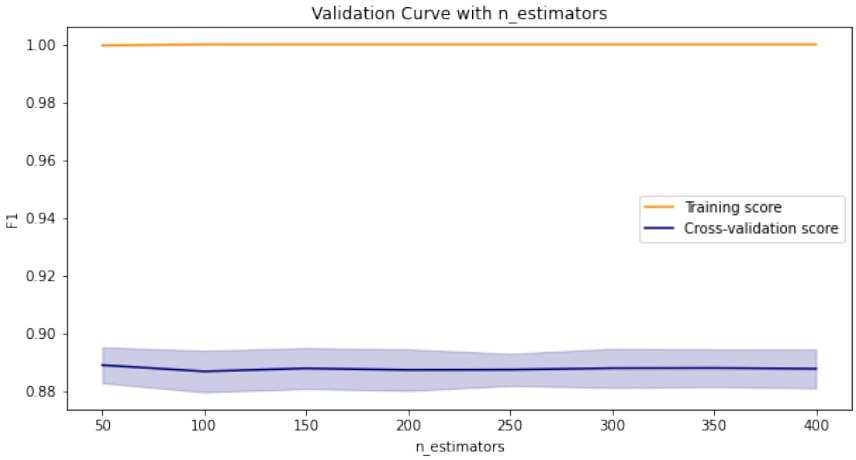


(b) F1 score for the number of trees outages over 5 minutes

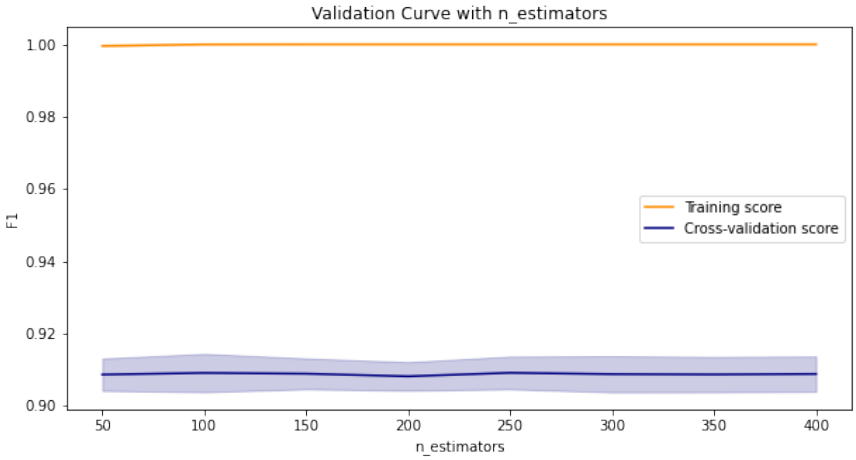


(c) F1 score for the number of trees outages over 15 minutes

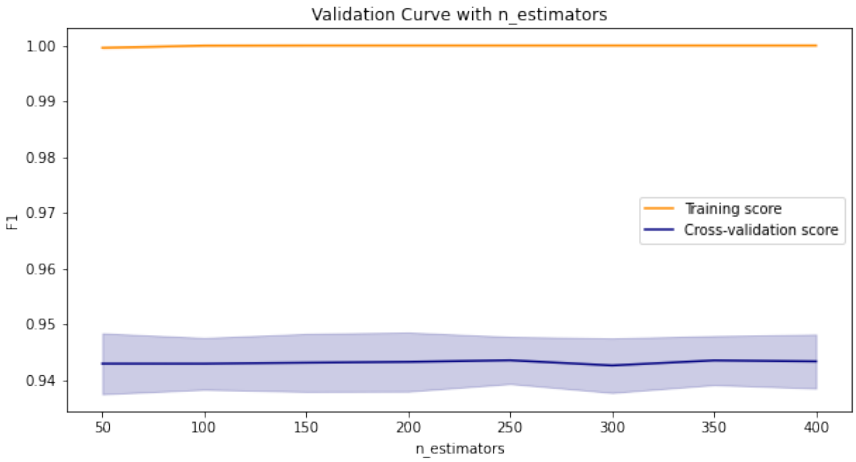
Figure C.7: F1 scores for the number of trees for outages over 0, 5, 15



(a) F1 score for the number of trees outages over 30 minutes



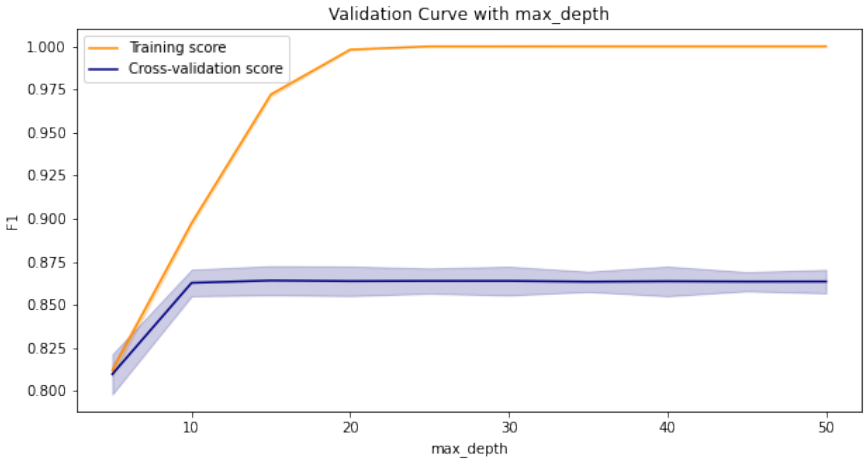
(b) F1 score for the number of trees outages over 60 minutes



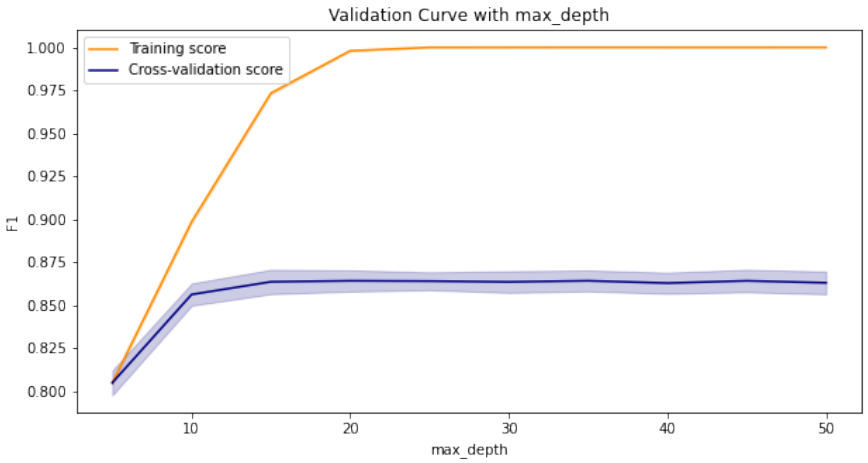
(c) F1 score for the number of trees outages over 120 minutes

Figure C.8: F1 scores for the number of trees for outages over 30, 60, 120

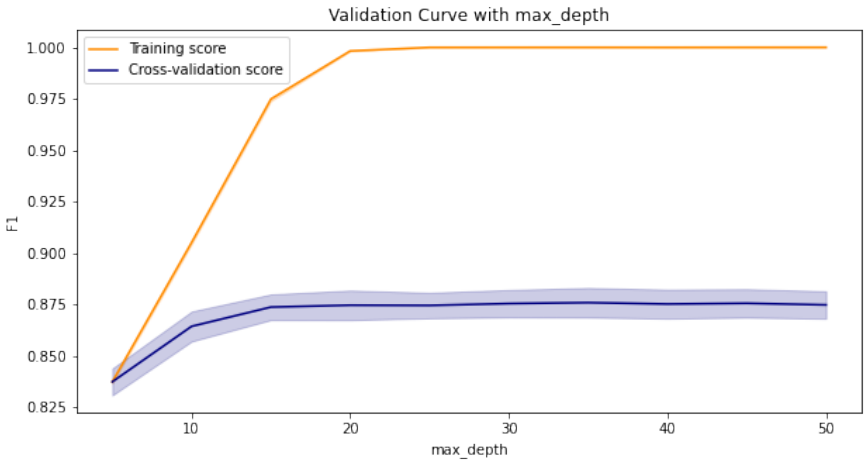
C.2.2. Max Depth



(a) F1 score for the max depth outages over 0 minutes

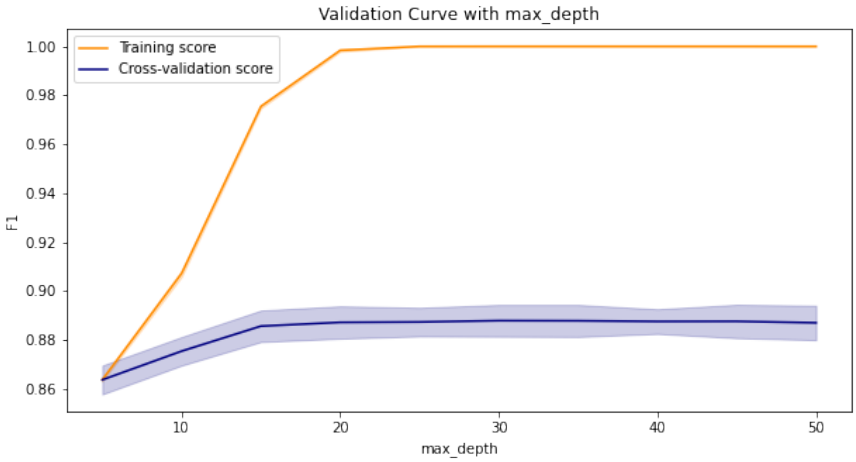


(b) F1 score for the max depth outages over 5 minutes

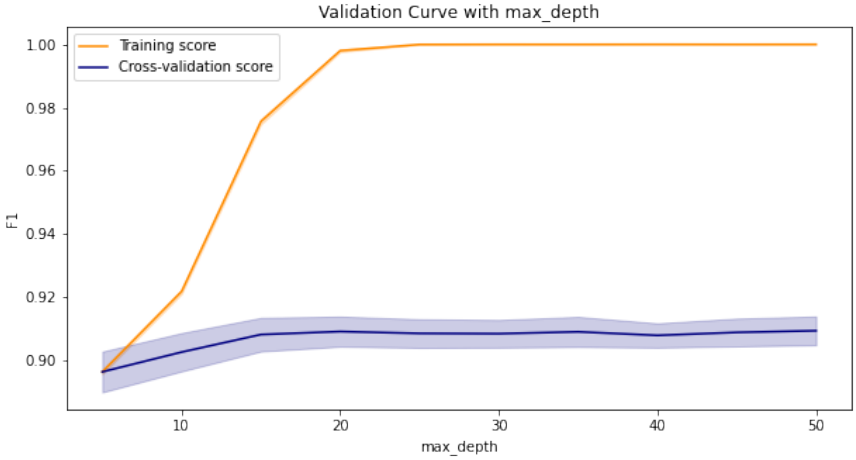


(c) F1 score for the max depth outages over 15 minutes

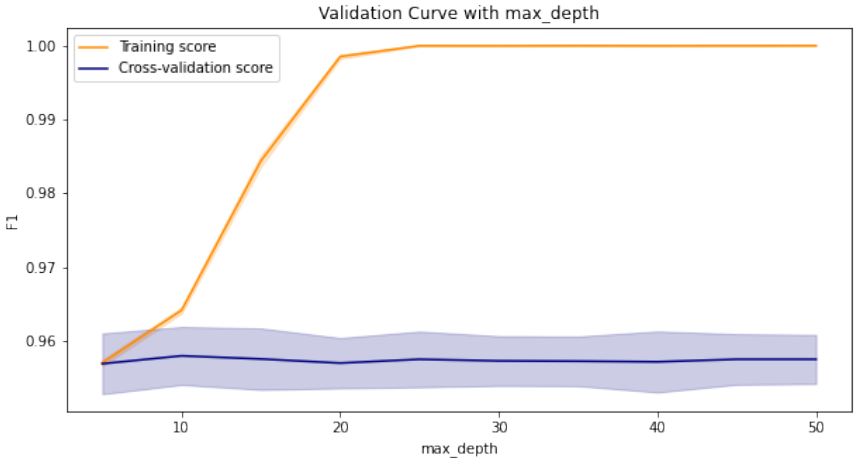
Figure C.9: F1 scores for the max depth for outages over 0, 5, 15



(a) F1 score for the max depth outages over 30 minutes



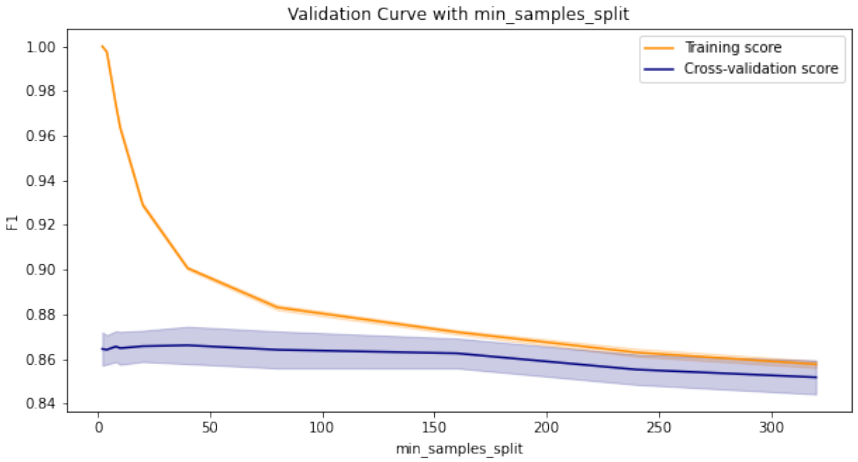
(b) F1 score for the max depth outages over 60 minutes



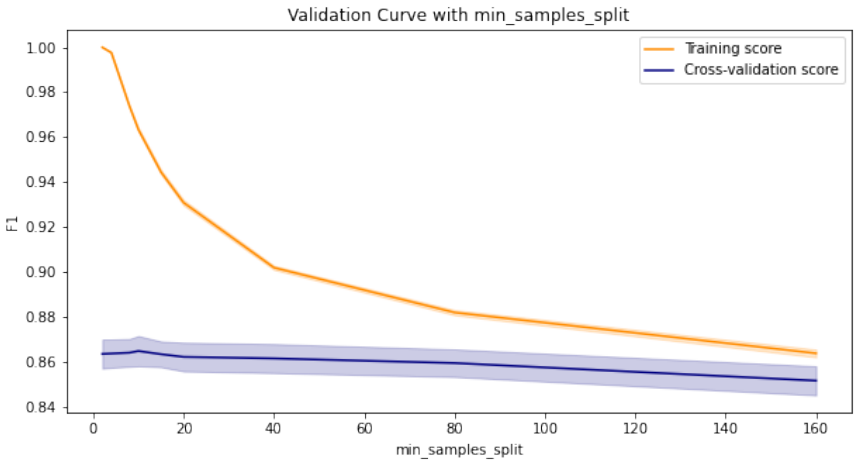
(c) F1 score for the max depth outages over 120 minutes

Figure C.10: F1 scores for the max depth for outages over 30, 60, 120

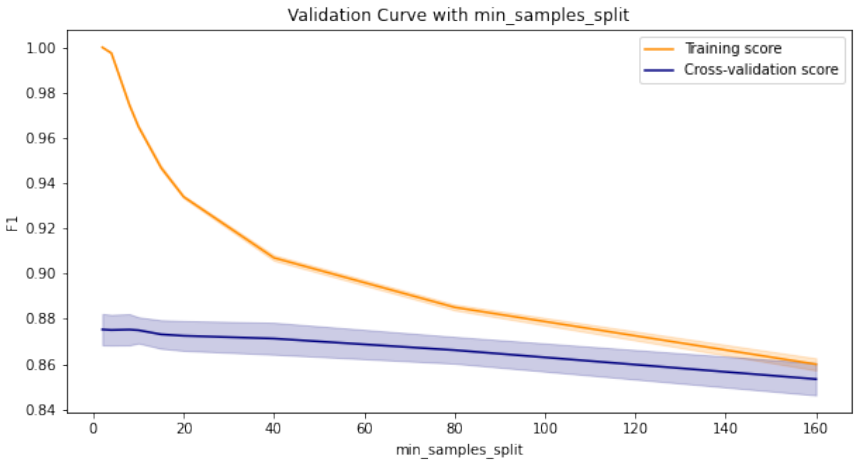
C.2.3. Min Samples Split



(a) F1 score for the min samples split over 0 minutes

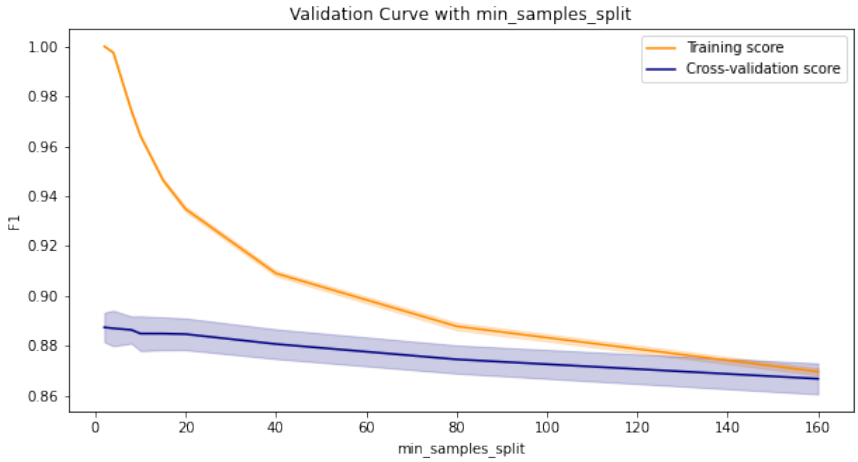


(b) F1 score for the min samples split over 5 minutes

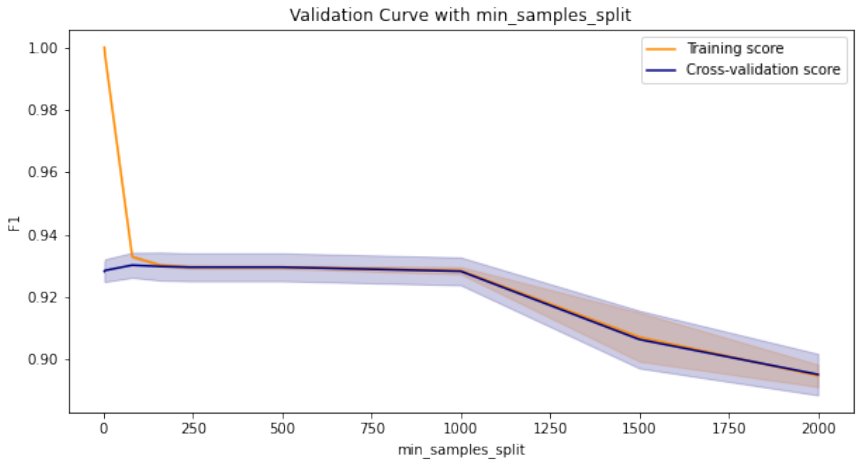


(c) F1 score for the min samples split outages over 15 minutes

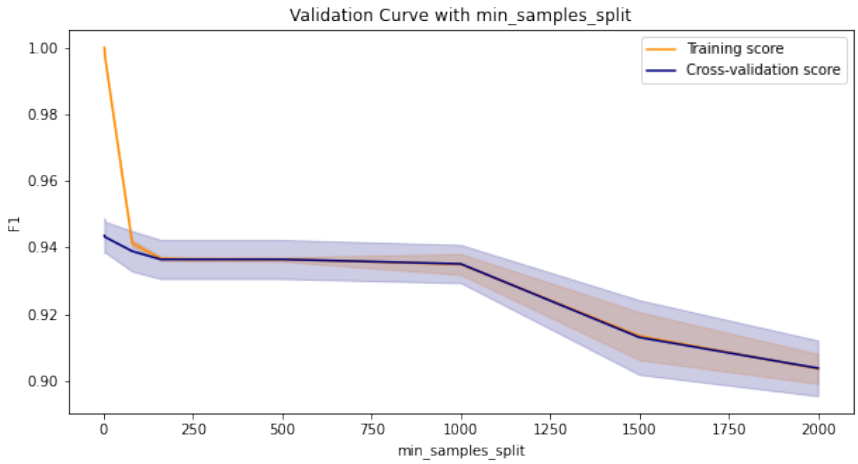
Figure C.11: F1 scores for the min samples split for outages over 0, 5, 15



(a) F1 score for the min samples split outages over 30 minutes



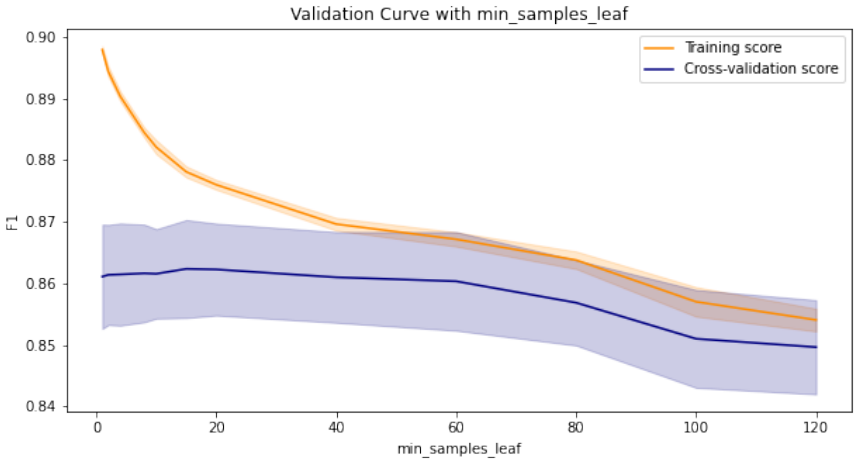
(b) F1 score for the min samples split outages over 60 minutes



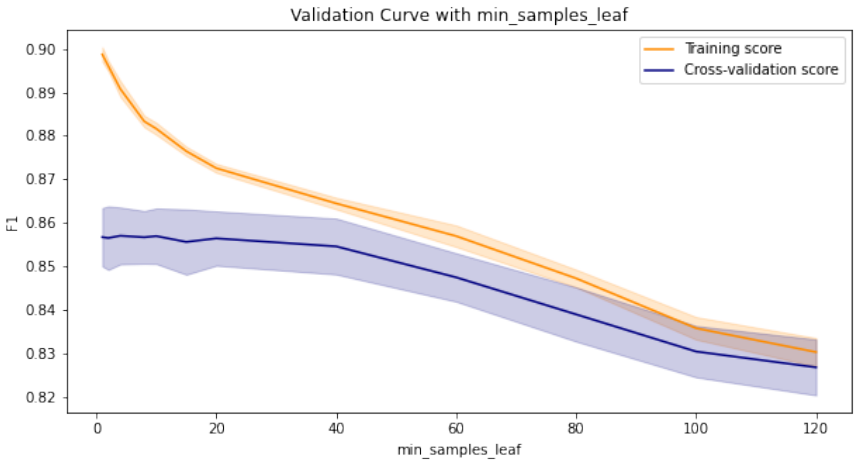
(c) F1 score for the min samples split outages over 120 minutes

Figure C.12: F1 scores for the min samples split for outages over 30, 60, 120

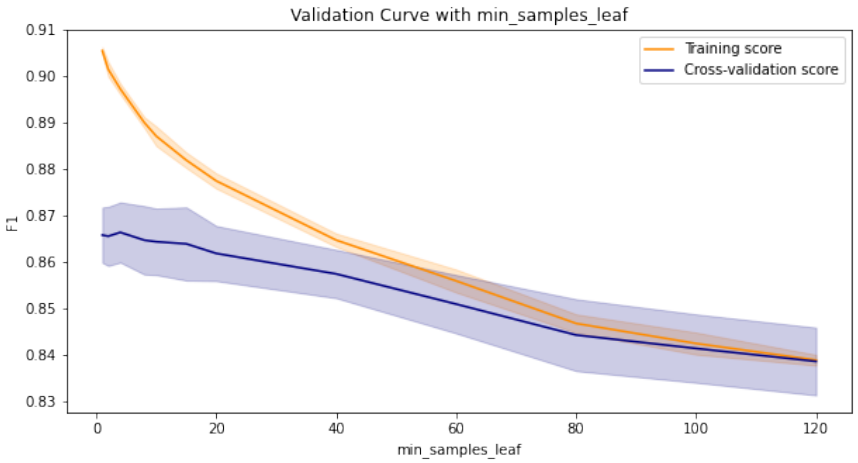
C.2.4. Min Samples Leaf



(a) F1 score for the min samples leaf over 0 minutes

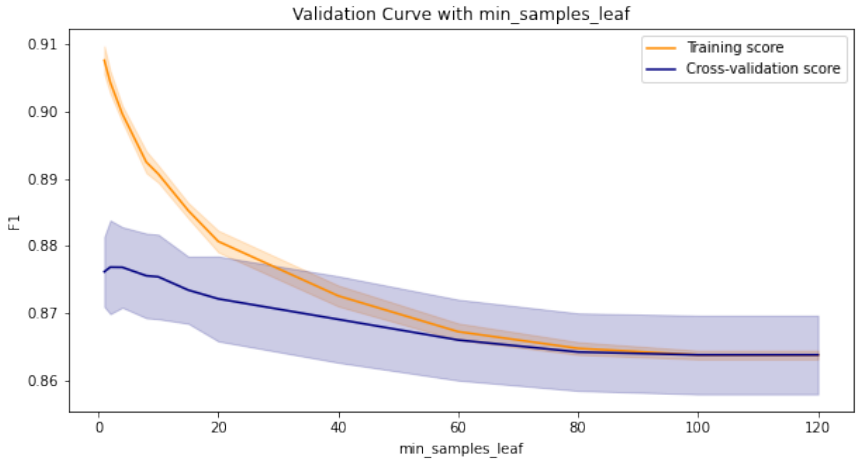


(b) F1 score for the min samples leaf over 5 minutes

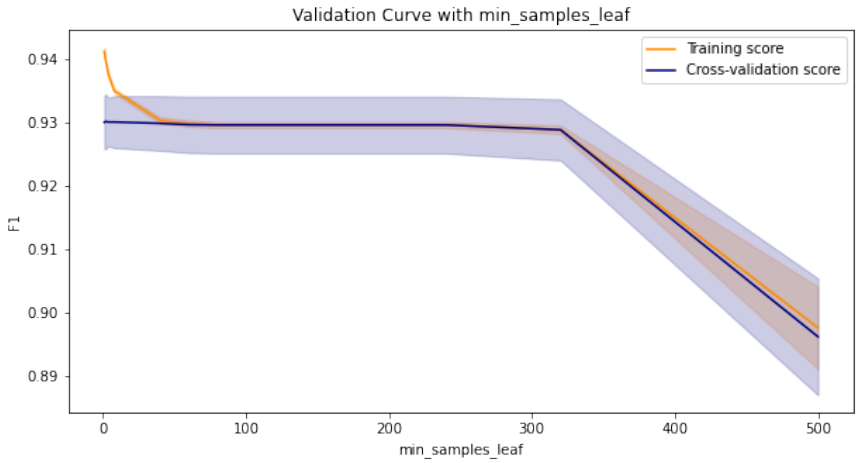


(c) F1 score for the min samples leaf outages over 15 minutes

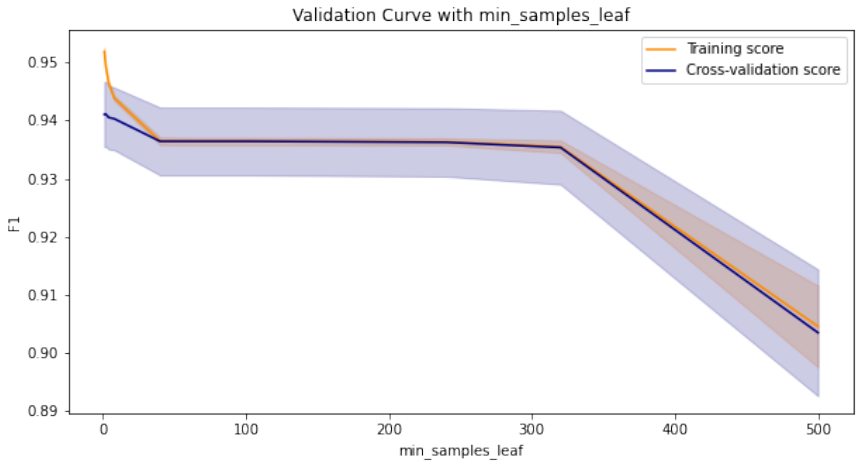
Figure C.13: F1 scores for the min samples leaf for outages over 0, 5, 15



(a) F1 score for the min samples leaf outages over 30 minutes



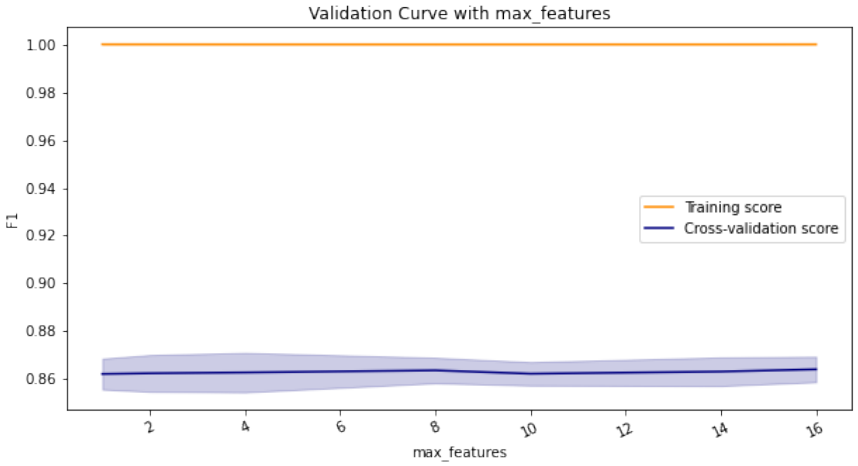
(b) F1 score for the min samples leaf outages over 60 minutes



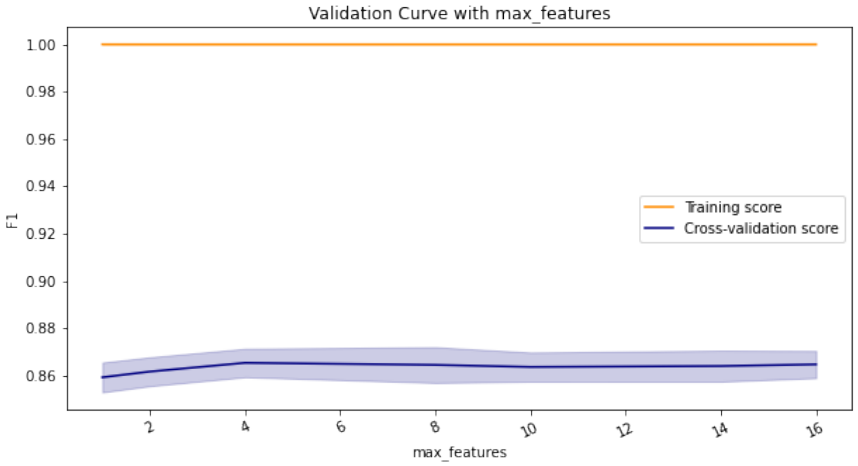
(c) F1 score for the min samples leaf outages over 120 minutes

Figure C.14: F1 scores for the min samples leaf for outages over 30, 60, 120

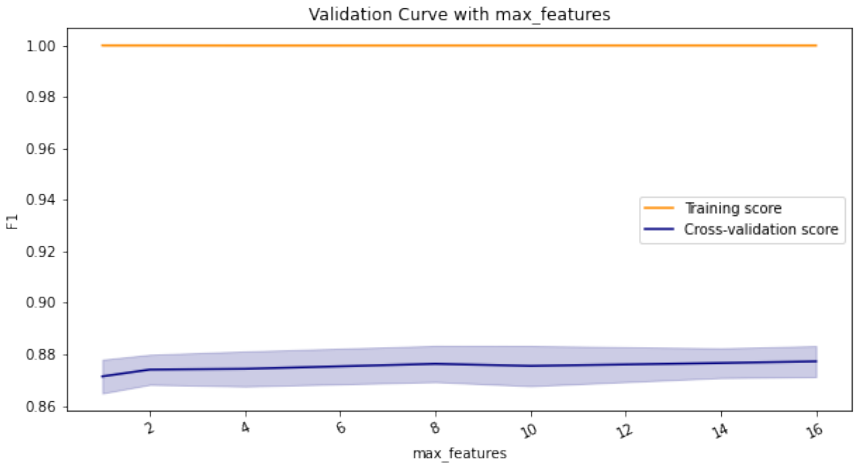
C.2.5. Max Features



(a) F1 score for the max features outages over 0 minutes

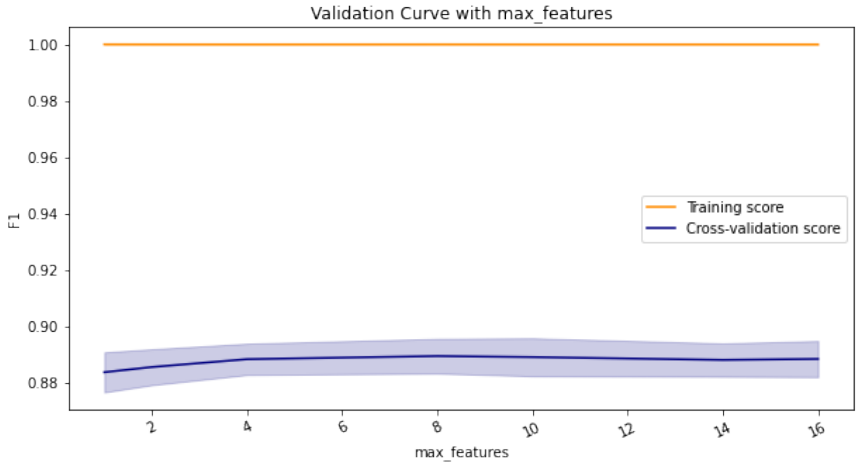


(b) F1 score for the max features outages over 5 minutes

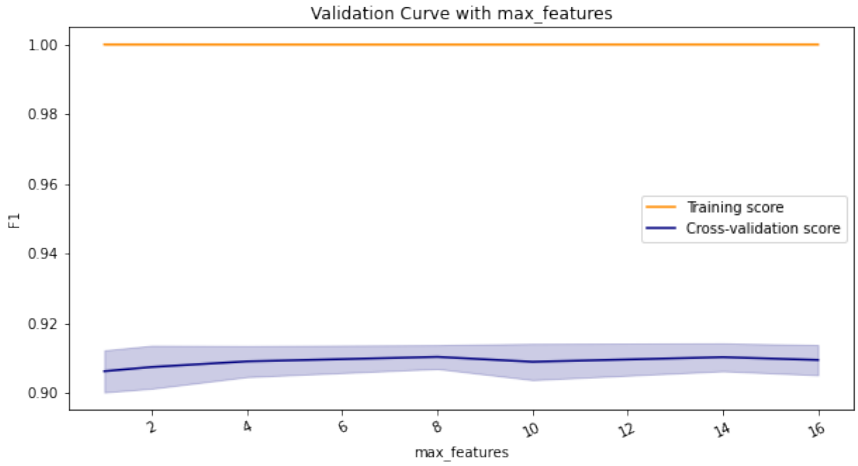


(c) F1 score for the max features outages over 15 minutes

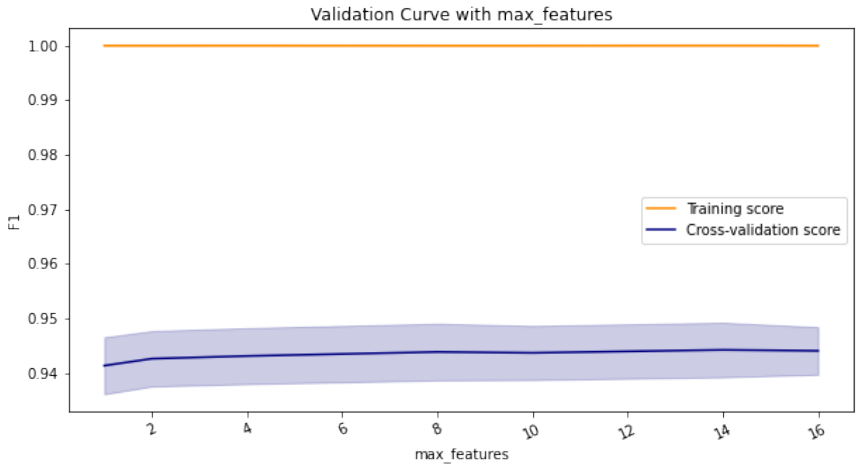
Figure C.15: F1 scores for the max features for outages over 0, 5, 15



(a) F1 score for the max features outages over 30 minutes



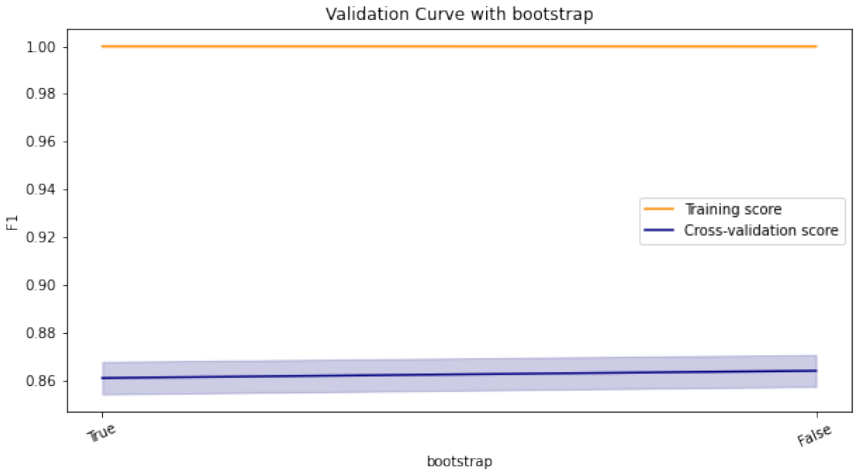
(b) F1 score for the max features outages over 60 minutes



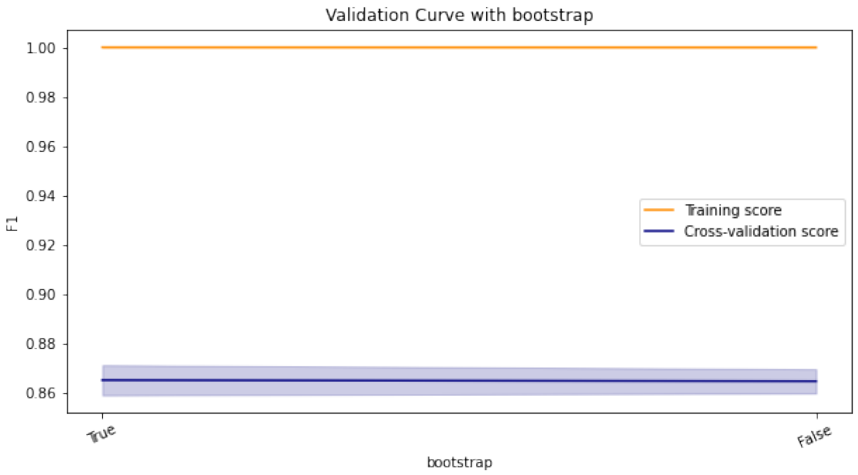
(c) F1 score for the max features outages over 120 minutes

Figure C.16: F1 scores for the max features for outages over 30, 60, 120

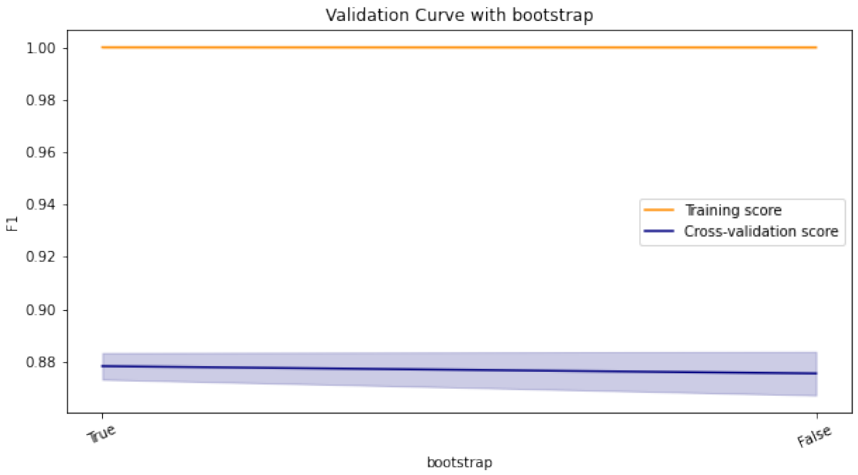
C.2.6. Bootstrap



(a) F1 score for the bootstrap outages over 0 minutes

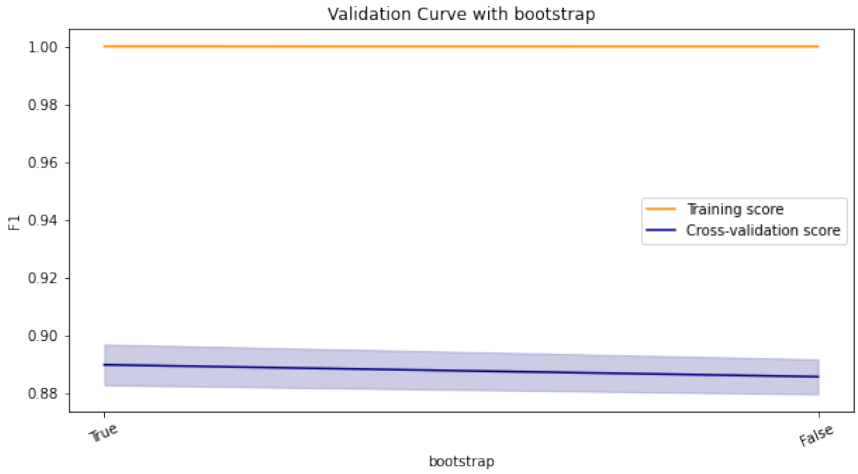


(b) F1 score for the bootstrap outages over 5 minutes

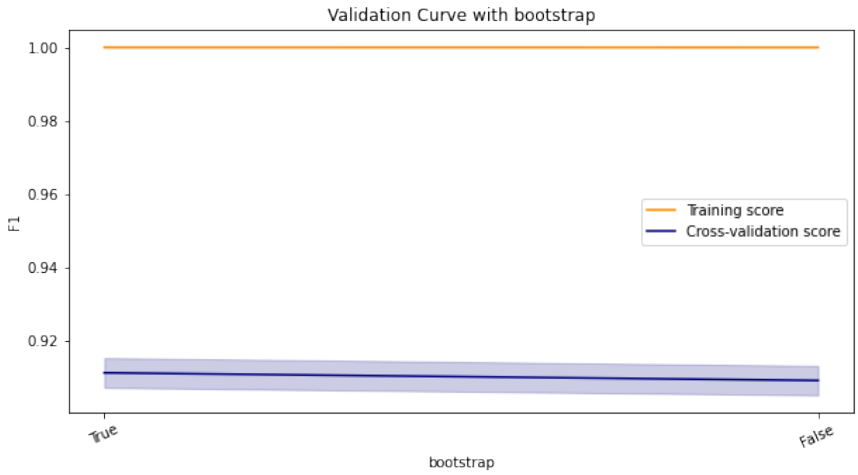


(c) F1 score for the bootstrap outages over 15 minutes

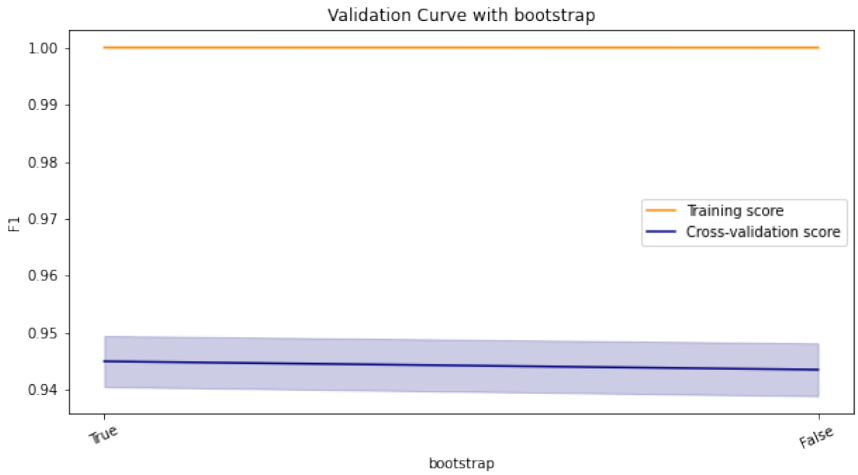
Figure C.17: F1 scores for the bootstrap for outages over 0, 5, 15



(a) F1 score for the bootstrap outages over 30 minutes



(b) F1 score for the bootstrap outages over 60 minutes



(c) F1 score for the bootstrap outages over 120 minutes

Figure C.18: F1 scores for the bootstrap for outages over 30, 60, 120

D

Results (stratified) 10-Fold All Target Variables

Table D.1: Scores 10-Fold cross-validation over 0

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.94	0.88
Outage	0.69	0.44	0.54

Table D.2: Scores stratified 10-fold cross-validation over 0

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.93	0.87
Outage	0.69	0.44	0.53

Table D.3: Scores 10-Fold cross-validation over 5

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.83	0.95	0.89
Outage	0.62	0.27	0.38

Table D.4: Scores stratified 10-fold cross-validation over 5

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.83	0.95	0.89
Outage	0.62	0.28	0.39

Table D.5: Scores 10-Fold cross-validation over 15

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.85	0.97	0.90
Outage	0.58	0.17	0.26

Table D.6: Scores stratified 10-fold cross-validation over 15

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.85	0.97	0.90
Outage	0.57	0.16	0.25

Table D.7: Scores 10-Fold cross-validation over 30

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.86	0.99	0.92
Outage	0.53	0.08	0.14

Table D.8: Scores stratified 10-fold cross-validation over 30

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.86	0.99	0.92
Outage	0.56	0.09	0.15

Table D.9: Scores 10-Fold cross-validation over 60

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.89	1.00	0.94
Outage	0.51	0.04	0.07

Table D.10: Scores stratified 10-fold cross-validation over 60

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.89	1.00	0.94
Outage	0.51	0.03	0.7

Table D.11: Scores 10-Fold cross-validation over 120

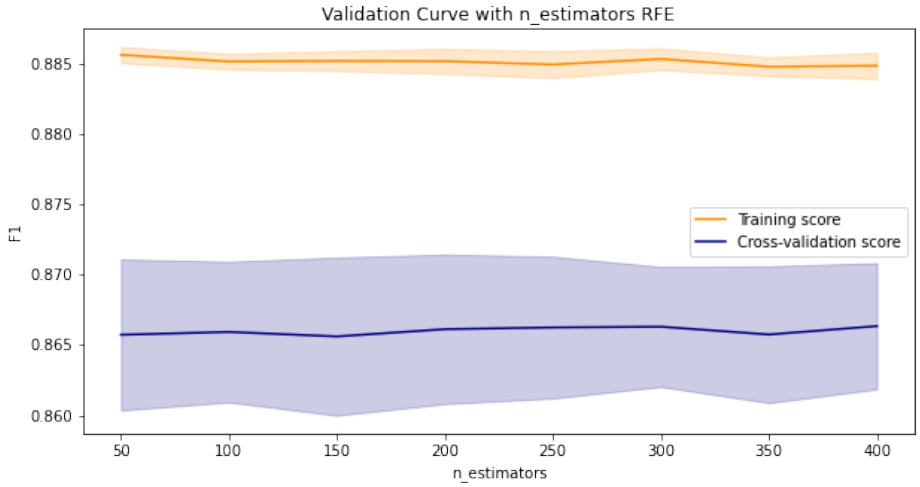
Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.93	1.00	0.96
Outage	0.51	0.04	0.07

Table D.12: Scores stratified 10-fold cross-validation over 120

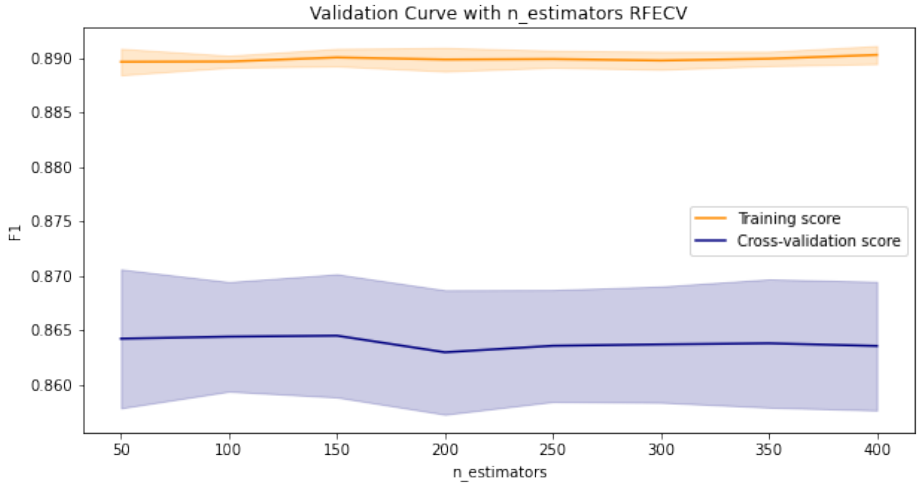
Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.93	1.00	0.96
Outage	0.60	0.04	0.07

E

Feature Selection Hyperparameter Tuning

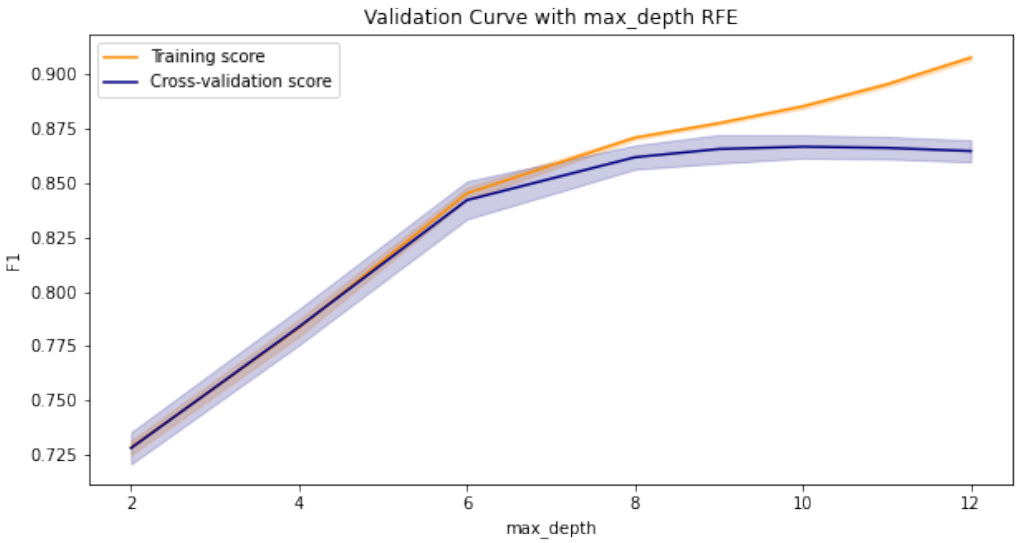


(a) F1 score for the number of trees RFE over 0 minutes

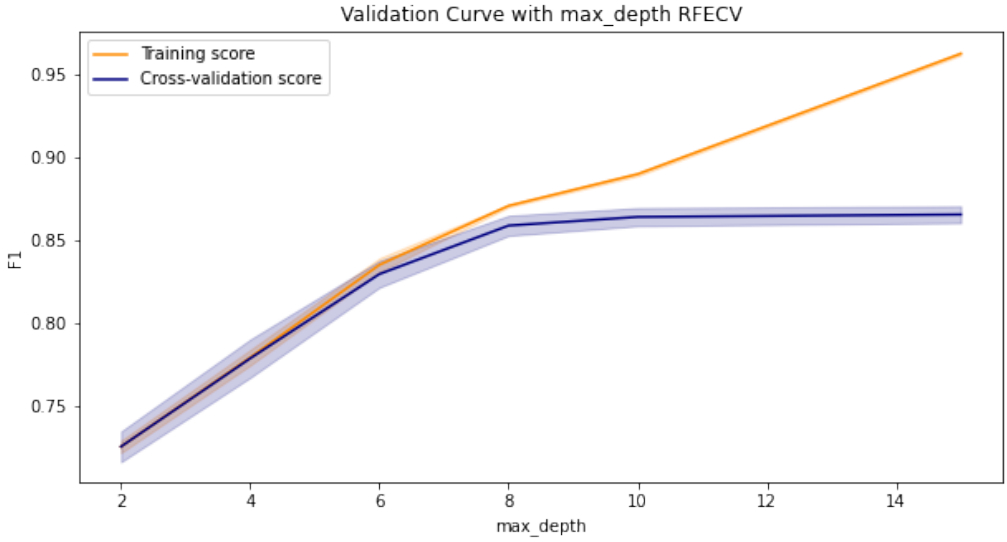


(b) F1 score for the number of trees RFECV over 0 minutes

Figure E.1: F1 scores for the number of trees of feature selections

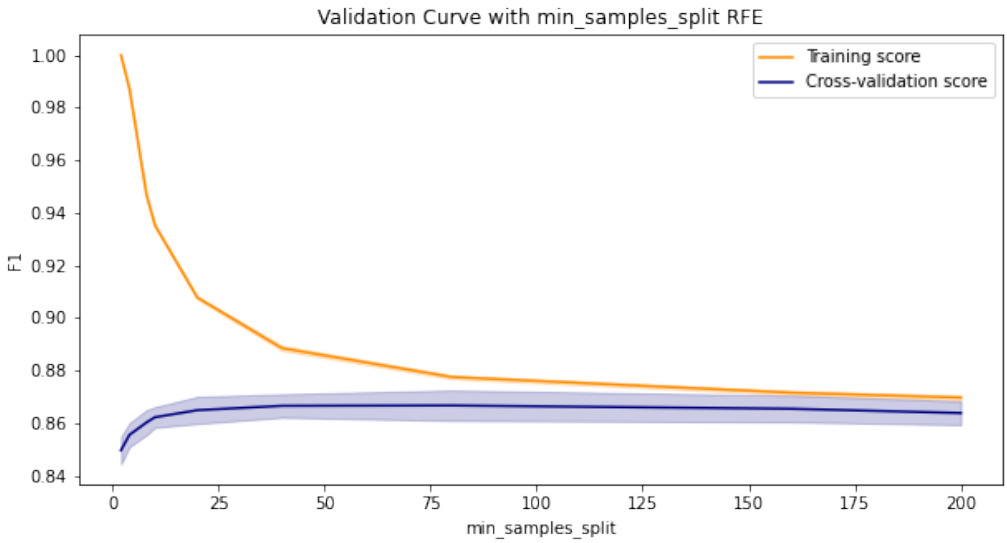


(a) F1 score for the max depth RFE over 0 minutes

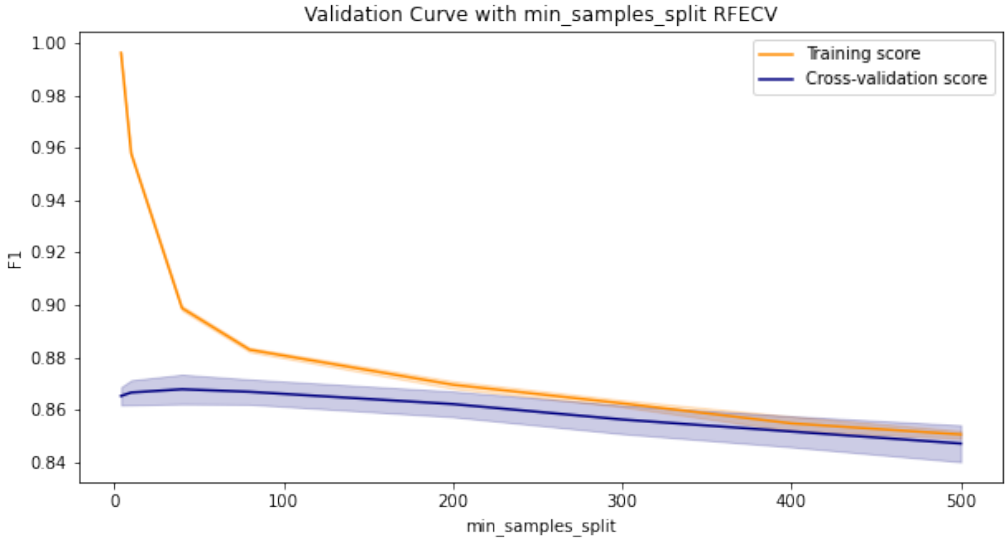


(b) F1 score for the max depth RFECV over 0 minutes

Figure E.2: F1 scores for the max depth of feature selections

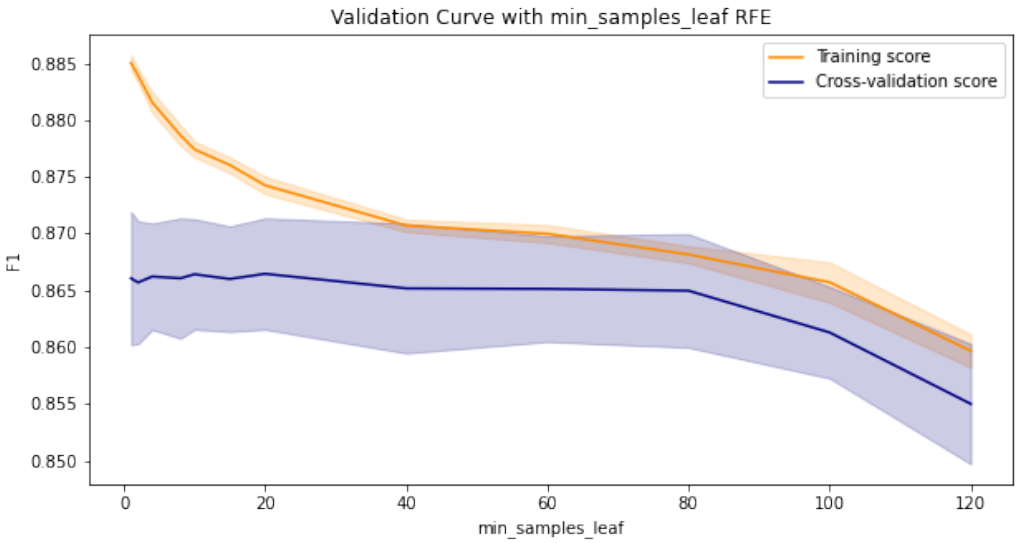


(a) F1 score for the min samples split RFE over 0 minutes

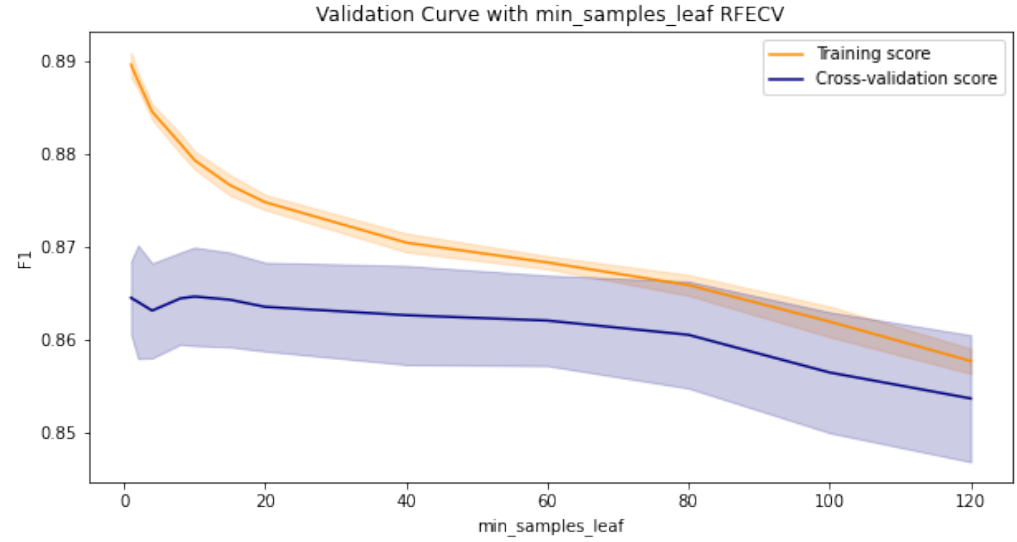


(b) F1 score for the min samples split RFECV over 0 minutes

Figure E.3: F1 scores for the min samples split of feature selections

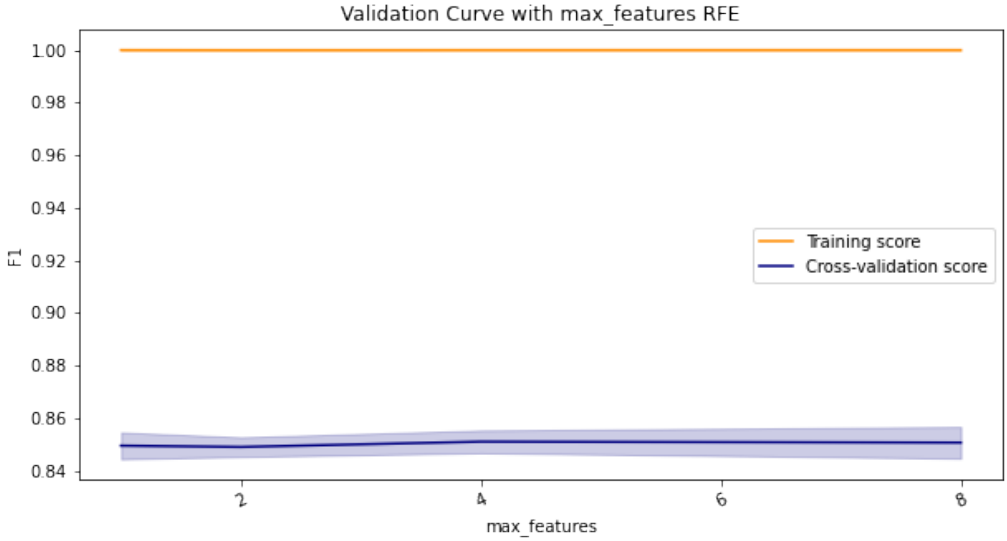


(a) F1 score for the min samples leaf RFE over 0 minutes

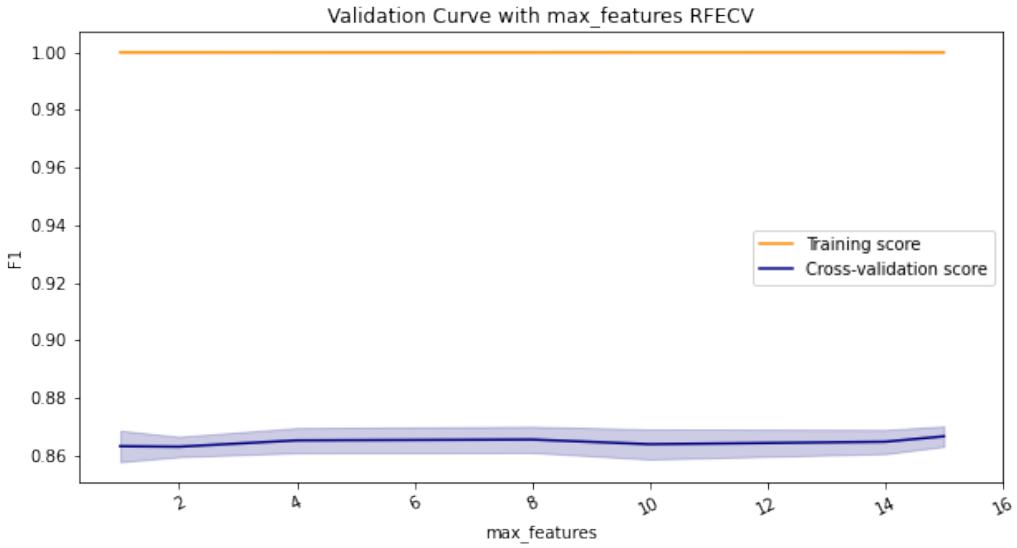


(b) F1 score for the min samples leaf RFECV over 0 minutes

Figure E.4: F1 scores for the min samples leaf of feature selections

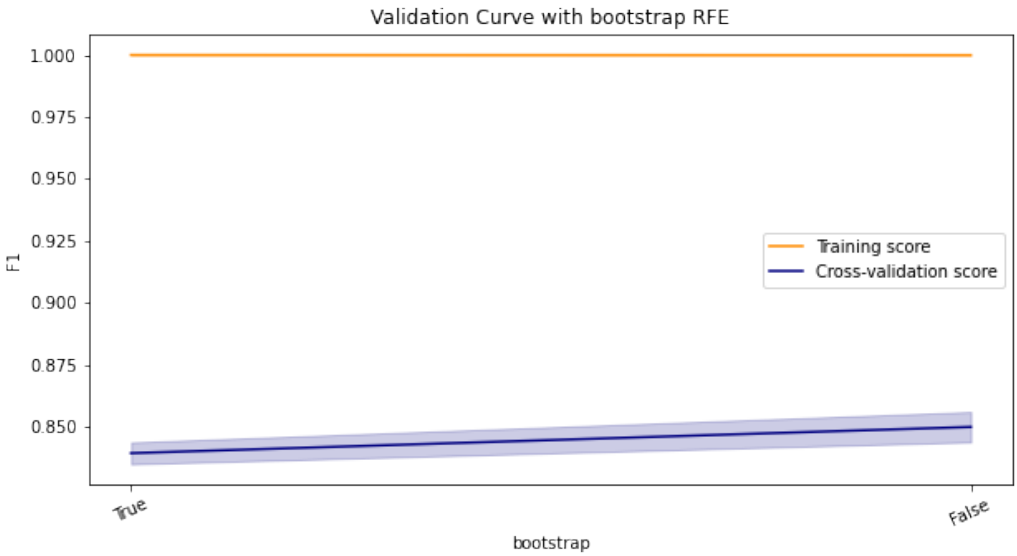


(a) F1 score for the max features RFE over 0 minutes

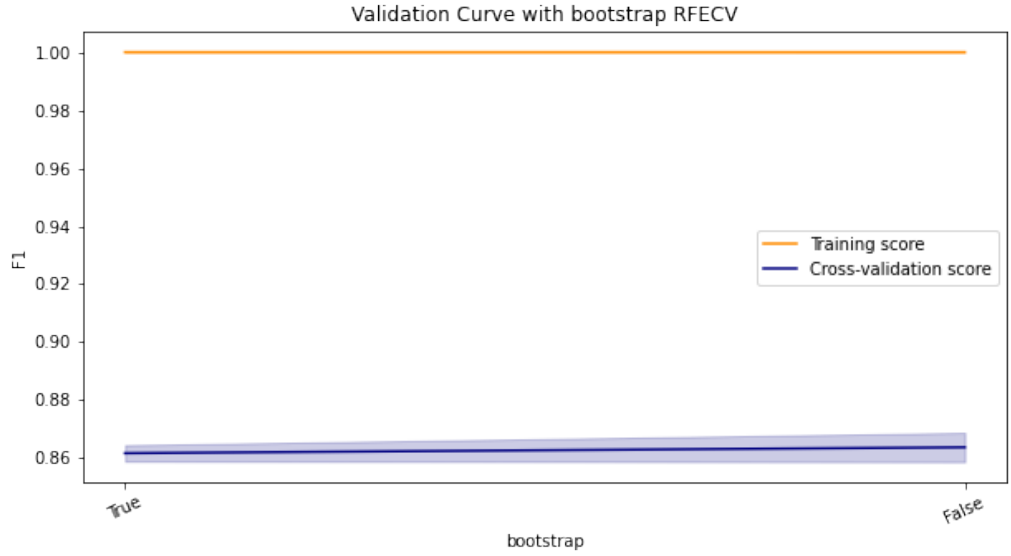


(b) F1 score for the max features RFECV over 0 minutes

Figure E.5: F1 scores for the max features of feature selections



(a) F1 score for the bootstrap RFE over 0 minutes



(b) F1 score for the bootstrap RFECV over 0 minutes

Figure E.6: F1 scores for the bootstrap of feature selections

F Resampling RFE and RFECV

F.1. RFE

Table F.1: Scores SMOTE resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.89	0.77	0.82
Outage	0.52	0.71	0.60

Table F.2: Scores Tomek resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.84	0.90	0.87
Outage	0.64	0.52	0.58

Table F.3: Scores SMOTETomek resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.88	0.77	0.82
Outage	0.52	0.71	0.60

Table F.4: Cohen's kappa scores SMOTE, Tomek, SMOTETomek RFE

Class	SMOTE	Tomek	SMOTETomek
Overall	0.76	0.79	0.76
Never Access	1.00	1.00	1.00
Normal Access	0.71	0.87	0.71
Outage	0.70	0.51	0.70

F.2. RFECV

Table F.5: Scores SMOTE resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.88	0.76	0.82
Outage	0.51	0.71	0.59

Table F.6: Scores Tomek resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.84	0.91	0.87
Outage	0.66	0.50	0.57

Table F.7: Scores SMOTETomek resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.88	0.76	0.82
Outage	0.51	0.71	0.59

Table F.8: Cohen's kappa scores SMOTE, Tomek, SMOTETomek RFECV

Class	SMOTE	Tomek	SMOTETomek
Overall	0.75	0.80	0.75
Never Access	1.00	1.00	1.00
Normal Access	0.70	0.88	0.70
Outage	0.70	0.48	0.70

G

Resampling Spatial and Temporal

G.1. All Features Spatial

Table G.1: Scores Spatial SMOTE resampling

Class	Precision	Recall	F1
Never Access	0.89	0.91	0.90
Normal Access	0.83	0.69	0.76
Outage	0.40	0.57	0.47

Table G.2: Scores Spatial Tomek resampling

Class	Precision	Recall	F1
Never Access	0.84	0.91	0.87
Normal Access	0.78	0.87	0.82
Outage	0.40	0.18	0.25

Table G.3: Scores spatial SMOTETomek resampling

Class	Precision	Recall	F1
Never Access	0.88	0.91	0.89
Normal Access	0.84	0.68	0.75
Outage	0.39	0.57	0.46

Table G.4: Cohen's kappa scores spatial SMOTE, Tomek, SMOTETomek

Class	SMOTE	Tomek	SMOTETomek
Overall	0.62	0.62	0.61
Never Access	0.90	0.89	0.89
Normal Access	0.62	0.83	0.61
Outage	0.56	0.18	0.55

G.2. RFE Spatial

Table G.5: Scores spatial no resampling RFE

Class	Precision	Recall	F1
Never Access	0.82	0.85	0.83
Normal Access	0.74	0.88	0.80
Outage	0.41	0.15	0.22

Table G.6: Scores spatial SMOTE resampling RFE

Class	Precision	Recall	F1
Never Access	0.87	0.92	0.90
Normal Access	0.84	0.69	0.76
Outage	0.39	0.55	0.47

Table G.7: Scores spatial Tomek resampling RFE

Class	Precision	Recall	F1
Never Access	0.81	0.86	0.83
Normal Access	0.76	0.83	0.80
Outage	0.43	0.23	0.30

Table G.8: Scores spatial SMOTETomek resampling RFE

Class	Precision	Recall	F1
Never Access	0.87	0.92	0.89
Normal Access	0.83	0.69	0.76
Outage	0.40	0.54	0.46

Table G.9: Cohen's kappa scores spatial no resampling, SMOTE, Tomek, SMOTETomek RFE

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.58	0.62	0.58	0.61
Never Access	0.83	0.91	0.84	0.91
Normal Access	0.83	0.63	0.78	0.63
Outage	0.14	0.55	0.22	0.55

G.3. RFECV Spatial

Table G.10: Scores spatial no resampling RFECV

Class	Precision	Recall	F1
Never Access	0.83	0.88	0.85
Normal Access	0.75	0.88	0.81
Outage	0.48	0.16	0.24

Table G.11: Scores spatial SMOTE resampling RFECV

Class	Precision	Recall	F1
Never Access	0.88	0.90	0.89
Normal Access	0.83	0.68	0.75
Outage	0.39	0.58	0.47

Table G.12: Scores spatial Tomek resampling RFECV

Class	Precision	Recall	F1
Never Access	0.83	0.89	0.86
Normal Access	0.77	0.86	0.81
Outage	0.46	0.23	0.30

Table G.13: Scores spatial SMOTETomek resampling RFECV

Class	Precision	Recall	F1
Never Access	0.88	0.89	0.88
Normal Access	0.84	0.69	0.76
Outage	0.38	0.57	0.45

Table G.14: Cohen's kappa scores spatial no resampling, SMOTE, Tomek, SMOTETomek RFECV

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.60	0.61	0.61	0.60
Never Access	0.86	0.88	0.87	0.87
Normal Access	0.84	0.61	0.80	0.62
Outage	0.15	0.57	0.22	0.55

G.4. All Features Temporal

Table G.15: Scores temporal SMOTE resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.76	0.81
Outage	0.49	0.68	0.57

Table G.16: Scores temporal Tomek resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.89	0.86
Outage	0.63	0.50	0.56

Table G.17: Scores temporal SMOTETomek resampling

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.76	0.81
Outage	0.49	0.68	0.57

Table G.18: Cohen's kappa scores temporal SMOTE, Tomek, SMOTETomek

Class	SMOTE	Tomek	SMOTETomek
Overall	0.74	0.78	0.74
Never Access	1.00	1.00	1.00
Normal Access	0.70	0.86	0.70
Outage	0.66	0.48	0.66

G.5. RFE Temporal

Table G.19: Scores temporal no resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.81	0.95	0.87
Outage	0.70	0.35	0.47

Table G.20: Scores temporal SMOTE resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.75	0.81
Outage	0.48	0.68	0.57

Table G.21: Scores temporal Tomek resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.83	0.88	0.85
Outage	0.62	0.51	0.56

Table G.22: Scores temporal SMOTETomek resampling RFE

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.76	0.81
Outage	0.49	0.68	0.57

Table G.23: Cohen's kappa scores temporal no resampling, SMOTE, Tomek, SMOTETomek RFE

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.78	0.74	0.78	0.75
Never Access	1.00	1.00	1.00	1.00
Normal Access	0.93	0.70	0.85	0.70
Outage	0.34	0.67	0.50	0.67

G.6. RFECV Temporal

Table G.24: Scores temporal no resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.81	0.95	0.87
Outage	0.71	0.33	0.45

Table G.25: Scores temporal SMOTE resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.76	0.81
Outage	0.49	0.68	0.57

Table G.26: Scores temporal Tomek resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.82	0.90	0.86
Outage	0.64	0.48	0.55

Table G.27: Scores temporal SMOTETomek resampling RFECV

Class	Precision	Recall	F1
Never Access	1.00	1.00	1.00
Normal Access	0.87	0.76	0.81
Outage	0.49	0.68	0.57

Table G.28: Cohen's kappa scores temporal no resampling, SMOTE, Tomek, SMOTETomek RFECV

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.78	0.74	0.78	0.75
Never Access	1.00	1.00	1.00	1.00
Normal Access	0.94	0.70	0.87	0.72
Outage	0.32	0.67	0.47	0.64

H Results Original Data

H.1. Original Data Normal 10-Fold

Table H.1: Scores no resampling original data

Class	Precision	Recall	F1
Never Access	0.96	0.99	0.97
Normal Access	0.81	0.92	0.86
Outage	0.70	0.37	0.49

Table H.2: Scores SMOTE resampling original data

Class	Precision	Recall	F1
Never Access	0.96	0.98	0.97
Normal Access	0.86	0.77	0.81
Outage	0.50	0.63	0.55

Table H.3: Scores Tomek resampling original data

Class	Precision	Recall	F1
Never Access	0.95	0.99	0.97
Normal Access	0.83	0.89	0.86
Outage	0.64	0.45	0.53

Table H.4: Scores SMOTETomek resampling original data

Class	Precision	Recall	F1
Never Access	0.96	0.98	0.97
Normal Access	0.86	0.76	0.81
Outage	0.50	0.64	0.56

Table H.5: Cohen's kappa scores no resampling, SMOTE, Tomek, SMOTETomek original data

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.76	0.72	0.76	0.72
Never Access	0.98	0.98	0.98	0.98
Normal Access	0.90	0.71	0.86	0.71
Outage	0.37	0.62	0.44	0.62

H.2. Original Data Spatial

Table H.6: Scores spatial no resampling original data

Class	Precision	Recall	F1
Never Access	0.85	0.90	0.87
Normal Access	0.74	0.87	0.80
Outage	0.46	0.15	0.23

Table H.7: Scores spatial SMOTE resampling original data

Class	Precision	Recall	F1
Never Access	0.86	0.90	0.88
Normal Access	0.83	0.65	0.73
Outage	0.37	0.56	0.45

Table H.8: Scores spatial Tomek resampling original data

Class	Precision	Recall	F1
Never Access	0.84	0.90	0.87
Normal Access	0.76	0.82	0.79
Outage	0.44	0.25	0.32

Table H.9: Scores spatial SMOTETomek resampling original data

Class	Precision	Recall	F1
Never Access	0.86	0.90	0.88
Normal Access	0.83	0.65	0.73
Outage	0.37	0.57	0.45

Table H.10: Cohen's kappa scores spatial no resampling, SMOTE, Tomek, SMOTETomek original data

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.60	0.58	0.60	0.58
Never Access	0.88	0.88	0.89	0.88
Normal Access	0.83	0.58	0.77	0.58
Outage	0.14	0.55	0.24	0.56

H.3. Original Data Temporal

Table H.11: Scores temporal no resampling original data

Class	Precision	Recall	F1
Never Access	0.89	0.96	0.92
Normal Access	0.78	0.90	0.84
Outage	0.65	0.23	0.34

Table H.12: Scores temporal SMOTE resampling original data

Class	Precision	Recall	F1
Never Access	0.90	0.94	0.92
Normal Access	0.84	0.70	0.76
Outage	0.44	0.61	0.51

Table H.13: Scores temporal Tomek resampling original data

Class	Precision	Recall	F1
Never Access	0.88	0.96	0.92
Normal Access	0.79	0.86	0.82
Outage	0.59	0.32	0.42

Table H.14: Scores temporal SMOTETomek resampling original data

Class	Precision	Recall	F1
Never Access	0.90	0.93	0.92
Normal Access	0.84	0.70	0.76
Outage	0.43	0.60	0.50

Table H.15: Cohen's kappa scores temporal no resampling, SMOTE, Tomek, SMOTETomek original data

Class	No	SMOTE	Tomek	SMOTETomek
Overall	0.68	0.65	0.68	0.64
Never Access	0.95	0.93	0.95	0.92
Normal Access	0.87	0.63	0.82	0.64
Outage	0.22	0.60	0.31	0.59