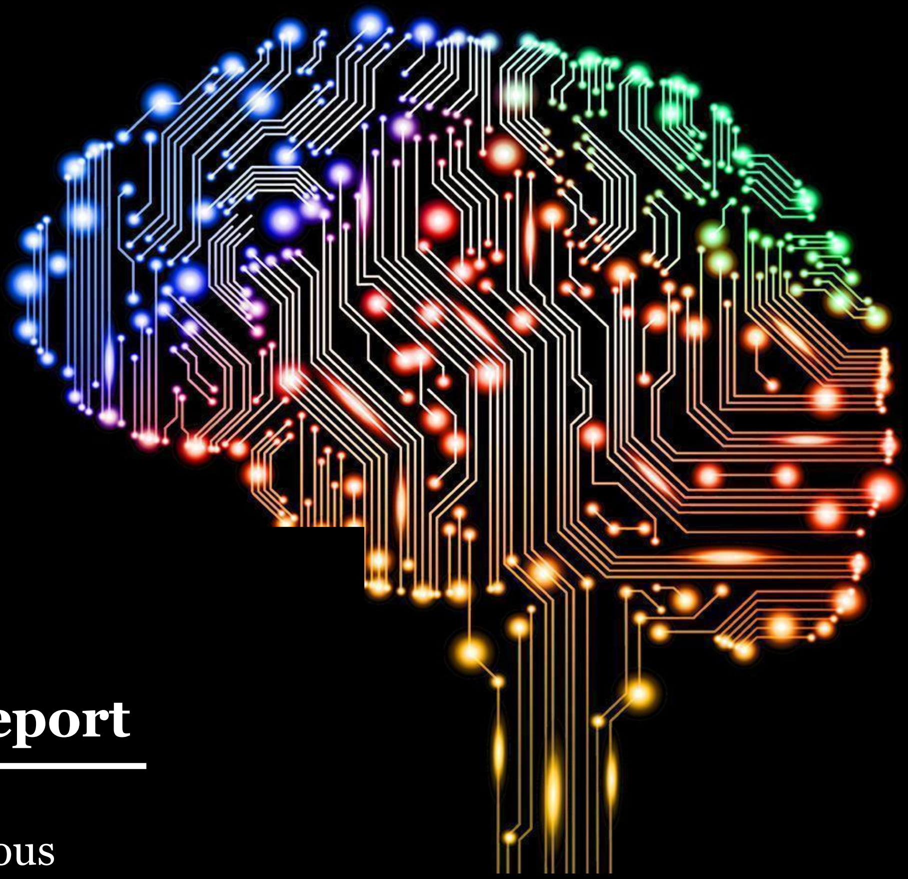

EE-401
Mid-Year
Evaluation Report

Applications of Deep Learning for Anomalous Energy Consumption Detection

Submitted By:

EE-163 – Saad Mashkoor Siddiqui
EE-164 – Faiq Siddiqui
EE-194 – Syed Abdul Haseeb Qadri
EE-177 – Muhammad Waleed Hasan

Section – **D**
Class – **BE**
Date – **2nd March, 2020**
Batch – **2016-17**

Internal: **Dr. Muhammad Ali Baig**
External: **Mr. Shahzeb Anwar**

ABSTRACT

Electricity theft in electrical grids is a pressing issue for distribution companies (DISCOs), resulting in estimated losses of PKR 45 billion in Pakistan during 2019 [8]. Prevalent methods of electricity theft detection in Pakistan rely on manual energy meter inspections which are expensive, ineffective, and fail to scale with increasing number of consumers.

This project aims to address these issues by assessing the feasibility of artificial intelligence (AI) techniques for anomalous energy consumption detection. Specifically, anomalous energy consumption detection is framed as a supervised binary classification machine learning problem to which both shallow and deep learning approaches are applied.

This report establishes anomalous energy consumption detection as a complex technical problem of substantial socioeconomic importance along with the relevance and scope of the project detailed herein. After summarizing existing research on AI-based solutions for energy theft detection, the report outlines the methodology followed for the design and development of shallow and deep learning models for electricity theft detection. Data preprocessing of the SGCC dataset is discussed, followed by a description of the training and tuning of learning algorithms applied to the binary classification task in question: logistic regression, support vector machine, random forest, and wide neural network. Test set performance of these learning algorithms is then compared and contrasted with each other and with benchmarks identified through literature review. The report concludes with an overview of outcomes expected to be achieved by the end of the academic year, as well as potential areas for future work beyond the current scope of the project.

Table of Contents

Table Of Figures

List of Tables

CHAPTER 1
INTRODUCTION

1. INTRODUCTION

1.1. Brief Introduction

Electricity has become a necessity of life and losses in electricity occur due to many reasons. These losses reduce the revenue of the power companies, previous studies shows that the losses due to electricity theft costs millions of dollars each year [2]. The abnormal pattern in energy consumption can be a good indicator of electricity theft. Nowadays, the abnormality in the energy consumption pattern have brought good attention to the data driven electricity theft detection approaches [3]. In this report we will discuss the abnormality found in the energy consumption of consumer’s kWh data provided by state grid corporation of China (SGCC) [3] and classify thieves and normal consumer using machine learning tools and libraries.

1.2. Problem Statement

Theft of electricity is an alarming problem for any state. This crime can be committed either by tampering meter or by using electricity without being recorded by energy meters. Our aim in this regard is to use machine learning techniques to identify the anomalies in power consumption pattern of consumers and also to verify the results using different toolkits of machine learning (Keras, Scikit Learn) for this purpose. Moreover, to classify the users (from thief and regular consumers) we have built different machine learning architectures Random Forrest, SVM (Support vector machine), logistic regression model and Wide Neural Network.

1.3. Objectives

1.3.1. To explore the feasibility of artificial intelligence-based methods for electricity theft detection.

One of the objective of this project is to explore if electricity thieves can be identified using Artificial Intelligence based methods, hence to carry out this investigation we will study previous theft detection techniques.

1.3.2. To identify the optimal machine and deep learning algorithm for AI-based electricity theft detection.

The second objective of this project is to devise most efficient AI-based method for electricity theft detection among the existing methods and assess their selectivity (Accuracy, speed of computation, cost). On the basis of these measures we will decide the best AI-based method for electricity theft detection.

1.4. Scope of work

1.4.1. Peer Review of an IEEE Journal Article

The project will verify the results of machine learning-based approaches to electricity theft detection presented in Zheng et. al’s research paper. Having been cited by 25 other IEEE papers [3], Zheng et. al’s research is a canonical work in the field of AI-based electricity theft detection. By systematically assessing and analyzing the assumptions, techniques, and results presented in the paper, the project will be an exercise in peer review which is a cornerstone of the academic research process.

1.4.2. Proof of concept of AI-based Electricity Theft Detection

The project will also serve as a proof-of-concept of deep learning-based solutions for residential electricity theft detection in smart grids. The project will establish that given a sufficiently large, labeled dataset of kWh values recorded over a time interval T, it is possible to train both shallow and deep learning models to identify potential electricity theft occurring within T with reasonable accuracy and reliability.

1.5. Significance of research

This project holds significance for two primary stakeholders in an electrical distribution system, namely consumers and distribution companies (DISCOs).

1.5.1. Consumers

Presently, DISCOs have a two-pronged approach to deterring electricity theft among residential consumers. On a micro level, consumers who are identified to have engaged in illegal abstraction of electrical energy have their connections terminated and are required to pay fines to the DISCO [17]. On a macro level, communities and locale s with substantial occurrences of electricity theft are more likely to face load shedding during periods of increased demand or shortfall [5] [6]. In the latter case, consumers who have paid their dues are also penalized along with thieves, thereby eliminating the incentive not to engage in electricity theft. On a second level, Jamil and Ahmad [7] have found that increasing electricity prices have a strong positive correlation with electricity theft, suggesting increase in electricity bills to compensate for non-technical losses in areas with widespread electricity theft tends to exacerbate, rather than mitigate, the issue. Blanket penalties for electricity theft detection are therefore both unnecessary and ineffective in preventing electricity theft.

The neural network presented in this paper is a first step towards incisive, accurate, and targeted electricity theft deterrence. It will empower DISCOs to remotely pinpoint specific consumers engaging in electricity theft without the need for manual, site inspection-based detection. This means DISCOs will be less likely to penalize the entire locale or neighborhoods' for the actions of the few consumers that are practicing electricity theft, thereby eliminating the idea of collective penalization of regular consumers.

1.5.2. DISCOs

A neural network capable of identifying potential electricity thieves using a window of kWh readings is of great operational and monetary value to DISCOs.

The neural network developed in this project is intended to be an autonomous, intelligent, and scalable solution for electricity theft detection that can form the core of a smart grid's theft deterrence operations. Prevalent methods of electricity theft detection rely on manual energy meter inspections which are often inaccurate, time consuming, and scale poorly with growing number of electricity consumers often found in urban centers. In contrast, a neural network-based theft detection system will lead to more accurate, reliable, and actionable identification of instances of energy theft, and will minimize costs associated with manual energy meter inspections, improving operational efficiency.

From a monetary perspective, the primary benefit of a neural network-based electricity theft detection procedure will be the mitigation of the non-technical losses associated with electricity theft, which were recently estimated to be as much as PKR 45 billion in 2019 [8]. On a second level, the neural network is capable of adding value to energy consumption data that is already being logged and collected by DISCOs. The neural network(s) developed as part of this project can either be used as a transfer learning solution for tuning parameters of a similar, proprietary network developed by the DISCO, or can be trained on the DISCO's data from scratch.

In either case, the network effectively monetizes electrical consumption data by deriving actionable insights about potential electricity thieves which can then be used to both minimize instances of theft as well as decrease costs associated with identifying such occurrences through manual inspections. Furthermore, the network may be modified to be an online learning system, enabling it to continually learn from new instances of electricity thieves' consumption data identified during operation, and thus continually improve its performance, and improving return on investment after deployment.

CHAPTER 2
LITERATURE REVIEW

2. LITERATURE REVIEW

Electrical energy generation and distribution is a lucrative business that plays an integral part in building a country’s economy. Electricity is now considered an essential part of our lives. With the demand for electrical energy increasing at a rate of 5.3% each year since World War II [4], it has become vital that the energy delivery network must be kept in proper working condition and any problems occurring in it must be addressed immediately.

In the energy sector the losses occurring in the generation, transmission and distribution is the foremost problem that degrades the efficiency of the system and causes utility companies 100 million Canadian dollars every year [2]. These losses fall into two main categories, namely Technical Losses (TLs) and Non-Technical Losses (NTLs). Technical losses are comparatively easier to resolve as the core cause and the location of the loss is known. This way they can be predicted and addressed by means of using better equipment. On the other hand NTLs are not only harder to solve but are nearly impossible to detect using manual inspections and other conventional methods.

Electricity theft is considered as one of the chief NTLs that plagues the distribution network in every energy delivery system. Especially in developing countries, the poor economic conditions encourages people to steal electricity causing huge revenue losses for the utility company and at the same time causing public safety hazard, like in case a fire or an electric shock occurs. Energy theft may occur primarily due to a number of reason which includes, physical tampering of the energy meter, sidestepping the energy meter (results in lower KWhs registered), false data dispatch from the host, malfunctions and misconfigurations in the communication scheme and improper synchronization of the meter after a scheduled maintenance job [2].

Now with the advent of smart girds and Advance Metering Infrastructure (AMI) the distribution sector is evolving. It is now possible to measure even very minute quantities of energy consumed by a customer and report them remotely using a smart meter. This data could serve as the basis for reducing NTLs, especially theft as the anomalous energy consumption trend can be reflected in the data transmitted by the smart meter. [2] highlights a technique that detects electricity theft based on data received by the energy meter. The 3 main features on which the mentioned system classifies a thief from a normal consumer are a) host based intrusion detection, that detects manipulating the internal software and hardware of the meter, b) on-meter anti tampering devices, and c) Non-intrusive load monitoring. The proposed model named AMIDs, integrated intrusion detection solution, is able to detect 62% of the cases as mentioned by the author.

In [10] devises a classification technique that uses Artificial Intelligence based methods to distinguish between thieves from non-thieves. An Artificial Neural network was proposed that learns on the basis of 7 different datasets and is able to recognize useful patterns leading to the final classification. The datasets include the socio-economic background of consumer, history of scrutiny, ownership exchange, debits and meter reading. These data bases are cleaned and fed into the neural network which then uses modern deep learning algorithms to map the input features to the desired output. The accuracy of the neural net achieved as quoted by the author is 87.17%, while the precision of classifying true energy thieves from normal consumer bumped up from 40% to 65.03%.

[11] proposes an innovative and fast hybrid algorithm named that comprises of Harmony Search in conjunction with a Optimum-path forest classifier to analyze patterns in consumption data to identify NTLs. Just like approaches discussed before the author of this article have showed it superiority over other methods for energy theft detection. The feature selection was performed on a dataset provided Brazilian electric power company with both commercial and industrial consumers. Accuracies of 92.60% were achieved using this method for the commercial consumers, while accuracy of 96.5% was published for industrial consumers. An advantage of this approach was that it was 9 times faster to train, in comparison to Particle Swarm optimization (PSO) algorithm, for a given dataset and identify appropriate features.

In [3] Zheng et al. propose a novel neural network architecture comprising wide and deep convolutional neural networks (WCNN and DCNN respectively). These CNNs are trained in tandem as a single neural network to identify electricity theft using a labeled, publicly available dataset from the State Grid Corporation of China (SGCC), which provides daily kWh consumption data of 42,372 residential consumers over 1,035 days.

2.1. Over View of “Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection”

The authors first establish electricity theft detection as a practically and economically pertinent problem, citing BC Hydro Tech Inc. in stating that electricity theft losses cost ~CAD 100 million every year, and lead to severe electrical faults and hazards.

The authors then proceed to review existing work on electricity theft detection, discussing both hardware and software-based anti-theft methods, the latter preferred by the authors for making use of data made available by the proliferation of Advanced

Metering Infrastructure (AMI). Specifically, the authors have discussed the application of game theory, anomaly detection, and machine learning models such as support vector machines (SVMs) and neural networks (NNs) in the domain of electricity theft detection.

In assessing these approaches, the authors found hardware-based methods to be expensive to operate, secure, and maintain. Machine learning approaches were analyzed to be more effective than other software alternatives. However, the authors note that these models either require artificial feature extraction or fail to take advantage of the inherent periodicity in the weekly kWh consumption patterns of regular consumers - a feature which the authors then proceed to establish to be absent in kWh consumption patterns of electricity thieves. Zheng et al.'s model aims to overcome this shortcoming by combining a WCNN and DCNN in a single neural network to simultaneously identify trends across all consumers on a single day and exploiting weekly periodicity in a single consumer's kWh consumption pattern. In this regard, Zheng et al. have provided a high-level overview of the WDCNN architecture along with analytical expressions for the custom convolutional kernels used in the DCNN component.

2.1.1. Data Preprocessing

The paper discusses how the SGCC data was preprocessed for use with machine learning models. Specifically how, Zheng et al. have replaced missing kWh values and scaled abnormal values present in the consumption distribution by applying a variant of the three-sigma rule and then finally mapping all kWh values to a [0, 1] range using min-max feature scaling.

2.1.2. Methodology and Results

Preprocessed data is then used to train various shallow learning models such as logistic regression (LR), SVMs, and random forest (RF) classifiers along with standalone WCNN and DCNNs for training ratios of 50%, 60%, 70, and 80%. These models are then evaluated on the basis of the Mean Average Precision (MAP) and the Area Under the receiver-operator characteristics Curve (AUC) over 100 and 200 test set samples.

Hyperparameters for the final models are provided, although the methodology for deriving them has not been discussed. The WDCNN model is trained and evaluated in the same way, and is shown to outperform all other approaches by all metrics across all training ratios, reaching a peak MAP of 96.86%, as summarized in table 1 [3]. Hyperparameter tuning for the WDCNN model is then discussed at length, with trends in both MAP and AUC used to identify optimal values for WDCNN's layer counts, layer sizes, and epochs for all training ratios.

Table 1: Performance Comparison with Conventional Schemes

Training ratio = 80%		
Methods	**AUC**	**Optimal Hyperparameters**
LR	0.7060	C = inverse Regularization Strength =1.0, L2 Regularization
SVM	0.7413	Kernel = RBF, Penalty for Errors: 1.0
RF	0.7385	Trees= 200, Splitting Quality function = Gini impurity
Wide	0.6965	Neurons = 90, Epochs = 20, Relu Activation
CNN	0.7797	Various Combinations

2.1.3. Future Work

Zheng et al.’s results show that while a WDCNN does indeed outperform all other approaches, exploration of the DCNN is a cheaper, less computationally intensive solution which may be worthwhile, as it offers comparable performance as the WDCNN. The paper’s methodology for identification of outliers is also worth re-evaluating, as the three sigma rule is more likely to identify outliers correctly when applied on a consumer, rather than daily, basis. Furthermore, despite identifying the kWh values as sequence data, Zheng et al. have not compared the performance of sequence models such as Recurrent Neural Networks (RNNs) with that of the WDCNN, which warrants further investigation.

CHAPTER 3
METHODOLOGY/OBSERVATION & CALCULATION

3. METHODOLOGY/OBSERVATION AND CALCULATION

3.1. Data Preprocessing

Data preprocessing is an important step before training and testing data on the models. This technique is used to convert raw data into a more efficient and useful format. The data preprocessing steps for the dataset of SGCC are as follows.

1. Loading raw Data
2. Checking for Negative Values
3. Sorting the data
4. Replacing missing values
5. Identifying and Replacing outliers
6. Feature scaling
7. Storing Finalized data

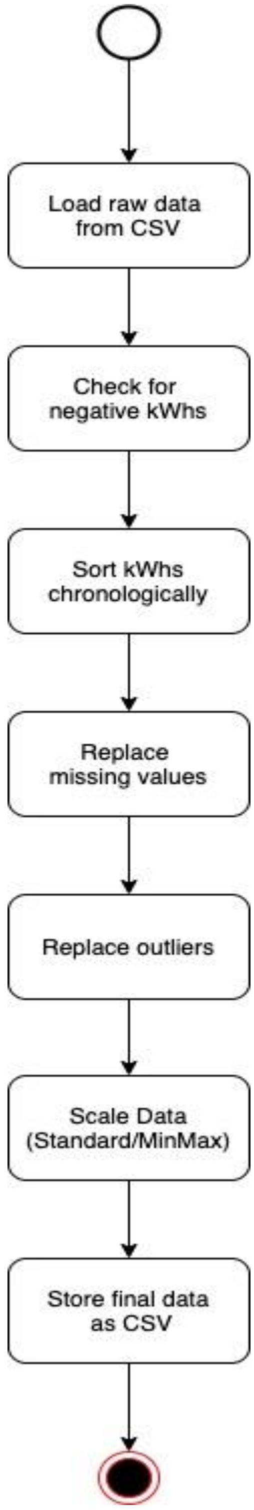

Figure 1:Data preprocessing Flow Chart

3.1.1. Loading Raw data from CSV

In the first step the data is read into pandas data frame to determine the number of columns and rows in the dataset, results show that the given dataset contains 42,372 rows and 1,036 columns, the first two columns consist of flag and consumer number remaining 1,034 columns contains the kilo watt hour reading of consumers for 1034 days.

3.1.2. Checking negative values

Since the energy consumption of a consumer can't be negative therefore we check for negative values in kwh programmatically. The dataset consists of large number of kwhs, checking negative values by inspection of head is intractable. After checking the results show that there are no negative kwh values in the entire dataset.

3.1.3. Sorting data

The dates in the header of each column containing daily Kwh readings are not in chronological order. Without sorting the data in chronological order the actual consumption pattern of a consumer with time can't be determined and even the normal consumption pattern can be identified as abnormal, therefore it is necessary to sort data in chronological order. For this purpose following steps are taken for sorting the data in chronological order.

- The columns containing kwh values are separated from the column containing FLAGs and Consumer numbers.
- The columns in the kwh dataframe are converted from string to datetime objects.

- After converting to datetime objects the kwhs dataframe is sorted in chronological order.
- After sorting data in chronological order the sorted kwhs dataframe is joined back with the columns containing consumer numbers and FLAGs.

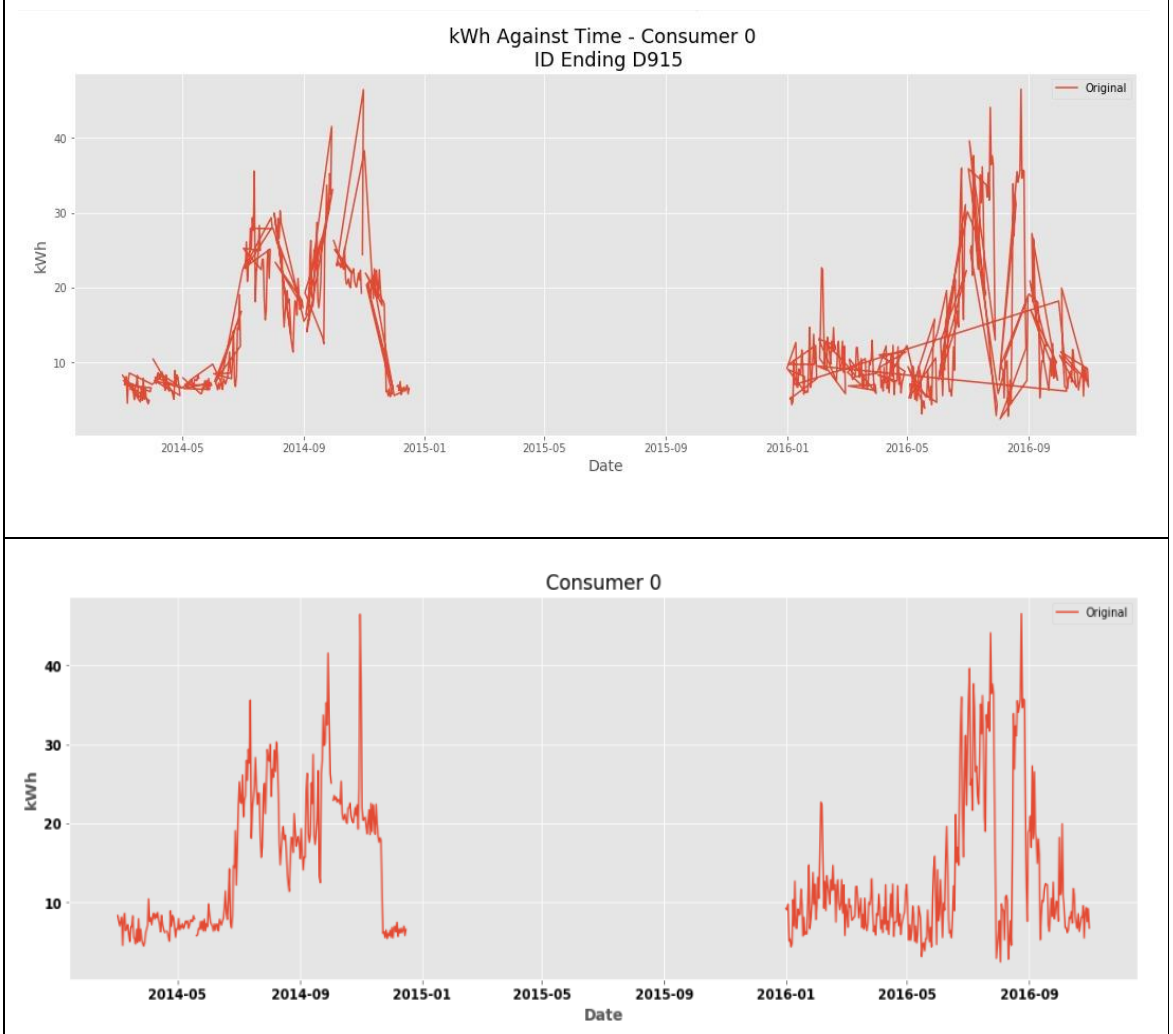

Figure 2:kwh Vs time before sorting (upper) and after sorting (lower)

3.1.4. Visualizing missing values

Since there are more than 1,000 columns in the dataset, using head to visualize missing values in each column is intractable. Instead of creating a new dataframe of non-null values in each column of the dataset, we have plotted it to show the trend in the number of missing values in the dataset. The figure given below shows the number of non-null values per day.

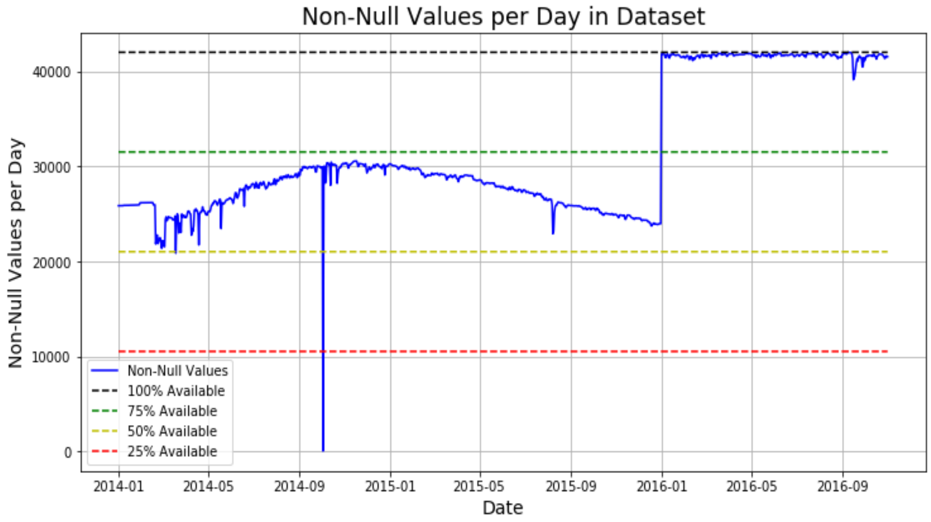

Figure 3: Non-Null/Non-missing values per day

From the figure shown it is clear that there is one day when all the kwh values are missing because number of non-null values dropped to zero. On most of the days, the number of non-null values lies in between 50% and 75% of the total number of consumers in the dataset and there is no missing values in the dataset after towards the end of the dataset i.e during the year 2016.

3.1.5. Replacement of missing values

The missing values are treated as NaN (Not a Number), to replace these values with a numeric value following techniques have been used in [3].

- If the value preceding the missing value and next to missing value are non-missing/non-null then it is replaced by the average of the Kwhs consumed on the next and previous days [3].

  If $x_{i-1}, x_{i+1} \notin$ NaN, then

$$x_i = \frac{x_{i-1} + x_{i+1}}{2} \qquad Equation\ (3.1)$$

- If either of the next or previous days' kwhs are undefined, the current day's kwhs are assumed to be 0. Since the kWh value before first day and after the last day remains unidentified so the missing values in the first and last columns are replaced by 0 [3].

  if $x_{i-1}, x_{i+1} \in$ NaN, then

$$x_i = 0$$

  Where;

  - x_{i-1} is the kwh value on the day before the missing value's day.
  - x_{i+1} is the kwh value on the day after missing value's day.
  - x_i is the day on which missing value exists

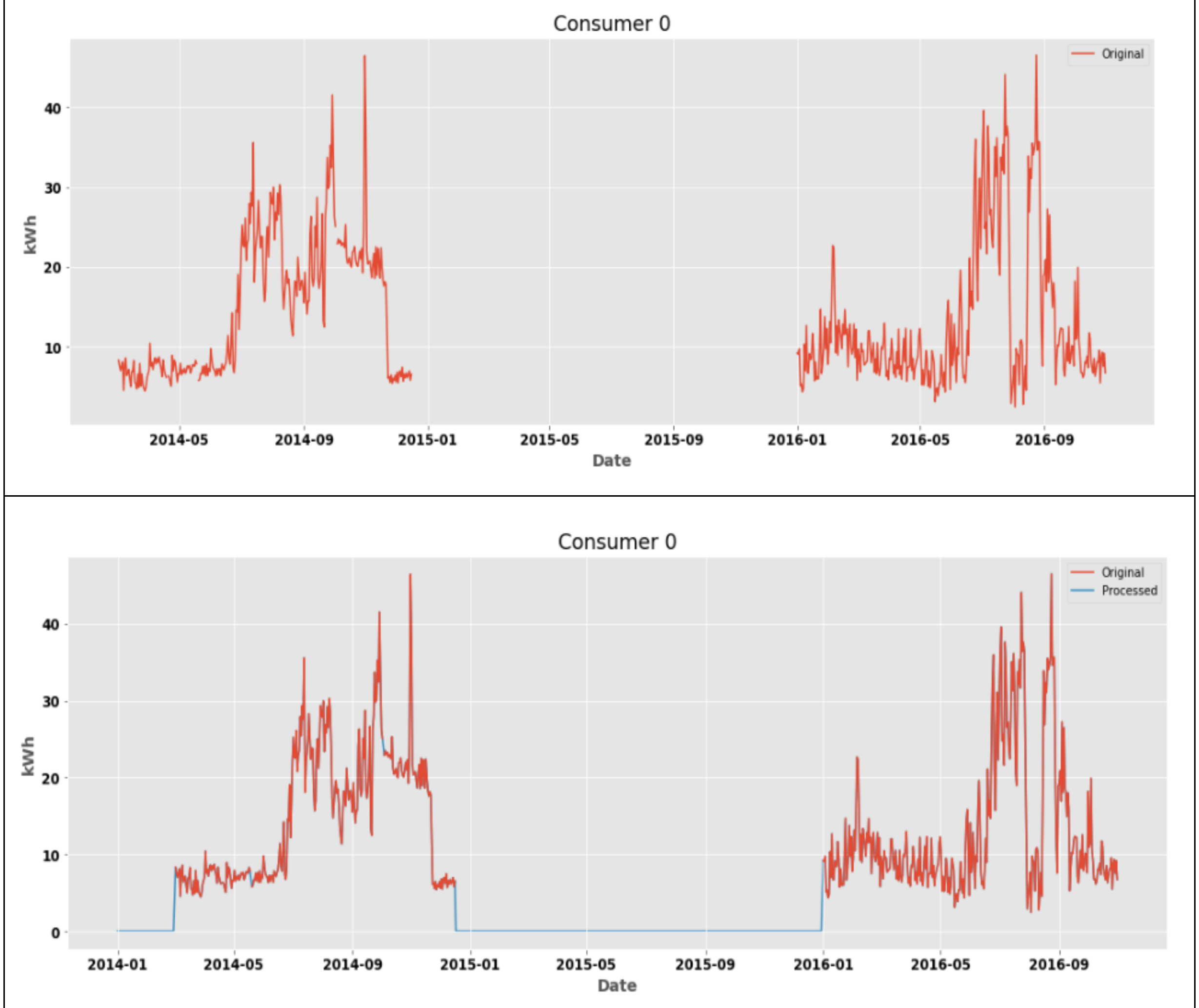

Figure 4:kwh Vs time before replacement (upper) and after replacement (lower) of NaN

3.1.6. Identification of outliers

Outliers are erroneous values that differs from other values in the data significantly. According to three sigma rule of thumb [16], any kWh value which is more than sum of average value and twice the standard deviation is considered as an outlier. Only positive deviations from the mean should be considered while identifying outliers because electricity consumption of each user is always greater than 0 after analyzing the electricity consumption data of 1,034 days.

Concretely, the kWh consumption on the i^{th} day x_i is considered to be an outlier if

$$x_i > avg(X) + 2 \times std(X) \qquad Equation\ (3.2)$$

where X is the vector containing all kWh consumption values of a single consumer.

3.1.7. Replacement of outlier with threshold values

The outliers have been replaced with the threshold values that is sum of average kwh value and twice the standard deviation [3].

$$f(x_i) = \begin{cases} avg(x) + 2std(x) & if\ x_i > avg(x) + 2std(x), \\ x_i & otherwise, \end{cases} \qquad Equation\ (3.3)$$

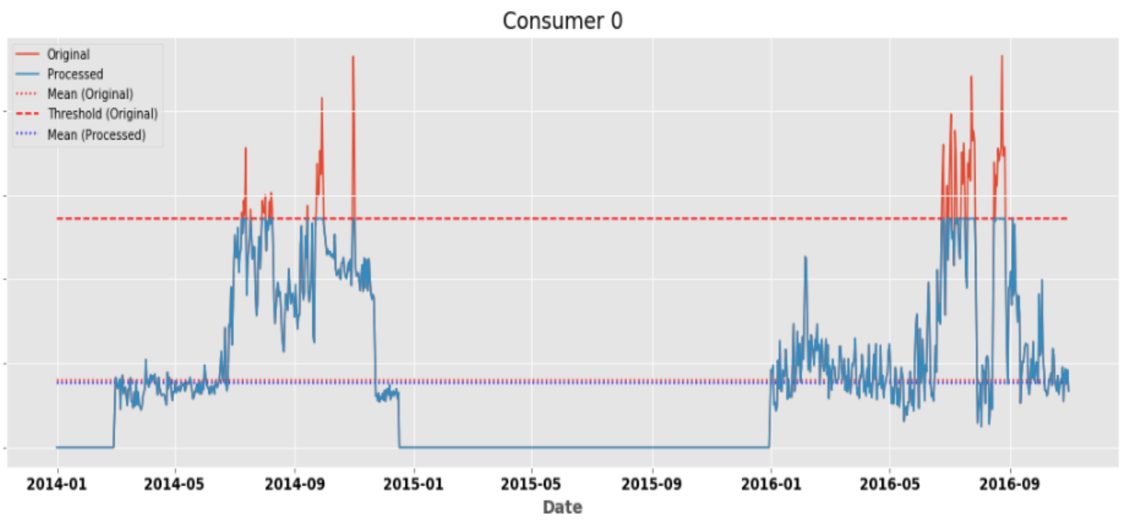

Figure 5:Consumption pattern of consumer number 0 after replacement of outliers

3.1.8. Feature Scaling

a. MinMax Scaling

`MinMaxScaler` is a built-in function which transforms each value in the column proportionally within the range [0,1]. It has been used to normalize the features for an individual consumer, concretely,

$$f(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad Equation\ (3.4)$$

- x_i is the kWh consumption of a single consumer on the i^{th} day.
- x is a vector x_i day-by-day
- $\min(x)$ and $\max(x)$ are the minimum and maximum values of x_i for that consumer

`MinMaxScaler` scales along the column axis i.e. it will find the minimum and maximum values in a single column and use them for scaling. Since we want to scale according to min and max values amongst all kWhs of a single consumer (row axis) and not the min/max values amongst all consumers on a single day (column axis), therefore the data is first transposed before scaling.

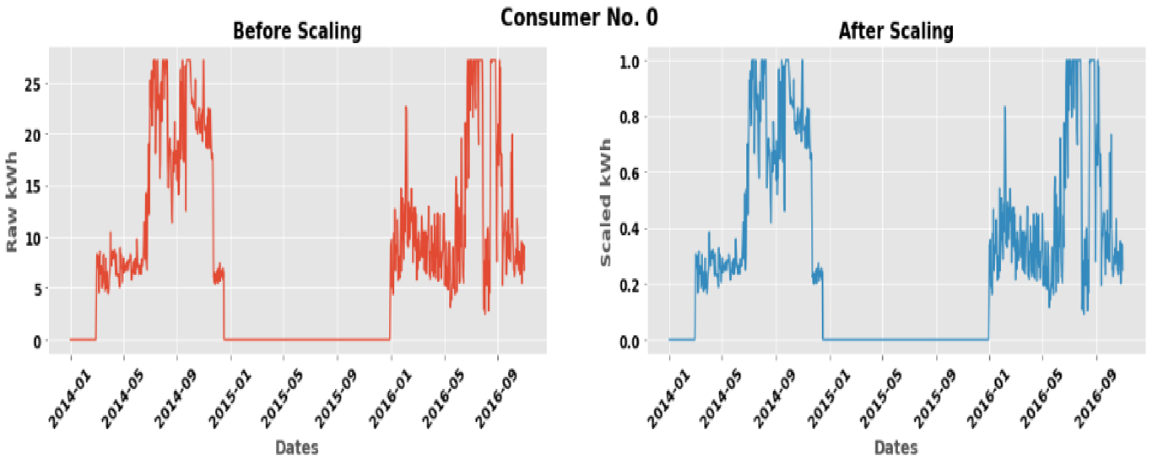

Figure 6: Consumer number 0 before and after minmax scaling

b. Standard Scaler

Standard scaler is another built-in function used for feature scaling. It removes the mean and scales the data to unit variance. The standard score of a sample x_i can be calculated as:

$$f(x_i) = \frac{x_i - mean(x)}{std(x)} \qquad Equation\ (3.5)$$

Where

- x_i is the kWh consumption of a single consumer on the i^{th} day.
- x is a vector x_i day-by-day
- $mean(x)$ is the mean of the training samples.
- $std(x)$ is the standard deviation.

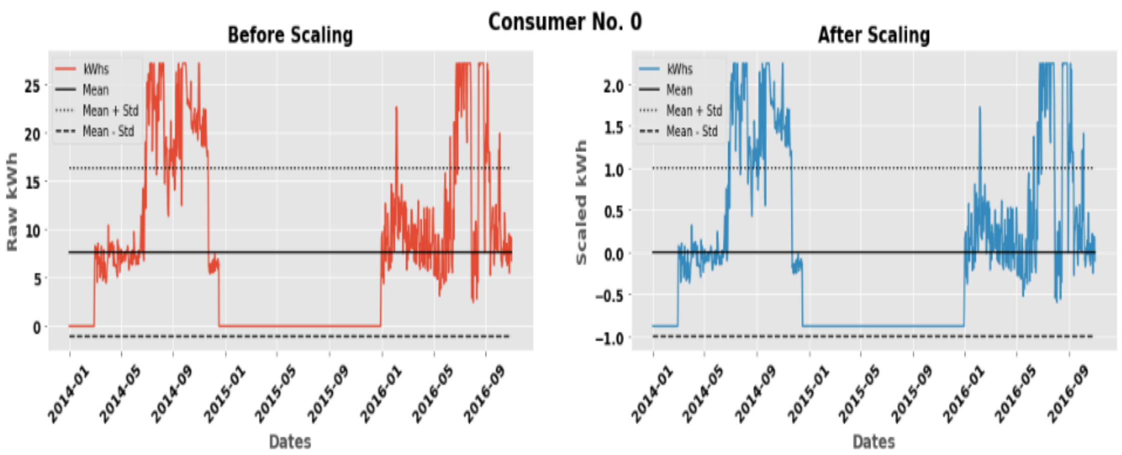

Figure 7:Consumer number zero before and after standard scaling

3.2. Shallow Learning Models

3.2.1. Logistic Regression (LR)

It is the most basic and effective learning algorithm for binary classification problems which has been used in this project for the classification of thieves and normal consumers of electricity by analyzing the preprocessed daily kwh consumption data. Python-based Deep Learning Libraries Scikit learn and Keras are used to develop, train and test this model and assess its selectivity (True positive rate). The steps involved in carrying out Logistic regression techniques are described below.

a. Train test Split

The preprocessed data produced by standard scaler is imported from the drive and split into training and test samples with the training ratio of 80% since LR gives best results at 80% training ratio in previous studies [3]. The proportion of electricity thieves and normal consumers is same in both training and test dataset for stratified sampling in train-test data split. The graph shown below confirms the proportion of thieves and non-thieves in the train, test and overall data.

3.2.2. Support Vector Machines (SVM)

It is another shallow learning model used for discriminative classification and regression problems. To train this model we have used the data produced by standard scaler at the training ratio of 20%, 60% and 80%. The computational speed of this model is the slowest among the tested models. At the training ratio of 80% the model took ~43 minutes to train. Following are the hyperparameters used for this model.

- Penalty parameter of error term is set to 1.0.
- Kernel: `rbf` – Radial Basis Function.

3.2.3. Random Forrest (RF)

The third shallow learning model used for electricity theft detection problem. The data produced by standard scaler has been used to train this model with the training ratio of 80%. This model is expected to outperform SVM and Logistic Regression, specified hyper parameters are,

- Number of trees: 200
- `gini` Function is used to measure quality of split

3.2.4. Wide Neural Network

Wide Neural network is a deep learning model, the model has been used for theft detection with the training ratio of 80%. Python-based Deep Learning Libraries scikit learn and keras are used to develop, train and test this model and assess its selectivity on the data produced by `standardscaler, minmaxscaler` and `maxabs`.

- **Number of Neurons**: In [3] the optimal number of neurons is 50. However, the highest AUC is obtained when the number of neurons is closer to 90.
- **Activation functions**: `Relu` in the first layer and `sigmoid` in the output layer

CHAPTER 4
PRELIMINARY RESULTS

4. PRELIMINARY RESULTS

Table 2: Preliminary Results of Models Trained

	Training Ratio = 80%			
Methods	**Scaling**	**AUC from our experimentation**	**Scaling**	**AUC from Zheng's Models**
Logistic Regression	Standard	0.71	Min-Max	0.706
SVM	Min-Max	0.77	Min-Max	0.7413
Random Forest	Min-Max	0.76	Min-Max	0.7385
Wide Neural Network	Standard	0.74	Min-Max	0.6965

4.1. Logistic Regression

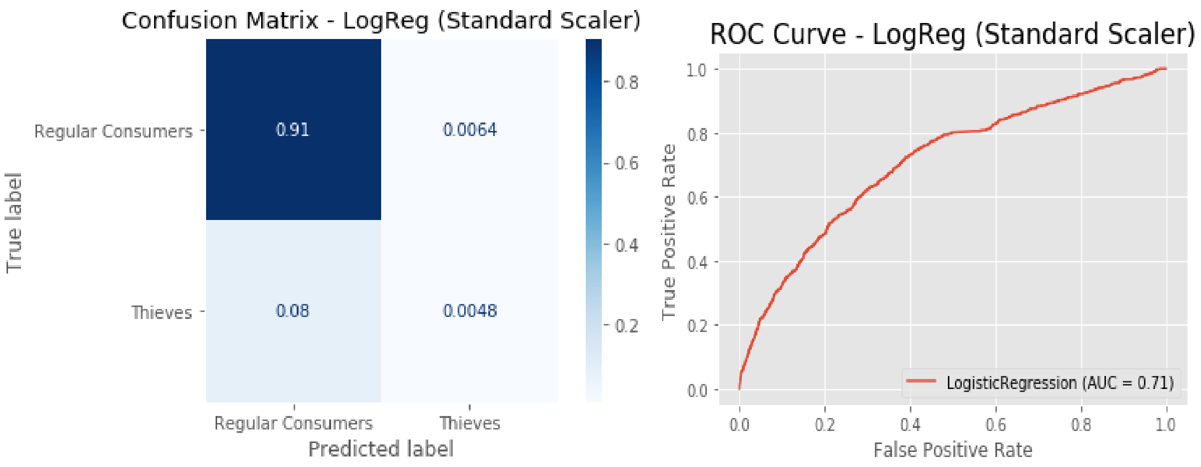

Figure 8:confusion matrix and ROC Curve of Logistic Regression

Extensive experimentation on logistic regression was performed as the results obtained were not satisfactory and lagged far behind results obtained using hyperparameters mentioned in [3]. With 500 iteration, 'liblinear' solver and an L2 penalty, the very first modeled trained on min-max scaled data gave an AUC of **0.52.** Similar hyperparameters but with a SGDC Classifier gave an AUC of **0.706**. In addition to this a grid search was also performed on the inverse regularization parameter 'c' which didn't improve the model much. Finally the standard scaled data was used in hopes of improving the AUC which had a marked effect on the classifier and bumped the AUC to **0.71.**

4.2. SVM

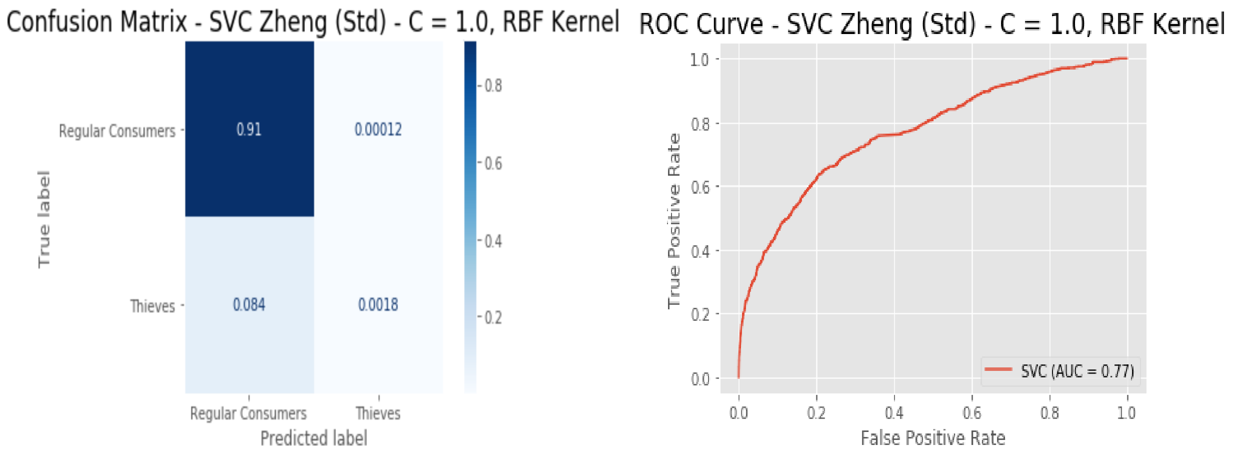

Figure 9:Confusion matrix and ROC curve of Support Vector Classifier

The Support Vector Classifier was trained using the standard scaled data. A very common Radial Basis Function (RBF) kernel was used with a penalty parameter C = 1.0. These hyper parameters were selected following previous investigations on SVMs for classifying electricity thieves from normal consumers [3]. A larger value of penalty parameter defines that the classification boundary opt by the SVM would be considerably small hence would lead to a better classification of the features. The results of SVM classifier for the purpose of theft detection as illustrated in [3] gave us an AUC of **0.7413**. This result was obtained by scaling the data according to Min-Max method. A selfexperimentation of the same model with the hyperparameters mentioned in [3] improved the AUC to **0.77** with a training ratio of 80% and using the same scaling procedure for the features. Further investigation in hopes of improving the AUC led us to train the model with standard scaled data as it was proven to slightly outperform the models trained on min-max scaled data. But after training the model it gave us an gave us an AUC of **0.75** with standard scaling and 80% training ratio.

4.3. Random Forest

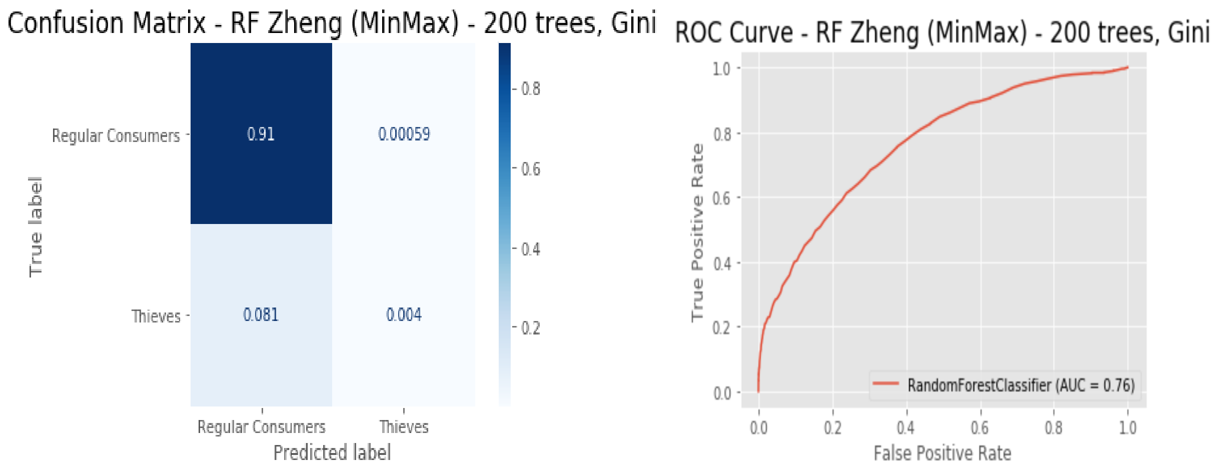

Figure 10:Confusion matrix and ROC curve of Random Forrest Classifier

For this experimentation 2 separate models were trained using the standard scaled and Min-Max scaled data. Other hyperparameters include 200 decision trees working on 'Gini' criterion which were selected as per previous investigations in this regard [3]. Note that the max_depth hyperparameter was not initialized in any training of the model. The model trained using the Mix-max scaled data gave an AUC of **0.76** while the same hyperparameters gave a marginally better AUC of **0.73** with standard scaled data. Just like the SVC classifier the Random Forest model performed slightly better on min-max scaled data. Increasing the number of estimators (trees) may produce better results but with a risk of overfitting the training data and a considerable increase in training time. With a satisfactory set of metrics obtained with 200 estimators, no parameter search was performed.

4.4. Wide Neural Network

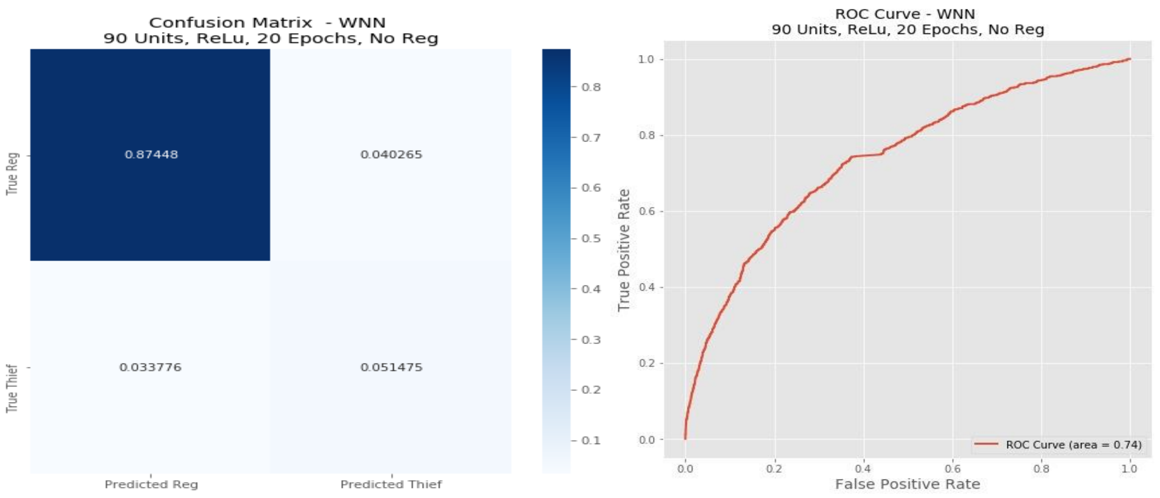

Figure 11: Confusion matrix and ROC curve of Wide Neural Network

A wide Neural Network with hyperparameters based on previous investigation [3] was also trained that outperformed other classifiers. The first dense layer had 90 neurons with a ‘Relu’ activation that fed a sigmoid classifier which performed the binary classification between thieves and normal consumers. The model was trained for 20 epochs only as there was no dropout layer in the initial models that would prevent overfitting of the model on the data. The AUC achieved using the above described model was **0.74** which was far better than the results produced by other shallow classifiers. Although the same standard scaled data was used for this experiment as well.

CHAPTER 5
EXPECTED OUTCOMES

5. EXPECTED OUTCOMES

5.1. Convolutional Neural Network

WNNs are particularly well-suited for *memorizing* how specific combinations of features in the global feature space correlate with a specific class. However, they are incapable of generalizing to it by identifying cross-feature transformations which may be correlated with the positive class but do not appear in the data [3]. Zheng et. al [4] observe it is difficult, if not impossible, for a WNN processing 1-D timeseries data to identify the periodicity inherent in the kWh consumption of regular consumers that distinguishes them from electricity thieves.

As such, the next logical step in this project is to train a Convolutional Neural Network (CNN) which will operate on 2-D weekly, rather than 1-D daily kWh data, thus *generalizing* well to the kWh data [4]. A CNN will likely be able to extract increasingly abstract feature representations from the 2D kWh data by convolving them with filter maps and reducing dimensionality through pooling layers [1]. The output of the CNN will then be fed to a densely connected classifier, much like the WNN, which will then use the generalized, abstract features for classification.

5.2. Wide And Deep Convolutional neural Network

A CNN is capable of *generalizing* to data and extracting abstract features, but will not *memorize* exceptions to these generic feature-output correlations that are often necessary in complex learning problems. While WNNs do not suffer from this shortcoming, they are incapable of generalizing well to training data. As the two network topologies complement each other, it will be worthwhile to investigate a wide and deep convolutional neural network (WDCNN) which combines them in a single, multi-input model [1] capable of both generalization and memorization [3]. Zheng et. al [4] have found the WDCNN to outperform its constituents on the electricity theft detection classification task, with Cheng et. al [4] reporting similar results on a recommender system trained and evaluated on the Google Play Store. This model will be built with the `keras` functional API and/or Tensorflow.

5.3. Hyper Parameter Tuning

Using grid search and cross validation, optimal hyperparameters for all shallow and deep learning models can be optimized to explore if further improvements in model AUC and other performance metrics is possible.

CHAPTER 6
CONCLUSION, FUTURE RECOMMENDATIONS

6. CONCLUSIONS, FUTURE RECOMMENDATIONS

This investigation concludes that shallow and deep learning-based approaches for electricity theft detection data are both viable and effective. All models have demonstrated the ability to extract meaningful transformations of kWh consumption patterns from 1D timeseries data in order to distinguish electricity thieves from regular consumers. Experimental results demonstrate that all models, with the exception of Random Forest classifiers, offer better classification performance when trained with standardized data rather than the min-max normalized data recommended by [3]. This difference in feature scaling techniques is also why experimental results of this investigation are approximately 1 - 5% higher than those presented in [3]'s research paper for the same training ratio.

6.1. On Specific Models

A WNN trained on standardized data with the same hyperparameters as those presented in [3]'s work has the second highest ROC AUC score among all investigated models. Trends in the model's validation set loss and ROC AUC scores show evidence of overfitting which, counterintuitively, has not been mitigated by the use of conventional regularization techniques such as Dropout, L1, and L2 kernel regularizers. This suggests the model may have too large a memorization capacity [12], which may be mitigated by lowering the number of neurons or further hyperparameter tuning.

Results also indicate that Random Forest (RF) classifiers are the best shallow learning approach to the problem. Despite not requiring the computational overhead of feature scaling common to almost all other models, RF classifiers have the highest ROC AUC score across all investigated models. However, this is likely due to the fact that RF classifiers are an example of ensemble learning [14]: specifically, a collection of Decision Tree classifiers learning and predicting independently on the same data. Therefore, any comparison between the performance of standalone models such as logistic regression (LR) and support vector machines (SVM)s must account RFs being an ensemble learning solution.

While SVMs demonstrate incrementally better performance on this classification problem than the Wide Neural Network (WNN), this improvement is far outweighed by the disproportionately large training and prediction times. SVM time complexity of $O(n \times m^2) - O(n \times m^3)$ [15] means a theft detection model based on an SVM classifier is highly unlikely to scale well with training examples of a real-world distribution system, although it may find use as an online learning classifier that improves its weights incrementally using individual samples [15].

Logistic Regression offers the worse performance on this classification task, which is a consequence of the fact that the model has high bias: it assumes the kWh data of consumers and regular consumers is linearly separable. Since the logistic regression model is based on a linear combination of features and weights, its hypothesis space is far too limited to capture all the non-linearities necessary for successfully distinguishing a regular consumer from a thief. That being said, its ROC AUC score is still of the same order as all other models, and the model itself is far more interpretable than WNN and SVM classifiers.

6.2. Future Work

6.2.1. Outlier/Novelty Detection

A fundamental issue intrinsic to this binary classification problem is that of a class imbalance in the data: there is a 91.5%-8.5% split between regular consumers (the negative class) and thieves (the target class). As such, the electricity thieves in the dataset are anomalous samples, which may be detected as either outliers or novelties. The former will transform this problem into an unsupervised learning task whereas the latter will be a semi-supervised learning task.

6.2.2. Ensemble Learning

Random Forest classifiers demonstrated the best performance on this binary classification task thus far, primarily because they are an ensemble model: they combine the predictions of multiple Decision Tree classifiers, which makes predictions more reliable. It is worthwhile to investigate the performance of an ensemble learning classifier consisting of a support vector machine, a random forest classifier, and other classification models to explore whether together these models can complement each other and lead to better predictions.

6.2.3. Periodicity as a Feature

Zheng et. al [3] have identified the periodicity of kWh consumption values as a promising signal to distinguish electricity consumers from thieves. Applying autocorrelation to the kWh consumption data of each consumer may therefore yield a new set of features which can be fed to both neural networks and shallow learning models. This will enable an assessment of the veracity of Zheng’s claims, and may also improve classification performance in shallow learning models. Deep learning models such as Convolutional Neural Networks (CNNs), however, will not benefit from this substantially as Zheng et. al [3] posit that such networks are already capable of extracting such features - and more - from 2D kWh consumption data

7. REFERENCES

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. S. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Science and Technology,* vol. 19, no. 2, pp. 105-120, 2014.

[2] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications,* vol. 31, no. 7, pp. 1319-1330, 2013.

[3] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics,* vol. 14, no. 4, pp. 1606-1615, 2017.

[4] R. M. Rotty, "Growth in global energy demand and contribution of alternative supply systems," *Energy,* vol. 4, no. 5, pp. 881-890, 1979.

[5] The News, "Areas of power theft will not be exempted of load shedding: Kh Asif", 2017. [Online]. Available: https://www.thenews.com.pk/latest/196271-Areas-of-power-theft-will-not-be-exempted-of-load-shedding-Kh-Asif [Accessed: 25-Feb-2020]

[6] Pakistan Today, "'Power theft' areas to brave load shedding, says minister", 2019. [Online] Available: https://www.pakistantoday.com.pk/2019/06/08/power-theft-areas-to-brave-load-shedding-says-minister/ [Accessed: 25-Feb-2020]

[7] F. Jamil and E. Ahmad, "An Empirical Study of Electricity Theft from Electricity Distribution Companies in Pakistan", Ph.D, National University of Sciences and Technology, Islamabad, 2019. [Online]. Available: https://www.pide.org.pk/psde/pdf/AGM29/papers/Faisal%20jamil.pdf [Accessed: 25-Feb-2020]

[8] ARY News, "Power theft causes financial loss of more than Rs45 bln: NEPRA", 2019. [Online]. Available: https://arynews.tv/en/power-theft-financial-loss/ [Accessed: 25-Feb-2020]

[9] "Smart meters help reduce electricity theft", Bchydro.com, 2020. [Online]. Available: https://www.bchydro.com/news/conservation/2011/smart_meters_energy_theft.html. [Accessed: 26- Feb- 2020].

[10] B.C.Costa, B. Alberto, A. M. Portela, M. W and E. O.Eler, "Fraud Detection in Electric Power Distribution Networks using an Ann-Based Knowledge-Discovery Process", *International Journal of Artificial Intelligence & Applications*, vol. 4, no. 6, pp. 17-23, 2013. Available: 10.5121/ijaia.2013.4602.

[11] C. Ramos, A. Souza, G. Chiachia, A. Falcão and J. Papa, "A novel algorithm for feature selection using Harmony Search and its application for non-technical losses detection", *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 886-894, 2011. Available:10.1016/j.compeleceng.2011.09.013.

[12] F. Chollet, *Deep learning with Python*. New York: Manning Publications Co., 2018, pp. 108-110.

[13] A. Géron and R. Demarest, *Hands-on machine learning with Scikit-Learn and TensorFlow*. Sebastopol (Clif.) [etc.]: O'Reilly, 2019, pp. 163-164.

[14] H. Cheng, "Wide & Deep Learning: Better Together with TensorFlow", *Google AI Blog*, 2016. .

[15] H. Cheng et al., "Wide & Deep Learning for Recommender Systems", *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 2016. Available: https://arxiv.org/pdf/1606.07792.pdf. [Accessed 25 February 2020].

[16] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection", *ACM Computing Surveys*, vol. 41, no. 3, pp. 1-58, 2009. Available: 10.1145/1541880.1541882.

[17] "K-Electric’s Drive Against Defaulters and Power-theft Continues Full Throttle - K-Electric", *K-Electric*, 2020. [Online]. Available: https://www.ke.com.pk/2019/01/03/k-electrics-drive-against-defaulters-power-theft-continues-full-throttle/. [Accessed: 28- Feb- 2020].