

Short thesis for the degree of doctor of philosophy (PhD)

Analytics of Non-Technical Loss Detection

Khawaja MoyeezUllah Ghori

Supervisor: Dr. Laszlo Szathmary



UNIVERSITY OF DEBRECEN
Doctoral School of Informatics

Debrecen, 2020

1. Table of Contents

2. Abstract.....	1
3. Introduction	2
3.1 Knowledge Gap in Existing Techniques for NTL Detection	3
3.1.1 Identification of Best Performance Evaluation Metrics for NTL Detection.....	3
3.1.2 Comparison of Performance of Different Types of Classifiers for NTL Detection	3
3.1.3 Identification of Best Features for NTL Detection in a Real Dataset	4
4. Aims of the Thesis.....	5
5. Thesis Overview	5
6. Thesis Points.....	7
7. State of the Art Methodologies in NTL Detection	7
8. Taxonomy of the Literature Review	8
9. Data Collection.....	10
10. First Contribution: Performance Analysis of Machine Learning Classifiers for Non-Technical Loss Detection	11
10.1 Results.....	11
11. Second Contribution: Analyzing Types of ML Classifiers for NTL Detection.....	13
11.1 Selecting Top-k Features.....	13
11.2 Performance Analysis of Classifiers with Respect to their Types	14
11.3 Best Performing Classifiers for NTL Detection	17
11.4 Performance Analysis of Deep Learning on Reduced Feature- Set	17
12. Third Contribution: Impact of Feature Selection on Non- Technical Loss Detection	18
12.1 The Incremental Feature Selection Algorithm.....	18

12.2	Results.....	19
13.	Conclusions and Future Work	21
13.1	Conclusions	21
13.2	Future Work.....	25
14.	Bibliography.....	24
15.	Certified Publication List (Khawaja MoyeezUllah Ghorl).....	26

2. Abstract

With the ever-growing demand of electric power, it is quite challenging to detect and prevent Non-Technical Loss (NTL) in power industries. NTL is committed by meter bypassing, hooking from the main lines, reversing and tampering the meters. Many countries suffer huge losses in billions of dollars due to NTL in power supply companies (Bhat, 2016), (Hussain, 2016). Manual on-site identification of NTL remains an unattractive strategy due to the required manpower and associated cost. The use of machine learning classifiers has been an attractive option for NTL detection.

The literature review has identified the knowledge gap in NTL detection. This gap is surrounded by first finding the best metrics that can identify the top performing classifiers considering the requirements of NTL detection. Secondly, using those metrics finding the best classifiers and the types of the classifiers for NTL detection. Finally, quantifying the impact of feature selection in a real dataset for NTL detection.

Firstly, we compare 14 performance evaluation metrics across the three classifiers and identify the key scientific relationships between them specifically related to NTL detection in a real dataset of an electric supplier containing approximately 80,000 monthly consumption records. We concluded that recall is the best performance measure for NTL detection.

Secondly, we evaluate 15 existing machine learning classifiers for NTL detection across 9 different types. Our work is validated using extensive simulations. Results show that ensemble methods and Artificial Neural Network (ANN) outperform the other types of classifiers for NTL detection. Moreover, we have also derived a procedure to identify the top-14 features out of a total of 71 features, which are contributing in 77% of the prediction in NTL.

In our next contribution, we propose the Incremental Feature Selection (IFS) algorithm, which first uses feature importance to identify the

most relevant features for NTL detection and then these features are used to test the three classifiers namely CatBoost, Decision Tree (DT) Classifier and KNN for NTL detection.

The results show that using the most relevant features identified by the IFS algorithm, the three classifiers have the same or slightly better efficiency as compared to using all features.

Overall, the thesis has contributed in three main processes of NTL detection i.e. the feature selection, the identification of suitable machine learning classifiers and their types, and the identification of suitable performance evaluation metrics for NTL detection.

3. Introduction

Over the last decade, the research community has been actively participating in an attempt to bring down the occurrences of NTL in power industry. For this, the use of physical devices as well as the data analytics on consumption patterns are major sources of NTL detection schemes. Many developed countries have installed Advanced Metering Infrastructure (AMI) in their network. The AMI is a smart metering system with a two-way communication between the supplier and the consumer that helps in better management of consumption history. The use of AMIs enables the company to record the consumption history over a different intervals of time like hourly, twelve-hourly, daily and bi-monthly consumption recordings. However, many under-developed countries still use the traditional meters for which monthly manual meter reading is required. Due to the use of old metering structure, the chances of NTL grow as the use of meter by-passing, magnet and false meter reading is easy in old metering infrastructure. In order to detect NTL, companies have used separate physical observer devices in meters and distribution poles which calculate the difference of energy supplied and used. However, this approach incurs heavy monetary cost. Another widely adapted method for NTL detection is the use of machine learning classifiers on a dataset of hourly, daily or monthly consumption records. These classifiers help in identifying potential theft which otherwise is a costly manual effort. However, still there is a need to improve the way we use classifiers and their evaluation

metrics for NTL detection. This knowledge gap is highlighted in the next section.

3.1 Knowledge Gap in Existing Techniques for NTL Detection

A substantial literature is available on the techniques used for NTL detection. However, there are open challenges and knowledge gaps which are still needed to be addressed. These gaps are shown in Figure 1 and described in the subsequent sections.

3.1.1 Identification of Best Performance Evaluation Metrics for NTL Detection

In a real dataset, the ratio of records with theft and normal consumption is vastly imbalanced. That is, the number of records of normal consumption is very high whereas the number of records of theft instances are very small. The problem becomes interesting when the focus is on least representative records (Awais M. a., 2016), i.e. the theft records. This type of problem belongs to class imbalance problem where the ratio of representation of the two classes is highly imbalanced. For this reason, not every metric is useful in evaluating the performance of the classifiers used. Keeping in mind the number of normal consumption identified as theft and the number of theft records identified as normal, there is a need to prioritize the available metrics which can best describe the performance of the classifiers used in NTL detection.

3.1.2 Comparison of Performance of Different Types of Classifiers for NTL Detection

A range of machine learning classifiers is available for training and testing. These classifiers belong to different types. Efforts have been made to compare different classifiers for NTL detection but there is a need to compare not only the performance of the individual classifiers but also the comparison of the types of the classifiers for NTL detection in some real dataset. Also, a recently developed CatBoost (Prokhorenkova, 2018), LGBBoost (Ke, 2017) and XGBoost (Chen,

2016) classifiers are also needed to be tested for NTL detection in a real dataset.

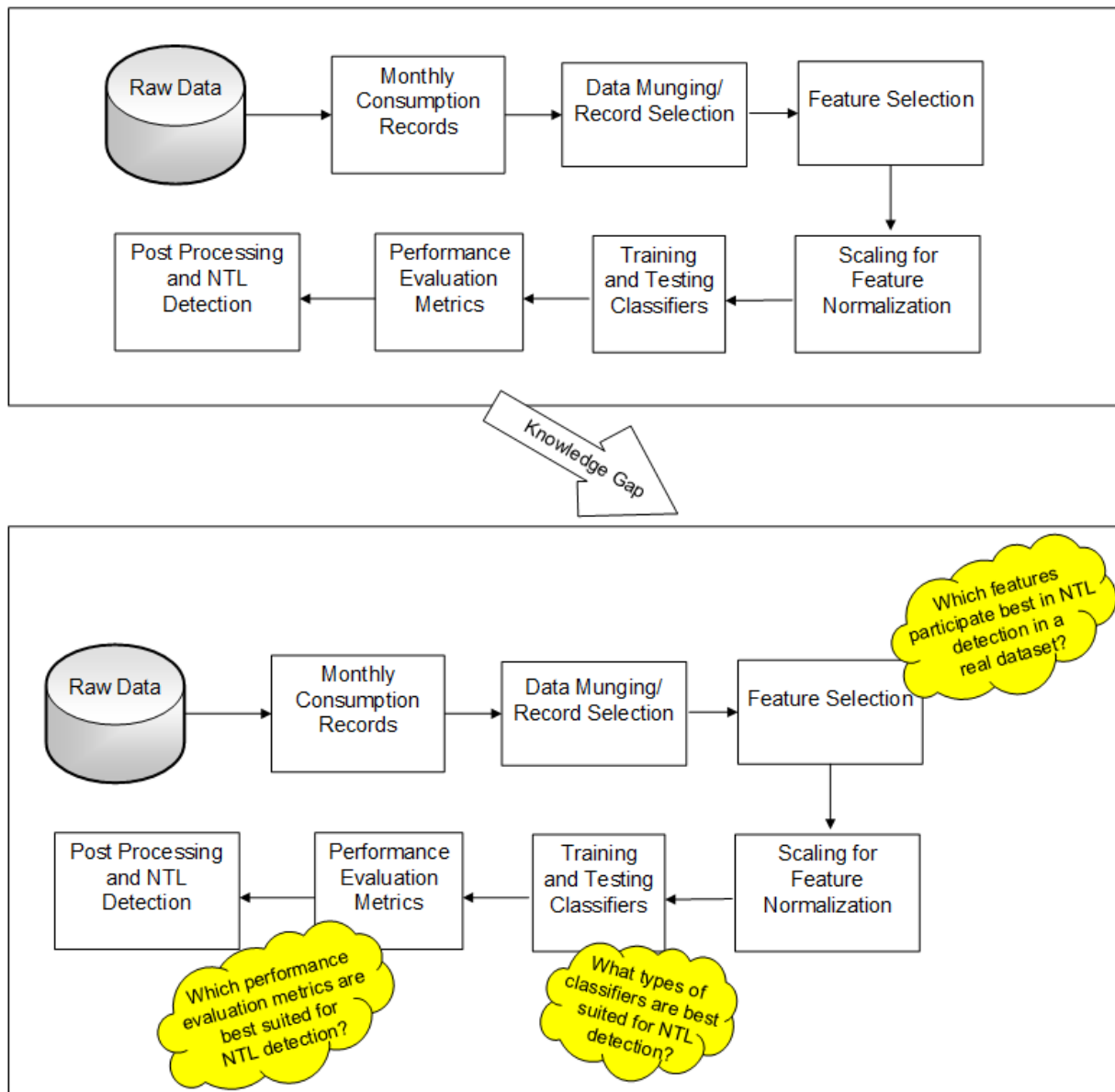


Figure 1: Knowledge Gap in NTL Detection

3.1.3 Identification of Best Features for NTL Detection in a Real Dataset

Feature selection is an important process before the selected data is used for the training and testing of the classifiers. Multiple techniques for feature selection are available. For NTL detection, there is a need to identify most relevant features in a real dataset. In other words, the feature importance of every feature in participating in NTL detection is needed to be precisely monitored and evaluated for some real dataset.

The performance of using only selected features is also needed to be compared with the performance when all features are used.

4. Aims of the Thesis

The main aim of the thesis is to first find the best metrics that can evaluate classifiers considering the requirements of NTL detection. Then, use those metrics for better evaluation of the classifiers and the types of the classifiers. The objectives of the thesis are outlined below:

1. Work on a real dataset for the identification of best metrics which can evaluate different machine learning classifiers considering the requirements of NTL detection.
2. Compare the performance of the different machine learning classifiers used for NTL detection.
3. Compare the performance of the types of machine learning classifiers for NTL detection.
4. Compare the performance of recently developed CatBoost, LGBost and XGBoost classifiers for NTL detection in a real dataset.
5. Identify the most relevant features for NTL detection in a real dataset.

5. Thesis Overview

The thesis is comprised of six chapters. The details of the chapters are highlighted in Figure 2.

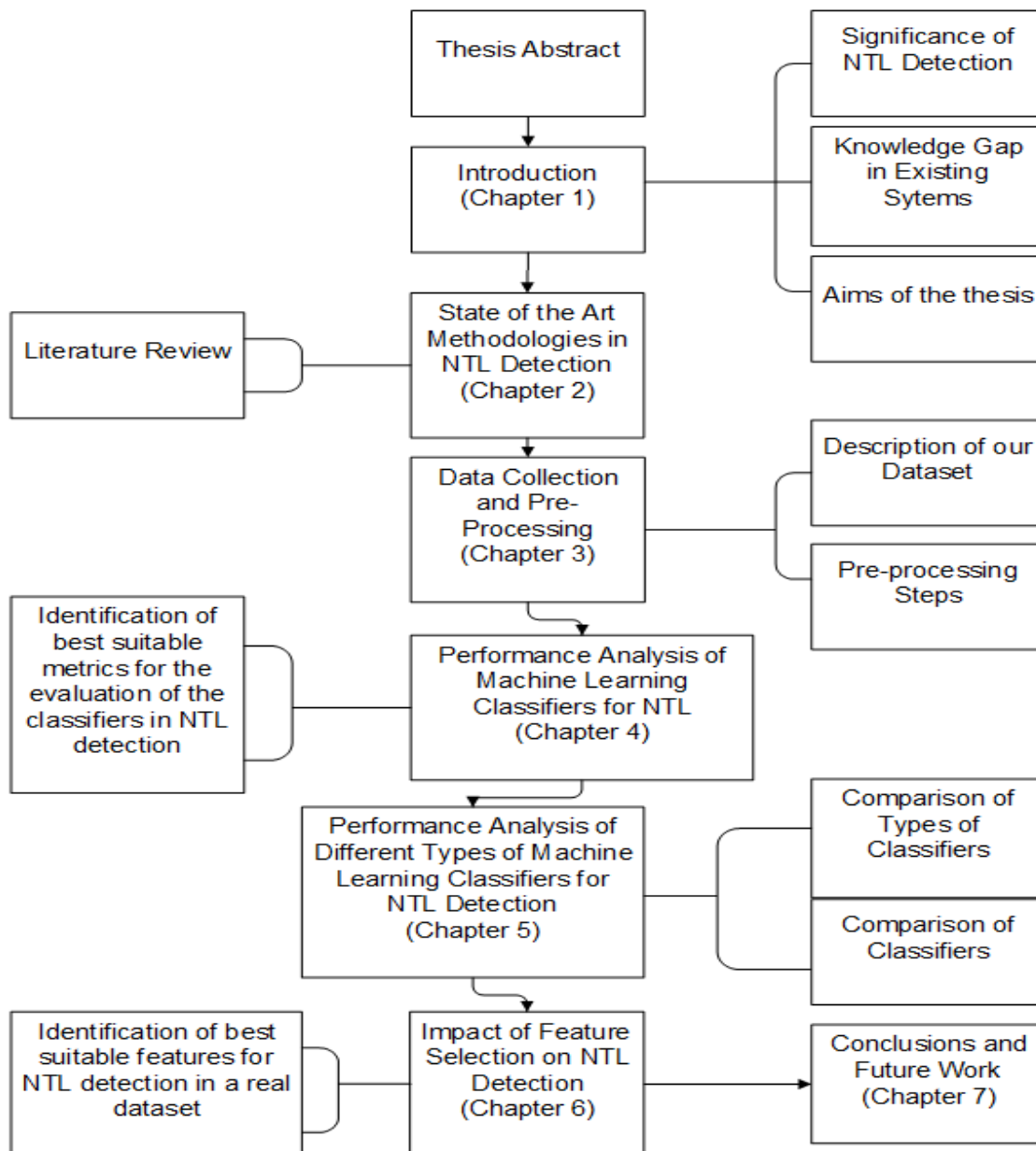


Figure 2: Thesis Overview

6. Thesis Points

The thesis points are summarized below:

1. The thesis contains a comprehensive literature review which has enabled to identify the limitations of recent works in NTL detection.
2. A total of 14 performance evaluation metrics are analyzed for NTL detection using different classifiers. We found that recall should be given higher priority for NTL detection, and random forest is the better algorithm for it having the highest recall.
3. We have performed the testing of 15 machine learning classifiers belonging to 9 different types. The MLP classifier was found the best individual classifier with respect to recall, and ANN was found the best type of the classifiers for NTL detection.
4. We proposed a novel framework to identify relevant features for NTL detection by using the Incremental Feature Selection (IFS) algorithm, which identifies the most relevant features for NTL detection in a real dataset using feature importance. The results have shown that with the use of the IFS algorithm, recall and F-Measure of KNN is increased by 120% and 60%, respectively, while the training time of KNN is reduced by 90%.

7. State of the Art Methodologies in NTL Detection

During the last few years, the research community has paid attention to the problem of NTL detection. To encounter this problem, supervised, unsupervised and semi-supervised learning methods have been used. Some of the authors have used customers' consumption history while others have used the grid and network data. Effort has also been made to use both types of data, i.e. consumers' consumption profile as well as the grid data which may contain current and voltage information supplied to different areas. At times, some additional data are also merged to the consumption data to see the effect of hit ratio of NTL detection. This additional data may comprise of environmental and temperature readings. Figure 3 depicts the historical state of the art methodologies used in NTL detection.

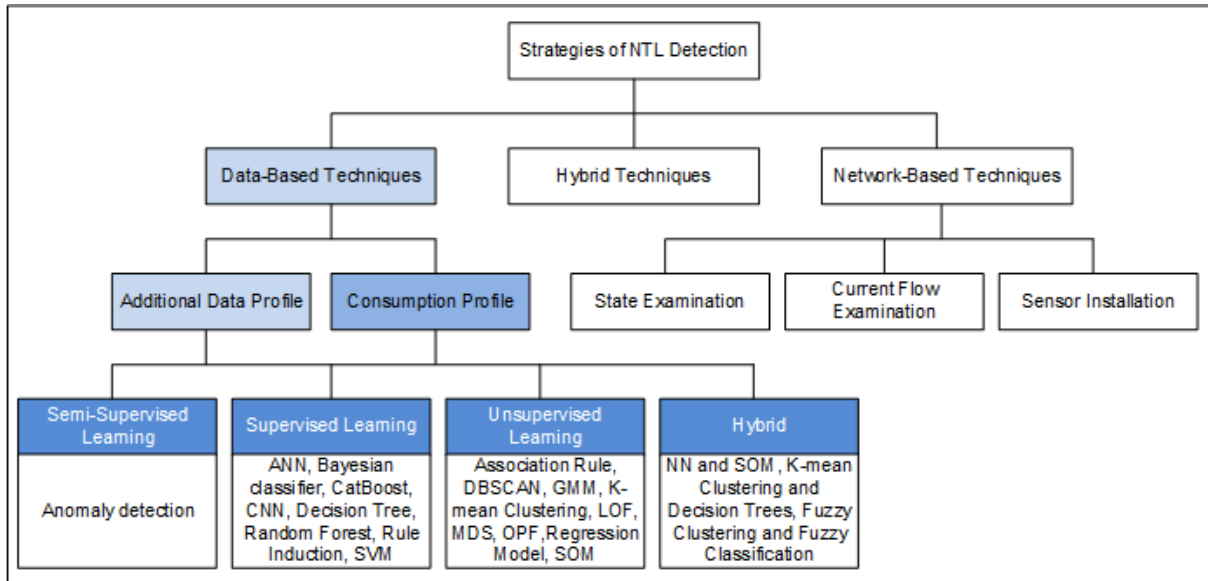


Figure 3: Strategies of NTL Detection

8. Taxonomy of the Literature Review

One of the contributions of the thesis is the derivation of a complete taxonomy for the strategies of NTL detection as described in Figure 4.

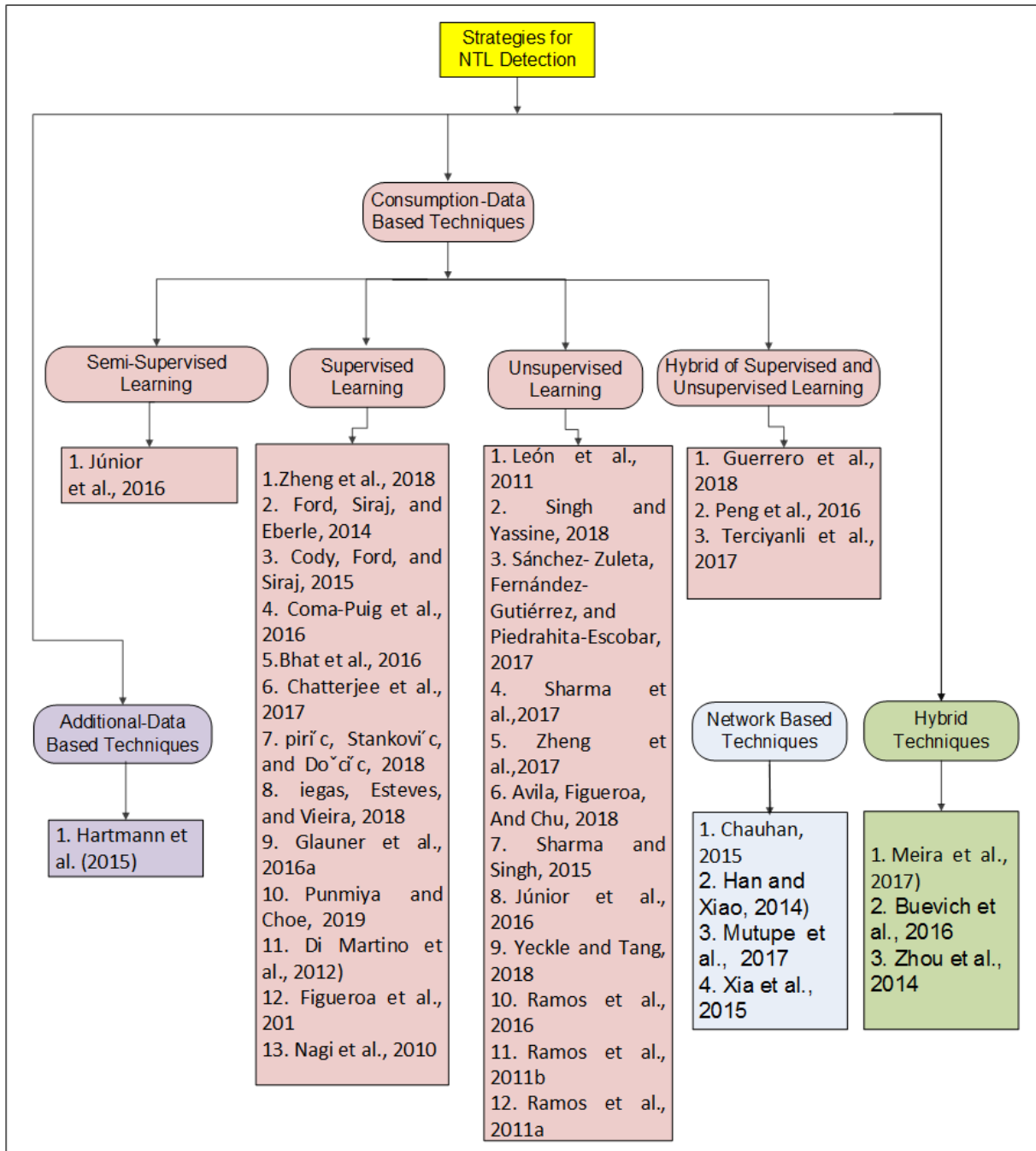


Figure 4: Taxonomy for Strategies of NTL Detection

9. Data Collection

Relevant data is the most important ingredient for a success or a failure of a solution to a problem. Percentage of success is increased with the increase of relevancy of data. For NTL detection problem, a major constraint is access to real data. For a power supply company, the client's data must be secured. For this reason, no real data is freely available from power supply companies. For NTL detection, access to real data becomes more important due to the fact that only real data can give real insights to the occurrences of NTL. On the other hand, synthesized data has its own limitations as it does not represent real instances. There have been many attempts for NTL detection in synthesized data. The techniques used to synthetically produce instances of NTL include decreasing the units consumed or decreasing the amount billed. However, in real-life, many factors affect the occurrences of NTL like the neighborhood area, amount of load shedding, the billed amount, etc. These factors are ignored in synthetically built data.

We have collected a dataset from an electric supply company in Pakistan. The collected data contains monthly consumption records of consumers which range between January 2015 and March 2016. It comprises of 80,244 monthly consumption records. For privacy reasons, we have changed the feature names and have omitted the clients' details.

The dataset is randomly split into train and test sets with the ratio of 80% and 20%, respectively. The training set contains 64,195 records out of which 61,456 are normal records with no theft and 2,739 are abnormal records where the users have committed stealing of electricity. The test set contains 16,049 records out of which 15,366 are normal consumption records and 683 records contain NTL. The percentage of NTL in both sets is 4%.

As the number of normal users is always much greater than the number of abnormal users (thieves), this data is considered imbalanced and

biased towards major representation of normal users. This behavior is shown in Figure 5.

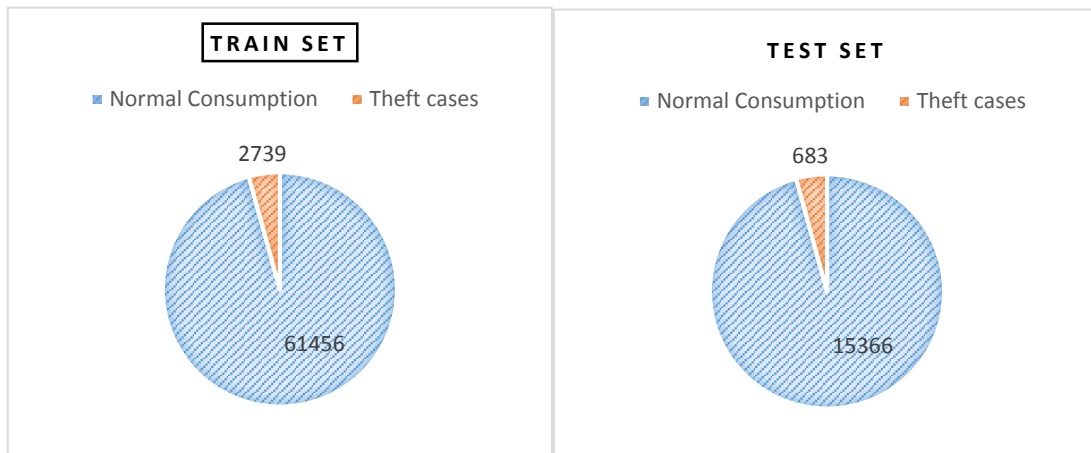


Figure 5: Imbalance Data Representation

10. First Contribution: Performance Analysis of Machine Learning Classifiers for Non-Technical Loss Detection

As the problem of NTL detection belongs to the class imbalance domain, a bias is observed between the number of observations for the normal consumption and the number of observations for the theft cases. Therefore, considering the requirements of NTL detection, there is a need to identify the metrics that are best suited for the performance evaluation of the classifiers. Chapter 4 compares the performance of three classifiers using 15 performance evaluation metrics. Finally, it presents a detailed analysis of the pros and cons of each metric for NTL detection. Thus, it identifies the relationship that exists between different metrics considering the specific requirements of the NTL detection. This contribution is cited in (Ghori K. M., 2020).

10.1 Results

An important observation regarding the problem of NTL detection is that the model which has a high recall is most suitable for theft detection. In order to understand this relation, consider the cases of FP

and FN. False Positives are those normal users that have been predicted by the classifier as thieves whereas False Negatives are those thieves that are predicted by the classifier as normal users. Considering the two cases, having a large FP value will only result in increasing the manual effort of on-site inspections whereas a high FN value will result in the failure of the classifier to correctly identify the thieves. Therefore, for NTL it is recommended to promote the classifier which has a low FN value. It can be observed that recall increases with the decrease of FN. This gives a nice measure of the selection of the classifier for NTL detection that both the precision and the recall should not have equal priority. In fact, for NTL detection, classifiers with high recall are most suitable regardless of what the precision value is.

It is observed that KNN has the lowest number of FN in our simulation. Consequently, it has the highest recall among the three classifiers as shown in Figure 6. The lowest recall is observed for SVM, which is 98.39%. Thus, the percent increase of recall from using SVM to KNN is 0.89%. This gives a clear indication that for our real dataset, KNN is the better choice for NTL detection.

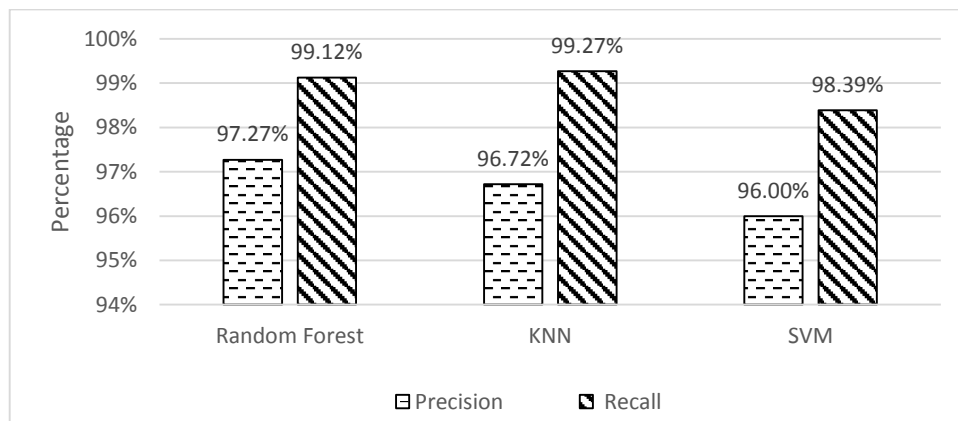


Figure 6: Comparison of Precision and Recall

This analysis can be used as a baseline for the accurate selection of the classifiers in NTL detection. This work will vastly benefit the electric supplier in detecting NTL. It will not only improve their abilities for NTL detection, but will also save huge amount of monetary losses which they are already bearing.

11. Second Contribution: Analyzing Types of ML Classifiers for NTL Detection

In recent years, different machine learning classifiers have been tested in NTL detection. However, there is a need to identify the types of the classifiers that perform best in NTL detection. This contribution compares not only the performance of the individual classifiers but also the performance of the types of the classifiers. It also introduces the recently developed CatBoost, LGBost and XGBost classifiers for NTL detection in the real dataset. Lastly, with the help of extensive simulations, it highlights the best classifiers and the best type of classifiers for NTL detection in the real dataset. In addition, this chapter compares the performance of deep learning with the performance of the other classifiers. This contribution is cited in (Ghori K. M., 2020).

11.1 Selecting Top-k Features

One of the contributions of this paper is to find the optimum number of k features that can provide best theft prediction in a real dataset. It is observed that not every feature has an equal or comparable participation in predicting NTL. Some features have a high role while others have a negligible role. Also, using all features to predict NTL will increase the computational complexity of the classifiers. It turns out that there should be a threshold for the contributing features beyond which including or excluding features should not affect the efficiency of the classifier. For this, we first sort the feature set in descending order with respect to feature importance. It is a measure that uses accuracy to filter attributes which are most suitable for correctly identifying the target variable. Thus, it gives an insight to the relative importance of every feature with respect to the target variable. Then, we apply Gini Index (Raileanu, 2004) to find the top- k number of features for which the F-measure is the highest where k ranges from 1 to the last feature. For our dataset, the value of k with best F-measure is 14. This indicates that using this set of top 14 features to find NTL has the same behavior as using all features. This simulation has not only identified key features that are participating in predicting NTL in the real dataset but it also has helped to significantly reduce the execution

time of the classifiers. The cumulative percentage of feature importance of top 14 features is presented in Figure 7.

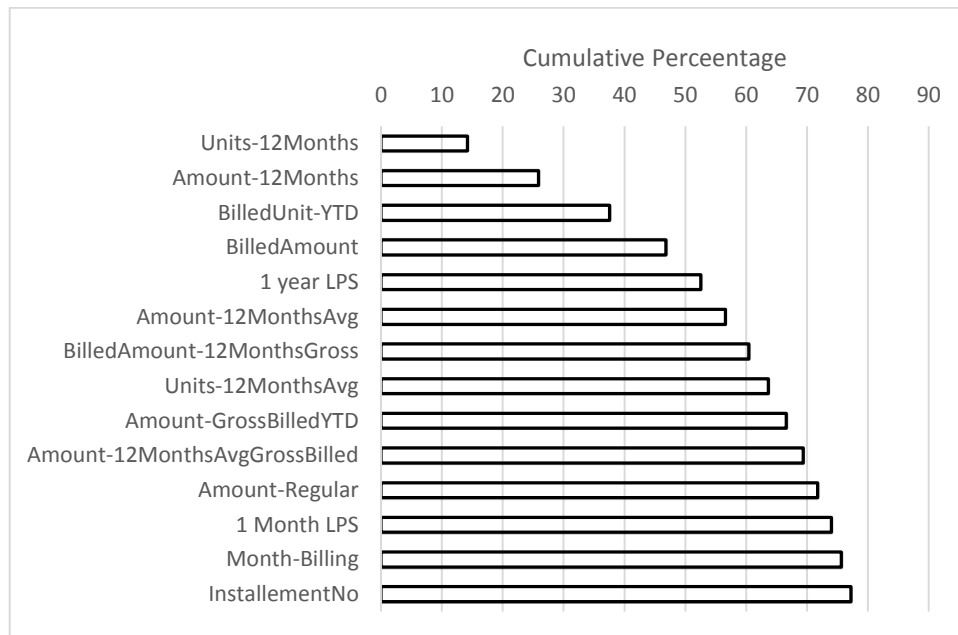


Figure 7: Cumulative Percentage of Feature Importance of Top 14 Features

11.2 Performance Analysis of Classifiers with Respect to their Types

It is interesting to observe that considering F-measure as the classifiers' efficiency measure, ensemble methods outperformed all other types of classifiers. In fact, the top three classifiers having the best F-measures belong to the ensemble methods namely CatBoost, Random Forest and XGBoost, while Naive Bayes performed worst with lowest two F-measures. A reason for this behavior is that ensemble methods are robust to overfitting as compared to Naive Bayes classifiers which tend to overfit the model. Thus, any method which overfits the model will suffer. Two classifiers are experimented in 'LDA and QDA' type. The F-measure of LDA is quite low, that is, 0.471 while the F-measure of QDA is observed to be 0.782 which shows a percent increase in the performance of 66% while the percent increase in the performance from Naive Bayes to 'LDA and QDA' type is 248%. The type 'Generalized Linear Models' performed no better than the type 'LDA and QDA'. The F-measure obtained from its classifier, that is, Logistic

Regression, is 0.787. The percent increase in the performance from 'LDA and QDA' to the type of 'Generalized Linear Models' is only 0.64%. One classifier from each of SGD, SVM, DT, NN and Nearest Neighbors is tested. Their F-measures are observed as 0.945, 0.972, 0.977, 0.979 and 0.980, respectively. Notably, all these readings are above 0.90. LGBost has the worst F-measure among ensemble methods which is 0.933 while the F-measure of AdaBoost (Punmiya, 2019) is 0.971. The percent increase in the performance from the worst to the best classifier in the ensemble methods is only 5.5%, which shows that the performance of all classifiers in ensemble methods is close to each other. The comparison of all types of classifiers using F-measure is shown in Figure 8 where T-1 to T-9 correspond to the types of Naive Bayes, LDA and QDA, Generalized Linear Models, SGD, SVM, DT, NN, Nearest Neighbors and Ensemble Methods, respectively.

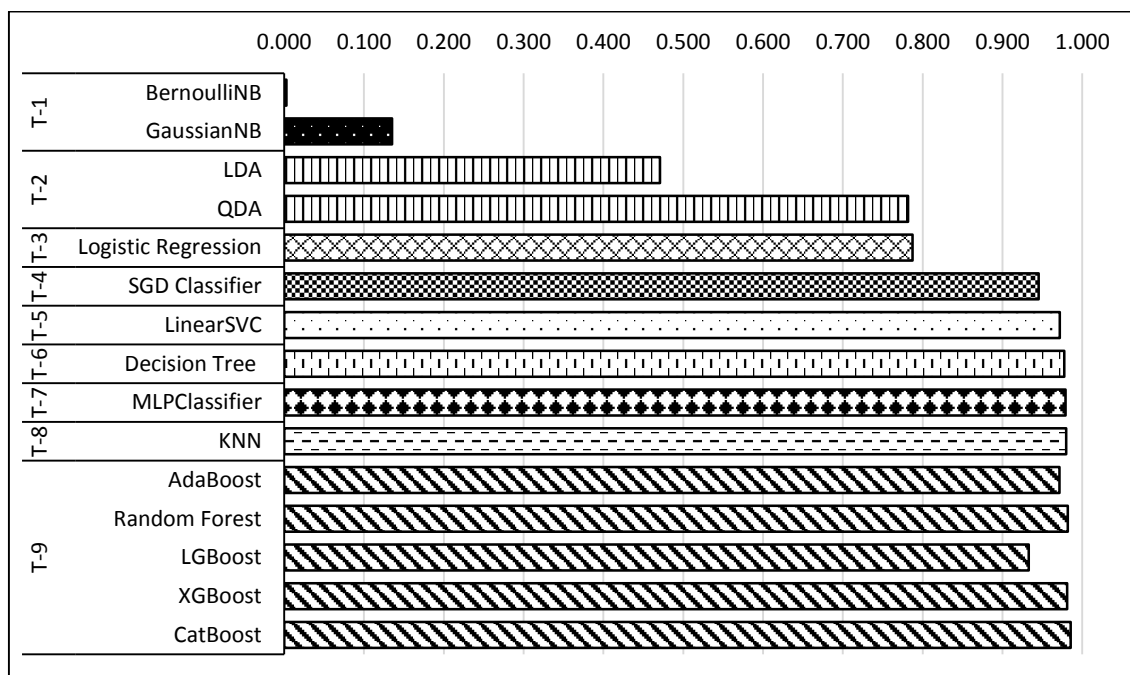


Figure 8: Comparison of Different Types of Classifiers using F-measure

Considering recall as the efficiency measure of the classifiers, NN (Ford, 2014) outperformed other types with MLP Classifier having a recall of 0.994. The worst two recalls are observed for the Naive Bayes. For the type 'LDA and QDA', LDA has a recall as low as 0.337, while the recall of QDA is 0.829, which shows a performance increase of 146%. The type 'Generalized Linear Models' performed no better than

'LDA and QDA'. The classifier used for this type is Logistic Regression. Its recall is 0.663. An interesting point is that the counterpart of logistic regression, that is, MLP Classifier (Raza, 2019) which belongs to the type of NN, has the highest recall. Thus, the percentage increase in performance from the Logistic Regression to MLP Classifier is 50%. The only difference between the two classifiers is the number of hidden layers between the input and the output layer. This observation has led us to a new future direction of testing deep learning in our real dataset. The recalls of each of the classifiers from SGD, SVM (Chang, 2011), DT and Nearest Neighbors are observed as 0.936, 0.984, 0.984 and 0.993, respectively. The recalls of ensemble methods are 0.966, 0.991, 0.887, 0.987 and 0.991. These recalls are for the classifiers AdaBoost, Random Forest, LGBost, XGBoost and CatBoost, respectively. Other than LGBost, the recalls of all ensemble methods are above 0.960, which shows that the performance of ensemble methods is very good for our data. The percent increase from the worst to the best classifier in the ensemble methods is 11.7%. The comparison of different types of classifiers using recall is shown in Figure 9.

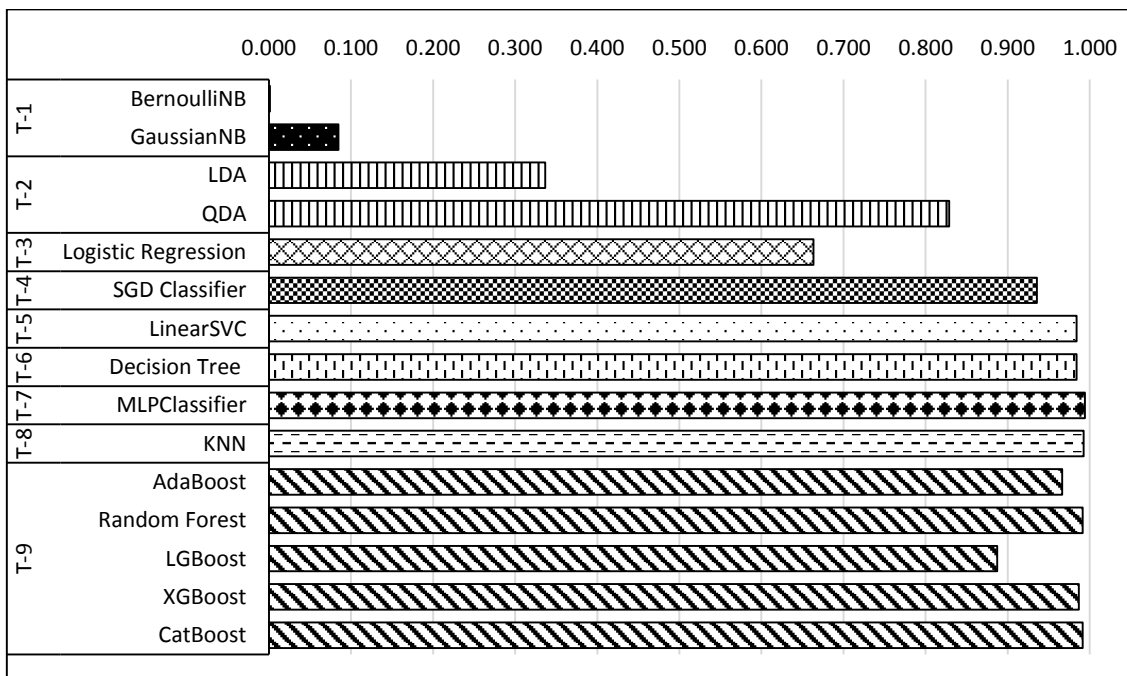


Figure 9: Comparison of Different Types of Classifiers using Recall

11.3 Best Performing Classifiers for NTL Detection

We have used precision, recall and F-measure as the performance evaluation metrics. The best F-measure is 0.985 for CatBoost classifier, which narrowly outperforms Random Forest and KNN. These three classifiers have corresponding high precision and recall values indicating small FP and small FN values, respectively.

The F-measure of LGBost classifier is 0.933, which is comparatively less than the F-measure of CatBoost classifier, i.e 0.985, while the corresponding figure for XGBost is 0.981. There is an increase of 5.6% in the F-measure from LGBost to CatBoost. Overall, precision and recall obtained for CabBoost, XGBost and LGBost classifiers are above 0.97 except that the recall of LGBost is 0.887.

The F-measures is significantly increased from 0.471 to 0.782 when choosing QDA instead of LDA, which indicates that QDA outperforms LDA. This is because when multiple classes have a different co-variance relationship then LDA suffers while QDA remains a better option. This gives an insight to the characteristics of features of this real dataset, that is, for NTL, there is a room to explore more about the co-variance relationship for individual classes.

11.4 Performance Analysis of Deep Learning on Reduced Feature-Set

In addition to finding the best classifiers and the types of the classifiers for NTL detection, a separate experiment is performed on the same dataset with different layers of neural network (MLP) while each layer consists of 100 units. Five MLPs are tested with layers of 2,3,4,5, and 10, respectively. As depicted in Figure 10, the maximum precision of 0.979 is observed for the MLP with 2 layers while the maximum recall of 1 is observed for the MLP with 3 layers. Considering recall, the MLP with 3 layers has outperformed all the other classifiers. The maximum F-measure of 0.981 is observed for the MLP with 5 layers which is comparable to the F-measures of the other classifiers. These results have encouraged us for experimenting an exploratory analysis of deep

learning with different parameters in NTL detection using a real dataset.

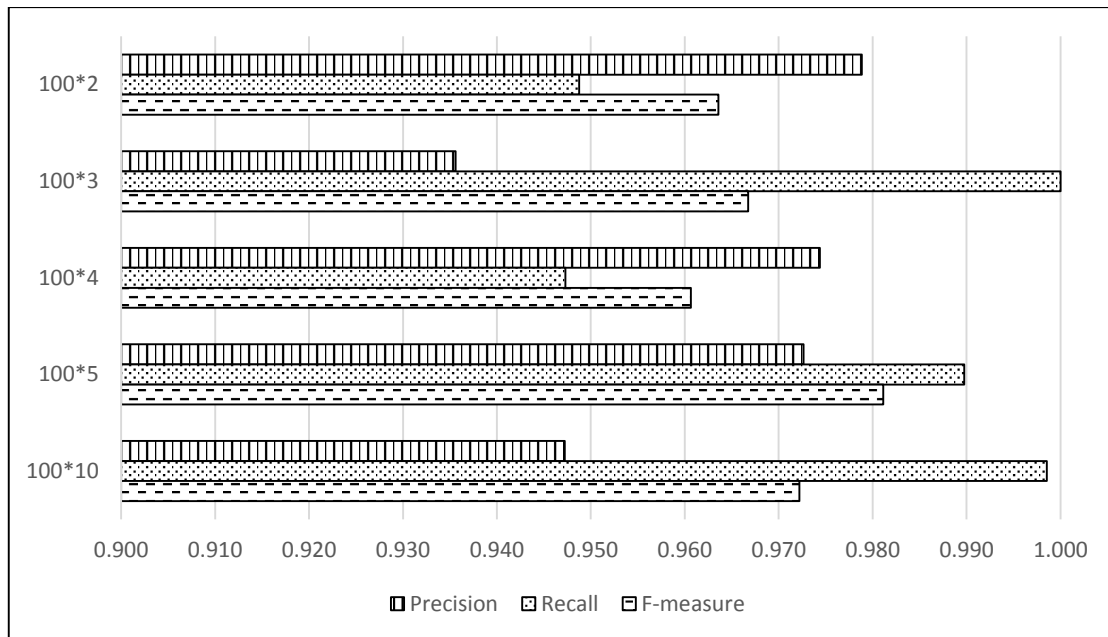


Figure 10: Precision, Recall and F-Measure of Neural Networks with Different Layers

12. Third Contribution: Impact of Feature Selection on Non-Technical Loss Detection

Not only the identification of the relevant records is important for correctly identifying the potential theft but the identification of the relevant features is equally important. This is because not all the features are relevant to the identification of NTL. This contribution identifies the best features for NTL detection in a real dataset. It also introduces our proposed Incremental Feature Selection (IFS) algorithm for feature selection. Finally, it compares the success ratio of NTL detection using all features with the success ratio of NTL detection using selected features from the IFS algorithm. This contribution is cited in (Ghori K. M., 2020).

12.1 The Incremental Feature Selection Algorithm

In this contribution, we present our Incremental Feature Selection (IFS) algorithm for selecting the best features across multiple classifiers. The IFS is shown in Figure 11. The set of selected features are first sorted

in reverse order with respect to the Gini index value. As a result, the feature with the highest information gain comes in the first place, the feature with the second highest information gain comes in the second place, and so on. The list of classifiers which is initially empty, is incrementally appended by adding the next feature. For every new feature set, the data is split into training and test set. Once the training and the test sets are ready for the features, we do the following for every classifier. First, train the classifier with the training set. Second, test the classifier with the test set. Third, compute the confusion matrix (Awais M. a., 2017) containing TP, True Negative (TN), False Positive (FP) and FN values. Finally, compute precision, recall and F-Measure. After complete execution of the algorithm, the F-Measure for the selected classifiers is obtained with initially only first feature, then with first two features, and so on. It continues until all features are selected.

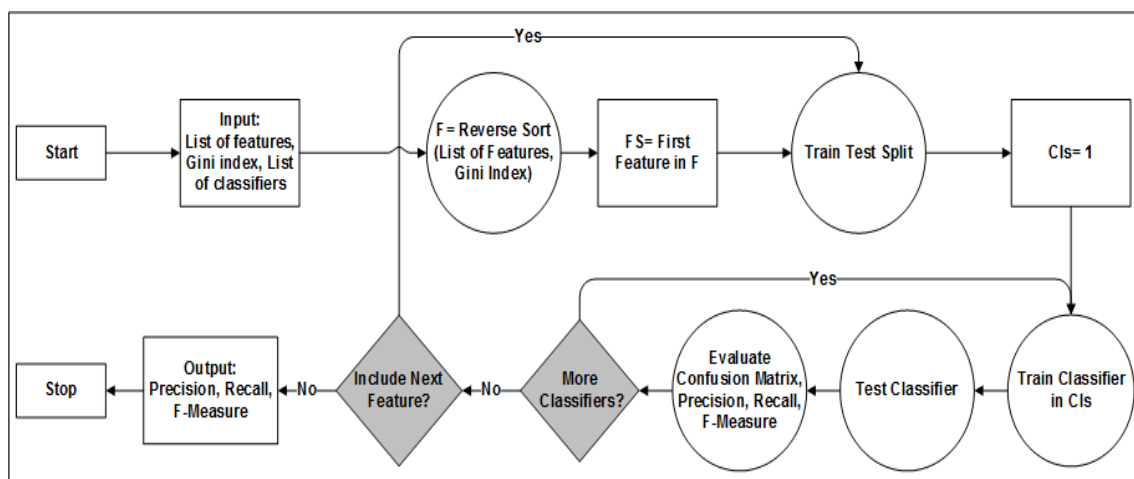


Figure 11: Incremental Feature Selection (IFS) algorithm

12.2 Results

The performance of the proposed framework has been analyzed by computing precision, recall and F-Measure of CatBoost, KNN (Goldberger, 2005) and Decision Tree Classifier (Mingers, 1989). Figure 12 shows the precision, Figure 13 shows the recall and Figure 14 shows the F-Measure of the three classifiers. The figures represent the top-9 features as identified by the IFS algorithm. These features are incrementally added with respect to their feature importance (Breiman, 2001).

A significant increase in overall performance is observed for all classifiers with the inclusion of the 3rd and 6th features. The total number of units consumed in current year is stored in the 3rd feature, while the 6th feature records average amount paid during the last 12 months. Performance of the Decision Tree Classifier is boosted with the addition of the 3rd feature, while the 6th feature has a significant impact on recall and F-Measure of KNN. With the inclusion of the 9th feature, the precision, recall and F-Measure of the three classifiers achieved near optimum values which are comparative to the corresponding figures when all features are included.

An interesting observation about the performance of the three classifiers is that it is comparative or slightly better when the features are selected using the IFS algorithm as compared to when all the features are selected. Precision and F-Measure of CatBoost using the IFS algorithm is decreased by 1% while recall remained the same. On the other hand, there is a major decrease in the training time of CatBoost with the use of the IFS algorithm. The training time of CatBoost with all features included is 149 secs and with the use of the IFS algorithm, it is reduced to 94 secs, which is a 37% reduction. Analyzing the three metrics, one can observe that the performance of KNN is improved significantly using the IFS algorithm. For KNN, precision is increased by 3%, while recall and F-Measure are increased by 120% and 60%, respectively. The training time of KNN using the IFS algorithm is also reduced by 90%.

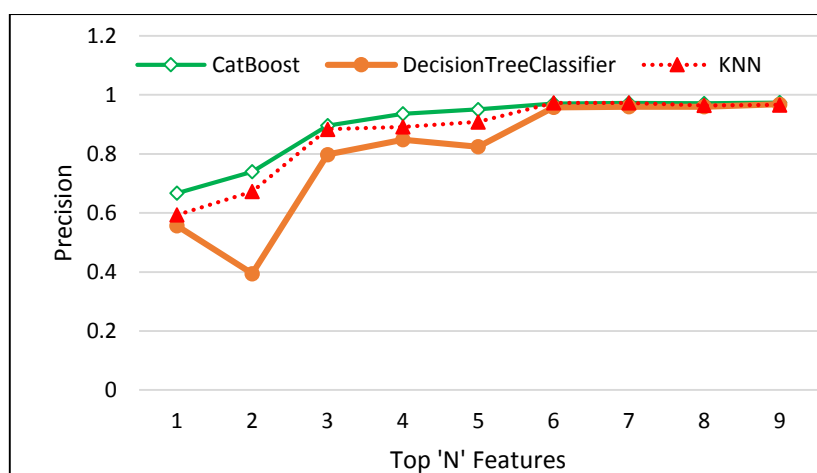


Figure 12: Precision of CatBoost, KNN and Decision Tree Classifier with 9 Features

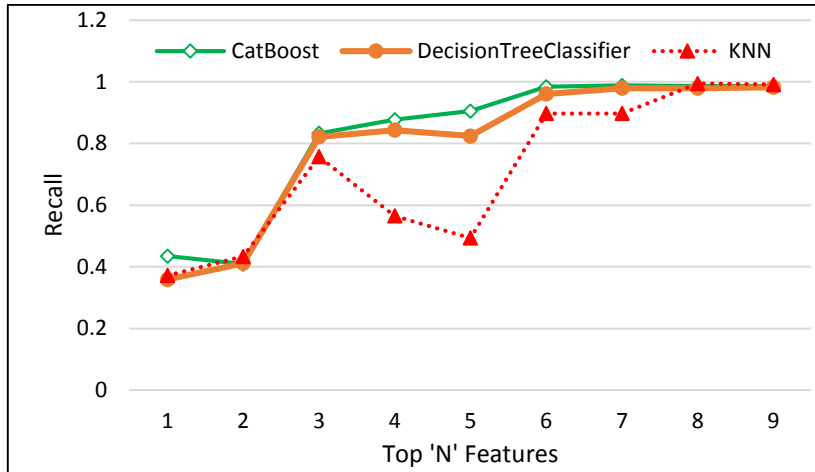


Figure 13: Recall of CatBoost, KNN and Decision Tree Classifier with 9 Features

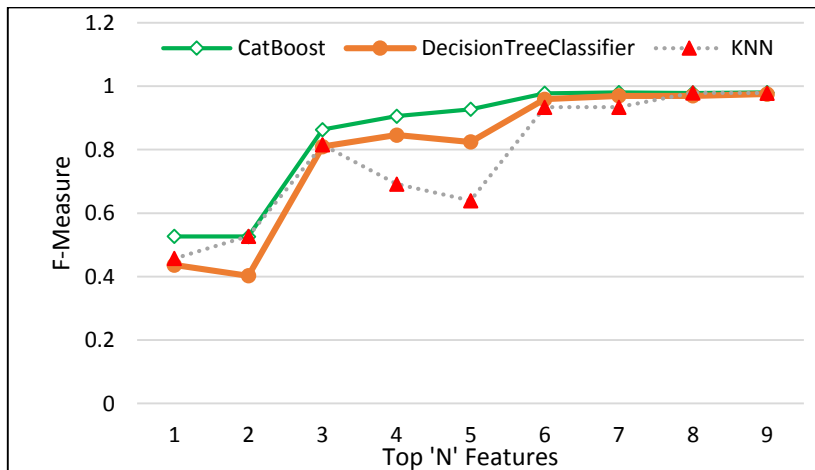


Figure 14: F-Measure of CatBoost, KNN and Decision Tree Classifier with 9 Features

13. Conclusions and Future Work

13.1 Conclusions

The objective of the dissertation was to first identify the open challenges and the knowledge gap in NTL detection through extensive literature review and then address those challenges by applying recently developed tools and techniques in a real dataset. The contributions of the dissertation are listed below:

1. We presented a comprehensive literature review of NTL.

2. Based on the literature review, we identified the limitations of recent work in NTL detection.
3. These limitations and gaps identified in the literature review were addressed by the following contributions:
 - a. A total of 14 performance evaluation metrics were analyzed for NTL detection using different classifiers. We found that recall should be given higher priority for NTL detection, and random forest is the better algorithm for it having the highest recall.
 - b. We performed the testing of 15 machine learning classifiers belonging to 9 different types. The MLP classifier was found the best individual classifier with respect to recall, and ANN was found the best type of the classifiers for NTL detection.
 - c. We proposed a novel framework to identify relevant features for NTL detection by using the Incremental Feature Selection (IFS) algorithm, which identifies the most relevant features for NTL detection in a real dataset using feature importance. The results have shown that with the use of the IFS algorithm, recall and F-Measure of KNN is increased by 120% and 60%, respectively, while the training time of KNN is reduced by 90%.

13.2 Future Work

There is a need to further extend the use of performance evaluation metrics that can estimate and compare error rates on the basis of which a combination of classifiers can be selected for a specified dataset for NTL detection. Currently, there is a small range of graphical metrics used for performance analysis. This includes receiver operating characteristic (ROC) and area under ROC curve (AUC). There is also a need of further exploration in the usage of graphical performance metrics.

There is a need of a benchmark dataset which can widely be used in NTL detection. Due to the highly sensitive nature of the data, currently no such benchmark dataset is available. There is a possible future direction of the creation of a real as well as a synthetic dataset for the problem of NTL detection. Another related future direction can be a comparison of the results achieved in this study with the experimental results performed on a different dataset. The new dataset can be synthesized or a real dataset.

Another future direction is using penalized machine learning models in which weighted classifiers are used. The best classifiers identified in this study can also be implemented on different feature selection approaches.

Moreover, with the amount of success deep learning has been receiving in the last few years, there is also a need to test the performance of deep learning in NTL detection using selected feature set and all feature set. Moreover, deep learning can also be tested in feature selection for a real dataset. Over sampling (Chawla, 2002) and under sampling (Liu, 2008) techniques have shown good results in many problems. NTL detection can also be thoroughly tested with over sampling and under sampling techniques with a real dataset.

Power industries face a closely related problem of defaulters. The defaulters do not pay the bill for many months. The dataset used in NTL detection can also be used in the prediction of defaulters. Rather, an interesting fact finding research activity can be proceeded which finds the co-relation between the defaulters and the NTL fraudsters.

14. Bibliography

- Awais, M. a. (2016). *Physical activity classification using body-worn inertial sensors in a multi-sensor setup*. IEEE.
- Awais, M. a. (2017). A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors*, 1991.
- Bhat, R. R. (2016). Identifying nontechnical power loss via spatial and temporal deep learning. *Machine Learning and Applications (ICMLA)* (pp. 272--279). IEEE.
- Breiman, L. (2001). Random forests. *Machine learning*, 5--32.
- Chang, C.-C. a.-J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, 27:1--27:27.
- Chawla, N. V. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 321--357.
- Chen, T. a. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785--794.
- Ford, V. a. (2014). Smart grid energy fraud detection using artificial neural networks. *Computational Intelligence Applications in Smart Grid (CIASG)*, (pp. 1--6).
- Ghori, K. M. (2019). Performance analysis of different types of machine learning classifiers for non-technical loss detection. *IEEE Access* , 16033-16048.
- Ghori, K. M. (2020). Impact of Feature Selection on Non-technical Loss Detection. *6th Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 19-24). Riyadh: IEEE.
- Ghori, K. M. (2020). Performance analysis of machine learning classifiers for non-technical loss detection. *Journal of Ambient Intelligence and Humanized Computing*, 1-16.
- Goldberger, J. a. (2005). Neighbourhood components analysis. *Advances in neural information processing systems*, 513--520.
- Hussain, Z. a. (2016). Methods and techniques of electricity thieving in Pakistan. *Journal of Power and Energy Engineering*, 1--10.
- Ke, G. a.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 3146--3154.

- Liu, X.-Y. a.-H. (2008). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 539--550.
- Mingers, J. (1989). An Empirical Comparison of Pruning Methods for Decision Tree Induction. *Machine Learning*, 227--243.
- Prokhorenkova, L. a. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 6639--6649.
- Punmiya, R. a. (2019). Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing. *IEEE Transactions on Smart Grid*, 2326--2329.
- Raileanu, L. E. (2004). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 77--93.
- Raza, M. a.-M. (2019). Diagnosis and Monitoring of Alzheimer's Patients Using Classical and Deep Learning Techniques. *Expert Systems with Applications*.



Registry number: DEENK/254/2020.PL
Subject: PhD Publication List

Candidate: Khawaja Moyeezullah Ghori
Doctoral School: Doctoral School of Informatics
MTMT ID: 10066585

List of publications related to the dissertation

Foreign language scientific articles in international journals (2)

1. **Ghori, K. M.**, Abbasi, R. A., Awais, M., Imran, M., Ullah, A., Szathmáry, L.: Performance Analysis of Different Types of Machine Learning Classifiers for Non-Technical Loss Detection. *IEEE Access*. 8, 16033-16048, 2020. ISSN: 2169-3536.
DOI: <http://dx.doi.org/10.1109/ACCESS.2019.2962510>
IF: 3.745 (2019)
2. **Ghori, K. M.**, Imran, M., Nawaz, A., Abbasi, R. A., Ullah, A., Szathmáry, L.: Performance analysis of machine learning classifiers for non-technical loss detection. *Ambient Intell. Humaniz. Comput. [Epub]*, 1-16, 2020. ISSN: 1868-5137.
DOI: <https://doi.org/10.1007/s12652-019-01649-9>
IF: 4.594 (2019)

Foreign language conference proceedings (1)

3. **Ghori, K. M.**, Abbasi, R. A., Awais, M., Imran, M., Ullah, A., Szathmáry, L.: Impact of Feature Selection on Non-Technical Loss Detection. In: 6th Conference on Data Science and Machine Learning Applications, IEEE, Piscataway, 19-24, 2020. ISBN: 9781728127460





List of other publications

Foreign language scientific articles in international journals (1)

4. Rahman, S., Irfan, M., Raza, M., **Ghori, K. M.**, Yaqoob, S., Awais, M.: Performance Analysis of Boosting Classifiers in Recognizing Activities of Daily Living.
Int. J. Environ. Res. Public Health. 17 (3), 1-15, 2020. ISSN: 1661-7827.
DOI: <http://dx.doi.org/10.3390/ijerph17031082>
IF: 2.849 (2019)

Total IF of journals (all publications): 11,188

Total IF of journals (publications related to the dissertation): 8,339

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

1 September, 2020

