SCIENCE PROCEEDINGS SERIES

http://readersinsight.net/SPS

THEFT DETECTION IN POWER UTILITIES USING ENSEMBLE OF CHAID DECISION TREE ALGORITHM

Muhammad Salman Saeed *

School of Electrical Engineering
University Technology Malaysia, Skudai, Johor Bahru 81310
Malaysia
salman@live.utm.my

Mohd. Wazir Mustafa

School of Electrical Engineering
University Technology Malaysia, Skudai, Johor Bahru 81310
Malaysia
wazir@fke.utm.my

Usman Ullah Sheikh

School of Electrical Engineering
University Technology Malaysia, Skudai, Johor Bahru 81310
Malaysia
usman@fke.utm.my

Attaullah Khidrani

School of Electrical Engineering
University Technology Malaysia, Skudai, Johor Bahru 81310
Malaysia
khidraniatta@gmail.com

Mohd Norzali Haji Mohd

Faculty of Electrical and Electronics Engineering
Universiti Tun Hussein Onn Malaysia
Malaysia
norzali@uthm.edu.my

***Corresponding author's Email:** salman@live.utm.my*

Peer-review under responsibility of 4th Asia International Multidisciplinary Conference 2020 Scientific Committee
http://connectingasia.org/scientific-committee/

Office # 6, First Floor, A & K Plaza, Near D Watson, F-10 Markaz, Islamabad. Pakistan,
editor@readersinsight.net

SCIENCE PROCEEDINGS SERIES

http://readersinsight.net/SPS

Abstract

Modern civilizations are greatly relying on the electrical energy for the normal living. Energy consumption is steadily increasing for the emerging markets due to increased market participants economic wealth. However, large parts of the produced and distributed electricity are not paid and therefore do not add to the profit margin of the Power Distribution Companies (PDCs). Theft of electricity and fraud in energy consumption billing are few of the major issues faced by the power utilities (1–3).As a result of these fraudulent activities, a huge revenue of power utilities is lost each year (4). Power utilities in the underdeveloped companies still continue to use inefficient and time-consuming methods for electricity theft detection. (5). The current research work tries to address the electricity theft detection problem faced by power distribution companies by classifying the fraudster customers in a power delivery system. The main inspiration for this current research is to support the power utilities in the battle against these fraudulent activities. Firstly, the proposed method uses the monthly energy customer usage data acquired from the power distribution company of Pakistan to classify the fraudsters and genuine customers. The Bagged Chi - square Automatic Interaction Detection (CHAID) based decision tree algorithm is then utilized to segregate the fraudster and genuine consumers.Furthermore, based on the standard performance measuring metrics, the superiority of the Bagged CHAID-based NTL detection method is validated by comparing its efficacy with that of the well-known machine learning algorithms such as Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Network (ANN), Discriminant Analysis and Bayesian Network (BN).

Research Highlights

Trying to explore state-of-the-art machine learning algorithms for electricity theft detection.

Exploring the bagging algorithm has been used to enhances the classification ability of the proposed CHAID decision tree fraud detection model, as verified by the finding of current research.

The simulation results reveal that the proposed Bagged CHAID Decision Tree outranks all well known AI-based algorithms in terms of classification accuracy, precision specificity, sensitivity, AUC and F1 score.

Graphical Abstract

The figure shows the DT formed by the CHAID algorithm for the classification of honest and fraudster customers based on the difference between energy consumption.

SCIENCE PROCEEDINGS SERIES

http://readersinsight.net/SPS

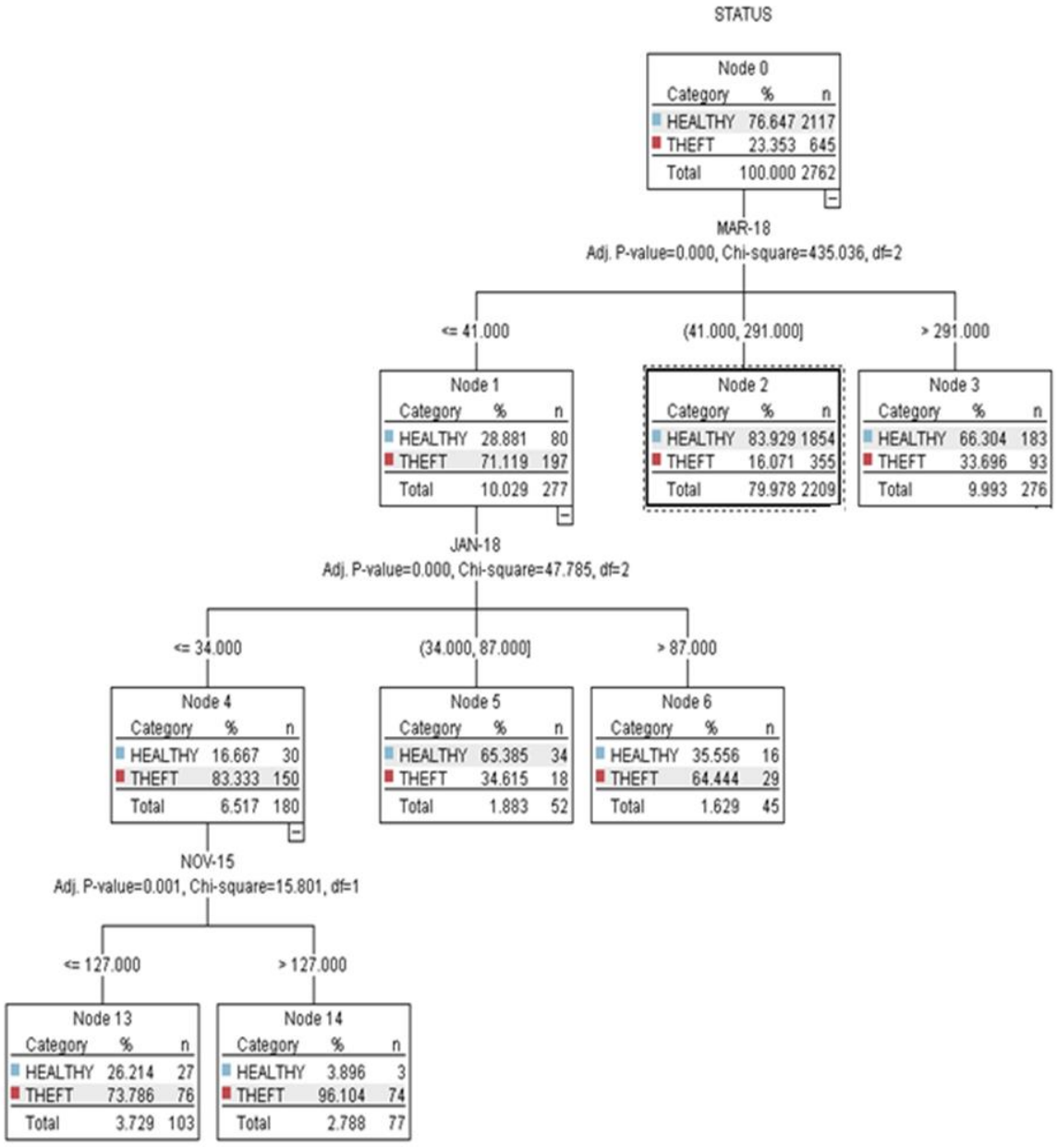

Fig. 1. CHAID Decision Tree algorithm for classification of honest and fraudster customers**.**

Research Objectives

This study proposes an Artificial-Intelligence based method called Bagged CHAID Decision tree algorithm for energy theft detection. The proposed scheme will generate a list of fraudulent customers which will help to effectively classify the fraudulent consumers. The Bagged CHAID Decision tree algorithm based classification approach makes use of MEPCO Multan, Pakistan's energy consumption data to classify the fraudster and honest customer. At the final

SCIENCE PROCEEDINGS SERIES
http://readersinsight.net/SPS

stage of the proposed method, shortlisted potential fraudsters customers must be inspected on-site to catch the perpetrators effectively

Methodology

The CHAID belongs to the group of the classification decision tree. CHAID Decision Tree (DT) can produce precise rules and has greater memory performance. Ensemble learning methods, on the other hand, train many machine learning algorithms to reach a final judgment (6). Ensemble Learning Systems (ELS) are inspired by human experience, which means that discovering and integrating multiple expert opinions can successfully solve any problem. The conclusion is taken based on those differing opinions. ELSs provides better performance by just using a single classifier. There are various ensemble algorithms with the bagging and boosting being the most common (7).

Results

The confusion matrix is usually used to assess the efficiency of machine learning algorithms that provide “True” output for all correctly classified data and "False" for all incorrectly classified datasets. It is evident that the bagging algorithm significantly improves the classification efficiency of the CHAID decision tree algorithm. The bagged CHAID algorithm's accuracy is 86.35 compared to a simple CHAID decision tree algorithm, which is about 84.142. The Neural Network model achieves 82.73 accuracy, which is significantly higher than Linear Regression, Support Vector Machine, Bayesian Network and Discriminant Analysis. The Receiver Operating Characteristic (ROC) curve is another significant performance assessor for a classifier. The Bagged CHAID Decision Tree achieves the highest AUC of 0.927 among all the studied methods..

Findings

This research offered a novel model for identifying NTL in Power utilies utilizing the Bagged CHAID DT algorithm, which is one of the most powerful classification algorithms. The efficacy of the proposed Bagged CHAID algorithm is compared with few well-known machine learning algorithms. The outcomes of this study show that the Bagged CHAID algorithm performs significantly better than the above-mentioned machine learning algorithms and attains an accuracy of 86.35 per cent and an AUC of 0.927, validating its efficiency supremacy.

References

1. Saeed MS, Mustafa MW Bin, Sheikh UU, Salisu S, Mohammed OO. Fraud Detection

SCIENCE PROCEEDINGS SERIES

http://readersinsight.net/SPS

for Metered Costumers in Power Distribution Companies Using C5.0 Decision Tree Algorithm. J Comput Theor Nanosci. 2020;17(2):1318–25.

2. Otuoze AO, Mustafa MW, Sofimieari IE, Dobi AM, Sule AH, Abioye AE, et al. Electricity theft detection framework based on universal prediction algorithm. Indones J Electr Eng Comput Sci. 2019;15(2):758–68.

3. Hussain S, Baloch SK, Jumani TA. A novel unsupervised feature-based approach for electricity theft detection using robust PCA and outlier removal clustering algorithm. 2020;(July):1–18.

4. Saeed MS, Mustafa MW, Hamadneh NN, Alshammari NA, Sheikh UU, Jumani TA, et al. Detection of Non-Technical Losses in Power Utilities — A Comprehensive Systematic Review. :1–25.

5. Otuoze AO, Mustafa MW, Mohammed OO, Saeed MS, Surajudeen-Bakinde NT, Salisu S. Electricity theft detection by sources of threats for smart city planning. IET Smart Cities. 2019;1(2):52–60.

6. Saeed MS, Mustafa MW, Sheikh UU, Jumani TA, Mirjat NH. Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan. Electron. 2019;8(8).

7. Salman Saeed M, Mustafa MW, Sheikh UU, Jumani TA, Khan I, Atawneh S, et al. An Efficient Boosted C5.0 Decision-Tree-Based Classification Approach for Detecting Non-Technical Losses in Power Utilities. Energies. 2020;13(12):3242.