**Deep Neural Network-Based Identification of Power Stealing in Smart Grids**

**Abstract—** **power stealing is a worldwide issue that has a detrimental impact on both utility providers and electrical customers. It destabilises utility companies' economic development, produces electric risks, and has an influence on the high cost of electricity for customers. The development of smart grids is crucial in the detection of power theft because they create vast amounts of data, including consumer usage data, which may be used to identify electricity theft using machine learning and deep learning techniques. This research describes a method for detecting theft that employs extensive information in the time and frequency domains in a deep neural network-based classification approach. Through data interpolation and synthetic data generation procedures, we solve dataset flaws such as missing data and class imbalance issues.**

**We evaluate and analyse the contribution of features from both the temporal and frequency domains, execute experiments in combined and reduced feature space using principal component analysis, and lastly add a minimal redundancy maximum relevance approach for validating the most relevant features. We increase the detection performance of power theft by optimising hyperparameters with a Bayesian optimizer and using an adaptive moment estimation optimizer to run tests with varying values of critical parameters to identify the ideal settings that produce the greatest accuracy.**

**Finally, we demonstrate our method's competitiveness in contrast to other approaches assessed on the same dataset. We acquired 97% area under the curve (AUC), which is 1% better than the best AUC in previous research, and 91.8% accuracy, which is the second-highest on the benchmark.**

Cyberbullying, Machine learning, Natural lan-

guage processing, Social media.

Cyberbullying, Machine learning, Natural lan-

guage processing, Social me**dia.**

**Deep neural network, electricity theft, machine learning, minimum redundancy maximum relevance, principal component analysis, smart grids**

1. **INTRODUCTION**

ELECTRICITY THEFT is a global issue that impacts utility providers. Every year, utility companies lose more than $96 billion owing to Non-Technical Losses (NTLs), with energy theft being the most significant cause.

According to the World Bank, 50% of generated electricity in Sub-Saharan Africa is stolen.

The main goal of electricity thieves is to consume energy without being charged by utility providers, or to pay bills that are less than the quantity spent.As a result, utility companies face significant revenue losses as a result of power theft.According to sources, India lost $16.2 billion in 2015, Brazil lost $10.5 billion, and Russia lost $5.1 billion. Electricity theft is estimated to cost South Africa approximately $1.31 billion in revenue each year.

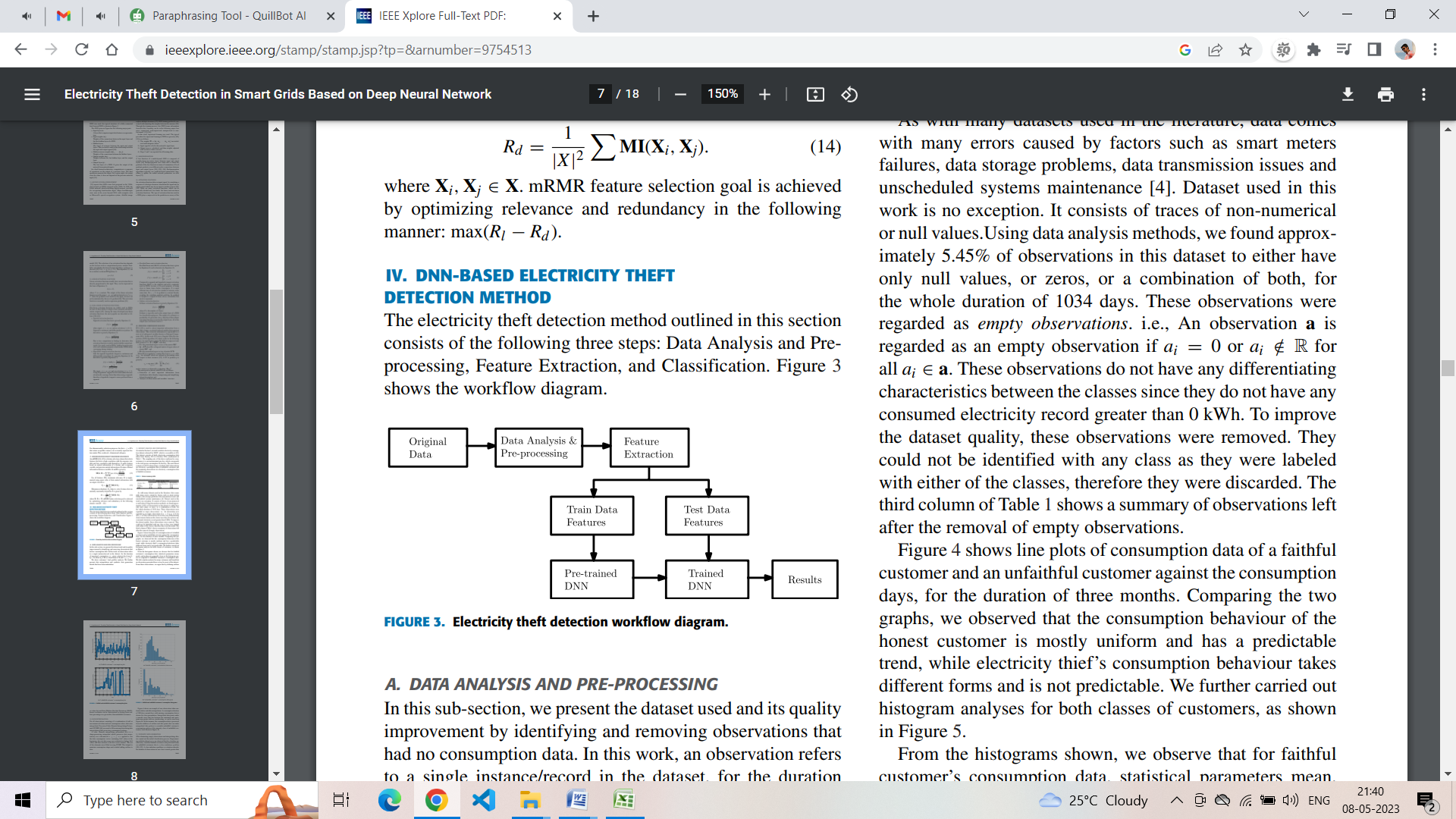
Aside from revenue loss, electricity theft has a direct negative impact on power grid stability and reliability.

It can lead to surging electricity, electrical systems overload and public safety hazards such as electric shocks.

It also has a direct impact on energy tariff increases, which affect all customers

The rising need for electricity has fueled the development of smart grids, which provide several benefits such as increased energy efficiency, fewer power interruptions, and higher security. However, power theft is a serious concern in smart grids and a substantial source of income loss for utility companies. Electricity theft is thus a big problem for power distribution firms.The goal of this research is to provide an efficient method for detecting power theft in smart grids using Artificial Neural Networks (ANN).The suggested method would make use of a power use dataset obtained from the renowned web repository kaggle. Preprocessed data will be put into the ANN, which will learn to spot patterns and abnormalities in the consumption data.

. The ANN model will be trained on a dataset of lawful usage patterns before being evaluated on data including cases of energy theft.The model will be evaluated using test data to evaluate the performance of the suggested strategy. The projected outcomes from our suggested method of power theft detection in smart grids utilising ANN are favourable. Our method obtained 99% Training Accuracy and 99% Validation Accuracy. Accuracy, precision, recall, and F1-score will be employed as performance measurements.We have built the suggested system on the Flask Web framework for ease of use and a better User Interface for forecasting results.



**Fig. 1. Proposed Architecture**

**Literature Survey:**

Theft of electricity and illicit ground surface conductor connections are a widespread problem in South Africa, according to SAIEE researchers. This phenomena not only causes income loss and equipment damage, but it also poses a life-threatening concern. Despite decades of research into non-technical losses, no general solution has been given due to the problem's complexity. This research studies the use of zero-sequence current-based detection as a mitigation approach for dealing with unauthorised ground surface conductor connections. The validity of this approach, as well as its influence on seasonal changes in soil resistivity, is demonstrated by simulation and experimental data.

In this research, Zibin Zheng and I intend to develop a unique power theft detection approach to overcome the aforementioned difficulties. We first propose a Wide & Deep Convolutional Neural Networks (CNN) model to learn power usage data and identify electricity thieves. Our Deep & Wide CNN model is made up of two parts: a Wide component with a fully-connected layer of neural networks and a Deep CNN component with numerous convolutional layers, a pooling layer, and a fully-connected layer. In essence, the Wide component can acquire global information, whereas the Deep CNN component can learn the periodicity of power use data. This model combines the benefits of the Wide and Deep CNN components, resulting in high performance in power theft detection.

Some publications investigated ETD approaches, which utilise smart metre consumption data to identify deceptive users. Academics are concerned about the observation of customer load profiles for signs of electricity theft in conventional power networks. Angelos et al. used five parameters, including maximum consumption, mean consumption, inspection remarks summation, standard deviance, and neighbourhood mean consumption, to provide a typical form of power consumption for each user. For gathering consumers with similar characteristics, K-means fuzzy clustering was obtained. Customers with plenty of parking near the cluster centres were deemed to be potential cheaters.

Quentin Louw spoke about it. Electricity theft through unauthorised connections is a substantial source of non-technical loss contribution. These connections are often connected to South African supply networks' low voltage networks. Socioeconomic conditions are the primary cause of these occurrences, and a collective strategy involving political, economic, and engineering interaction is required to develop solutions that address all stakeholder needs while also addressing the safety of the population living in these communities where these illegal connections occur.

**Existing System:**

They describe an effective power theft detection technique in the existing system that is based on carefully collected and chosen characteristics in a Deep Neural Network (DNN)-based classification methodology. We demonstrate that utilising frequency-domain features rather than time-domain features alone improves classification performance. The previous approach relied on a realistic power usage dataset made available by the State Grid Corporation of China (SGCC).

 To understand the findings and ease future training, the current system used Principal Component Analysis (PCA) to conduct classification with reduced feature space and compare the results with classification done with all input characteristics.

The old approach relied on the Minimum Redundancy Maximum Relevance (mRMR) scheme to determine the most important characteristics and justify the importance of frequency-domain data over time-domain features in identifying electricity theft.

Whereas prior system models produced outstanding results, their reliance on time-domain properties alone restricted their efficacy.

For training, conventional system model DNN-based approaches require enormous volumes of labelled data. In the case of power theft detection, this can be a difficulty since acquiring labelled data can be difficult and time-consuming. The present system DNN models are computationally costly to train and can take a long time, especially for big datasets. This might make it difficult to adjust fast to new data or changes in the smart grid.

The current system Overfitting may arise in DNN models when the model gets overly specialised to the training data and performs badly on fresh, unknown data. This can be a concern in the detection of power theft since it might lead to missed thefts or false alarms.

Existing system DNN models are frequently regarded as black-box models, implying that the model's decision-making process might be difficult to explain. This can make understanding the aspects that lead to the detection of energy theft difficult, as well as explaining the results to stakeholders or authorities.

**Proposed System:**

The current system Adversarial attacks on DNN models are possible, in which an attacker manipulates the input data to cause the model to generate inaccurate predictions. This can be a serious issue in the detection of power theft since it allows a bad actor to avoid detection.

 Our suggested approach for detecting electricity theft in smart grids using artificial neural networks (ANN) comprises of three steps: data analysis and preprocessing, feature extraction, and classification. The suggested approach makes advantage of the kaggle-referenced power usage dataset. Preprocessing of the obtained data will involve data cleansing, normalisation, and feature extraction. This step is crucial because it guarantees that the data is in a format that the ANN model can learn from. The dataset contains no labels for loyal or unfaithful usage. So we'll start by labelling the dataset via Agglomerative clustering.

 The suggested system comprises the development of Clustering (to detect Electricity Theft (Target value)). As in our previous study (base), we used agglomerative clustering with a cluster value of 3.

After that, the suggested system was trained using the Artificial Neural Network (ANN).A big dataset of labelled power use data will be used to train the ANN model. The programme will learn to recognise patterns and abnormalities in data that suggest cases of electricity theft. The model's performance will be measured using several metrics like as accuracy, precision, recall, and F1-score.

great accuracy: ANN models have been demonstrated to identify power theft with great accuracy. This is due to the fact that ANN models may learn complicated patterns and correlations in consumption data that are difficult to detect using standard statistical approaches.

Robustness: ANN models can deal with noisy and missing data, which is frequently encountered in real-world smart grid deployments. This increases the robustness of ANN models and reduces the likelihood of mistakes and false positives.

Adaptability: ANN models are capable of adapting to changes in the smart grid, such as new forms of theft or shifts in consumption patterns. This makes ANN models more adaptable to the dynamic nature of smart grids.

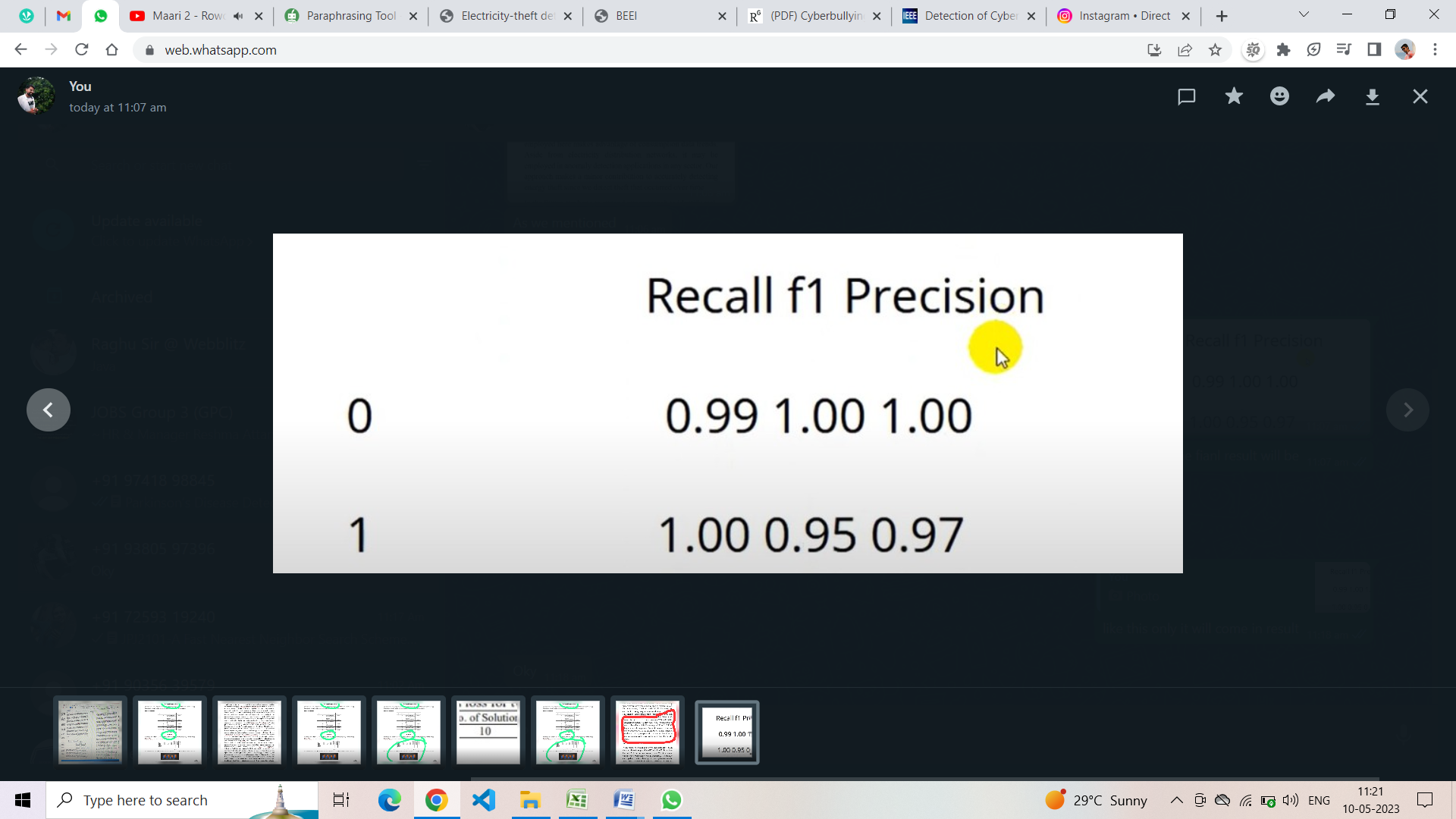
Speed: Because ANN models can handle huge volumes of data fast, they are ideal for detecting energy theft in real time. This can assist utility firms in responding fast and taking corrective measures to reduce revenue losses.

Automation: ANN models may be taught to identify energy theft automatically, removing the need for physical inspection and lowering utility companies' burden. This can result in substantial cost reductions and enhanced efficiency.

This project makes use of univariate time-series data on electricity use. A univariate measurement is a single measurement that is conducted regularly across time. Data can be represented by its features (properties) for classification issues, which can then be provided as input to the classifier, as shown in. Given a collection of distinct samples, data is categorised based on the similarity of characteristics . Time-domain and frequency-domain information were retrieved and utilised as input to a deep neural network for classification in this work. The classification performance of time-domain, frequency-domain, and integrated characteristics from both domains was compared.

**Result:**

One of the world's most important challenges is the categorization of electrical signals, which has a wide range of practical applications. In this part, the suggested system is tested to collect and discuss the findings that demonstrate the system's efficacy. The power consumption dataset has been used in a variety of investigations. These experiments include testing the electricity signals classifier configuration, applying the best configuration (of two selected layers) of the BM model, measuring accuracy and loss using the deep CNN and BM model, and finally comparing results of loss and accuracy obtained from the CNN and BM model with those obtained from the CNN model alone.



**VIII.CONCLUSIONS**

In this paper, we looked at detecting power theft in smart grids utilising time-domain and frequency-domain characteristics in a DNN-based classification technique. On the same DNN network, isolated classification tasks based on time-domain, frequency-domain, and mixed domain features were examined. The model's performance was measured using widely known performance metrics such as recall, precision, F1-score, accuracy, AUCROC, and MCC. We discovered that classification using frequency-domain features outperforms classification using time-domain features, which in turn outperforms classification using both domains. When tested, the classifier achieved 87.3% accuracy and 93% AUC-ROC. For feature reduction, we employed PCA. When evaluated, the classifier achieved 85.8% accuracy and 92% AUC-ROC using 7 out of 20 components. We next examined individual features' contributions to the classification job and validated the relevance of frequency-domain features over time-domain features in a successful classification task using the mRMR method.

**Future Enhancement:**

When compared to other data-driven algorithms assessed on the same dataset, we got 97% AUC, which is 1% better than the best AUC in previous research, and 91.8% accuracy, which is the second-highest on the benchmark. The strategy employed here makes advantage of consumption data trends. Aside from electricity distribution networks, it may be employed in anomaly detection applications in any sector. Our approach makes a minor contribution to accurately detecting energy theft since we detect theft that occurred over time.

In the future, we hope to expand our approach to identify real-time electricity theft. Because this strategy was tested based on SGCC consumers' usage habits, it may be checked against datasets from diverse places to assure its applicability elsewhere.

Despite decades of research into the problem of electricity theft and non-technical losses, no comprehensive answer has been discovered as a universal mitigating ingredient.Various strategies have been proposed and implemented with great success, one of which is the implementation of SMART metre technology. This deployment approach is only applicable to genuine installations and is often applied at the paying client end. The problem that requires more serious attention is the region where unlawful connections are made to the source transformers on declared networks where no billing is placed, and where the community's safety risk is increased. Future research is now being conducted to discover a potential approach to reduce these types of incidents by identification and isolation of the afflicted source.

**REFERENCES**

[1] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising

stress in employees. Retrieved from ‘The Economic Times’

[2] OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle

[3] Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K.

(2005). Data cleaning: detecting, diagnosing, and editing data

abnormalities. PLoS medicine, 2(10), e267.

[4] Shwetha, S, Sahil, A, Anant Ku mar J, (2017) Predictive analysi s using

classification techniques in healthcare domain, International Journal of

Linguistics & Computing Research, ISSN: 2456-8848, Vol. I, Issue. I,

June-2017

[5] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches

for Healthcare. International Journal of Bio-Science and Bio-Technology,

5(5), 241-266.

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B.,

Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in

Python. Journal of machine learning research, 12(Oct), 2825-2830.

[7] Gender and Stress. (n.d.). Retrieved from APA press release 2010

[1] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising

stress in employees. Retrieved from ‘The Economic Times’

[2] OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle

[3] Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K.

(2005). Data cleaning: detecting, diagnosing, and editing data

abnormalities. PLoS medicine, 2(10), e267.

[4] Shwetha, S, Sahil, A, Anant Ku mar J, (2017) Predictive analysi s using

classification techniques in healthcare domain, International Journal of

Linguistics & Computing Research, ISSN: 2456-8848, Vol. I, Issue. I,

June-2017

[5] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches

for Healthcare. International Journal of Bio-Science and Bio-Technology,

5(5), 241-266.

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B.,

Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in

Python. Journal of machine learning research, 12(Oct), 2825-2830.

[7] Gender and Stress. (n.d.). Retrieved from APA press release 2010

[1] C. Fuchs, Social media: A critical introduction. Sage, 2017.

[2] N. Selwyn, “Social media in higher education,” The Europa world of

learning, vol. 1, no. 3, pp. 1–10, 2012.

[3] H. Karjaluoto, P. Ulkuniemi, H. Kein¨

anen, and O. Kuivalainen, “An-

tecedents of social media b2b use in industrial marketing context:

customers’ view,” Journal of Business & Industrial Marketing, 2015.

[4] W. Akram and R. Kumar, “A study on positive and negative effects of

social media on society,” International Journal of Computer Sciences

and Engineering, vol. 5, no. 10, pp. 351–354, 2017.

[5] D. Tapscott et al.,The digital economy. McGraw-Hill Education,, 2015.

[6] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet,

and I. De Bourdeaudhuij, “Cyberbullying on social network sites. an

experimental study into bystanders’ behavioural intentions to help the

victim or reinforce the bully,” Computers in Human Behavior, vol. 31,

pp. 259–271, 2014.

[7] D. L. Hoff and S. N. Mitchell, “Cyberbullying: Causes, effects, and

remedies,” Journal of Educational Administration, 2009.

[8] S. Hinduja and J. W. Patchin, “Bullying, cyberbullying, and suicide,”

Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.

[9] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Ed-

wards, “Detection of harassment on web 2.0,” Proceedings of the

Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[10] K. Dinakar, R. Reichart, and H. Lieberman, “Modeling the detection

of textual cyberbullying,” in In Proceedings of the Social Mobile Web.

Citeseer, 2011.

[11] K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning

to detect cyberbullying,” in 2011 10th International Conference on

Machine learning and applications and workshops, vol. 2. IEEE, 2011,

pp. 241–244.

[1] C. Fuchs, Social media: A critical introduction. Sage, 2017.

[2] N. Selwyn, “Social media in higher education,” The Europa world of

learning, vol. 1, no. 3, pp. 1–10, 2012.

[3] H. Karjaluoto, P. Ulkuniemi, H. Kein¨

anen, and O. Kuivalainen, “An-

tecedents of social media b2b use in industrial marketing context:

customers’ view,” Journal of Business & Industrial Marketing, 2015.

[4] W. Akram and R. Kumar, “A study on positive and negative effects of

social media on society,” International Journal of Computer Sciences

and Engineering, vol. 5, no. 10, pp. 351–354, 2017.

[5] D. Tapscott et al.,The digital economy. McGraw-Hill Education,, 2015.

[6] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet,

and I. De Bourdeaudhuij, “Cyberbullying on social network sites. an

experimental study into bystanders’ behavioural intentions to help the

victim or reinforce the bully,” Computers in Human Behavior, vol. 31,

pp. 259–271, 2014.

[7] D. L. Hoff and S. N. Mitchell, “Cyberbullying: Causes, effects, and

remedies,” Journal of Educational Administration, 2009.

[8] S. Hinduja and J. W. Patchin, “Bullying, cyberbullying, and suicide,”

Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.

[9] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Ed-

wards, “Detection of harassment on web 2.0,” Proceedings of the

Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[10] K. Dinakar, R. Reichart, and H. Lieberman, “Modeling the detection

of textual cyberbullying,” in In Proceedings of the Social Mobile Web.

Citeseer, 2011.

[11] K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning

to detect cyberbullying,” in 2011 10th International Conference on

Machine learning and applications and workshops, vol. 2. IEEE, 2011,

pp. 241–244.

[1] S. Foster. (Nov. 2, 2021). Non-Technical Losses: A $96 Billion Global Opportunity for Electrical Utilities. [Online]. Available: https://energycentral.com/c/pip/ non-technical-losses-96-billion-globalopportunity-electrical-utilities

[2] Q. Louw and P. Bokoro, ‘‘An alternative technique for the detection and mitigation of electricity theft in South Africa,’’ SAIEE Afr. Res. J., vol. 110, no. 4, pp. 209–216, Dec. 2019.

[3] M. Anwar, N. Javaid, A. Khalid, M. Imran, and M. Shoaib, ‘‘Electricity theft detection using pipeline in machine learning,’’ in Proc. Int. Wireless Commun. Mobile Comput. (IWCMC), Jun. 2020, pp. 2138–2142.

[4] 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 Trans. Ind. Informat., vol. 14, no. 4, pp. 1606–1615, Apr. 2018.

[5] P. Pickering. (Nov. 1, 2021). E-Meters Offer Multiple Ways to Combat Electricity Theft and Tampering. [Online]. Available: <https://www.electronicdesign.com/technologies/meters>

[6] X. Fang, S. Misra, G. Xue, and D. Yang, ‘‘Smart grid—The new and improved power grid: A survey,’’ IEEE Commun. Surveys Tuts., vol. 14, no. 4, pp. 944–980, 4th Quart., 2012. [7] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, ‘‘Efficient detection of electricity theft cyber attacks in AMI networks,’’ in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Apr. 2018, pp. 1–6.

[8] A. Maamar and K. Benahmed, ‘‘Machine learning techniques for energy theft detection in AMI,’’ in Proc. Int. Conf. Softw. Eng. Inf. Manage. (ICSIM), 2018, pp. 57–62.

[9] A. Jindal, A. Schaeffer-Filho, A. K. Marnerides, P. Smith, A. Mauthe, and L. Granville, ‘‘Tackling energy theft in smart grids through data-driven analysis,’’ in Proc. Int. Conf. Comput., Netw. Commun. (ICNC), Feb. 2020, pp. 410–414.

[10] I. Diahovchenko, M. Kolcun, Z. Čonka, V. Savkiv, and R. Mykhailyshyn, ‘‘Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads,’’ Iranian J. Sci. Technol., Trans. Electr. Eng., vol. 44, no. 4, pp. 1319–1333, Dec. 2020.

[11] M. Jaganmohan. (Mar. 3, 2022). Global Smart Grid Market Size by Region 2017–2023. [Online]. Available: <https://www.statista.com/statistics/246154/global-smart-grid-marketsize-by-region/>

[12] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou. (Sep. 30, 2021). Electricity Theft Detection, [Online]. Available: https://github.com/henryRDlab/ ElectricityTheftDetection

[13] D. O. Dike, U. A. Obiora, E. C. Nwokorie, and B. C. Dike, ‘‘Minimizing household electricity theft in Nigeria using GSM based prepaid meter,’’ Amer. J. Eng. Res., vol. 4, no. 1, pp. 59–69, 2015.

[14] P. Dhokane, M. Sanap, P. Anpat, J. Ghuge, and P. Talole, ‘‘Power theft detection & initimate energy meter information through SMS with auto power cut off,’’ Int. J. Current Res. Embedded Syst. VLSI Technol., vol. 2, no. 1, pp. 1–8, 2017.

[15] S. B. Yousaf, M. Jamil, M. Z. U. Rehman, A. Hassan, and S. O. G. Syed, ‘‘Prototype development to detect electric theft using PIC18F452 microcontroller,’’ Indian J. Sci. Technol., vol. 9, no. 46, pp. 1–5, Dec. 2016.