IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 15, May 2023

Impact of Machine Learning on Electricity Theft Detection

Prof. Pankaj Phadtare[1], Priti Chavan[2], Jay Kadam[3] , Rugved Shinde[4] , Velankanni Yadavar[5]

Department of Computer Engineering[1,2,3,4,5]

Trinity College of Engineering, Pune, India

Abstract: *The principal source of electrical power loss that has a substantial impact on both the quantity and quality of electrical power is electricity theft. However, the approaches currently in use for detecting this theft-related criminal activity are varied and complex since it is difficult to extract useful information from time-series data due to the uneven nature of the dataset. This research develops a novel approach for detecting electricity theft by combining three algorithms into a pipeline. The suggested approach first balances the dataset using the synthetic minority oversampling technique (SMOTE), then integrates kernel function and principal component analysis (KPCA) to extract features from highly dimensional time-series data, and uses support vector machines (SVM) to classify the data. Additionally, the effectiveness of the system.*

Keywords: ESP826612E, ATmega328, Thingsspeak IOT, Energy Meter, Display.

I. INTRODUCTION

Modern existence requires electricity as a basic necessity. It powers electric machinery and appliances and is used for lighting, cooling, and heating. Electricity has changed modern medical and surgical procedures, entertainment, communication, and transportation to comfort people. A variety of steps are being done to make power capable of meeting the demands as demand and usage grow daily [1]. However, the largest danger to the electrical management system continues to be power outages.

An intelligent energy system and smart grid have been created in response to the requirement to minimise power losses and maximise the usage of electricity (SG). Advanced metering infrastructure (AMI) is the foundation of SG [2]. The energy system now uses smart metres (SM), which have been replaced by AMI.

Transmission, distribution, and consumption are possible points of energy loss in the power system [1], [8]. Technical losses (TLs) and commercial losses, commonly referred to as non-technical losses (NTLs) [1, [5,] [9], are the two groups into which these losses are split. The energy loss in the conductors, distribution lines, and transmission lines is what causes TLs. NTLs can occur for a variety of reasons, including poor installation, broken metres, billing problems, tampering with the metres, hacking into smart metres, manipulating the data, direct hooking on other homes, etc. [5], [10]. The power used by customers but not billed by the utility is known as NTLs, according to utilities [11].

Energy system damage from NTLs is considerable. Customers' dishonest behaviour is to blame for the economy's troubles. The primary reason for the power utility's revenue loss is NTL [8]. It also has an impact on supply quality. NTLs are also to blame for the rise in energy prices that impacts all consumers. Because when tariffs are calculated, such losses are distributed among all consumers [12]. Additionally, it makes the electricity grid less stable and reliable [11]. The globe loses $89.3 billion annually to electricity theft, claims Northeast Group LLC [13]. NTLs are a serious problem for rich countries as well as underdeveloped countries.

II. PROBLEM STATEMENT

India loses more money to theft than any other country in the world. In this proposed system we use dataset having electricity usage of a smart grid (SG) meter (or simply smart meter). Using this dataset we does feature selection and preprocessing on dataset. When we have large number of features in dataset then feature selection is very important part in our Machine Learning. As we use feature selection it gives us most important feature and this feature selection gives

DOI: 10.48175/568

us more accuracy. Then we perform the preprocessing on that data. After that we use the superiority of OCR and sarimax over other ML algorithms To detect the theft.

III. OBJECTIVES

Prevention: The primary objective is to implement measures that discourage and deter electric theft. This may involve improving the physical security of electrical infrastructure, such as substations and distribution networks, to make unauthorized access difficult. Implementing smart metering systems and advanced metering in- frastructure can also help detect tampering and prevent unauthorized consumption.

Detection: The objective is to promptly identify instances of electric theft. This includes using advanced metering technologies and analytics to detect abnormal patterns of electricity usage that may indicate theft. Automated monitoring systems and data analysis techniques can assist in flagging suspicious activities, such as meter tampering or energy diversion.

Investigation: Once electric theft is detected, the objective is to conduct thorough investigations to gather evidence, identify the culprits, and establish legal cases against them. This involves collaboration between electricity utility companies, law enforcement agencies, and regulatory authorities to ensure effective prosecution and deterrence.

Legal Enforcement: The objective is to enforce laws and regulations related to electric theft, ensuring appropriate penalties and legal consequences for offenders. This may involve close cooperation with law enforcement agencies and judicial systems to expedite legal proceedings and increase the likelihood of convictions.

Public Awareness and Education: Educating the public about the dangers, consequences, and ethical implications of electric theft is an important objective. Raising awareness about the impacts of electric theft on society, the economy, and the reliability of electrical systems can help in reducing the occurrence of such activities.

IV. SYSTEM ARCHITECTURE

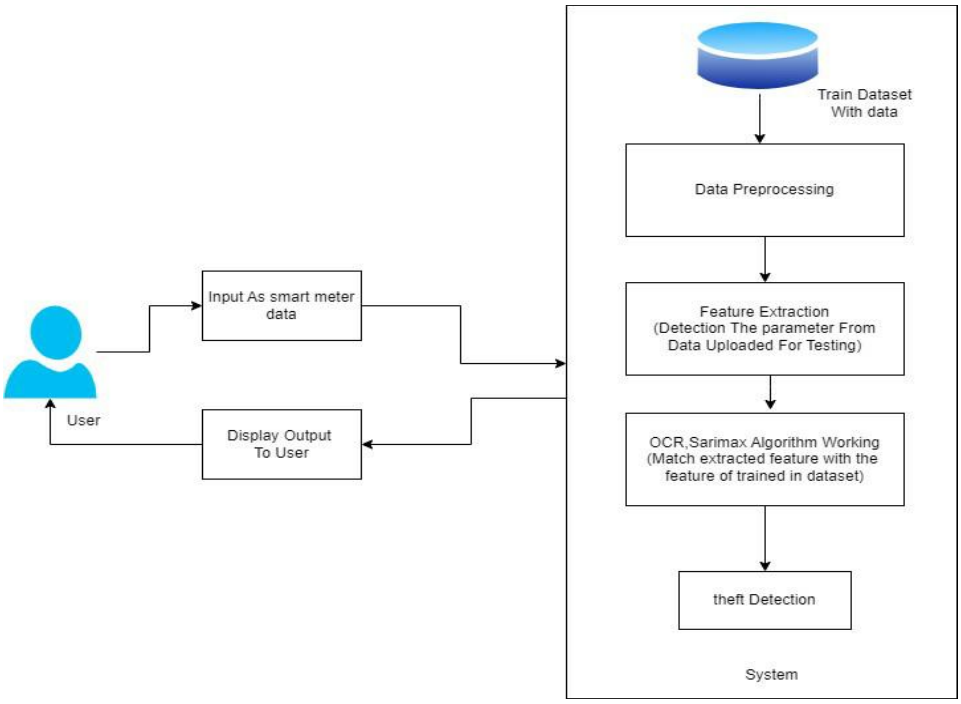

Train Dataset
With data
Data Preprocessing
Feature Extraction
(Detection The parameter From
Data Uploaded For Testing)
OCR,Sarimax Algorithm Working
(Match extracted feature with the
feature of trained in dataset)
theft Detection
System
Input As smart meter
data
Display Output
To User
User

Figure1: system Architecture

V. ALGORITHM

Optical character recognition (OCR) : Optical character recognition (OCR) algorithms allow computers to analyze printed or handwritten documents automatically and prepare text data into editable formats for computers to efficiently process them. It is another way to extract and leverage business-critical data.

SARIMAX -SARIMAX(Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors): SARIMAX -SARIMAX(Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors) is an updated version of the ARIMA model. ARIMA includes an autoregressive integrated moving average, while SARIMAX includes seasonal effects and eXogenous factors with the autoregressive and moving average component in the model.

VI. SOFTWARE REQUIREMENT SPECIFICATION

6.1 Functional Requirements

6.1.1 System Feature1 (Functional Requirement)

a) Feature point extraction: Feature points of each Dataset parameters gets detected.

b) Feature correspondence matching: Matching of selected feature points across various parameters.

6.1.2 System Feature1 (Functional Requirement)

In system we have used SARIMAX and OCR algorithm.

6.2 Non Functional Requirements

6.2.1 Performance Requirements

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work evidently. Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement. The application is designed in modules where errors can be detected. This makes it easier to install and update new functionality if required.

6.2.2 Safety Requirement

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

6.2.3 Software Quality

Attributes Our software has many quality attribute that are given below:-

- Adaptability: This software is adaptable by all users.
- Availability: This software is freely available to all users. The availability of the software is easy for everyone. college name, Department of Computer Engineering 2022-23 17
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
- Reliability: The performance of the software is better which will increase the reliability of the Software.
- User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior. Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled. Security: Users are authenticated using many security phases so reliable security is provided.
- Testability: The software will be tested considering all the aspects.

6.3 Constraints

User Interface Constraints - Desktop Application of System.

Hardware Constraints - In system, Laptop is 64 Bit and RAM is 8 GB.

Software Constraints - In system, used Anaconda Navigator Software and IDE Spyder. Programming Language is used Python.

Assumptions and dependencies

Assumption - We have to assume that the dataset which is going to use accurate. Input as dataset of Electricity.

Dependencies - Dependence on Python language and Machine Learning Technique. Used Python Library is Pandas Numpy, Tkinter, Pillow Etc.

DOI: 10.48175/568

6.4 Hardware Requirements

- RAM : 8 GB
- Hard Disk :500 GB
- Processor : Intel i5 Processor

6.5 Software Requirements

- IDE : Spyder Coding
- Language : Python Version 3.5
- Operating System : Windows 10 and Above

6.6 Interfaces

6.6.1 User Interface

- Application Based Cricket Shot Detection.

6.6.2 Hardware Interfaces:

- RAM : 8 GB
- Hard Disk :500 GB
- Processor : Intel i5 Processor

6.6.3 Software Interfaces :

- IDE : Spyder
- Coding Language : Python Version 3.5
- Operating System : Windows 10 and Above

VII. MODELING AND DESIGN

ER DIAGRAM:-

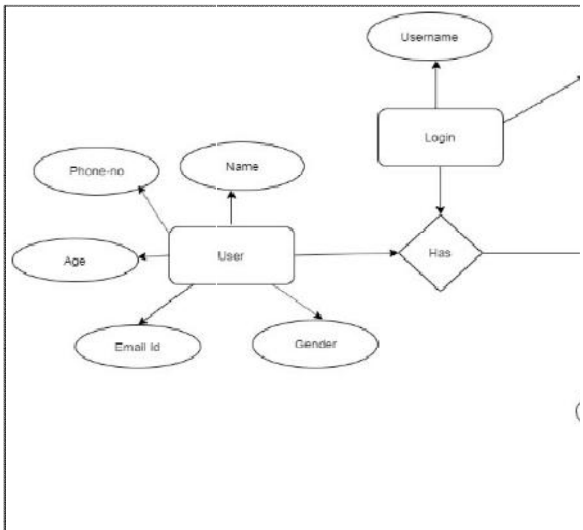


Fig: ER Diagram

USE CASE DIAGRAM:-

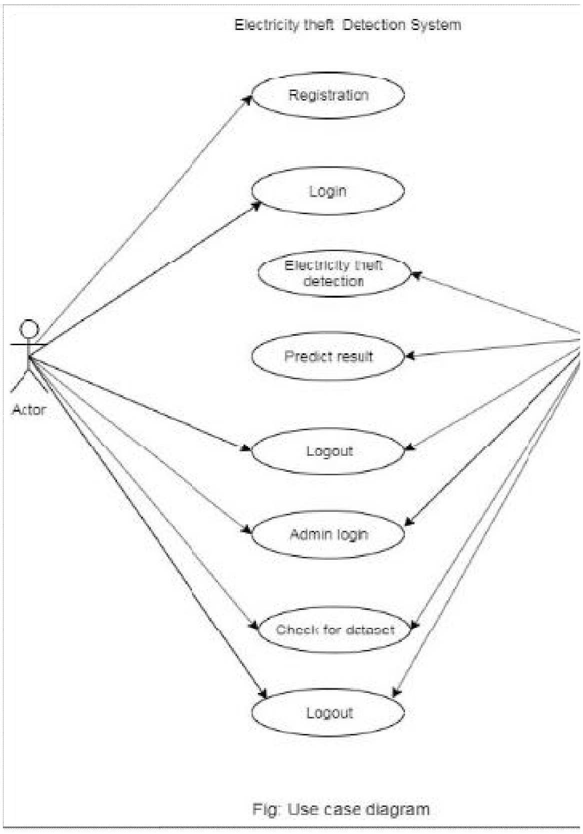
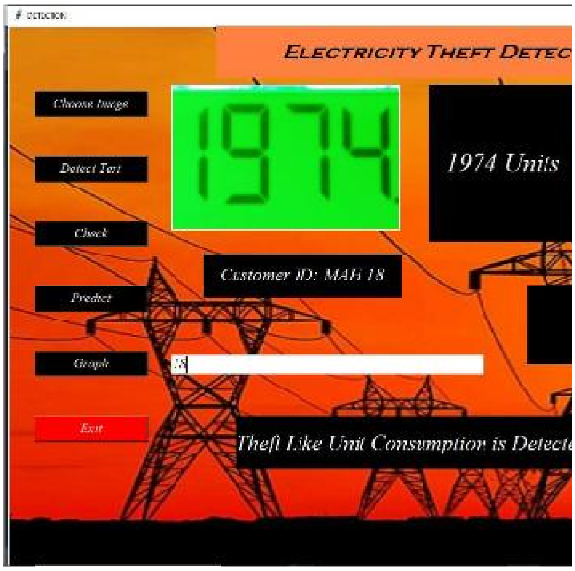


Fig: Use case diagram

VIII. RESULTS



1974 Units

Customer ID: MAH 18

Theft Like Unit Consumption is Detecte

FIGURE:- RESULT

IX. CONCLUSION

This proposed system detects the electricity theft using OCR and SARIMAX machine learning method. However, SARIMAX appeared to be the fastest classifier. This proposed system helps to electricity utilities to detect electricity theft and they will not have to bare loss. This is most important application of this project.

X. FUTURE SCOPE

In future Scope, The experimental results indicated that the adaptive TSRNN architecture in fusion with the SMOTE balancing technique is feasible to extract data features for monitoring the abnormal electricity theft behavior. The methodology framework is prospectively promoted to be used for online monitoring on big data analysis for a large scale of electricity power consumption.

REFERENCES

[1]. J. Nagi, K. Yap, S. Tiong, S. Ahmed, and M. Mohamad, “Nontechnical loss detection for metered customers in power utility using support vector machines,” IEEE Transactions on Power Delivery, vol. 25, no. 2, pp. 1162–1171, 2010, cited By 104.

[2]. S. McLaughlin, D. Podkuiko, and P. McDaniel, “Energy theft in the advanced metering infrastructure,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6027 LNCS, pp. 176–187, 2010, cited By 99.

[3]. G. Tsekouras, N. Hatziargyriou, and E. Dialynas, “Twostage pattern recognition of load curves for classification of electricity customers,” IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 1120–1128, 2007, cited By 122. [Online].

[4]. Y. Zhang, W. Chen, and J. Black, “Anomaly detection in premise energy consumption data,” 2011, cited By 12. [Online].

[5]. S. Dua and X. Du, Data Mining and Machine Learning in Cybersecurity, 1st ed. Boston, MA, USA: Auerbach Publications, 2011.

[6]. V. Barnett and T. Lewis, Outliers in Statistical Data, ser. Wiley Series in Probability Statistics. Wiley, 1994. [Online].

[7]. Available: https://books.google.com.pr/books?id=B44QAQAAIAAJ

[8]. N. Billor, A. Hadi, and P. Velleman, “Bacon: Blocked adaptive computationally efficient outlier nominators,” Computational Statistics and Data Analysis, vol. 34, no. 3, pp. 279–298, 2000, cited By 154.

[9]. E. Eskin, “Anomaly detection over noisy data using learned probability distributions,” in Proceedings of the Seventeenth International Conference on Machine Learning, ser. ICML ’00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 255–262.

[10]. E. M. Knorr and R. T. Ng, “Algorithms for mining distance-based outliers in large datasets,” in Proceedings of the 24rd International Conference on Very Large Data Bases, ser. VLDB ’98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 392–403.

[11]. C. Aggarwal and P. Yu, “Outlier detection for high dimensional data,” 2001, pp. 37–46, cited By 433

DOI: 10.48175/568