International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105318 |

A Statistical Framework for Detecting Electricity Theft Activities in Smart Grid Distribution Networks

Chaitanya Karnakar, Krishna Patil, Yash Mulik, Aditya Karkhele

UG Student, Dept. of Computer Science, Sinhgad Institute of Technology and Science, Narhe, Pune, India

Dr.Geeta Navale

Dept. of Computer Science, Sinhgad Institute of Technology and Science, Narhe, Pune, India

ABSTRACT: Power robbery and electricity losses is one of the serious issues of electric utilities. Such power robbery produce monetary misfortune to the service organizations. Agricultural nations, particularly India, are dealing with a major issue of force misfortunes because of Electricity burglary and unlawful obstruction in electrical meters. In developed and under developed countries this point is normal to such an extent that there is generally conversation going on it, cause of electricity losses includes tempering meter to show a low meter, by-passing of electric meter, bills not paid consistently. Abnormalities in billing includes incorrect meter readings taken by administration men, unlawful fixing of bill in office by the workplace employees. In the system a smart meter is installed at every user side and the main server is maintained at the admin side, we are using OCR(Optical Character Recognition) algorithm for image processing to cover smart meter image into digital format and SERIMAX classifier to deal with seasonal dataset to detect electricity losses. In this proposed system, we are detecting electricity losses by taking smart meter reading image as input using OCR(Optical Character Recognition) and SERIMAX algorithms. This model provides an accuracy of 95% compared to previously proposed models which provide much lower accuracy

KEYWORDS - OCR, SARIMAX, Image Processing, ML

I INTRODUCTION

Electricity losses has developed as a severe issue with in electricity sector, particularly in the developing countries. Electricity theft costs the government a lot of money. In certain nations, the situation is so bad that governments are losing money instead of making money. In some circumstances, the government must pay subsidies to the power sector in able to preserve electricity prices affordable[1]. financial loss, there are fewer funds available for investments to increase traditional power capacity, and administrations are still unable to meet the ever-increasing demand for electricity. In those other situations, the problem has gotten so bad that the afflicted electrical systems are on the verge of passing bankrupt. However, power theft is so frequent in developing and under developed countries that it is also kept out of the publicspotlight[4]. Tampering with metres to low metre reading, stealing electricity via bypassing a meter, billing inconsistencies, and unpaid bills all are symptoms of electricity losses. Billing irregularities include erroneous reading collected by bribed service personnel and purposeful bill manipulation by office personnel in exchange for illicit payment from the customer. Various nontechnical and technical approaches for detecting electrical pilfering have been proposed in the past. Non-technical methods may include inspecting clients with suspect load profiles, in whose case we develop a supervised machine learning based theft detection model that determines whether an anomalous usage pattern has happened in the smart meter or not by taking smart meter reading image as a input.[5] In this proposed system, we are implementing electricity losses detection model using OCR and SARIMAX algorithm. This model provides an accuracy of 97%compared to previously proposed models which provide much lower accuracy.

II PROBLEM DEFINITION

To detect electricity losses by giving smart meter reading image as input using machine learning techniques like OCR and SARIMAX methods and to get maximum accuracy within minimum training time

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105318 |

III INDENTATIONS AND EQUATIONS

Agricultural nations, particularly India, are dealing with a major issue of power misfortunes because of electricity losses and unlawful obstruction in electrical meters. The motivation here is to create web app to detect electricity losses using smart meter reading image as input. The user can use the system to detect electricity losses. Detection of electricity losses is very difficult work to do .But our system uses such algorithm and the techniques which give maximum accuracy with minimum training time. So it becomes very easy to user to get respective output. The system provides an improved mechanism to detect electricity losses using image processing.OCR and SERIMAX algorithm works and how using some classifiers, feature extracting techniques ,python features and various python libraries which will provide more accurate results than previously implemented electricity detection model[2].

IV PROPOSED ALGORITHM

OCR

Optical character recognition (OCR) is a technology that facilitates the translation of many forms of texts or images into analyzable, editable, and searchable data. During the last decade, researchers have employed artificial intelligence/machine learning techniques to automatically evaluate handwritten and printed materials in order to convert them to electronic format. An OCR system is primarily dependent on feature extraction and discrimination/classification of these characteristics. Handwritten OCR is gaining popularity as a subset of OCR. Based on this, it is further divided into offline systems and online systems. data input The offline system is a static system with input data in the form of scanned photographs, but the type of input in the online system is more dynamic and is based on the movement of a pen tip with a certain velocity, projection angle, location, and locus point. As a result, an online system is deemed more complicated and advanced than an offline system since it overcomes the overlapping problem of input data that exists in the offline system.

OCR detects the reading from the image and converts it into machine-readable format.

Figure 2: Extracted Meter Reading

SARIMAX Algorithm

The ARIMA model (Seasonal Auto-Regressive Integrated Moving Average) has been updated. SARIMAX, as SARIMA and Auto ARIMA, is an occasional comparable model. It's additionally equipped for managing outside impacts. This part of the model separates it from others. To foresee, System can anticipate() or gauge() techniques for SARIMAX on the item returned by fitting the information.[3].Underneath we use anticipate() and give the beginning and end, alongside the exog variable based on which the forecasts will be made. We can likewise utilize figure() and give step furthermore, exogparameters.In our framework we are utilizing occasional dataset for expectation system got great precision by utilizing the serimax model[9]. Advantage:- It is having high adaptability. Optical person acknowledgment (OCR) calculations permit PCs to investigate printed or on the other hand written by hand records consequently and get ready text information into editable arrangements for PCs to handle them productively. It is one more method for extricating and influence business-basic information[10]. OCR is short for optical person acknowledgment. ... During OCR checking, a calculation perceives characters from printed sources and converts them into computerized design. Whenever this is done, the computerized design is effectively accessible also, editable.[6] Text acknowledgment There are two essential sorts of center OCR calculation, which may favorable to duce a positioned rundown of competitor characters. Lattice matching includes contrasting a picture with a put away glyph on a pixel-by-pixel premise; it is otherwise called

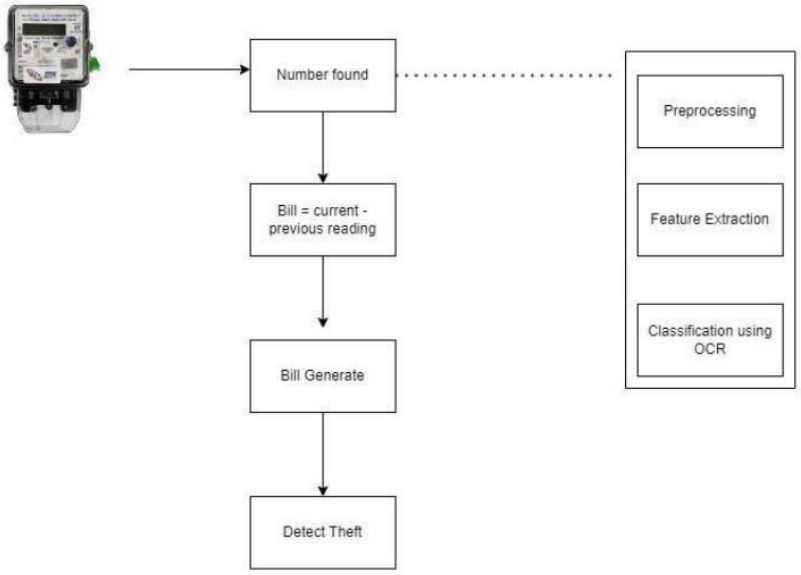
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105318 |

"design coordinating", "design acknowledgment", or "picture connection"[8]. Benefits of OCR :- Cheaper than paying somebody to enter enormous sums physically of text and a lot quicker than somebody physically entering a lot of tex output

In the system architechure the input would be an smart meter reading image in any format link .jpg , .png and etc. For classification the data set will go through feature extraction, model training and will give a trained classifier. Now the uploaded image by the user will go through feature extraction and process it through trained classifier and give the prediction as output.Once the number is found the bill will be generated and in detect theft section there will be detection of theft like pattern occured in smart meter or not and will give output



The SARIMAX model is an extension of the SARIMA model that incorporates exogenous (explanatory) factors to improve forecasting performance[9]. Seasonal ARIMA with an eXogenous component (i.e. SARIMAX) is a multivariate variation of the SARIMA model that is mathematically stated as:

$$\varphi p\,(B)\,\Phi p\,(Bs)\,\nabla d\nabla D\,s\,yt = \beta kxk,t + \theta q(B)\Theta Q(Bs)\varepsilon t$$

wherekt, x ′ is the vector comprising the kth explanatory input variables at time t and Stationarity and invertibility requirements are the same as in ARMA models. The SARIMAX model's response variable in the analyzed application is PV production, whereas the explanatory variable is a time series of unit projections.Because the data is nonstationary (12-months seasonality included), the differentiation of both response and exogenous time series prior to model estimate is critical for effective SARIMAX model development; otherwise, the risk of "spurious regression" exists. The seasonal (P, D, Q) and non-seasonal (p, d, q) keywords were chosen mostly based on information criteria (AIC and FPE), although autocorrelation (ACF) and partial autocorrelation (ACF) were also considered (PACF)[9].For estimating power usage, p=1 (nonseasonal autoregressive), d=1 (non-seasonal differencing), and q=MA order will be 12

V.RESULTS

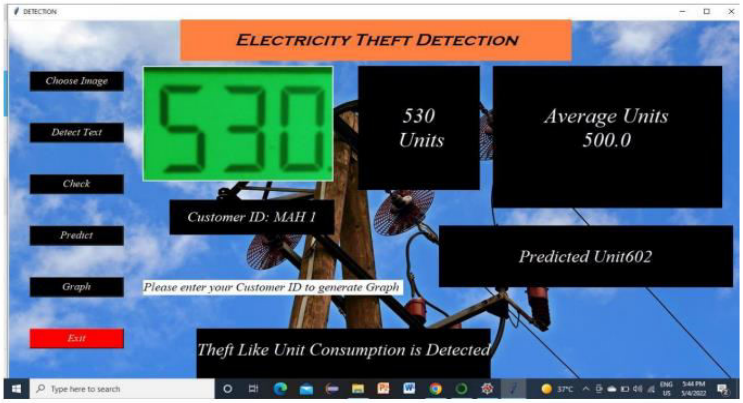
- Detect Electricity losses without any human control.
- To detect text in smart meter reading image using OCR. ● To apply and adapt a variety of problem solving strategies to solve problems.
- To apply and adapt a variety of problem solving strategies to solve problems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

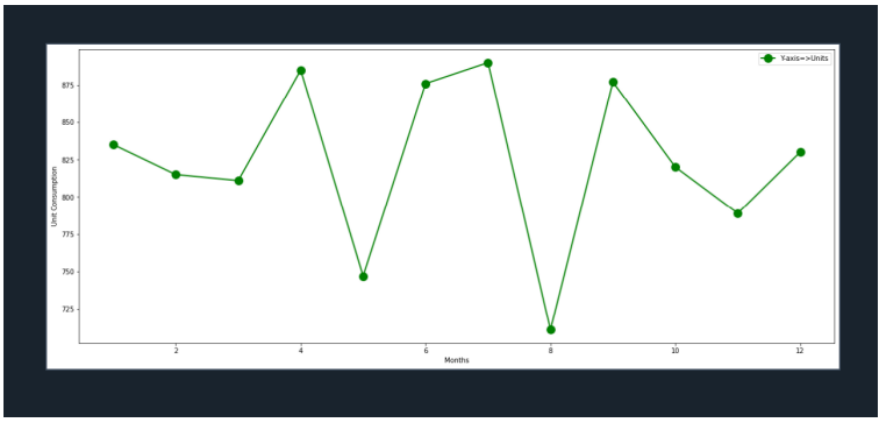
|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105318 |

Electricity losses detected:-



Graph of the electricity losses detected image:-



VI.CONCLUSION AND FUTURE WORK

This paper presents the system whicj implements electricity losses detection model using machine learning algorithms like OCR (optical character recognition) and SERIMAX. Our model provides an maximum accuracy of 97% compared

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105318 |

to previously proposed models which provide much lower accuracy. Our work is based on designing an application that detecteseletrcity losses by taking smart meter reading image aa input

In future The XG Boost model can be applied to generate the forecasts based on the best fit model during hyper parameter tuning. Big Datasets can be used for prediction the electricity theft in the city. More features can be added in this project to show consumption graph to user on front end and also show them range of current range of their genuine consumption.

REFERENCES

[1] J. Nagi, K. Yap, S. Tiong, S. Ahmed, and M. Mohamad, GuodongGu, Qingsu He "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 Muhammad Ismail, MostafaShahin, Mostafa F. Shaaban, ErchinSerpedin, and Khalid Qaraqe,", "Energy theft in the advanced metering infrastructure," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture.

[3] Pandurang G .Kate,"ISCX Android botnet dataset. Available from https://www.unb.ca/cic/datasets/androidbotnet.html

[4] Eslahi, R. Salleh, , Mahmoud Nabil and N. B. Anuar, "Bots and botnets: An overview of characteristics, detection and challenges," in Proceedings of the IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2012,pp. 349-354

[5] N. Hatziargyriou, and E. Dialynas, "Twostage R. E. Ogu and G. A. Chukwudebe 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].

[6] Nikhil V. Patil, Rohan S. Kanase, P D Bamane"Dissecting Android Malware: Characterization and Evolution," in Proceedings of the Symposium on Security and Privacy (SP), 2012,pp. 95-109.

[7] N. Billor, A. Hadi, and Muhammad Saad, Muhammad Faraz Tariq, Amna Nawaz, Muhammad Yasir Jamal 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.

[8] Muhammad Tariq and H. Vincent Poor, "Outlier detection for high dimensional data," 2001, pp. 37–46, cited By 433.

[9] Daniel Nikovsk, "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] Sandeep Kumar Singh, "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..