Survey on Identify Electricity Theft using Data mining Technique

Mane Kuldip C[1], Dhendwal Diksha H[2], Yadav Mayur V[3], Prof. Miss. Bala Kumkum[4]

B.E. Student, Dept. of Computer Engineering, Bharati Vidyapeeth Lavale, Pune, India[1,2,3]

Assistant Professor, Dept. of Computer Engineering, Bharati Vidyapeeth Lavale, Pune, India[4]

ABSTRACT: India has reaches to its highest point in terms of endemic use of electrical energy. Due to this use of electrical energy the overall growth of the country is getting harmed. As we know that the total power loss it equal to its transmission power losses. Distribution of electricity contains` significant Technical as well as Non-Technical Losses (NTL). Illegal consumption of electricity or electricity theft generates a major share of NTL. The transmission of power is mainly done through the network operation and problem in distribution power arises because of issues like unpaid bills, meter tempering , if famous has taken illegal connections. This project represents that by using data mining we can detect or the electricity theft in both public and private sectors as data mining can be used in both sectors. This project mainly focuses on an encoding technique that simplifies the received customer energy consumption readings (patterns) and maps them on corresponding irregularities in consumption. It also explains or clear the operation of classification techniques on customer energy consumption and illegal use of electricity, illegal consumer and also how much use of the electricity is theft by illegal connections

KEYWORDS: Data Pre-processing, Artificial Neural Network, K-Means Clustering, Gaussian Distribution, Fuzzy Logic.

I. INTRODUCTION

Today's world is started with all smart things. Such smart things are smart economy, smart lifestyle, smart phones, smart cities etc. To improve this smartness we propose the electricity theft detection using data mining technique. Electricity plays an important in development of any country. Every country should modernize when its industrial sector should grow rapidly. Electricity plays a vital role in country growth. Electricity generated in conventional method means to generate through nuclear power and thermal power generation and non-conventional method means solar power, wind power, hydro power etc. Government has spent so much money on generating and maintaining electricity. But so many cases of electricity theft are recorded hence some actions are needed to overcome this problems. In virtue of this, we use data mining technique to implement for theft detection and electricity billing in properly. Distribution of electricity involves significant Technical as well as Non-Technical Losses (NTL). Illegal consumption of electricity or electricity theft constitutes a major share of NTL. This project discusses several methods implemented by illegal consumers for stealing. With the advent of advanced metering technologies, real-time energy consumption data will be available at the utilities end, which can be used to detect illegal consumers. This project presents an encoding technique that simplifies the received customer energy consumption readings (patterns) and maps them into corresponding irregularities in consumption. Then, this project elucidates operation of intelligent classification techniques on customer energy consumption data to classify genuine and illegal consumers. These classification models are applied on regular energy consumption data as well as the encoded data to compare corresponding classification accuracies and computational overhead.

II. GOALS & OBJECTIVES

➢ Objective of our project is to create a system that will be helpful to avoid the electricity theft by end user as well as to reduce detection cost manually.
➢ The aim our project is to elucidate operation of intelligent classification techniques on customer energy consumption data to classify genuine and illegal consumers.

➢ These classification models are applied on regular energy consumption data as well as the encoded data to compare corresponding classification accuracies and computational overhead.

III.LITERATURE SURVEY

The requirement of electricity is huge than electricity being generated. In most countries either developed or under developed its economy is affected by mismanagement of electricity generation, transmission, distribution. Wireless technology like ZigBee used for theft detection and proper electricity billing. This system requires ZigBee technology, Remote collection of unit count and LCD display for sending bills. Future scope of ZigBee technology in electricity theft detection are automatic meter reading and billing in water management system, LPG gas management system in domestic housing society etc.[1]. Non-Technical losses comprise important concern for electricity distribution utilities worldwide .In this to identify and detect in electric utility market. To identify fraud activities an automatic feature extraction method for load profiles with a combination of support vector machine (SVM). Using data mining and statistical technique customer consumption patterns are extracted which represents customer load profiles. Load profile contains abnormalities when a fraud event occurs, SVM classifies load profiles of customers for detection of fraud suspects [2].Implementation of Advanced Metering Infrastructure (AMI) in smart grids reduces reduces the risk of energy theft through its monitoring capabilities and fine grained usage measurements. In traditional mechanical meters physical tampering mostly happened, but in AMI locally and remotely metering data can be tampered. An Energy theft detection system (ETDS) that detects energy theft against AMI is urgently needed [3].

IV.METHODALOGY

A. DATA PRE-PROCESSING-:

Data cleaning

1. Fill in missing values (attribute or class value):
    o   Ignore the tuple: usually done when class label is missing.
    o   Use the attribute mean (or majority nominal value) to fill in the missing value.
    o   Use the attribute mean (or majority nominal value) for all samples belonging to the same class.
2. Identify outliers and smooth out noisy data:
    o   Clustering: group values in clusters and then detect and remove outliers (automatic or manual)
    o   Regression: smooth by fitting the data into regression functions.
3. Correct inconsistent data: use domain knowledge or expert decision.

Data transformation

1. Normalization:
    o   Scaling attribute values to fall within a specified range.
        ▪   Example: to transform V in [min, max] to V' in [0,1], apply V'=(V-Min)/(Max-Min)
    o   Scaling by using mean and standard deviation (useful when min and max are unknown or when there are outliers): V'=(V-Mean)/StDev
2. Aggregation: moving up in the concept hierarchy on numeric attributes.
3. Generalization: moving up in the concept hierarchy on nominal attributes.
4. Attribute construction: replacing or adding new attributes inferred by existing attributes.

Data reduction

1. Reducing the number of attributes
    o   Data cube aggregation: applying roll-up, slice or dice operations.
    o   Removing irrelevant attributes: attribute selection (filtering and wrapper methods), searching the attribute space (see Lecture 5: Attribute-oriented analysis).

2. Reducing the number of attribute values
   o Binning (histograms): reducing the number of attributes by grouping them into intervals (bins).
   o Clustering: grouping values in clusters.
   o Aggregation or generalization
3. Reducing the number of tuples
   o Sampling

B. ARTIFICIAL NEURAL NETWORKS:

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. Using neural networks as a tool, data warehousing firms are harvesting information from datasets in the process known as data mining. The difference between these data warehouses and ordinary databases is that there is actual manipulation and cross-fertilization of the data helping users makes more informed decisions.

C. K-MEANS CLUSTERINGALGORITHM:

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more.

D. FUZZY LOGIC:

A fuzzy control system is a control system based on fuzzy logica mathematical system that analyze analog input values in terms of logical variables that take on continuous values between 0 and 1, in contrast to classical or digital logic, which operates on discrete values of either 1 or 0.

Fuzzy logic in electricity theft detection shows unbalanced use of electricity by the user. It does not predict answer only Gives result in percentage.

V. CONCLUSION AND FUTURE WORK

Presented Article is literature survey examination on electricity theft detection. Article reviews existing techniques in data mining domain. Focus of article remains at deriving problem definition. Review examination concludes requirement of artificial intelligence to find unseen frauds. Artificial neuron network are found to be successful solution .Future scope is to define technique to overcome shortcoming in existing system and implement system with artificial intelligence.

REFERENCES

[1] P. G. Kate, J. R. Rana. "ZigBee Based Monitoring Theft Detection and Automatic Electricity Meter Reading"2015 International Conference on Energy Systems and Applications (ICESA 2015)

[2] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, Member, IEEE,S. K. Ahmed, Member, IEEE "Non-Technical Loss Analysis for Detection of Electricity Theft using Support Vector Machines" 2 IEEE International Conference on Power and Energy (PECon08),December 1-3,2008, Johor Bahru, Malaysia.

[3] Alvaro A. Cardenas. University of Texas, Dallas, Saurabh Amin. MIT, "A Game Theory Model For Electricity Theft Detection and Privacy-Aware Control in AMI System " 15 Annual Allerton Conference Allerton house, llliois, USA October 1-5, 2012.

[4] Dr. Yashpal Singh, Alok Singh Chauhan, "Neural Networks In Data Mining ", Reader Bundelkhand Institute Of Engineering & Technology, Jhansi, India.

BIOGRAPHY

Mane Kuldip C, Dhendwal Diksha, Yadav Mayur V.all are students of Bharati Vidyapeeth Collage of engineering lavale,Pune. Studied in last year of Computer Engineering.

Miss. Bala Kumkum.is a Assistant professor in the Department ofcomputer engineering, College of of Bharati Vidyapeeth Collage of engineering lavale,Pune,Savitribai phule Pune University.