ISSN: 2278-4632
Vol-11 Issue-02 2021

FRAUDULENT BEHAVIOUR IN WATER CONSUMPTION

[1]Mrs.D.Uma, Asst.Professor -CSE, Sri Indu Institute of Engineering And Technology, Telangana, India,email id:umadonthagani@gmail.com

[2]K.Abhinav, Student (B. Tech, CSE), Sri Indu Institute of Engineering And Technology, Telangana, India

[3]K.Satish, Student (B. Tech, CSE), Sri Indu Institute of Engineering And Technology, Telangana, India

[4]K.Shiva, Student (B. Tech, CSE), Sri Indu Institute of Engineering And Technology, Telangana, India

[5]J.Sandeep, Student (B. Tech, CSE), Sri Indu Institute of Engineering And Technology, Telangana, India

Abstract:

Fraudulent behaviour in drinking water consumption is a significant problem facing water supplying companies and agencies. This behaviour results in a massive loss of income and forms the highest percentage of non-technical loss. Finding efficient measurements for detecting fraudulent activities has been an active research area in recent years. Intelligent data mining techniques can help water supplying companies to detect these fraudulent activities to reduce such losses. This research explores the use of two classification techniques (SVM and KNN) to detect suspicious fraud water customers. The main motivation of this research is to assist Yarmouk Water Company (YWC) in Irbid city of Jordan to overcome its profit loss. The SVM based approach uses customer load profile attributes to expose abnormal behaviour that is known to be correlated with non-technical loss activities. The data has been collected from the historical data of the company billing system. The accuracy of the generated model hit a rate of over 74% which is better than the current manual prediction procedures taken by the YWC. To deploy the model, a decision tool has been built using the generated model. The system will help the company to predict suspicious water customers to be inspected on site

I. INTRODUCTION

1. INTRODUCTION

Water is an essential element for the uses of households, industry, and

ISSN: 2278-4632
Vol-11 Issue-02 2021

agriculture. Jordan, as several other countries in the world, suffers from water scarcity, which poses a threat that would affect all sectors that depend on the availability of water for the sustainability of activities for their development and prosperity.

According to Jordan ministry of water and irrigation, this issue always has been one of the biggest barriers to the economic growth and development for Jordan. This crisis situation has been aggravated by a population increase that has doubled in the last two decades. Efforts of the ministry of Water and irrigation to improve water and sanitation services are faced by managerial, technical and financial determinants and the limited amount of renewable freshwater resources.

To address these challenges, Jordan ministry of water and irrigation as in many other countries is striving, through the adoption of a long-term plan, to improve services provided to citizens through restructuring and rehabilitation of networks, reducing the non-revenue water rates, providing new sources and maximizing the efficient use of available sources. At the same time, the Ministry continues its efforts to regulate the water usage and to detect the loss of supplied water.

Water supplying companies incur significant losses due to fraud operations in water consumption. The customers who tamper their water meter readings to avoid or reduce billing amount is called a fraud customer. In practice, there are two types of water loss: the first is called technical loss (TL) which is related to problems in the production system, the transmission of water through the network (i.e., leakage), and the network washout. The second type is called the non-technical loss (NTL) which is the amount of delivered water to customers but not billed, resulting in loss of revenue.

The management of the Yarmouk Water Company (Jordan) has a significant concern to reduce its profit losses, especially those derived from NTLs, which are estimated over 35% in the whole service area in the year 2012. One major part of NLT is customer's fraudulent activities; the commercial department manages the detection processes with the absence of an intelligent computerized system where the current process is costly, not effective nor efficient.

NTL is a serious problem facing Yarmouk Water Company (YWC). In 2012 the NTL reached over 35%, ranging from 31% to 61 according to districts, which results in a loss of 13 million dollars per year. Currently, YWC follows random inspections for customers, the proposed model in this paper provides a valuable tool to help YWC teams to detect theft customers, which will reduce the NTL and raise profit.

Literature has abundant research for Non-Technical Loss (NTL) in electricity fraud detection, but rare researches have been conducted for the water consumption sector. This paper focuses on customer's historical

ISSN: 2278-4632
Vol-11 Issue-02 2021

data which are selected from the YWC billing system. The main objective of this work is to use some well-known data mining techniques named Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) to build a suitable model to detect suspicious fraudulent customers, depending on their historical water metered consumptions.

SYSTEM ANALYSIS

3.1 Existing System:

Literature has abundant research for Non-Technical Loss (NTL) in electricity fraud detection, but rare researches have been conducted for the water consumption sector. Water supplying companies incur significant losses due to fraud operations in water consumption. The customers who tamper their water meter readings to avoid or reduce billing amount is called a fraud customer. In practice, there are two types of water loss: the first is called technical loss (TL) which is related to problems in the production system, the transmission of water through the network (i.e., leakage), and the network washout. The second type is called the non-technical loss (NTL) which is the amount of delivered water to customers but not billed, resulting in loss of revenue. To address these challenges, Jordan ministry of water and irrigation as in many other countries is striving, through the adoption of a long-term plan, to improve services provided to citizens through restructuring and rehabilitation of networks, reducing the non-revenue water rates, providing new sources and maximizing the efficient use of available sources. At the same time, the Ministry continues its efforts to regulate the water usage and to detect the loss of supplied water.

3.2 Disadvantages of Existing System:

- Complexity in analyzing the data.
- Prediction is challenging task working in the model
- Coding is complex maintaining multiple methods.

ISSN: 2278-4632
Vol-11 Issue-02 2021

- Libraries support was not that much familiar.

3.3 Proposed System:

This paper focuses on customer's historical data which are selected from the YWC billing system. The main objective of this work is to use some well-known data mining techniques named Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) to build a suitable model to detect suspicious fraudulent customers, depending on their historical water metered consumptions. The CRISP-DM (Cross Industry Standard Process for Data Mining) was adopted to conduct this research. The CRISPDM is an industry standard data mining methodology developed by four Companies; NCR systems engineering, DaimlerChrysler AG, SPSS Inc. and OHRA. The CRISP-DM model consists of business understanding, data understanding, data preparation, model building, model evaluation and model deployment. To extract the fraud customers' profile, a new table is created containing the client's number, the water consumption, and a new attribute for fraud class. This attribute is filled with a value of 'YES'. Another table for the normal clients is created, and the fraud class attribute is filled with the value "NO". The two tables are then consolidated into one table containing the customer ID, consumption profile, and fraud class attributes. To filter the data, some preprocessing operations were performed such as Eliminate redundancy, Eliminate customers having zero consumption through the entire period, Eliminate new clients who are not present during the whole targeted period, and Eliminate customers having null consumption values. Filtering the data resulted in a reduced original dataset of the non-fraud customer to 16114 record and the fraud customers to 647 records.

3.4 Advantages:

- Libraries help to analyse the data.

ISSN: 2278-4632
Vol-11 Issue-02 2021

- Statistical and prediction is very easy comparing to existing technologies.

  Results will be accurate compared to other methodologies

SOFTWARE REQUIREMENT SPECIFICATION

The reason for this SRS record is to distinguish the necessities and functionalities for Intelligent Network Backup Tool. The SRS will characterize how our group and the customer consider the last item and the attributes or usefulness it must have. This record additionally makes a note of the discretionary prerequisites which we intend to execute yet are not required for the working of the venture.

This stage assesses the required necessities for the Images Processing for an orderly method for assessing the prerequisites a few procedures are included. The initial step associated with dissecting the prerequisites of the framework is perceiving the idea of framework for a solid examination and all the case are defined to better comprehend the investigation of the dataset.

INTENDED AUDIENCE AND READING SUGGESTIONS

This record is proposed for extend engineers, directors, clients, analyzers and documentation journalists. This report goes for examining plan and execution imperatives, conditions, framework highlights, outside interface prerequisites and other non utilitarian necessities.

IDENTIFICATION OF NEEDS

The first and imperative need for a business firm or an association is to know how they are performing in the market and parallelly they have to know how to conquer their rivals in the market.

To do as such we have to investigation our information in view of all the accessible variables

ISSN: 2278-4632
Vol-11 Issue-02 2021

4.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- **Economical Feasibility**
- **Technical Feasibility**
- **Social Feasibility**

4.1.1 Economical feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

4.1.2 Technical feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

ISSN: 2278-4632
Vol-11 Issue-02 2021

4.1.3 Social feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

5. SYSTEM DESIGN

The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces.

This section describes the system in narrative form using non-technical terms. It should provide a high-level system architecture diagram showing a subsystem breakout of the system, if applicable. The high-level system architecture or subsystem diagrams should, if applicable, show interfaces to external systems. Supply a high-level context diagram for the system and subsystems, if applicable. Refer to the requirements trace ability matrix (RTM) in the Functional Requirements Document (FRD), to identify the allocation of the functional requirements into this design document.

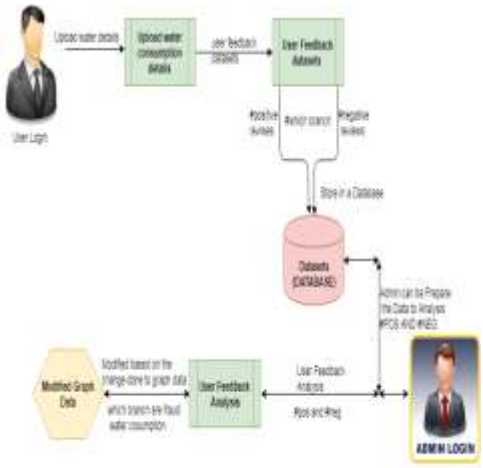
This section describes any constraints in the system design (reference any trade-off analyses conducted such, as resource use versus productivity, or conflicts with other systems) and includes any

ISSN: 2278-4632
Vol-11 Issue-02 2021

assumptions made by the project team in developing the system design.

The organization code and title of the key points of contact (and alternates if appropriate) for the information system development effort. These points of contact should include the Project Manager, System Proponent, User Organization, Quality Assurance (QA) Manager, Security Manager, and Configuration Manager, as appropriate

SYSTEM ARCHITECTURE



Future Work

The conducted experiments showed that a good performance of Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) had been achieved with overall accuracy around 70% for both. In Future accuracy of the same can be improved with the help of improved techniques. The model hit rate is 60%-70% which is apparently better than random manual inspections held by YWC teams with hit rate around 1% in identifying fraud customers. This model introduces an intelligent tool that can be used by YWC to detect fraud customers and reduce their profit losses. The suggested model helps saving time and effort of employees of Yarmouk water by identifying billing errors and corrupted meters. With the use of the proposed model, the water utilities can increase cost recovery by reducing administrative Non-Technical Losses (NTL's) and increasing the productivity of inspection staff by onsite inspections of suspicious fraud customers.

VI. CONCLUSION

In this research, we applied the data mining classification techniques for the purpose of detecting customers' with fraud behaviour in water consumption. We used SVM and KNN classifiers to build classification models for detecting suspicious fraud customers. The models were built using the customers' historical metered consumption data; the Cross Industry Standard Process for Data Mining (CRISP-DM). The data used in this research study the data was collected from Yarmouk Water Company (YWC) for Qasabat Irbid ROU customers, the data covers five years customers' water consumptions with 1.5 million customer historical records for 90 thousand customers. This phase took a considerable effort and time to pre-process and format the

ISSN: 2278-4632
Vol-11 Issue-02 2021

data to fit the SVM and KNN data mining classifiers

REFERENCES

[1] N/A, "Jordan Water Sector Facts & Figures, Ministry of Water and irrigation of Jordan". Technical Report. 2015.

[2] N/A, "Water Reallocation Policy, Ministry of Water and irrigation of Jordan". Technical Report. 2016.

[3] C. Ramos , A. Souza , J. Papa and A. Falcao, "Fast non-technical losses identification through optimum-path forest". In Proc. of the 15th Int. Conf. Intelligent System Applications to Power Systems, 2009, pp.1-5.

[4] E. Kirkos, C. Spathis and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements", Expert Systems with Applications, 32(2007): 995–1003.

[5] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines", IMECS, 2011, Vol I, pp. 16 – 18.

[6] S. Panigrahi, A. Kundu, S. Sural and A. Majumdar, "Credit card fraud detection: a fusion approach using dempster–shafer theory and bayesian learning, information fusion", 2009, 10(4): 354–363.

[7] N. Carneiro, G. Figueira and Costa M., "A data mining based system for credit-card fraud detection in e-tail decision support systems", Decision Support Systems, 2017, 95(C): 91-101.

[8] Ortega P., Figueroa C., and Ruz G. "A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile", In proc of DMIN, 2006.

[9] B. Kusaksizoglu, "Fraud detection in mobile communication networks using data mining", Bahcesehir University, The Department of computer engineering, Master Thesis. 2006.

[10] C. Liang-Chun, H. Chien-Lung, L.Nai-Wei, Y. Kuo-Hui and L. Ping-Hsien, "Fraud analysis and detection for real-time messaging communications on social networks", IEICE Trans. Inf. & Syst., 2017, Vol. E100–D, No.10, pp: 2267-2274.