



MODEL FOR IDENTIFYING FRAUDULENT WATER CONSUMPTION ACTIVITY BASED ON DATA MINING

¹Ranjit Patnaik Sekharamantri, ²Narala Narasimha

¹Computer Science and Engineering,
¹Lovely Professional University, Phagwara, India

Abstract: As a result of financial frauds, financial losses are on the rise. The challenge of fraud detection is identifying losses and suspicious behavior. It is important to understand the underlying business objectives in order to apply data mining objectives. It is critical to apply data mining objectives to financial statements in order to point out fraudulent usage. All water and power companies worldwide that manage their billing systems using a financial system are faced with consumer dishonesty. Research has focused on identifying fraudulent electricity consumption for several years. This thesis uses data mining techniques to develop a new model for detecting non-technical losses (NTL) in water consumption utilities. As part of developing a fraud detection model for water consumption, this study applied a suitable data mining technique based on a financial billing system for water consumption in the city. An accepted method for testing and evaluating the model's efficiency and accuracy has been developed. As part of this research, an intelligent model has been developed that predicts and selects suspicious customers for inspection by the department of water theft combat (DWTC) teams at the municipality of India (MOI) in order to detect fraud. It is estimated that by using this model, 80% of the random manual detection hit rate will be replaced by an intelligent detection hit rate of 1-10%. In this approach, features are selected and extracted from historical water consumption data of customers, which is a form of data mining. In this study, SVC is used to identify abnormal customer load profile behaviors based on customer load profile information.

Keywords: Data Mining, Water, Intelligence System, financial fraud, FFD.

1. INTRODUCTION

In distributing water fairly to citizens, water theft can be a big problem, decreasing an organization's revenue. As a non-profit foundation responsible for providing water services to all citizens, the MOI must balance its expenses with its revenue to provide fair water services to all citizens. Consequently, fraudulent water consumption resulting in an incorrect consumption amount is essential for achieving revenue improvement. Approximately 8000 buildings in the city do not have the water meters that the MOI uses to calculate each customer's monthly water consumption. As a result of meter tampering, more than 3000 cases of water breach have been reported over the past five years. An irregular bill, a malfunctioning meter, an illegal connection. DWTC at MOI conducts manual investigations to detect customer irregularities and metered water consumption theft, which is slow, costly, and done randomly, which is why it is important to speed up and increase the accuracy of investigations so that suspicious customer consumption can be detected and costs reduced.

In various domains, there are several groups of researchers who devote considerable time and effort to studying fraud detection. Several of them applied data mining techniques successfully to solve the problem, but each work had pros and cons. In order to identify fraud based on customer water consumption profiles using classification techniques, the research focuses on the cons of the related works while avoiding their pros. SVM is applied in the research and is compared with other detection methods like ANNs and KNNs. In the underlying business fraud detection, the method SVM achieved the highest performance and accuracy. Based on historical consumption data managed by a computerized billing system, the model improves manual hit rate detection from 1- 10% random detection to 80% intelligent detection.

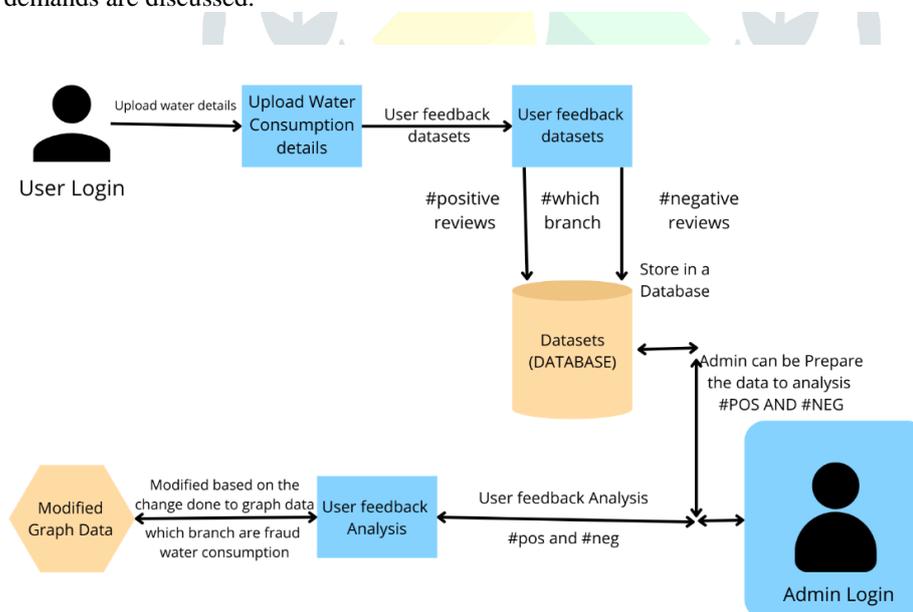
In load profiling, variations in load are explained by the time of day, the time of the year, by the type of day or season of the year, as well as by the type of day. Thus, it is a representation of customer information and known or observable weather conditions as well as habits and usual weather conditions. Therefore, load profiles take into account the type of activity a customer is involved in, as well as their behavior. This thesis presents three types of customer load profiles (monthly, seasonal, and yearly) derived from historical water consumption databases and tested using the proposed classification method.

According to this research, a methodology will be developed for determining a typical normal load profile and a representative fraud load profile based on the type of water consumption of customers. It is the responsibility of the financial billing system to calculate and issue water invoices based on the water consumption of customers (load profiling). Currently, an employee of a utility reads the meter at the customer's premises, records the reading in a book, transfers it to the computer system, and generates a bill based on the reading.

2. APPROACH

As a result of this study, the following objectives will be met:

- Identify opportunities to reduce the MOI's NTLs in the water distribution sector.
- Analyze the possibility of detecting and identifying NTL activities with data mining classification techniques, and then resolve the existing drawbacks in that they deal with 20% of the available load profile data for customers. In order to detect fraud activities, MOI DWTC teams performed on-site inspections manually and randomly, so the intelligent model developed in this study can predict suspicious customers.
- It is believed that the financial losses caused by water consumption result from the big difference between water wells located within the city and water consumers, as stated in the new statistics provided by the department of water distribution. Approximately 15 million cubic meters of water were lost in 2011 as a result of this difference. This research assumption applies only to financial billing systems that manage and calculate water consumption. It focuses on historical transaction data of MOI and uses SVM classification data mining techniques, which are widely applied internationally. An illegal connection without a consumption meter can result in water theft, or an illegal connection within a metered consumption meter can result in water theft. This research examines customers whose water consumption is metered and who participate in at least one water agreement. Classification methods dominate this field of study. So in this study we tested and analyzed three classification techniques on the underlying business datasets. These techniques were support vector machines (SVM), artificial neural networks (ANN), and K-nearest neighbor (KNN).
- This thesis research is divided into six chapters. Chapter 1 presents general overview, research objectives, problem definition, motivation for the research, and limitations and scope. A literature review of the thesis literature is presented in Chapter 2, which includes definitions of non technical losses, load profiling, data mining, classification methods, and performance evaluation methods. A few related fraud detection projects are discussed and introduced in Chapter 3. In chapter 4, the model development methodology is presented and the business data sets are highlighted. The final analysis of the classification techniques on all datasets is presented in chapter 5. In chapter 6, the thesis research is summarized and future demands are discussed.



Utility of linear and nonlinear model

3. RELATED WORK

Data mining techniques have been used to detect fraud in many published papers. There has been significant research into FFD based on the same data by several groups of researchers, so they chose FFD to build a model for predicting the probability of a current transaction based on the transaction information and historical trading patterns. Providing a scientific basis for anti-fraud refusal strategies and refusal strategies to launch, or refusals to investigate suspicious transactions, is the objective of this study. Several domains are presented in this chapter, including banks, energy, and health, where the state-of-the-art for fraud detection techniques is introduced. There are three types of technologies used in fraud detection in this chapter: the RDBMS (unsupervised

learning), the Data Mining (supervised learning) and the Data Mining (unsupervised learning). The authors conclude by discussing and demonstrating the weaknesses of the related works.

Three classification methods are used to detect credit card fraud in this study, including decision trees, neural networks, and logistic regression. Based on a transactional database with over 40 fields, the researchers generated a predictive model by revealing a few variables, which are common data schema used by most banks. Due to a nondisclosure agreement, they reveal a few variables that most banks use. As far as fraud and nonfraud records are concerned, the bank already labeled the data used. In order to measure the performance of targeting models, they then evaluated the three models using a lift chart metric. The results show that a neural network outperformed a logistic regression and decision tree.

As a result of designing and applying a database application to record and control prescription data in a health care center, some restrictions and constraints were implemented to prevent fraud. In addition to administrative restrictions, medical rules were employed to ensure that prescribed drugs correspond to the diagnoses that appear on prescriptions. 20000 prescription records were used for the application. As shown in Figure 1, an interesting result of statistical analysis found that 2% of insured people were diagnosed with a high number of different diseases over the course of one year, which needs to be investigated further. Data mining techniques can be applied with real correct results with the help of a powerful relational database.

4. Proposed Work

YWC's billing system is used to select historical data about customers. Based on their historical water metered consumptions, this study seeks to build a suitable model to detect suspicious fraudulent customers, using some well-known data mining techniques such as Support Vector Machines (SVM) and K-Nearest Neighbor (KNN). In this research, we used CRISP-DM (Cross Industry Standard Process for Data Mining). Four companies have developed the CRISPDM data mining methodology; NCR systems engineering, DaimlerChrysler AG, SPSS Inc., and OHRA. Among the components of the CRISP-DM model are understanding the business, understanding the data, preparing the data, building the model, evaluating the model, and deploying it. We create a new table with the client's number, water consumption, and a fraud class attribute for extracting the fraud customer's profile. In this attribute, 'YES' is entered. A new table is created for normal clients, and the fraud class attribute is entered as "NO". The two tables are then consolidated into one table, containing customer IDs, consumption profiles, and fraud class attributes. To filter the data, some pre-processing operations were performed, including eliminating redundancy, eliminating customers with zero consumption through the entire period. We eliminated new clients who were not present for the whole period, and we eliminated customers with null consumption values. Filtering the original dataset resulted in a reduced dataset of 16114 non-fraud records and 647 fraud records.

5. Modules

- Customer Data

Customers who want water via agencies must register with the system. This registration is the only way that customers can get water and receive bills. The only way for the customer to consume water is by contacting it.

- Verify Feedback

As soon as field executives check the limit, bills are generated. In this way, the quantity consumed by each customer must match the quantity noted by the admin. Fraud details can be checked through this process. Bills were uploaded after this and fraud among customers was found.

- Action against fraudulent

An admin can locate fraudulent customers who consume more water than they need or may be required, and the bills can also be verified by them. By blocking the fraud details, the user prevents them from receiving any more water from them, and the cops are notified.

- Graph Analysis

This analysis gives admins an overview of the data, which allows them to find fraudulent customers. By understanding where exactly the problem arises and finding where improvements and lacks occur, the business gradually improves. By using this dataset, a clear picture will be provided regarding the current and past picture.

6. Conclusions

An investigation and application of SVM to fraud detection has been presented in this study. This study focuses primarily on using SVM, specifically SVC, to classify patterns (load consumption profiles) into two categories: normal (No) customers and fraud (Yes) customers based on their load consumption profiles. The following are the contributions of the thesis research study:

According to this thesis, the main findings and advantages of the research study are listed as follows.

- The researcher considers the analysis and mining of a new business domain as a research contribution as it pertains to managing customers' water consumptions in the municipality.
- Utilize various data sources such as computerized historical water consumption data and paper files containing irregularities of customers to design and reveal a fraud detection model
- With experiments, a novel, efficient, and effective real-world classifier model for fraud detection of water consumption profiles was developed to identify relevant attributes and relevant periodical consumption datasets that improved classification accuracy for data, allowing fraudsters' profiles to be better modeled.

REFERENCE:

- [1] Approach to Detection of Tampering in Water Meters", In *Procedia Computer Science*, 2015, 60: PP 413-421. [10] B. Boser, I. Guyon, and V. Vapnik, "A coaching formula for best margin classifiers," in *Proceedings of the 5th Workshop on Computational Learning Theory*.
- [2] E. Kirkos, C. Spathis and Y. Manolopoulos, "Data mining techniques for the detection of fallacious money statements", *skilled Systems with Applications*, 32(2007): 995-1003
- [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] Approach to Detection of Tampering in Water Meters", In *Procedia Computer Science*, 2015, 60: PP 413-421. [10] B. Boser, I. Guyon, and V. Vapnik, "A coaching formula for best margin classifiers," in *Proceedings of the 5th Workshop on Computational Learning Theory*.
- [5] Jeyaranjani J and, Devaraj D "Machine Learning Algorithm for efficient power theft detection using smart meter data" *International Journal of Engineering & Technology*, 7 (3.34) (2018) 900-904.
- [6] Ishmael S. Nelwamondo Msiza, and Fulufhelo V. TshilidziMarwala "Artificial Neural Networks and Support Vector Machines for Water Demand Time Series Forecasting", 2007.
- [7] "Monedero, Félix Biscarri, Juan I. Guerrero, MoisésRoldán, Carlos León "An Approach to Detection of Tampering in Water Meters" *Procedia Computer Science* 60 (2015) 413-421.
- [8] E. Kirkos, C. Spathis and Y. Manolopoulos, "Data mining
- [9] Juan Ignacio, Carlos Leon "Real Application on Nontechnical losses detection", *The 2011 World Congress in Computer Science, Computer and Applied Computing (WORLDCOMP 11)*, Volume: The 2011 International Conference on Data Mining.