International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Special Issue: | Apr 2023 www.irjet.net p-ISSN: 2395-0072

International Conference on Recent Trends in Engineering & Technology- 2023 (ICRTET-3)
Organised by: VSM College of Engineering, Ramachandrapuram

DATA MINING MODEL BASED ON DISCRIMINATIONOF AMBIGUOUS BEHAVIOUR IN WATER CONSUMPTION

[1] M. Naga Jyothi Assistant Professor, [2] K. Surekha Associate Professor, [3]P.Ramya Assistant Professor

Dept of Computer Science, V.S.M. College of Engineering Ramachandrapuram

---

ABSTRACT

Water supply companies and government struggle greatly with drinking water fraud. The majority of non-technical losses are caused by this activity, which has a considerable negative financial impact. Recentresearch has placed a lot of emphasis on developingreliable criteria for identifying dishonest behavior's. Waterdelivery firms have access to sophisticated data mining techniques that can be used to identify fraud andcut losses. The purpose of this study istoinvestigate the use of three classification algorithms (SVM,KNN,ANN) to identify potentialwater scam victims. This study's primary goal is to help the Yarmouk Water Company (YWC) in Irbid, Jordan, fill its revenue gap. Known anomalous activity is exposed by the SVM-based approach using customerload profile characteristics.

I. INTRODUCTION

Water theft reduces the organization's revenue and makes it more difficult to give water to all citizens equally. Despite the fact that the MOI is a nonprofit organisation tasked withproviding water to all residents, maintaining a balance between costs and income is essential to providing an equitable water supply to everyone. A false consumption amount is produced by fraudulent water use since it is crucial to properly quantify the energyused in order to increase profits. Around 8000 buildings in the city, according to the MOI, receive water without the aid of a water metre, which is what the MOI uses to determine each client's monthly usage.

There are advantages and disadvantages to each strategy that has been developed by the many peoplewho are affected by this fraudulent behaviour and who want to stop it in addition to recovering from the losses brought on by it. We developed a data mining model to get around some drawbacks and increase the advantages. In comparison to SVM, other approaches like KNN and ANN are less accurate and perform worse. The SVM performs and is accurate really well. The SVM achieves an intelligent detection rate of 80% and a detection rate for random data of 1–10%. Depending on the customer's activities and behaviour, the customer profile can be altered. According on the client's choices, this strategy divides the customer data into three categories: monthly, seasonally, and yearly. To determine the client's behaviour, past customer data is gathered and put to the test. This study contributes to the provision of customer load profiles according to water consumption patterns. The primary goal of this study is to distinguish between genuine customers and scam customers. Customers' water consumption is tracked in this system.

II.PROCEDURE

The primary goals of this study are to

1. Help the MOI reduce its NTL's for the water distribution sector.

2. Consider the challenges of dealing with only 20% of the available customer load profile data, as well as the ability to detect and identify NTL actions using data mining classificationalgorithms.

3. Criminal activities In order for the intelligent model developed in this research study to forecast suspicious customers and help them uncover fraud activities, MOI DWTC staff manually and randomly inspect clients on-site. Recent data from the water distribution agency show that the financial losses brought on by water use are due to a significant discrepancy between the city's water well production and water use, as seen in Fig 2. The difference increased to 15 in2011.

This method is used to categorize clients based on their conduct. The technique determines whether a customer is fraudulent or not by comparing their data with information that has already been saved. When classifying data, SVM,KNN and ANN classification models are used because of their high accuracy and superior performance.

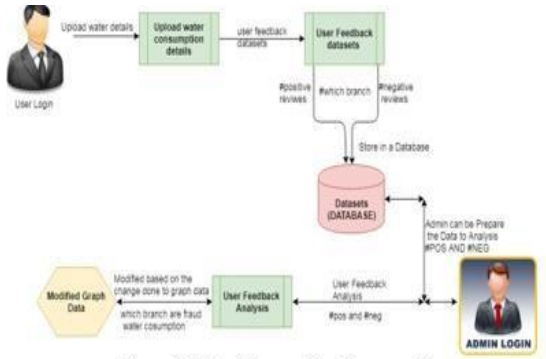


Figure 1 Utility of linear and nonlinear model

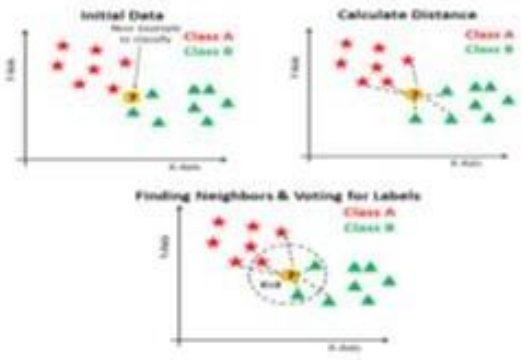
Customers need to register with the system in order to purchase water from agencies. You can upload data about your water consumption after registering. Field executives use this technique to gather client feedback andmonitor branch-specific restrictions. Administrators can arrange the data to examine both positive and negative feedback. Based on the user's comments, fraud details are blocked,prohibiting them from receiving any more water. The problem's root cause can be found by the admin, who can find the fraudulent consumers. This will give a clear picture of the dataset'scurrent and historical state.

A. Algorithm

In this study, experiments are conducted using the K-Nearest Neighbour (KNN), Artificial Neural Network (ANN) and SVM classification algorithms, which are applied to a data set of water customers4 obtained from theExecutives. The datasets are evaluated to determine their accuracy. Each method has been applied to the trainingdataset, and its accuracy performance has been assessed together with the forecasting performed on the testing dataset. One of the most popular and widely applied open source data mining methods worldwide. It offers a pleasant user interface, whereconfigurations for analyses are made inthe process view. It employs a modularconcept where the analysis process applies the appropriate operators. These operators can interface with other operators via I/O ports to receive input data or to transfer data and models created by one operator to the next. In this approach, data flow is produced throughout the analytical process. Forecasts are produced using the K nearest neighbouring results (K-Nearest Neighbour).5 We must therefore develop a measure formeasuring the distance between the query point and the instances in theexamples sample in order to make predictions with KNN and ANN.

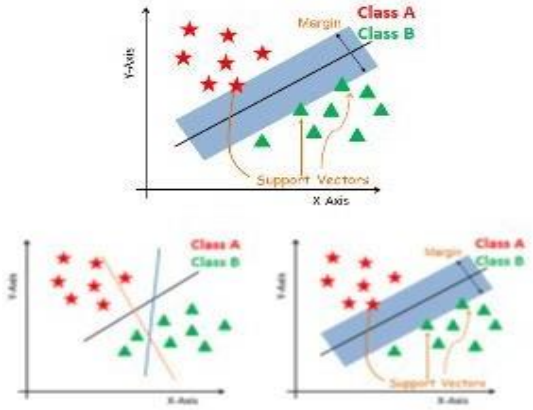
1. KNN Algorithm

(Image)[11]

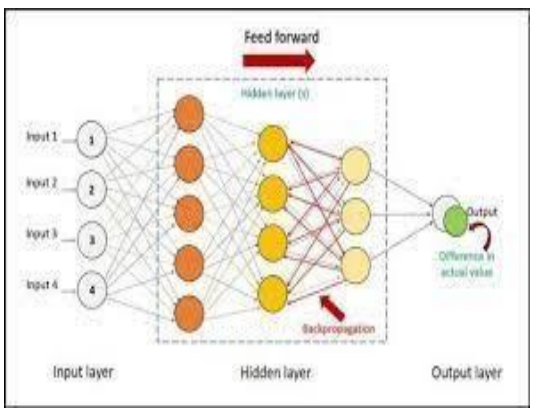


2. SVM Algorithm

(Image)[10]



3. ANN Algorithm



ANN-ALGORITHM(Artificial Neural Network):

A neural network is a method of information processing. It's kind of likethe way in which our brains process information. The ANN system is composed of a large number of interconnected processing units which work together to collect and store information. They're also delivering important results. Not only can we use the neural network to classify, we can use it to classify. It can also apply for regression of continuous target attributes.

For data mining that is used in the sector, neural networks are of great use. In the area of economics, forensics, etc.and pattern recognition for example. It's also possible, after a thorough training, to use it for the classification of huge amounts of data. The research forecasts that in the forecast period 2019 to 2026, there will be a projected 33.5% growth of the network software sector.

KNN,

The following 3 layers can be included in a neural network:

Input layer : The raw information which can be passed on to the network is represented by the operation of inputunits.

Hidden layer : To determine theactivity of each hidden unit. The operations of the input units, as well as weightings for connections between inputs and hidden units. Maybe there's just one or two concealed layers.

Output layer : The behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and the output units.

B. Modules

1. Customers must register with the system in order to obtain water through agencies. (Users' only choice to drink water from customers is to register.)6 Customer requests for the creation of water bills and administration.

2. Verify Feedback: After checking the limit, on-field executives generate the bills. They must consume the same amount of food as reported by admin. This process allows us to check the specifics of fraud. The bills were then posted, and the fake clients were located.

3. Fraudulent customers that illegally consume more water than they are allowed or require can be identified by the admin, and their bills can be verified. The user has the option to block fraud details, preventing them from obtaining any further water, and to turn over the details to law enforcement for possible criminal prosecution.

4. Graph analysis: The graphs make it easier for the administrator to understand the data. Based on this analysis, fraud clients can be located. The company gradually gets better as they gain a better grasp of where the issue arises and where flaws and improvements are required. This will give a clear picture of the dataset's current and historical state.

Conclusion

In this study, we used data mining classification approaches to identify clients who engaged in fraudulent water usage. We developed classification models for identifying suspect fraud clients using SVM, and ANN classifiers. The Cross Industry Standard Process for Data Mining (CRISP-DM) was used to create the models utilizing previous metered consumption data from the clients. The data used in this research study was gathered from Yarmouk Water Company (YWC) for Qasabat Irbid ROU subscribers, and it comprises five years' worth of water consumption statistics for 90,000 users. The pre-processing and formatting of the data to match the SVM and KNN data mining classifiers required a significant amount of work and time in this step.

References:

[1] AihuaShen, Rencheng Tong "Application of classification Models on credit card Fraud Detection", 2007.

[2] Anastassios Tagaris "Implementation of Prescription Fraud Detection Software Using REDBMS Tools and ATC Coding", 2009.

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

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

[5] 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.

[6] 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.

[7] Juan Ignacio, Carlos Leon "Real Application on Nontechnical losses detection", The 2011 World Cogress in Computer Science, Computer Engineering,

and Applied Computing (WORLDCOMP 11), Volume: The 2011 International Conference on Data Mining.

[8] Ishmael S. Msiza, Fulufhelo V. Nelwamondo and Tshilidzi Marwala "Artificial Neural Networks and Support Vector Machines for WaterDemand Time Series Forecasting",2007.

[9] 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.

[10] "Monedero, Félix Biscarri, Juan I. Guerrero, Moisés Roldán, Carlos León "An Approach to Detection of Tamperingin Water Meters" Procedia Computer Science 60 (2015) 413 –421.

[11] https://www.analyticsvidhya.com/ blog/ 2021/06/build-an-image-classifier-