iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44661

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A Data Mining Based Model for Detection of Fraudulent Behaviour in Water Consumption

Alavala Sathwik[1], B. Yugendar Nayak[2], Dr. Y. Srinivasulu[3]
[1, 2]Students, [3]Guide & Coordinator , Department of Electronics and Communication, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

Abstract: Drinking water fraud is a major issue for water delivery businesses and authorities. This behaviour results in a significant loss of income and is responsible for the bulk of non-technical losses. Finding appropriate criteria for detecting fraudulent behaviour has been a prominent focus of research in recent years. Intelligent data mining methods can be used by water delivery companies to detect fraudulent activity and reduce losses. This study looks into the usage of two classification techniques (SVM and KNN) to discover suspicious water fraud clients. The main purpose of this research is to assist the Yarmouk Water Company (YWC) in Irbid, Jordan, in overcoming their earnings shortfall. The SVM-based technique uses customer load profile attributes to expose known abnormal behaviour.
Keywords: SVM, KNN, Prediction.

I. INTRODUCTION

Water theft makes it harder to provide water to all citizens equally and diminishes the organization's revenue. Although the MOI is a non-profit organisation responsible for delivering water to all residents, it is necessary to maintain a balance between expenses and earnings in order to offer an equal water supply to all individuals. "In order to improve income, it is important to precisely quantify the energy expended," hence fraudulent water usage results in an incorrect consumption quantity. According to the MOI, around 8000 structures in the city receive water without the use of a water metre, which the MOI utilises to calculate each client's monthly usage.

Many people who are effected and who wants to stop the this fradulent behoviour in addition to overcome loss in their bussiness due to these fraudulent behaviour created many methods to reduce this behaviour and in every methods there are pros and cons. To decrease the cons and enhance the benefits we created a data mining model to overcome some of disadvantages. The study employs the

The other methods like KNN and ANN has low accuracy and performance compared to SVM. The SVM has high performance and accuracy. The SVM hits detection reate of 1-10% for random data and 80% intelligent detection. The profile of the customer can be changed according to the activity of customer and customer behaviour. In this method the customer data is classified into three types (montly, seosonlly, yearly) according to the customer preferences. The customers historical data is collected and tested to find the behaviour of the customer. This reasearch helps to provide customers load profile by their form of water consumption. The main theme of this research is to find the fraud customer and true customer. In this system customer water consumption are recorded in computer as a financial billing system which generates the water consumption invoice. In this method a utility staff visits customer place records the data and issue the water bill.

II.PROCEDURE

The main motivations of this study is to

1) Assist the MOI in lowering its NTLs in the water distribution industry.
2) Investigate the capacity to detect and identify NTL activities using data mining classification approaches, as well as the difficulties associated with dealing with only 20% of the available customer's load profile data.
3) Fraudulent activity MOI DWTC teams manually and randomly inspect clients on-site so that the intelligent model produced in this research study can forecast suspicious customers and assist them in detecting fraud activities.According to recent data from the water distribution agency,, the financial losses caused by water use result from a large disparity between water well production in the city and water consumption, as seen in Fig 2. In 2011, the difference reached 15 million cubic metres of water, which is considered a water loss.

This technique employs to classify the customers according to their behaviour. The method compares the customer data with already saved data by that it can tell whether the customer is fraud or not. The SVM andKNN classification models are used to classify data these technique has high accuracy and greater performance..

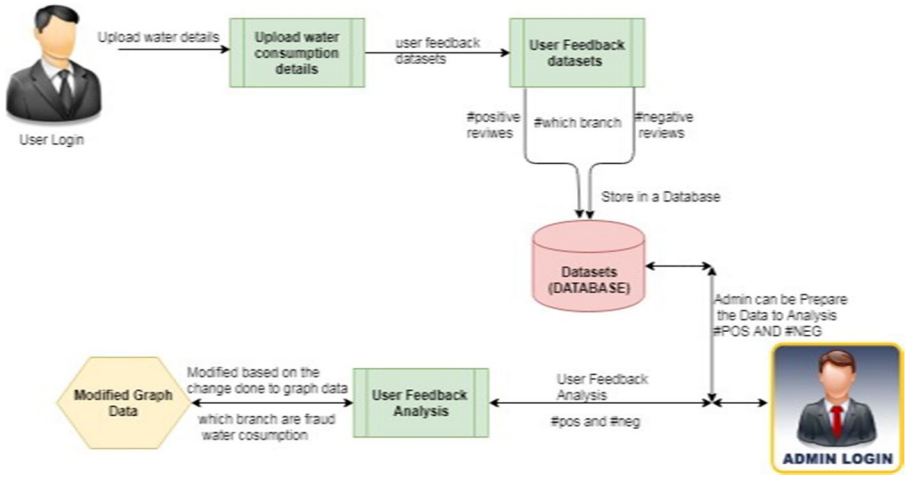


Figure 1 Utility of linear and nonlinear model

Customers who want water through agencies need to register with the system. This registration allows you to upload information about your water consumption. Through this method, field executives monitor the restriction according to branch and collect feedback from clients. Admin can prepare the data so that good and negative feedback can be analysed. Fraud details are blocked by the user based on their feedback, preventing them from receiving any further water. Admin can locate the fraudulent customers and determine the cause of the problem. This will provide a clear view of the present and previous state of the dataset.

A. Algorithm

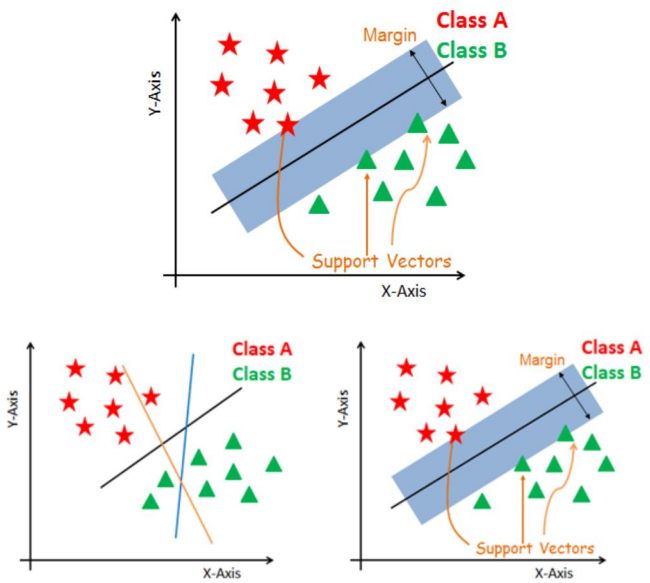
In this work the experiments are performed on two major and well known classification algorithms (K-Nearest Neighbor (KNN)) &SVM are applied to the water customer's dataset)[4] which is taken from the Executives. There accuracy is obtained by evaluating the datasets. Each algorithm has been run over the training dataset and their performance in terms of accuracy is evaluated along with the prediction done in the testing dataset. It is one of the most well-known and widely used open source data mining technologies in the world. It has a comfortable user interface, where in a process view analyses are configured. It uses a modular concept, where respective operators are used in the analysis process. These operators with the help of I/O ports can communicate with the other operators in order to receive input data or pass the data and generated models over to the following operator. The entire analysis process creates a data flow in this way. (K-Nearest Neighbor generates forecasts based on the results of the K closest neighbours.)[5] Therefore, to make predictions with KNN, We'll need to come up with a metric for calculating the distance between the query point and the cases in the examples sample.

1) KNN Algorithm

(Image)[11]

2) SVM Algorithm

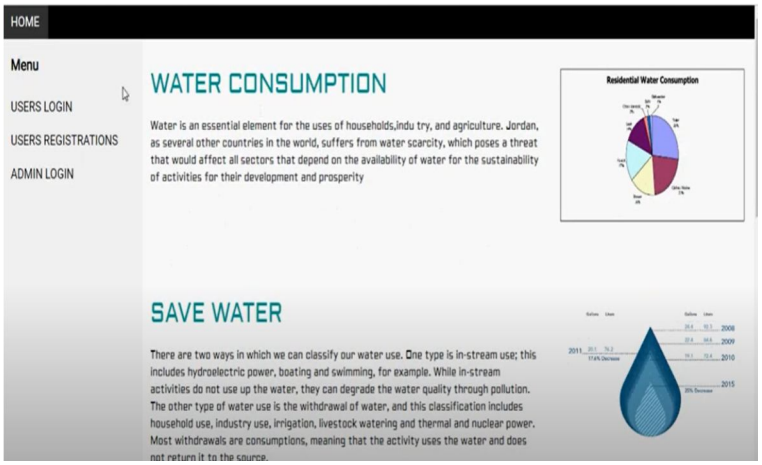
(Image)[10]



B. Modules

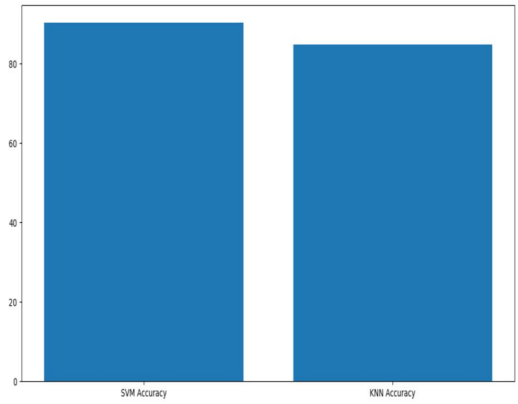
1) *Customer Data:* Customers whowant water through agencies must register with the system. (This registration is the sole option for users to consume water from customers.)[6] Customer requests for water administration and bill generation.
2) *Verify Feedback:* Bills are generated by on-field executives once they have checked the limit. The amount of food they consumed must match the information provided by admin. By this procedure we can check fraud details. After that, the bills were uploaded, and the bogus clients were identified.
3) *Action Against Fraudlent:* Admin can track down fraudulent consumers who illegally drink more water than they need or are entitled to, and their bills can be confirmed. The user can set fraud details to be blocked, preventing them from receiving any more water, and the details can be handed over to cops to be prosecuted..
4) *Graph Analysis:* The graphs help admin grasp the data, and based on this analysis, fraud clients can be identified. The firm steadily improves as they have a better understanding of where the problem emerges and where improvements and deficiencies are needed. This will provide a clear view of the present and previous state of the dataset.

C. Webpage

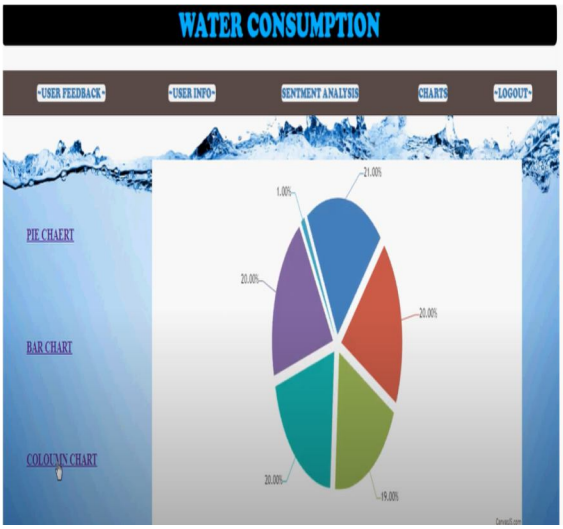


D. Interface Page

E. Graph Analysis



F. Chart Analysis

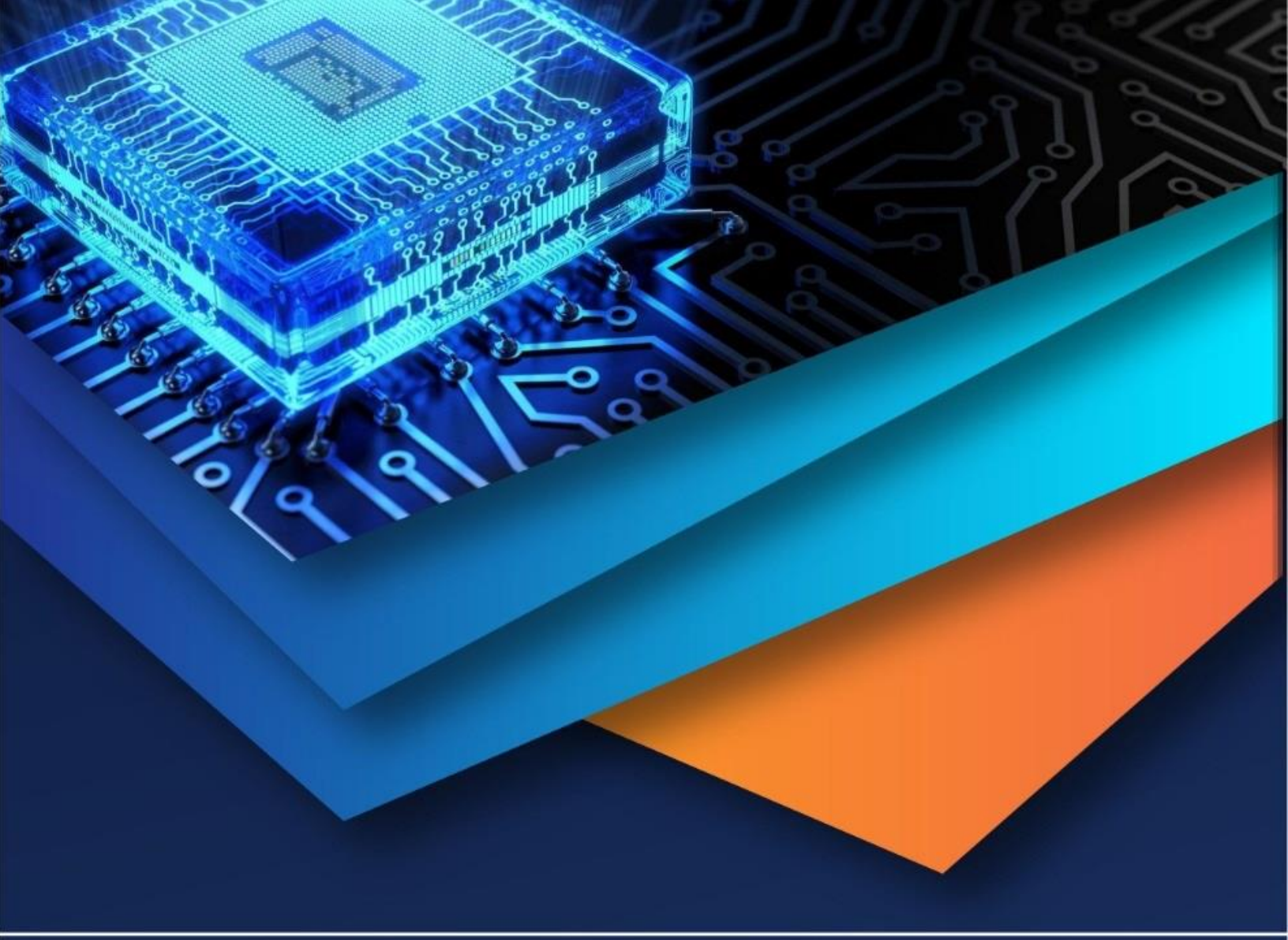


III. CONCLUSION

To reduce the fradulent behaviour in water consumption the SVM and KNN classification models helps. These data mining models has many advatages compared to other methods it increase profits. To reduce fradulent behour we used this data mining technique.

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 Water Demand 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.
[10] https://www.analyticsvidhya.com/blog/2021/06/build-an-image-classifier-with-svm/
[11] https://www.google.com/imgres?imgurl=http%3A%2F%2Fres.cloudinary.com%2Fdyd911kmh%2Fimage%2Fupload%2Ff_auto%2Cq_auto%3Abest%2Fv1531424125%2FKNN_final1_ibdm8a.png&imgrefurl=https%3A%2F%2Fwww.datacamp.com%2Ftutorial%2Fk-nearest-neighbor-classification-scikit-learn&tbnid=GqQENpctyuBCXM&vet=12ahUKEwjl-4arl7L4AhWRxqACHco8DOMQMygCegUIARDZAQ..i&docid=z7VK3BwXsYgp9M&w=591&h=515&q=knn%20algorithm%20images&ved=2ahUKEwjl-4arl7L4AhWRxqACHco8DOMQMygCegUIARDZAQ



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)