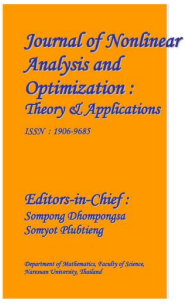
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**



FRAUDULENT BEHAVIOUR DETECTION USING DATA MINING BASED MODEL

Dr. Thanveer jahan

Mrs. G. VijayaLaxmi

Dr. Thanveer jahan, Associate professor, Head of the Department of CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)

Mrs. G. VijayaLaxmi, Assistant professor, CSE(AI&ML), *Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)*

G Kruthan Reddy (20641A6667), UG student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)

K Sai Manish (20641A6679), UG student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)

K Haritej (20641A6681), UG student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)

N Shekar (20641A66A8), UG student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban – 506005(T.S)

ABSTRACT

Fraudulent behavior in drinking water consumption is a significant problem facing water supplying companies and agencies. This behavior 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 behavior 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

Water theft makes it harder to provide water to all citizens equally and diminishes the organizati on's revenue. Although the MOI is a non-profit organisation responsible for delivering water to a ll residents, it is necessary to maintain a balance between expenses and earnings in order to offer an equal water supply to all individuals [1]. "In order to improve income, it is important to precis ely quantify the energy expended," hence fraudulent water usage results in an incorrect consumpt ion 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 [2].

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 red uce this behaviour and in every methods there are pros and cons. To decrease the cons and enhan ce the benefits we created a data mining model to overcome some of disadvantages. The study e mploys the

The other methods like KNN and ANN has low accuracy and performance compared to SVM. T he SVM has high performance and accuracy [3] The SVM hits detection reate of 1-10% for rand om data and 80% intelligent detection. The profile of the customer can be changed according to t he activity of customer and customer behaviour. In this method the customer data is classified int o 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 t

o provide customers load profile by their form of water consumption [4]. The main theme of this research is to find the fraud customer and true customer. In this system customer water consumpt ion are recorded in computer as a financial billing system which generates the water consumptio n invoice. In this method a utility staff visits customer place records the data and issue the water bill.

2. LITERATURE SURVEY

Water supply businesses and organizations have a big problem with fraudulent drinking water use. This conduct causes a significant loss of money and accounts for the majority of nontechnical losses. In recent years, finding effective measures for identifying fraudulent activity has been a hot topic of research. Water supply businesses can use intelligent data mining tools to detect fraudulent activity and decrease losses. The use of two classification approaches (SVM and KNN) to detect suspected fraud water consumers is investigated in this study. The major goal of this study is to help the Yarmouk Water Company (YWC) [6] in Jordan's Irbid overcome its earnings deficit. Customer load profile characteristics are used in the SVM-based method to highlight aberrant behaviour that is known to be associated with non-technical loss activities. The information was gathered from the company's billing system's historical data. The developed model's accuracy was above 74%, which is significantly higher than the YWC's existing manual prediction processes. A decision tool based on the produced model was created to deploy the model. The technology will assist the firm in anticipating questionable water consumers for on-site inspection [7].

Drinking water fraud is a major issue for water delivery businesses and authorities. This behavior results in a significant loss of income and is responsible for the bulk of non-technical losses. Finding appropriate criteria for detecting fraudulent behaviors 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 behavior [10].

Spurious(dishonorable) behavior in drinking water utilization is a compelling problem facing water supplying companies and agencies. This behaviour results in a enormous loss of income and forms the highest percentage of non-technical loss. Finding competent analysis for identifying spurious activities has been an active research area in recent years. Intelligent data mining techniques can service water supplying companies to detect these spurious activities to decrease such losses. This analysis explores the use of two distribution techniques (SVM and KNN) to identify suspicious cheat water customers. The main encouragement of this analysis is to benefit Yarmouk Water Company (YWC) in Irbid city of Jordan to affect its profit loss. The SVM based approach uses customer load profile characteristics to display anomalous behaviour that is well-known to be interacted with non-technical loss activities. The data has been collected from the ancient data of the company billing system. The efficiency of the developed model hit a rate of over 74% which is better than the present standard prediction steps taken by the YWC. To expand the model, a opinion tool has been built using the develop model. The system will help the company to predict careful water customers to be authorized on site.

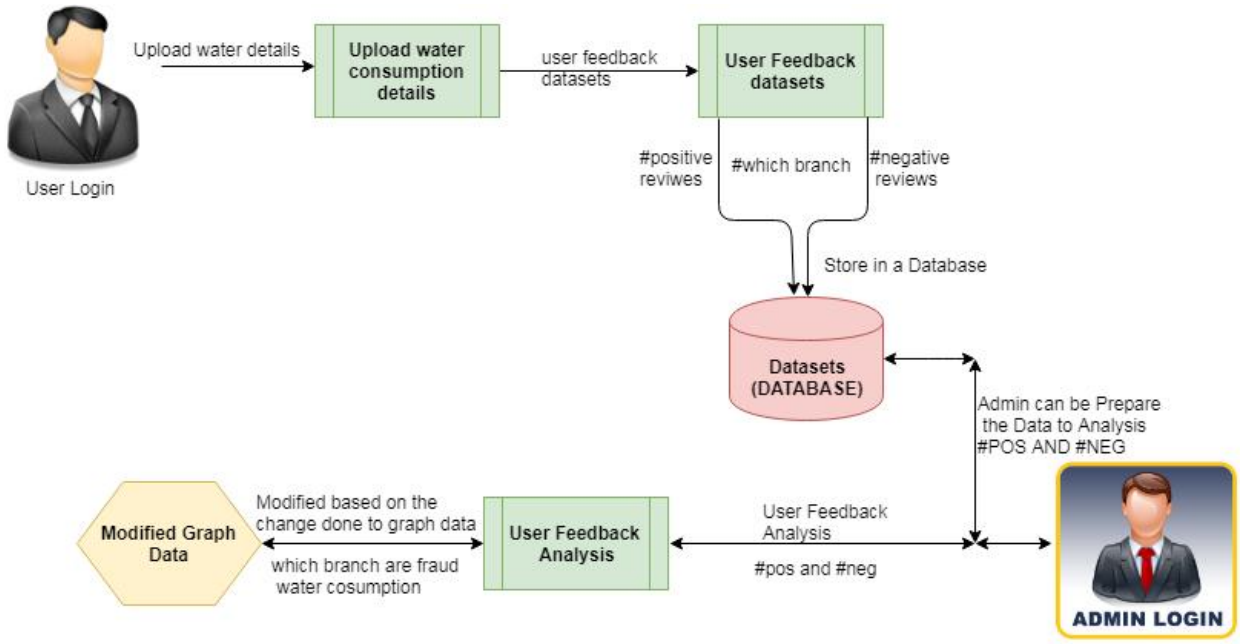
3. PROBLEM STATEMENT

Literature has abundant research for Non-Technical Loss (NTL) [5] 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) [6] 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 tocitizens 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.

4. 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) [7]to build a suitable model to detect suspicious fraudulent customers, depending on their historical water metered consumptions.The CRISP-DM [7] (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 [8] 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 zeroconsumption 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 [9].

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

6.1 CUSTOMER DATA

The customers those who are willing to get water through agencies are registered with system. The only ways for user to consume water by customers are through this registration. Customer request for admin for water and to generate bills.

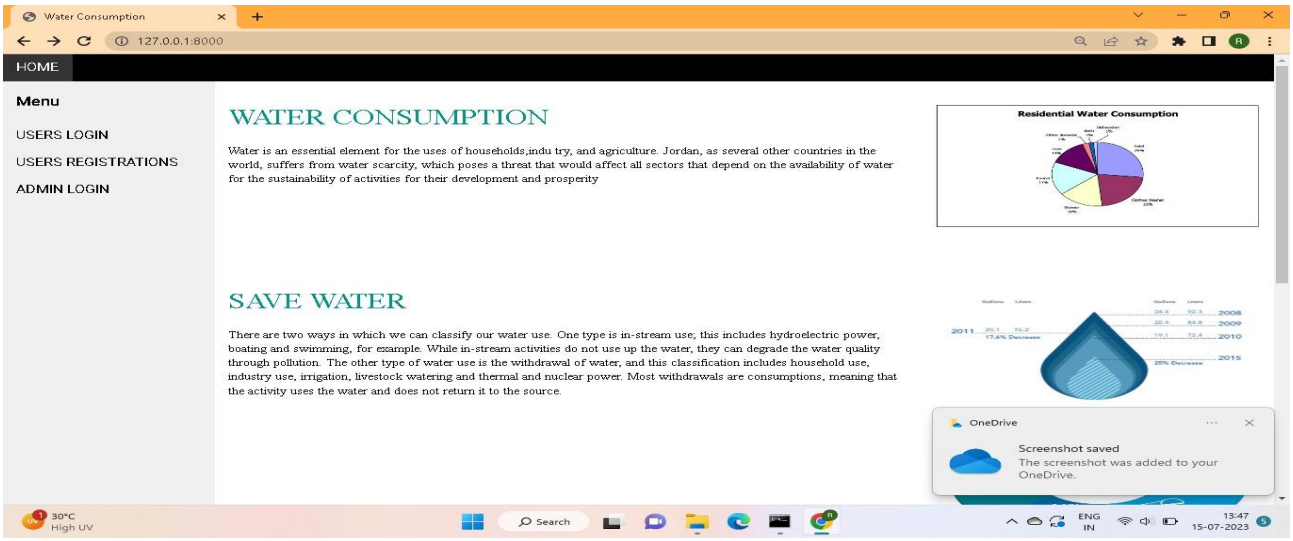
6.2 VERIFY FEEDBACK

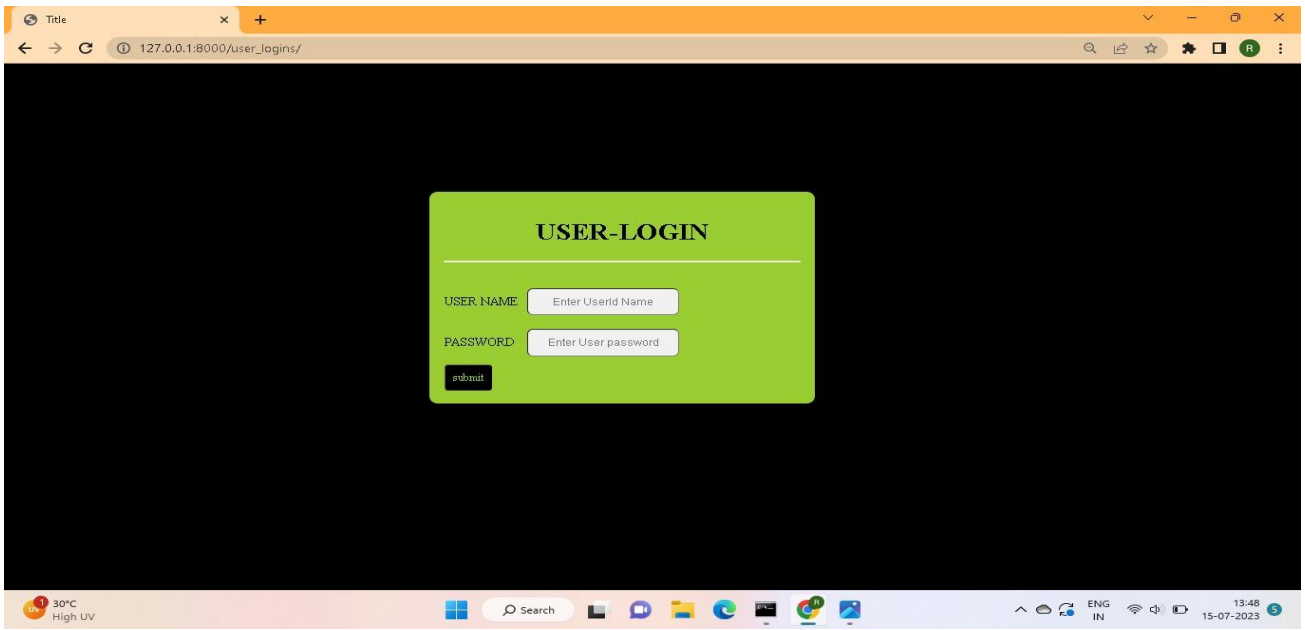
Bills are generated after checking the limit by on field executives after check the limit. The quantity that they consumed must be equal to noted details by admin. The fraud details can be check through this process. The bills were uploaded after this and find the fraudulent among the customers.

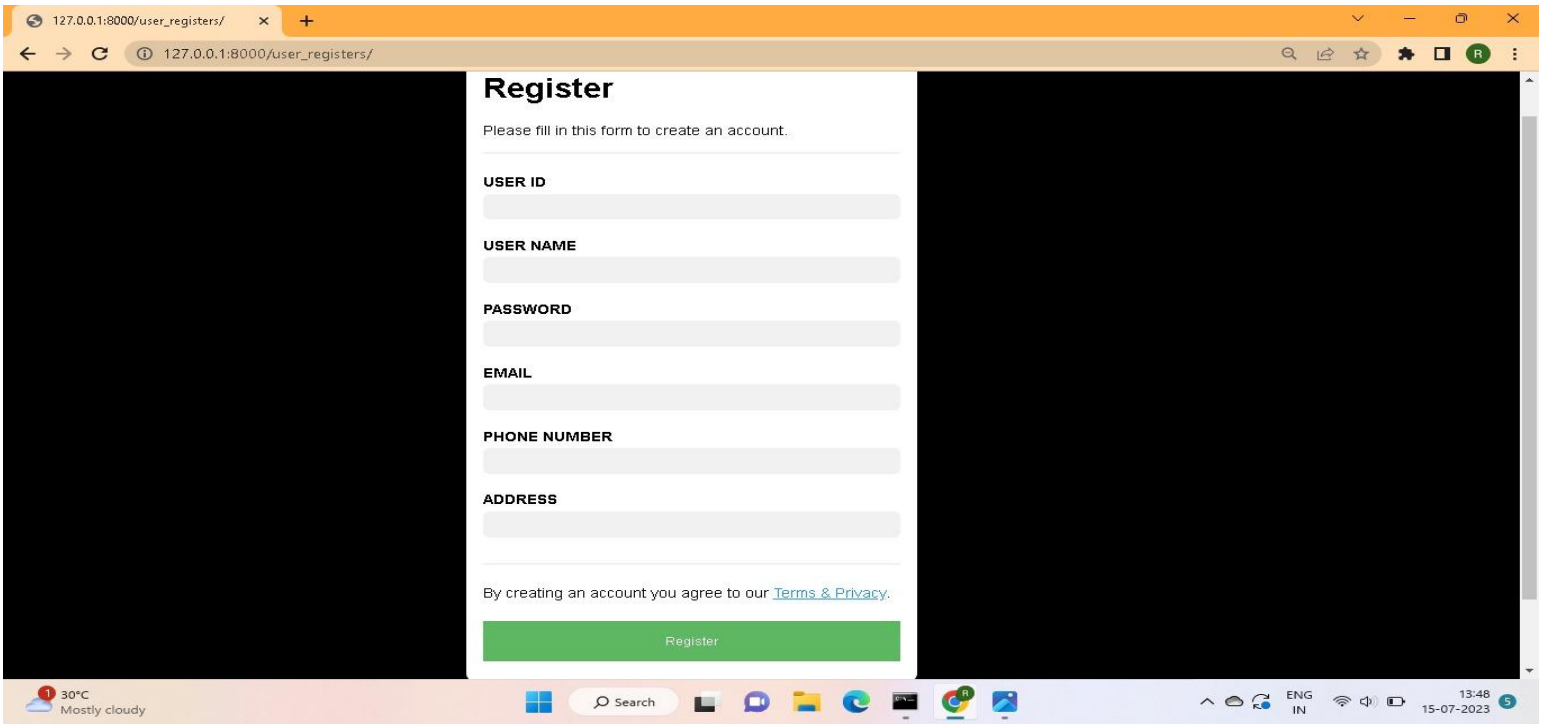
6.3 ACTION AGAINST FRAUDLENT

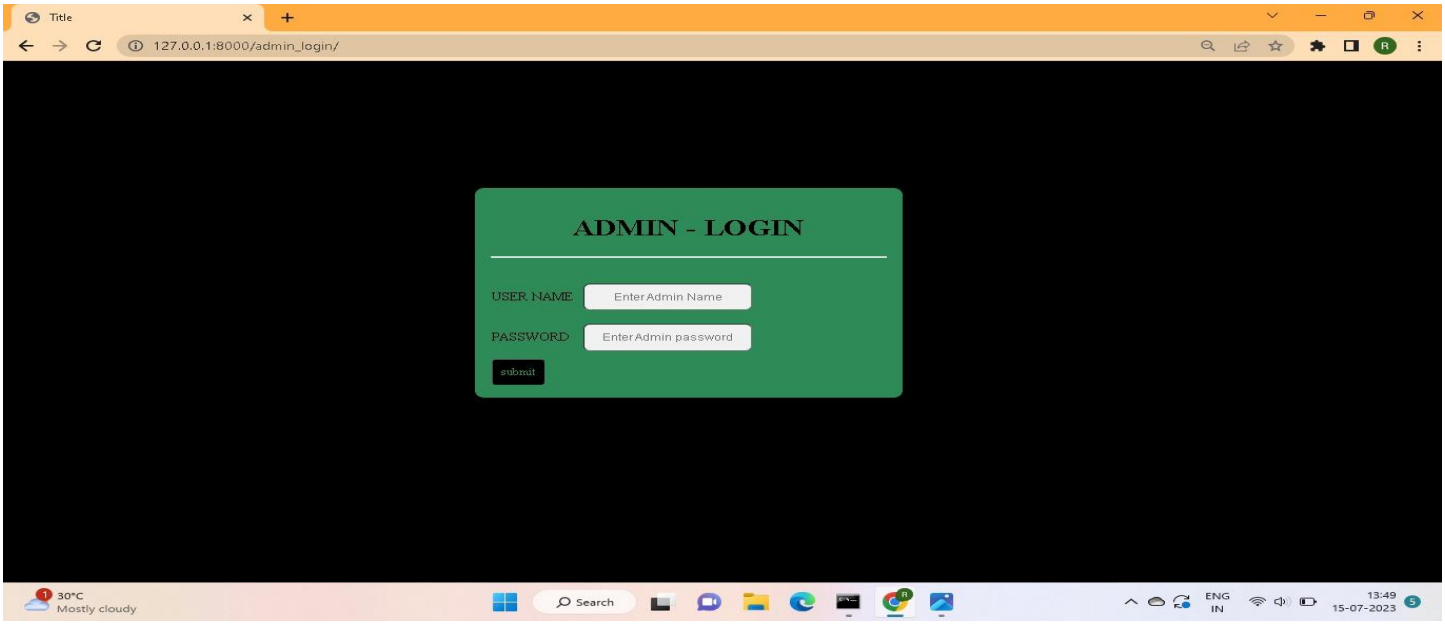
The fraud customers who illegally consumes more water than they used or may be requires can be found by admin and bills also verified by them. Fraud details are set to block by the user and let them not provide any more water to them again and the details handover to cops to punish them with legally.

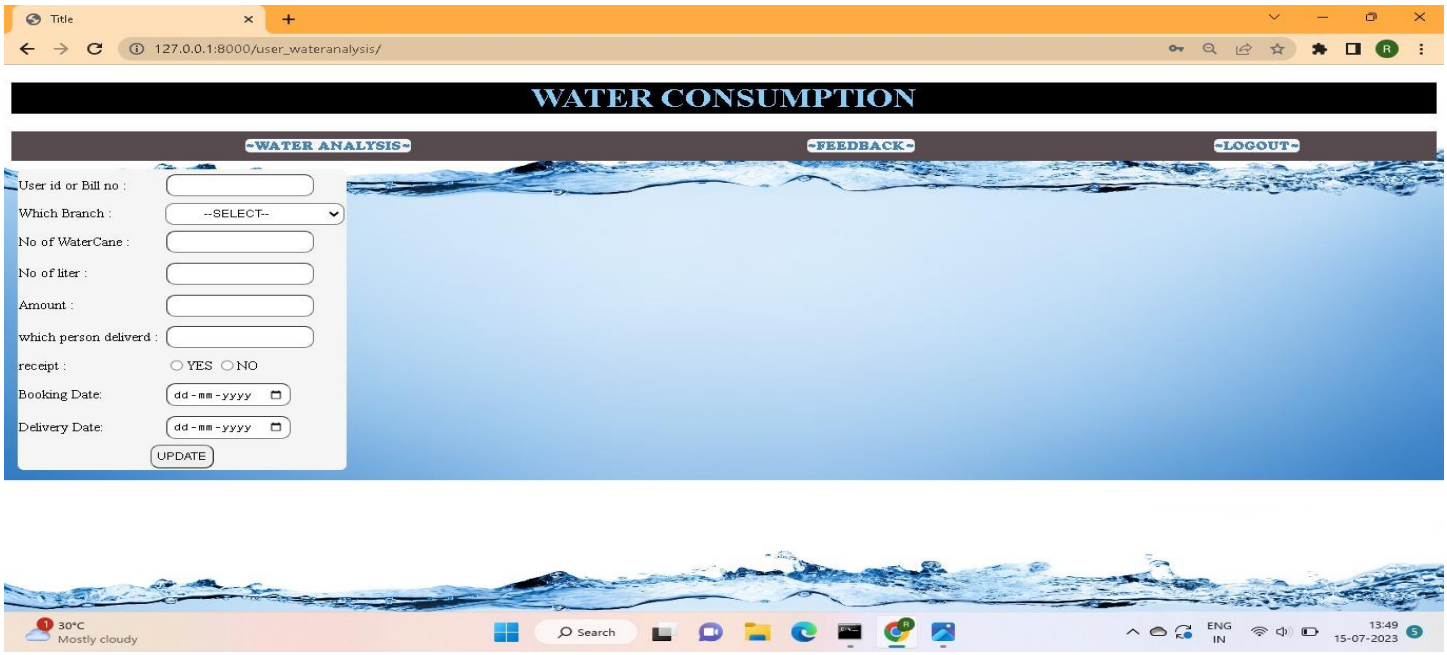
7. OUPUT RESULT

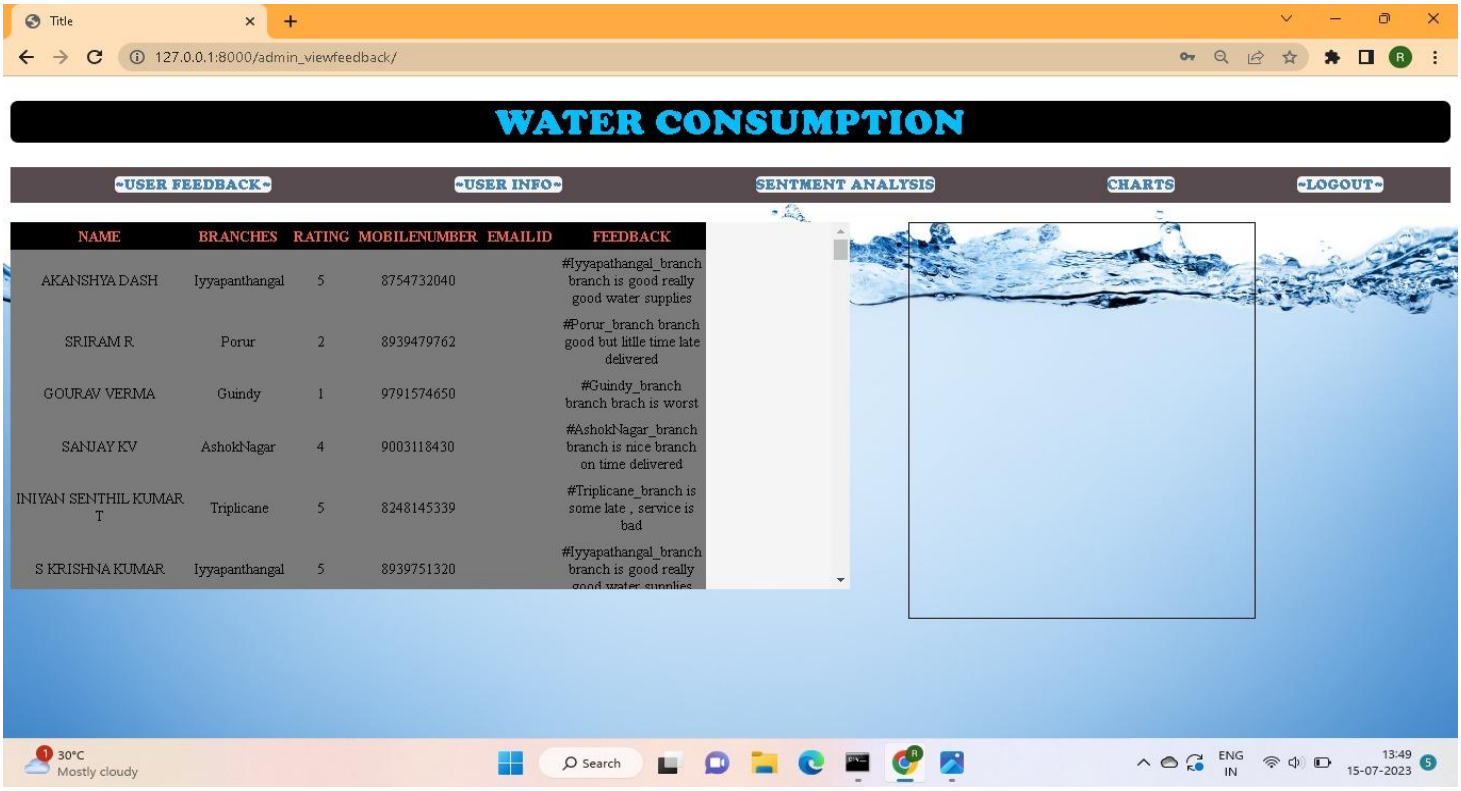


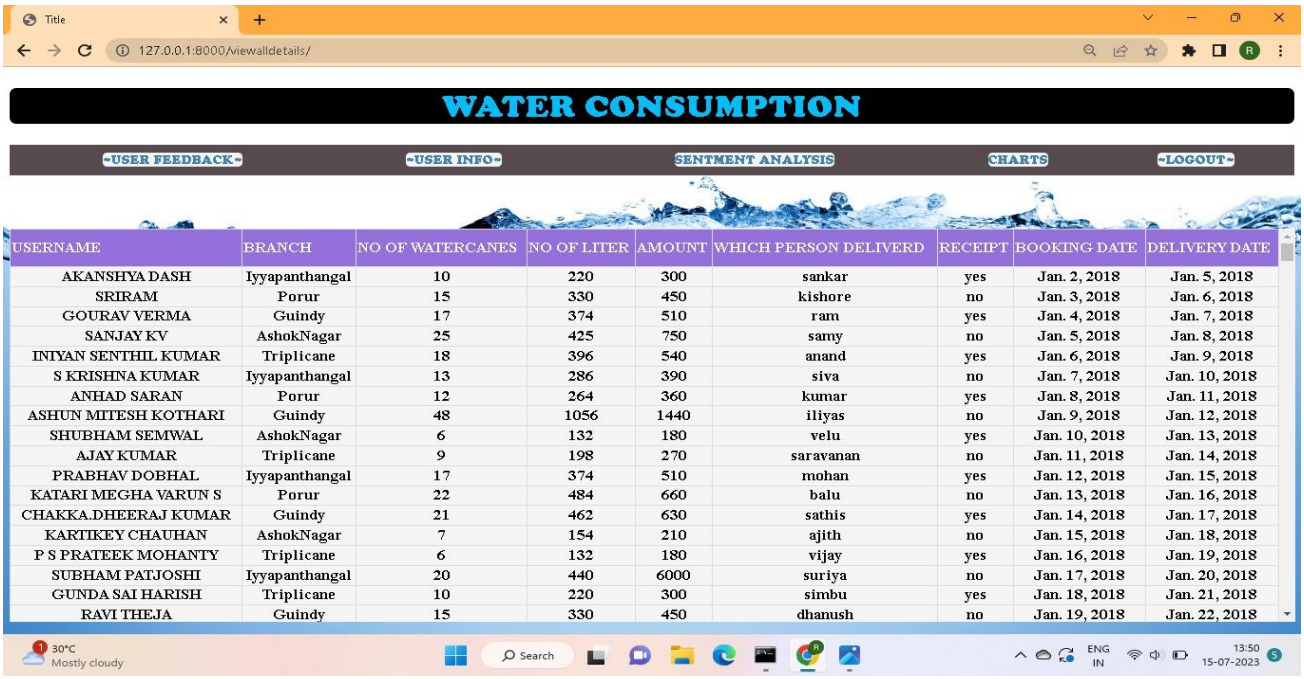



127.0.0.1:8000/user_registers/
Register
Please fill in this form to create an account.
USER ID
USER NAME
PASSWORD
EMAIL
PHONE NUMBER
ADDRESS
By creating an account you agree to our Terms & Privacy.
Register

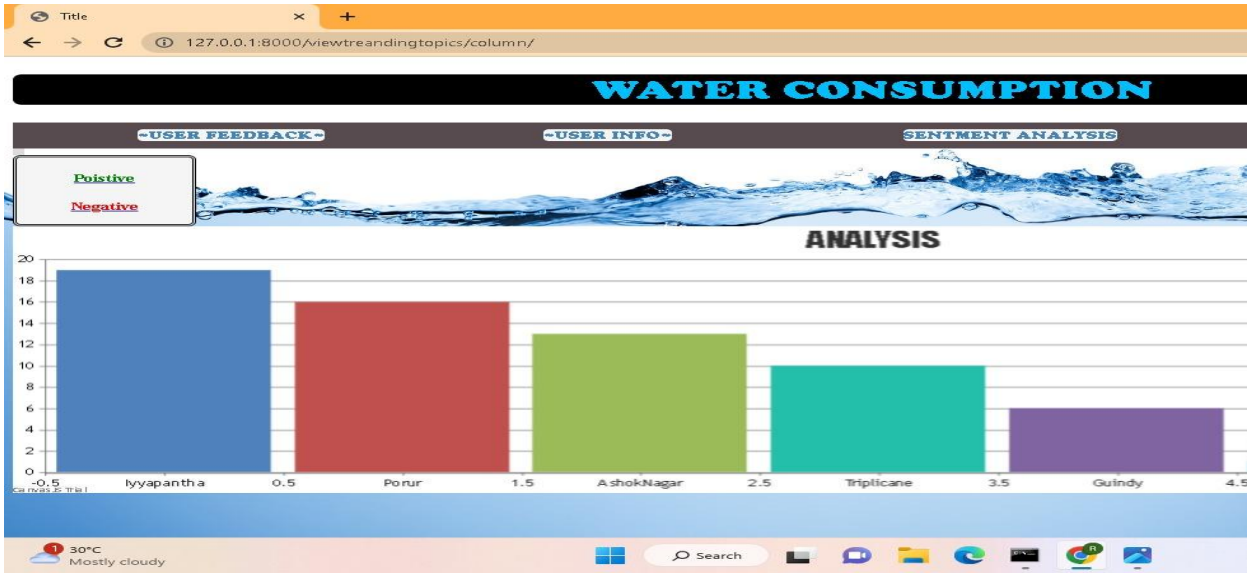

Title
127.0.0.1:8000/admin_login/
ADMIN - LOGIN
USER NAME
Enter Admin Name
PASSWORD
Enter Admin password
submit


Title
127.0.0.1:8000/user_wateranalysis/
WATER CONSUMPTION
WATER ANALYSIS
FEEDBACK
LOGOUT
User id or Bill no :
Which Branch :
--SELECT--
No of WaterCane :
No of liter :
Amount :
which person deliverd :
receipt :
YES NO
Booking Date:
dd-mm-yyyy
Delivery Date:
dd-mm-yyyy
UPDATE
30°C
Mostly cloudy
Search
ENG
IN
13:49
15-07-2023


Title
127.0.0.1:8000/admin_viewfeedback/
WATER CONSUMPTION
USER FEEDBACK
USER INFO
SENTMENT ANALYSIS
CHARTS
LOGOUT
NAME BRANCHES RATING MOBILENUMBER EMAILID FEEDBACK
AKANSHYA DASH Iyyapanthangal 5 8754732040 #Iyyapathangal_branch branch is good really good water supplies
SRIRAM R Porur 2 8939479762 #Porur_branch branch good but litlle time late delivered
GOURAV VERMA Guindy 1 9791574650 #Guindy_branch branch brach is worst
SANJAY KV AshokNagar 4 9003118430 #AshokNagar_branch branch is nice branch on time delivered
INIYAN SENTHIL KUMAR T Triplicane 5 8248145339 #Triplicane_branch is some late , service is bad
S KRISHNA KUMAR Iyyapanthangal 5 8939751320 #Iyyapathangal_branch branch is good really
30°C
Mostly cloudy
Search
ENG
IN
13:49
15-07-2023



USERNAME	BRANCH	NO OF WATERCANES	NO OF LITER	AMOUNT	WHICH PERSON DELIVERD	RECEIPT	BOOKING DATE	DELIVERY DATE
AKANSHYA DASH	Iyyapanthangal	10	220	300	sankar	yes	Jan. 2, 2018	Jan. 5, 2018
SRIRAM	Porur	15	330	450	kishore	no	Jan. 3, 2018	Jan. 6, 2018
GOURAV VERMA	Guindy	17	374	510	ram	yes	Jan. 4, 2018	Jan. 7, 2018
SANJAY KV	AshokNagar	25	425	750	samy	no	Jan. 5, 2018	Jan. 8, 2018
INIYAN SENTHIL KUMAR	Triplicane	18	396	540	anand	yes	Jan. 6, 2018	Jan. 9, 2018
S KRISHNA KUMAR	Iyyapanthangal	13	286	390	siva	no	Jan. 7, 2018	Jan. 10, 2018
ANHAD SARAN	Porur	12	264	360	kumar	yes	Jan. 8, 2018	Jan. 11, 2018
ASHUN MITESH KOTHARI	Guindy	48	1056	1440	iliyas	no	Jan. 9, 2018	Jan. 12, 2018
SHUBHAM SEMWAL	AshokNagar	6	132	180	velu	yes	Jan. 10, 2018	Jan. 13, 2018
AJAY KUMAR	Triplicane	9	198	270	saravanan	no	Jan. 11, 2018	Jan. 14, 2018
PRABHAV DOBHAL	Iyyapanthangal	17	374	510	mohan	yes	Jan. 12, 2018	Jan. 15, 2018
KATARI MEGHA VARUN S	Porur	22	484	660	balu	no	Jan. 13, 2018	Jan. 16, 2018
CHAKKA.DHEERAJ KUMAR	Guindy	21	462	630	sathis	yes	Jan. 14, 2018	Jan. 17, 2018
KARTIKEY CHAUHAN	AshokNagar	7	154	210	ajith	no	Jan. 15, 2018	Jan. 18, 2018
P S PRATEEK MOHANTY	Triplicane	6	132	180	vijay	yes	Jan. 16, 2018	Jan. 19, 2018
SUBHAM PATJOSHI	Iyyapanthangal	20	440	6000	suriya	no	Jan. 17, 2018	Jan. 20, 2018
GUNDA SAI HARISH	Triplicane	10	220	300	simbu	yes	Jan. 18, 2018	Jan. 21, 2018
RAVI THEJA	Guindy	15	330	450	dhanush	no	Jan. 19, 2018	Jan. 22, 2018



8. 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 data to fit the SVM and KNN data mining classifiers.

9. FUTURESCOPE

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.

10. 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/