iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46783

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 IX Sep 2022- Available at www.ijraset.com

The Data Mining Based Model for Detection of Fraudulent Behavior in Water Consumption

Ch. Veena[1], I. Shweshitha[2], N. Nitheesha[3], Harsha Priya Raulo[4]
[1]*Assistant Professor,* [2, 3, 4]*UG Students, Department of CSE, Sridevi Women's Engineering College, Hyderabad, T.S, India.*

Abstract: ***There are many significant problems facing by water supplying companies and agencies because of fraudulent water consumption. Which is resulting a higher loss of income to water supplying agencies. Finding efficient measurements for detecting fraudulent activities has beenan active research area in recent years. To detect this fraudulent behaviour faced by water companies' intelligent datamining techniques can be used to reduce the loss. This research explores the use of two classification techniques SVM and KNN to detect suspicious fraud water customers. 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 obtained 74% which is better than the current manual prediction procedures.The system will help the company to predict suspicious water customers.***

I. INTRODUCTION

Water is necessary for residential, industrial, and agricultural purposes. Many nations across the world, has water shortage and loss due to fraud behaviour. There are two kinds of loss facing by the water supplying companies one is technical loss and the other is non-technical loss. Non-Technical losses are irregularities resulting from electricity theft and other consumer misbehaviour are a concern.

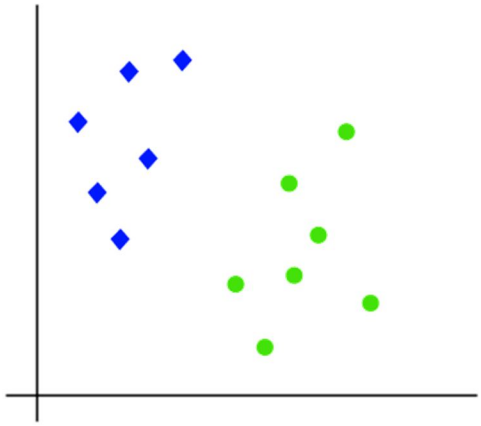
The following activities are included in NTLs:

1) Losses caused by malfunctioning meters and equipment.
2) Tampering with meters to make them reflect low usage rates.
3) Bribing meter readers to take false readings.

II. COMPONENTS

A. Support Vector Machine

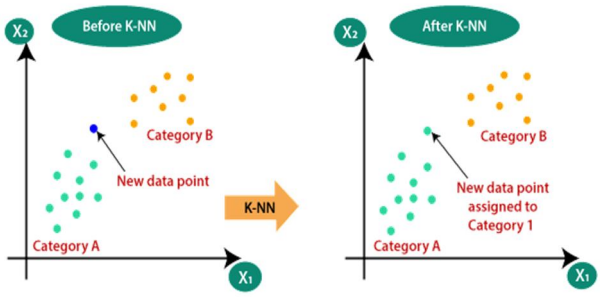
Support Vector Machine (SVM) is one of the most robust and accurate methods in all machine-learning algorithms. It primarily includes Support Vector Classification (SVC) and Support Vector Regression (SVR). The SVC is based on the concept of decision boundaries. A decision boundary separates a set of instances having different class values between two groups. The SVC supports both binary and multi-class classifications. The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue. Consider the below image:



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

B. K-Nearest Neighbor

The KNN classifier is based on a distance function that measures the difference or similarity between two instances. The standard Euclidean distance d(x, y) between two instances x and is defined as : n 2 k k k=1 d(x,y)= (x -y) where, xk is the kth featured element of instance x, yk is the kth featured element of the instance y and n is the total number of features in the dataset. Assume that the design set for KNN classifier is U. The total number of samples in the design set is S. Let C = {C1 ,C2 ,…CL} are the L distinct class labels that are available in S. Let x be an input vector for which the class label must be predicted. Let yk denote the kth vector in the design set S. The KNN algorithm is to find the k closest vectors in design set S to input vector x. Then the input vector x is classified to class Cj if the majority of the k closest vectors have their class as Cj.Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:



C. PYTHON

Python is a general-purpose interpreted, interactive, object-oriented, and high level programming language. An integrated language, Python has a design philosophy that emphasizes code readability, and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++ or java. It provides constructs that enable clear programming on both small and large 4 The data mining model based on detection of fraudulent behavior in water consumption scales. Python interpreters are available for many operating system. Cpython, the reference implementation of Python, is open source software and has a community based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit python software foundation.

D. Django

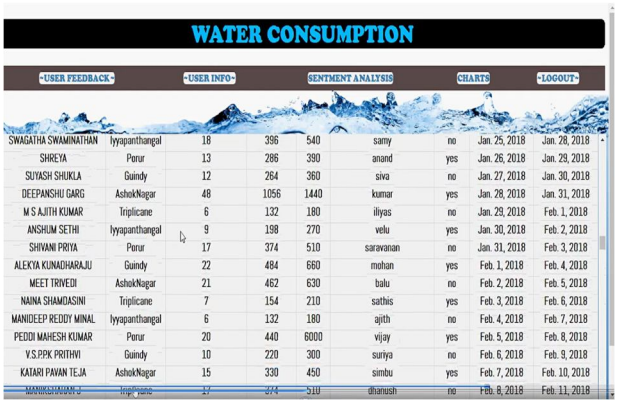
Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source. Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development, and the principle of don’t repeat yourself. Python is used throughout, even for settings files and data models.

III. PROPOSED SYSTEM

In this project, we will use data mining classification approaches to discover consumers that engage in fraudulent water use behavior. This project focuses on customer’s historical data which are selected from the YWC billing system. The main objective of this work is to use best 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.

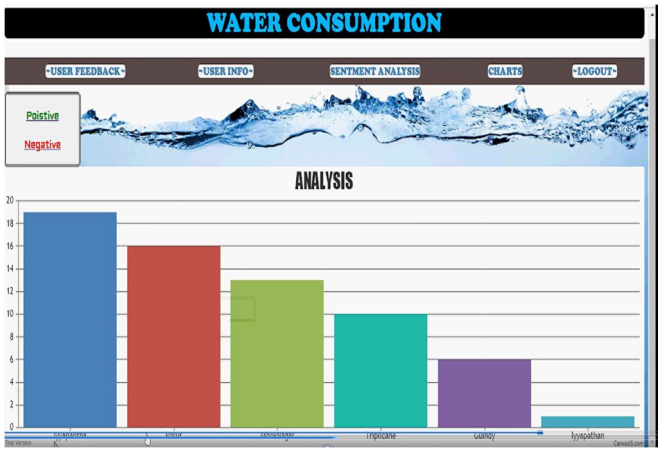
IV. RESULTS

Two types of unauthorised water consumption have been identified: Illegal connections to the water network (especially in the detached houses of the Rural District) and practices to handle the meter tempering with the installation of a magneton. Besides, according to the comments of the technicians of AMAEM, they are detecting new ways of handling.



SWAGATHA SWAMINATHAN	Iyyapanthangal	18	396	540	samy	no	Jan. 25, 2018	Jan. 28, 2018
SHREYA	Porur	13	286	390	anand	yes	Jan. 26, 2018	Jan. 29, 2018
SUYASH SHUKLA	Guindy	12	264	360	siva	no	Jan. 27, 2018	Jan. 30, 2018
DEEPANSHU GARG	AshokNagar	48	1056	1440	kumar	yes	Jan. 28, 2018	Jan. 31, 2018
M S AJITH KUMAR	Triplicane	6	132	180	iliyas	no	Jan. 29, 2018	Feb. 1, 2018
ANSHUM SETHI	Iyyapanthangal	9	198	270	velu	yes	Jan. 30, 2018	Feb. 2, 2018
SHIVANI PRIYA	Porur	17	374	510	saravanan	no	Jan. 31, 2018	Feb. 3, 2018
ALEKYA KUNADHARAJU	Guindy	22	484	660	mohan	yes	Feb. 1, 2018	Feb. 4, 2018
MEET TRIVEDI	AshokNagar	21	462	630	balu	no	Feb. 2, 2018	Feb. 5, 2018
NAINA SHAMOASINI	Triplicane	7	154	210	sathis	yes	Feb. 3, 2018	Feb. 6, 2018
MANIDEEP REDDY MINAL	Iyyapanthangal	6	132	180	ajith	no	Feb. 4, 2018	Feb. 7, 2018
PEDDI MAHESH KUMAR	Porur	20	440	6000	vijay	yes	Feb. 5, 2018	Feb. 8, 2018
V.S.PPK PRITHVI	Guindy	10	220	300	suriya	no	Feb. 6, 2018	Feb. 9, 2018
KATARI PAVAN TEJA	AshokNagar	15	330	450	simbu	yes	Feb. 7, 2018	Feb. 10, 2018
MANIKSHAMM J	Triplicane	17	374	510	dhanush	no	Feb. 8, 2018	Feb. 11, 2018

The above results are customer details who are consuming the water.



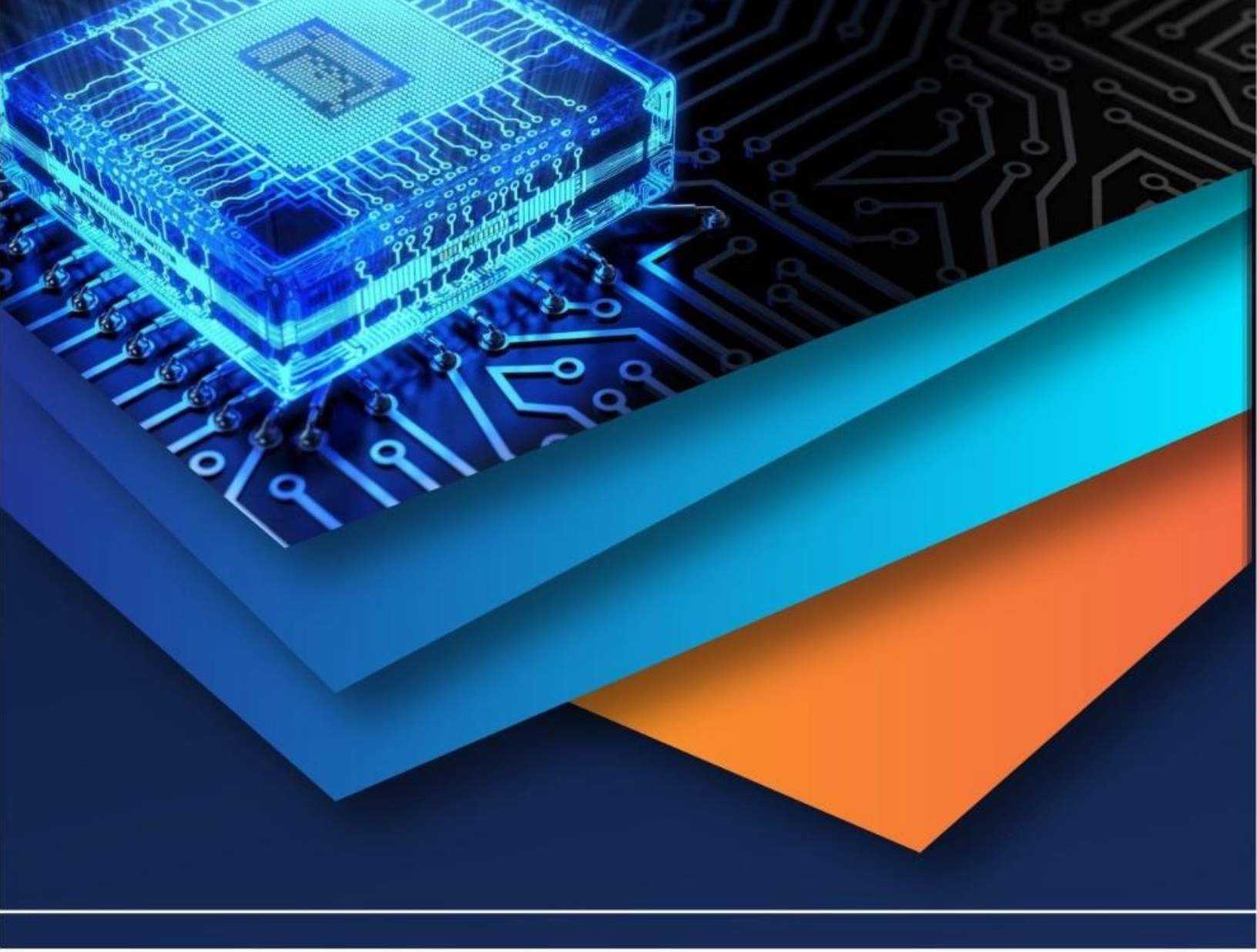
The results are analysis of consumption of water who are giving positive feedback according to area wise and number of contracts) makes it possible to incorporate new qualitative and quantitative elements in the analysis of this process. As said before, in the relationship between domestic contracts and unauthorised domestic water consumption, the lowest values, that is, the highestlevel of fraud, are recorded in the North District (115 contracts per fraud), which would corroborate the dynamics described. However, in second place is the Rural District (detached houses) with 208 contracts per fraud, which situates it closer in percentage terms to the North District than to the Beach Area. In the Beach Area (low density urban model), to detect a fraud, 6992 domestic contracts are required, the highest of the city, which would corroborate the relationship between higher level of income, lower level of fraud.

V. CONCLUSION

The analysis of unauthorised water consumption offers interesting information regarding the current status and characteristics of the management and control of water consumption ina city. However, it is a subject that has rarely been considered in socio-regional studies since ithas traditionally been analysed from a strictly economic or technical perspective, It aims to contribute a socio-economic and regional view to the international and national scientific literature analysing the causes that explain unauthorised domestic water consumption, since itis one of the few studies carried out on this topic and the first in the area under study. Besides, with this study some factors could be taking into account in other cities for improve the management and the reduction of the unauthorised consumption such as the use of treated waterand the implementation of the technology (smart meters).

REFERENCES

[1] 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
[2] E. Kirkos, C. Spathes and Y. Manolo Poulos, Data mining techniques for the detection of fallacious money statements, skilled Systems with Applications, 32(2007): 9951003
[3] Software Engineering- Sommerville, 7th edition, Pearson Educatio[4]The unified modeling language user guide Grady Booch, James Rambaugh, IvarJacobson, Pearson Education



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY