Datamining Based Model for Detection of Duplicitous Behaviour in Water Management

M. Geetha Lakshmi Priyanka[1], K. Serena Evanglin[2], J. Pavani Sri Sai[3] & Dr. G. Murali[4]

[1 2 3] *B.Tech Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, AP, India.*

[4]*Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, AP, India.*

mglpriyanka98@gmail.com[1], sereblessy@gmail.com[2], pavanisrisai.j@gmail.com[3], m_gudipati@yahoo.com[4]

Abstract

The main Artificial Intelligence techniques used for duplicitous detection include Data mining. It is used to classify, cluster and segment the data. It automatically finds associations and rules in the data. These patterns may include those related to fraud. We can use some techniques like SVM (Support Vector Machine) and KNN (K Nearest Neighbour) to identify characteristics of fraud. Duplicitous (Dishonest) water management is a problem faced by all water companies worldwide. It can be solved by using some techniques, which are discussed in this paper.

Keywords: SVM, KNN, Duplicitous, Water management, Fraud.

I. Introduction

Water is the basic necessity for existence of all life forms on the earth. It's been important for people from thousands of years, without water there would be no life on earth. For every household need we use water as main source. Drinking water supply is a big business in our country where many people depend on it. But water theft is a big problem at many places, where customers book water amount and they get the less amount than booked and the bills get tampered. The distribution of water from water supplying agency to the customer is not direct, there is a distributor/dealer in middle to supply the water to a particular customer. In this process we come across many losses. These are divided into two, one is technical loss and the other is non-technical loss.

Technical loss is completely regarding the production system in water supplying agency like leakage due to breakout of pump or it may be any network issue or it may be anything related to the engine or power supply for production of water in the water supplying agency. The below mentioned irregular activities are known as Non-technical losses. Non-Technical losses originate from water theft where dealers deliver the amount of water less than the customer ordered for and also there is no correct billing system.

Non-technical loses include the following irregularities:

1. Losses due to duplicitous dealers.
2. Tampering of meters, so that the meters show less amount of water utilization.
3. Ignorance of unpaid bills.
4. Customers feedback is not taken regularly.
5. Distributors are bribing the meter readers for arranging false readers.

II. Literature survey

A systematic and thorough search of all published papers as well as other sources including thesis, says that, the data mining process helps in extracting and analyzing various data patterns, information or trends from large databases. This research explores the use of two classification techniques (SVM and KNN) to detect duplicitous water dealers. The SVM based approach uses the details provided by user to expose duplicitous behavior that is known to be similar to non-technical loss activities."

In the existing system, customer books amount of water which is delivered him through the dealer. Here, billing is not done properly. False meter readings are shown and amount is collected for more litres than what they have delivered.

Proposed system:

Here, we propose a direct communication between the customer and water supplying agency which is through a web interface. In this web interface we have 2 other sub interfaces, one is user interface in which user provides all the required details. The other is admin interface which can be accessed only by admin, he is able to analyse the user provided data.

III. Proposed work

In this project, we can identify the dealer with Duplicitous behaviour in the water management by using some data mining classification techniques. Here, the entire user provided data is stored in the database. We apply the classification techniques like SVM for linearly separable data and KNN for linearly inseparable data in order to classify the provided feedback as Positive feedback or Negative Feedback. The model is trained using the default classifiers parameters.

This model requires the following user interactions:

1) Upload Service details
2) Provide Genuine Feedback

The admin is able to view all the details provided by the user and he performs some analysis like Service Analysis, Feedback Analysis and Graph Analysis in order to find the duplicitous dealer of the water management. This model helps us to detect the duplicitous dealer by providing the data given by user in form of patterns. The result inspection is done by using the confusion matrix.

a) Architecture:

In this model, we have 4 modules. One is registration module where the user can create his own account by completing the registration process. The second one is feedback module, in this module the user has to give feedback regarding the dealer who delivered the water. The third module is billing module, here the customer will pay the bill directly to agency for amount of water he asked. The last module is admin module only admin can login into that and he will get graphs of different areas how their dealers are working and he can check the feedback of the users.

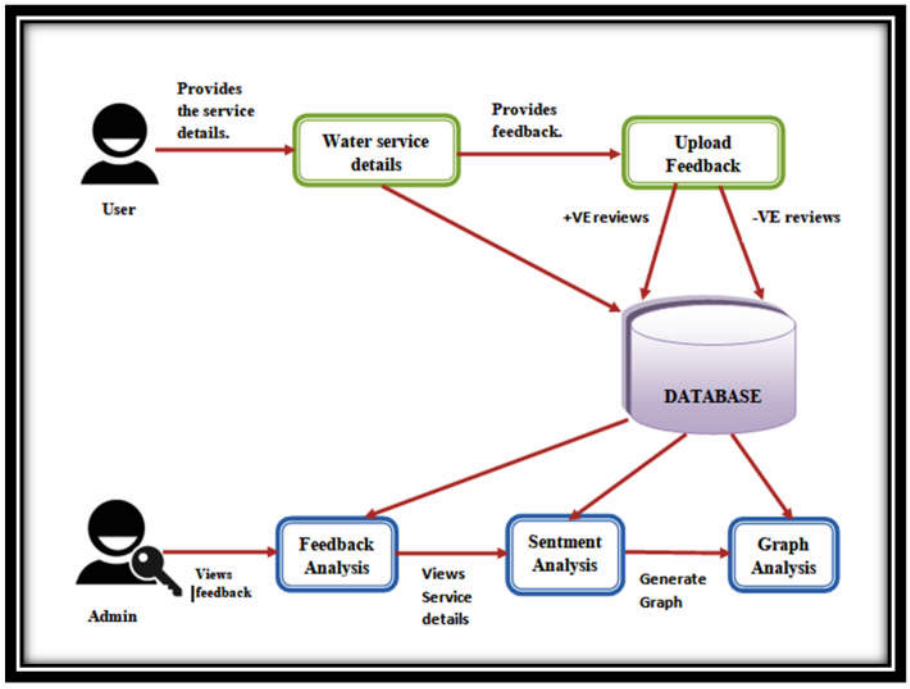

Figure 1: Architecture of the System

b) Algorithms:

In this work, we have used two classification algorithms K-Nearest Neighbor and Support Vector Machine. These algorithms are the world's commonly used data mining algorithms for the purpose of classification.

These algorithms are applied to the customer's dataset to classify them depending on some labels.

i. Support Vector Machine(SVM):

Binary data separation is done using SVM. Unlabeled classes can also be predicted if the model gets trained.

Input:

Data: Dataset with P variables and binary outcome*

Output: Ranked list of variables according to their relevance

Find the values for tuning parameters of the SVM model;

Train the SVM model;

P←p;*

While** p≥2 **do

SVMp ←SVM with optimized tuning parameters for the p variables and observations in Data;

wp ←calculate the weight vector of the SVMp(wp1,……,wpp)

rank.criteria ←(w_{p1}^2*,……..,* w_{pp}^2*);*

min.rank.criteria ←variable with lowest value in rank.criteria vector;

*Remove min.rank.criteria from **Data**;*

$Rank_p$ *←min.rank.criteria;*

P ←p-1;

***end**;*

$Rank_1$ *← variable in* $\boldsymbol{Data} \notin (Rank_2, \ldots\ldots, Rank_{p*})$*;*

***return** (*$Rank_1, \ldots\ldots\ldots, Rank_{p*}$*);*

ii. K Nearest Neighbor(KNN):

KNN uses distance as a metric for data classification. Generally, it is used as a multi-class classifier. Here, the distance metric is calculated whenever we come across unlabeled data.

***TRAIN-**kNN(C,D)*

1 D'←PREPROCESS(D')
2 K←SELECT-K(C,D')
*3 **return** D',K*

***Apply**-kNN(C,D',k,d)*

1 Sk← ComputeNearestNeighbours(D ',k,d)
*2 **for each** cj ∈ C*
*3 **do** pj ←| Sk ∩ cj |/k*
*4 **return** arg maxj pj*

c) Technologies:

We used the following technologies in our project to build the model more efficiently and user friendly.

Python: It is a simple and most populous language. It is mainly used in server scripting, web development, and software development. It works on various platforms. It allows the developer to write programs by using a simple syntax.

Django: It is a framework mainly used for web based applications. This framework is a toolkit that contains all the necessary components for application development. It helps the developer to concentrate on the new components than the components which are already developed.

SQL: It means Structured Query Language. By using SQL we can communicate with the database. SQL statements are used to update data on a database and retrieve data from the database. Most preferred commands in SQL are "insert", "delete", "update", "create", "drop" and "select".

IV. RESULT

The following are the output screens of our project:

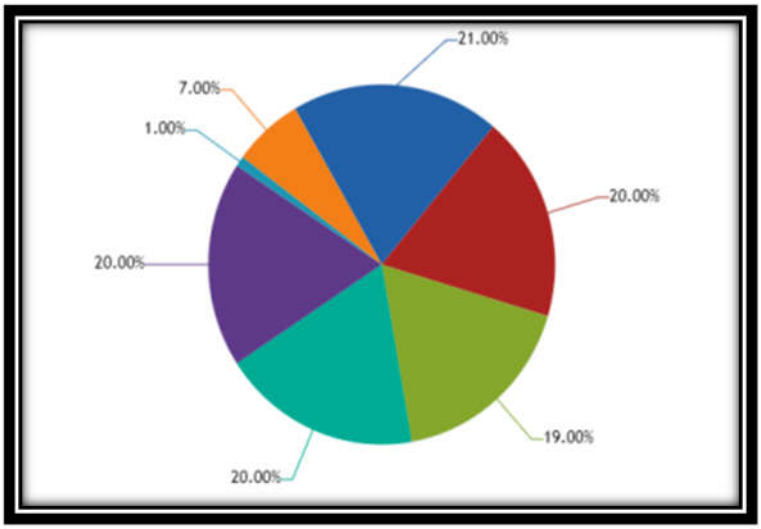

Figure 2: Admin Performs Graph Analysis Based on Data Provided by User

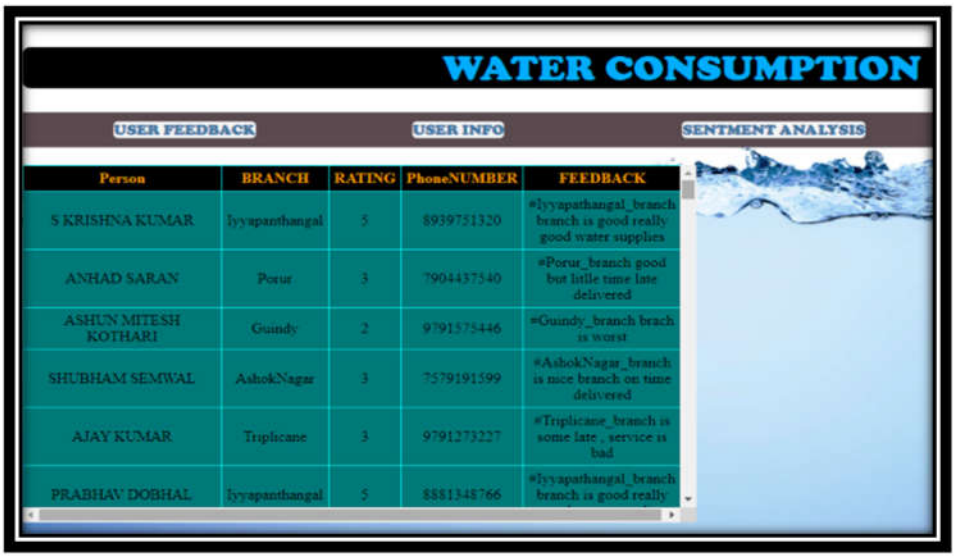

Figure 3: Feedback Analysis is Done by the Admin, it Shows Different Feedbacks Provided by the User

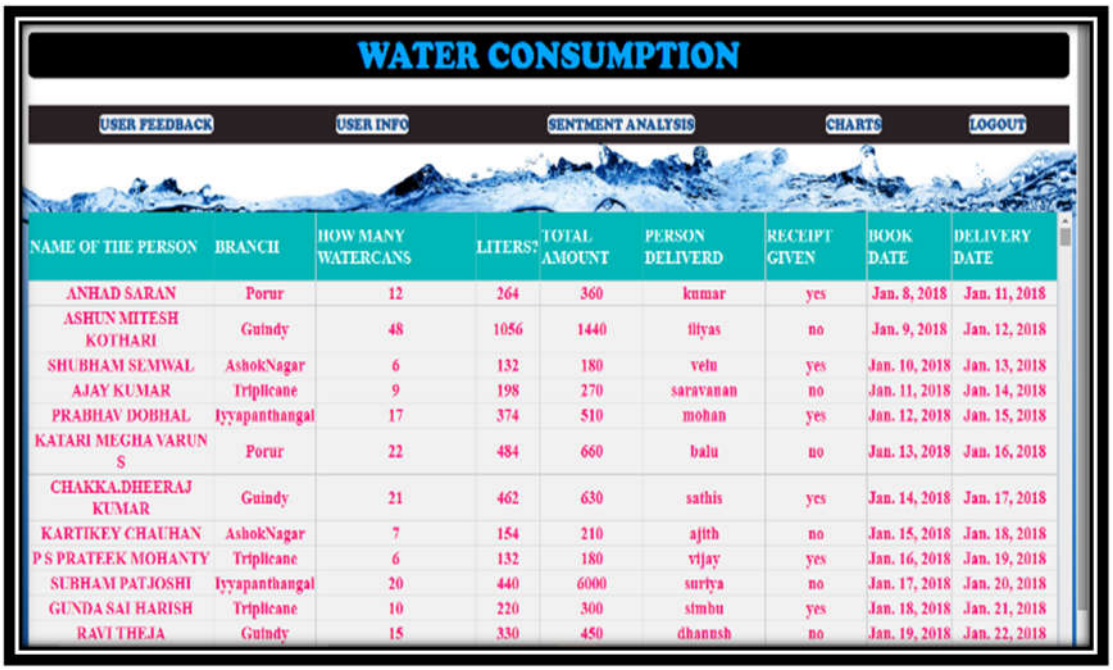

Figure 4: Service Analysis is Done Based on the User Provided Details and Admin Performs this Task

V. CONCLUSION AND FUTURE WORK

In this project, we applied the data mining techniques like SVM and KNN to detect the duplicitous behaviour of dealers in water management. The result analysis is done by using the confusion matrix. In order to convert the data into a specific format that suits the

mining classifiers, it took some effort and time. Here, in the process of detecting the duplicitous dealer, some part is done by the admin manually.

An accuracy of around 70% has been achieved due to the better performance of both classification techniques. In Future accuracy of the model can be improved by using better techniques and better technologies or resources. Our Future work is to automate the manual work done by the admin.

VI. REFERENCES

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

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

[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] 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.

[5] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines", IMECS, 2011, Vol I, pp. 16 – 18.

[6]https://scholar.google.co.in/scholar?q=detection+of+fraudulent+behavior+in+water+consumption&hl=en&as_sdt=0&as_vis=1&oi=scholart

[7]https://ieeexplore.ieee.org/document/8355440

[8]https://www.ijraset.com/fileserve.php?FID=23351

[9]https://www.amazon.com/Mining-Detection-Consumption-Billing-System-ebook/dp/B075C86QCD

[10]https://prezi.com/p/wlxnlmb9k1xm/detection-of-fraudulent-behaviour-of-water-consumption/