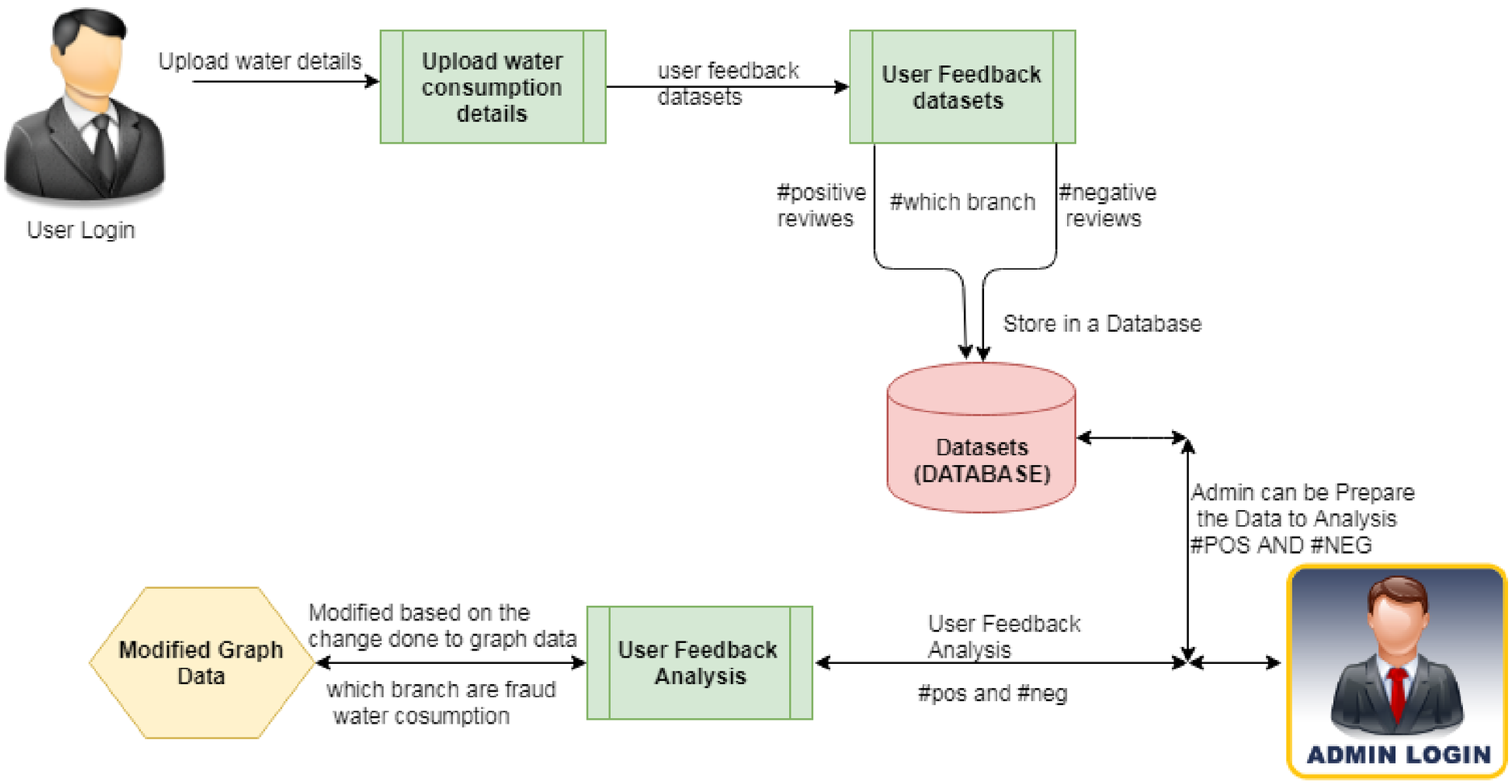
AK Tech Training and **Placements**

Transform Dreams into Reality

A DATA MINING BASED MODEL FOR DETECTION OF FRAUDULENT BEHAVIOUR IN WATER CONSUMPTION

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.

Architecture



Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/

EXISTING SYSTEM

Literature has abundant research for Non-Technical Loss (NTL) 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) 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 to citizens 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.

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) to build a suitable model to detect suspicious fraudulent customers, depending on their historical water metered consumptions. The CRISP-DM (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 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 zero consumption 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.

ALGORITHM

SUPPORT VECTOR MACHINE

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

MODULES

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.

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.

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.

Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/

4.GRAPH ANALYSIS

The graphs are handy to understand the data and based on this analysis admin can find the fraud customers. The business gradually improves as per their understand of where exactly problem arises and to find the place improve and lack. This will gives the clear picture about the current and past picture from the dataset.

FUTUREWORK

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.

REQUIREMENT ANALYSIS

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

REQUIREMENT SPECIFICATION

Functional Requirements

Graphical User interface with the User.

Software Requirements

For developing the application the following are the Software Requirements:

1. **Python**
2. **Django**
3. **Mysql**
4. **Wampserver**

Operating Systems supported

1. **Windows 7**
2. **Windows XP**
3. **Windows 8**

Technologies and Languages used to Develop

1. **Python**

Debugger and Emulator

1. **Any Browser (Particularly Chrome)**

Hardware Requirements

For developing the application the following are the Hardware Requirements:

1. **Processor: Pentium IV or higher**
2. **RAM: 256 MB**
3. **Space on Hard Disk: minimum 512MB**

CONCLUSION

In this research, we applied the data mining classification techniques for the purpose of detecting customers' with fraud behavior 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 Qasab at 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.

Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/