A

Mini Project

On

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

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

by

MUGALA YESHWANTH REDDY (207R1A05A1)

VANGALA SHRUTHI (207R1A05B7)

SUNKARA NIKHIL (207R1A05A6)

Under the Guidance of

Dr. B. Laxmaiah

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2020-2024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"A DATA MINING BASED MODEL FOR DETECTION OF FRADULENT BEHAVIOUR IN WATER CONSUMPTION"** being submitted by **MUGALA YESHWANTH REDDY (207R1A05A1), VANGALA SHRUTHI (207R1A05B7) & SUNKARA NIKHIL (207R1A05A6)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2022-23.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. B. Laxmaiah
(Associate Professor)
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express my profound gratitude and deep regard to my guide **Dr. B. Laxmaiah**, Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Committee Review (PRC) **G. Vinesh Shankar, Dr. J. Narsimha Rao, Ms. Shilpa, Dr. K. Maheshwari & Saba Sultana** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

MUGALA YESHWANTH REDDY (207R1A05A1)

VANGALA SHRUTHI (207R1A05B7)

SUNKARA NIKHIL (207R1A05A6)

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.

LIST OF FIGURES/TABLES

FIGURE NO	FIGURE NAME	PAGE NO
Figure 3.1	Project Architecture for a data mining based model for detection of fraudulent behaviour in water consumption	7
Figure 4.1	Use Case diagram for a data mining based model for detection of fraudulent behaviour in water consumption	9
Figure 4.2	Class diagram for a data mining based model for detection of fraudulent behaviour in water consumption	10
Figure 4.3	Sequence diagram for a data mining based model for detection of fraudulent behaviour in water consumption	11
Figure 4.4	Activity diagram for a data mining based model for detection of fraudulent behaviour in water consumption	12
Table 7.3.1	Classification (Test Cases)	24

LIST OF SCREENSHOTS

SCREENSHOT NO.	SCREENSHOT NAME	PAGE NO
Figure 6.1	User Registration page	17
Figure 6.2	User Login page	18
Figure 6.3	Admin Login page	18
Figure 6.4	User Login Water Analysis	19
Figure 6.5	Customer Feedback	19
Figure 6.6	Analysis using Pie chart	20
Figure 6.7	Analysis using Bar chart	20
Figure 6.8	Analysis using Colum chart	21

TABLE OF CONTENTS

ABSTRACT i

LIST OF FIGURES ii

LIST OF SCREENSHOTS iii

1. INTRODUCTION

1.1 PROJECT SCOPE 1

1.2 PROJECT PURPOSE 1

1.3 PROJECT FEATURES 1

2. SYSTEM ANALYSIS

2.1 PROBLEM DEFINITION 2

2.2 EXISTING SYSTEM 2

2.2.1 LIMITATIONS OF THE EXISTING SYSTEM 3

2.3 PROPOSED SYSTEM 3

2.3.1 ADVANTAGES OF PROPOSED SYSTEM 3

2.4 FEASIBILITY STUDY 4

2.4.1 ECONOMIC FEASIBILITY 4

2.4.2 TECHNICAL FEASIBILITY 5

2.4.3 SOCIAL FEASIBILITY 5

2.5 HARDWARE & SOFTWARE REQUIREMENTS 5

2.5.1 HARDWARE REQUIREMENTS 5

2.5.2 SOFTWARE REQUIREMENTS 6

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE 7

3.2 DESCRIPTION 8

3.2.1 USER LOGIN 8

3.2.2 UPLOAD WATER CONSUMPTION DETAILS 8

3.2.3 USER FEEDBACK DATASETS 8

3.2.4 ADMIN LOGIN 8

3.2.5 USER FEEDBACK ANALYSIS 8

3.2.6 MODIFIED GRAPH DATA 8

4. DESIGN

4.1 USE CASE DIAGRAM 9

4.2 CLASS DIAGRAM 10

4.3 SEQUENCE DIAGRA 11

4.4 ACTIVITY DIAGRAM 12

TABLE OF CONTENTS

5. IMPLEMENTATION

5.1 SAMPLE CODE 13

6. SCREENSHOTS

6.1 USER REGISTRATION 17

6.2 USER LOGIN PAGE 18

6.3 ADMIN LOGIN PAGE 18

6.4 USER LOGIN WATER ANALYSIS 19

6.5 CUSTOMER FEEDBACK 19

6.6 ANALYSIS USING PIE CHART 20

6.7 ANALYSIS USING BAR CHART 20

6.8 ANALYSIS USING COLUMN CHART 21

7. TESTING

7.1 INTRODUCTION TO TESTING

7.2 TYPES OF TESTING 22

7.2.1 UNIT TESTING 22

7.2.2 INTEGRATION TESTING 23

7.2.3 FUNCTIONAL TESTING 23

7.3 TEST CASES 24

7.3.1 CLASSIFICATION 24

8. CONCLUSION & FUTURE SCOPE

8.1 PROJECT CONCLUSION 25

8.2 FUTURE SCOPE 25

9. REFERENCES

9.1 REFERENCES 26

9.2 GITHUB LINK 26

1. INTRODUCTION

1.INTRODUCTION

1.1 PROJECT SCOPE

This project aims to develop a data mining-based model for detecting fraudulent behavior in water consumption. It involves collecting and preprocessing historical water consumption data, engineering relevant features, and implementing anomaly detection algorithms. The model will be fine-tuned and evaluated for accuracy, and a real-time monitoring system will be established for continuous fraud detection. Compliance with water regulations and privacy laws will be ensured, with regular reporting to stakeholders.

1.2 PROJECT PURPOSE

The purpose of the project is to develop a data mining-based model for the detection of fraudulent behavior in water consumption. This project aims to enhance the management of water resources by identifying and flagging anomalous patterns in water consumption data, which may indicate unauthorized usage or water theft. By implementing this fraud detection system, the project seeks to reduce water-related losses, promote fair and legal water consumption practices, and ensure compliance with water regulations.

1.3 PROJECT FEATURES

This project encompasses data collection, preprocessing, and feature engineering to prepare historical water consumption data. It involves the implementation of data mining algorithms for anomaly detection, along with thresholding for classification. Model evaluation metrics will be employed for performance assessment. Continuous monitoring and alerting mechanisms will ensure real-time fraud detection, with a feedback loop for model updates. Interpretability and compliance with water regulations are essential, along with robust documentation and reporting. Integration into the existing water consumption monitoring infrastructure is a key feature, along with risk assessment and resource allocation for successful project execution.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

2.1 PROBLEM DEFINITION

The project aims to address the challenge of detecting fraudulent behavior in water consumption. This involves developing a data mining model to identify unauthorized water usage or abnormal consumption patterns. The objective is to protect water resources, prevent revenue loss, and ensure compliance with water regulations by implementing an efficient and accurate fraud detection system.

2.2 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.

2.2.1 DISADVANTAGES OF EXISTING SYSTEM

- Relies mainly on manual inspection and analysis of water usage data.
- Time-consuming, costly, and prone to errors.
- May not be able to detect complex fraud schemes or patterns involving multiple users or properties.
- May not be able to adapt to changes in water usage patterns over time.
- May lack the ability to provide real-time alerts or notifications to the appropriate authorities or stakeholders.
- Delays in detecting and addressing fraudulent behaviour may occur as a result of these limitations.

2.3 PROPOSED SYSTEM

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 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.

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

- More accurate and effective detection of fraudulent behaviour.
- Ability to handle large and complex datasets.
- Faster detection of fraudulent behaviour, potentially in real-time.
- Improved ability to adapt to changes in water usage patterns over time.
- Better identification of patterns and anomalies that may indicate fraudulent behaviour.
- Reduced reliance on manual inspection and analysis.
- Potentially lower costs and fewer errors compared to the existing system.
- Ability to generate automated alerts or notifications for timely response

2.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- **ECONOMICAL FEASIBILITY**
- **TECHNICAL FEASIBILITY**
- **SOCIAL FEASIBILITY**

2.4.1 ECONOMIC FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2.5 HARDWARE & SOFTWARE REQUIREMENTS

2.5.1 HARDWARE REQUIREMENTS:

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. For developing the application the following are the Hardware Requirements:

- **System** : Pentium IV 2.4 GHz.
- **Hard Disk** : 40 GB.
- **Floppy Drive** : 1.44 Mb.
- **Monitor** : 14’ Colour Monitor.
- **Mouse** : Optical Mouse.
- **Ram** : 4 GB.

2.5.2 SOFTWARE REQUIREMENTS:

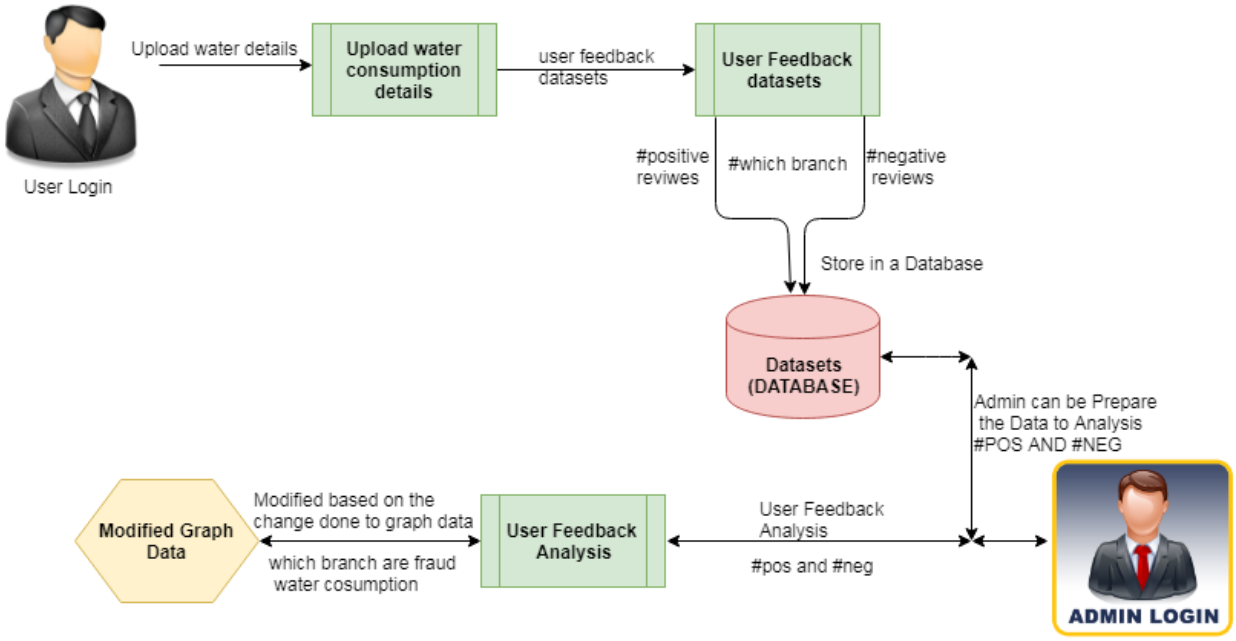
- **Operating system** : Windows 10.
- **Coding Language** : Python.
- **Front-End** : Python.
- **Designing** : Html, CSS, JavaScript.
- **Data Base** : MySQL.

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

This project architecture shows the procedure followed for classification,starting from input to final prediction.



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

Figure 3.1: Project Architecture of a data mining based model for detection of fraudulent behaviour in water consumption.

3.2 DESCRIPTION

3.2.1 User Login:

User must register initially by the filling the details such as user id, user name, password, email, phone number, address. In the next step, user must login using user name and password.

3.2.2 Upload water consumption details:

User must enter the bill no, branch, number of cans, number of litres, amount, receipt, booking date, delivery date.

3.2.3 User feedback datasets:

User must enter the name, branch, services rating, mobile number, feedback. This details are then stored in the dataset.

3.2.4 Admin Login:

The admin has to enter the username, password to login. Admin can prepare the data to analyze.

3.2.5 User feedback analysis:

Here the admin analyzes the positive and negative feedback given by the customers.

3.2.6 Modified graph data:

After the analysis of the user info and feedback the output can be viewed in the form of charts such as pie chart, Bar chart, Column chart.

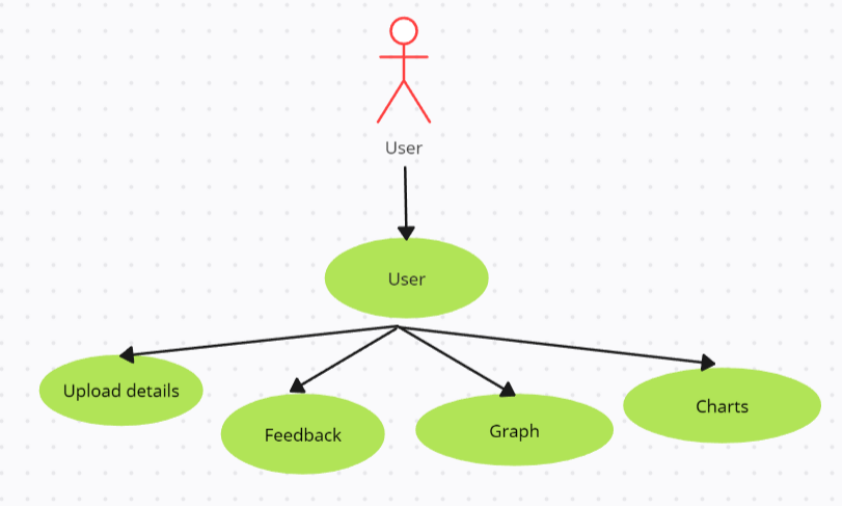
4. DESIGN

4.DESIGN

4.1 USE CASE DIAGRAM

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of usersthe system has. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

a. User



b. Admin

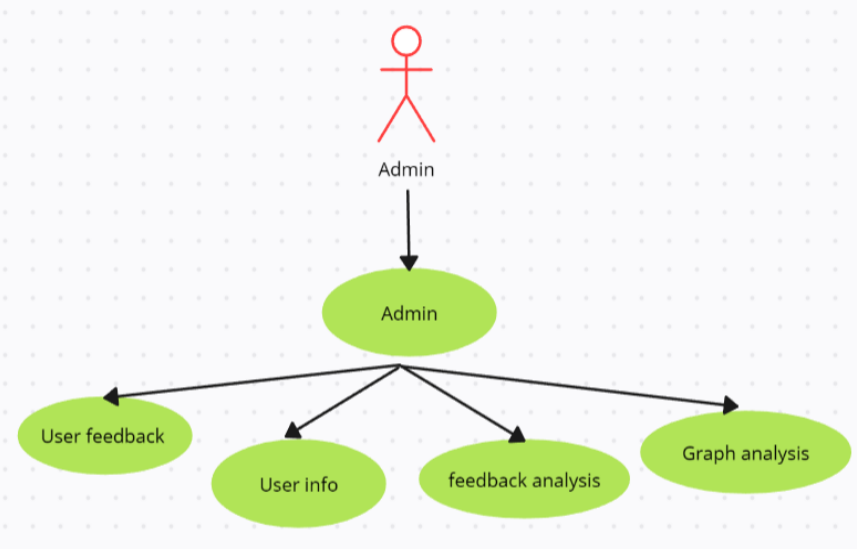
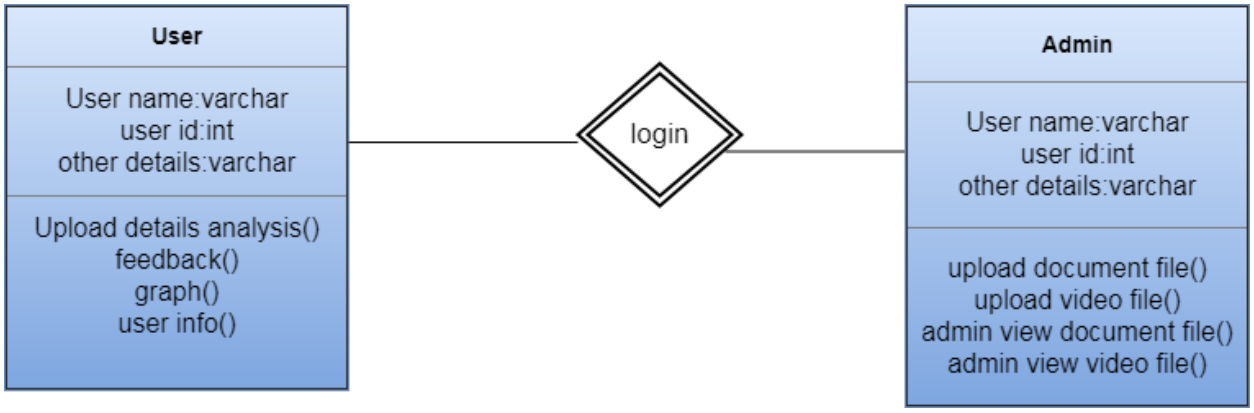


Figure 4.1: Use Case Diagram for a data mining based model for detection of fraudulent behaviour in water consumption.

4.2 CLASS DIAGRAM

Class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.



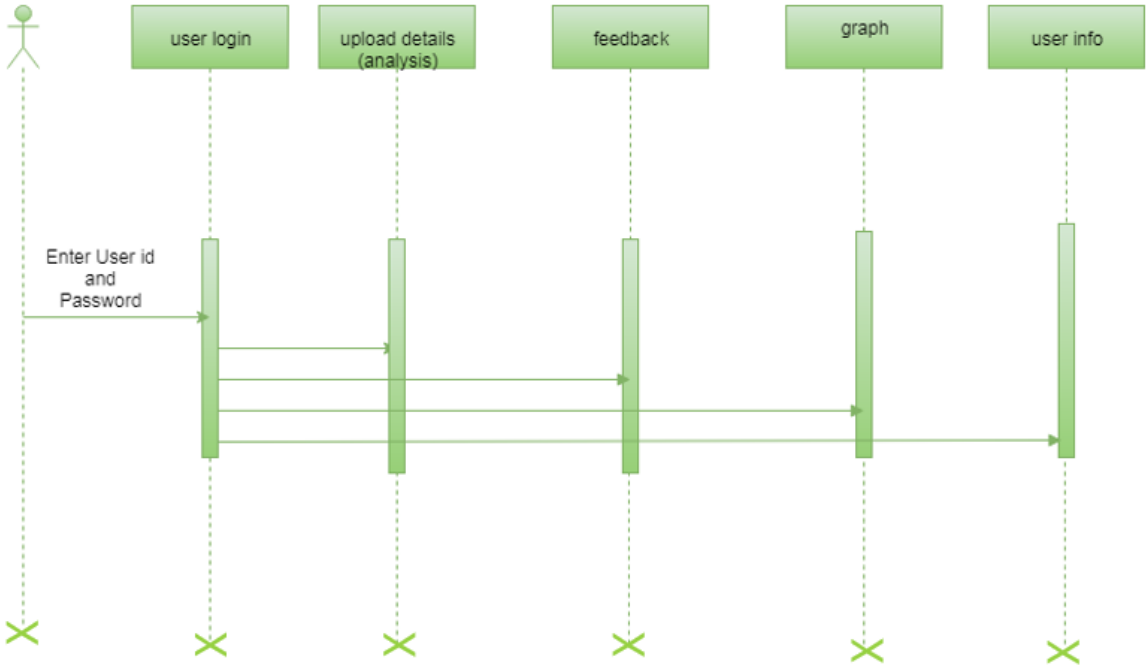
CLASS-DIAGRAM OF A DATA MINING BASED MODEL FOR DETECTION OF FRAUDULENT BEHAVIOUR IN WATER CONSUMPTION

Figure 4.2: Class Diagram for Spammer detection and fake user identification on social network

4.3 SEQUENCE DIAGRAM

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development.

a. User



b. Admin

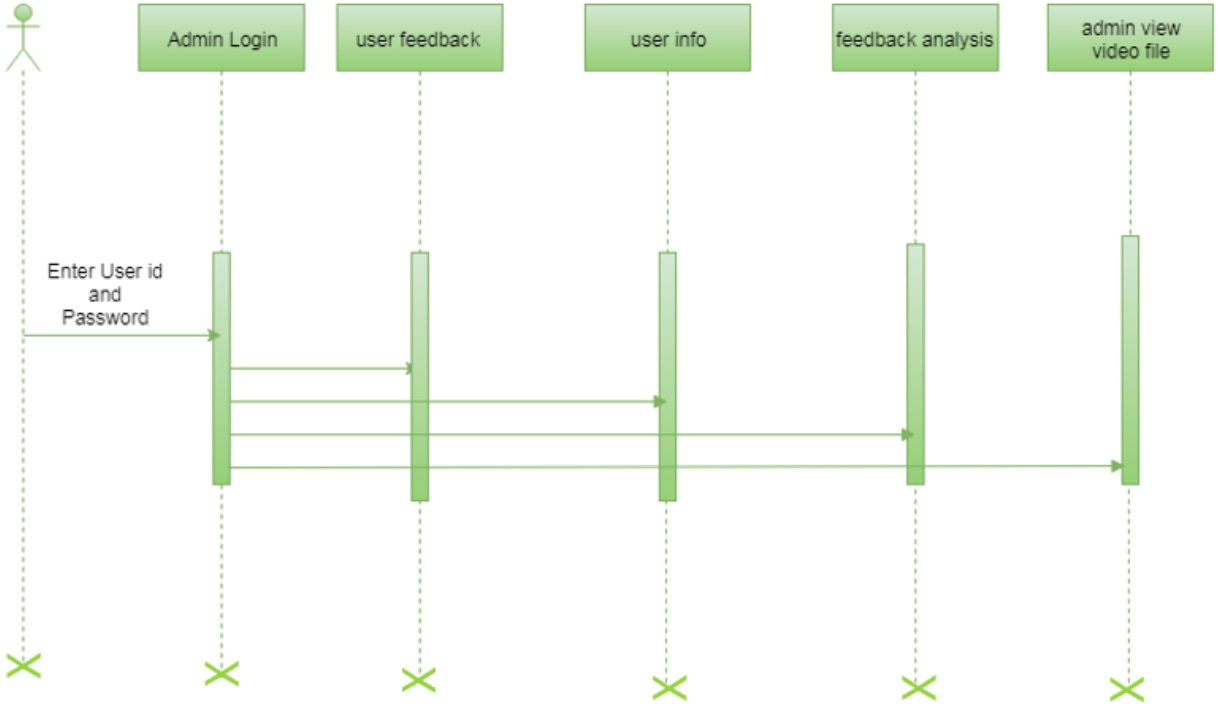
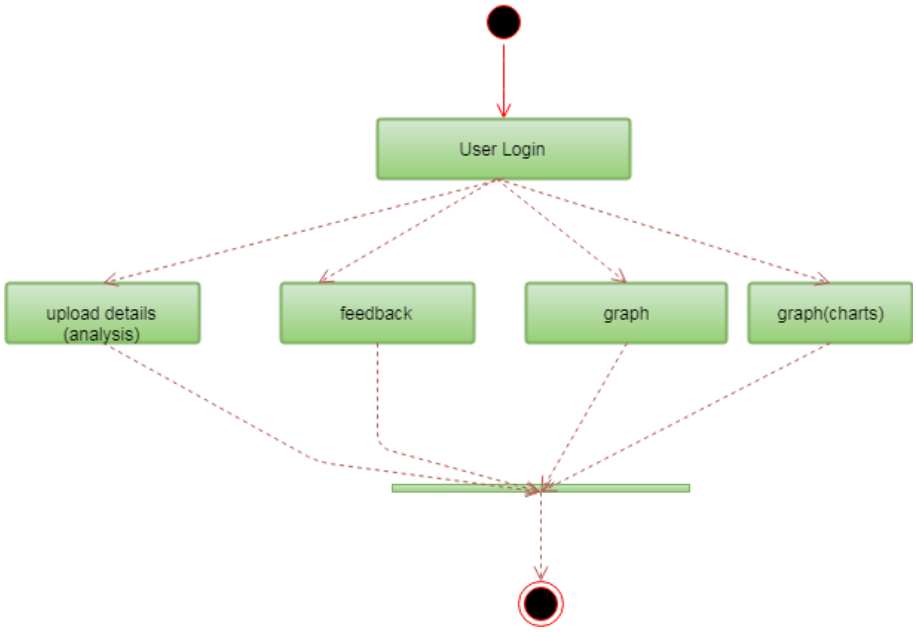


Figure 4.3: Sequence Diagram for a data mining based model for detection of fraudulent behaviour in water consumptio

4.4 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency.

a. User



b. Admin

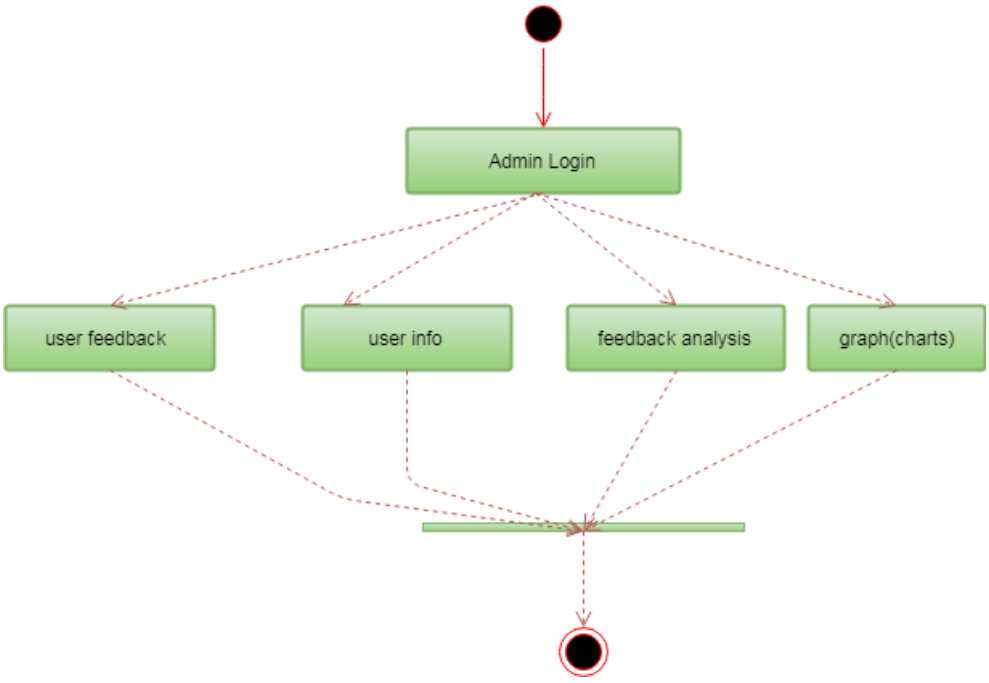


Figure 4.4: Activity Diagram of a data mining based model for detection of fraudulent behaviour in water consumption.

5. IMPLEMENTATION

5. IMPLEMENTATION

5.1 SAMPLE CODE

```
#!/usr/bin/env python
"""Django's command-line utility for administrative tasks."""
import os
import sys


def main():
    os.environ.setdefault('DJANGO_SETTINGS_MODULE',
'spmmer_detection.settings')
    try:
        from django.core.management import
execute_from_command_line
    except ImportError as exc:
        raise ImportError(
            "Couldn't import Django. Are you sure it's installed
and "
            "available on your PYTHONPATH environment variable?
Did you "
            "forget to activate a virtual environment?"
        ) from exc
    execute_from_command_line(sys.argv)


if __name__ == '__main__':
    main()
```

```
from django.db.models import Count
from django.shortcuts import render, redirect

# Create your views here.
from Users_Controller.models import Userregisters_Model, Userwateranalysis_Model,
Userfeedback_Model


def index(request):
    return render(request, 'users/index.html')
def base(request):
    return render(request, 'users/base.html')

def user_logins(request):
    if request.method == "POST":
        name = request.POST.get('name')
        password = request.POST.get('password')
        try:
            enter = Userregisters_Model.objects.get(name=name,password=password)
            request.session["name"]= enter.id
            return redirect('user_wateranalysis')
        except:
            pass

    return render(request, 'users/user_logins.html')

def user_registers(request):

    if request.method == "POST":
        userid= request.POST.get('userid')
        name = request.POST.get('name')
        password = request.POST.get('password')
        email = request.POST.get('email')
        phoneno = request.POST.get('phoneno')
        address = request.POST.get('address')
        Userregisters_Model.objects.create(userid=userid,name=name,password=passw
ord,email=email,phoneno=phoneno,address=address)

    return render(request, 'users/user_registers.html')

def user_wateranalysis(request):
    name = request.session['name']
    obj = Userregisters_Model.objects.get(id=name)
```

```
    if request.method =="POST":

        useridorbillno = request.POST.get("useridorbillno")
        branch = request.POST.get("branch")
        now = request.POST.get("now")
        nol = request.POST.get("nol")
        amount = request.POST.get("amount")
        wpd = request.POST.get("wpd")
        receipt = request.POST.get("receipt")
        bookdate = request.POST.get("bookdate")
        deliverydate = request.POST.get("deliverydate")
        Userwateranalysis_Model.objects.create(uregid=obj,useridorbillno=useridor
billno,branch=branch,now=now,nol=nol,amount=amount,wpd=wpd,receipt=receipt,bookda
te=bookdate,deliverydate=deliverydate)


    return render(request, 'users/user_wateranalysis.html')

def user_feedback(request):
    name = request.session['name']
    userObj = Userregisters_Model.objects.get(id=name)
    result = ''
    pos = []
    neg = []
    oth = []
    se = 'se'
    if request.method == "POST":
        name = request.POST.get('name')
        branches = request.POST.get('branches')
        rating = request.POST.get('rating')
        mobilenumber = request.POST.get('mobilenumber')
        emailid = request.POST.get('emailid')
        twt = request.POST.get('feedback')

        if '#' in twt:
            startingpoint = twt.find('#')
            a = twt[startingpoint:]
            endingPoint = a.find(' ')
            title = a[0:endingPoint]
            result = title[1:]
        # return redirect('tweetpage')

        for f in twt.split():
```

```
            if f in ('good', 'nice', 'beteer', 'best', 'excellent',
'extraordinary', 'happy', 'wonder', 'love', 'greate',):
                pos.append(f)
            elif f in ('worst', 'waste', 'poor', 'worsttaste', 'advance amount',
'extrachage', 'late', 'imporve', 'bad', 'unhealthy'):
                neg.append(f)
            else:
                oth.append(f)
        if len(pos) > len(neg):
            se = 'positive'
        elif len(neg) > len(pos):
            se = 'negative'
        else:
            se = 'nutral'
        Userfeedback_Model.objects.create(uregsid=userObj,name=name,branches=bran
ches,rating=rating,mobilenumber=mobilenumber,sentiment=se ,topics=result,
feedback=twt, )
    obj = Userfeedback_Model.objects.all()

return render(request, 'users/user_feedback.html',{'list_objects':
obj,'result':result,'se':se})


def ucharts(request,chart_type):
    chart =
Userfeedback_Model.objects.values('topics').annotate(dcount=Count('sentiment'))

    return render(request,"users/ucharts.html",
{'form':chart'chart_type':chart_type})
```

6. SCREENSHOTS

6. SCREENSHOTS

6.1 User Registration page :

User must register initially by the filling the details such as user id, user name, password, email, phone number, address.

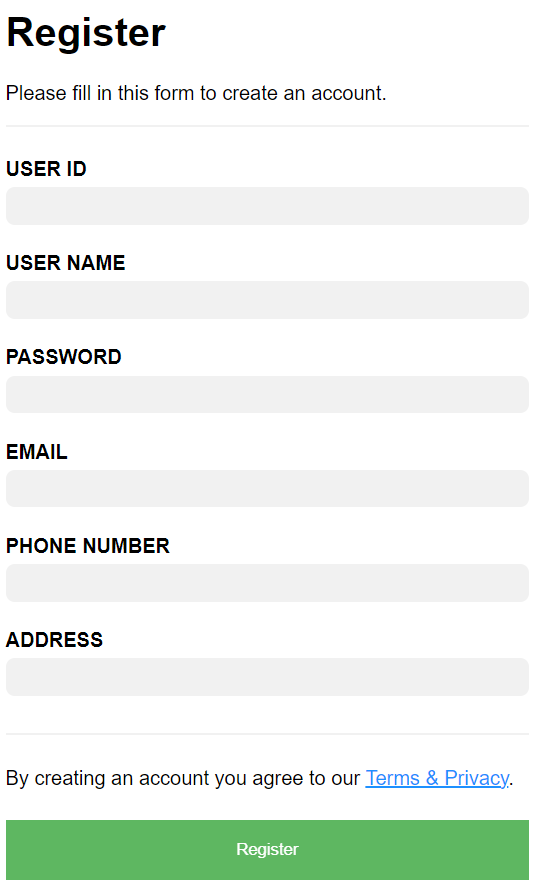


FIGURE 6.1: User Registration page

6.2 User Login:

The user must login with username and password.

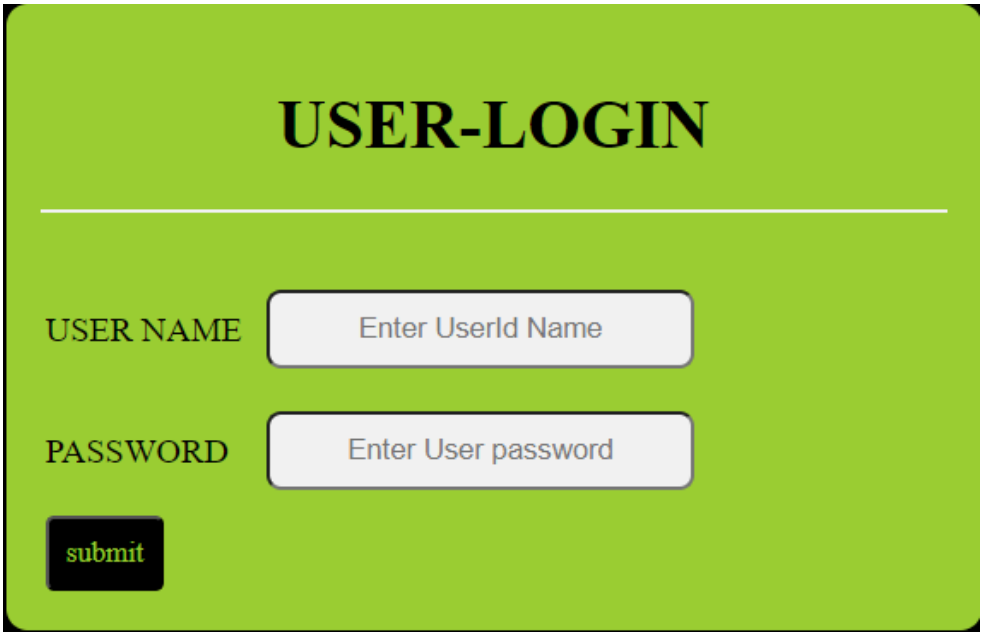


FIGURE 6.2: User-Login page

6.3 Admin Login:

The admin has to enter the username, password to login. Admin can prepare the data to analyze.

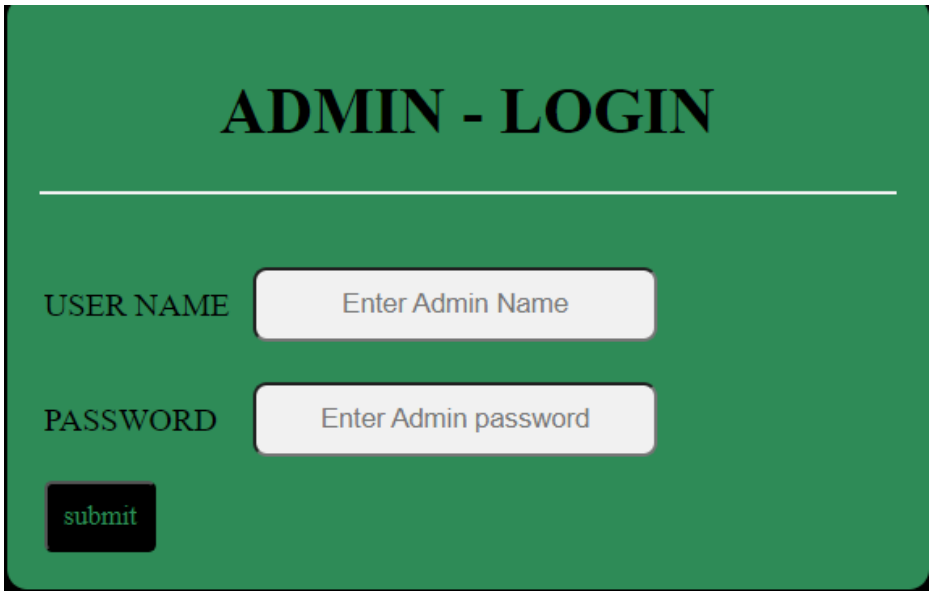


FIGURE 6.3: Admin Login page

6.4 User Login Water Analysis:

User must enter the bill no, branch, number of cans, number of litres, amount, receipt, booking date, delivery date.

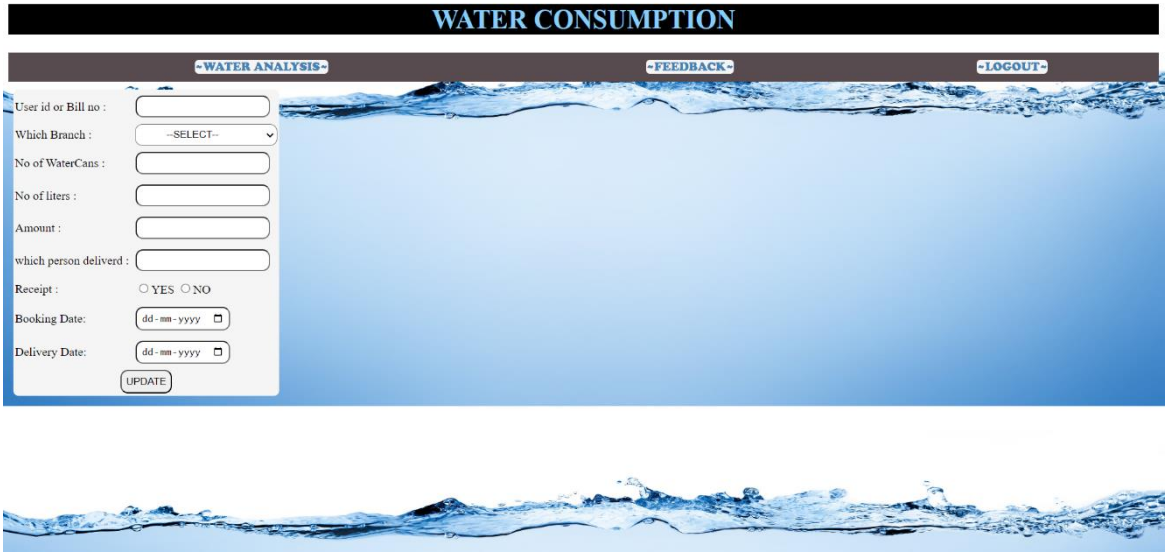


FIGURE 6.4: User Login Water Analysis

6.5: Customer feedback:

User must enter the name, branch, services rating, mobile number, feedback. This details are then stored in the dataset.

FIGURE 6.5: Customer Feedback

6.6 Analysis using pie chart:

After the analysis of the user info and feedback the output can be viewed in the form of charts such as pie chart.

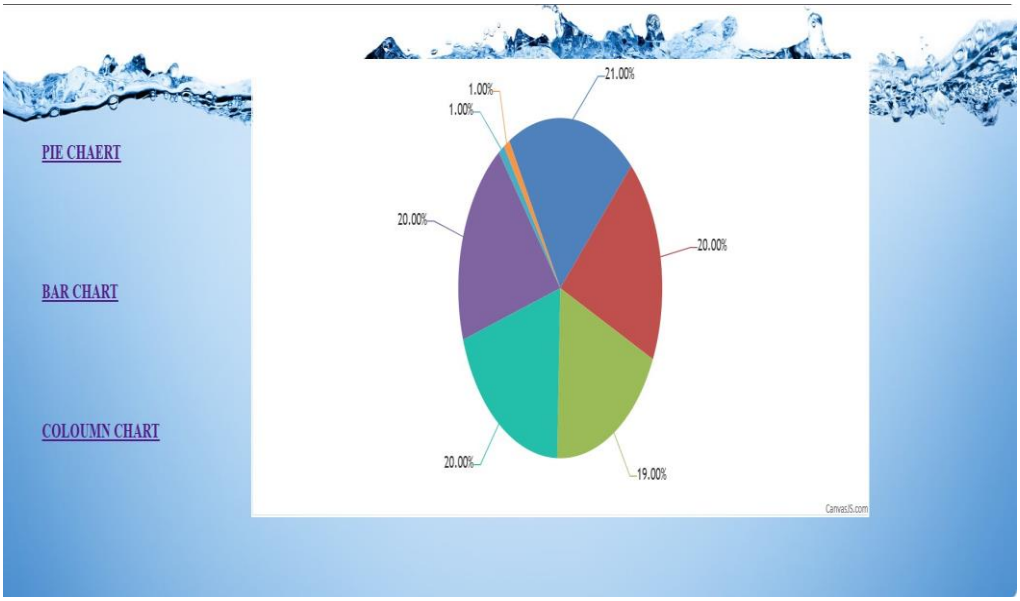


FIGURE 6.6: Analysis using Pie Chart

6.7 Analysis using Bar Chart:

After the analysis of the user info and feedback the output can be viewed in the form of charts such as bar chart.

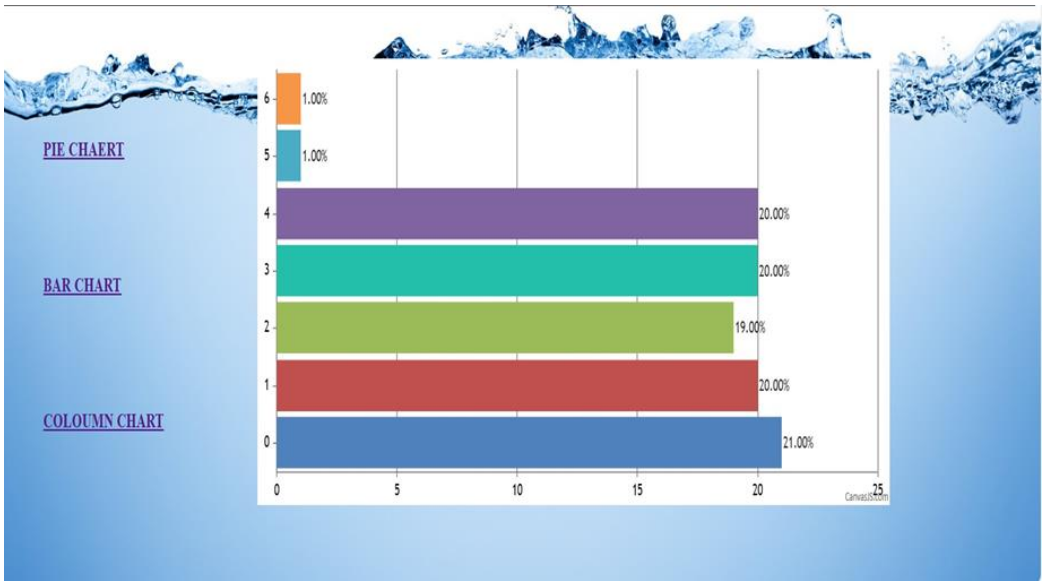


FIGURE 6.7: Analysis using Bar Chart

6.8 Analysis using Column chart:

After the analysis of the user info and feedback the output can be viewed in the form of charts such as column chart.

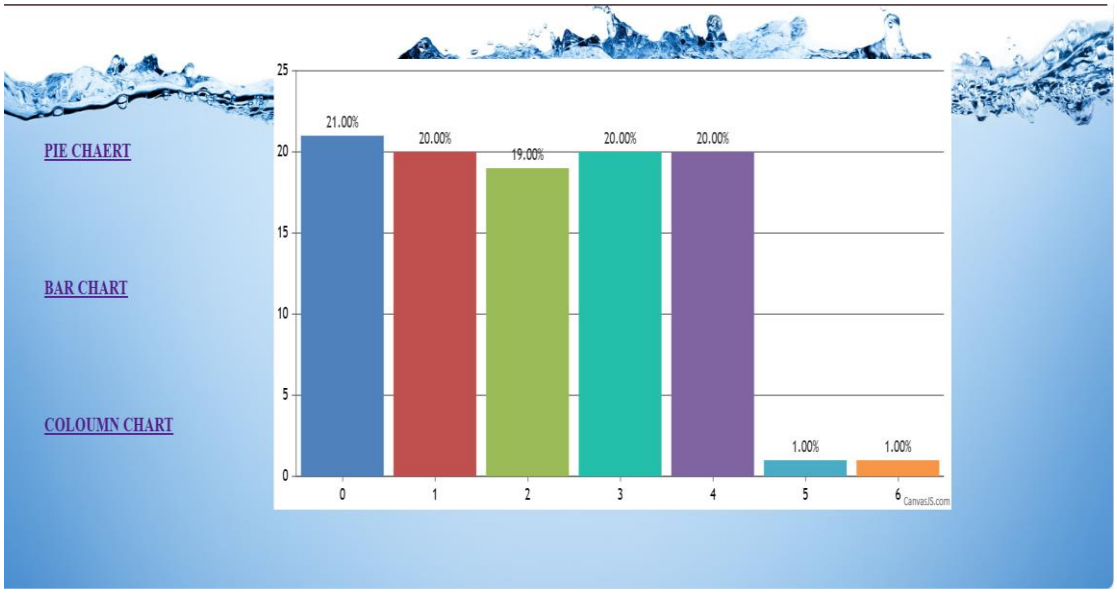


FIGURE 6.8: Analysis using Column Chart

7. TESTING

7.TESTING

7.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

7.2 TYPES OF TESTING

7.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

7.2.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

7.2.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input	: identified classes of valid input must be accepted.
Invalid Input	: identified classes of invalid input must be rejected.
Functions	: identified functions must be exercised.
Output	: identified classes of application outputs must be exercised.
Systems/Procedures	: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

7.3 TEST CASES

7.3.1 CLASSIFICATION

Test Case ID	Description	Input Data/Scenario	Expected Output/Result	Status (Pass/Fail)
01	Normal Consumption	Simulate normal water consumption data for a typical customer over time.	No fraud alert should be triggered.	Pass
02	Sudden Spike in Consumption Alert	Introduce a sudden and significant increase in water consumption over a short time period for a customer.	A fraud alert should be triggered.	Pass
03	Consistent High Consumption Alert	Simulate a customer who consistently consumes a high amount of water, well above their historical average.	A fraud alert should be triggered.	Pass
04	Meter Tampering Detected	Simulate a scenario where meter tampering is reported or detected for a customer.	A fraud alert should be triggered.	Pass
05	Customer Behavior Change Alert	Introduce a significant change in a customer's consumption pattern, such as shifting from low usage to high usage.	A fraud alert should be triggered.	Pass

8. CONCLUSION

8. CONCLUSION & FUTURE SCOPE

8.1 PROJECT 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 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.

8.2 FUTURE SCOPE

Studies have shown that the efficiency of Support Vector Machines (SVM) and its KNearest (KNN) neighbors achieved an average of about 70% of both. In the future similar accuracy can be improved with the help of advanced techniques. Through the use of the proposed model, water services can increase revenue recovery by reducing Non-Technical Loss Management (NTL's) and increase the productivity of inspecting staff by checking in the area of suspicious fraudulent customers.

9. REFERENCES

9. REFERENCES

9.1 REFERENCES

[1] ‘Jordan Water Sector Facts & Figures Ministry of Water and irrigation of Jordan’, Technical Report, 2015.

[2] ‘Water Reallocation Policy Ministry of Water and irrigation of Jordan’, Technical Report, 2016.

[3] "Fast non-technical losses identification through optimum-path forest",C Ramos, A. Souza, J. Papa and A. Falcao, Proc. of the 15th Int. Conf. Intelligent System Applications to Power Systems, pp. 1-5, 2009.

[4] ‘Data mining techniques for the detection of fraudulent financial statements’, E. Kirkos, C. Spathis and Y. Manolopoulos, Expert Systems with Applications, vol. 32, pp. 995- 1003, 2007.

[5] ‘Detecting credit card fraud by decision trees and support vector machines’, Y. Sahin and E. Duman, IMECS, vol.I, pp. 16-18, 2011.

9.2 GITHUB LINK

https://github.com/ShruthiVangala/water-consumption