Financial Bank Statement Analysis

Farman Ansari[1], Arsil Zunzunia[2], Aamir Thekiya[3], Er. Farzana Shaikh[4]

[1,2,3]Student, Information Technology Engineering Department,

Anjuman-I-Islam's M. H. Saboo Siddik College of Engineering

[4]Professor, Information Technology Engineering Department,

Anjuman-I-Islam's M. H. Saboo Siddik College of Engineering

Abstract: For our financial statement analysis project which we are planning to develop there are many free websites available which basically analyse your statement and give you a response and the response is also divided into various categories, but they are not able to provide accuracy. What paid versions are giving is basically they are giving accuracy but the issue with that is that they are very much costly. What we are planning is basically to make a system which is providing accuracy and that is to be free for all. We would like to achieve the maximum accuracy and the service should be free for all. This is a huge problem for all of those who are basically working in any organisation, the people who are managing the finances of the organisations they have to do all these things manually. We are trying to automate this thing. Secondly it is very much helpful for those who are doing too many online transactions and those individuals who are in their last year their financial year are very much confused or very much concerned about how and why they are spending their money. Since this project looks very simple and there is very little user interaction but tricky and interesting machine learning models involved for building the system. We have to train on a model for future use. The categories are not so simple for dividing them individually since the uploaded file will be in PDF format and we have to convert that into text format and then we have to make the data consistent and finally we have to provide the data to the machine learning model. After that machine learning model will be trained in such a way that it will give maximum possible accuracy.

Keywords: Financial Bank Statement, Maximum Possible Accuracy.

I. INTRODUCTION

Financial statement analysis is an idea where bank statements are uploaded and getting the expenditure categorized into various categories, this helps the users to manage their expenses and get personalized reports for keeping track of expenses. This reduces the burden on users, they are free to not think on the money and where their money is going. Our aim is to build a system that analyses the bank statement of the user and gives a systematic and very much accrete analysis of the user's expenditures, meaning where they spend their money like how much their personal expenditure, capital expenditures, interest expenses, etc. There are many portals or systems that are available for analysing bank statements but in their analysis, accuracy is too bad or some of the portals and systems provide good accuracy but they are very much costly. Our main aim is to achieve maximum accuracy with minimum cost.

II. LITERATURE SURVEY

Sr. No	Paper Title	Author publishing date	Methodologies	Hardware and software requirements
1	pandas: a Foundational Python Library for Data Analysis and Statistics.	Wes Mckinney,2013	**Advantages:-** Python library of rich data structures and tools for working with structured data sets common to statistics, finance, social sciences, and many other fields. The library provides integrated, intuitive routines for performing common data manipulations and analysis on such data sets.	**Hardware specification** Processor (CPU) with 2 gigahertz (GHz) or above A minimum of 2 GB of RAM Monitor Resolution 1024 X 768 or higher A minimum of 20 GB of available space on the hard disk Internet Connection Broadband (high-speed) Internet connection with a speed of up to 2 Mbps or

			Disadvantages:- One major issue for would-be statistical Python programmers in the past has been the lack of libraries implementing standard models and a cohesive framework for specifying models.	higher. **Software specification**:- Markup Language & Scripting Languages Web Browser PostgraseSQL Database
2	Machine Learning-Based Techniques for Financial Data Analysis and Forecasting Purposes	Hakob Grigoryan, Professor Cătălina-Lucia Cocianu,2009	**Advantages**:- We attempt to answer the important questions regarding the effectiveness of technical and fundamental analysis together with data pre-processing techniques on performance of time series prediction models. **Disadvantages**:- Later on, the effectiveness of both - technical 2 and fundamental analysis in financial prediction tasks was disputed by the efficient market hypothesis (EMH) stating that it is impossible to outperform the market, and stock prices follow a random walk.	**Hardware specification** Processor (CPU) with 2 gigahertz (GHz) or above A minimum of 2 GB of RAM Monitor Resolution 1024 X 768 or higher A minimum of 20 GB of available space on the hard disk Internet Connection Broadband (high-speed) Internet connection with a speed of up to 2 Mbps or higher. **Software specification**:- Markup Language & Scripting Languages Web Browser PostgraseSQL Database
3	Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labelled Corpus	Silvia García-Méndez,2020.	**Advantages**:- linear classifiers, which are efficient, robust and easy to interpret, have been successful at sentiment analysis. Diverse complex features have been added to these text classification models. **Disadvantages**:- Despite considerable efforts to introduce more sophisticated techniques for document representation such as those based on higher-order word statistics , NLP , string kernels and word clusters, simple bag-of-words (BOW) approaches are still popular.	**Hardware specification** Processor (CPU) with 2 gigahertz (GHz) or above A minimum of 2 GB of RAM Monitor Resolution 1024 X 768 or higher A minimum of 20 GB of available space on the hard disk Internet Connection Broadband (high-speed) Internet connection with a speed of up to 2 Mbps or higher. **Software specification**:- Markup Language & Scripting Languages Web Browser PostgraseSQL Database

III. OBJECTIVE

One of the solutions to Communication with the deaf-mute people is by using the services of sign language interpreter. But the usage of sign language interpreters could be expensive. [3] Cost-effective solution is required so that the deaf-mute and normal people can communicate normally and easily. [3] Our strategy involves implementing such an

application which detects pre-defined American Sign Language (ASL) through hand gestures. Detection of movement of gesture, we would use basic level of hardware component like camera and interfacing is required. Our application would be user-friendly Based system built on PyQt5 module [6] This application will comprise of two core module one is that simply detects the gesture and displays appropriate alphabet. The second is after a certain amount of interval period the scanned frame would be stored into buffer so that a string of character could be generated forming a meaningful word. Additionally, an-addon facility for the user would be available where a user can build their own custom-based gesture for a special character like period (.) or any delimiter so that a user could form a whole bunch of sentences enhancing this into paragraph and likewise. Whatever the predicted outcome was, it would be stored into a .txt file.

IV. PROBLEM STATEMENT

To develop a system which performs analysis of a user's bank statement. We are developing this financial statement analyser to reduce their headache to where they had spent their money and where they had spent. A system which gives more accurate results to the end user for their statement so they can easily see and understand their personal expenses, interest expense, etc also they will see from where their money is credited. The main problem in this system is minimisation of cost for their analysis. In the market there are many portals, applications which analyse the bank statement, but they are paid or some portals, applications which do not provide accuracy. By performing analysis on the bank transactions, we need to categorise every bank transaction. Categorizing/Filtering/Sorting enables us to perform category specific operations and answer questions such as "how much does he spend on which items?" or "what are the different channels of earning?". A category can be Shopping or ATM or NEFT, etc. Statement analyser helps in analysing the buying, spending and saving behaviour of the user, checking whether the user is doing any side business or calculating growth of the business and checking whether user has any running loans and their payment status, to calculate repay etc Analysing recurring transactions.

V. SYSTEM REQUIREMENT

HARDWARE REQUIREMENTS
- Processor (CPU) with 2 gigahertz (GHz) or above
- A minimum of 2 GB of RAM • Monitor Resolution 1024 X 768 or higher •
- A minimum of 20 GB of available space on the hard disk
- Internet Connection Broadband (high-speed) Internet connection with a speed of up to 2 Mbps or higher..

SOFTWARE REQUIREMENTS
- Markup Language & Scripting Languages
- Web Browser
- PostgraseSQL Database

VI. SYSTEM ARCHITECTURE
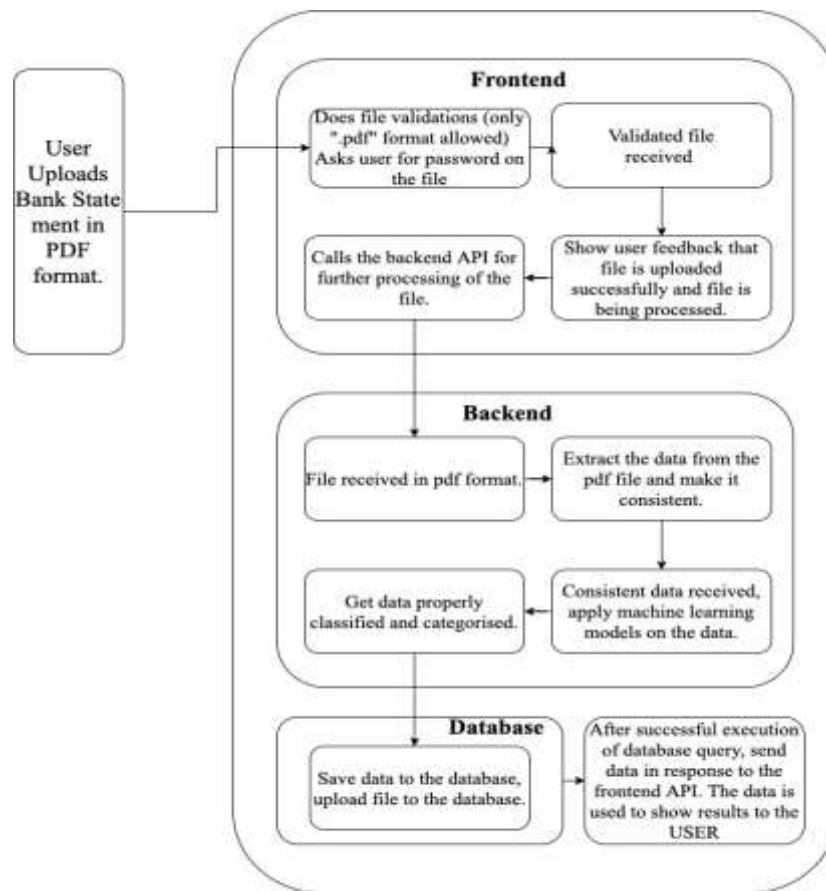
Architecture of the proposed system
Fig 1 : Architecture of the system.

VII. MODULES IN THE SYSTEM

- When user access the bank statement analyser portal the user can upload his/her bank statement and they can analyse their expenditures
- User can see their savings and the amount where he/she is spent. They can see their unnecessary expenditures and reduce them to save their money.
- Users can register them self and after that they have to login and then they can upload the bank statement.
- Users can see their expenditures in pie chart for better understanding

VIII. USE CASE DIAGRAM
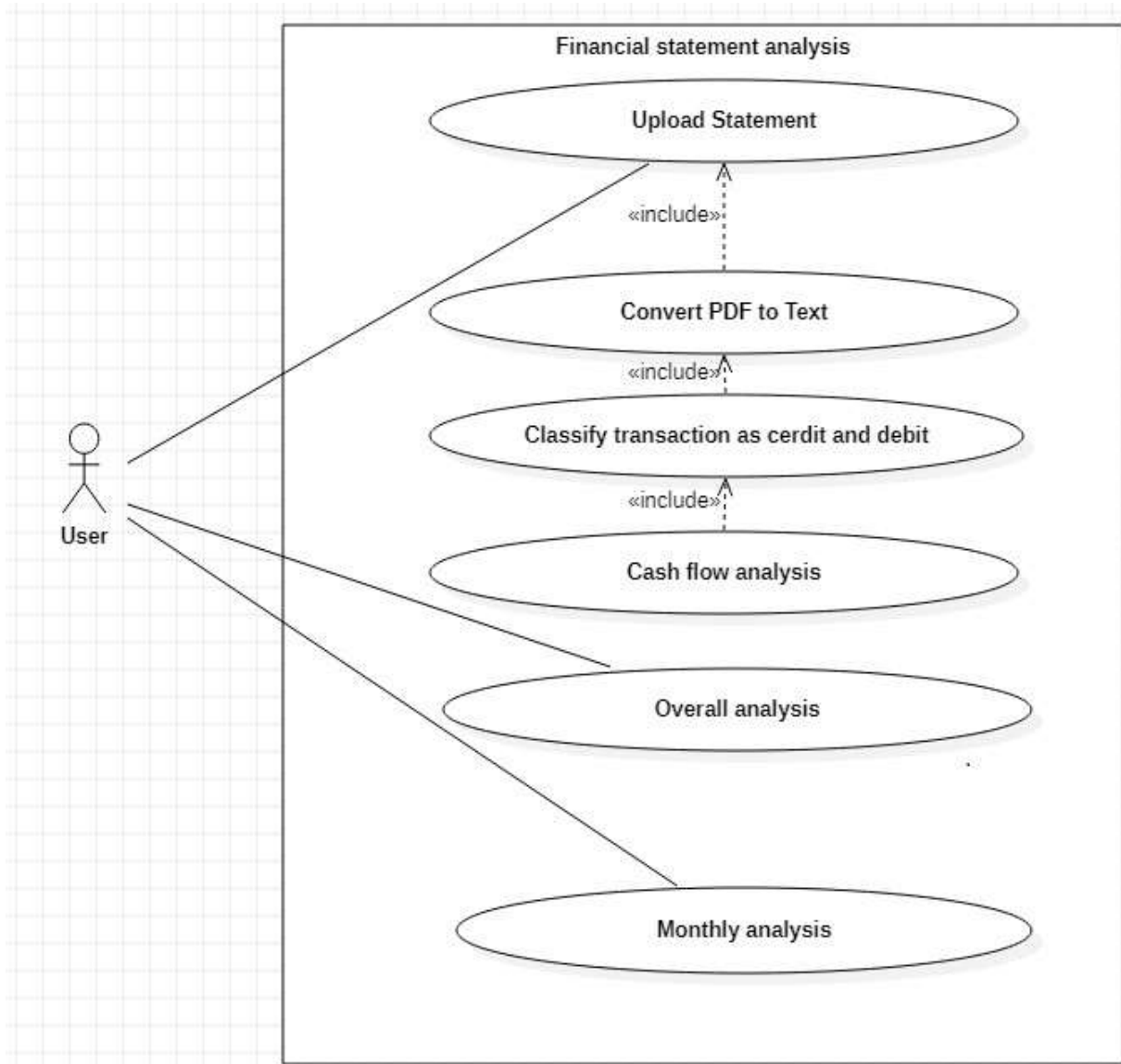
Fig 2: Use Case Diagram for Financial Statement Analysis.

IX. ACTIVITY DIAGRAM
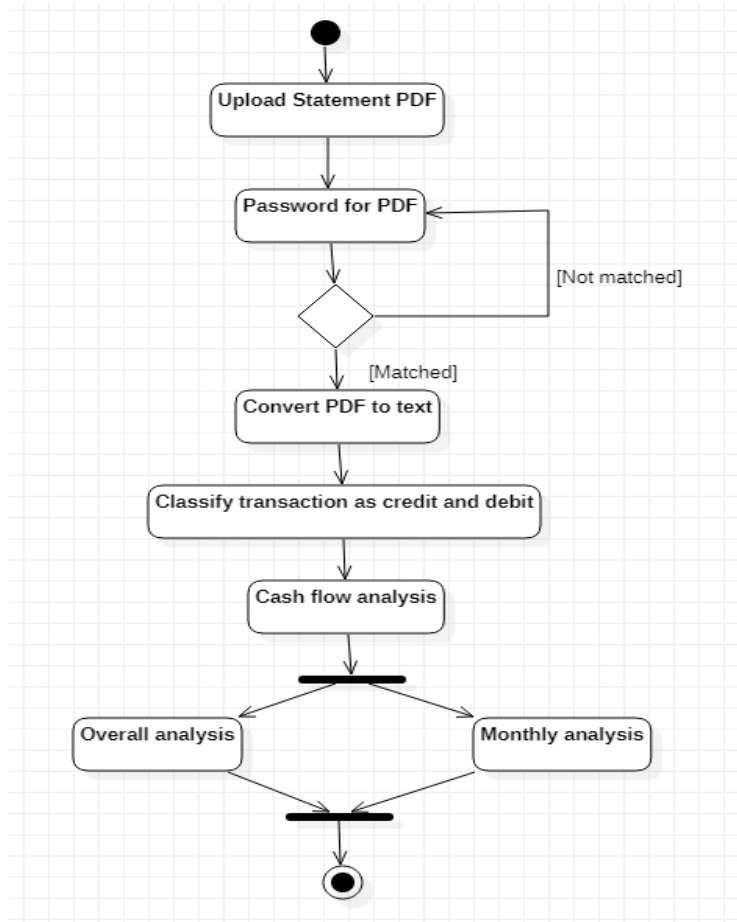
Fig 3: Activity Diagram for Financial Statement Analysis.

X. RESULT

We have successfully studied the bank statement analysis project and we are able to achieve below results. We are able to categorise transactions from bank statements in various categories like Cash, EMI, Shopping, etc. We are showing data visualizations like pie charts and line graph. We are not just showing the transaction amount in money but also, we are showing the number of times the transaction appears in the bank statement. We also have combined as well as different pie charts and line graphs for "Credited" and "Debited" transactions. With the above achievements we are also showing basic information present in the bank statements like name of the account holder, Bank name, total number of transactions. Along with that we also have one section for Balances which shows Average closing balance, Average weekly closing balance, etc. Below there are some screenshots of out project which contains all the information mentioned above.

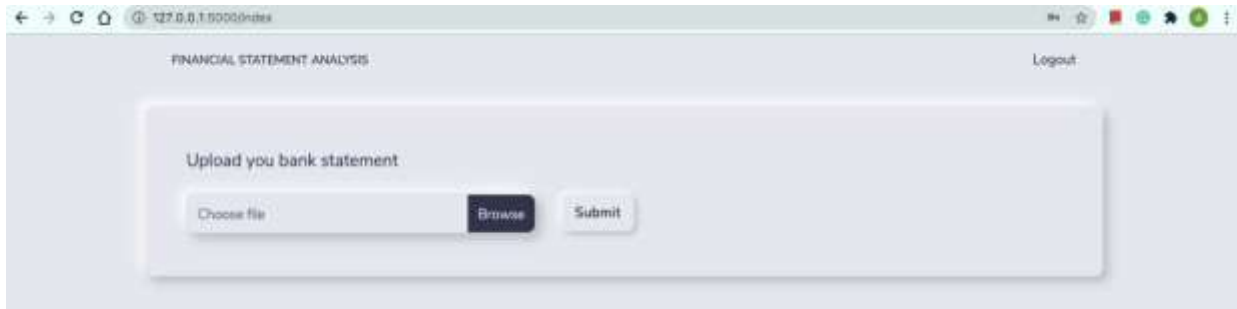Fig 4 : Home page for Financial Statement Analysis.

Fig 5 : This screenshot allows users of our system to upload their bank statements for analysis.

Fig 6 : After analysis this is the basic information and
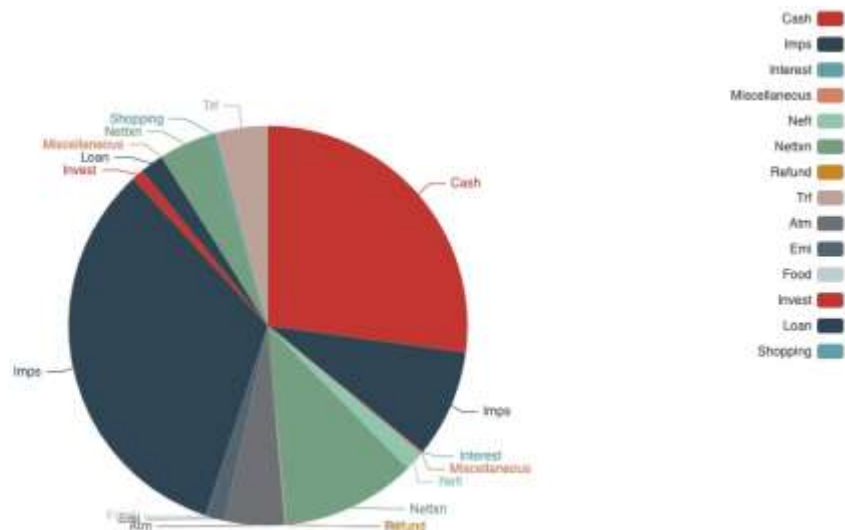balances information which we are able to extract from the bank statement.

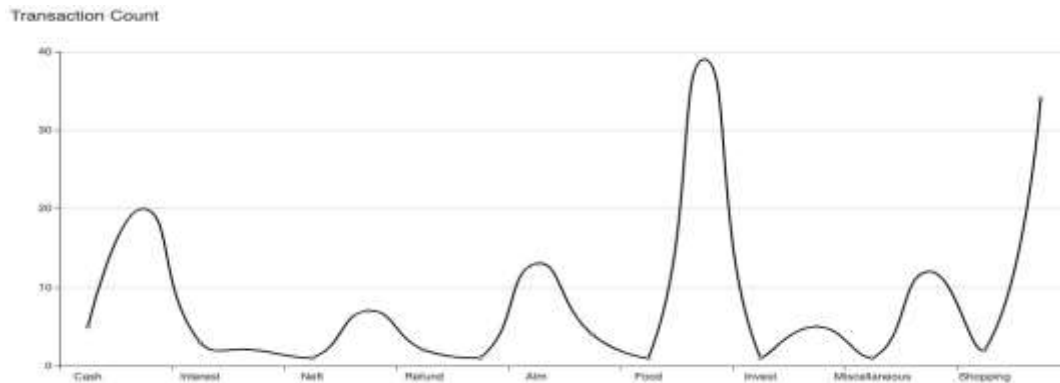Fig 7 : After analysis this is the total transactions categorised in rupees.

Fig 8 : After analysis this is the total transactions categorised in number of occurrence in the bank statement.
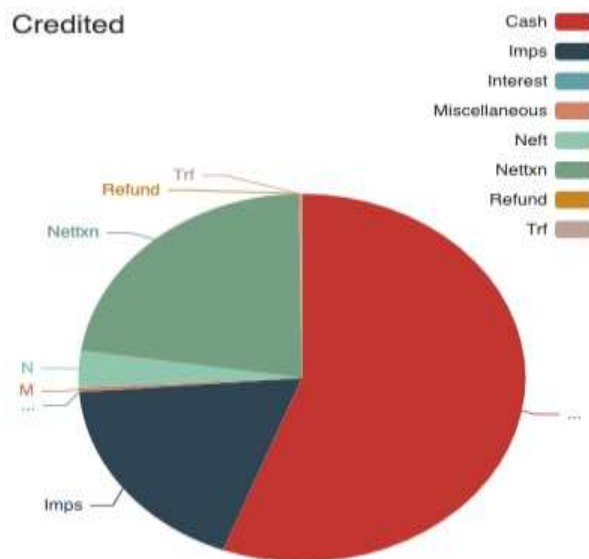

Fig 9 : After analysis this is the total credited transactions categorised in rupees.
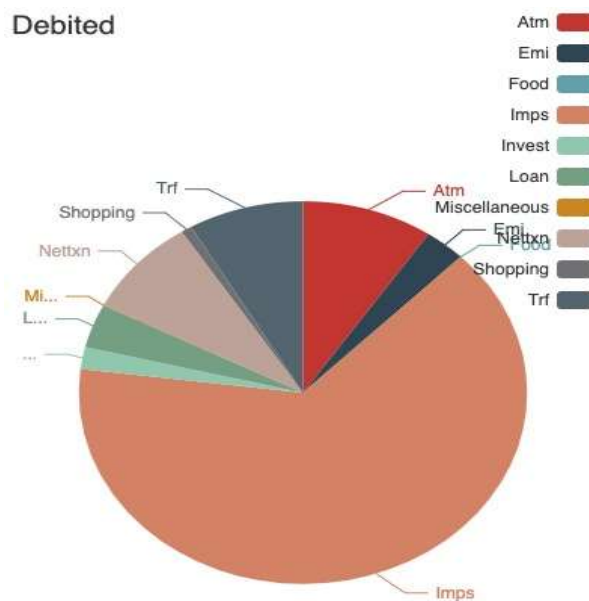

Fig 10 : After analysis this is the total debited transactions categorised in rupees.
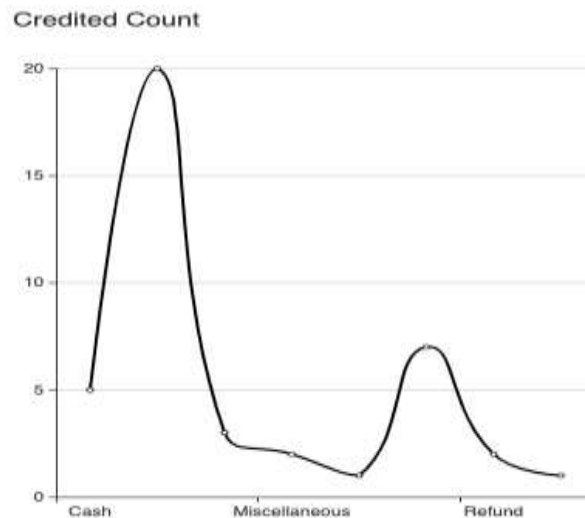
Fig 11 : After analysis this is the total credited transactions categorised in number of occurrence in the bank statement.
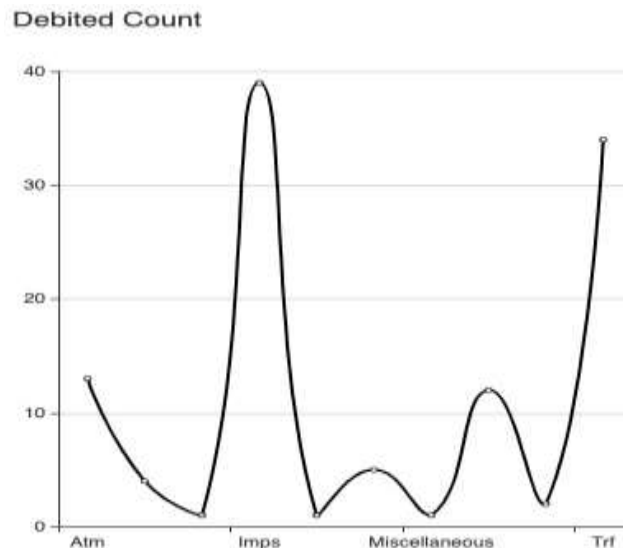

Fig 12 : After analysis this is the total debited transactions categorised in number of occurrence in the bank statement.

XI. CONCLUSION

This application analyses bank statements and provides analytical reports on the account expenses. This application takes a financial statement (provided by the bank) and provides analysis of the expense.
A bank statement containing previous six months of a person running a business is usually more than 20 pages long with around 1k transactions. Columns are generally of date, particular balance, deposit, withdrawal, etc.
Count of columns, positioning of columns, separators, text format and abbreviations vary.

XII. FUTURE SCOPE

Basically, we will be trying in future to enhance our system globally by making it efficient for all the banks, now we have restricted it to only two banks. We will also make our system convenient for all the National as well as International organisations. Till date there is no such system which gives 100% accuracy in future this will be our main motto to give complete accuracy.
There are some ways that should improve the performance of the classifier. They are as following:

● **Get more data and use it for training** –
In machine learning problems it is often the case when getting more data significantly improves the performance of a learning algorithm. We believe that a dataset with a couple of millions of news articles would be of a great help for the learning process.

- **Achieve more accuracy**–

i. **Add more data:** In future we will be making our system more efficient and accurate by providing all the relevant data.

ii. **Feature Engineering:** This step helps to extract more information from existing data. New information is extracted in terms of new features. These features may have a higher ability to explain the variance in the training data. Thus, giving improved model accuracy.

iii. **Multiple algorithm:** Hitting at the right machine learning algorithm is the ideal approach to achieve higher accuracy. But,it is easier said than done. This intuition comes with experience and incessant practice. Some algorithms are better suited to a particular type of data sets than others. Hence, we should apply all relevant models and check the performance.

REFERENCES

1. Wes Mckinney, "pandas: a Foundational Python Library for Data Analysis and Statistics", ResearchGate, Jan 2011.
2. Silvia García-Méndez "Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labelled Corpus." , IEEE, Mar 2020.
3. Fatih Ertam, Galip AydÕn "Data Classification with Deep Learning using Tensorflow", IEEE, 2017.
4. V. Singh, L. Freeman, B. Lepri, and A. Pentland, "Predicting spending behaviorusing socio-mobile features,".
5. KIM Zi Won Comparison between different reinforcement learning algorithms on open AI Gym environment (Cart-Pole v0) 2017.
6. 3.Ankit Choudhary A Hands-On Introduction to Deep Q-Learning using Open AI Gym in Python IIT Bombay EEE.
7. Andrew NG Machine Learning Course Coursera. Available: https://www.coursera.org/learn/machine-learning
8. J. Filliben et al., "Introduction to time series analysis," in NIST/SEMATECH Handbook of Statistical Methods, National Institute of Standards and Technology, 2003.