FINANCIAL SUPERVISION AND MANAGEMENT SYSTEM USING ML ALGORITHM

Atishay Jain[#1], Yash Sharma[*2], Kaushal Kishor[#3]

[#1] *Under Graduate Student ABES Institute of Technology, Ghaziabad, U.P, India*

[*2] *Under Graduate Student ABES Institute of Technology, Ghaziabad, U.P, India*

[#3] *Associate Professor ABES Institute of Technology, Ghaziabad, U.P, India*

[#1]*atishay23@gmail.com,* [*2]*Yash.october27@gmail.com,* [#3]*kaushal.kishor@abesit.in*

I.   ABSTRACT

In this paper, we aim to put forward highly beneficial features in banking and money tracking web-based application. We make use of machine learning algorithms mathematical tools available for Python which is our programming language of choice. As per the ML part of our paper, we have used Random Forest Algorithm as a base on which we have built a proposed algorithm which will predict the user's financial portfolio. For the prediction to take place we are taking the help of tools like RNN, LSTM, MLP. We are aiming to also include a few modern technologies in order to make the financial supervision system, a more usable and interactive experience i.e., by including features like Virtual assistant and a chatbot using technologies like NLTK. This paper shows a proposed algorithm which can be taken into consideration when designing algorithms for banking and financial needs.

Keyword- Random Forest Algorithm, NLTK, RNN, LSTM, MLP.

II.   INTRODUCTION

In most countries like India, which is a developing country, money management, accounting, budget execution or other financial things are still being in an incredibly old way. We currently use so much powerful stuff as compared to the previous generation such as a smartphone, laptops etc but still, we use the traditional ways of managing our finances.

As we have seen and done a lot of research on how small businesses work in our country, we have concluded that people still lack technological advancements when it comes to knowing how to predict their business's future trend or market. They lack prediction tools which are not easily accessible in order to see future trends. This has had deleterious effects on the functioning of their public expenditure management, they also tend to spend on other areas which might not be that essential for their business currently without analysing on what areas they should focus and put their money on.

The lack of reliable and timely revenue and expenditure data for budget planning, monitoring, expenditure control and

reporting has negatively impacted the budget management of these small businesses. India is a developing nation with a lot of young potentials when it comes to manufacturing great products for the entire world, but they lack a good budget and resource management. We look forward to taking a step and help these small businesses for doing so by providing them with great tools for the prediction of market trends and their own business trends.

III.   FEATURE EXTRACTION

A. Methodology:
1) *Deep Learning:*

Deep Learning, the most important methodology, is the solution of all the complex inputs.

Prediction of Financial Market:

Using the historical data and different parameters of the current market scenario, the neural networks in deep learning predicts the financial values. As deep learning takes too much data into consideration, taking the hidden layers as well, the accuracy and efficiency of the prediction improve up too many extents.

Some techniques used in prediction like:

Recurrent Neural Network (RNN)

RNN is used for organizing data in a sequential order e.g., time-series database.

Long Short-Term Memory Models (LSTM)

In order to help with longer memory so that the forecasted time horizon can be longer, LSTM technique is used.

Multilayer Perceptron (MLP)

MLP consists of an Input layer, Hidden layer and Output layer. This is also suitable for time series forecasting because it is: Non-linear modelling, Multi-step forecasting etc.

Security of Finance:

As taking care of the security regarding finance, deep learning is the best approach. Deep learnings are used in such a way to automate the searching data process for anomalies that could be a security threat.

2) Machine learning and Algorithm based Machine Reasoning -

Machine Learning is a subset of data science that uses statistical insights and make predictions. The main motive of this methodology is to provide the stabilised, efficient and accurate prediction.

In terms of getting the efficiency and accuracy, the algorithm named Random Forest, which is a big win has been used up, supervised learning which has an advantage of solving classification and regression relating problems which is helping a lot in making the almost perfect prediction.

Random Forest algorithm builds multiple decision trees and combines them to give a stable and efficient prediction.

Random forest algorithm reduces the overfitting as if there are so many trees present in the forest, the classifier will not be able to overly fit the model.

Random Forest is simple, fast to execute, flexible in terms of stable prediction.

Random Forest algorithm uses Sklearn (formerly known as Scikit- learns), which is a free machine learning python library which is a biggest helping hand in terms of making a stable and efficient prediction.

Sklearn some built-in functions for random forest algorithm: -

- n_estimators- stabilizes the prediction
- max_features- splits the nodes
- min_sample_leaf - determines the minimum number of leaves required to split an internal node.

The Mathematics involved in Random Forest:

Regression Problems: When to solve regression problems, using the Random Forest Algorithm.

$$ MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2 $$

Where N is the number of data points,
fi is the value returned by the model and
yi is the actual value for data point i.

This formula calculates the distance of each node from the predicted actual value, helping to decide which is the better decision for your forest. Here, yi is the data point value we

are testing at a certain node and fi is the value given by the decision tree.

Classifications Problems: When applying Random Forests based on classification data, we should keep this thing in mind that we are often using the Gini index, or the formula applied to decide how nodes on a decision tree branch.

$$ Gini = 1 - \sum_{i=1}^{C} (p_i)^2 $$

This formula uses the class and probability to determine the Gini value of each branch on a node, determining which of the branches is more likely to occur. Here, pi represents the class relative frequency we are observing in the dataset and c represents the number of classes.

3) NLTK:
Natural Language Toolkit is a toolkit for developing Python programs to work with human language data. It provides interfaces such as WordNet, along with text processing libraries suite for a different piece of works like classification, tokenization, etc. for NLP (Natural Language Processing) libraries.

NLTK is used in the development of virtual assistants and chatbots wherein the virtual assistant, the responses were recorded in accordance with the casual users relating problem. Casual users facing problem relating with access to the system and its usage can help to sort it out very easily.

IV.   LITERATURE REVIEW

Following are a few literature reviews papers we discovered and read. These papers contain similar implementations and ideas whereas they have some or the other limitations.

We reviewed the following papers in order to come to a few conclusions as well as define their limitations when compared to our approach:

1. Javier A. Varela, Nobert When (2017). In this paper, the author comes up with various analysis of how risk can be managed using PyOpenCL in Python. The author emphasis on mathematical optimizations in managing risk and how we can make use Jupyter Notebook in creating an environment for analysing the state of the financial portfolio. [2]
2. Peng Zhang, Yuxiang Gao, Xiang Shi (2018). In this paper, we read that is the Quant Cloud, trade and design algorithmic studies were signified using Python. They took consideration of Machine Learning algorithms and used the cloud for data exchange facility. [3]
3. Puneet Mathur (2019). In this paper, the author takes into consideration various other fields in which Python has shown wonderful success using ML technologies, these fields include healthcare industry, Oil & Gas, government, transportation etc. with keeping that in mind, he comes to various

conclusions about how python can be especially useful when it comes to financing. [4]

4. Anish Gupta (2019). In this paper the author emphasis upon how data can be classified in various structured, semi structured and unstructured forms. He also gives examples of companies like Yahoo and Google who finance through supervised and unsupervised machine learning algorithms. Also, in this paper they analysed and predicted stocks using ML algorithms. [5]

5. G. Biau," Analysis of a Random Forests Model. In this paper we concluded that how Random Forest works, the author describes it in a very elaborative terms through a model. The author describes how randomization in it takes place, how sparsity happens etc. [10]

We concluded a few limitations from the existing model:

Lack of Prediction feature which is a core part of our paper. These given papers do not mention how we can use Machine Learning Algorithms such as Logical Regression or Random Forest in order to predict the customer's future financial status.

1. Lack of modern requirements for a business-oriented web app such as the implementation of chatbot and a virtual assistant which will enable customers to interact with the system in a more intrusive way.

2. No special mentions of graphical representation of the user's financial data by using technologies such as XlsxWriter.

V. PROPOSED SYSTEM OVERVIEW

The given model shows how the proposed system will work, all the steps we will take to implement it and they use them to predict the data based on the performance of it.
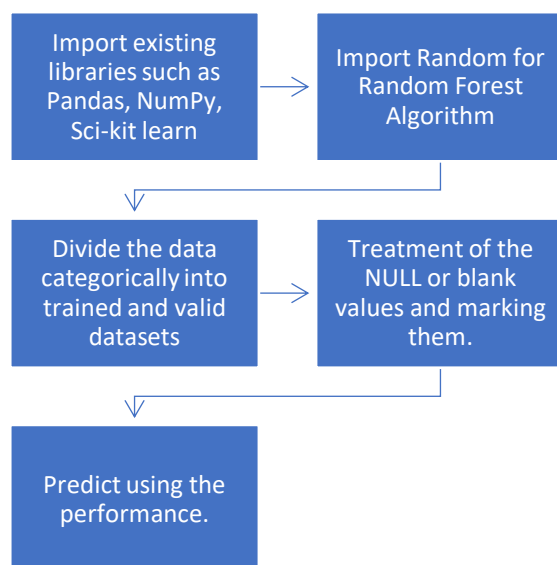

Figure 1: Proposed System overview

VI. IMPLEMENTATION

For the implementation part of there, we are using the Random Forest Machine Learning Algorithm as for the base on which we are going to propose our algorithm.

PROPOSED ALGORITHM

Basic steps to proposing the Algorithm:

Step 1- Descriptive Analysis such as identifying inputs by the user, identifying categorical & numerical values, eliminating missing values, etc.
Step 2- Using the Random Forest techniques, create a benchmark solution which models the data in an efficient manner.
Step 3- Performance estimation by dividing your train data set into train and validate sections (70:30 as a ratio).
Step 4- Importing necessary libraries, tools and train the data set. Libraries and tools such as Pandas, NumPy, Sklearn, random etc.
Step 5- Create summary datasets and in columns classify them accordingly.
Step 6- Identify different variables according to the proposed dataset model such as ID, Target variables, categorical variables or numerical variables.
Step 7- Treating the variables with missing values and create a flag for them. Mark them as false or indicate them with a coloured coded tag.
Step 8- Impute missing values and pass them into the modelling process by using the random forest to predict the classes of them.
Step 9- Check the performance of it and make predictions accordingly.

Let us consider a scenario of US revenue data report.

This table is consisting of different dataset values which are as follows to estimate future values of US revenues:

Revenues- The inputs taken in this table are in either increasing or decreasing order from one of the US revenues reports.
3-mo MA (month Moving Average)- These inputs are the averages of 3 months, calculated by the inbuilt math module. e.g. (=Average (Jan: Mar)).
5-mo MA (month Moving Average)- These inputs are the averages of 5 months, calculated by the inbuilt math module. e.g. (=Average (Jan: May)).

Here, in the table below we can see some blocks left in the starting of the 3-mo MA and 5-mo MA, these gaps represent the average of 3 month and 5-month revenues, respectively.

For example, the average of Jan, Feb, March is represented in the month of March in the 3-mo MA column in the below table, similarly for the 5-mo MA.

($5.0(Jan)+ $8.0(Feb) + $7.0(mar))/3 = $6.7, *which is represented in 3-mo MA.*

Similarly,($5.0(Jan)+$8.0(Feb)+ $7.0(Mar)+$8.0(Apr)+$8.0(May))**/5**=$7.2,*which is represented in 5-mo MA.*

TABLE 1
: The US revenue & 3 and 5 mo Moving Averages

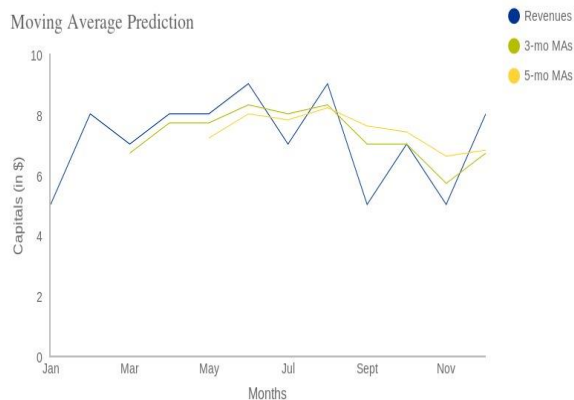|	*Revenues*	*3-mo MAs*	*5-mo MAs*
Jan	$5.0		
Feb	$8.0		
Mar	$7.0	$6.7	
Apr	$8.0	$7.7	
May	$8.0	$7.7	$7.2
Jun	$9.0	$8.3	$8.0
Jul	$7.0	$8.0	$7.8
Aug	$9.0	$8.3	$8.2
Sept	$5.0	$7.0	$7.6
Oct	$7.0	$7.0	$7.4
Nov	$5.0	$5.7	$6.6
Dec	$8.0	$6.7	$6.8

The output of the above table is :

Figure: 2 Moving Average Prediction

🔵 Revenue

🟢 3-MO MAs

🟡 5- MO MAs

The graph represents the input variables, in which we input revenues starting from the month of January to December.

Here we have predicted the graph using our proposed algorithm in which these are the following points to note down:

1. X-axis is representing the months (Jan to Dec).
2. Y-axis is representing the capital in US $(Dollars).
3. The 3-month MA varies to a greater degree, with a prominent increase or decrease in historic revenues in comparison to the 5- month MA.
4. The 5-month MA varies to a greater extent in comparison to 3-month MA.

The above model is a clear representation of the working of our proposed algorithm based on the random forest algorithm.

In this, the revenue is considered as a possible feature for the tree in the random forest which results in 3 month and 5-month predictions. Revenue also represents a part of the historical data of US financial portfolio.

VII. ADVANTAGES

1. This paper would be able to predict future trends for businesses and customers.
2. The paper would be able to diagrammatically represent the given data into the form of excel oriented charts and diagrams.
3. The virtual assistant and the chatbot would help the customers navigate through the application and easily do the task they want to do.

VIII. APPLICATION

The financial management system is utilitarian in many different applications. Especially in collecting information, budgeting, reporting, predicting, analysing etc.in so many sectors. This can be run and used in any kind of financial sectors. Lots of people use these types of application but not able to be used efficiently as they failed to communicate and work rightly. That is why this system is proposed to go beyond in terms of helping every kind of person whether they are casual or not. AI virtual assistants and chatbots can help in communicating with the users so that they can work over their problems easily and efficiently.

Secondly, no system has been proposed till, yet which can predict the financial market whether it will rise or fall. But using this system this prediction will take place.

IX. CONCLUSION

Financial supervision system all studies the casualties taken by the casual user in accordance to control it efficiently and easily through AI virtual assistant and chatbots. It also studies the prediction to be predicted through the old and trending markets and predicts whether the financial market will seek to high or low in the upcoming time. Through this review paper, we try to communicate all the problems which are usually faced by the people whether they are literate or illiterate.

REFERENCES

[1] "Yuxing Yan", "Using Python for Finance 101". Published in January 2013, at Canisius College. SSRN Electronic Journal.

[2] "Javier A. Varela, Nobert When", "Real-Time Financial Risk Measurement of Dynamic Portfolios with Python and PyOpenCL". DOI: 10.1145/3149869.3149872, 7th Workshop, November 2017.

[3] "Peng Zhang, Yuxiang Gao, Xiang Shi", "Quant Cloud using Python", Published in July 2018 at 2018 IEEE International Conference on Software Quality, Reliability and security (QRS), Lisbon, Portugal. The project name is: Big Data Infrastructure for Quantitative Finance on the Cloud.

[4] "Puneet Mathur", "ML application Using Python", DOI: 10.1007/978-1-4842-3787-8_9, January 2019. [Book retrieved from Research gate]. ISBN: 978-1-4842-37-86-1.

[5] "Anish Gupta, Manish Gupta, Prateek Chaturvedi", "Investing Data with ML using Python", July 2019, at International Conference on Quality, Reliability, Infocom Technology and business Operation, New Delhi.

[6] "Yuan Tao", "Analysis on Financial Fraud cases by Python", DOI: 10.2991/icesed-19.2020.8, at Proceedings of the 2019 International Conference on Education Science and economic development (ICESED 2019). Published in Advances in Economics, Business and Management Research, volume 116.

[7] X.H. Hu. Financial Management, Shanghai: Shanghai Lixin Accounting Press, 2009.

[8] L. Zhang Financial Management System Mode, Research Management, 2006.

[9] L. Breiman. Random forests, Machine Learning, 45:5–32, 2001.

[10] G. Biau ," Analysis of a Random Forests Model", Journal of Machine Learning Research 13 (2012) 1063-1095.

AUTHOR PROFILE
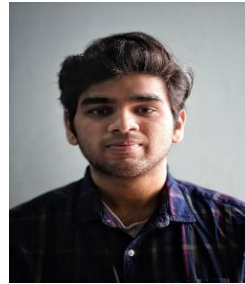
Atishay Jain currently pursuing his B.Tech (4th year) degree in Information Technology through ABESIT, Ghaziabad associated by AKTU Lucknow, Uttar Pradesh. His area of interests are frontend development and database.

Yash Sharma currently pursuing his B.Tech (4th year) degree in Information Technology through ABESIT, Ghaziabad associated by AKTU Lucknow, Uttar Pradesh. His area of interests are backend development and database.

Dr. Kaushal Kishor received his Ph.D. in Computer science and engineering from AKTU Lucknow, in the domain of Mobile Ad hoc Network. MTech & B.Tech in Computer Science & Engineering Currently, he is working in ABES Institute of Technology, Ghaziabad as Associate Professor of Information Technology. He has supervised more than 50 projects for B.Tech students. He has more than 17 years of experience teaching. He is Gate Qualified 2003 score 94.5 percentile. He has reviewed no. of books for publications and book published (1) Design and Analysis of Algorithm (2) Computer Networks (3) Design and Analysis of Algorithms Techniques and Control Management (4) Computer Networks a system approach for various engineering field like B.Tech. and MCA student. He has published more than 20 papers in peer reviewed international/national journals and conferences. His research interest includes Computer Network, Algorithm, Compiler Design Wireless and Sensor networking.