A Project Report

on

Stock closing price prediction

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

Under The Supervision of
Mr. s. Rakesh kumar
Assistant Professor
Department of Computer Science and Engineering

Submitted By

19SCSE1010267 – PRIYANKA DHAR DWIVEDI

19SCSE1010161 – PRIYANSHU KUMAR PANDEY

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA

DECEMBER - 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **"Stock closing price prediction"** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**
submitted in the **School of Computing Science and Engineering** of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **Mr. S. Rakesh kumar, Assistant Professor**, **Department of Computer Science and Engineering** of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

19SCSE1010267 - PRIYANKA DHAR DWIVEDI
19SCSE1010161 – PRIYANSHU KUMAR PANDEY

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

(Mr.V.Arul, Assistant Professor)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **19SCSE1010267 – PRIYANKA DHAR DWIVEDI, 19SCSE1010161 – PRIYANSHU KUMAR PANDEY** has been held on _____and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date:

Place:

ACKNOWLEDGEMENT

This project becomes a reality with the kind support and help of many individuals. I would like to extend my sincere thanks to all of them. Foremost, I am highly indebted to the faculties of Galgotias University for encouraging us to the highest peak and to provide me the opportunity to prepare the project. I am immensely obliged to my friends for their elevating inspiration and help in the completion of the project. I would like to extend my special gratitude and thanks to my guide. MR S. RAKESH KUMAR for imparting his knowledge and experience in this project and her kind supervision given to me throughout the course which shaped the present work as its show. My thanks and appreciations also go to my colleague and people who have willingly helped me out with their abilities. Last but not the least, my parents are also an important inspiration for me. So with due regards, I express my gratitudes to them.

PRIYANKA DHAR DWIVEDI (19SCSE1010267)
PRIYANSHU KUMAR PANDEY (19SCSE1010161)

ABSTRACT

Stock rates using machine learning help you determine the future value of a company's stock and other financial assets traded on an exchange. The whole idea of predicting stock prices to make a big profit. Predicting how the stock market will perform is a daunting task. Accurate forecasting or prediction of stock market acquisition is a very difficult and challenging task due to the volatile nature and inconsistencies of the stock market. With the introduction of smart installations and the proliferation of computer skills, systematic forecasting methods have been shown to be very effective in predicting price levels. We are now able to predict stock prices to some extent using modern techniques. In this process, Artificial Neural Network and Random Forest strategies were used to predict the next day's closing price for other companies in various sectors. Financial details: Open, high, low and stock prices are used to create new variables that are used as inputs in the model. Models are tested using standard strategic indicators: RMSE and MAPE. The low values of these two indicators indicate that the models are effective in predicting the closing price of the stock. We aim to make a project so that it gets easier for people to look where to invest and when, just like mutual fund scheme but less risky. Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are widely used to predict stock prices and their movement. Every algorithm has its own way of learning patterns and predictions.

Table of Contents

List of Figures

List of Tables

Acronyms

RMSE	Root mean square error
MAPE	mean absolute percentage error
SVM	Support Vector Machines
ANN	Artificial Neural Networks

CHAPTER-1

Introduction

Stock market is trading platform where different investors sale and buy shares per stock availability. stock exchange ups and downs effects the profit of stakeholders. If market prices growing with available stock then stakeholders get profit with their purchased stocks. In other case, if market taking place with available stock prices then stakeholders need to face losses. Buyers buy stocks with low prices and sell stocks at high prices and take a look at to urge huge profit. Similarly, sellers sell their products at high prices for profit purpose (Tae Kyun Lee et al, 2019). securities market (SM) work as trusty platform among sellers and buyers. Advances in computer science (AI) supporting lots in each field of life with its intelligent features. Several algorithms present in AI that performing their role in future predictions (Eunsuk Chong et al, 2017). Machine learning (ML) could be a field of AI (AI) which will be considered as we train machines with data and analysis future with test data. Machines is trained on the premise of some standard that are called algorithms. securities market predictions are often great beneficial to businessman. SMP provide future trend of stock prices on the idea of previous history (Bruno et al,2019). If stakeholders get future predictions then investment can lead him toward profit. Predictions is 50% correct and 50% wrong because it is risk of business. Risks facing capability in business filed can lead toward success. In any field of life, we take risks for fulfillment . Similarly, we depend on ML predictions about future prices of stock. during this chapter we'd wish to explain these ML algorithms with the assistance of their working methodologies and examples.

Forecasting and analyzing stock markets is one of the most difficult tasks to complete. There are many causes for this, including market volatility and other independent and volatile fluctuations that affect the value of a particular stock market. These changes make it very difficult for any stock market expert to anticipate the ups and downs of the market with great accuracy. However, with the introduction of Machine Learning and its robust algorithms, recent market research and Stock Market Prediction developments have begun to integrate these approaches into stock market data analysis. The stock market seems strong, unpredictable and out of line by nature. Assessing is a difficult task as it depends on a variety of factors including but not limited to politics, the global economy, corporate financial reports and performance etc. Analyzing trends over the past few years, can greatly assist in stock market movements, near prices etc. stock to predict future prices. The second type of analysis is quality, which is done on the basis of external factors such as company profile, market

situation, political and economic factors, textual information in the form of new financial articles, social media and blogs by economic analysts. Nowadays, advanced strategies are based on technical or basic analysis used to predict prices. Especially in stock market analysis, data size is large and linear to deal with this variability in the active data model it is necessary to identify hidden patterns and complex relationships in this large data set. Mechanical learning methods in this area have been shown to improve efficiency by 60-86 percent compared to previous methods.

Importance of stock market:

1. Stock markets help companies raise money.
2. It helps to produce personal wealth.
3. Stock markets serve as an indicator of economic status.
4. It is a widely used source of investment in potential companies.

MERITS OF PROPOSED SYSTEM:

Stock forecasting has been an important area of research for a long time. Although the promoters of a successful market hypothesis believe that it is impossible to accurately predict stock prices, there are valid suggestions that accurate modeling and design flexibility may lead to models using stock prices and stock price patterns that can be accurately predicted. In this activity, we propose a mixed stock price model to create a unique learning machine with deep learning-based models. For the purposes of our research, we have used NIFTY 50 reference numbers on the National Stock Exchange (NSE) of India, for the period 29 December 2014 until July 31, 2020. NIFTY 50 dated 29 December 2014 to 28 December 2018. Using these retrospective models, we predict NIFTY 50 open prices for the period 31 December 2018 to 31 July 2020. We are therefore improving our forecast framework by creating four in-depth learning models using long-term and long-term memory. short term (LSTM) connects a new way to ensure continuity. We use the power of LSTM recovery models to predict future NIFTY 50 open prices using four different models of their properties and input data structure. Extensive results are presented with different metrics for all regression models. The results clearly show that the

LSTM-based mathematical model uses the first week's data as a prediction of the next week's open pricing series for the most accurate NIFTY 50 timeline series. In the age of big data, in-depth stock market research trends and trends have become more popular than ever. We collected 2-year data on the INDIAN stock market and recommended a fully customizable feature of the feature engineering and in-depth stock market research model. The proposed solution is comprehensive as it incorporates pre-market stock market analysis, application of many feature engineering techniques, combined with an in-depth customized learning program to predict stock market prices. We performed a thorough evaluation of the most widely used machine learning models and concluded that the proposed solution was highly efficient due to the wide range of engineering features we created. The system achieves the highest accuracy of stock market estimates. With a detailed design and evaluation of the length of the forecast period, feature engineering, and previous data processing methods, this work contributes to the community of stock analysis research in both the financial and technical fields.

CHAPTER-2

Literature Survey/Project

Design

ANN, is one of the most ingenious methods of data mining that identifies basic processes from data and normalization from it. ANN is able to mimic and analyze complex patterns in random data compared to most of common methods. The model uses the basic structure of the Nervous Network with different neurons layers. The model works in three stages. Contains the input layer, the hidden layer and the output layer. Installation layer contains new variants H-L, O-C, and 7 DAYS MA, 14 DAYS MA, 21 DAYS MA, 7 DAYS STD DEV and Volume . Weights in each input load are repeated and added and transmitted to neurons. Hidden layer or the activation layer contains these sensors. The total weight is calculated and subtracted by the third layer the output layer. The output layer contains only one neuron that will provide the predicted value about closure stock price. It shows a detailed representation of the ANN structures and the new variations that function as input. Random Forest (RF) is an integrated machine learning program. It can do both conversions as well separation functions. The idea is to combine multiple decision trees to determine the final result of reliance on each decision tree to minimize model variability. In this work, something new has been created flexibility is provided by the training of each decision-making tree in the areas of tree. Noise in stock market data is usually very high due to its large size and can cause trees to grow on them a completely different approach compared to the expected growth. It aims to reduce the risk of medical prediction stock market analysis as a divisive issue and based on training variables predicted the closing of the next day the stock price of a particular company. Forecasting and analyzing stock markets is one of the most difficult tasks to complete. There are many causes for this, including market volatility and other independent and volatile fluctuations that affect the value of a particular stock market. These changes make it very difficult for any stock market expert to anticipate the ups and downs of the market with great accuracy. However, with the introduction of Machine Learning and its robust algorithms, recent market research and Stock Market Prediction developments have begun to integrate these approaches into stock market data analysis. The stock market seems strong, unpredictable and out of line by nature. Assessing is a difficult task as it depends on a variety of factors including but not limited to politics, the global economy, corporate financial reports and performance etc. Analyzing trends over the past few years, can greatly assist in stock market movements, near prices etc. stock to predict future prices. The second type of analysis is quality, which is done on the basis of external factors such as company profile, market situation, political and economic

factors, textual information in the form of new financial articles, social media and blogs by economic analysts. Nowadays, advanced strategies are based on technical or basic analysis used to predict prices. Especially in stock market analysis, data size is large and linear to deal with this variability in the active data model it is necessary to identify hidden patterns and complex relationships in this large data set. Mechanical learning methods in this area have been shown to improve efficiency by 60-86 percent compared to previous methods. Most of the previous work in this area uses older algorithms such as the backbone ,The Random Walk Theory. (RWT) , Moving Average Convergence / Divergence (MACD) and using certain line models such as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) ,for predicting stock prices. Recent work suggests that stock market speculation could be improved using machine learning. Strategies like these Vector Support Machine (SVM), Random Forest (RF). Other strategies based on neural networks such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and depth. neural networks such as Long Short Term Memory (LSTM) have also shown promising results . ANN is able to discover hidden features through the self-study process. These are good equals too they are able to access the input and output relationships of the largest complex database. So, ANN proves to be good option to predict the stock price of the organization. Selvin et al predicted the shares of companies listed on the NSE in comparative analysis of different in-depth reading strategies. Hamzaebi et al. most of the time stock is checked market forecasting using repetitive and directing methods such as the ANN model . Route et al. predicted stock market using a sophisticated RNN cheap model and tested Bombay stock exchange and S&P 500 index database . Roman et al. use RNN models in stock market data from five countries: Canada, Hong Kong, Japan, UK and USA, training networks and then these networks are used to predict the stock return process . In 2014, Yunus et al. application ANN on the NASDAQ predicts stock closing price. Mizuno et al. use ANN to perform technical analysis is the TOPIX website and its function in the buy and sell time forecast system . Some jobs have it It is suggested that RF Forest be used for forecast purposes. RF is a combination method. It's normal capable of performing both restoration and editing tasks. It works by building many decision-making trees the resulting training period refers to the retraction of trees per decision . Mei et al. use RF correctly predicting real-time prices on the New York electricity market . A similar work was done by Yand et al. who does short-term load prediction on power system performance using the RF model. Herrera et al. used RF as a predictive model for predicting urban water demand per hour . In this project, two strategies namely ANN and RF used to predict the closing value of an organization. Models using a new set of power created using Financial website containing Open, High, Low and Close for a particular company. These new directions will pay off your debts contribute to the improved accuracy of the models in predicting the closing price of the next day in a particular

company. I Model performance is assessed using two operating modes: RMSE and MAPE.

Given the impact and power of news on stock market performance, there is an urgent need to analyze, evaluate, and evaluate strategies for predicting critical stock market news. This paper is recommended to address this need. In the context of predicting critical stock trends, we have three questions to guide this research work: (i) What is the most common way to make predictions? (ii) What methods have been used to analyze the text and how can existing methods be improved? (iii) What machine learning algorithms have the potential to model a selected domain?

Based on the above research questions, comprehensive book reviews related to stock forecasting are discussed and guidelines are developed to predict the impact of news events on the stock market. The main findings of this paper are as follows:

1. A comprehensive review of forecasts to predict stock trends in three areas:

a) Stock forecasting research based on timeline financial data and preliminary financial data analysis.
b) Research based on the analysis of the original text and methods of removal of features.
c) Research on forecasting algorithms to analyze the impact of stocks and numbers on the stock market.

2. Based on the revised literature, solutions are identified that guide to solving the challenges found in the general steps of the stock forecasting process.

3. Discussions about open-ended issues and research guidelines, which can be investigated and monitored by the public.

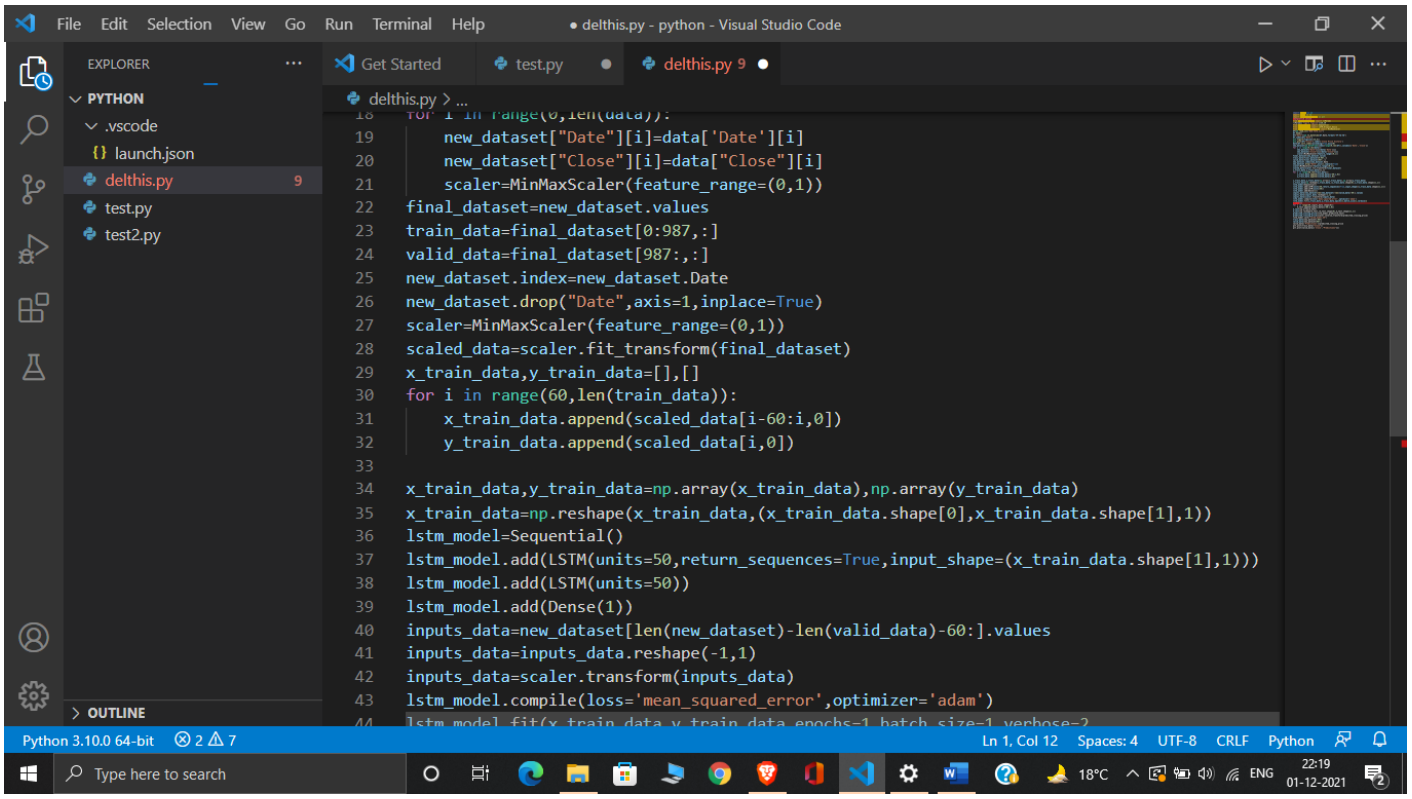
This research is useful in understanding the needs, foundations, frameworks, and strategies for analyzing stock market trends. This research project is organized into eight phases. The order of the research is shown in Fig. 1 and the test method is discussed in "Test Method". Normal course and major challenges are discussed in “Common Ways to Predict a Direct Stock Trend”. “Main Points” discuss key words and ideas. Book reviews are featured in "Book Reviews". In addition, "In the discussion", a discussion of revised literature strategies was introduced along with suggested opportunities to address the challenges posed by the "General Method of Predicting Sensitive Stock News". update.

The dataset:

This section provides details on data extracted from public data sources, as well as the latest updated data. Stock market-related data is different, so we first compared the related activities from financial analytics analysis to stock market data analysis to clarify data collection guidelines. After collecting the data, we described the design of the website. Given below, we describe the database, which includes the data structure, and data tables in each data section and definition descriptions.

Description of our website
In this section, we will explain the details. This database contains 3558 shares from the Chinese stock market. In addition to daily price data, daily basic ID data for each stock, we also collected stop and restart history, top 10 shareholders, etc. We list two reasons for choosing 2 years as the duration of this site: (1) majority. of investors who make stock market analysis using data over the past 2 years, (2) using the most recent data may benefit the analysis result. We collected data through an open source API, Tushare , while we also used a web-based writing strategy to collect data from Sina Finance web sites, the SWS Research website.

Project design (screenshot of code)



CHAPTER 3

FUNCTIONALITIES AND

SYSTEM REQUIREMENTS

Six new variables have been created for the prediction of stock closing price. These variables have been used to train the model. The new variables are as follows:

1. Stock High minus Low price (H-L)
2. Stock Close minus Open price (O-C)
3. Stock price's seven days' moving average (7 DAYS MA)
4. Stock price's fourteen days' moving average (14 DAYS MA)

5. Stock price's twenty-one days' moving average (21 DAYS MA)
6. Stock price's standard deviation for the past seven days (7 DAYS STD DEV)

Requirements:

- Visual studio code
- Pytssx3 module
- Excel sheet
- Audio framework
- Normalization
- Api.ai
- Wit.ai

Chapter 4
Result and discussion:

To test the effectiveness of the models, comparisons were made between two five different strategies sector companies namely, JP Morgan, Nike, Johnson and Johnson, Goldman Sachs and Pfizer use both ANN and RF models. Estimated closing prices are below the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE) errors for detecting errors less than predicted. RMSE is calculated using eq:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Oi-Fi)^2}{n}}$$

where 'Oi' refers to the original closing price, 'Fi' refers to the predicted closing price and 'n' refers to the total window size. MAPE has also been used to evaluate the performance of the model and is computed using eq 2

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Oi-Fi}{Oi}\right)\times 100$$

where 'Oi' refers to the original closing price, 'Fi' refers to the predicted closing price and 'n' refers to the total window size. MBE has also been used to evaluate the performance of the model and is computed using eq 3

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(Oi - Fi)$$

where 'Oi' refers to the initial closing value, 'Fi' refers to the predicted closing value and 'n' refers to the total value.
window size. Figure 2 represents graphs showing the actual closing price of the

stock in relation to the predicted closing price
stock of five different companies using ANN. Figure 3 represents graphs showing the actual closing price of a stock
predicted the closing price of shares using RF. Comparative analysis of RMSE, MAPE and MBE values obtained
using the ANN and RF model shown in Table 2, it can be noted that ANN shows better stock speculation results.

This section briefly describes the background of stock forecasting using news analysis and outlines key concepts that provide the basis for all discussions. First, it introduces the background and conditions associated with stock market predictions. Second, it discusses stock quotes or financial data, features obtained and their use. It discusses in detail the techniques that purify text data and extract the key features from it. Then it discusses speculative algorithms that use a combination of numerical and textual data to predict. Lastly, it presents the most basic test metrics to test the performance of guessing algorithms.

Chapter 5
Conclusion and Future Scope

The work consists of three parts: data extraction and preliminary analysis of the Chinese stock market databases, feature engineering, and a long-term stock price-based estimates model (LSTM). We collected, processed and edited 2 years of Chinese stock market data. We reviewed the various strategies commonly used by real-world investors, developed a new algorithmic component, and named it Feature Extensions, which proved to be effective. We applied the Fee Augmentation (FE) method with recursive feature elimination (RFE), followed by principal component analysis (PCA), to create a more efficient and effective feature engineering process. The system is developed by integrating feature engineering process with LSTM speculation model, achieving high-precision speculation that makes it more than a leading model in many related applications. We have also done a thorough evaluation of this work. By comparing the most widely used machine learning models with our proposed LSTM model under the engineering part of our proposed system, we conclude a few heuristic potential questions for future research in both the technical and financial research fields. it could be.

Our proposed solution is a unique performance improvement compared to previous works because instead of proposing another modern LSTM model, we have developed a system for predicting in-depth and advanced learning and the use of comprehensive feature engineering. This is combined with LSTM to make predictions. With the research observed in previous works, we close the gaps between investors and researchers by raising the expansion algorithm feature before the completion of the repetitive feature and found significant improvements in model performance.

Although we have found a positive effect on our proposed solution, this study has great potential for future research. During the testing process, we also found that the RFE algorithm is not sensitive to the period except for 2 days, weekly, twice a week. A possible guide for future research would be to find in-depth research where technical indicators will affect the duration of an unusual period. In addition, by combining the latest sensory analysis techniques and feature engineering with in-depth learning models, there is also a great potential to develop a comprehensive guessing system that is trained in different types of

information such as tweets, news and other text-based data.

We have analyzed the best possible way to predict short-term value trends from a variety of factors: feature engineering, financial domain information, and forecasting algorithm. We then answered three research questions on each issue, respectively: How can we integrate the accuracy of an engineering benefit model? How does the financial institution's findings benefit the modeling model? And what is the best algorithm for predicting short-term price trends?

The first research question is about feature engineering. We would like to know how the feature selection method benefits the performance of the guessing models. Based on many previous operations, we can conclude that stock price data is embedded in high volume, and there is a correlation between factors, which makes price predictions worse. That is also the main reason for many previous works that introduce an engineering feature as a performance module.

The second research question is to evaluate the effectiveness of the results we have released in the financial sector. Unlike previous operations, in addition to the regular testing of data models such as training costs and points, our evaluation will emphasize the effectiveness of additional new features that we have released in the financial sector. We present some features from the financial domain. Although we found only some of the findings in previous works, and the related raw data needs to be processed into useful features. After removing the related elements from the financial system, we combine features with other common technical indicators for voting for high-impact features. There are many things that are said to work well in the financial sector, and we cannot put them all together. Therefore, the approach to successfully transforming the findings from a financial institution into the module data analysis of our system design is a hidden research question we are trying to answer.