**Synopsis Report**

**On**

**CUSTOMER CHURN PREDICTION**

**Submitted as partial fulfillment for the award of**

**BACHELOR OF TECHNOLOGY DEGREE**

**Session 2023-24**

**in**

**CSE-Data Science**

**By:**

**Vikas Kumar Singh**

**2200321549018**

**Under the Guidance of**

**Mr. Vishal Kanaujia**

**Assistant Professor**

**DEPARTMENT OF CSE-DS**

**ABES ENGINEERING COLLEGE, GHAZIABAD**

|  |  |  |
| --- | --- | --- |
|  |  | naac A Grade Engineering College |
| **AFFILIATED TO**  **DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW**  **(Formerly UPTU)** | | |

# Student’s Declaration

I / we hereby declare that the work being presented in this report entitled **“Customer Churn Prediction.”** is an authentic record of my/ our own work carried out under the supervision of Mr. Vishal Kanaujia **, Assistant Professor, CSE-DS.** The matter embodied in this report has not been submitted by us for the award of any other degree.

This is to certify that the above statement made by the candidate(s) is correct to the best of my knowledge.

**Date:**

**Signature of Student Signature of Supervisor**

(Name: Raj Pratap Singh) (Name)

(Roll No.:- 2000321540043) (Post)

Department: CSE-DS

**Signature of Project Coordinator Signature of HOD**

Dr, Dimple Tiwari Mr. Prabhat Singh

Assistant Professor Assistant Professor

# Acknowledgement

We would like to convey our sincere thanks to **Mr. Vishal Kanaujia** for giving the motivation, knowledge and support throughout the course of the project. The continuous support helps in a successful completion of project. The knowledge provided is very useful for us.

We also like to give a special thanks to the department of Information and Technology for giving us the continuous support and opportunities for fulfilling our project.

We would also like to extend our sincere obligation to Mr. Prabhat Singh, Head of Department, Information Technology for providing this opportunity to us.

# Signature of student

**Vikas Kumar Singh**

**(Roll No. 2200321549018)**

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Contents** | **Page No.** |
|  | Student’s Declaration | i |
|  | Acknowledgement | ii |
|  | Abstract | III |
| Chapter 1: | Introduction | 1 |
| Chapter 2:  2.1:  2.2: | Related Work/Methodology  Existing Approaches  Comparative Analysis of Existing Works | 2 |
| Chapter 3: | Project Objective | 3 |
| Chapter 4: | Proposed Methodology | 4 |
| Chapter 5:  5.1: | Design and Implementation  Work Flow Diagram | 5 |
| Chapter 6: | Results and Discussion | 6 |
| Chapter 7: | Conclusion | 7 |
|  | References | 8 |

# ABSTRACT

# the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price. Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our aim is to create software that analyses previous stock data of certain companies, with help of certain parameters that affect stock value. We are going to implement these values in data mining algorithms and we will be able to decide which algorithm gives the best result. This will also help us to determine the In values that particular stock will have in near future. We will determine the patterns in data with help of machine learning algorithms

# Introduction

# OVERVIEW In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profits gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. More over the predicted trends of the market will help the regulators of the market in taking corrective measures. 1.2 AIM AND OBJECTIVE The aim of the project is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data. 1.3 STOCK MARKET A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds. Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques. Financial institutions and active traders have created various proprietary models to beat the market for 2 themselves or their clients, but rarely did anyone achieve consistently higher than the average returns on investment. The challenge of stock market price forecasting is so appealing because an improvement of just a few points of percentage can increase the profit by millions of dollars. This paper discusses the application of Support Vector Machines and Linear Regression in detail along with the pros and cons of the given methods. The paper introduces the parameters and variables which can be used to recognize the patterns in stock prices which can be helpful in future stock prediction and how boosting can be integrated with various other machine learning algorithms to improve the accuracy of our prediction

# 

# **Chapter 2**

Thank you for sharing the information about yfinance. It is indeed a Python library that provides a simple way to download historical market data from Yahoo! Finance’s API . It offers a threaded and Pythonic way to download market data from Yahoo! Finance, including historical prices (open, high, low, close), volume, dividends, stock splits, and various other financial indicators .

The code snippet you shared is correct. It downloads the historical data for Microsoft Corporation (MSFT) for the past month . You can also use yfinance to download data for multiple stocks at once .

Please note that yfinance is not affiliated, endorsed, or vetted by Yahoo, Inc. It’s an open-source tool that uses Yahoo’s publicly available APIs, and is intended for research and educational purposes . You should refer to Yahoo!'s terms of use for details on your rights to use the actual data downloaded . Remember - the Yahoo! finance API is intended for personal use only .

# Chapter 3

# Project Objective

**3.1 OBJECTIVES** The aims of this project are as to identify factors affecting share market, To generate the pattern from large set of data of stock market for prediction of NEPSE and to predict an approximate value of share price to provide analysis for users through web application The objective of the system is to give a approximate idea of where the stock market might be headed. It does not give a long term forecasting of a stock value. There are way too many reasons to acknowledge for the long term output of a current stock. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

**3.2 EXISTING SYSTEM** Nowadays, as the connections between worldwide economies are tightened by globalization, external perturbations to the financial markets are no longer domestic. With evolving capital markets, more and more data is being created daily. The intrinsic value of a company’s stock is the value determined by estimating the expected future cash flows of a stock and discounting them to the present, which is known as the book value. This is distinct from the market value of the stock, that is determined by the company’s stock price. This market value of a stock can deviate from the intrinsic value due to reasons unrelated to the company’s fundamental operations, such as market sentiment. The fluctuation of stock market is violent and there are many complicated financial indicators. Only few people with extensive experience and knowledge can understand the meaning of the indicators and use them to make good prediction to get fortune. Most people have to rely solely on luck to earn money from stock trading. However, the advancement in technology, provides an opportunity to gain steady fortune from stock market and also can help experts to find out the most informative indicators to make better prediction. The prediction of the market value is 10 of paramount importance to help in maximizing the profit of stock option purchase while keeping the risk low.

## Chapter 4

## Proposed Methodology

## Stage 1 : Raw data collection

## Stage 2: Data pre-processing

## Stage 3: Feature Extraction

## Stage 4: splitting data into train set and test set.

* Extraction of Dataset from Twitter API
* Pre-processing of Data to remove special characters, punctuations, Stop Words and Images
* Processing of Data to analyze the polarity of the Dataset
* To use Machine Learning Algorithm and find which fits best for performing Sentiment Analysis
* Results are represented using tables and graphs.

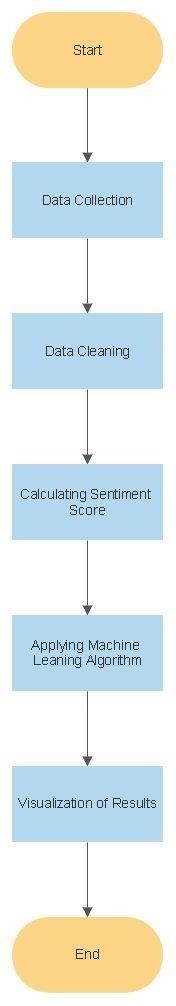
**Fig.1.** Proposed Approach

# Chapter 5

## Design and Implementation

The design and implementation of our project is as follows:

### 5.1. Work Flow Diagram



**Fig.2.** Work Flow Diagram

**4.4 MODULE DECRIPTION** The implementation of this project is divided into following steps **1. Data Preprocessing**

**2. Feature selection**

**3. Data Preprocessing:**

The entries are present in the dataset. The null values are removed using df = df.dropna() where df is the data frame. The categorical attributes (Date,High,Low,Close,Adj value) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like total which can be used as feature for the model.

4.4.2 Feature selection: Features selection is done which can be used to build the model. The attributes used for feature selection are Date,Price,Adj close,Forecast X coordinate , Y coordinate, Latitude , Longitude, Hour and month,

4.4.3 Building and Training Model: After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain ,ytrain and xtest, y test. The algorithms model is imported form skleran. Building model is done using model. Fit (xtrain, ytrain). This phase would involve supervised classification methods like linear regression, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.

4.5PYTHON TECHNOLOGY Python is an interpreted, object- oriented programming language similar to PERL, that has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in a number of operating systems, including UNIX- based systems, Mac OS, MS- DOS, OS/2, and various versions of Microsoft Windows

code easier to read. Python offers dynamic data type, ready- made class, and interfaces to many system calls and libraries. It can be extended, using the C or C++language. Python can be used as the script in Microsoft's Active Server Page (ASP) technology. The scoreboard system for the Melbourne (Australia) Cricket Ground is written in Python. Z Object Publishing Environment, a popular Web application server, is also written in the Python language’s 4.5.1 Python Platform Apart from Windows, Linux and MacOS, CPython implementation runs on 21 different platforms. IronPython is a .NET framework based Python implementation and it is cabable of running in both Windows, Linux and in other environments where .NET framework is available.

4.5.2 Python Library Machine Learning, as the name suggests, is the science of programming a computer by which they are able to learn from different kinds of data. A more general definition given by Arthur Samuel is –“Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.” They are typically used to solve various types of life problems. In the older days, people used to perform Machine Learning tasks by manually coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in Machine Learning are:

**4.5.2.1 NumPy** NumPy is a very popular python library for large multi- dimensional array and matrix processing, with the help of a large collection of high- level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High- end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

**4.5.2.2 SciPy:** SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation. 4.5.2.3 Skikit: Skikit- learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit- learn supports most of the supervised and unsupervised learning algorithms. Scikit- learn can also be used for data- for their own projects.

**4.5.2.5 TensorFlow:** TensorFlow is a very popular open- source library for high performance numerical computation developed by the Google Brain team in Google. As the name suggests,

**Tensorflow** is a framework that involves defining and running computations involving tensors. It can train and run deep neural networks that can be used to develop several AI applications. TensorFlow is widely used in the field of deep learning research and application.

**4.5.2.6 Keras:** Keras is a very popular Machine Learning library for Python. It is a high- level neural networks API capable of running on top of TensorFlow, CNTK, or Theano. It can run seamlessly on both CPU and GPU.Keras makes it really for ML beginners to build and design a Neural Network. One of the best thing about Keras is that it allows for easy and fast prototyping

**4.5.2.7 PyTorch:** PyTorch is a popular open- source Machine Learning library for Python based on Torch, which is an open- source Machine Learning library which is implemented in C with a wrapper in Lua. It has an extensive choice of tools and libraries that supports on Computer Vision, Natural Language Processing(NLP) and many more ML programs. It allows developers to perform computations on Tensors with GPU acceleration and also helps in creating computational graphs

**. 4.5.2.8 Pandas:** Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In

**4.5.2.9 Matpoltlib**: Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc,

# Chapter 6

## Results and Discussion

The result we got from analyzing the tweets is given below in Fig.3.



# Chapter 7

**6.1 CONCLUSION** By measuring the accuracy of the Linear Regression algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

**6.2 FUTURE WORK** Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee.

# References

* <https://www.projectpro.io/article/stock-price-prediction-using-machine-learning-project/571#mcetoc_1ftklmk3qi>
* https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning

<https://bvmengineering.ac.in/NAAC/Criteria1/1.3/1.3.4/18CP808_Thesis.pdf>

https://ieeexplore.ieee.org/document/8703332