 Final Year Project - UG

**School of Computing Science and Engineering (SCOPE)**

**B.Tech. CSE/CPS/AL AND ML Capstone Project** **IN HOUSE**

**General Report (Sept 2022 – November 2022)**

**Program: B.Tech. CSE/CPS/AI AND ML Batch: 2019-2023 Course Code: CSE1904**

**Register No.: 19BAI109 Name of the Student: Arnab Karmakar Mobile No. 9721866757**

|  |  |  |
| --- | --- | --- |
| **Project Title:**  **STOCK MARKET PREDICTION USING MACHINE LEARNING** | | |
| **General Report (Sept 2022 – November 2022)**  **The model should be able to predict stocks that are expected to appreciate in the near future. For correct prediction, we will be giving out promoter buying data combined with other attributes of the company. The rationale being a promoter is the best judge of the company and will not buy a substantial amount of stock if he doesn't expect it to do well in the future. The other attributes of the company are fed to increase the accuracy of the model. A corporate promoter is a firm or person who does the preliminary work related to the formation of a company, including its promotion, incorporation, and flotation, and solicits people to invest money in the company, usually when it is being formed. In most cases, the promoter is the founder and operator of the company. Therefore, a promoter is expected to have the best knowledge in terms of both company and the market.**  **PROJECT GOALS**  **The finalized project should be able to achieve the following purposes:**  **● Check if it is possible to create a successful portfolio with the promoter buying sheet.**  **● Create models that give buy/not buy decisions.**  **● Check how the model works in different time frames.**  **● Check if the portfolio gives alpha during the holding period.**  **● Check the impact of changing the return benchmark from zero on accuracy and portfolio**  **return.**  **PROJECT CHALLENGES**  **This section discusses the challenges that were encountered in completing this project. Each problem is briefly described and discussed in detail in subsequent sections of the study. The first and foremost challenge was extracting and managing the historical data of insider trading. The historical data should be sufficiently large enough to train the model and back test the model for its accuracy. Managing the data included filtering the data to only include relevant data, and amalgamating the various transactions company-wise in the monthly time frame to make the data fit for training. This was achieved using Excel functions like filter and pivot table. We also had to extract the historical attributes of companies that were relevant at the time of buying. It was difficult to scrap data through different sites especially stock-related sites as there is too much useless information available, picking the one you need is really difficult. Finding the sites which contain the information you need is another challenging task. Once we had the filtered promoter buying data and attributes of the company, we needed to decide on the structure of the model. The structure that needed to be created was based on financial theory. We had to decide on questions like what timeframes should be included, should there be a benchmark rate, and if yes, what should be the benchmark rate for different time periods.** **PROBLEM STATEMENT** **This part examines the difficulties that were experienced in finishing this undertaking. Every issue is momentarily depicted and examined exhaustively in resulting areas of the review. The most importantly challenge was removing and dealing with the authentic information of insider exchanging. The verifiable information ought to be adequately sufficiently enormous to prepare the model and back test the model for its exactness. Dealing with the information included separating the information to just incorporate important information, and amalgamating the different exchanges organization wise in the month-to-month time period to make the information fit for preparing. This was accomplished utilizing Excel capabilities like channel and turn table. We likewise needed to remove the authentic qualities of organizations that were pertinent at the hour of purchasing. It was hard to scrap information through various locales particularly stock-related destinations as there is a lot of futile data accessible, picking the one you really want is truly troublesome. Finding the locales which contain the data you want is another difficult assignment. When we had the sifted advertiser purchasing information and properties of the organization, we expected to settle on the design of the model. The construction that should have been made depended on monetary hypothesis. We needed to settle on questions like what time periods ought to be incorporated, should there be a benchmark rate, and if indeed, what ought to be the benchmark rate for various time spans?** | | |
| **Implementation** | Patent / SCI / Scopus Indexed Journal Paper / Scopus Indexed Conference Paper/ Scopus  Indexed Book Chapter DATA PREPARATION Data preparation has a very important role in building models. The first and foremost step in this project was extracting insider trading data. The data was extracted from the website of the NSE. The data contained historical insider trading data from the year 2015 to the year 2021. The downloaded data was insider transactions made by different types of parties during the year. It included pledge data, derivatives data, and a lot of other irrelevant data.  After downloading the data, the first step was to filter the transactions. Irrelevant transactions had to be deleted. The data filtering tool available in Excel was used to achieve this. First of all, the transactions that were made by promoters were kept. All other types of transactions were removed. Similarly, other irrelevant transactions were removed like transactions other than market purchase/sale.  Now we have the transactions that are relevant to us. The next step was to amalgamate the transactions of the same company in the same month. Pivot table function of Excel was used to do this. After adding up the transactions for the same month for the same companies, the companies that had net selling positions were removed as we are only concerned with companies that have net positive buying. The next step was a few mathematical calculations like net buying value, average buying price, etc.  The data of other attributes of the stock was also needed to be extracted. For this purpose, the Google finance function of the spreadsheet was used. Also, data was extracted using web scraping. However, this data needed to be placed correctly with all the entries. In order to do these various functions of Excel was used including Index, Match, Left, Char, Substitute, etc. | |
| **Work Status** | ***Excellent / Good / Satisfactory / Needs improve*** | |
| ***Attendance Status*** | **Regular / Irregular** | ***CAM – Max. 5 Marks per week*** |

**Arnab Karmakar (19BAI1090)**

**Signature of the Student with date Name & Signature of the Guide with date**

