|  |  |
| --- | --- |
| Project Name | *Modelling direction detection in selected stocks in the Indian BFSI sector* |
| Background Information | *Back Testing is still becoming a grim prospect, because of several things like value variations, quiet news, and existing noise. Hence, a feasible solution could be to identify and implement more than a few popular stock evaluation strategies.*  *Several Machine-Learning associated techniques that are developed have created the potential to predict the market to an extent.*  *The requirement is to overcome the ambiguities of Fundamental and technical evaluation, and additionally, the glaring development in the modelling strategies has pushed several researchers to check new strategies for stock value forecasting. New innovative strategies are being used for stock price predictions.*  *Daily Trading NSE Data of HDFC, KOTAK, and SBI Bank from the year 2000 to 2022 is being used for this capstone project which would broadly come under BFSI.BFSI comprises of Banking, Financial Services, and Insurance sector. Also, the BFSI industry includes financial service firms such as Broking and Asset Management. The BFSI industry is growing year on year at a 27% rate.* |
| Literature Review | *Documents reviewed for capstone2 included topics namely the impact of algorithmic trading and Stock market prediction using machine learning techniques and analysis thereafter.*  *The effect of Technical and fundamental analysis on investment decisions was researched which included technical analysis for HDFC, KOTAK, and SBI stocks.*  *Supervised and unsupervised machine learning methods were gone through. Other studies included a Decision tree for classification and regression, random forest algorithm, Logistic regression, XGBoost, and KNN.*  *Effects of volatility, trend, and momentum indicators for the prediction of the stock market were researched. Confusion matrix method of evaluation of error metrics was gone through.*  *Some of the research gaps observed were that Feature expansion and elimination techniques in data preparation were lacking details, fundamental analysis wasn't explored enough, and research on volume indicators in the technical analysis was missing. Hyper parameter tuning while discussing machine learning algorithms should have been discussed in more detail.* |
| Statement of the Problem | *There are plenty of Regression algorithms that can be utilized to detect the closing price of any stock. However, risks are more predominant in predicting the exact closing price using both Linear and Non-linear Regression algorithms.*  *When we have lesser data to build a regression model, under fitting scenarios may destroy the accuracy of our machine learning model especially when we are trying to build a linear model with a complex dataset.*  *At certain times, while trying to cater to all kinds of both existent and nonexistent possibilities in data points, over fitting scenarios in regression models may again destroy the accuracy of test data while the accuracy of the trained data may work perfectly fine.*  *The situation requires not completely relying only on regression algorithms to quantitatively predict the exact closing price of any stock. Investors can find plenty of algorithms that detect the exact closing price of any stock but will not tell the direction of the closing price.*  *Therefore, we should try alternate approaches as well which enables us to decide objectively whether say the price of any stock will move up or move down or remain neutral.* |
| Objectives | *•Firstly, the objective of this project is to get the right stock and collect all relevant data to make correct forecasting. Understand the data pattern using Exploratory Data Analysis and perform data preparation which enables the production of clean and well-curated info with extra Features addition that results in more sensible and correct model outcomes.*  *•Secondly, the objective of the project is to build the right models by using multiple Classification Modelling techniques namely LR Classifier, DT Classifier, RF Classifier, KNN Classifier, and XG Boost Classifier to determine the Modelling algorithm which would provide the best accuracy in direction prediction.*    *•Thirdly the objective of the project is to explore state-of-the-art solutions to minimize errors in direction prediction. For every forecasting Technique, there will be errors, and since the stock market has high volatility, hence the chances of errors are more. Therefore, given the historical data, it should be correctly predicted whether the price will move up or move down utilizing precision, recall, and accuracy Metrics used in classification modelling techniques.* |
| Methodology | *The CRISP-DM framework has been used for the project. The process of CRISP-DM is split into Business Understanding, data understanding, data preparation, Modelling, Evaluation, and Deployment. Business understanding provides Fundamental and Technical analysis of HDFC, KOTAK, and SBI stock to demonstrate why the HDFC, KOTAK, and SBI stock dataset has been used for this project. Data understanding explains the different columns used in the HDFC, KOTAK, and SBI dataset. Data preparation explains that Handling Missing values, Features Addition, and Data Scaling using MinMax Scaler were the steps used for processing the dataset before being used for Modelling.LR Classifier, DT Classifier, RF Classifier, KNN Classifier, and XG Boost Classifier were used in the Data Modelling phase. The data evaluation phase examines the results of different Modelling techniques which were used in the Data Modelling phase. Deployment speaks about developing a front-end API for the deployment Dashboard.* |
| Proposed Solution/  Expected Results | The objective of the project is to build various classification models to predict the direction of the closing price for the stock under consideration. The direction of the closing price is to be detected to determine the extent of accuracy of Modelling predictions.  The 6-day consecutive closing price for the stock under consideration is being taken. These 6 days' consecutive closing prices will be getting tabulated week on week for the entire dataset and will be utilized as 6 different feature variables for building the classification Model.  Computation is being done to evaluate whether it is a positive change, negative change, or no change between the 7th and 8th-day closing price. The rule is being set to determine what has to be seen as a direction change.0.5% difference,1% difference, and 1.5% difference -these are different classes of direction for which the rule is being set which is to be followed for computing the direction change as either positive change, negative change, or no change. 0.7% difference as a class of direction can also be used in place of 0.5% change if that gives a better directional indicator.  Therefore, given 6-day data, it will be predicted whether on 8th data the closing price of the stock under consideration is going to increase or decrease or remain the same. Based on the close price, the direction of the next day's closing price is to be predicted as to whether it is going to increase or decrease. Several target variables can be created based on whether the change is 0.5 or 0.7%, 1%, or 1.5%, and then it is to be determined for each of these target variables what would be the prediction accuracy.  Once it is determined say for example 0.7% change has the best prediction accuracy among all different classes of direction namely 0.7% change, 1% change, and 1.5% change then the range of consecutive days to be utilized as feature variable is increased to 10 days and 14 days consecutively. These 10 days and 14 days consecutive closing prices will be getting tabulated week on week for the entire dataset and will be utilized as different feature variables for again building the classification Model for reconfirmation of the directional indication.  Similarly, all technical indicators can be utilized in Technical Analysis to build another set of classification Models. All different types of technical indicators namely momentum indicators, trend indicators, volatility indicators, and volume indicators can be utilized as feature variables based on the input dataset and different classification models can be built to determine their prediction accuracy. Generally, Open price, High price, low price, close price, and volume for the stock under consideration will be utilized to derive feature variables from technical indicators. These derived feature variables will then be used as the feature variables to predict the direction of the close price.  If for example say 10000 is invested in HDFC stock, and say it is predicted as a positive change for the next day. The same prediction process is repeated say 100 times and evaluated how much is the net gain and loss based on that.  For a stop loss of 2.0 reward-risk ratio for approximately 0.8 Precision would be 2\*.8/2\*.2=4:1 if a 0.5% difference in consecutive day close price for any stock is only 2.0.for higher percentage difference reward to risk ratio would be higher.  Similarly, it can be shown that average returns using the Go Long Direction Prediction would be much higher compared to Average stock market returns using Bank interest returns which range from 7.0-7.5%. This would have immense benefits in doing this project from the perspective of the stakeholders. |
| Detailed Scope of Work: | *Daily Trading Data of HDFC company from the year 2000 to 2022 is being used for this study. This study uses NSE Data.*  *The 6-day consecutive closing price for the stock under consideration is being taken. These 6 days' consecutive closing prices will be getting tabulated week on week for the entire dataset and will be utilized as 6 different feature variables for building the classification Model.*  *The data is prepared week on week to determine how exactly computation is being done for what is up, what is down, and what is neutral. Say, for example, anything more than 0.7% change can be positive up, anything less than -0.7% change can be positive down, and anything between 0.7% and -0.7% change can be taken as Neutral.*  *The difference between the 7th and 8th-day Closing price is determined. If the 8th-day closing price is seen as an increase from the 7th day by 0.7% or more, the direction of the closing price can be made positive. If the 8th-day closing price is seen as a decrease from the 7th day by -0.7% or less, the direction of the closing price can be made negative. Between -0.7% and 0.7% that the direction of the closing price for the stock under consideration can be treated as sideways.*  *For data within the 0.7% and -0.7% band, usually, the advice to the investor will be to hold on to existing portfolios and wait for the direction of the closing price to show as either a negative or positive change. If there is a negative change, usually the advice to the investor will be to not invest in such a circumstance. If there is a positive change the investor will be suggested to invest.*  *It is to be determined how many times the positive changes are identified by predicting and how many times positive changes are there in the actual data. This will be utilized to evaluate how many times true positives were detected and how many times the false positives were predicted in the prediction. A similar process is to be followed for detecting true negatives and false negatives. A similar process is to be followed for detecting true neutrals and false neutrals. Based on prediction accuracy, it can be suggested whether to invest or not to invest to the prospective investor.*  *Computation is being done to evaluate whether it is a positive change, negative change, or no change between the 7th and 8th-day closing price. The rule is being set to determine what has to be seen as a direction change.0.5% difference,1% difference, and 1.5% difference -these are different classes of direction for which the rule is being set which is to be followed for computing the direction change as either positive change, negative change, or no change. 0.7% difference as a class of direction can also be used in place of 0.5% change if that gives a better directional indicator.*  *Therefore, given 6-day data, it will be predicted whether on 8th data the closing price of the stock under consideration is going to increase or decrease or remain the same. Based on the close price, the direction of the next day's closing price is to be predicted as to whether it is going to increase or decrease. Several target variables can be created based on whether the change is 0.5 or 0.7%, 1%, or 1.5%, and then it is to be determined for each of these target variables what would be the prediction accuracy.*  *It will be identified regarding the extent of accuracy by which positive, negative, or neutral changes can be predicted based on 0.3 of the existing test data. Based on the prediction, the prediction accuracy is determined.*  *Once it is determined to say for example 0.7% change has the best prediction accuracy among all different classes of direction namely0.7% change, 1% change, and 1.5% change then the range of consecutive days to be utilized as a feature variable is increased to 10 days. Therefore, a 10-day consecutive closing price for the stock under consideration is being taken. These 10 days' consecutive closing prices will be getting tabulated week on week for the entire dataset and will be utilized as different feature variables for building the classification Model.*  *The difference between the 11th and 12th-day Closing price is determined. If the 12th-day closing price is seen an increase from the 11th day by 0.7% or more, the direction of the closing price can be made positive. If the 12th-day closing price is seen a decrease from the 11th day by -0.7% or less, the direction of the closing price can be made negative. Between -0.7% and 0.7%, the direction of the closing price for the stock under consideration can be treated as sideways. The prediction accuracy is determined to confirm that say 0.7% change has the best prediction accuracy among all different classes of direction even when the range of consecutive days to be utilized as a feature variable is increased to 10 days.*  *A similar process is again repeated for a range of consecutive days to be utilized as the feature variable increases to 14 days. The prediction accuracy is determined to confirm that say 0.7% change has the best prediction accuracy among all different classes of direction even when the range of consecutive days to be utilized as a feature variable is increased to 14 days.*  *Similarly, all technical indicators can be utilized in Technical Analysis to build another set of classification Models. All different types of technical indicators namely momentum indicators, trend indicators, volatility indicators, and volume indicators can be utilized as feature variables based on the input dataset and different classification models can be built to determine their prediction accuracy. Generally, Open price, High price, low price, close price, and volume for the stock under consideration will be utilized to derive feature variables from technical indicators. These derived feature variables will then be used as the feature variables to predict the direction of the close price. Four different Classification models based on four different types of technical indicators are being built.*  *For momentum indicators, the Awesome Oscillator Indicator, KAMA Indicator, PercentagePriceOscillator, PercentageVolumeOscillator, ROCIndicator, RSIIndicator, Stochastic Oscillator, TSI Indicator, Ultimate Oscillator, WilliamsR Indicator are being utilized as the feature variables to predict the direction of the closing price and determine the prediction accuracy.*  *For trend indicators,* *ADX Indicator, Aroon Indicator, CCI Indicator, Ichimoku Indicator,* *KST Indicator, MACD, PSAR Indicator, EMA Indicator, WMA Indicator, and Vortex Indicator are being utilized as the feature variables to predict the direction of the closing price and determine the prediction accuracy.*  *For volatility indicators, AverageTrueRange, Bollinger Bands, Donchian Channel, Keltner Channel, and Ulcer Index are being used as feature variables. The lower and upper bands of these volatility indicators are also utilized as feature variables and the direction of the closing price is predicted to determine what the prediction accuracy is.*  *For volume indicators, AccDistIndex Indicator,* *ChaikinMoneyFlow Indicator, EaseOfMovement Indicator, ForceIndex Indicator,* *MFI Indicator, OnBalanceVolume Indicator, VolumePriceTrend Indicator,* *VolumeWeightedAveragePrice, NegativeVolumeIndex Indicator, DailyLogReturn Indicator are used as feature variables, and the direction of the closing price is predicted as to whether it is a positive change or Not a positive change to determine what is the prediction accuracy.*  ***Go Long Direction Prediction using Technical Indicators:*** *The direction of the close price is estimated as the percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is more than 0.5%, the direction of the closing price is treated as positive, and suitable for long Trading in the stock market. Otherwise, the direction of the close price is treated as non-positive and not suitable for long Trading in the stock market.*  *When the majority of the 15 various models or all of them move in the same direction, a choice on whether to invest or not to invest in the stock under consideration must be made. What works in the Indian stock market must be proven with evidence.*  *The entire process is needed to be tried and tested for a different dataset altogether to ensure that Any stock on the stock market can utilize the same procedure to forecast whether to invest or not to invest, which is helpful.*  *Daily Trading Data of SBI and Kotak Mahindra company from the year 2000 to 2022 is being used to repeat the entire process which had been implemented for the HDFC dataset.* |
| Support needed from the Program Office | *Program Office(JBS sir)* |
| References | 1. *Alhomadi, A. (2021). Forecasting stock market prices : A machine learning approach. Digital Commons, 11(2), 16–36.* 2. *Anjani, T., & Syarif, A. D. (2019). The Effect of Fundamental Analysis on Stock Returns using Data Panels ; Evidence Pharmaceutical Companies listed on IDX. International Journal of Innovate Science and Research Technology, 4(7), 500–505.* 3. *Elbialy, B. A. (2019). The Effect of Using Technical and Fundamental Analysis on the Effectiveness of Investment Decisions of Traders on the Egyptian Stock Exchange. International Journal of Applied Engineering Research, 14(24), 4492–4501. http://www.ripublication.com* 4. *Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. Big Data and Society, 7(1). https://doi.org/10.1177/2053951720926558* 5. *Huang, Y., Capretz, L. F., & Ho, D. (2021). Machine Learning for Stock Prediction Based on Fundamental Analysis. 2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings. https://doi.org/10.1109/SSCI50451.2021.9660134* 6. *Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. ACM International Conference Proceeding Series, 412–419. https://doi.org/10.1145/3453892.3461323* 7. *Mohapatra, S., & Misra, A. K. (2020). Momentum returns: A portfolio-based empirical study to establish evidence, factors and profitability in Indian stock market. IIMB Management Review, 32(1), 75–84.* [*https://doi.org/10.1016/j.iimb.2019.07.007*](https://doi.org/10.1016/j.iimb.2019.07.007) 8. *Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics (Switzerland), 10(21). https://doi.org/10.3390/electronics10212717* |