


Dissertation Proposal Form

Date of Submission: 31st May 2024

Name	Gaurab Shrestha
Student Id	230188112
Module Code	COM7040M
Project Title	Performance Evaluation and Analysis of Machine Learning Algorithms in Stock Price Prediction
Supervisor Name	Sangita Pokhrel
Supervisor Approval	Yes
Supervisor Signature	 27-05-2024

Section 1: Academic

This section helps Academic staff assess the viability of your project. It also helps identify the most appropriate supervisor for your proposed research. This proposal will be referred to as a point of discussion by your supervisor in seminar sessions.

NAME: Gaurab Shrestha	STUDENT NUMBER: 230188112
PROPOSED TITLE OF PROJECT: Performance Evaluation and Analysis of Machine Learning Algorithm in Stock Price Prediction	
BRIEFLY DESCRIBE YOUR FIELD OF STUDY: My dissertation topic is related to Machine Learning and Stock Market. Stock Market A stock market, equity market, or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as stock that is only traded privately, such as shares of private companies that are sold to investors through equity crowdfunding platforms. Investments are usually made with an investment strategy in mind. Well, with the growth of broad financial market, nowadays, people are more concerned about buying and selling stocks, trading in forex and so on. But, with the less knowledge and idea about the stock market, most of the people are losing their money. So, with the advancement of technology and computing, now it is possible to predict the stock market trends which will happen in future. Machine Learning This is made possible with the help of machine learning and historical data. The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data is called machine learning. The historical stock market data has some trends and patterns with the advancement of time. So, we can find the patterns and trends with the help of machine learning algorithms and use this model to predict the future stock market price and its trends. Also, the field of machine learning is advancing and new algorithms are developing for the purpose of training the model with high accuracy. So, I am going to use different machine learning algorithms like SVM, Linear Regression, Random Forest, ARIMA, LSTM, etc. to predict the stock price, analyse and evaluate which machine learning algorithms is best for the prediction of stock market price.	
WHAT QUESTION DOES YOUR PROJECT SEEK TO ANSWER? How effectively and accurately can different machine learning algorithms and models predict future stock price trends and which algorithm performs best in terms of accuracy and computational efficiency.	
WHAT HYPOTHESIS ARE YOU SEEKING TO TEST? Prediction Accuracy Hypothesis: Different machine learning algorithms have varying levels of accuracy in predicting stock prices. <ul style="list-style-type: none">• Null Hypothesis (H0): There is no significant difference in the prediction accuracy of different machine learning algorithms.• Alternative Hypothesis (H1): There is a significant difference in the prediction accuracy of different machine learning algorithms.	

Model Efficiency Hypothesis: Some machine learning algorithms are more computationally efficient than others in stock price prediction.

- Null Hypothesis (H0): All machine learning algorithms have similar computational efficiency when used for stock price prediction.
- Alternative Hypothesis (H1): There is a significant difference in the computational efficiency of different machine learning algorithms.

WHAT ARE THE PROBABLE PROJECT OUTCOMES?

The probable outcomes of the project are:

- The project outcome helps investors make informed decisions by forecasting future price movements, better investment strategies, identify the potential risks and early warnings about unfavourable market conditions.
- Outcome help to identify mispriced stocks by comparing predicted prices with current market prices.
- ‘Detailed Evaluation of Prediction Accuracy for Each Machine Learning Algorithm using Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-Squared (R2)’
- ‘Identification of the Algorithm with the Highest Prediction Accuracy’
- ‘Feature Importance and Selection in Prediction of Stock Price’
- ‘Effects of Sentiments in Stock Price Prediction’.

PLEASE PROVIDE A BRIEF BIBLIOGRPAHY OF 2-4 KEY TEXTS FOR YOUR STUDY (USE HARVARD REFERENCE STYLE)

Khanna, M. et al., 2022. Performance Evaluation of Machine Learning Algorithms for Stock Price and Stock Index Movement Prediction Using Trend Deterministic Data Prediction. *International Journal of Applied Metaheuristic Computing*, 13(1), pp. 1-30.

Rizvi, D. R. & Khalid, M., 2024. Performance Analysis of Stocks using Deep Learning Models. *5th International Conference on Innovative Data Communication Technologies and Application*, 233(2024), pp. 753-762.

Saud, A. S. & Shakya, S., 2019. Analysis of Gradient Descent Optimization Techniques with GRU for Stock Price Prediction: A Case Study on Banking Sector of Nepal Stock Exchange. *Journal of Institute of Science and Technology*, 24(2), pp. 17-21.

Saxena, S., 2024. *What is LSTM? Introduction to Long Short-Term Memory*. [Online] Available at: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> [Accessed 15 May 2024].

Sonkavde, G. et al., 2023. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, 11(94).

PLEASE NAME ANY MEMBER OF THE ACADEMIC TEAM YOU HAVE DISCUSSED THIS POTENTIAL PROJECT:

<i>(staff use only) Project Approved by Academic Team?</i>	YES		NO
Any other Academic Staff comments			

Section 2: Technical

This section is designed to help the technical team ensure the appropriate equipment to support each project has been ordered. It also exists to help you fully ascertain the technical requirements of your proposed project. In filling out this section please note that we do not 'buy' major items of equipment for student projects. However, if a piece of equipment has a use to the department beyond the scope of a single project, we will consider purchasing it. Though purchasing equipment through the university is often is a slow process.

PLEASE DESCRIBE YOUR PROJECT IN TECHNICAL TERMS:

This project is machine learning based project where different machine learning algorithms are going to be used to forecast, analyse and evaluate the performance of these algorithms in stock price in future. Talking about the algorithms, SVM, Linear Regression, Random Forest, ARIMA, LSTM, etc are going to be used.

Talking about the datasets, I am going to use two datasets differently to train the model. One is **Alphabet Inc. (Google) Stock Historical Prices and Data** which is available in Yahoo Finance website (<https://finance.yahoo.com/quote/GOOG/history/?period1=1093996800&period2=1716267975>) and another one is **NABIL** bank stock price which is from **Nepali Stock Market (NEPSE)** that can be found in Kaggle and also, I am going to scrap the data from ShareSansar website (<https://www.sharesansar.com/company/nabil>).

For the web scraping of data from Sharesansar, got the approval from the sharesansar and discussed it with the supervisor as well.

WHAT EXISTING LAB EQUIPMENT DO YOU NEED ACCESS TO UNDERTAKE YOUR PROPOSED PROJECT:

High-end desktop to train the model.

PLEASE LIST ANY **MINOR** EQUIPMENT YOU MUST PURCHASE TO COMPLETE YOUR RESEARCH PROJECT: (e.g., Switches, Resistors, Raspberry pi, Arduino etc)

No

PLEASE LIST ANY **MAJOR** EQUIPMENT YOU REQUIRE TO COMPLETE YOUR RESEARCH PROJECT ALONG WITH LINKS TO WHERE IT MAY BE PURCHASED (eg a Drone, mobile phone etc).

No

HAVE YOU DISCUSSED THE FEESIBILITY OF YOUR PROJECT WITH A MEMBER OF THE TECHNICAL TEAM? IF SO, WHO?

No

(Staff use only) Project Approved by Technical Team?

YES

NO

Please comment on the Feasibility of the project:

Section 3: Ethics Approval

This section of the form will help ascertain if you need to complete and undergo the universities research ethics approval process. Please answer all questions honestly.

Question	Yes	No
Does your Research involve any of the following? Human participants / subjects, Human tissue, Documents		✓
Will the research require the collection of primary source material that might be considered offensive or illegal to access or hold on a computer? (e.g. studies related to state security, pornography, abuse, illegal behaviour or terrorism).		✓
Does your research concern group which may be construed as terrorist or extremist?		✓
Will the research involve visual/vocal methods where participants may be identified?		✓
Will the research involve the use of genetic data (inherited/acquired genetic characteristics resulting from the analysis of a biological sample)?		✓
Will the study require the co-operation of a gatekeeper to give access to, or to help recruit, participants? (e.g., headteacher or group leaders publicising your work)		✓
Will it be necessary for participants to take part in the study without their knowledge or consent at the time?		✓
Will the study involve recruitment of patients through the NHS?		✓
Will inducements be offered to participants? (eg the offer of being entered into a prize draw)		✓
Does the study involve participants who are particularly vulnerable or unable to give informed consent? (e.g. participants under 18. Adults with learning disabilities, the frail elderly, or anyone who may be easily coerced due to lack of capacity)		✓
Is there a possibility that the safety of the researcher may be in question?		✓
Will the study require participants to commit extensive time to the study?		✓
Are drugs, placebos or any other substances to be administered to participants, or will the study involve invasive, intrusive or potentially harmful procedures of any kind?		✓

If there are experimental and control groups, will being in one group disadvantage participants?		✓
Is an extensive degree of exercise or physical exertion involved?		✓
Will blood or tissue samples be obtained from participants?		✓
Could the study induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life?		✓

*This part of Section 3 requires you to thoroughly **identify** and **mitigate** the ethical challenges of your research project. This is required to enable the computer Science ethics panel to properly consider if your proposed project requires you to submit a formal proposal to the university ethics panel.*

With your answers to the previous questions in mind, please describe the main ethical challenges of your research project and how you propose to mitigate them. Your discussion may include material not covered in the above questions. Please be as thorough as possible:

Challenges:

1. Market Manipulation and Unfair Advantage

Challenge: The application of machine learning algorithms in stock price prediction can lead to market manipulation or provide an unfair advantage to certain market participants. Advanced predictive models might be used to exploit market inefficiencies, potentially disadvantaging retail investors and leading to market distortions.

2. Bias and Fairness

Challenge: Machine learning models can inherit biases present in historical data, leading to unfair or discriminatory outcomes. For example, a model might favour certain stocks or sectors based on historical trends that do not reflect future potential or fairness considerations.

3. Data Privacy and Security

Challenge: The use of historical stock data and potentially sensitive market information poses risks related to data privacy and security. Unauthorized access or data breaches could lead to significant financial and reputational damage.

Mitigation:

- **Transparency:** Promoting transparency in the development and deployment of predictive models. Publishing research findings and methodologies to ensure that the broader financial community is aware of the techniques being used.
- **Regulatory Compliance:** Adhere to regulations set by financial authorities (e.g., SEC, FCA) to prevent insider trading and market manipulation. Implementing controls to ensure that the use of these models complies with legal and ethical standards.

- **Ethical Guidelines:** Developing and following a strict code of ethics for the use of machine learning in financial markets, ensuring that the technology is used to improve market efficiency rather than exploit it.
- **Data Audits:** Conducting thorough audits of training data to identify and mitigate biases. Ensure that data pre-processing steps include techniques for bias correction.
- **Fairness Metrics:** Incorporating fairness metrics into model evaluation to monitor and mitigate any unfair biases that may arise in predictions.
- **Diverse Data Sources:** Using a diverse set of data sources to train models, ensuring that predictions are not skewed by any single data perspective.
- **Data Encryption:** Using robust encryption methods to protect data at rest and in transit. Ensure that all sensitive data is securely stored and accessed.
- **Access Controls:** Implementing strict access controls and audit trails to monitor and restrict access to sensitive financial data.