A red and black logo

Description automatically generated

**Project Definition Document (PDD)**

**Degree title:** BSc. Computer Science

**Project name:** Forcastock

**Project title:** Machine Learning-Enhanced Trading: Applying Machine Learning Models to Predict the Stock Market Prices

**Name:** Alex Elemele

**Email:** [Alex.Elemele@city.ac.uk](mailto:Alex.Elemele@city.ac.uk)

**Consultant:** Atif Riaz

**Proposed by:** Alex Elemele

***Arrangements for proprietary interests:*** *None*

**Word count:** 1500(Without Ethics Checklist)

# Problem to be solved

The challenge addressed by Forcastock includes the need to conduct research within the sector of machine learning to receive stock market predictions. Essentially, the results of the research should yield an engaging environment for individuals interested in (day-) trading by developing an Interface which will apply the stock predictions. Current trading platforms may lack features that include selecting stock price prediction models on certain stocks (Vantage, 2023)I have appended a research paper at the end of this proposal in the reference section (Ananda Chatterjee, 2024). While the research does explore machine learning algorithms (supervised and unsupervised), it does not dive into areas like reinforcement learning. Furthermore, the need to explore different targets (aside from closing price) and possibly automate popular trading patterns which experienced traders use should be considered as well. All this information should be made available for (day-) traders in the form of a user interface. Key issues include:

1. **Insufficient Day Trading Practice Tools:**

* Existing trading platforms may not cater to users seeking a dedicated space for day trading practice and skill enhancement. Forcastock adds something unique by combining the functionality of a trading platform with a stock prediction model. While existing trading platforms (Vantage) offer demo accounts for practice, they do not include integrated stock prediction capabilities. By incorporating machine learning-based stock prediction into Forcastock, the user gains access to valuable insights and forecasts that can inform their trading by aiding them in making unbiased decisions.

1. **Motivational Gaps:**
   * The absence of gamified elements and motivational structures may lead to a lack of sustained interest and commitment among users to regularly participate in day trading simulations.

Solving the above issues aims to transform (day-) trading practice into a dynamic and educational experience, motivating users to develop their skills and stay engaged with the platform daily while increasing their knowledge on increasing financial capital. Furthermore, the stock prediction will involve machine learning techniques to ensure reliable test results. Users will not only enhance their day trading skills but also gain insights into market trends, contributing to a comprehensive and dynamic learning experience. As for the dataset, Yahoo Finance provides comprehensive stock market data, including historical and real-time information. For this project, we will utilize the yfinance API, a Python library that facilitates easy access to Yahoo Finance's data. (Yahoo finance, 2024)

# Objectives

## Single Main Objective

The main objective of this research project (Forcastock) is to develop and evaluate machine learning models for stock prediction, leveraging techniques such as supervised learning, unsupervised learning and reinforcement learning. Following the research, an interactive user interface employing the machine learning models shall be implemented. The user shall be able to input search criteria (e.g stock name) and be given a stock prediction.

### Sub Objectives

The sub-objectives which must be met to achieve this are:

* Forcastock shall conduct extensive research on machine learning (possibly statistical) models to achieve accurate stock price predictions.

How it is tested: Accuracy can be achieved by testing and training procedures using historical stock price data.

* Forcastock shall implement a User Interface which will allow users to input search parameters and receive predictions

How it is tested: The User Interface successfully accepts and displays expected results.

* Forcastock shall have a server to serve client requests

How is it tested: The backend processes input and sends response back to the user

After these basic sub-objectives are met, the project shall implement the following additional features:

* Forcastock shall implement a demo trading environment where the user can allocate a specific amount of money to a demo account and simulate trades using stock predictions

How is it tested: A User will navigate the user interface and simulate trades on Forcastock.

* Forcastock shall implement basic trading indicators which will serve as an unbiased trading confirmation for the user.

How is it tested: A user will interact with the system and evaluate the behaviour

* Forcastock shall implement a reminder (streak) system to keep the user engaged with the system.

How it is tested: Have the user place a trade and not interact with the system for a certain amount of time.

*Project Beneficiaries*

The trading application outlined in this project will benefit various stakeholders.

1. **End Users (Individual Traders):**
   * *How They Benefit:* Individual traders will experience an enriched trading experience with access to a platform powered by machine learning for live stock predictions. The simulated trading environment on the demo account enables risk-free skill development, strategy testing, and continual learning, enhancing their proficiency in day trading.
2. **Developers in Financial Technology and Machine Learning:**
   * *How They Benefit:* Developers specializing in financial technology and machine learning will gain valuable insights and hands-on experience in implementing innovative features. The project involves intricate technical aspects, contributing to the growth of expertise within the developer community.
3. **Academic Community:**
   * *How They Benefit:* The academic community gains from the practical application of machine learning in financial technology. The project results contribute to academic endeavours, serving as a valuable resource for students, researchers, and projects exploring the intersection of machine learning and stock trading.
4. **Stock Market Enthusiasts:**
   * *How They Benefit:* The community of stock market enthusiasts benefits from the trading application's features, including educational content, advanced analytics, and simulated trading. The project aims to elevate the overall understanding of day trading and market dynamics.

# Work Plan

The development of this project will be guided by the Agile software development methodology. The project will be organized into sprints, each lasting between 10 to 14 days, to accomplish specific project goals within these defined timeframes. Throughout the entire project duration, notes will be taken during consultations, research findings, and discussions on programming methodologies. Regular testing of the Forcastock software will be conducted, with detailed documentation of the testing process.

I plan to write and update the final project report after the completion of each sprint. This iterative reporting process ensures a real-time reflection on progress, challenges faced, and lessons learned during the development phases.

While I am confident in the feasibility of this project, I acknowledge the necessity for thorough research on frameworks, particularly Streamlit used for this project and machine learning models.

The Gantt Chart and associated dates provide a visual representation of the project timeline. The order of responsibilities is respective to the bar order displayed on the Gantt Chart.

**A screenshot of a computer screen

Description automatically generated**

Note: **Streaks**, refer to a consecutive series of actions or events. In the context of Forcastock, streaks are used to encourage user engagement by letting the user know how many subsequent days a trade was placed. The user might not want to lose a streak and for that reason, the user will try to keep the streak alive by placing a trade every day.

Also Note, that the Integration of the Stock Prediction into the web application is the last functionality implemented to ensure a Minimum Viable Product (MVP).

*Project Risks*

There are several project risks when it comes to developing Forcastock:

1. **Technical Complexity:** The integration of machine learning models and real-time market data into the platform may introduce technical complexities. Unforeseen challenges in implementing these technologies could delay the project. // add time complexity concerns
2. **Research of new Technologies:** While it is highly unlikely that I have underestimated the time required to learn specific aspects of the project if this occurs, there may be a need to prioritize the essential features of the platform, potentially sacrificing additional functionalities like adding streaks or additional trading functionality to ensure a minimum viable product.
3. **Data loss:** Over the whole course of the project, the code will be constantly pushed to a private GitHub repository
4. **Lack of Expertise:** If the developer lacks experience in financial markets, trading, or machine learning, there is a risk of overlooking critical aspects of the project. I believe to have sufficient knowledge to conduct the necessary research and to implement the user interface as I have some experience as a trader and have navigated the real market before. I also undertook both AI elective modules (IN3062 and IN3063) and gained extensive knowledge of machine learning procedures.
5. **User Financial Risk Considerations:** Forcastock poses no risk to its potential users. However, should users interpret its predictions as financial advice and make consequential decisions, such as investments, there is a potential risk of financial loss. To mitigate this, it will be explicitly communicated that Forcastock is designed as an educational tool and should not be seen as an encouragement for users to make investment decisions based on its findings.

# Works Cited

Ananda Chatterjee. (2024, 02 01). *arvix*. Retrieved from arvix: https://arxiv.org/pdf/2111.01137.pdf

Vantage. (2023, 02 01). *Vantage*. Retrieved from Vantage: https://www.vantagemarkets.co.uk/?utm\_source=TMBTPS01&utm\_medium=cpc&utm\_campaign=RET-Brand&utm\_content=text&utm\_term=vantage&ls=uk\_uk\_en\_tmbtps01\_cpc\_RET-Brand\_text&gad\_source=1&gclid=CjwKCAiA8YyuBhBSEiwA5R3-E5CSOSvXGucVhoCJ1fHNRqBwSzb-WeZQHHLVL2857Wy9gC

Yahoo finance. (2024, 02 01). *Yahoo finance*. Retrieved from Yahoo finance: https://finance.yahoo.com/sectors/financial-services/financial-data-stock-exchanges?.tsrc=fin-srch

Top of Form

Bottom of Form

*Additional References*

LSTM for stock price prediction:

<https://www.analyticsvidhya.com/blog/2021/10/machine-learning-for-stock-market-prediction-with-step-by-step-implementation/>

|  |
| --- |
| **Research Ethics Review Form for BSc and MSci Projects**  **Computer Science Research Ethics Committee (CSREC)**  <http://www.city.ac.uk/department-computer-science/research-ethics> |

Undergraduate students undertaking their final project in the Department of Computer Science must consider the ethics of their project work and ensure that it complies with research ethics guidelines and the law for data protection. In some cases, a project will need approval from an ethics committee before it can proceed. Usually, but not always, this will be because the student is involving other people (“participants”) in the project.

To ensure that they give appropriate consideration to ethical issues, all students must complete this form and attach it to their project definition document (PDD). There are two parts:

***PART A: Ethics Checklist***. All students must complete this part. The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval.

***PART B: Ethics Proportionate Review Form****.* Students who have answered “no” to all questions in A1, A2 and A3 and “yes” to question 4 in A4 in the ethics checklist must complete part B as well. The project supervisor or consultant has delegated authority to provide approval in such cases that are considered to involve MINIMAL risk. The approval may be ***provisional*** *– identifying the planned work with human end user participants as*likely to involve MINIMAL RISK. In such cases you must additionally seek ***full approval*** from the supervisor or consultant as the project progresses and details are established. You must obtain ***full approval*** in writing, before recruiting and engaging with human end users participants for your project.

**Part A: Ethics Checklist**

|  |  |  |
| --- | --- | --- |
| **A.1 If you answer YES to any of the questions in this block, your consultant/supervisor must have obtained approval for the project from an appropriate external ethics committee, and you need to have received written confirmation of this from him/her. Students cannot themselves apply for ethics approval in this case as the project is considered high risk". This type of research is not covered by City’s process, and external approval from an appropriate institution is required.** | | *Delete as appropriate* |
| 1.1 | Does your research require approval from the National Research Ethics Service (NRES)? | **NO** |
| 1.2 | Will you recruit participants who are covered by the Mental Capacity Act 2005? | **NO** |
| 1.3 | Will you recruit any participants who are covered by the Criminal Justice System, for example, people on remand, prisoners and those on probation? | **NO** |
| **A.2 If you answer YES to any of the questions in this block your consultant/supervisor must have obtained appropriate ethics committee approval** | | *Delete as appropriate* |
| 2.1 | Does your research involve participants who are unable to give informed consent?  For example, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf. | **NO** |
| 2.2 | Is there a risk that your research might lead to disclosures from participants concerning their involvement in illegal activities? | **NO** |
| 2.3 | Is there a risk that obscene and or illegal material may need to be accessed for your research study (including online content and other material)? | **NO** |
| 2.4 | Does your project involve participants disclosing information about protected characteristics (as identified by the Equality Act 2010)?  *For example: racial or ethnic origin; political opinions; religious beliefs; trade union membership; physical or mental health; sexual life; criminal offences and proceedings* | **NO** |
| 2.5 | Does your research involve you travelling to another country outside of the UK, where the Foreign & Commonwealth Office has issued a travel warning that affects the area in which you will study?  *Please check the latest guidance from the FCO -* [*http://www.fco.gov.uk/en/*](http://www.fco.gov.uk/en/) | **NO** |
| 2.6 | Does your research involve invasive or intrusive procedures?  These may include, but are not limited to, electrical stimulation, heat, cold or bruising. | **NO** |
| 2.7 | Does your research involve animals? | **NO** |
| 2.8 | Does your research involve the administration of drugs, placebos or other substances to study participants? | **NO** |
| **A.3 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee or the Senate Research Ethics Committee (SREC), you must apply for approval from the Computer Science Research Ethics Committee (CSREC) through Research Ethics Online -** [**https://researchmanager.city.ac.uk/**](https://researchmanager.city.ac.uk/)**. Depending on the level of risk associated with your application, it may be referred to the Senate Research Ethics Committee (SREC).** | | *Delete as appropriate* |
| 3.1 | Does your research involve participants who are under the age of 18? | **NO** |
| 3.2 | Does your research involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)?  This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people. | **NO** |
| 3.3 | Are participants recruited because they are staff or students of City, University of London?  For example, students studying on a particular course or module.  If yes, then approval is also required from the Head of Department or Programme Director. | **NO** |
| 3.4 | Does your research involve intentional deception of participants? | **NO** |
| 3.5 | Does your research involve participants taking part without their informed consent? | **NO** |
| 3.5 | Is the risk posed to participants greater than that in normal working life? | **NO** |
| 3.7 | Is the risk posed to you, the researcher(s), greater than that in normal working life? | **NO** |
| **A.4 If you answer YES to the following question and your answers to all other questions in sections A1, A2 and A3 are NO, then your project is deemed to be of MINIMAL RISK.**  **If this is the case, then you can apply for approval through your supervisor under PROPORTIONATE REVIEW. You do so by completing PART B of this form.**  **If you have answered NO to all questions on this form, then your project does not require ethical approval. You should submit and retain this form as evidence of this.** | | *Delete as appropriate* |
| 4 | Does your project involve human participants or their identifiable personal data?  *For example, as interviewees, respondents to a survey or participants in testing.* | **YES** |

**PART B: Ethics Proportionate Review Form**

If you answered YES to question 4 and NO to all other questions in sections A1, A2 and A3 in PART A of this form, then you may use PART B of this form to submit an application for a proportionate ethics review of your project. Your project supervisor has delegated authority to review and approve this application under proportionate review. You must receive final approval from your supervisor in writing before beginning the planned research.

However, if you cannot provide all the required attachments (see B.3) with your project proposal (e.g. because you have not yet written the consent forms, interview schedules etc), the approval from your supervisor will be ***provisional***. You **must** submit the missing items to your supervisor for approval prior to commencing these parts of your project. Once again, you must receive written confirmation from your supervisor that any provisional approval has been superseded by with ***full approval*** of the planned activity as detailed in the full documents. **Failure to follow this procedure and demonstrate that final approval has been achieved may result in you failing the project module and/or result in an academic misconduct investigation.**

Your supervisor may ask you to submit a full ethics application through Research Ethics Online, for instance if they are unable to approve your application, if the level of risks associated with your project change, or if you need an approval letter from the CSREC for an external organisation.

|  |  |  |
| --- | --- | --- |
| **B.1 The following questions must be answered fully.**  **All grey instructions must be removed.** | | *Delete as appropriate* |
| 1.1. | Will you ensure that participants taking part in your project are fully informed about the purpose of the research? | **YES** |
| 1.2 | Will you ensure that participants taking part in your project are fully informed about the procedures affecting them or affecting any information collected about them, including information about how the data will be used, to whom it will be disclosed, and how long it will be kept? | **YES** |
| 1.3 | When people agree to participate in your project, will it be made clear to them that they may withdraw (i.e. not participate) at any time without any penalty? | **YES** |
| 1.4 | Will consent be obtained from the participants in your project?  Consent from participants **MUST** be obtained if you plan to involve them in your project or if you plan to use identifiable personal data from existing records. “Identifiable personal data” means data relating to a living person who might be identifiable if the record includes their name, username, student id, DNA, fingerprint, address, etc. | **YES** |
| 1.5 | Have you made arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential? | **YES** |

|  |  |  |
| --- | --- | --- |
| **B.2 If the answer to the following question (B2) is YES, you must provide details** | | *Delete as appropriate* |
| 2 | Will the research be conducted in the participant’s home or other non-University location? | **NO** |

|  |  |  |  |
| --- | --- | --- | --- |
| **B.3 Attachments**  **ALL of the following documents MUST be provided to supervisors if applicable.**  **All must be considered prior to final approval by supervisors.**  **A written record of final approval must be provided and retained.** | ***YES*** | ***NO*** | ***Not Applicable*** |
| Details on how safety will be assured in any non-University location, including risk assessment if required (see B2) |  | **NO** |  |
| Details of arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential (see B1.5) | **YES** |  |  |
| Full protocol for any workshops or interviews\*\* | **Provisional** |  |  |
| Participant information sheet(s)\*\* | **YES** |  |  |
| Consent form(s)\*\* | **YES** |  |  |
| Questionnaire(s)\*\* | **YES** |  |  |
| Topic guide(s) for interviews and focus groups\*\* |  |  | **N/A** |
| Permission from external organisations or Head of Department\*\* | **Provisional** |  |  |

*\*\*If these items are not available at the time of submitting your project proposal, then* ***provisional approval*** *can still be given, under the condition that you must submit the final versions of all items to your supervisor for approval at a later date.* ***All*** *such items* ***must*** *be seen and approved by your supervisor before the activity for which they are needed begins. Written evidence of* ***final approval*** *of your planned activity must be acquired from your supervisor before you commence.*