**RESEARCH PROPOSAL FORM**

*(also referred to as the ‘Statement of Intent Form’, or SOI)*

***To be submitted by the researcher to the Institute Research Sub-Committee (IRC)***

|  |  |
| --- | --- |
| **Research Title:**  Applying Machine Learning Technique to Forecast Foreign Exchange Market. | |
| **Institute name**  Institute of Information & Communication Technology | |
| **Course / Programme:**  MCAST Bachelor of Science (Honours) in Business Analytics | |
| **Level and year of study**  Level 6 - Year 3 (2022-2023) | |
| **Main area of study being proposed:**  The main area of study proposed is Machine Learning. Machine learning is a branch of artificial intelligence which focuses on the use of data and algorithms to imitate the way that a user can learn and improve accuracy (Education, 2020).  The researcher will concentrate on the return-on-investment ROI of market prices in this study by using supervised learning, also known as supervised machine learning, which is a subclass of artificial intelligence and machine learning. It is distinguished by the use of labelled datasets to train algorithms to properly categorize data or predict outcomes (Education, 2019).  The researcher aims to forecast the forex exchange market with sentimental analysis by creating a bot using Python script and generate social media data. This would help better understand the forex market and provide a more refined experience for future trading. The research method will be quantitative, as the data will be collected from existing forex data to monitor and analyse how the market is changing and developing creating a detailed outlook and unique perspective on how the market is moving over a period. The experiment will make use of a past datasets to analyse the factors and types of markets that can be used for predictions. The experiment will be developed using Python to deploy a bot and use it to scrape data for market insight which can be made to establish different datasets per each market. | |
| **Name of Researcher:**  Ismael Ben Daoud | **Researcher’s I.D. Number:**  0293600L |
| **Signature of Researcher** | **Date of submission of Form**  23/05/2022 |
| **Name of Tutor (or Recommended Tutor):**  Jean Paul Tabone | |

|  |
| --- |
| **Personal Motivation for the Choice of Research Theme.** |
| The motivations of this research are trifold which contributes to the researcher’s personal motivation for the choice of the research theme:  Firstly, the researcher has a great interest in both software and business aspects. As a business analytics student:   * It is a great opportunity to develop a project-based on both software sector and business sector combined. * Using different techniques of analysing the effects of forecasting on the market. * Over the years a huge difference has been seen in the forex market, so this project can show different types of development and present figures of how a market is doing.   Secondly, Machine Learning is an emerging technology within the Artificial Intelligence field. The researcher has an interest into understanding how machine learning techniques work, the requirements of making a good machine learning system, their methods and uses.  Finally, Machine Learning has been applied in many sectors even in Bioinformatics and for various applications like weather predictions. There is few research on the study of Machine Learning when it comes to predicting potential trading but not in a great depth. Therefore, this will allow the researcher to delve in more profundity into this. |

|  |
| --- |
| **Outline of Key Literature and Theoretical Framework or Propositions.** |
| This section will give an overview of the literature that is related to the research topic. This section will include Principles of Machine Learning, Branches of Machine Learning, Advantages and Disadvantages of Machine learning, Deep Learning, Current Implementations and Suggested Solution.   1. **Principles of Machine Learning**   According to (Wilpon et al., 2017), the purpose of machine learning is to create models that mimic and generalize data. These models learn how to differentiate between different items in order to reach the intended result.   1. **Branches of Machine Learning**   Machine learning is classified into three types: supervised, unsupervised, and reinforcement learning. In my situation, I plan to use supervised learning (Anwar, 2021). Machine learning has the benefit of being able to evaluate massive amounts of data and uncover unique trends and patterns by comprehending different sorts of behaviours to assist cater to the proper outcomes. Machine learning has the problem of requiring a large dataset to train on, which might delay the process while waiting for fresh data to be created (Fiair, 2022).   1. **Advantages and Disadvantages of Machine Learning**   Machine learning has the benefit of being able to evaluate massive amounts of data and uncover unique trends and patterns by comprehending different sorts of behaviours to assist cater to the proper outcomes. Machine learning creates predictions while also improving algorithms on its own. It learns to filter new risks as they are identified.  Machine learning algorithms are adept at dealing with multidimensional and diverse data, and they can do so in dynamic or unpredictable contexts. (Fiair, 2022).  Machine learning has the problem of requiring a large dataset to train on, which might delay the process while waiting for fresh data to be created. Machine learning is self-sufficient, although it is prone to mistakes. In the case of machine learning, such mistakes can lead to a chain of errors that can go undiscovered for a long time. When they are recognized, it takes a long time to identify the root of the problem, and much longer to repair it. (Fiair, 2022).    Figure 1: Advantages and Disadvantages of Machine learning  (Fiair, D., 2022)   1. **Deep Learning**   As seen in Figure 2, deep learning employs the neural network design approach, thus the term "deep neural networks." It only has 2-3 hidden layers, but deep networks may have up to 150. Deep learning models are taught utilizing massive amounts of labelled data and neural network topologies that learn features directly from the data, eliminating the requirement for human feature extraction (Deep Learning, n.d.).    Figure 2: Deep learning structure  (Deep Learning, n.d.)   1. **Current Implementations**   Machine learning algorithms are used in a variety of industries, including some of the most prevalent issues. For example, data mining, content screening, and product recommendations are internet-related fields. According to (Varun, 2021), Microsoft stated ambitions to use artificial intelligence into its customer support procedures in 2014. Customers may input inquiries into an online chat and receive automatic replies based on AI-generated scripts, thanks to the company's usage of a chatbot.   1. **Suggested Solution**   The researcher aims to forecast the forex exchange market with sentimental analysis by creating a bot using Python script and generate social media data. This would help better understand the forex market and provide a more refined experience for future trading. The research method will be quantitative, as the data will be collected from existing forex data to monitor and analyse how the market is changing and developing creating a detailed outlook and unique perspective on how the market is moving over a period. The experiment will make use of a past datasets to analyse the factors and types of markets that can be used for predictions. The experiment will be developed using Python to deploy a bot and use it to scrape data for market insight which can be made to establish different datasets per each market. |

|  |
| --- |
| **Significance of the Study.** |
| The key objective of this research is how sentimental analysis maximises profits by implementing machine learning to improve decision making from predictions using past data and current exchange rates. According to (Lasod and Pawar, 2019), many machine learning methods are employed in sentiment analysis for sentiment categorization. More individuals are expressing their feelings and thoughts on the internet and through various social media platforms. This has resulted in a rise in the amount of user created phrases carrying sentiment information, making it hard for humans to read and analyse them all. Thus, automatic analysis of opinions expressed on various web platforms is becoming progressively important for making effective decisions. However, it is also difficult because extracting emotions is a complex activity that requires understanding the content and capturing hidden sentiments in written text, which necessitates the study of new methods (Jemai, Hayouni and Baccar, 2021).  Also, more educated decisions could be made to boost the predicted number, returned by the system. When it comes to the actual implementation, the predictions would also be checked by industry experts in this field, to give feedback on the accuracy on the implemented solution.  Furthermore, this study can be used as a reference by other researchers conducting similar studies, in other words to obtain the relevant and sustainable information about the use of machine learning techniques or it can be also used as a reading material for any person who is interested in this area of study. |

|  |
| --- |
| **Hypotheses and/or Research Question/s** |
| **Hypothesis:** Sentimental analysis maximises profits by implementing machine learning to improve decision making from predictions using past data and current exchange rates.  **Research Questions:**   * RQ 1: Are movements in rates of Forex exchange predictable when taking into consideration past datasets with exchange rates? * RQ 2: Can Machine learning contribute to an optimal decision when dealing with Forex exchange? |

|  |
| --- |
| **Target Participants and Research Methods for Data Collection and Analysis** |
| This section will present a detailed idea about how the proposed study will be conducted. This section will include: Target Participants and Sample size, Data Collection Methods, and Data Analysis.   1. The target participants for this project are employee with related fields and the sample size will be of 20 participants. 2. Data Collection Methods    1. Online Survey   For online survey google forms will be used to evaluate the data sets.   * 1. Experiment   Use the Python programming language to construct a machine learning algorithm to forecast particular foreign exchange markets.   1. Data Analysis    1. Descriptive Statistics   Descriptive data will be used to characterize the data set's properties (Brock and Logan, 2022). The researcher will be using Pie Charts for the demographic section.   * 1. Inferential Statistics   Inferential statistics' purpose is to create generalizations about a population, which may be described as a field of statistics that employs analytical methods to draw inferences about a population from random samples (Inferential Statistics - Definition, Types, Examples, Formulas, n.d.)   1. Chi Squared   The Chi Squared test is a statistical test that compares observed and predicted outcomes. The goal of this test is to identify whether a disparity between actual and predicted data is due to chance or to a link between the variables being analysed.   1. T-Test   A t-test is a statistical test that compares two samples' means. It is used in hypothesis testing, using a null hypothesis that the difference in group means is zero and an alternative hypothesis that the difference is not zero. The difference in group means is calculated by dividing the difference in group means by the pooled standard error of the two group means. In this manner, it computes a number of t-values that illustrate the size of the difference between the two groups being compared and evaluates the likelihood that this difference arises just by chance in p-value.   1. Anova Test   Anova, which stands for analysis of variance, is a statistical test used to compare the means of more than two groups. A one-way anova employs a single independent variable, whereas a two-way anova employs two independent variables.  In order to address the mentioned hypotheses and research questions, the below methods will be adopted:   * An evaluation of existing research on the subject and review of related techniques adopted by different markets will take place to study their implementation and effectiveness determining how different markets are created, understanding how they work and how it can be utilized, how generated content can be implemented. * From the research conducted a combination of various techniques will be analysed and used to develop a prototype of the proposed project. The prototype will focus more on the forecasting of the market. * Will provide a detailed description of the data being used and how it was cleaned and split into training and testing models, additionally, gives a better understanding of the variables present within the research. * A detailed description of the implementation developed will be provided, in addition to the algorithms utilized. Further to this, metrics and accuracy measures are listed. |

|  |
| --- |
| **Anticipated Contributions of the Study.** |
| This research can contribute to investigating the potential of various machine learning techniques within forex exchange and to detect whether exchange rates may be used to forecast stock movements. An additional contribute is to learn whether machine learning could potentially contribute to effective decision making within a forex exchange.  Moreover, this study will contribute by:   1. Introducing a solution to business/ individuals to have better insights about forex exchange. 2. Perceiving benefits and drawbacks of machine learning within the exchange market. |

|  |
| --- |
| **Dissertation Project Plan.** |
|  |

|  |
| --- |
| **Ethical Considerations.**  **Refer to *guidance points below. You are also additionally required to read MCAST Document 074 ‘Research Ethics Policy and Procedure’ that is available on the College website***  *Research shall be conducted in such a manner so as to avoid any psychological and physical harm to humans and animals and financial damage to organizations:*   1. *Only the supervisor and examiners will have access to any data gathered.* 2. *Participants will remain free to withdraw from the study at any time without having to provide any reason. In the case of withdrawal, all the records and information collection will be deleted.* 3. *The participant, who is the sole proprietor of the data provided, is granting that such data would be processed for this study purposes only.* 4. *The data collection process will be a transparent process.* 5. *All transcriptions and/or electronic recordings reflecting the data collected, once exhausted, are to be deleted* 6. *Confidentiality, anonymity and data protection procedures are to be ethically abided by.* 7. *The researcher would provide a soft copy of the study to the participant, if required.* |
| *Enter details here regarding possibility of issues regarding confidential personal data:*  The researcher will ensure that no personal data or confidential data is divulged in this study. The data gathered from the study will be stored on the researcher’s computer and will be accessible only to the researcher.  The researcher will ensure that the participants will be kept anonymous as names and addresses are not necessary for the study therefore personal data will not be asked and stored.  Data will not be kept longer than necessary. The maximum amount of time it will be kept is for one year, and afterwards it will be shredded.  The researcher will emphasize that the data will be used solely for the purpose of this study only, and that the identity of the participants will be kept anonymous*.* |
| *Enter details here regarding possibility of physical harm:*  The researcher will ensure that no physical harm will occur to the participants during the study. The researcher will provide the participants with materials that will be used for the study and are safe and have no way in harming the participants physically.  The environment in which the target participants will take part in the study will be safe and will not endanger the participants physically. |
| *Enter details here regarding possibility of moral harm:*  The researcher will ensure that the participants will not suffer from any moral harm. The researcher will provide a clear and understandable explanation to the participant about the conducted study, by using simple words that the participants are knowledgeable of and highlighting their part in the study.  The researcher will ask participants if they have any questions regarding the study and their part in it.  The researcher will make sure to answer all and any questions related to the conducted study.  The researcher will ensure that he asks simple and straight to the point questions to the participants so that they will understand better what is being asked from them.  The researcher will take into consideration the sensitivity and privacy of the participants. What will be gathered from the participants will only be used for the sole purpose of the study and what happens in the session will remain private.  In the case of disabled persons and other social groups, if they will be within the target population, they would still be taking part in the study, and they will receive special care.  The participants are provided with the right to withdraw from participating in this study, without the need to provide any reasons to the researcher. |
| *Enter details here regarding possibility of business harm:*  The researcher will ensure that there will be no business harm as a result of the conductive study.  The researcher will provide any businesses who may have a query about business harm with a soft copy of the study.  The researcher will make sure that only relevant data will be gathered. The data will remain private and used only for the sole purpose for feedback purposes and improvement of the solution.  The researcher will make sure to abide by the new legislation (GDPR) regarding business and data protection. |

|  |
| --- |
| *Please see* ***Annex 1*** *for a sample Participant Information Letter and* ***Annex 2*** *for a sample Participant Consent Form. Student is to submit a copy of the proposed Participant Information Letter and Participant Consent Form where applicable. Both documents should be attached to the end of the SOI that is being submitted by the student.* |

**List of Key References:**

|  |
| --- |
| References  1. Anwar, T., 2021. *Different Branches of Machine Learning | Artificial Intelligence Part 3/4 (Episode 5 | CVFE) | Bleed AI*. [online] Bleed AI. Available at: <https://bleedai.com/different-branches-of-machine-learning-artificial-intelligence-part-3-4-episode-5-cvfe/#:~:text=Machine%20Learning%20can%20be%20further,Unsupervised%20Learning%2C%20and%20Reinforcement%20Learning.> [Accessed 18 May 2022]. 2. Brock, T. and Logan, M., 2022. *Descriptive Statistics Definition*. [online] Investopedia. Available at: <https://www.investopedia.com/terms/d/descriptive\_statistics.asp> [Accessed 23 May 2022]. 3. Cuemath. n.d. *Inferential Statistics - Definition, Types, Examples, Formulas*. [online] Available at: <https://www.cuemath.com/data/inferential-statistics/> [Accessed 23 May 2022]. 4. Education, I., 2019. *What is Supervised Learning?*. [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/supervised-learning#:~:text=Supervised%20learning%2C%20also%20known%20as,data%20or%20predict%20outcomes%20accurately.> [Accessed 20 May 2022]. 5. Education, I., 2020. *Machine Learning*. [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/machine-learning> [Accessed 20 May 2022]. 6. Fiair, D., 2022. [online] Available at: <https://data-flair.training/blogs/advantages-and-disadvantages-of-machine-learning/> [Accessed 19 May 2022]. 7. Jemai, F., Hayouni, M. and Baccar, S., 2021. *Sentiment Analysis Using Machine Learning Algorithms*. [ebook] Available at: <https://www.researchgate.net/publication/353807644\_Sentiment\_Analysis\_Using\_Machine\_Learning\_Algorithms> [Accessed 21 May 2022]. 8. Lasod, A. and Pawar, R., 2019. *Sentiment Analysis Using Machine Learning Techniques*. [ebook] Available at: <https://www.academia.edu/41359036/Sentiment\_Analysis\_Using\_Machine\_Learning\_Techniques> [Accessed 21 May 2022]. 9. n.d. *Deep Learning*. [online] Available at: <https://www.mathworks.com/discovery/deep-learning.html> [Accessed 20 May 2022]. 10. n.d. *Machine Learning solutions*. [online] Available at: <https://www.arm.com/glossary/machine-learning-solutions> [Accessed 21 May 2022]. 11. Raosoft.com. n.d. *Sample Size Calculator by Raosoft, Inc.*. [online] Available at: <http://www.raosoft.com/samplesize.html> [Accessed 22 May 2022]. 12. Varun, B., 2021. *7 Real-World Examples of Machine Learning in Current Times - DataScienceCentral.com*. [online] Data Science Central. Available at: <https://www.datasciencecentral.com/7-real-world-examples-of-machine-learning-in-current-times/> [Accessed 19 May 2022]. 13. Wilpon, J., Thomson, D., Bangalore, S., Haffner, P. and Johnston, M., 2017. *THE FUNDAMENTALS OF MACHINE LEARNING*. [ebook] Available at: <https://www.interactions.com/wp-content/uploads/2017/06/machine\_learning\_wp-5.pdf> [Accessed 21 May 2022]. |

***This section is to be filled in by the representative of the Institute Research Sub-Committee (IRC) prior to forwarding of this Form to the ‘MCAST Research Ethics Committee’ for final ethics approval:***

|  |  |  |
| --- | --- | --- |
| ***Nature of Ethical Consideration*** | ***Outcome (Tick)*** | ***Comments/Advice*** |
| All ethical issues have been adequately tackled. |  |  |
| Possibility of issues regarding misuse of data or some form of harm. |  |  |

|  |  |
| --- | --- |
| **Details of Representative to the Institute Research Sub-Committee.** | |
| Name | Signature |
| Designation | Date |

**Annex 1: Participant Information Letter**

***Sample:*** Graphical user interface

Description automatically generated with medium confidence

**Title of Research: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

You are being invited to take part in a research study. Before you decide to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

**What is the purpose of the study?**

This research is being undertaken on…

**Why have I been chosen?**

You have been chosen because…

**Do I have to take part?**

It is up to you to decide whether or not your take part. If you decide to take part you will be given this information sheet to keep and be asked to sign a corresponding consent form.

**What will happen to me if I take part?**

You will then be given a questionnaire on.../your data will be used…/your image will be used…

**What are the possible disadvantages and risks of taking part?**

There are no disadvantages or risks foreseen in taking part in the study.

**What are the possible benefits of taking part?**

By taking part you will be contributing to the development of a set of recommendations for…

**What if something goes wrong?**

If you wish to complain or have any concerns about any aspect of the way in which you have been approached or treated during the course of this study, please contact…(researcher is to give his/her MCAST email as a contact)

**Will my details be kept confidential?**

All information which is collected about you during the course of the research will be kept strictly confidential so that only the researcher carrying out the research will have access to such information and will not be shared with any other individuals. Participants should note that data/images collected from this project may be retained and published in an anonymized form. By agreeing to participate in this project, you are consenting to the retention and publication of data.

**What will happen to the results of the research study?**

The results will be written up into a dissertation for my final project of my Bachelor…

**Who is organizing the research?**

The research is conducted as part of a degree in …

**Who may I contact for further information?**

If you would like more information about the research before you decide…(researcher is to give his/her MCAST email as a contact)

*Thank you for your interest in this research…*

**Annex 2: Participant (or Guardian) Consent Form**

***Sample:*** Graphical user interface

Description automatically generated with medium confidence

**Title of Research: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Name of Researcher: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# Please initial box

1. I confirm that I have read and understand the Information Letter

for the above study and have had the opportunity to ask questions.

1. I understand that my/my charge’s participation is voluntary and that I/my charge am/are free to withdraw at any time without giving any reason.

3. I agree to allow my daughter/son/charge to take part in the above study.

*(Statement 3 is to be included only when guardians/parents are involved in giving consent)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

Name of Participant/

Guardian Date Signature

Researcher Date Signature

*1 for participant; 1 for researcher*