MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	Examining Feature engineering techniques for time series prediction in forex markets
Student ID:	22265978, 23270600
Student name:	Shreyash Ramtekkar, Pooja Ashok
Student email	shreyash.ramtekkar2@mail.dcu.ie pooja.ashok2@mail.dcu.ie
Chosen major:	Data Analytics
Supervisor	Andrew Mccarren
Date of Submission	

Section 2: About your Practicum

What is the topic of your proposed practicum? (100 words)

The United States Dollar (USD) stands as a linchpin in the intricate web of the foreign exchange (FX) market, renowned as one of the largest financial markets globally. With trillions in daily transactions, the USD's role is pivotal in shaping international trade, investment, and risk management. The complex task of predicting exchange rate movements involves navigating through a myriad of economic, political, and psychological factors. Fundamental indicators, such as GDP, inflation, interest rates, retail sales, and foreign reserves, exert significant influence on the value of the US dollar. Additionally, technical factors, employing chart and price analysis tools like moving averages, RSI, and Bollinger bands, contribute to the intricate dance of currency prices. Within this multifaceted landscape, exploring methodologies becomes imperative to discern and comprehend the various forces shaping the trajectory of the US dollar in global currency markets.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment (https://www.sciencedirect.com/science/article/pii/S0957417414004801#b0135)

2. Short-term stock market price trend prediction using a comprehensive deep learning system (https://link-springer-com.dcu.idm.oclc.org/article/10.1186/s40537-020-00333-6)

3. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators (https://link-springer-com.dcu.idm.oclc.org/article/10.1186/s40854-020-00220-2)

4. Forecasting foreign exchange markets: further evidence using machine learning models(https://link-springer-com.dcu.idm.oclc.org/article/10.1007/s00500-021-05830-1)

5.Algorithmic Forex Trading using Combination of Numeric Time Series and News Analysis(https://ieeexplore.ieee.org/document/9058285)

How does your proposal relate to existing work on this topic described in these papers? (200 words)

This proposal aligns with existing research on predicting foreign exchange (forex) market movements by emphasizing the significance of sentiment and semantics analysis, feature extraction, and the application of machine learning models. The literature review highlights the integration of algorithms for sentiment and semantic analysis from forex fundamental news headlines, showcasing the relevance of text mining in forecasting prices. It focuses on feature extraction and reduction corresponding with efforts in the literature to employ techniques such as Heuristic-Hypernym, Synchronous Targeted Feature-Reduction, and Sum Score-Weighting for effective feature selection. Both our proposal and the literature explore the integration of machine learning models, including Recursive Feature Elimination, Principal Component Analysis, and Long Short-Term Memory models, emphasizing the importance of selecting appropriate features to enhance predictive accuracy.

Furthermore, our proposal shares common ground with the literature in advocating for the combination of fundamental and technical analyses, as seen in the discussion on creating hybrid models like the Macroeconomic and Technical LSTM model. The recognition of the need to evaluate model performance using metrics such as Mean Absolute Percentage Error and Root-Mean-Square Error is consistent with the literature's emphasis on rigorous assessment. Additionally, the literature's insight into the variability of model performance across different currencies resonates with your proposal's acknowledgement of the importance of choosing a suitable model contextually. Overall, our proposal builds upon existing methodologies and suggests further exploration of feature engineering techniques, model selection, and context-specific strategies to enhance the accuracy of forex market predictions.

What are the research questions that you will attempt to answer? (200 words)

The research questions that can be inferred from the given paper are:

1. What feature engineering techniques significantly impact changes in market prices within the forex market?

2. Is it possible to accurately predict forex rates by identifying the key features affecting prices?

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

The paper poses several pertinent research inquiries:

1. Exploring Feature Engineering Techniques: Uncovering the primary feature engineering strategies that exert influence on market price variations within the forex market.

2. Role of Sentiment and Semantics Analysis: Evaluating the significance of sentiment and semantics analysis in the prediction of forex market movements and investigating effective methods for integrating these analyses into the feature engineering process.

3. Contribution of Machine Learning Models: Investigating the specific contributions of machine learning models—namely Recursive Feature Elimination, Principal Component Analysis, and Long Short-Term Memory models—towards augmenting predictive accuracy within the realm of forex market predictions.

4. Efficiency of Hybrid Models: Assessing the effectiveness of hybrid models, such as the Macroeconomic and Technical LSTM model, and understanding how they harmonize fundamental and technical analyses to enhance the precision of forex market predictions.

5. Model Performance Across Currencies: Analyzing how the selection of a predictive model influences performance variations across diverse currencies, with a focus on identifying contextual factors pivotal for model suitability in forex market predictions.

What software and programming environment will you use?

Jupyter Notebooks offer interactive computing with support for multiple languages, seamless integration with data visualisation libraries, and widgets for enhanced interactivity. They enable code modularity and easy sharing through various export formats and are widely used in collaborative environments.

What coding/development will you do?

Python for data analysis provides a versatile and powerful ecosystem with libraries like Pandas, NumPy, and Matplotlib for efficient data manipulation, research, and visualisation. Its simplicity and readability make it a preferred language for data scientists.

What data will be used for your investigations?

We will use GDELT dataset and the currency data for different country currencies, which will be fetched from yfinance. These data will contribute to performing the feature engineering and analysis

Is this data currently available, if not, where will it come from?

We are going to fetch the data for extracting the feature from gdealt. However, the currency price movement dataset is unavailable for that specific period. So, we will use the Yahoo Finance Library to get the currency data.

Currency Prices Dataset: yfinance

Gdelt Dataset: http://data.gdeltproject.org/events/index.html

What experiments do you expect to run?

Conduct experiments to assess the impact of diverse feature engineering techniques on market price changes in the forex market. Implement and compare methods like Heuristic-Hypernym, Synchronous Targeted Feature-Reduction, and Sum Score-Weighting for efficacy. Explore the contributions of machine learning models, such as Recursive Feature Elimination, Principal Component Analysis, and Long Short-Term Memory models, in enhancing predictive accuracy. Run comparative experiments to evaluate individual and combined model performances. Assess the hybrid model against fundamental and technical analysis models. Experiment to understand how predictive model choices affect performance across diverse currencies. Identify and evaluate contextual factors influencing the suitability of predictive models for specific currencies.

What output do you expect to gather?

The anticipated outputs include detailed insights into influential feature engineering techniques, quantification of sentiment and semantic analysis impact, performance metrics for Recursive Feature Elimination, Principal Component Analysis, and Long Short-Term

Memory models. Evaluation of the Macroeconomic and Technical LSTM hybrid model's effectiveness in combining fundamental and technical analyses, along with assessments of model performance variations across different currencies. The aim is to derive comprehensive conclusions on feature engineering, sentiment analysis, and model selection strategies, refining approaches for more accurate forex market predictions.

How will the results be evaluated?

The results will be evaluated based on the effectiveness of feature engineering techniques in capturing market price changes, and the significance of predicting accuracy of machine learning models. The evaluation will assess the success of the Macroeconomic and Technical LSTM hybrid model in combining fundamental and technical analyses. Contextual variations in model performance across diverse currencies will be considered, providing a comprehensive understanding of the impact of feature engineering and model selection on forex market predictions.