SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCSD 6215
Sem:…01……. Session:……20242025……

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCSD 6215)**
Student Name:	RAIAN HAFIZ NILOY
Metric Number:	MCS241008
Student Email & Phone:	raian@graduate.utm.my ; +60182802938
Project Title:	Predicting Stock Market Trends: A Novel Approach Using Multi-Source Sentiment Analysis And advanced deep learning algorithms
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

Stock price prediction has traditionally been reliant on various statistical models without considering market sentiment and consequently, less accurate. To include public sentiment in stock price prediction, NLP has been used to analyze Textual data and yield insights. However, integrating sentiment analysis comes with the challenge of choosing between traditional ML models and advanced DL models as one offers computational efficiency and the other one is better at understanding complex data patterns. To solve this issue, this project uses a hybrid approach where it combines advanced NLP models such as FinBERT, GPT-4 with advanced DL algorithms like LSTM. The NLP models ensure sentiment classification from textual data with more precision & LSTM networks will be used for time series prediction.

Problem Background:

The financial market is complex and volatile per se; thus, predicting the movement of the market is very sensitive. .

Classically, statistical methods like time series analysis or regression models have been used for forecasting stock prices. However, prediction purposes cannot be fully objective as it does not consider the ever-changing market sentiments. In recent years, several Natural Language Processing (NLP) models have been imported to extract insights from sentiment data and this has improved the accuracy of such predictions. However, issues with the application of sentiment analysis continue to exist; that is, it leads to a choice between existing classical methods of Machine Learning (ML) or the newer Deep Learning (DL) algorithms. Classic ML models such as logistic regression are very simple and computationally efficient but fail miserably at unraveling the ambiguous nature of data patterns against those of the DL models

with their high-end computational power.

This project thus intends toward solving this problem utilizing a hybrid model that reaches the best of both worlds-collects its data from diverse sources to arrive at a complete view of market sentiment, analyzes market using advanced deep learning architectures like FinBERT and GPT-4 while using Long Short Term Memory (LSTM) network for time-series prediction.

Problem Statement:

The central issue that this project tackles goes in consonance to the pretas of the choice between traditional Machine Learning models and advanced Deep Learning algorithms for the purpose of doing sentiment analysis for predicting movement in the stock market. Recent research indicates that traditional ML algorithms such as logistic regression outperform leading-edge DL algorithms like FinBERT or GPT-4, but as far as processing more complex data is concerned, these DL algorithms Show a greater promise. In spite of being resource demanding & performing moderately, their analysis is more Sophisticated. That's why this project takes a hybrid approach to combine Advanced DL algorithms like FinBERT, GPT-4 with a specific recurrent neural network called LSTM. Also, this project gives emphasis on collecting sentiment data from multiple sources to get a holistic view of the market sentiment. As a result, this project aims to ensure more accuracy in predicting stock price.

Aim of the Project:

- To collect data from multiple sources and preprocess them
- To combine state-of-the-art NLP models and DL algorithms and create a hybrid model
- To increase the accuracy of stock price prediction

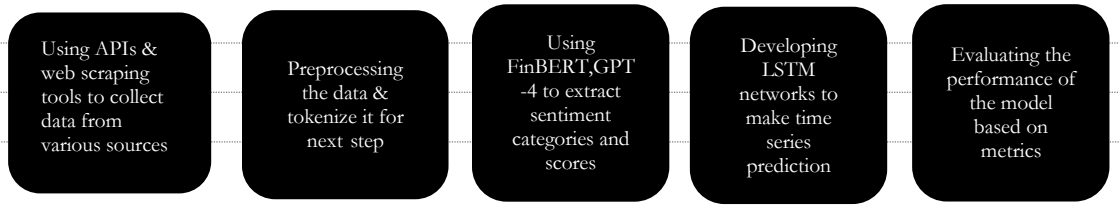
Objectives of the Project:

- To enhance the comprehensiveness of prediction by collecting and preprocessing sentiment data from multiple sources
- To implement FinBERT & GPT-4 for detailed sentiment analysis and develop LSTM networks for improving prediction accuracy
- To rigorously assess the model's performance with a number of evaluation metrics such as Accuracy, Precision, Recall, F1 score etc

Scopes of the Project:

This project combines sentiment analysis of financial news and social media posts and performs time series forecasting to make effective stock price predictions. This covers data collection preprocessing and model development using LSTM networks and the evaluation of predictive performances. However, the research does not include real-time data analysis. It doesn't perform a full-fledged market analysis and doesn't recommend any particular stock.

Methodology:



Expected Contribution of the Project:

- Introduction of a novel approach combining advanced NLP models and LSTM networks
- Improvement of stock price prediction accuracy
- Providing valuable insights to the investors and mitigate the risk factor
- Developing a robust framework in financial analysis sector

Project Requirements:

Software: Pyhton, Jupyter notebook, Libraries: NLP(NLTK, Spacy, Transformers), DL(TensorFlow, Keras), data collection(beautifulsoup, requests), data analysis (Pandas, numpy), data visualization(matplotlib, seaborn) .

Hardware: High-performance computing system with sufficient storage and processing power.

Technology/Technique/ Methodology/Algorithm: Natural Language Processing(NLP), Deep Learning(DL)

Type of Project (Focusing on Data Science):

- ✓ [ ] Data Preparation and Modeling
- ✓ [ ] Data Analysis and Visualization
- [ ] Business Intelligence and Analytics
- ✓ [ ] Machine Learning and Prediction
- [ ] Data Science Application in Business Domain

Status of Project:

- ✓ [ ] New
- [ ] Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

✓ [ ] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: Raian Hafiz Niloy

Raian Hafiz Niloy — **Signature**

15/11/2024 — **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1: .....

..... — **Signature**

..... — **Date**

Name of Supervisor 2 (if any): .....

..... — **Signature**

..... — **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL

[ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor)

[ ] FAIL*

***** Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

Signature **Date**

Name of Evaluator 2:

Signature **Date**