

RaianHafizNiloy_ProjectPropo sal.docx

by Raian Hafiz Niloy MCS241008

Submission date: 29-Nov-2024 04:27AM (UTC-0800)

Submission ID: 2534544560

File name: RaianHafizNiloy_ProjectProposal.docx (72.5K)

Word count: 948

Character count: 5632



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

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 inherently complex and volatile. So, predicting market movement has always been very critical. Traditionally statistical methods like time series analysis or regression models have been used to forecast stock prices. But without considering the ever-fluctuating market sentiment, accurate prediction is not possible. To gain insights from sentiment data, various Natural Language Processing (NLP) models have been used and it has improved accuracy of prediction. However, the difficulties to implement sentiment analysis do not end here as it makes us to choose from traditional Machine Learning (ML) models and advanced Deep Learning (DL) algorithms. While traditional ML models Such as logistic regression are simple and computationally efficient, they often fail to capture the intricate data patterns. Conversely the DL models are good at handling such complex data, they still fail to outperform the ML models.

This project attempts to tackle this challenge by using a hybrid approach which offers the best of both worlds. It collects data from various sources to get a holistic view of market sentiment, uses advanced deep learning architectures like FinBERT, GPT-4 to analyze the sentiment & Long Short Term Memory(LSTM) networks for time series prediction.

Problem Statement:

The primary problem this project addresses, originates from the dilemma of choosing between traditional Machine Learning models and advanced Deep Learning algorithms to integrate sentiment analysis for predicting stock market movements. A recently released research work shows that traditional ML algorithms like logistic regression outperform the advanced DL algorithms like FinBERT or GPT-4 but while it comes to handling complex data, the DL algorithms Show stronger potential. In spite of being resource demanding & performing moderately, their analysis was 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

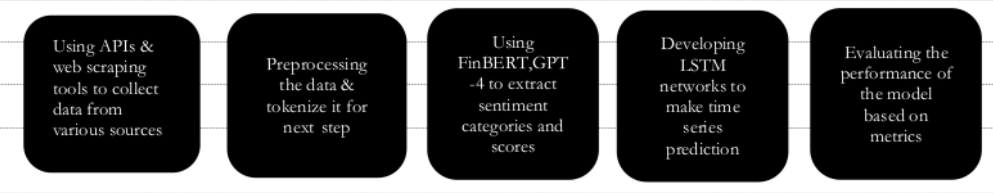
Objectives of the Project:

- 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:

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

Software: _____

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 Niloy15/11/2024

SignatureDate

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.

3

Name of Supervisor 1:

SignatureDate

Name of Supervisor 2 (if any):

SignatureDate

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:

RaianHafizNiloy_ProjectProposal.docx

ORIGINALITY REPORT

7%

SIMILARITY INDEX

7%

INTERNET SOURCES

0%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

[dokumen.pub](#)

Internet Source

2%

2

[ijsrst.com](#)

Internet Source

1%

3

[www.coursehero.com](#)

Internet Source

1%

4

[www.scfhs.org.sa](#)

Internet Source

1%

5

[arxiv.org](#)

Internet Source

1%

6

[github.com](#)

Internet Source

1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On