THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Large Language Model (LLM) Informed Graph Neural Network for Industrial Commodities Trend Pattern Prediction

Tse-Wei (Jefferson) Lee

Student Number: 43150498

Student Email: jefferson.lee.project@outlook.com

Supervisor: Professor Xue Li

School of Information Technology and Electrical Engineering,

The University of Queensland.

Submitted for Master of Data Science Capstone 1

CONTENTS

INTRODUCTION

In this capstone project, I propose a novel approach that leverages the power of the Large Language Model (LLM) (Kerner, 2023) and Graph Neural Network (GNN) (Dagli, 2022) to achieve accurate forecasts on commodities price trends. The importance of commodity price forecast has risen significantly due to inflation and high price volatility. Inflation and high volatility are caused by regional instabilities and supply-demand in economics. In the modern economy, crude oil and wheat are the most traded commodities (Rivero, 2023), so I will predict the price of crude oil and wheat, which have had unstable price in recent years.

In response to the unpredictability, institutions and governments have strived to enhance the accuracy of commodity price forecasts. To enhance the accuracy, analysts have applied several machine learning (ML) techniques, including Graph Neural Networks (GNN) and Recurrent Neural Networks

(RNN) (Zhu, 2020). Although each ML technique has its pros and cons, one of the general disadvantages is ML method cannot automatically evaluate the impact of the news media or changing geopolitical events. Therefore, I propose to use the latest developed LLM to interpret impacts.

I propose that the LLM with ML algorithm is the solution for more accurate prediction, so I chose the topic for this capstone project. This capstone project covers several objectives. Major objectives include fine-tuning the LLM with financial information, integrating LLM output into the ML algorithm, and predicting correct long-term, mid-term, and short-term price trends. I can achieve these objectives based on the proposed solutions.

The solution starts with using LLM. LLM has had a breakthrough in recent years, particularly since the release of ChatGPT. ChatGPT 4.0 is the best LLM available at the time I am writing this proposal, mid-2023, and ChatGPT 4.0 is made and trained by a company called OpenAI (Ephrem , 2023). OpenAI provides ChatGPT's LLM for public access through OpenAI's Application Programming Interface (API) at a cost. It is called OpenAI API.

Therefore, I will fine-tune the OpenAI API with global news. After fine-tuning, the next step is to feed it into GNN. GNN will receive output from OpenAI API and perform ML prediction on the commodity price trends. I expect the prediction outcome from LLM informed GNN (LLM-GNN) can perform a realistic price trend because LLM informs more realistic situation judgement to GNN. GNN outcome will be compared to the price ground truth and test under criteria. The comparison will provide insights into the feasibility and viability of this capstone project. I will ensure this capstone project is realistic and is built on the proposal outlines.

The proposal outline includes the introduction, current situations and methods, problems with the present system, objectives, proposed solutions, technical methods, operation methods, outcomes, and conclusions, so I will discuss the current situation and methods first.

CURRENT SITUATION & METHODS

To identify the need for using LLM-GNN to predict commodity prices, I have analysed the current industrial situation and methods, and have classified them into three parts, global situation, commodity situation, and existing prediction method. Firstly, I propose that the global situation is affecting the instability of commodities. The instability of commodities is the new normal that we are experiencing. The new normal is characterized by increased volatility and unpredictability, whether it is in the realm of commodity prices or food prices (Nelson et al., 2023). For example, the price of crude oil is based on the supply and demand in the market, and the Ukraine-Russian war and post-COVID China reopening can make a huge impact on price. To forecast the possible prices, institutions like JPMorgan, and Goldman Sachs constantly adjust their price predictions based on the global situation. (Albano, 2022) Furthermore, the Federal government of Australia also needs to forecast the commodity price each year for budget purposes. (John, 2022) All these examples show the demand for accurately forecasting commodity prices.

Secondly, I propose to analyse crude oil and wheat commodities. Commodities include raw materials and agricultural products, such as cotton, copper, natural gas, and crude oil, that can be traded in commodity markets (WTI or Brent for crude oil). The trading price is directly related to the global economy because raw materials and agriculture are key ingredients to produce goods and products. Goods and products affect society and inflation. While fighting inflation is one of the most important targets for the Federal Reserve Bank (Fed), studying commodity prices is an essential part for analysts. (Financial, 2023) Analysts prefer some wildly used and traded commodities like crude oil and wheat, so I selected crude oil and wheat. While crude oil and wheat have increasing volatility and unpredictability due to regional instabilities and demand changes, how to create accurate forecasting is getting more important.

Thirdly, to enhance the prediction accuracy, my research shows that there are several applied machine learning (ML) techniques, including Graph Neural Networks (GNN), Support Vector Machines (SVM), Graph Convolutional Networks (GCN) and Recurrent Neural Networks (RNN) (Zhu, 2020). Each technique has its advantages and disadvantages, but I have chosen GNN as the primary method. GNN is one of the artificial neural networks (ANN) that is capable of processing complex data represented as graphs. The graph is a type of data structure that contains nodes and edges, so I have an educated guess that a graph network is the best method to perform regional relationships. These are three parts of the current situation and methods, but now I will explore the current problem and research gap.

PROBLEMS WITH THE PRESENT SYSTEM & RESEARCH GAP

I propose the problem and research gap is the GNN method cannot automatically evaluate the impact of geopolitical incidents. Not only GNN method cannot provide agile incident evaluation, but the general ML algorithm also cannot automatically evaluate the effect of the latest geopolitical news. The analyst needs to manually interpret impacts into the algorithm. Therefore, I propose to use the latest developed LLM to attend to this research gap.

The LLM is a Deep Learning (DL) technology that's used to perform natural language processing (NLP) LLM is trained by massive datasets, enabling them to have the ability to interpret text contents. The interpretation can resolve the ML shortage by evaluating the impact of geopolitical incidents. To resolve the shortage, there are several milestones to reach.

OBJECTIVES

To validate whether LLM-GNN can predict commodity price trends is an attainable research problem, I have listed several deliverable milestones to ensure the project is successful. The project objectives include three major parts, technical objectives, business objectives and evaluation criteria objectives.

Technical objectives:

- Fine-tuned Large Language Model (LLM) will classify the news article correctly.
- The classification result will apply and merge with the Graph Neural Network (GNN).
- GNN will train and predict reasonable prices with price datasets and LLM classification results.

Business objectives:

- LLM-GNN approach will use the reasonable business dataset.
- The trend prediction will be similar compared to the ground truth (actual commodity price).

Evaluation criteria objectives:

- The LLM-GNN approach will provide a generalized output on different commodities.
- LLM-GNN model prediction can provide reasonable statistical accuracy under different methods:
    - Mean Squared Error (MSE)
    - Cross-validation (CV)
    - R-squared Score (R2)
    - Visualization

My goal is to deliver these objectives to the stakeholders, the goal is to correctly predict the price trend using LLM-GNN method. To reach this goal, I propose the following solutions.

PROPOSED SOLUTIONS

To reach the goal of predicting price trends for commodities using LLM and GNN, I propose the following solutions for each objective, and there are three parts, resource collection, method implementation, and outcome evaluation. Firstly, resource collection includes computation power and datasets. I will collect the required datasets, including all the required historical index prices, which Brent crude oil, WTI crude oil, wheat and S&P 500 (Brower, 2023). I will also collect crude oil-related stock prices, such as ExxonMobil or Aramco. The stock, index, and commodity prices interact with each other, so it is beneficial for prediction. However, the limitation is we can only obtain daily prices instead of hourly prices, so I can't complete a model based on hourly prediction with the LLM-GNN method.

To compile an efficient GNN model training, I will register for the cloud computing platform for computation power. I will pay to access OpenAI's LLM resources and fine-tune them with financial news collected from the Financial Times (Brower, 2023). After tuning in with Financial Times, I have completed the first resource collection part.

Secondly, the second method implementation part includes GNN and LLM interaction. LLM from OpenAI is fine-tuned and provides classification results in measurable impact scales. I will feed the measurable impact scales into the GNN ML model and tune the GNN method parameters to perform the best results. Through trial-and-error implementation, I will refine the outcome of the LLM-GNN method.

Thirdly, the outcome evaluation includes comparing with ground truths, generalisation, and data visualisation. I will apply data visualisation technology to ensure all the stakeholders can understand the outcome clearly. The outcome will be compared to the ground truths including long-term, mid-term, and short-term price time frames. The comparison will have to pass the statistical criteria. After I identify a stable LLM-GNN model, I will test the method on the other commodities to check the generalization ability.

These are the proposed solutions that I will deliver in this capstone project. The capstone project will be performed based on proposed solutions and detailed methods. I will explain the detailed method in the next section.

METHOD: QUALIFICATIONS, COST AND SCHEDULE

To deliver a viable price trend prediction with a large language model (LLM) and Graph Neural Network, I have a plan of steps, datasets, and a workable schedule, and I will explain each of them in this method section. I will start with a workable schedule, including the qualification, cost, and project schedule.

Qualification

Firstly, qualification evaluates the ethical and legal concerns of analysing the data. All the individuals including myself does not have legal concern. I have checked that there is no legal or sensitive information on the dataset, so there is no legal or application fee involved.

Cost

Secondly, the cost of this capstone project mainly comes from our time cost and technical costs. The technical cost includes access cost to OpenAI and computing power cost. Furthermore, Professor Xue Li's time and my time are the most valuable costs for this project. I will spend 10 hours each week on the project, and we will have weekly meetings and discuss the project. The time cost and technical costs are funded by the University of Queensland (UQ), so we will follow the academic calendar for our project schedule.

Project schedule

Thirdly, I have generated a Gantt chart based on the schedule, and the Gantt chart covers the expected outcome and timeline. Although this timeline only shows the expected situation, the basic framework is covered in the Gantt chart presented in the reference section. The next section is going to cover the dataset.

METHOD: DATASET

To predict the trend price for commodities with LLM-GNN, there are two major parts of the dataset, the price dataset and the news dataset. The price dataset covers the related index, commodity, and stock prices. The indexes are from the countries that affect the price of oil, including Saudi Arabia, the US, China, and Russia (Bajpai, 2022). Stock data are from the most influential petroleum companies, and these companies and countries are shown below. We also collected a similar dataset from the most influential countries and companies (Harvey, 2022)on wheat prices.

Price Dataset Profile

Item	Type	Field	Note
WTI crude oil	Commodity	Oil	West Texas Intermediate price
Brent crude oil	Commodity	Oil	Brent crude price
Wheat future	Commodity	Oil	Wheat future trading price
S&P500 US	Index	Oil, Wheat	National stock index in the US
MOEX Russia	Index	Oil, Wheat	National stock index in Russia
China A50	Index	Oil, Wheat	National stock index in China
TASI Saudi Arabia	Index	Oil	National Stock Index in Saudi Arabia
BP PLC	Stock	Oil	Multinational oil and gas company
TotalEnergies SE	Stock	Oil	Multinational petroleum company
Shell	Stock	Oil	Multinational oil and gas company
ExxonMobil	Stock	Oil	Multinational petroleum company
Sinopec	Stock	Oil	China Petroleum & Chemical Corp.
PetroChina	Stock	Oil	Chinese oil and gas company
Saudi Aramco	Stock	Oil	Saudi Arabian Oil Group
Archer Daniels Midland (ADM)	Stock	Wheat	Multinational grain trading company
Bunge	Stock	Wheat	Multinational grain trading company
Cargill	Stock	Wheat	Multinational grain trading company
Louis Dreyfus	Stock	Wheat	Multinational grain trading company
Table 1. Price dataset (Reiff, 2023)			

Detailed Dataset Profile (Major Dataset)

Item	Type	Field	Note
WTI crude oil	Commodity	Oil	West Texas Intermediate price
Date Range:	From: 1983-March-03 To: 2023-August-31		
Tuples:	10274 tuples before data cleaning Non-IID (Non-Independently and Identically Distributed) (Khanna, 2021)		
Attributes:	Date: Date in the format of DD/MM/YYYY Price: Daily closing price in US dollars Open: Daily opening price High: Daily highest price Low: Daily lowest price Change %: Percentage change compared to the previous day.		
Table 2. Detailed dataset profile (Ferlito, 2023)			

I have downloaded all the datasets from investment.com, so the owner of the data is investment.com, but the datasets are publicly accessible information. Each of the historical prices has its time frame. I will focus on daily data as a major time frame because the daily historical price is the best available commodity data. I have identified daily historical price as a limitation because the news dataset is available in a much smaller time frame.

News Dataset Profile

I have downloaded the news dataset from Factiva by Dow Jones (Newswire, 2023). It is an aggregated news platform providing business information and research tools for organizations. The University of Queensland provides access to Factiva, and Factiva provides news from Financial Times. I will use news from the Financial Times because the Financial Times has a good reputation and a long historical database from the 1980s until now. I will use a historical database and conduct a keyword search, including the following keywords in the table. I will feed the keyword search results into the LLM and GNN following the technical steps in the next section.

News Dataset Profile for commodities

<table>
<tr><td>Key words</td><td>Tuples (No. articles)</td><td>Attribute</td></tr>
<tr><td>Crude oil production</td><td>1288</td><td rowspan="7">• Article title
• Article source
• Word count
• Date & time
• Language
• Article content
• Each article has approximately 500 words.</td></tr>
<tr><td>Crude oil supply</td><td>132</td></tr>
<tr><td>Crude oil demand</td><td>109</td></tr>
<tr><td>Wheat production</td><td>906</td></tr>
<tr><td>Wheat supply</td><td>65</td></tr>
<tr><td>Wheat demand</td><td>26</td></tr>
<tr><td>Wheat consumption</td><td>74</td></tr>
<tr><td colspan="3">Table 3. News Dataset (Newswire, 2023)</td></tr>
</table>

METHOD: TECHNICAL

To predict the trend price with LLM and GNN, I will apply datasets to technical methods presented in this section. There are three major parts for technical method, LLM, GNN and interaction. I will first explain the concept of LLM.

Large language model (LLM)

LLM is a type of deep learning neural network, and it can provide human-like interpretation and classification. ChatGPT 4.0 is the best LLM now, and it is made and trained by a company called OpenAI (Mauran, 2023). OpenAI provides ChatGPT's LLM for public access through OpenAI's Application Programming Interface (OpenAI API) at a cost. Therefore, I will fine-tune the OpenAI API to perform the classification of impact from news headlines and content. I will feed news articles from the Financial Times (Brower, 2023) to OpenAI API. OpenAI API will classify the news extract features and sentiments, and then provide impact scales. Impact scales interpret whether the news will affect commodity pricing. After collecting the impact scales, the next step in the method is to feed it into GNN.

Graph Neural Network (GNN)

GNN is a type of deep learning including nodes and edges. It has been used on commodity prices such as agricultural commodities (Özden & Bulut, 2024). GNN creates a graph-based algorithm with nodes and edges (Guan et al., 2022). I will set events, countries, or factors as nodes, and set production, consumption, and relation as edges.

- Nodes: Events, countries, or factors.
- Edges: Production, consumption, or relation.

Nodes and edges are combined into the three types of layers, input, hidden, and output layers. Layers can learn the graph features and provide recommendation systems and network analysis. The simplified process is shown in the infographic explanations done by freeCodeCamp (Dagli, 2022).

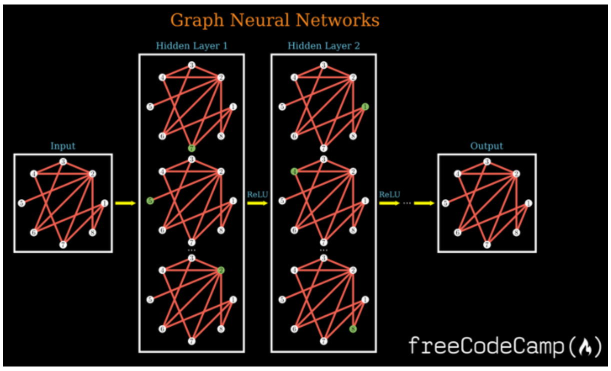

Figure 1. freeCodeCamp-GNN-1

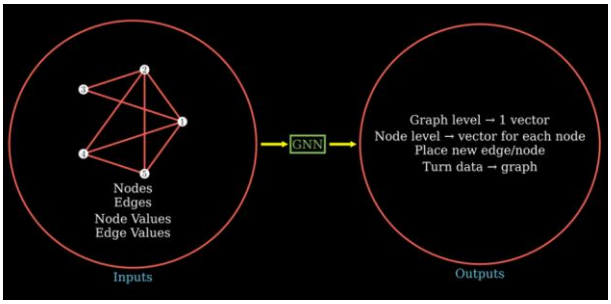

Figure 2. freeCodeCamp-GNN-2

I will use a similar GNN structure for ML forecasting. Although ML and GNN have been used for forecasting, none of the ML mythologies can integrate immediate news resources as the Large Language Model does. Therefore, I want to integrate LLM into GNN, so prediction can be more agile. The news impact scales from LLM will feed into a mechanism called Long-Short Term Memory (LSTM) (Chen et al., 2023). LSTM is an ML mechanism for the model to remember long-term input with a long-term memory. I will put impact scales into long-term memory, and price datasets into short-term memory, so GNN can have the ability to interpret news impact with LLM. Both LLM informed GNN (LLM-GNN) and capstone frameworks are presented in figures below.

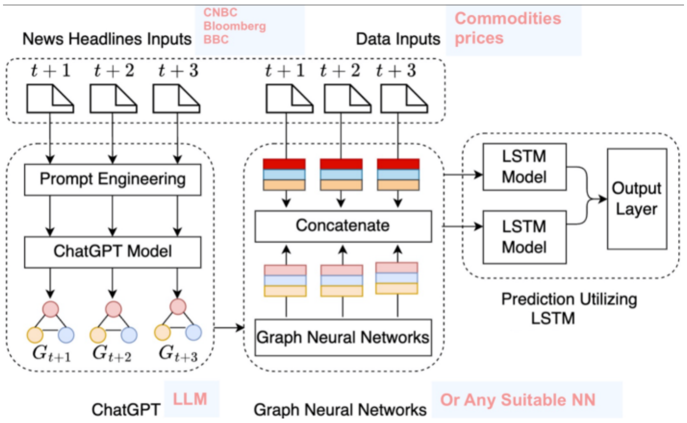

Figure 3. LLM-GNN (Machine learning)

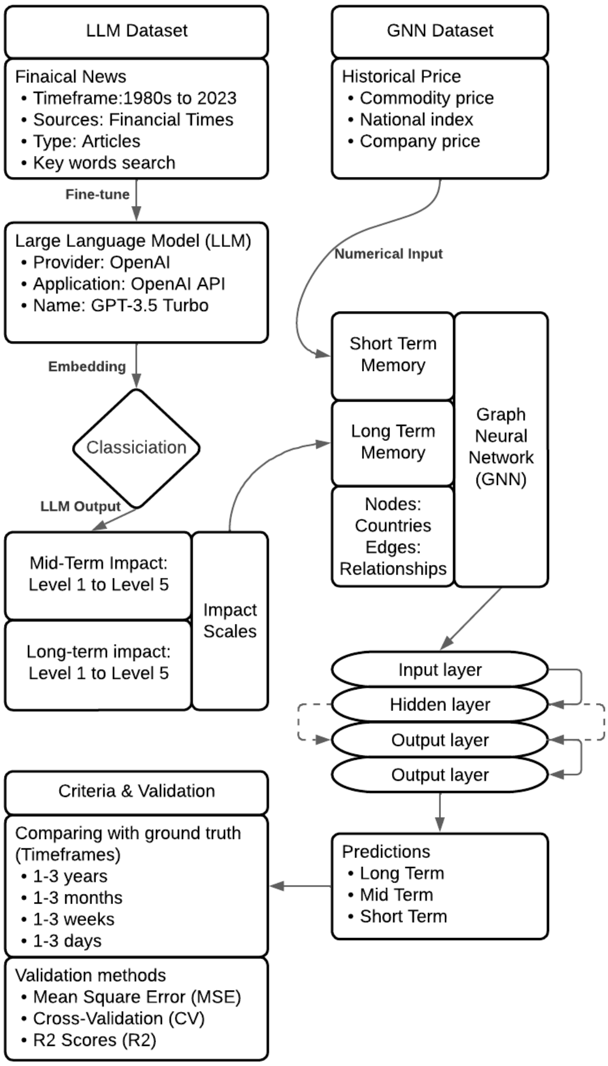

Figure 4. LLM-GNN Dataset + ML + Criteria Flow Chart

I have followed the flow chart and completed several method steps, so I have listed out the completed section and to-do section for the project. According to the Gantt chart, I have completed one-third of the project, and these are the completed tasks.

Completed Tasks

Identification:

- Identify types of commodities: I have selected crude oil and wheat as primary commodities because both have a huge impact on global food and energy security.
- Research news sources: I have chosen the Financial Times as the primary news source and gained access through Factiva with UQ library.
- Select machine learning methods: My first choice is GNN, because I have found previous research on commodity forecast using GNN. (Chen et al., 2023)

Resource collection:

- Collect all historical price datasets: I have collected datasets from investing.com (Ferlito, 2023).
- Collect news from Financial Times (1980s to present): I have collected news from Factiva.
- Ensure there are enough resources to train the ML method: I have applied computing resources though UQ.
- Apply access to the latest LLM. OpenAI API: Professor Xue Li has granted me the access to OpenAI API.

Implementation:

- Implement LLM fine-tuning on OpenAI API by feeding financial news into LLM: I have applied Embedding with features and sentiments extracted from headlines and news articles.

To-do Tasks

Implementation: (Continued)

- Perform classification on LLM and convert it into measurable impact scales.
- Integrate the impact scales to the GNN method through Long-term memory.
- Separate the price dataset into the training set and test set. (70% training and 30% test)
- Tune the ML method to the best parameters.
- Evaluate whether there is a better way to feed LLM and tune GNN.
- Improve the model based on trial and error.

Outcome evaluation:

- Create data visualization.
- Compare outcome with ground truth with different time frames. (Test dataset)
- Validate the method with suitable criteria (MSE, CV, R2) (See evaluation criteria section)
- Test the LLM-GNN model on other commodities.

These method steps ensure the approach is repeatable, and other researchers can follow the methods to obtain outcomes. The outcome will be a model framework to fill up the research gap, enabling existing machine learning models to automatically interpret the impact of the news and global events. The outcome model will be able to forecast accurate commodity prices for different time frames and pass the measurement criteria.

OUTCOME: EVALUATION CRITERIA

The LLM-GNN model will prove its ability to solve the research gap by passing the evaluation criteria. There are three evaluation criteria that I have selected based on the data science principle. Three criteria are Mean Squared Error (MSE), Cross-Validation (CV), and R2 score (Chugh, 2022).

I have chosen MES because the dataset is Non-IID (Non-Independently and Identically Distributed) data with time series property. Each data is not independent, and data often exhibits continuous trends. MSE can quantify the average squared difference between predicted and actual values, so it can capture the changing variance in the time series.

Cross-validation can perform time series cross-validation, so I can evaluate how will a model generalize to unseen data. R2 measures the proportion of variance and captures trends and patterns. The trend pattern can be uptrend, downtrend or sideways in a certain time frame, such as 1 to 3 years, 1 to 3 months, or 1 to 3 days. The prediction of smaller daily data is one of our research challenges.

OUTCOME: RESEARCH CHALLENGES & POTENTIAL DIRECTIONS

I have identified several research challenges related to the research gap. The research gap is to enable ML methods to interpret important news with LLM, and one research challenges is LLM cannot access the real-time news, because I can only access the database from Factiva through the UQ library, the capstone project does not have paid access to web scrap from the Financial Times.

On the other hand, the research challenges also include the limited detailed data for the smaller time frame. I only have access to daily price information, so any prediction smaller than daily prediction is impossible, and that's why this capstone focuses on forecasting the price trend instead of price for quantitative trading. I suggest any stakeholder who is interested in quantitative trading price forecast refine the model framework based on my work.

OUTCOME: ECONOMIC IMPACT & SOCIAL IMPACT

I am aware that quantitative price forecasts can create higher business impact for the finance industry, but this capstone focuses on creating more economic and social impact. The economic and social impact includes inflation control based on commodity prices and building more realistic budget estimation for stakeholders. Stakeholders can acquire a better prediction method with LLM-GNN method.

CONCLUSION

In summary, I propose the capstone project can utilize the large language model (LLM) and Graph Neural Network (GNN) to predict the price trend of commodity prices and fill up the research gap. The research gap is that machine learning (ML) algorithms cannot interpret geopolitical incidents from the news. For example, ML algorithms like GNN cannot evaluate the Russian-Ukraine war's impacts on crude oil prices. I have collected two commodities prices, crude oil and wheat, and related financial news as datasets. I will follow the proposed repeatable methods and schedules to use the training set to fine-tune a LLM called OpenAI API from OpenAI. My goal is to fine-tune the OpenAI API to classify the impact scale on the news. I will feed the impact scale to long-short-term memory (LSTM) for GNN. GNN will perform a prediction outcome, and I will compare the prediction with the ground truth dataset (test set), and utilize mean squared error (MSE), cross-validation (CV), and R2 score for different time frames as statistical criteria. By evaluating the criteria, I will ensure the model framework can fill up the research gap. In the end, I will present a method framework that can use LLM to inform GNN to provide an agile forecast on commodity price trends.

REFERENCE

Albano, A. (2022, November 30). *Investment banks divided on oil prices in 2023: Gs sees Brent at $110, JPM more pessimistic*. *Yahoo! Finance.* https://finance.yahoo.com/news/investment-banks-divided-oil-prices-043841817.html

Bajpai, P. (2022, March 7). *What makes oil prices go up or down*. Investopedia. https://www.investopedia.com/articles/investing/072515/top-factors-reports-affect-price-oil.asp

Brower, D. (2023, May 8). *Falling crude prices and recession fears bring US oil and gas rally to a halt*. Financial Times. https://www.ft.com/content/067a49c9-e068-4553-abba-33c7e4f9fe50

Chen, Z., Zheng, L., Lu, C., Yuan, J., & Zhu, D. (2023). Chatgpt informed graph neural network for stock movement prediction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4464002

Chugh, A. (2022, March 16). *Mae, MSE, RMSE, coefficient of determination, adjusted R squared - which metric is better?*. Medium. https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e

Dagli, R. (2022, June 16). *What are graph neural networks? how GNNS work, explained with examples*. freeCodeCamp.org. https://www.freecodecamp.org/news/graph-neural-networks-explained-with-examples/

Ephrem , J. (2023, July 7). *Open AI makes GPT-4 API generally available; new wave of AI-powered apps on the way*. Proactiveinvestors UK. https://www.proactiveinvestors.com.au/companies/news/1020050/open-ai-makes-gpt-4-api-generally-available-new-wave-of-ai-powered-apps-on-the-way-1020050.html

Fazio, F. (2023, March 10). *How to use CHATGPT to unlock new levels of Innovation*. Entrepreneur. https://www.entrepreneur.com/leadership/how-to-use-chatgpt-to-unlock-new-levels-of-innovation/446151

Ferlito, A. (2023, September 12). *Crude oil: Could the price move above $90?*. Investing.com. https://www.investing.com/analysis/crude-oil-could-the-price-move-above-90-200641751

Financial, M. (2023, July 30). *Commodity correlation, currency effects, bitcoin rallies and the Fed*. Livewire Markets. https://www.livewiremarkets.com/wires/commodity-correlation-currency-effects-bitcoin-rallies-and-the-fed

Guan, D., Liu, S., Li, Y., & Yi, P. (2022). Graph-based algorithm for self-trading accounts detection in bulk commodity trade. *2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)/2022 IEEE 8th International Conference on Edge Computing and Scalable Cloud (EdgeCom)*. https://doi.org/10.1109/cscloud-edgecom54986.2022.00033

Harvey, F. (2022, August 23). *Record profits for grain firms amid food crisis prompt calls for windfall tax*. The Guardian. https://www.theguardian.com/environment/2022/aug/23/record-profits-grain-firms-food-crisis-calls-windfall-tax

Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J.-J., & Naranpanawa, A. (2019). *Long-Term Forecast of Energy Commodities Price Using Machine Learning*, *179*, 214–221. https://doi.org/10.1016/j.energy.2019.04.077

John, D. (2022, October 26). *Federal budget October 2022: Commodity prices forecasts too “conservative”: CBA*. CommBank. https://www.commbank.com.au/articles/newsroom/2022/10/federal-budget-commodity-prices.html

Kerner, S. M. (2023, September 13). *What are large language models?: Definition from TechTarget*. WhatIs.com. https://www.techtarget.com/whatis/definition/large-language-model-LLM

Khanna, C. (2021, January 17). *Independent and identically distributed*. Medium. https://towardsdatascience.com/independent-and-identically-distributed-ce250ad1bfa8

Mauran, C. (2023, June 8). *The 5 best ai chatbots of 2023 (so far)*. Mashable. https://mashable.com/article/best-ai-chatbots-2023-list-chatgpt-bing-bard

Nelson, E., Swanson, A., & Smialek, J. (2023, August 10). *Heat, war and trade protections raise uncertainty for food prices*. *The New York Times*. https://www.nytimes.com/2023/08/10/business/global-food-prices-volatility.html

Newswire, D. J. (2023, January 18). *Dow Jones launches a reimagined Factiva Experience, providing business intelligence for the next generation of leaders*. *Yahoo! Finance*. https://finance.yahoo.com/news/dow-jones-launches-reimagined-factiva-133700377.html

Reiff, N. (2023, April 8). *10 biggest oil companies*. Investopedia. https://www.investopedia.com/articles/personal-finance/010715/worlds-top-10-oil-companies.asp

Rivero, R. (2023, August 23). *10 of the most traded commodities in the world*. Admirals. https://admiralmarkets.com/education/articles/general-trading/most-traded-commodities

Zhu, Y. (2020). Stock price prediction using the RNN model. *Journal of Physics: Conference Series*, *1650*, 032103. https://doi.org/10.1088/1742-6596/1650/3/032103

Özden, C., & Bulut, M. (2024). Spectral temporal graph neural network for multivariate agricultural price forecasting. *Ciência Rural*, *54*(1). https://doi.org/10.1590/0103-8478cr20220677

REFERENCE: GANTT CHART

Project: Capstone Project	Large Language Model (LLM) Informed Graph Neural Network for Industrial Commodities Trend Pattern Prediction																														
Start date: 24/07/2023			Semester 2 2023													Summer			Semester 1 2024												
Items/Date			W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	SM1	SM2	SM3	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13
Steps	**Category**	**Start**																													
Identification:																															
Identify types of commodities	Planning	S2W1																													
Research news sources	Planning	S2W2																													
Select machine learning methods	Planning	S2W3																													
Resource collection:																															
Collect all historical price datasets	Resource	S2W4																													
Collect news from Financial Times	Resource	S2W4																													
Ensure computing power	Resource	S2W3																													
Access to the latest LLM	Resource	S2W5																													
Implementation:																															
Implement LLM fine-tuning	Technical	S2W5																													
Perform classification	Technical	S2W8																													
Integrate the impact scales	Technical	S2W9																													
Separate the price dataset	Technical	S2W9																													
Tune the ML method	Technical	S2W11																													
Evaluate a better ML method	Technical	S2W13																													
Improve the model	Technical	S1W1																													
Outcome evaluation:																															
Create data visualization	Evaluation	S1W2																													
Compare outcomes	Evaluation	S1W3																													
Validate the method with criteria	Evaluation	S1W6																													
Test on other commodities	Evaluation	S1W6																													
Report Writing:																															
Proposal writing	Report	S2W2																													
Report writing	Report	S1W3																													