MGT 6203 Group Project Proposal

TEAM INFORMATION (1 point)

Team 9:

Team Members:

1. Rade Bajic; radeosimbio@gmail.com
    a. Engineer working in Construction Industry
    b. Education Masters in Construction Engineering and Masters in BA
2. Eddie Morrissey; emorrissey@nuvasive.com
    a. Engineer working in Medical Device Manufacturing
    b. Education Masters in Biomedical Engineering & Business
3. Stijn Astrid Leon Jorissen; stijnjorissen@gmail.com
    a. Brewing technologist, experience in the fast moving consumer goods sector
    b. Education in Chemical Engineering
    c. Experience related to analytics: finished the MITx micromaster in statistics and data science
4. Nurly Kuzdikbay; nurlykuzdikbay@gmail.com
    a. Business Analyst Intern working at Amazon
    b. Education in Civil & Environmental Engineering
5. David Basler; david.basler@ch.ey.com

OBJECTIVE/PROBLEM (5 points)

Project Title: Commodity price forecasting via machine learning methods between 1980 and 2016

Background Information on chosen project topic:

Commodity price forecasting represents a key concern for market participants, such as producers, processors, brokers, and hedgers. It is important for producers in fixing sales prices ahead of production, processors and exporters in covering costs, and speculators in generating profits. Due to high price volatilities of these commodities in times of climate change, crises like the ukrainian war and COVID-19, significant influences on decision-making, and hence on resource allocation and economic welfare, the importance of commodity price forecasting is evident.

Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):

The goal is to leverage historical price data, along with relevant market indicators and potentially weather data, to provide reliable predictions of future commodity price movements. By doing so, the project aims to assist stakeholders in making informed decisions, managing risks, and optimizing strategies in commodity trading, investment, and related sectors.

State your Primary Research Question (RQ):

How can historical price data and relevant market indicators, including potentially weather data, be effectively utilized to develop a reliable model for forecasting commodity prices?

Add some possible Supporting Research Questions (2-4 RQs that support problem statement):

1. What are the key historical price patterns and trends observed in different commodity markets, and how can they be incorporated into the forecasting model?
2. Which market indicators, such as supply and demand factors, and macroeconomic indicators have the most significant impact on commodity prices, and how can they be integrated into the forecasting model?
3. To what extent does weather data, such as temperature, precipitation, and extreme weather events, affect specific commodities, and how can it be incorporated into the forecasting model?

Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)
Financial Impact:

Accurate commodity price forecasts have substantial financial implications for businesses. Companies involved in commodity trading, manufacturing, retail, or procurement can optimize their pricing strategies, negotiate favorable contracts, and manage their input costs more effectively. This can lead to improved profitability, cost savings, and better financial performance.

Investment Decision Making:

Commodity price forecasts are crucial for businesses considering investments in sectors influenced by commodity prices. Accurate forecasts assist in evaluating the profitability and viability of investment projects, determining potential returns, and assessing risk exposure. This supports informed decision-making, capital allocation, and portfolio optimization, leading to improved investment outcomes.

DATASET/PLAN FOR DATA (4 points)

Data Sources (links, attachments, etc.):

Primary data:

- Worldwide Commodity Prices (https://www.kaggle.com/datasets/vagifa/usa-commodity-prices)

Secondary data:

- US macroeconomic data (https://www.kaggle.com/datasets/sarthmirashi07/us-macroeconomic-data)
- US weather data (https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily)

Data Description (describe each of your data sources, include screenshots of a few rows of data):

Worldwide Commodity Prices

The dataset contains monthly prices for 53 commodities and 10 indexes, starting from 1980 to 2016, Last updated on March 17, 2016.

Date	Beverage	Industrial	Agricultur	Metals Pri	Fuel Ener	Crude Oil	Aluminum	Bananas
1/1/1980	189.31	81.88965	78.90015	84.049		72.08931	2054.86	401.9608
2/1/1980	190.3879	83.04837	75.71515	88.34523		69.83942	2131.009	372.186
3/1/1980	194.0604	75.2289	69.00247	79.72631		70.98153	1978.379	422.9135
4/1/1980	186.1379	72.47125	67.87711	75.78966		70.40037	1932.456	395.8956

US macroeconomic data

The dataset contains monthly macroeconomic indicators such as the consumer price index (CPI), unemployment rate, NASDAQ Composite index, etc.

date	CPI	Mortgage_rate	Unemp_rate	NASDAQ	disposable_income	Personal_consumption_expenditure	personal_savings
11/1/1980	85.6	14.205	7.5	200.6855556	4976.5	1826.8	11.6
12/1/1980	86.4	14.79	7.2	198.3986364	4999.8	1851.7	11.4
1/1/1981	87.2	14.904	7.5	198.817619	4980.4	1870	10.9
2/1/1981	88	15.1325	7.4	194.8521053	4965	1884.2	10.8
3/1/1981	88.6	15.4	7.4	203.5931818	4979	1902.9	10.8
4/1/1981	89.1	15.58	7.2	215.12	4965.1	1904.4	10.9
5/1/1981	89.7	16.402	7.5	216.5435	4974.8	1913.8	11

US weather data

GHCN (Global Historical Climatology Network)-Daily is a database that addresses the critical need for historical daily temperature, precipitation, and snow records over global land areas. The archive includes over 40 meteorological elements, including temperature daily maximum/minimum, temperature at observation time, precipitation, snowfall, snow depth, evaporation, wind movement, wind maximums, soil temperature, cloudiness, and more.

STATION	NAME	LATITUDE	LONGITUDE	ELEVATION	DATE	PRCP	PRCP_ATTRIBUTES	TAVG	TAVG_ATTRIBUTES	TMAX	TMAX_ATTRIBUTES	TMIN	TMIN_ATTRIBUTES
RSM00023412	UST USA, RS	65.97	56.9194	77	1/1/1980	0.03	,,r	10	,,r	14	,,r	9	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/2/1980	0	T,,r	6	,,r	9	,,r	3	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/3/1980	0.04	,,r	12	,,r	19	,,r	3	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/4/1980	0.02	,,r	15	,,r	22	,,r	8	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/5/1980	0.05	,,r	11	,,r	15	,,r	8	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/6/1980	0.03	,,r	12	,,r	17	,,r	8	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/7/1980	0.02	,,r	13	,,r	16	,,r	11	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/8/1980	0.05	,,r	14	,,r	22	,,r	10	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/9/1980	0.05	,,r	14	,,r	26	,,r	3	,,r
RSM00023412	UST USA, RS	65.97	56.9194	77	1/10/1980	0.04	,,r	4	,,r	8	,,r	-2	,,r

Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)

Target variables: The variables in the commodity price dataset will be considered dependent.

Independent variables: The independent variables are chosen from the two datasets “US macroeconomic data” and “US weather data”. Before defining the variables, additional steps will be taken to determine which variables are suitable for the analysis:

- Understand the domain
- Conduct Exploratory Data Analysis (EDA)
- Feature selection techniques
- Literature review

APPROACH/METHODOLOGY (8 points)

Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))

1. Data Collection and Preparation:
    a. Gather historical commodity price data, macroeconomic data, and weather data. Preprocess and transform the data, handling missing values, normalizing, and performing feature engineering.
2. Model Selection:
    a. Explore and evaluate various models, including econometric models (ARMA, VAR, VECM) and machine learning techniques (neural networks, LSTM, ensembles).
3. Training and Hyperparameter Optimization:
    a. Train models on historical data using techniques like cross-validation. Optimize hyperparameters through grid search or Bayesian optimization to improve forecasting accuracy and generalization.
4. Model Comparison and Evaluation:
    a. Compare models based on forecasting performance. Assess accuracy, precision, and robustness to identify the most suitable model for commodity price forecasting.

Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement

1. The selected model for commodity price forecasting are expected to outperform baseline models (basic exponential smoothing over a time period of several years and/or returns of commodities index fund) and provide explanation why certain commodities can be predicted easier and some are more complex.
2. Where weather data is considered, it is expected to reveal correlations with specific commodities, contributing to improved forecasting accuracy for those commodities (e.g. agricultural commodities).
3. The optimized models, trained on historical data and fine-tuned through hyperparameter optimization, will yield improved forecasting accuracy, generalization, and robustness.

What business decisions will be impacted by the results of your analysis? What could be some benefits?

<u>Investment decision:</u>

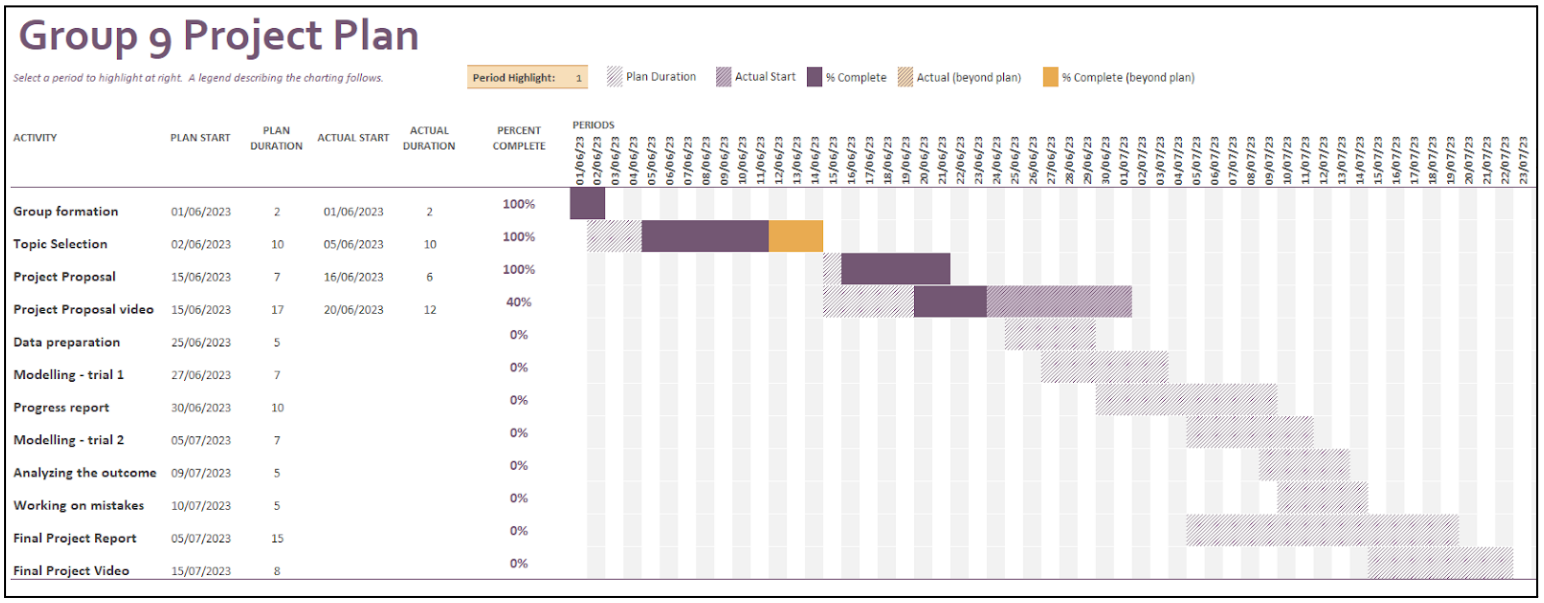
- The analysis results can influence investment decisions for commodity traders. By having a price forecast, traders can evaluate the viability and profitability of investment projects and assess the potential returns (considering the limitation of the prediction). This supports effective capital allocation, risk management, and strategic planning.

<u>Inventory management:</u>

- Reliable commodity price forecasts aid in optimizing inventory management. Businesses can anticipate price fluctuations and adjust their procurement strategies accordingly. This helps in minimizing inventory costs, aids in hedging and ensuring efficient supply chain operations. It also facilitates negotiation with suppliers, enabling businesses to secure favorable pricing and terms.

PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention key dates you hope to achieve certain milestones by:



ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Group formation	01/06/2023	2	01/06/2023	2	100%
Topic Selection	02/06/2023	10	05/06/2023	10	100%
Project Proposal	15/06/2023	7	16/06/2023	6	100%
Project Proposal video	15/06/2023	17	20/06/2023	12	40%
Data preparation	25/06/2023	5			0%
Modelling - trial 1	27/06/2023	7			0%
Progress report	30/06/2023	10			0%
Modelling - trial 2	05/07/2023	7			0%
Analyzing the outcome	09/07/2023	5			0%
Working on mistakes	10/07/2023	5			0%
Final Project Report	05/07/2023	15			0%
Final Project Video	15/07/2023	8			0%

Group 9