Predictive Analysis to Determine Future Price of Base Metals

TORONTO METROPOLITAN UNIVERSITY

Project By: Arif Virani
Student ID: 501148440

Abstract

Background:
The pricing of commodities, particularly industrial metals like Tin, Lead, Copper and Nickel, plays a crucial role in global economic dynamics. These metals are integral to various industries, including manufacturing and construction. Understanding and predicting their price movements are of importance for investors, businesses, and policymakers. The past decade has witnessed significant fluctuations in metal prices influenced by factors such as inflation rates, Manufacturing Purchasing Managers Index (PMI), currency value and metal-specific variables like their import data. In addition to these factors listed above, other global events like wars, Covid-19 pandemic have had adverse effects on mining activities as well as disruption of global supply chain, impacting the metal prices. These factors will be out of scope for this project and will be listed as a limitation.

Method:
This project leverages a comprehensive dataset comprising of historical metal prices spanning from January 2008 to August 2023, sourced from investing.com. The dataset encompasses 14,928 records, providing a substantial foundation for analysis. In addition to metal prices, key explanatory variables include inflation data, Manufacturing PMI, USD currency exchange rates, and import statistics for each metal category. These variables are selected based on their known influence on metal price dynamics.

Analysis:
To achieve the project's objectives of predicting commodity prices accurately, we will employ a data-driven approach. Apart from exploratory data analysis using pandas library in python, initial data preprocessing steps will involve handling missing values, normalization and/or feature engineering. Subsequently, a suite of predictive modeling techniques will be applied. This will include regression models, time series analysis, and machine learning algorithms, tailored to the specific characteristics of each metal. Our analysis will aim to uncover the relationships between metal prices (daily open price, close price, high price and low price) and the explanatory variables, that will enable us to develop robust predictive models. Later on, we will assess model performance through various metrics and cross-validation techniques to ensure their reliability.

This project's significance lies in its potential to provide valuable insights into the price dynamics of important industrial metals, enhancing decision-making processes for a wide range of stakeholders. The inclusion of different variables and a substantial dataset from 2008 to 2023 offers a comprehensive perspective on metal price movements in the context of global economic fluctuations and specific metal market dynamics.

Literature Review

Literature review done for this project aims to understand the existing research done on the base metal price prediction, different approaches taken by various researchers and analysts across the globe as well as understanding the dynamics of the mining industries. This literature review analysed publicly available research reports, news articles, reports from mining industry analysts to understand the topic better. There are various factors that impact the price of base metals, and each factor can be measured by different parameters.

Research done by University of Ottawa authors Anh, N.B. and Semenov, A. (2015) model the real prices of base metals as the equilibrium of aggregate supply and demand, aiming to understand the determinants of the decreasing trend in base metal prices over time. They find that the trend in base metal prices is influenced by factors such as technological progress, resource scarcity, natural resource taxes, and the interest rate. The main driver of this decreasing trend is the substitution effect and technological progress. Notably, a high natural resource tax results in higher prices but slower rate changes over time.

The research combines the supply side, considering a Cobb-Douglas production function for base metals, and the demand from the manufacturing sector, revealing that both supply and demand are decreasing, justifying the possibility of a downward sloping price trend. The model accounts for various features of base metals, including their recyclability, common deposits, positive cross-elasticities of demand, and durability.

The conclusions emphasize that existing theoretical literature does not comprehensively consider all factors influencing base metal prices, particularly neglecting market demand and speculative traders. The research answers the question of what drives the trends in base metal prices, attributing these trends to factors like total factor productivity in the mining industry, natural resource availability, taxes, interest rates, and economic demand. Additionally, they find that prices may decline over time based on numerical simulations, which is a novel finding compared to other models in the field.

We also examined research done by Wang, T., & Wang, C. (2019) that aims to uncover the spillover effects of China's industrial growth on price changes of base metal. This research paper used SVAR (Structural Vector Autoregression) models, based on monthly data from 1999 to 2016. The findings indicate that China's industrial growth has both direct lagged positive effects and indirect contemporaneous positive effects on the price changes of base metals, except for zinc. This implies that China's real activity has a substantial positive influence on the prices of base metals, excluding zinc. However, the study also reveals that the impact of the US dollar exchange rate on base metal prices is significantly larger than that of China's industrial growth, suggesting that China's influence on base metal prices might be overstated.

In conclusion, China's industrial growth has a significant and positive impact on base metal prices, particularly aluminum, copper, lead, nickel, tin, and zinc. These effects are both direct and indirect, operating through energy prices. While China's influence on base metal prices is notable, the study underscores that fluctuations in the US dollar exchange rate have a more significant influence on base metal prices. The paper suggests that China's influence on metal prices may have been somewhat exaggerated, and the focus should also be on the monetary side of the world economy, particularly the US dollar exchange rate, for predicting metal prices. Additionally, it is noted that as China's economy slows down and undergoes energy-saving reforms, the global demand for metals may experience downward pressure, potentially requiring the emergence of other major economies, such as India, to affect world metal markets positively in the future.

Another research paper by Brown, P. P., & Hardy, N. (2023) investigates the predictability of base metal prices, specifically aluminum, copper, lead, nickel, tin, and zinc, as well as the London Metal Exchange Index, using expectations about the future evolution of the Chilean exchange rate. The paper employs in-sample and out-of-sample analyses to determine the predictive ability of exchange rate expectations compared to the first lag of the Chilean currency's returns.

The findings indicate that exchange rate expectations have the ability to predict returns of the mentioned base metals and the LME Index. Notably, the predictability of exchange rate expectations is, in many cases, stronger than that of the first lag of the Chilean exchange rate. The paper also highlights that exchange rate expectations capture most of the predictive information contained in the cumulative 2-month return of the Chilean peso. This is significant because exchange rate expectations are often released to the public later than the other predictors used in the study, suggesting their robustness as forecasting tools.

The results have implications for understanding the predictive relationships between commodity currencies and commodity prices, challenging the efficient market hypothesis and demonstrating that predictability can coexist with market efficiency. The study encourages further research to explore similar predictive relationships with other commodity currencies and exchange rate fundamentals. In summary, the research provides new evidence supporting the idea that exchange rate expectations, particularly those related to commodity currencies like the Chilean peso, can predict base metal prices effectively, offering a new approach to forecasting commodity markets.

Research done by Mahmoodreza Modirroosta focuses on the analysis of long-term behavior in base metal prices using decomposition analysis. The study is divided into two main sections. In the first section, econometric methods are employed to identify long-term trends in base metal prices, considering stationarity tests like the Dickey-Fuller test. The results help determine the most suitable trend identification method.

The second section delves into the cyclical component of base metal prices, exploring various detrending approaches, including linear, quadratic trends, and more advanced techniques like Hodrick-Prescott and Baxter-King filters. The study evaluates these methods for their ability to extract cycles without altering their duration, identify time trend components, and eliminate unit roots. Ultimately, this research seeks to enhance the understanding of base metal price dynamics, benefiting the metals industry.

Another research titled ‘Theory of storage, inventory and volatility in the LME base metals’ investigates the relationship between inventory levels and price volatility for six base metals traded on the London Metal Exchange (LME), which include aluminum, copper, lead, nickel, tin, and zinc. The study is rooted in the Theory of Storage, which posits that two predictions hold true for commodities: when inventory is low (scarcity), spot prices will exceed futures prices, and spot price volatility will exceed futures price volatility; conversely, during periods of no scarcity, both spot prices and spot price volatility will remain relatively subdued.

The research validates both predictions of the Theory of Storage for the six base metals on the LME. The key findings include:

- A strong non-linear relationship between the forward curve spread (based on the ratio between spot and futures prices) and inventory levels.
- Spot volatility consistently exceeds futures volatility, with this difference being amplified in times of scarcity.
- Inventory figures from LME warehouses are sufficient to demonstrate the strong relationship between inventory and price/volatility. Adding inventory figures from the Shanghai Futures Exchange (SHFE) further strengthens the relationship, emphasizing China's growing influence in metal demand and trade.
- There is no evidence supporting recent allegations of major market players withholding inventory, suggesting that LME prices behave as if full inventory figures are available to the market.

Lastly, in terms of Literature review, a project done by Marcus Chan explores the topic of predicting precious metals like gold, silver, platinum, copper and Palladium instead of base metals. This project includes an Exploratory Data Analysis focusing on investigating correlations between prices of different commodities as well as the daily open price, high price, low price and close price. The researchers have also employed an augmented Dickey-Fuller test, to examine whether the stock prices are stationary. As the open prices were stationary, log transformations are applied to make the variable non-stationery and difference between the log of prices are calculated. Later, various decision tree-based models are employed including Random Forest, XGBoost, Gradient Boosting, and Histogram Based Gradient to all the response variables to make predictions.

Finally, training and validation sets are created to check if the models mentioned above are effective in predicting the prices. The study concludes that the models used do not do a great job at predicting the prices due to high inconsistency and variability in the prices. It recommends adding more variables to mix to improve the performance of the model.

These articles and research papers delve into a range of subjects concerning the application of machine learning in forecasting future prices of metals, encompassing both base metals and precious metals. They leverage a diverse set of data sources, combining existing and newly collected data, to approach this challenge from various angles. Given the continuous evolution of the machine learning field and the inherent volatility of metal prices, influenced by multiple factors, none of the cited studies can claim precise price predictions. Instead, these studies serve as valuable guide for shaping the direction of this project. Consequently, even if a project explores questions like those addressed in existing research, it holds the potential to contribute additional insights and build upon the findings of these studies.

Methodology:

The study will utilize base metal price data sourced from Investing.com spanning the period from 2008 to 2023, with a daily frequency. Furthermore, it will incorporate macroeconomic indicators such as the Consumer Price Index (CPI) to measure inflation and China's Purchasing Managers Index (PMI). The primary objective of the study is to employ various machine learning models for the purpose of predicting metal prices. Subsequently, an evaluation will be conducted to determine which model yields the most favorable results in terms of efficiency, effectiveness, and stability. This assessment will be performed systematically, following a structured approach known as the data science process.

The data science process serves as a framework that ensures a seamless workflow, reducing the risk of overlooking critical steps and enhancing the accuracy of the outcomes. This process offers guidance for the stages involved in collecting data, transforming it into high-quality input, constructing and assessing models, and interpreting and sharing the findings. The process encompasses the following stages:

1. Problem Understanding
2. Data Collection
3. Data Understanding and Preparation
4. Exploratory Data Analysis
5. Modeling and Validation
6. Result Communication
7. Next Steps

Structure Data Science Process:

Problem Statement description

This stage delves in defining the problem, its context and why is it important to solve.

Data Collection

This stage delves in identifying various sources of high-quality data sources suitable to answer the research problem.

Data understanding and Preparation

Involves data cleaning, standardization and brining the data in a form which can be used for analysis.

Exploratory Data Analysis

Initial summary statistics for variables, checking correlations, patterns within variables and identifying key features.

Modeling and Validation

Employing Machine learning supervised and unsupervised models including regression/classification/clustering and checking the performance of each model.

Communication and Next Steps

Presenting the conclusion of the study to stakeholders and delving on next step to expand the research question.

Data Science steps pertaining to this project are as follows:

1. Understanding the problem

In order to gain insight into topic, I undertook a comprehensive literature review focusing on base metal prices and the various factors influencing them. This literature review facilitated my exploration of the following key inquiries:

- What is my existing knowledge on this subject?
- How can I offer critical insights into the established knowledge?
- Has any prior work identical to mine been undertaken?
- Where does my research fit within the existing body of research?

Within the literature review section, I detailed how different researchers employed diverse machine learning algorithms to analyze the data.

2. Data Collection and Understanding

   This study will be using base metal price data from investing.com for 4 key base metals. To enrich the dataset, other macro-economic indicators like monthly US CPI, China CPI from International Monetary Fund and Purchasing Manager's index data from investing.com is included.

3. Data Preparation

   I utilized Panda profiling, alongside Python in the Jupyter Notebook environment, to examine and address issues like missing values, duplicates, null entries, data types, erroneous records, and special characters.

4. Exploratory Data Analysis (EDA)

   This step will include univariate and bivariate analysis and Panda profiling can be used to achieve this.

5. Modeling and Validation

   Following the literature review, I intend to perform three machine learning algorithms, and conduct predictive analyses, and produce classification reports (Random Forest, XGBoost, Gradient Boosting). Additionally, I will be computing the execution time for each classifier and assessed feature importance.

6. Communication and Next Steps:

   The last step would be to present the evaluation of the models based on their efficiency, effectiveness and stability. Also, ideas for further enhancing the research will be shared at this stage.

Data Description and Exploratory Data Analysis

Data Description:

This study aims at predicting the base metal price based on macro-economic factors like Inflation (measured by CPI) and demand/economic activity measured by Purchasing Manager's Index. The base metal prices are available at a daily frequency and can be modeled at a daily level whereas macro-economic data like CPI and PMI are captured at a monthly level. In order to manage this limitation, the CPI and PMI data is extrapolated at a daily level for each month. As per the literature review it is understood that demand from China and US has a bigger impact on the prices and hence this study will consider CPI and PMI index of US and China.

The dataset contains 14,928 records distributed among the 4 key base metals against 11 attributes including price data as well as macro-economic indicators. Below are the key data attributes considered for the study along with their description.

Attribute	Description	Variable Type
Date	Represents daily date from 2008 to 2023	Ordinal
Price	Close price of the corresponding base metal in USD measured at the end of the trading day	Ordinal
Open	Open price of the corresponding base metal in USD measured at the beginning of the trading day	Ordinal
High	Highest daily price of the corresponding base metal in USD measured at any time during the trading hours	Ordinal
Low	Lowest daily price of the corresponding base metal in USD measured at any time during the trading hours	Ordinal
Vol.	Represents the volume of the base metal traded in metric tonne in a trading day	Ordinal
Commodity	Name of the base metal (Copper, Tin, Nickel and Lead)	Categorical
US CPI	Consumer Price Index for US including all items (indexed as 100 for year 2010)	Ordinal
China CPI	Consumer Price Index for China including all items (indexed as 100 for year 2020)	Ordinal
US PMI	Represents the ISM PMI i.e., Institute of Supply Management Manufacturing Purchasing Index. It captures data compiled from executives in the industry on activities like new orders, backlog of orders, Export orders, Imports, Production etc.	Ordinal
China PMI	China's Manufacturing Purchasing Managers Index (PMI) that provides an early indication of economic activities in manufacturing sector. It is compiled by China Federation of Logistics & Purchasing (CFLP) and China Logistics Information Centre (CLIC), based on data collected by the National Bureau of Statistics (NBS)	Ordinal

Exploratory Data Analysis

A screen shot of Jupyter notebook output showcasing the basic structure and first few rows of the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14928 entries, 0 to 14927
Data columns (total 12 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       14928 non-null  datetime64[ns]
 1   Price      14928 non-null  float64
 2   Open       14928 non-null  float64
 3   High       14928 non-null  float64
 4   Low        14928 non-null  float64
 5   Vol.       7661 non-null   object
 6   Change %   14928 non-null  object
 7   Commodity  14928 non-null  object
 8   US CPI     14928 non-null  float64
 9   China CPI  14928 non-null  float64
 10  US PMI     14928 non-null  float64
 11  China PMI  14928 non-null  float64
dtypes: datetime64[ns](1), float64(8), object(3)
memory usage: 1.4+ MB
```

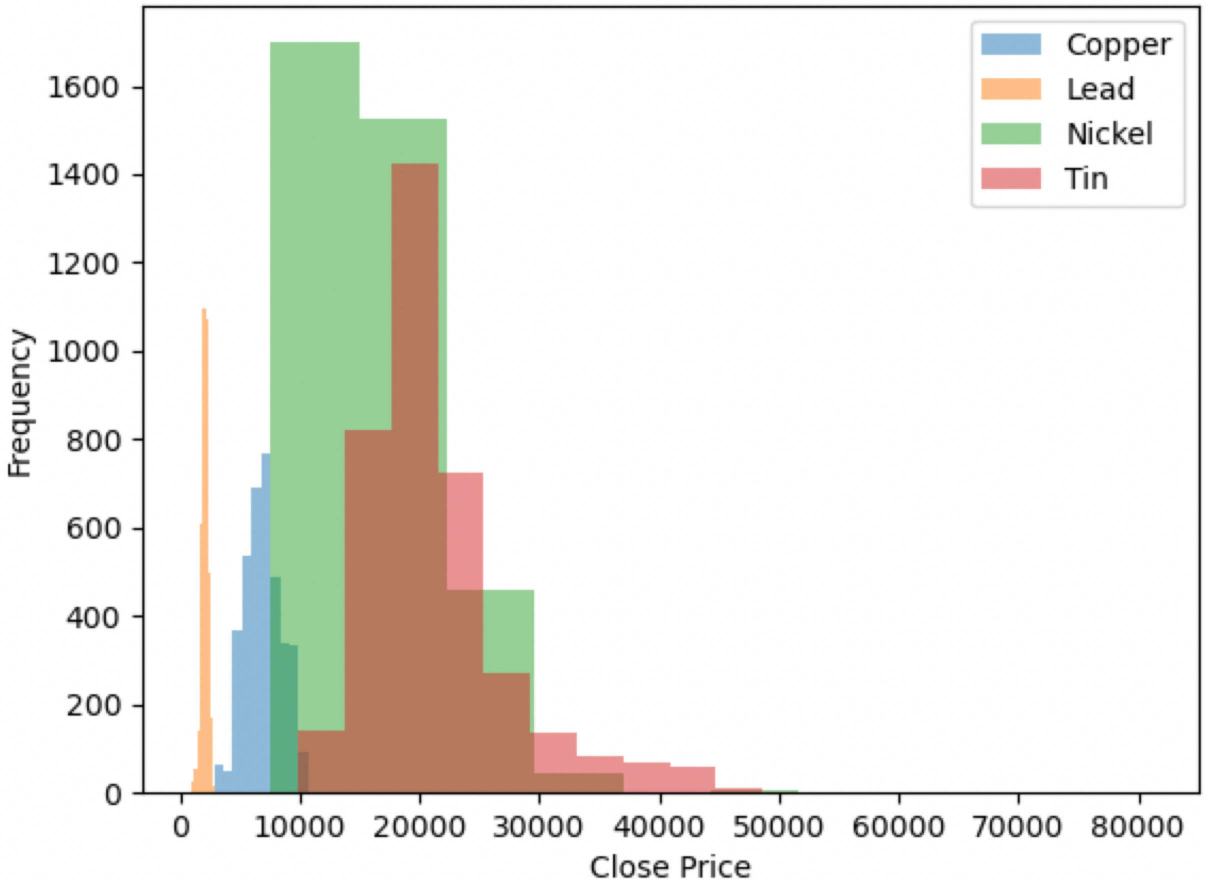
	Date	Price	Open	High	Low	Vol.	Change %	Commodity	US CPI	China CPI	US PMI	China PMI
0	2023-08-31	25,396.00	25,300.00	25,150.00	25,110.00	1.25K	-0.31%	Tin	140.801768	103.2	47.6	49.7
1	2023-08-30	25,475.00	25,400.00	25,300.00	25,325.00	1.27K	0.32%	Tin	140.801768	103.2	47.6	49.7
2	2023-08-29	25,395.00	25,200.00	25,050.00	25,055.00	1.32K	-0.40%	Tin	140.801768	103.2	47.6	49.7
3	2023-08-25	25,497.00	25,600.00	25,500.00	25,575.00	0.59K	-1.44%	Tin	140.801768	103.2	47.6	49.7
4	2023-08-24	25,870.00	25,900.00	25,825.00	25,850.00	0.39K	-0.98%	Tin	140.801768	103.2	47.6	49.7

Below code checks for duplicates and null values:

```
In [13]: #Check the duplicates
         dataset.duplicated().sum()

Out[13]: 0

In [15]: #Check the null values

         dataset.isnull().sum()

Out[15]: Date            0
         Price           0
         Open            0
         High            0
         Low             0
         Vol.         7267
         Change %        0
         Commodity       0
         US CPI          0
         China CPI       0
         US PMI          0
         China PMI       0
         dtype: int64
```

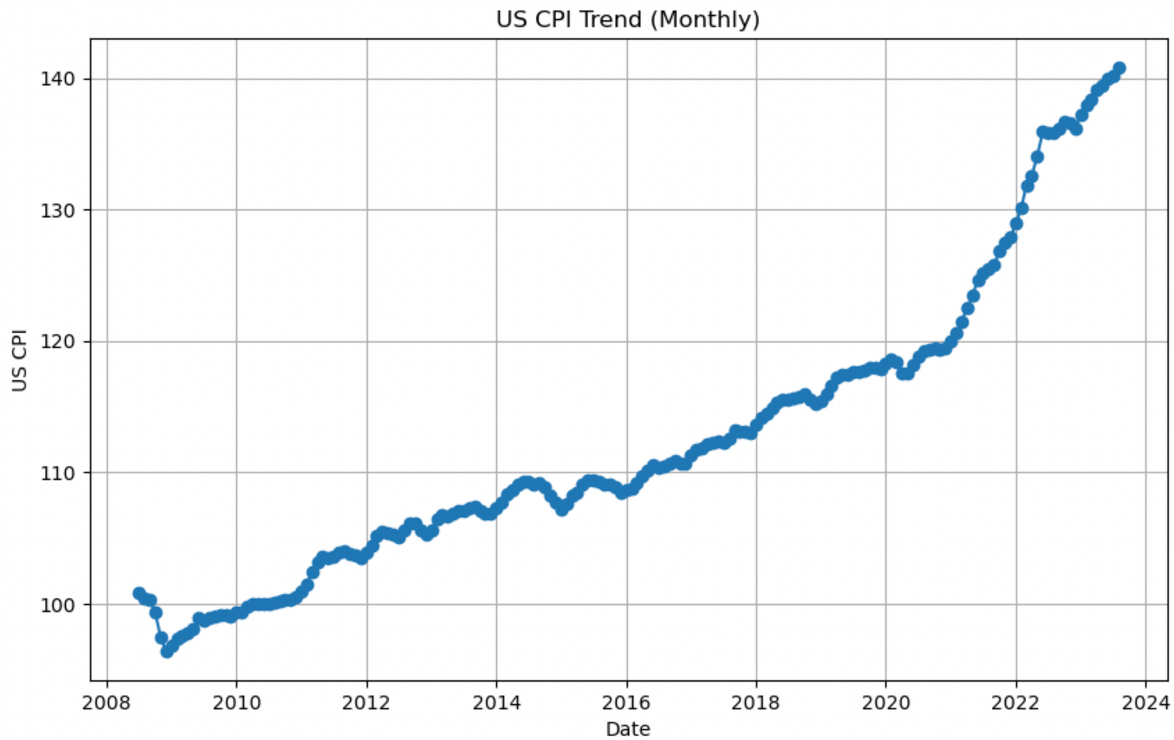
As ‘volume’ data has many null values (~50%), it will be excluded from further analysis.

Below histogram shows the distribution of close price by different metals:



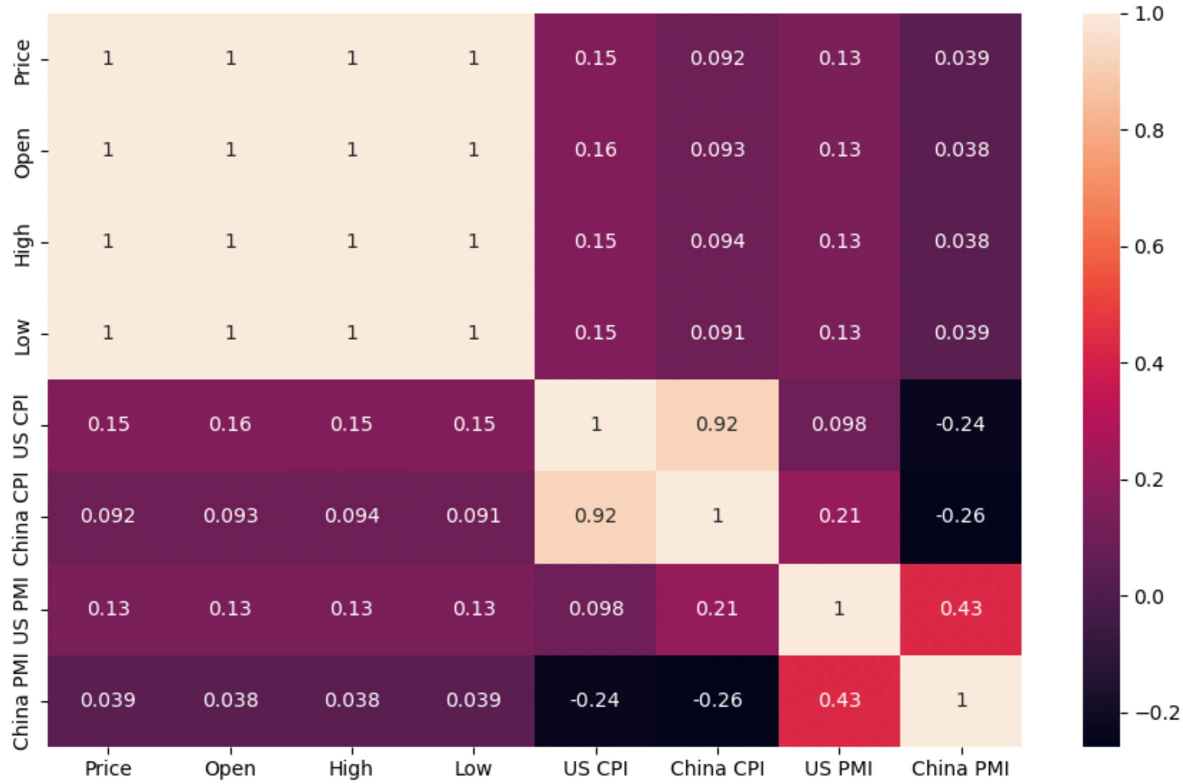
It shows that lead has the lowest average close price (per metric tonne) and variability whereas tin has the highest average close price (per metric tonne).

Next, we will plot the CPI Index for both China and US. One of the limitations of CPI index is that, for US data it is indexed to 2010 whereas for China data it is indexed to 2020. Hence the incremental value should be considered instead of comparing the index directly.


US CPI Trend (Monthly)
US CPI
Date
140
130
120
110
100
2008
2010
2012
2014
2016
2018
2020
2022
2024


China CPI Trend (Monthly)
China CPI
Date
105
100
95
90
85
80
75
2008
2010
2012
2014
2016
2018
2020
2022
2024

Next, we look at correlations between various numeric attributes of the dataset. It can be observed that similar attributes like prices, CPI (US and China) and PMI (US and China) show positive correlations.



The GitHub link to the entire initial analysis is as follow: https://github.com/Arif-Virani/Base-Metal-Price-Prediction/blob/aae70f05d2eb9d75d2c2bde1738bc232d7f7b237/MetalFuturesHistorical.ipynb

Modelling Techniques

For predicting future metal prices, we have used the following modelling techniques:

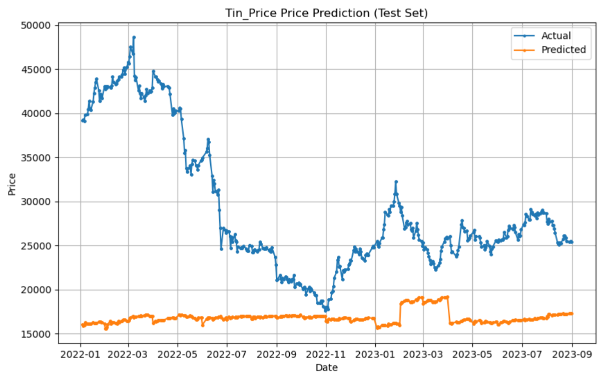
1. Random Forest
2. XGBoost
3. KNN

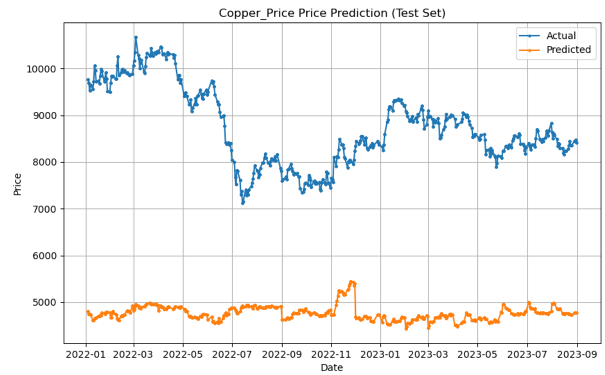
Following are key results achieved for after running the 3 models:

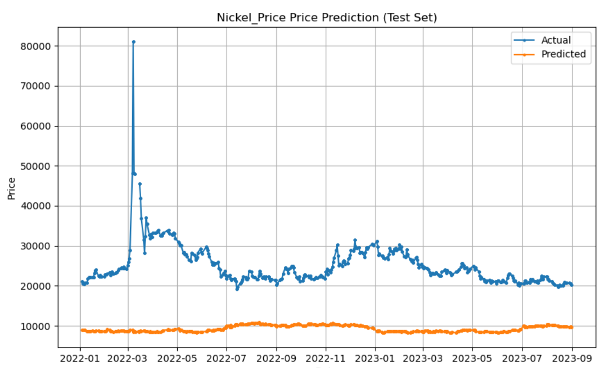
Evaluation Measures	Metal	Random Forest	XGBoost	KNN
RMSE	Tin	6746	6181	6531
R-Squared	Tin	-0.399	-0.174	-0.3119
Inference Time (Validation)	Tin	0.9510 s	0.1480 s	0.0184 s
Inference Time (Test)	Tin	0.0097 s	0.003 s	0.015 s
RMSE	Copper	2352	2361	2290
R-Squared	Copper	-2.04	-2.06	-1.88
Inference Time (Validation)	Copper	0.8372 s	0.1163 s	0.0058 s
Inference Time (Test)	Copper	0.0108 s	0.0025 s	0.0147 s
RMSE	Nickel	5231	5155	5463
R-Squared	Nickel	-2.128	-2.037	-2.411
Inference Time (Validation)	Nickel	0.8437 s	0.0807 s	0.0049 s
Inference Time (Test)	Nickel	0.0109 s	0.0027 s	0.0417 s
RMSE	Lead	356	310	352
R-Squared	Lead	-1.350	-0.781	-1.29
Inference Time (Validation)	Lead	0.8493 s	0.077 s	0.0053 s
Inference Time (Test)	Lead	0.0101 s	0.0024 s	0.0164

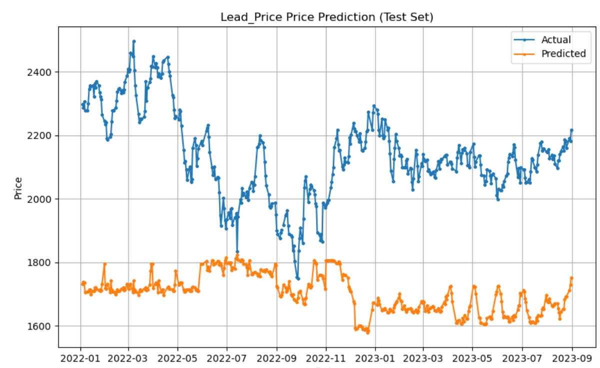
Github link to initial results: https://github.com/Arif-Virani/Base-Metal-Price-Prediction/blob/InitialResults/README.md

Predicted price vs. Actual Price – Random Forest

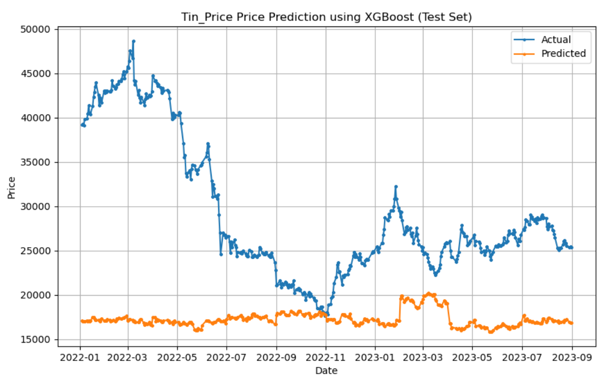


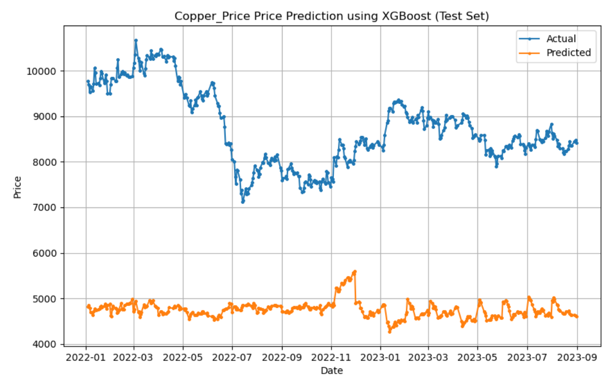


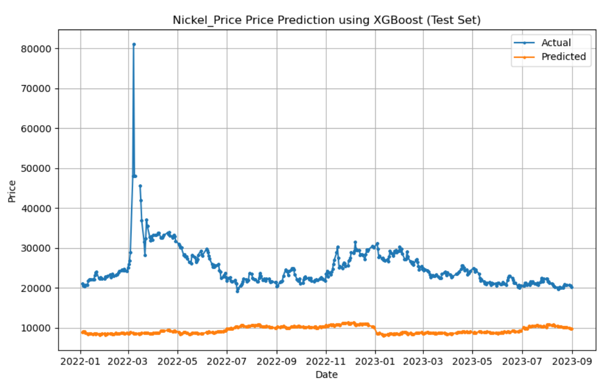


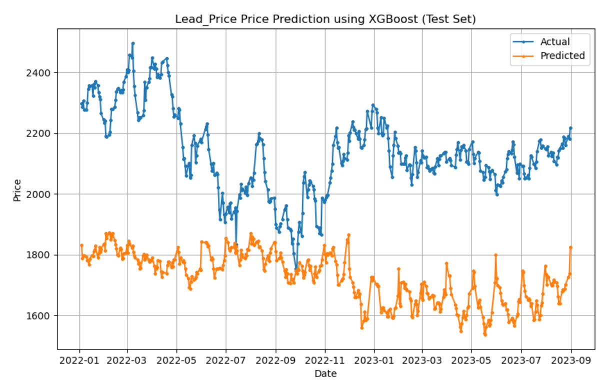


Predicted price vs. Actual Price – XGBoost

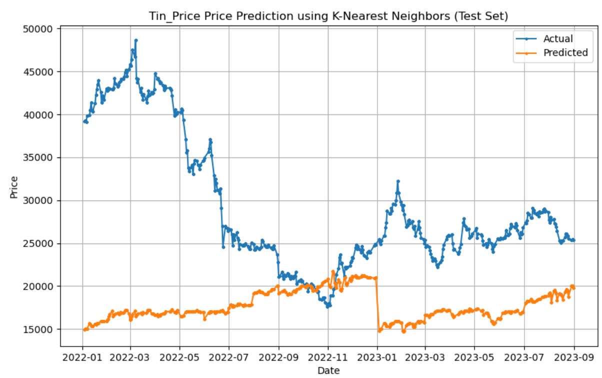


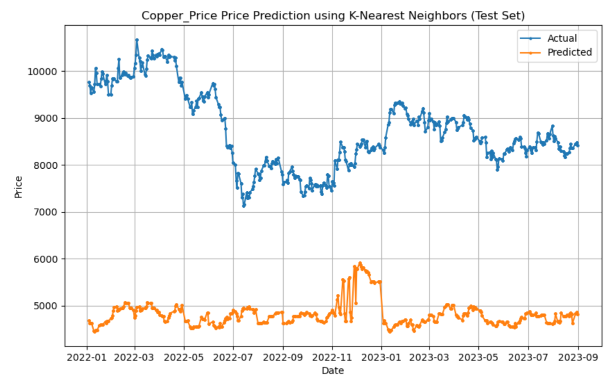


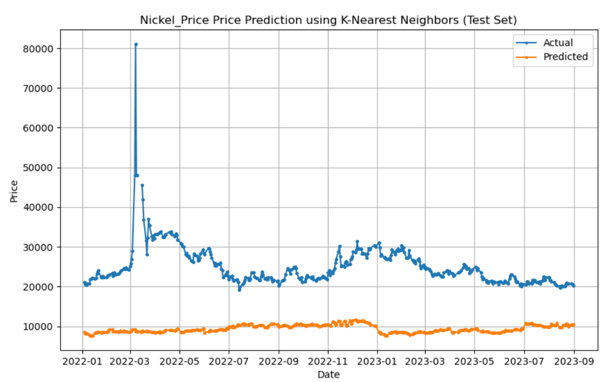


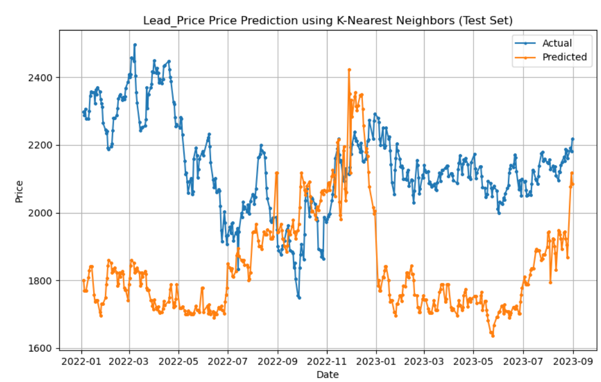


Predicted price vs. Actual Price – KNN









Conclusion and Next steps

Looking at the initial results, XGBoost seems to perform better except for copper dataset where KNN seems to be leading in terms of better R squared and inference time. For all the three modelling techniques, R-Squared is negative which indicates that the models do not do a good job at predicting stock price. This could be because:

- Due to the high inconsistency and variability in the metal prices, which might make it impossible for a model to predict future stock prices.
- The test data is from 2022-23 which was an unusual period and difficult for the model to predict the correct prices based on past data.
- Since it is a time series data, the dataset could be stationery or seasonality could be impacting the prices. Other research studies have performed Dickey-Fuller test to check the same. This can be done as a next step for the study.

References:

Tin Futures Historical prices. Investing.com Canada. (n.d.). https://ca.investing.com/commodities/tin-historical-data

Lead Futures Historical prices. Investing.com Canada. (n.d.). https://ca.investing.com/commodities/lead-historical-data

Copper Futures Historical prices. Investing.com Canada. (n.d.). https://ca.investing.com/commodities/copper-historical-data

Nickel Futures Historical prices. Investing.com Canada. (n.d.). https://ca.investing.com/commodities/nickel-historical-data

China Manufacturing Purchasing Managers Index. Investing.com Canada. (n.d.). https://www.investing.com/economic-calendar/chinese-manufacturing-pmi-594

U.S. ISM Manufacturing Purchasing Managers Index. Investing.com Canada. (n.d.). https://ca.investing.com/economic-calendar/ism-manufacturing-pmi-173

IMF Country CPI data. Imf.org. (n.d.). https://data.imf.org/regular.aspx?key=61015892

Vasishtha, G. (2022, January 24). Commodity price cycles: Causes and consequences. World Bank Blogs. https://blogs.worldbank.org/developmenttalk/commodity-price-cycles-causes-and-consequences

Home, A. (2023, July 4). Old economy drags base metals lower in first half of 2023. Reuters. https://www.reuters.com/markets/europe/old-economy-drags-base-metals-lower-first-half-2023-2023-07-04/

Pincheira, Pablo M. and Hardy, Nicolas, Forecasting Base Metal Prices with Commodity Currencies (January 2, 2018). Available at SSRN: https://ssrn.com/abstract=3095448 or http://dx.doi.org/10.2139/ssrn.3095448

Modirroosta, Mahmoodreza, "A DECOMPOSITION ANALYSIS OF BASE METAL PRICES: COMPARING THE EFFECT OF DETRENDING METHODS ON TREND IDENTIFICATION AND CYCLICAL COMPONENTS", Master's Thesis, Michigan Technological University, 2013. https://doi.org/10.37099/mtu.dc.etds/674

Anh, N.B. and Semenov, A. (2015) Trends of Base Metals Prices. Theoretical Economics Letters, 5,531-540. http://dx.doi.org/10.4236/tel.2015.54062

Pincheira Brown, P., & Hardy, N. (2019). Forecasting base metal prices with the Chilean exchange rate. Resources Policy, 62, 256-281. 10.1016/j.resourpol.2019.02.019

Wang, T., & Wang, C. (2019). The spillover effects of China's industrial growth on price changes of base metal. *Resources Policy, 61*, 375-384. 10.1016/j.resourpol.2017.11.007

Chan, M. (2023). (rep.). *Precious Metals' Stocks - EDA and Prediction*. Retrieved October 16, 2023, from https://www.kaggle.com/code/mcpenguin/precious-metals-stocks-eda-and-prediction

Hélyette Geman, William O. Smith, Theory of storage, inventory and volatility in the LME base metals, Resources Policy, Volume 38, Issue 1,2013, Pages 18-28, ISSN 0301-4207, https://doi.org/10.1016/j.resourpol.2012.06.014