



# Gold Price Prediction

Capstone



# Agenda

**01**

**Background  
& Problem Statement**

**02**

**Datasets  
& Data Dictionary**

**03**

**Data Exploratory**

**04**

**Feature Selection  
& Engineering**

**05**

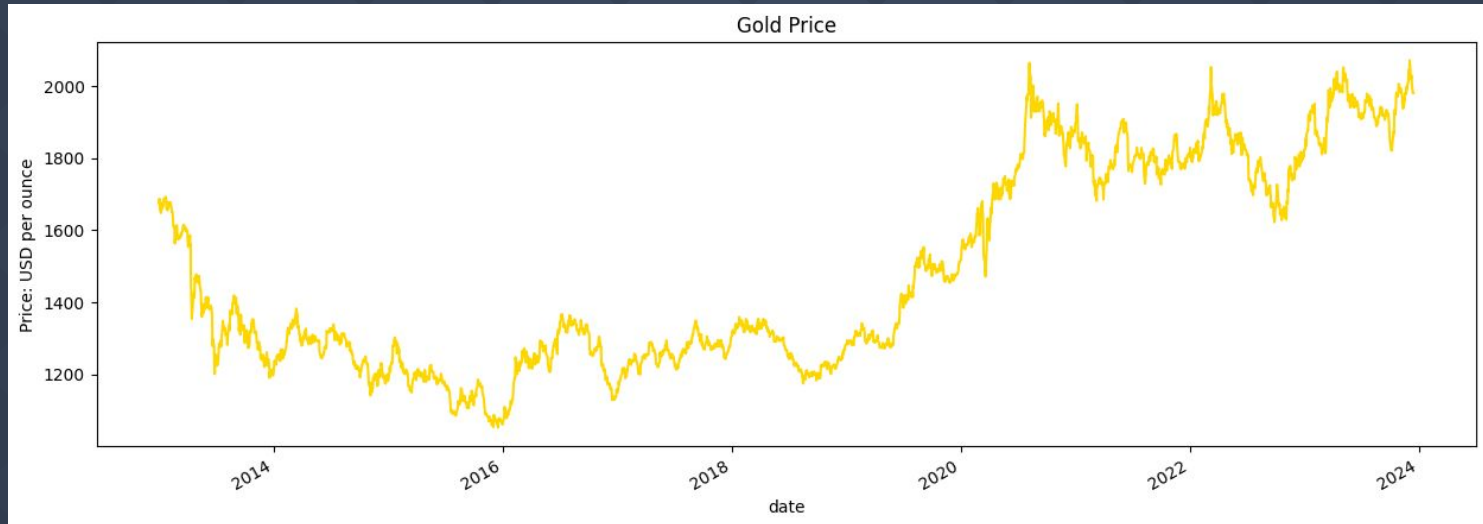
**Modeling**

**06**

**Summary**

# Background

- Gold is “Safe Haven” – Tangible assets used in everything from electronics to jewelry
  - Always in demand
- Gold World Council announce gold price



# Problem Statement

The objective is to analyze and determine the key factors or combinations of factors that significantly influence the trends in gold prices.

Investigate the relationships and interdependencies among these variables to understand the factors driving changes in gold price trends. The aim is to create a model capable of accurately classifying the future direction of gold prices, thereby providing valuable insights for decision-making and risk management in gold investment strategies.

Focus Users: Investors and stakeholders who are interest in gold investment

# Data Sets

## Gold Price

Daily data set: **2581** days  
(close Saturday and Sunday)



## Crude Oil

XAU Index  
(Philadelphia Gold and Silver mines)



higher oil prices can impact gold mining costs, potentially affecting gold supply and prices



## Federal Fund Rate



Return from Interest rate of bond and saving accounts, demand of gold decreased and lowering price



## Stock Index:

Dow Jones, NASDAQ, S&P500



Inverse Relationship that when stock market perform wells, gold demand is dropped, On the other hand, market downturns, gold become safe haven



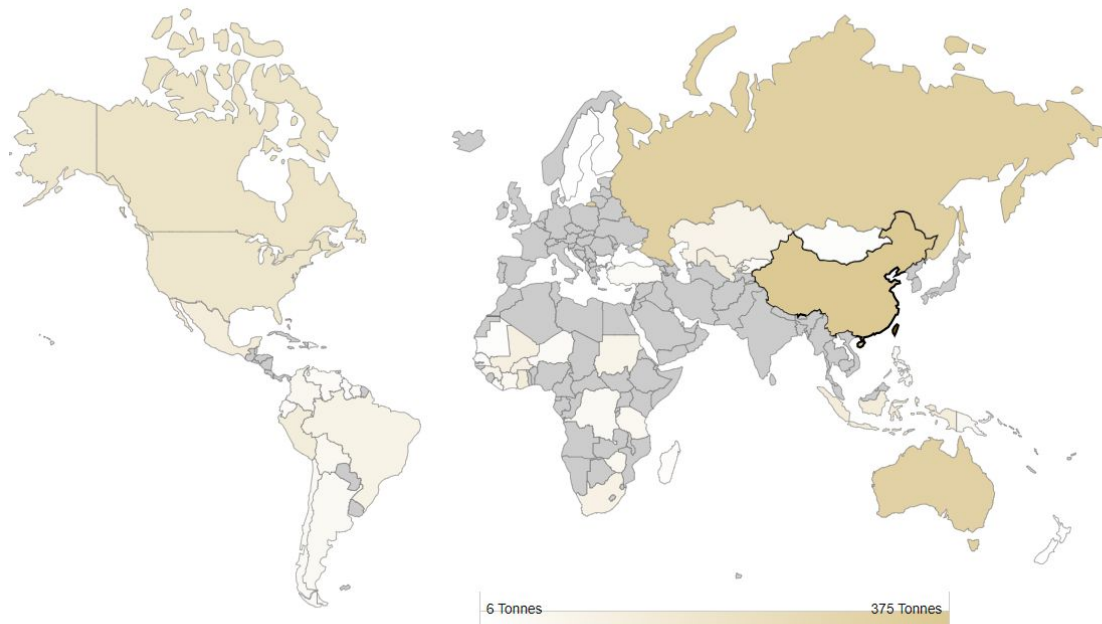
## Exchange Rate

THB, RUB, AUD, CNY to USD



Weak currency is going to be high demand of gold

# Data Dictionary



Filter and zoom

Country ▲ ▼ Tonnes ▲ ▼

China	375.0
Russian Federation	324.7
Australia	313.9
Canada	194.5
United States	172.7
Ghana	127.0
Peru	125.7
Indonesia	124.9
Mexico	124.0

Source: Metals Focus; World Gold Council

# Data Dictionary (Stock Index & Economic)

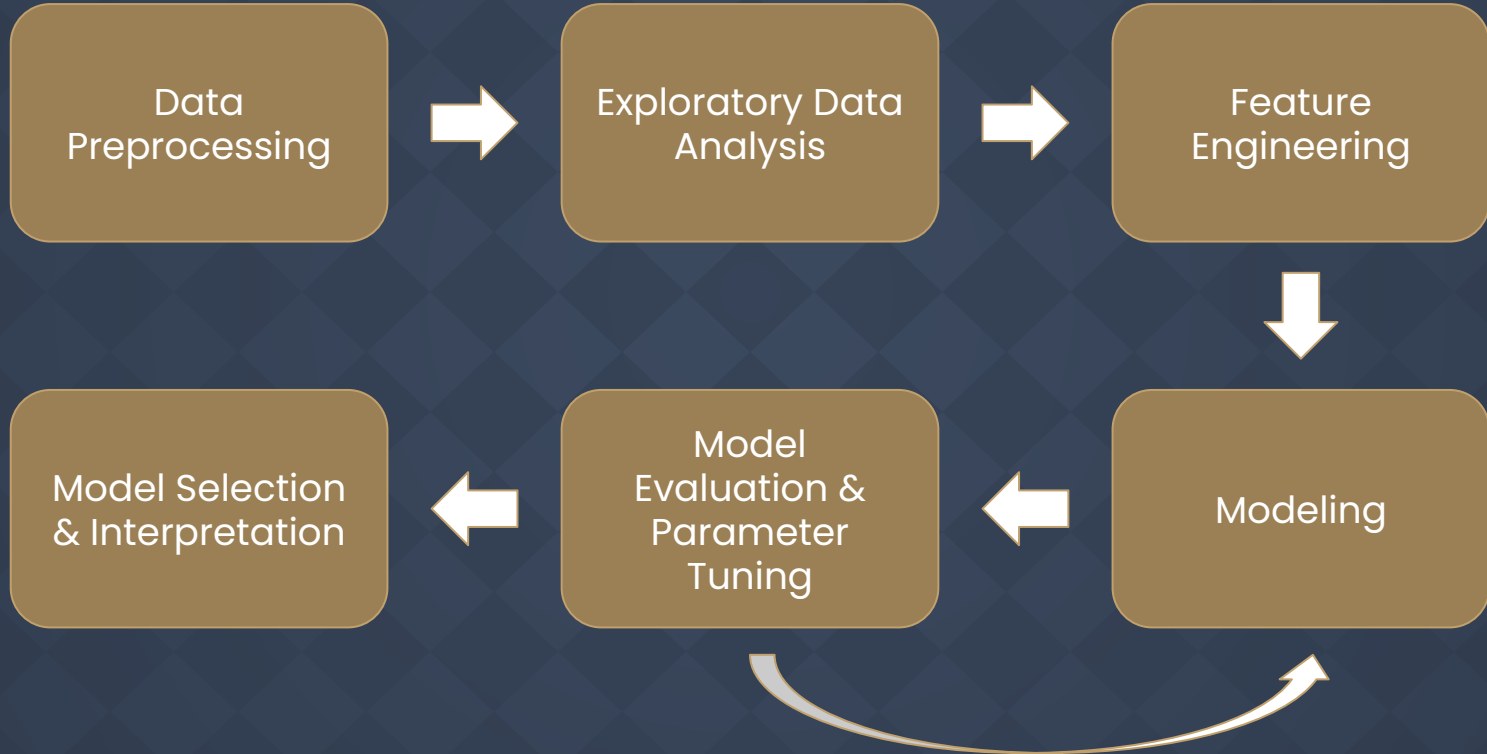
Data Name	Unit	Example
Date	DateTime	YYYY-MM-DD
Gold Price	US Dollar per Ounce	1234.12 USD per ounce
DowJones	US Dollar	12583.12
Nasdaq	US Dollar	3156.12
Crude Oil WTI	US Dollar / Barrel	69.12
Fed Fund (FF)	Percentage	8%
S&P500	US Dollar	1491.65
XAU Index (Gold Mine)	US Dollar	71.60
Unemployment Rate	Percentage	8%

# Data Dictionary (Exchange Rate)

Data Name	Unit	Exchange Rate (From Other Currency to 1 dollar)
USD-THB	Thai Baht	32.65 THB
USD- CHY (Chinese Yuan)	Chinese Yuan	23.45 Chinese Yuan
USD-AUD (Australia)	Australian Dollar	6.56 Australian Dollar
USD-RUB (Russia)	Russian Ruble	30.56 Russian Ruble



# Workflow



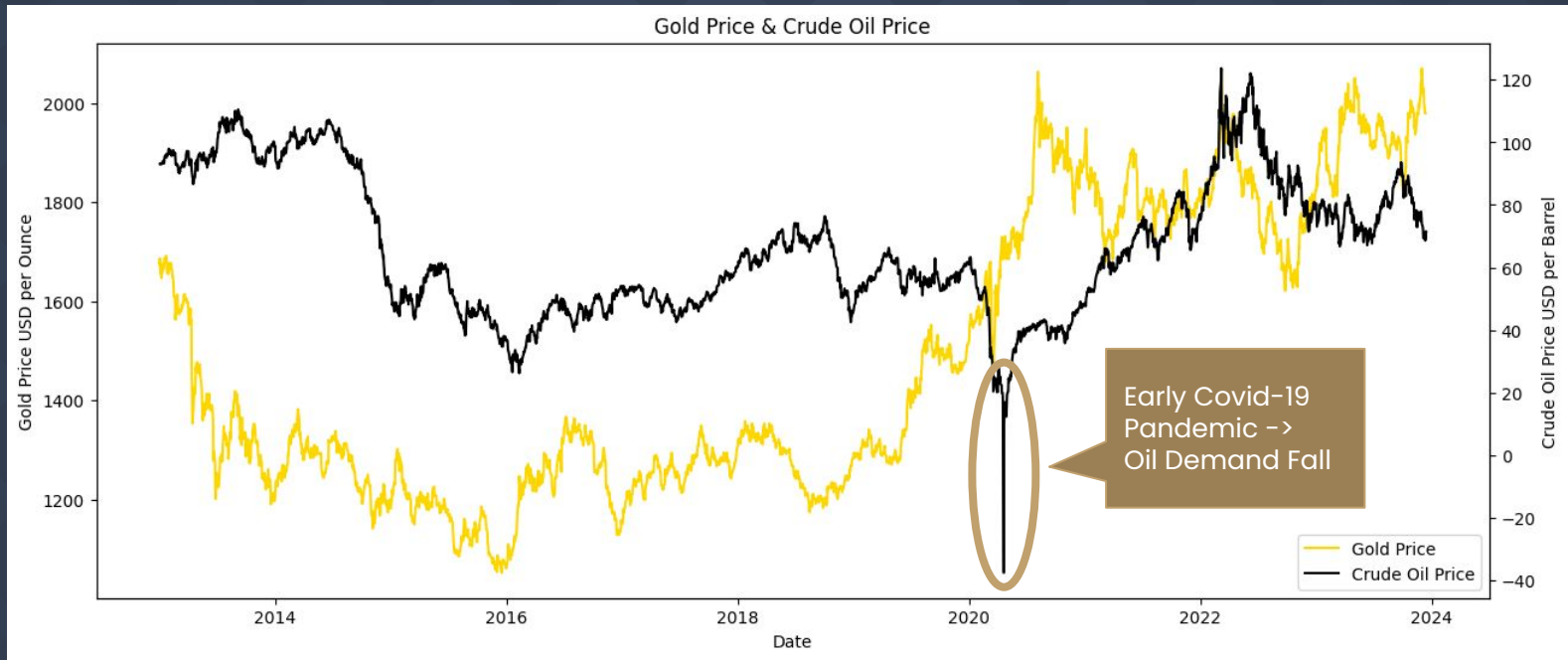
# Data Preprocessing

- Merge data set
- Drop some missing value on daily dataset
  - From 2851 rows to 2745 rows
- Weekly Aggregation



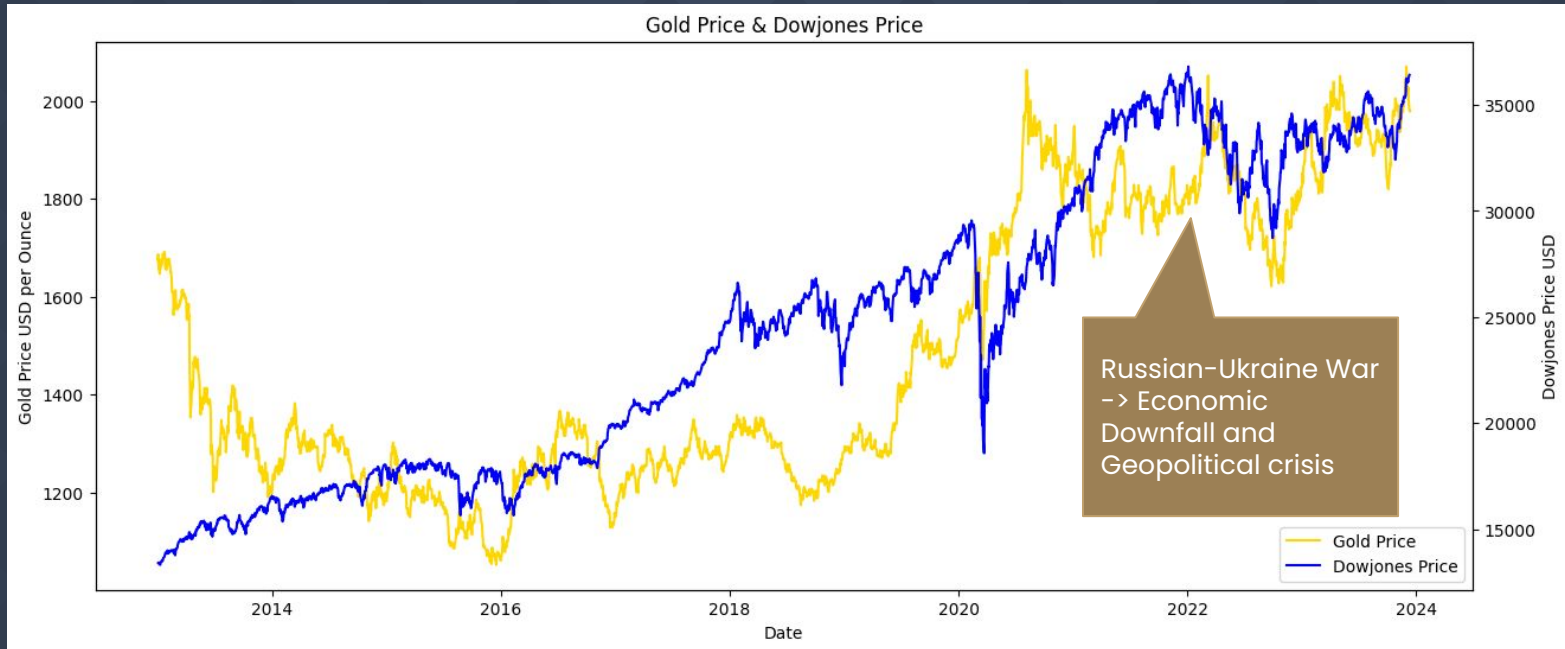
# Crude Oil & Gold Price

- Gold Price and crude oil price tend to move in opposite directions



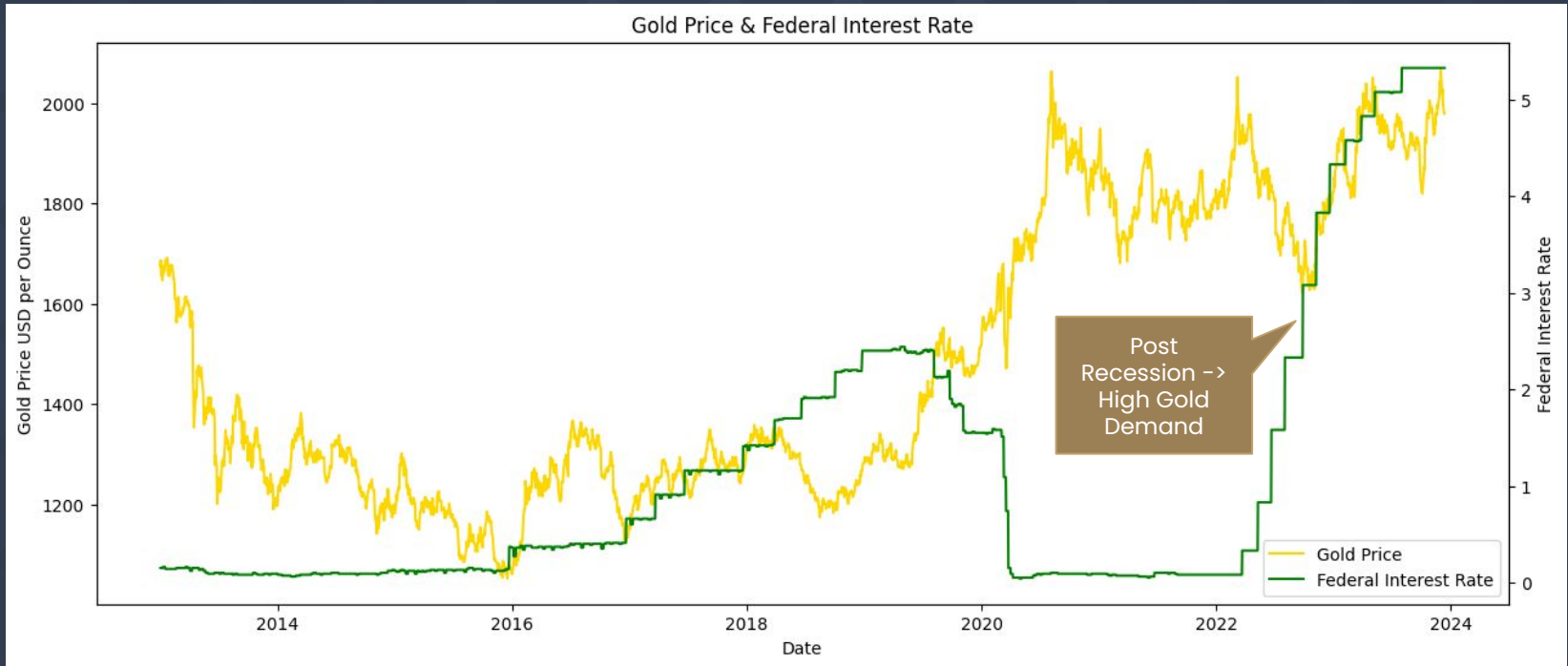
# Gold Price & Stock Exchange (Dowjones)

- Both gold price and stock price rise gradually since 2016



# Gold Price & Federal Fund

- Gold Price and Fed Fund rise gradually, except for Covid-19 Period
- Central bank always set Federal Interest rate 8 times a year (Ladder line trend)



# Feature Selection

- From the result of EDA
  - Crude Oil price
  - Federal Fund
  - Stock Index
    - Dow Jones, NASDAQ, S&P500
  - Exchange Rate
    - USD-THB
    - USD-CNY
    - USD-AUD
    - USD-RUB

# Feature Engineering

Independent Variable (previous week):

- Gold Price
- Crude Oil
- Federal Interest Rate
- XAU Index
- Stock Index
  - Dow Jones
  - Nasdaq
  - S&P500
- Exchange Rate
  - USD-CNY
  - USD-AUD
  - USD-RUB
  - USD-THB

Target Variable:

- Whether Gold price of next week rises or not  
(Binary Classification Problem)
  - Rise: 50 & Fall: 50 (almost balanced data)

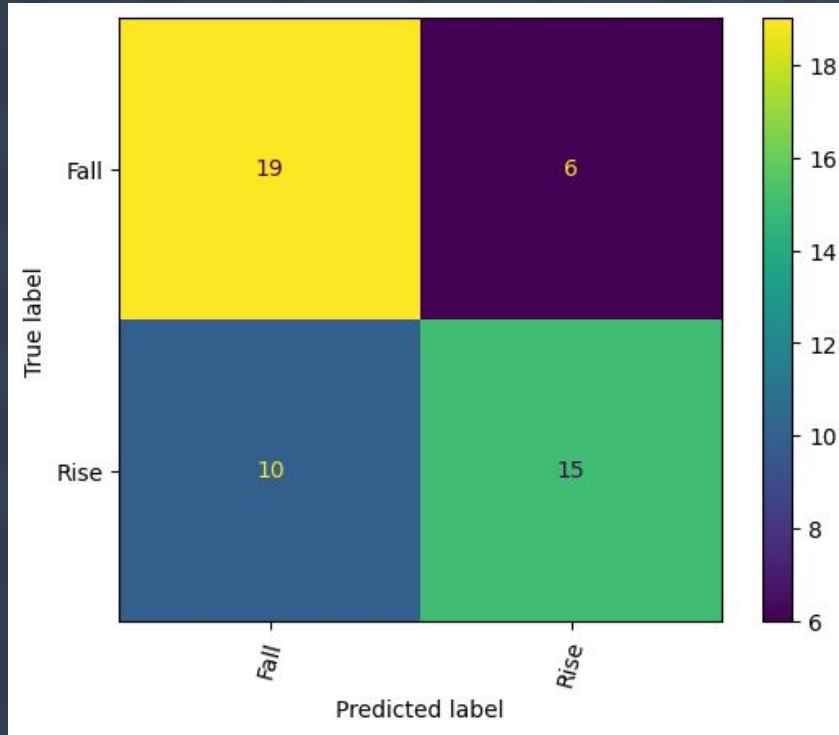
# Model Selection

- Train Data: 2013 - 2022
- Test Data: 2023

Model Name	Accuracy
Logistic Regression	0.64
Random Forest	0.62
XGBoost	0.68
XGBoost (with hyper Parameter)	0.62



# Best Model – XGBoost



- True Negative: 19
- False Negative: 10
- False Positive: 6
- True Positive: 15 is actual rise predicted correctly

# Summary

	Importance
xau_close_rise	0.132751
usdcny_close_rise	0.112717
crudeoil_close_rise	0.099885
dowjones_close_rise	0.099767
usdthb_close_rise	0.097123
sp500_close_rise	0.085091
ff_close_rise	0.081180
usdrussia_close_rise	0.077359
nasdaq_close_rise	0.076902
usdaud_close_rise	0.072678
gold_close_rise	0.064546

- Baseline: 0.50
- Model for prediction – XGBoost
  - Accuracy: 0.68
- Feature Importance – Using XGBoost
  - Xau close price: 0.13
  - USD – CNY: 0.11
  - Crude oil: 0.099

# Recommendation

- These features can estimate gold price movement
  - Investors should spend time on News
    - Global events
  - Regularly monitor on market indicators
    - US Dollar Strength
    - Stock Market
- Future Forecast – 2024 Prediction
  - Upfront: Week, Quarter, Year

# Limitation & Further Works

- Regression Prediction
  - This model only predicts direction of gold price (fall or rise)
  - Gold holders (who has gold) would like to know **value** of gold in future, which this classification model cannot predict
- Data Insufficiency (because not daily)
- COVID-19 situation can cause extraordinary movement in many economic indices
- Time Series Analysis
  - Some Time Series models may be available
    - ARIMA
    - SARIMAX
    - Prophet
    - Recurrent Neural Network (e.g. LSTM)

**THANK  
YOU**