GOLD PRICE PREDICTIONS

SUBMITTED BY:
KANAK (B210049EC)
AVANTASH RANJAN (B210076EC)
SONU KUMAR BHAGAT (B210066EC)

Contents:

Introduction:

- Predictive analytics includes many of statistical techniques from predictive modeling and machine learning, data mining, to analyze current and historical facts to make predictions about future or otherwise unknown events.
- Generally speaking, gold and equity have an inverse relationship. This means, that when the gold price goes up, prices in the stock market will fall and vice versa. Historically it has been observed that when the stock market is most pessimistic, gold performs very well.
- Predictive modeling is the general concept of building a model that is capable of making predictions

Problem Statement:

- Gold is a precious yellow good it use as money. The gold prices are time series data of gold prices fixed. Factors influencing gold prices are many and we have to be selective in this study to ensure that the model developed is significant.
- In this study, we proposed the development for predicting future gold price using Multiple Linear Regression, Decision Tree, Random Forest, K – Nearest Neighbour (KNN). The data used in this part are the Gold Prices from the Kaggle.

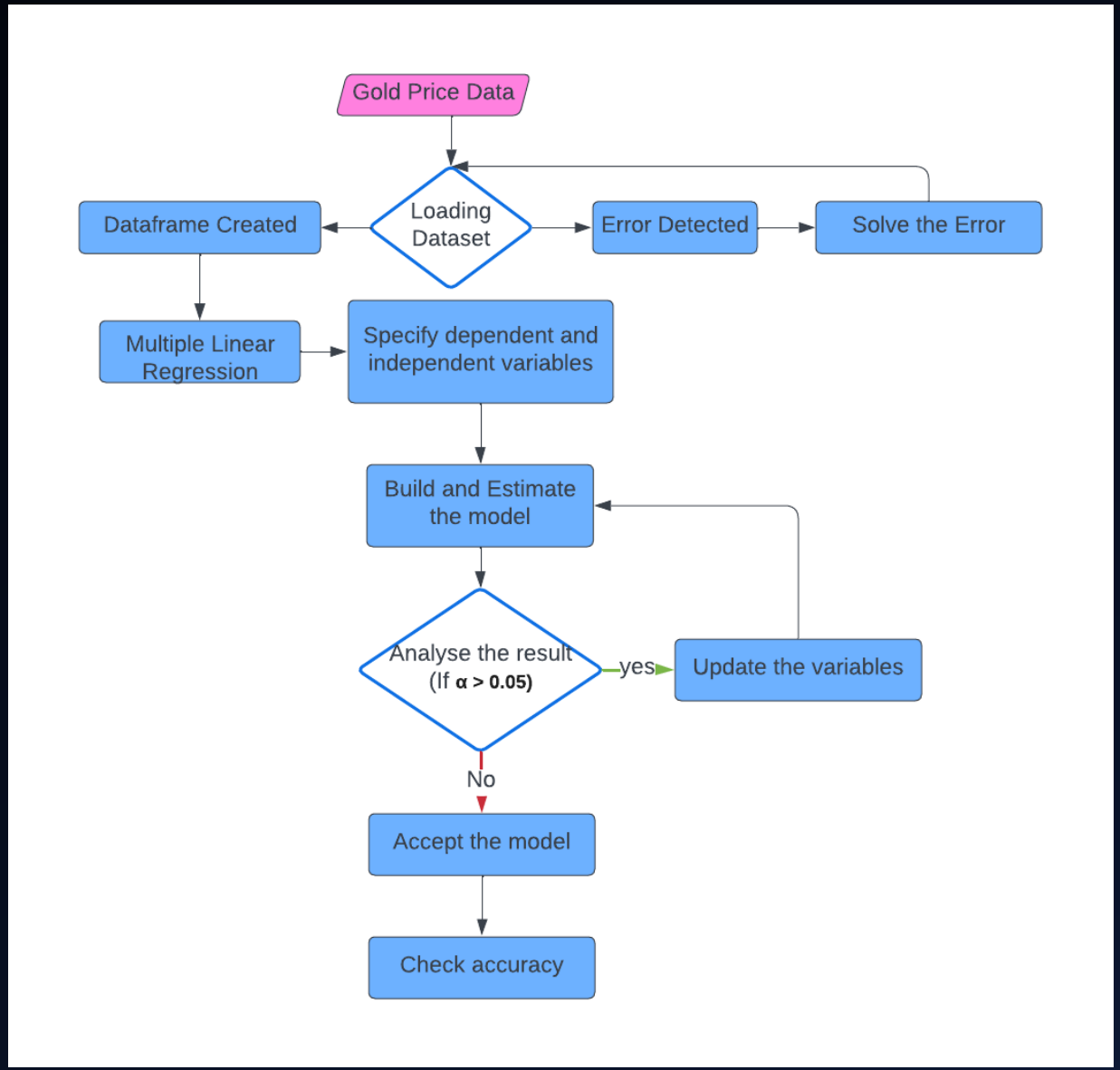
Algorithm:

- Multiple Linear Regression
- Decision Tree
- Random Forest
- KNN

	Date	SPX	GLD	USO	SLV	EUR/USD
1	1/2/2008	1447.160034	84.860001	78.470001	15.18	1.471692
2	1/3/2008	1447.160034	85.57	78.370003	15.285	1.474491
3	1/4/2008	1411.630005	85.129997	77.309998	15.167	1.475492
4	1/7/2008	1416.180054	84.769997	75.5	15.053	1.468299
5	1/8/2008	1390.189941	86.779999	76.059998	15.59	1.557099
6	1/9/2008	1409.130005	86.550003	75.25	15.52	1.466405
7	1/10/2008	1420.329956	88.25	74.019997	16.061001	1.4801
8	1/11/2008	1401.02002	88.580002	73.089996	16.077	1.479006
9	1/14/2008	1416.25	89.540001	74.25	16.280001	1.4869
10	1/15/2008	1380.949951	87.989998	72.779999	15.834	1.48021
11	1/16/2008	1373.199951	86.699997	71.849998	15.654	1.466405
12	1/17/2008	1333.25	86.5	71.029999	15.717	1.464
13	1/18/2008	1325.189941	87.419998	71.540001	16.030001	1.461796
14	1/22/2008	1310.5	88.169998	70.550003	15.902	1.464794
15	1/23/2008	1338.599976	87.889999	69.5	15.9	1.463208
16	1/24/2008	1352.069946	90.080002	70.93	16.299999	1.47741
17	1/25/2008	1330.609985	90.300003	71.910004	16.298	1.467502
18	1/28/2008	1353.959961	91.75	72.349998	16.549999	1.478809

MULTIPLE LINEAR REGRESSION:

- Multiple linear regression (MLR) is a statistical method used to model the relationship between the gold price (dependent variable) and predictors (multiple independent variables).
- In gold price prediction, MLR can be useful for identifying which factors have the most influence on the price of gold and for creating a predictive model based on those factors.



Contd….

- **Data Collection**: Collect data on the historical gold prices and potential predictors. Predictors might include variables such as: Stock market indices (S&P), Commodity prices (Silver prices), Supply and demand data for gold, Other market-related data (USD).
- **Model Building:** Create a multiple linear regression model using the chosen features. The model takes the form: $Y=\beta0+\beta1X1+\beta2X2+\beta3X3+\epsilon$

Where: Y is the gold price

X1, X2, X3 are the predictors.
β0, β1, β3 are the regression coefficients.

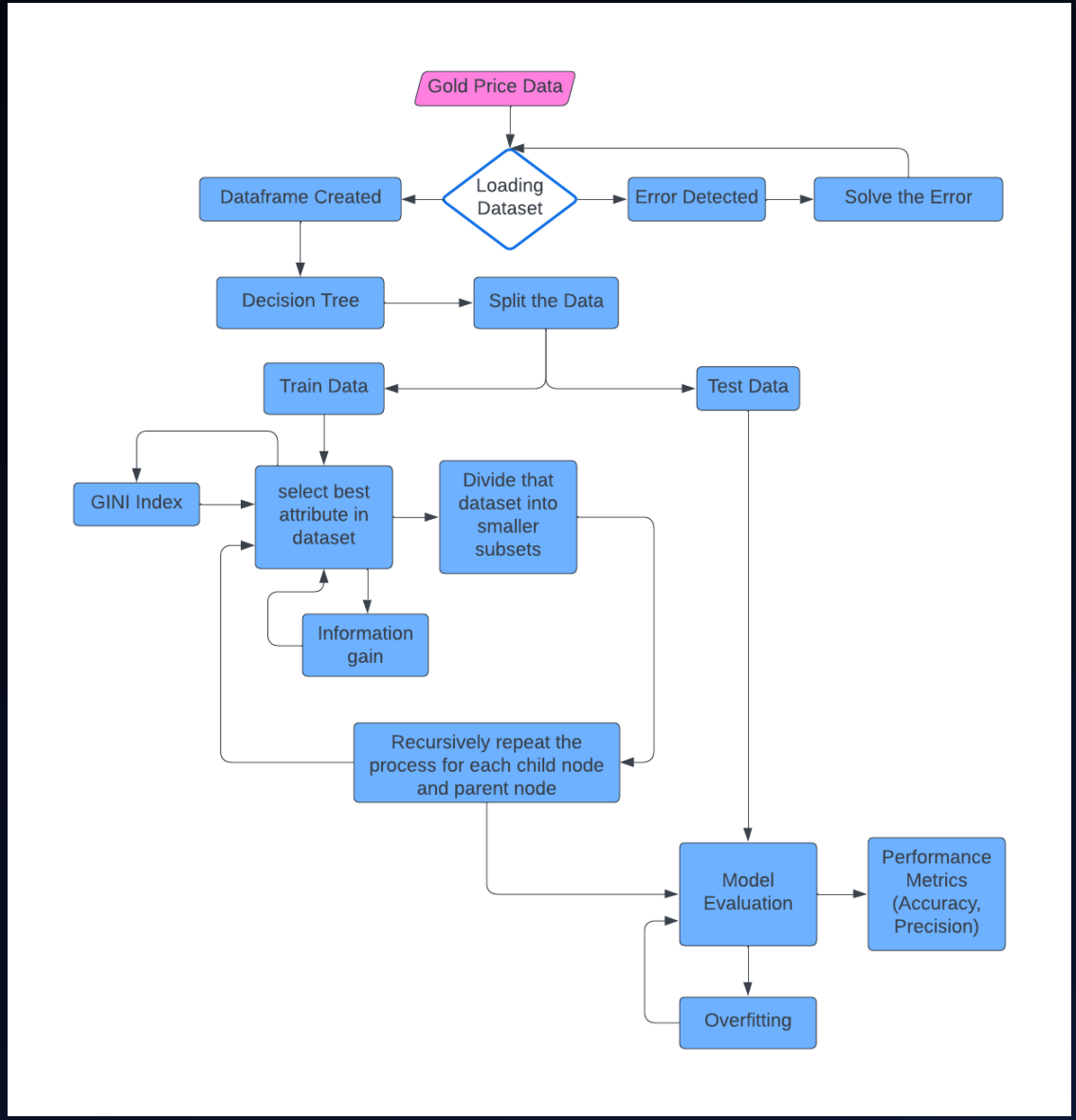
ϵ is the error term.

Contd…

- **Model Training:** Train the model using a training dataset. This involves finding the optimal regression coefficients (β0,β1,β3) that minimize the error between predicted and actual values of gold price.
- **Model Evaluation:** Assess the model's performance using a testing dataset. Common evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.
- **Predicting Gold Prices :** Once the model has been trained and evaluated, it can be used to make predictions on new data.

DECISION TREE:

- Decision trees can be used for gold price prediction. They work by recursively partitioning the data into smaller subsets based on conditions on the input features, eventually leading to a prediction at each leaf node.
- It can be a useful tool for predicting gold prices due to their ability to model non-linear relationships and handle a variety of features.



Contd…

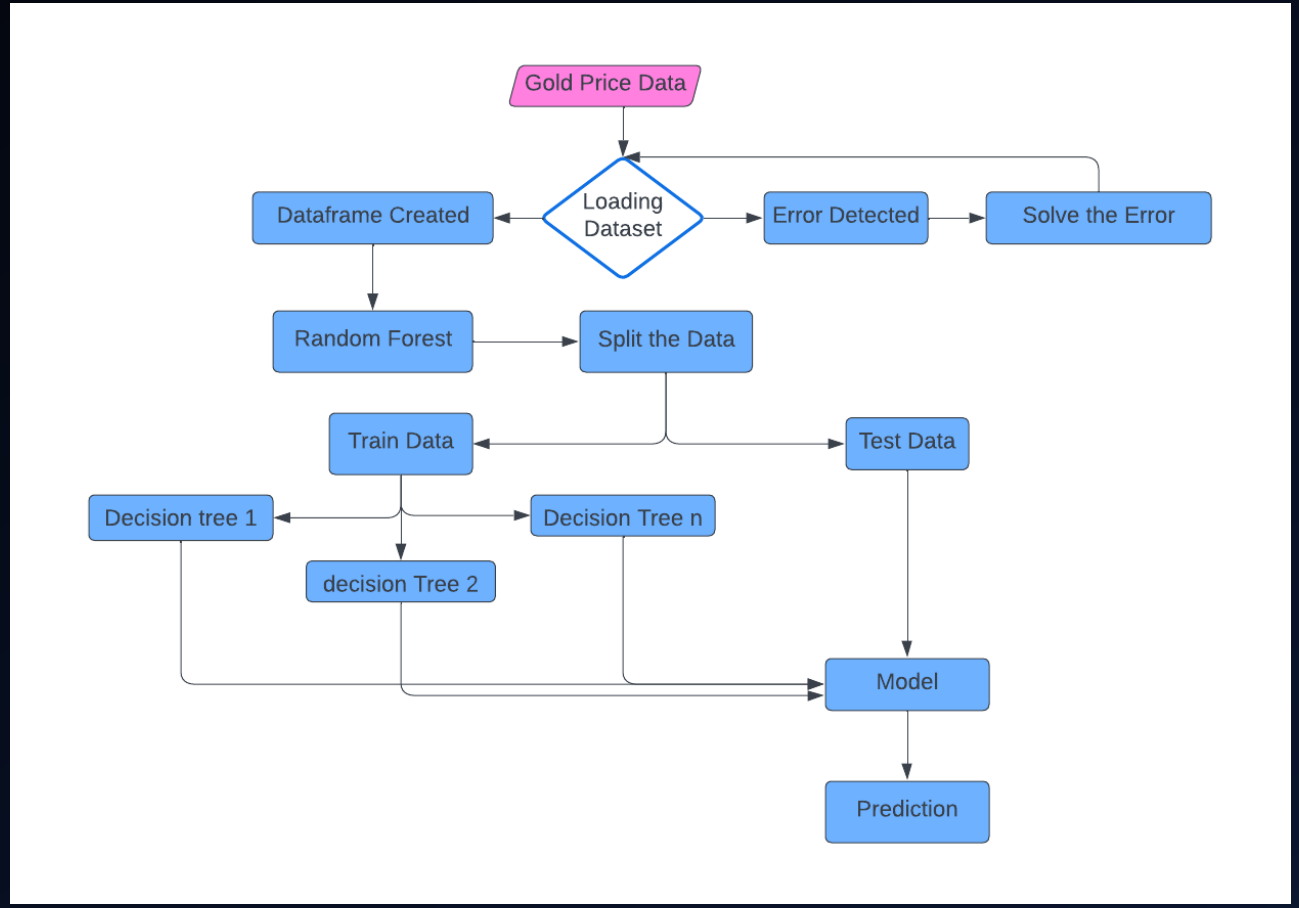
- **Data Collection**: Collect data on the historical gold prices and potential predictors. Predictors might include variables such as: Stock market indices (S&P), Commodity prices (Silver prices), Supply and demand data for gold, Other market-related data (USD).
- **Splitting the Data:** Split the data into training and test sets. This allows you to train the decision tree on the training set and evaluate its performance on the test set.
- **Model Training:** Use the training data to train the decision tree. The decision tree algorithm will create a model that can be used to predict future gold prices based on the features.

Contd…

- **Model Evaluation**: Evaluate the model's performance using metrics such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE). This will help you understand how well the model is performing.
- **Predicting Gold Prices** : Once you are satisfied with the model's performance, you can deploy it for gold price prediction in a real-world setting.

RANDOM FOREST:

- Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression.
- It operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.



Contd…

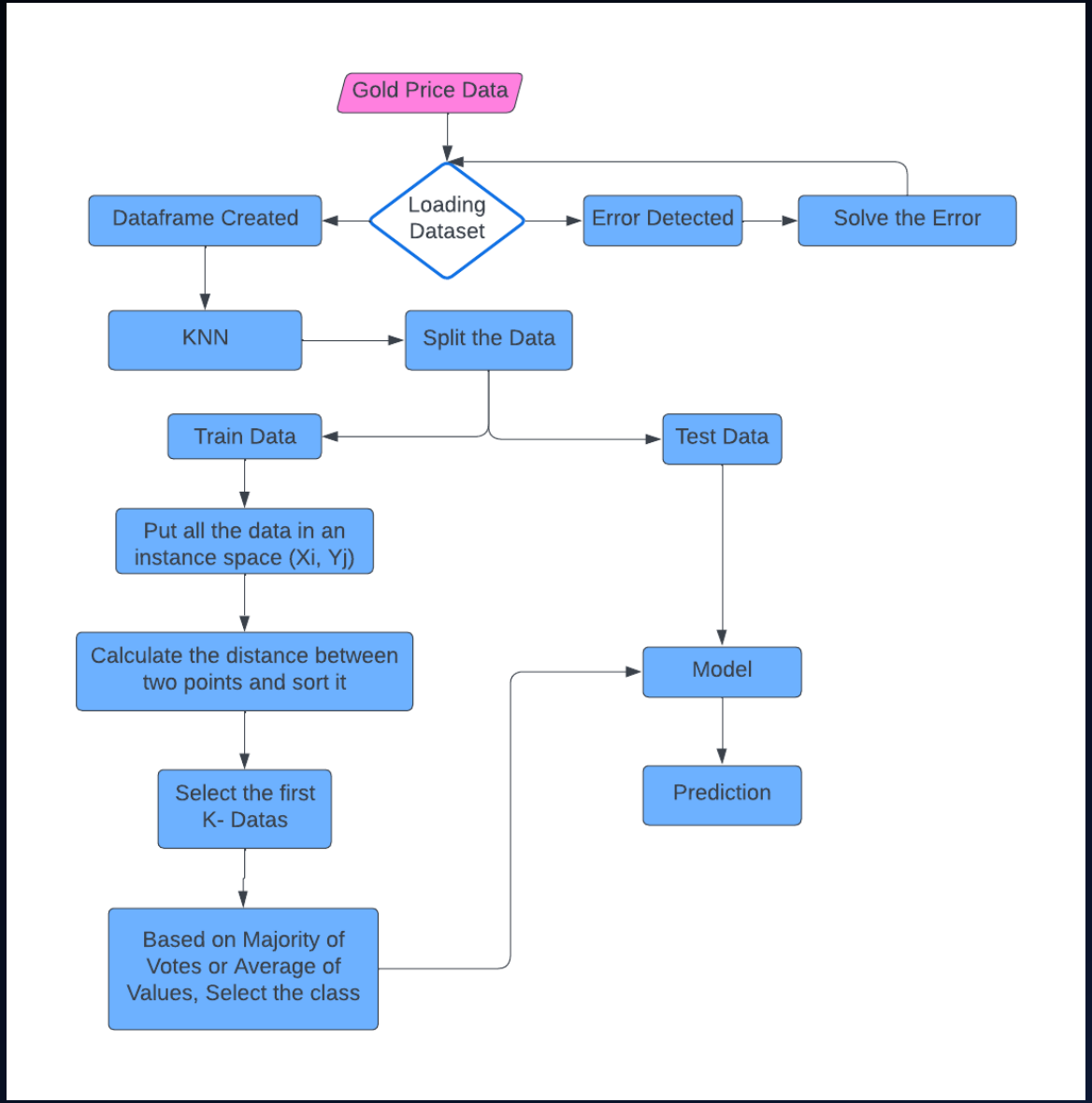
- **Data Collection**: Collect data on the historical gold prices and potential predictors. Predictors might include variables such as: Stock market indices (S&P), Commodity prices (Silver prices), Supply and demand data for gold, Other market-related data (USD).
- **Splitting the Data:** Split the data into training and test sets. This allows you to train the decision tree on the training set and evaluate its performance on the test set.
- **Random Forest Regressor**: Use the Random Forest Regressor model from libraries such as scikit-learn in Python. This algorithm is a powerful and flexible ensemble learning method that consists of multiple decision trees.

Contd...

- **Model Training:** Train the Random Forest model on your training set. Monitoring the training process and check for overfitting.
- **Model Evaluation**: Evaluate the model's performance using metrics such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE). This will help you understand how well the model is performing.
- **Predicting Gold Prices** : Once we are satisfied with the model's performance, we can deploy it for gold price prediction in a real-world setting.

KNN:

- KNN is a distance-based model, so the choice of distance metric (e.g., Euclidean distance) and how we weigh neighbors can have a significant impact on the predictions. It's also sensitive to outliers, so proper preprocessing and feature selection are key.
- In this project, KNN can be applied as a regression model to forecast the future price of gold based on historical data and other relevant features.



Contd…

- **Data Collection**: Collect data on the historical gold prices and potential predictors. Predictors might include variables such as: Stock market indices (S&P), Commodity prices (Silver prices), Supply and demand data for gold, Other market-related data (USD).
- **Splitting the Data:** Split the data into training and test sets. This allows you to train the decision tree on the training set and evaluate its performance on the test set.
- **Model Selection**: Choose KNN as the regression model. You need to decide on the number of neighbors (k) to use. This is a crucial hyperparameter that we can tune.

Contd…

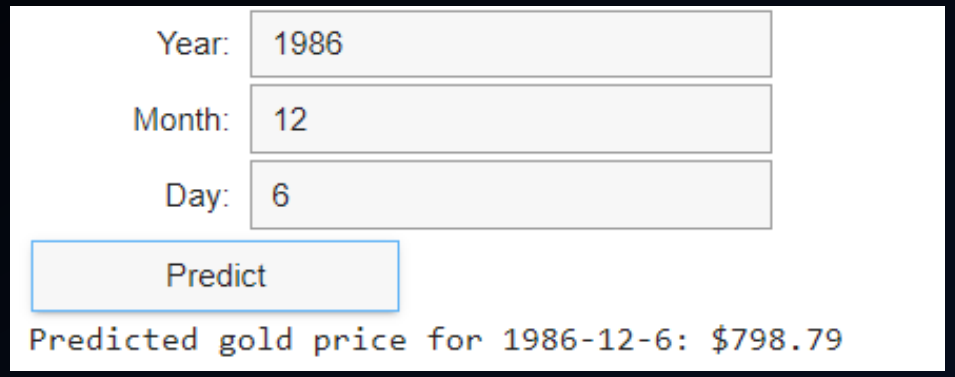
- **Model Training**: Train the KNN regression model using the training dataset.
- **Model Evaluation**: Evaluate the model's performance using metrics such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE). This will help you understand how well the model is performing.
- **Predicting Gold Prices**: Once we are satisfied with the model's performance, we can deploy it for gold price prediction in a real-world setting.

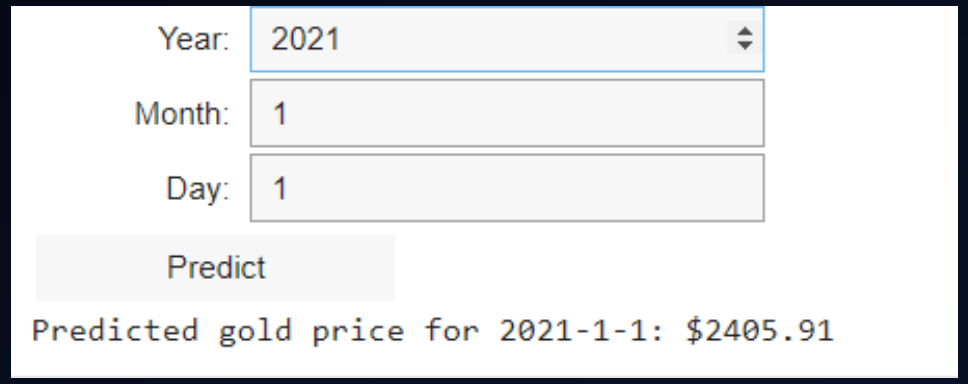
Results:

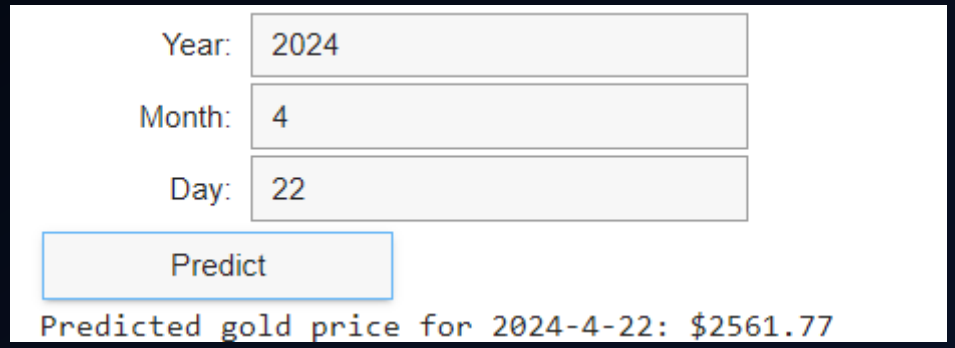
Algorithms	R2_Score	Mean Square Error
Multiple Linear Regression (MLR)	0.8975640982991402	56.16559421500603
Decision Tree	0.9831109316901961	9.260274392156793
Random Forest	0.9892766469309671	5.655962881475242
KNN	0.9941851307615867	3.188292196731859

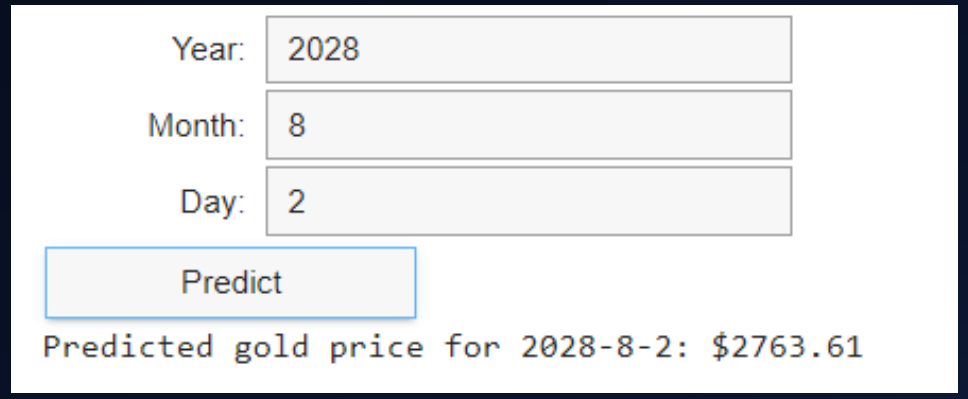
- From above table, KNN is the best choice for Gold Price Prediction.

Test Output :


Year: 1986
Month: 12
Day: 6
Predict
Predicted gold price for 1986-12-6: $798.79


Year: 2021
Month: 1
Day: 1
Predict
Predicted gold price for 2021-1-1: $2405.91


Year: 2024
Month: 4
Day: 22
Predict
Predicted gold price for 2024-4-22: $2561.77


Year: 2028
Month: 8
Day: 2
Predict
Predicted gold price for 2028-8-2: $2763.61

Conclusion:

- MLR might be a good choice for understanding linear relationships but may not be the best option for Gold Price Prediction as it has non-linear relationships.
- Decision Tree can be useful for capturing complex decision boundaries and visualizing the decision-making process. However, overfitting can be a concern, so techniques like pruning may be needed.
- Random Forest tends to perform well in practice for gold price prediction due to its ability to handle complex relationships and interactions in the data. It often provides better predictive performance than individual decision trees.
- KNN can be a good choice for Gold Price Prediction if simplicity and interpretability are desired. However, its performance may be impacted by the choice of k and distance metric, as well as the presence of outliers.

References:

- https://www.lucidchart.com
- https://www.slideshare.net
- https://www.analyticsvidhya.com
- https://link.springer.com/article/10.1007/s44196-023-00190-0
- https://www.kaggle.com/

Thank You!

FEEL FREE TO ASK QUESTIONS.