

A PROJECT REPORT

ON

“GOLD PRICE PREDICTION”

SUBMITTED TO

KIIT DEEMED TO BE UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIRMENT FOR THE AWARD OF

BACHELOR’S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

BY

| | |
|-----------------------|----------------|
| TANAY PAUL | 1605326 |
| SHUBHAM KUAMR | 1605231 |
| KOUSTUV SARKAR | 1605281 |
| SAURAV DAS | 1605304 |

**UNDER THE GUIDANCE OF
DR PRASANT KUMAR PATTNAIK**



**SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024
MARCH 2020**

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024



CERTIFICATE

This is certify that the project entitled
“GOLD PRICE PREDICTION”

Submitted by

| | |
|-----------------------|----------------|
| TANAY PAUL | 1605326 |
| SHUBHAM KUMAR | 1605231 |
| KOUSTUV SARKAR | 1605281 |
| SAURAV DAS | 1605304 |

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of **Bachelor of Technology (Computer Science & Engineering Or Information Technology)** at **KIIT Deemed to be University**, Bhubaneswar. This work is done during year **2020-2021**, under our guidance.

Date: **29/03/2020**

DR PRASANT KUMAR PATTNAIK
Project Guide

ACKNOWLEDGEMENTS

We are profoundly grateful to **DR PRASANT KUMAR PATTNAIK** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

TANAY PAUL (1605326)
SUBHAM KUMAR(1605231)
KOUSTUV SARKAR(1605281)
SAURAV DAS(1605304)

ABSTRACT

A complexity of business dynamics often forces decision-makers to make decisions based on subjective mental models, reflecting their experience. However, research has shown that companies perform better when they apply data-driven decision-making. This creates an incentive to introduce intelligent, data-based decision models, which are comprehensive and support the interactive evaluation of decision options necessary for the business environment. Recently, a new general explanation methodology has been proposed, which supports the explanation of state-of-the-art black-box prediction models. Uniform explanations are generated on the level of model/individual instance and support what-if analysis. We present a novel use of this methodology inside an intelligent system in a real-world case of business to-business (B2B) sales forecasting, a complex task frequently done judgmentally. Users can validate their assumptions with the presented explanations and test their hypotheses using the presented what-if parallel graph representation. A separation of the machine learning model selection from model explanation is another significant benefit for expert and intelligent systems. Explanations unconnected to a particular prediction model positively influence acceptance of new and complex models in the business environment through their easy assessment and switching.

Machine learning has often been applied to the prediction of financial variables, but usually with a focus on stock prediction rather than commodities. In my project, we chose to apply supervised learning to the prediction of gold prices in order to see what kind of success we could achieve. For the purposes of profitability, it is more important to predict the relative change in price tomorrow (i.e. whether the price will go up or down) than to predict the absolute price tomorrow, so we have formulated this as a classification problem: given historical price data up to a given day, my algorithm attempts to predict whether the gold price tomorrow will be higher or lower than it is today.

The use of **LINEAR REGRESSION MODEL** has been used to predict the future prices of the prices of gold in the future. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

If the goal is prediction, forecasting, or error reduction, clarification needed linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response. If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response. Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares cost

function as in ridge regression (L^2 -norm penalty) and lasso (L^1 -norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms "least squares" and "linear model" are closely linked, they are not synonymous.

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). Hence, the name is Linear Regression.

CONTENTS

| CHAPTERS | PAGE NO |
|-------------------------------------|----------------|
| 1.INTRODUCTION | 8 |
| 2.LITERATURE REVIEW | 10 |
| 3.TECHNOLOGY USED | 12 |
| 4.PROPOSED MODEL AND WORKING | 15 |
| 5.IMPLEMENTATION AND RESULTS | 19 |
| 6.CONCLUSION | 21 |

CHAPTER 1

1.1 INTRODUCTION

Data-driven decision-making is the practice of basing decisions on the data analysis, rather than on intuition (Provost and Fawcett, 2013). In the use of data-driven decision-making, companies in the top third of their industry are, on average, 5% more productive and 6% more profitable than their competitors (McAfee and Brynjolfsson, 2012; Brynjolfsson et al., 2011). The research into the acceptance of decision support systems (DSS) shows that users are more likely to adhere to recommendations when an explanation facility is available (Arnold et al., 2006; Gönül et al., 2006). However, the top performing, non-transparent black-box machine learning (ML) models, such as random forests, boosting, support vector machines (SVM), and neural networks achieve significantly better predictive performance than simple, interpretable models such as decision trees, naïve Bayes, or decision rules (Caruana and Niculescu-Mizil, 2006). This is one of the reasons for low usage and acceptance of predictive ML models in areas where transparency and comprehensibility of decisions are required.

Machine learning has often been applied to the prediction of financial variables, but usually with a focus on stock prediction rather than commodities. In our project, we chose to apply supervised learning to the prediction of gold prices in order to see what kind of success we could achieve. For the purposes of profitability, it is more important to predict the relative change in price tomorrow (i.e. whether the price will go up or down) than to predict the absolute price tomorrow, so we have formulated this as a classification problem: given historical price data up to a given day, our algorithm attempts to predict whether the gold price tomorrow will be higher or lower than it is today.

The spot price is the current market price at which commodity is purchased or sold for immediate payment and delivery. It is differentiated from the futures price, which is the price at which the two parties agree to transact on future date. Gold spot rates are decided twice a day based on supply and demand in gold market. Fractional change in gold price may result in huge profit or loss for these

investors as well as government banks. Forecasting rise and decline in the daily gold rates, can help investors to decide when to buy (or sell) the commodity. Various studies have been conducted by researchers to forecast gold rates, each of them insightful in their own right. We in this study forecast gold rates using a) the most comprehensive set of features than any of the previous studies, which for the first time includes the performance indicators of Russian, Chinese, and Indian economies (as they are the biggest purchaser of gold) and as well as the stock price of leading gold producing/trading companies, and b) apply various machine learning algorithms for forecasting and compare their results. We also identify which attributes influence the gold rates the most, some of which were not even used before.

CHAPTER 2

2.1 LITERATURE REVIEW

Historically, gold had been used as a form of currency in various parts of the world including USA [5]. In recent times also, gold has maintained its value and has been used as a means for assessing the financial strength of a country. Big investors have also been attracted to this precious metal and invested huge amounts in it. Recently, emerging world economies, such as China, Russia, and India have been big buyers of gold, whereas USA, South Africa, and Australia are among the big seller of this commodity [8]. Chinese and Indian traditional events also affect the price of the gold. In that time more money is poured for purchase of this commodity. Small investors also find this commodity for safe investment rather than alternate investment options, which bear in-built investment risks. Internal financial conditions of the aforementioned countries play a vital role for setting spot rates for gold. Governmental investments in gold are largely decided by their financial conditions, and interest rates, as they are indicators of the strength of their economy. When US interest rates become lower, more economic activity is witnessed in US, thus capital inflows in gold market are observed. Similarly, when interest rates lowered in China from 5.31 (2010) to 4.35 (2016), it bought gold aggressively [8]. Global investors, either countries or giant companies, tend to invest elsewhere if they foresee a significant decline in gold prices. In such a scenario, some investors turn to some other form of investment, such as US Bonds, or stock exchange. Fig. 1 shows that New York Stock Exchange (NYSE) and S&P 500 tend to do better when gold rates are at their low. S&P 500 is an American stock market index based on the market capitalizations of 500 large companies listed on the NYSE or NASDAQ. It implies that capital flow was noticed from stock exchanges to gold market. On the other hand, some stakeholders convert their gold reserves to USD, therefore EuroUSD index (exchange rate from Euro to USD) tends to rise with the decline in gold rates. Value of USD itself is dependent on various factors including the interest rates decided by US Government. Performance of the leading stock exchanges such as NASDAQ, and Dow Jones also reflects the strength of the US economy. Therefore, various phenomena are interconnected with gold rates and affect the price also. The spot price is the current market price at which commodity is purchased or sold for immediate payment and delivery. It is differentiated from the futures price, which is the price at which the two parties agree to

transact on future date. Gold spot rates are decided twice a day based on supply and demand in gold market. Fractional change in gold price may result in huge profit or loss for these investors as well as government banks. Forecasting rise and decline in the daily gold rates, can help investors to decide when to buy (or sell) the commodity. Various studies have been conducted by researchers to forecast gold rates, each of them insightful in their own right. We in this study forecast gold rates using a) the most comprehensive set of features than any of the previous studies, which for the first time includes the performance indicators of Russian, Chinese, and Indian economies (as they are the biggest purchaser of gold) and as well as the stock price of leading gold producing/trading companies, and b) apply various machine learning algorithms for forecasting and compare their results. We also identify which attributes influence the gold rates the most, some of which were not even used before.

CHAPTER 3

3.1 TECHNOLOGY USED

MACHINE LEARNING

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

3.2 PYTHON

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released in 2008, was a major revision of the language that is not completely backward-compatible,

and much Python 2 code does not run unmodified on Python 3. The Python 2 language, i.e. Python 2.7.x, is "sunsetting" in less than a month on January 1, 2020 (after extension; first planned for 2015), and the Python team of volunteers will not fix security issues, or improve it in other ways after that date. With the end-of-life, only Python 3.5.x and later will be supported.

3.3 JUPYTER NOTEBOOK

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension.

A Jupyter Notebook can be converted to a number of open standard output formats (HTML, presentation slides, LaTeX, PDF, ReStructuredText, Markdown, Python) through "Download As" in the web interface, via the nbconvert library or "jupyter nbconvert" command line interface in a shell. To simplify visualisation of Jupyter notebook documents on the web, the nbconvert library is provided as a service through NbViewer which can take a URL to any publicly available notebook document, convert it to HTML on the fly and display it to the user.

Jupyter Notebook can connect to many kernels to allow programming in many languages. By default Jupyter Notebook ships with the IPython kernel. As of the 2.3 release (October 2014), there are currently 49 Jupyter-compatible kernels for many programming languages, including Python, R, Julia and Haskell.

The Notebook interface was added to IPython in the 0.12 release (December 2011), renamed to Jupyter notebook in 2015 (IPython 4.0 – Jupyter 1.0). Jupyter Notebook is similar to the notebook interface of other programs such as Maple, Mathematica, and SageMath, a computational interface style that originated with Mathematica in the 1980s. According to *The Atlantic*, Jupyter interest overtook the popularity of the Mathematica notebook interface in early 2018.

3.4 ANACONDA (Python Distribution)

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system *conda*. The Anaconda distribution includes data-science packages suitable for Windows, Linux, and MacOS.

Anaconda distribution comes with 1,500 packages selected from PyPI as well as the *conda* package and virtual environment manager. It also includes a GUI, **Anaconda Navigator**, as a graphical alternative to the command line interface (CLI). The big difference between *conda* and the **pip** package manager is in how package dependencies are managed, which is a significant challenge for Python data science and the reason *conda* exists. When *pip* installs a package, it automatically installs any dependent Python packages without checking if these conflict with previously installed packages. It will install a package and any of its dependencies regardless of the state of the existing installation. Because of this, a user with a working installation of, for example, Google Tensorflow, can find that it stops working having used *pip* to install a different package that requires a different version of the dependent *numpy* library than the one used by **Tensorflow**. In some cases, the package may appear to work but produce different results in detail. In contrast, *conda* analyses the current environment including everything currently installed, and, together with any version limitations specified (e.g. the user may wish to have Tensorflow version 2.0 or higher), works out how to install a compatible set of dependencies, and shows a warning if this cannot be done.

CHAPTER 4

4.1 PROPOSED MODEL AND WORKING

In this project we mainly predict the price of the ETF GLD (Exchange Trade Fund GOLD) price of the gold with the best accuracy using **MACHINE LEARNING** in PYTHON using a few **MACHINE LEARNING** models like “**LINEAR REGRESSION MODEL**” . We will discuss about the **LINEAR REGRESSION MODEL**

LINEAR REGRESSION MODEL

Here is the formal definition, “Linear Regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X ”

Let me explain the concept of regression in a very basic manner, so imagine that you run a company that builds cars and you want to understand how the change in prices of raw materials (let's say Steel) will affect the sales of the car. The general understanding is this, the rise in the price of steel will lead to a rise in the price of the car resulting in lesser demand and in turn lesser sales. But how do we quantify this? And how do we predict how much change in sales will happen based on the degree of change in steel price. That's when the regression comes into the picture.

In our project we have used the **LINEAR REGRESSION MODEL**

The equation used in this model

$$y = \theta_1 + \theta_2 \cdot x$$

IMAGE 1

While training the model we are given :

x: input training data (univariate – one input variable(parameter))

y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ_1 and θ_2 values.

θ_1 : intercept

θ_2 : coefficient of x

Once we find the best θ_1 and θ_2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update θ_1 and θ_2 values to get the best fit line ?

Cost Function (J):

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

IMAGE 2

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

Gradient Descent:

To update θ_1 and θ_2 values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ_1 and θ_2 values and then iteratively updating the values, reaching minimum cost.

4.2 WORKING

Based on the model mentioned above our project has been made .

First we collect the data set and use the following data set to find the accuracy and the predict the price of the gold for the next/following day.

The given data set for the past 8years (2009-2017).

(only 1st 10 rows of data are shown) source-[kaggle.com](https://www.kaggle.com)

| | OPEN | HIGH | LOW | CLOSE | WAP | NO OF SHARES | NO OF TURNOVER | TOTAL TURNOVER | DELIVERABLE | % DELIVERABLE QUANTITY TO TRADE QUANTITY | SPREAD-HL | SPREAD-CO |
|----|------|------|------|-------|----------|--------------|----------------|----------------|-------------|--|-----------|-----------|
| 0 | 0.79 | 0.79 | 0.76 | 0.76 | 0.790000 | 7430 | 7 | 5848 | 7430 | 100.00 | 0.03 | -0.03 |
| 1 | 0.79 | 0.79 | 0.79 | 0.79 | 0.790000 | 310 | 4 | 244 | 310 | 100.00 | 0.00 | 0.00 |
| 2 | 0.83 | 0.83 | 0.83 | 0.83 | 0.830000 | 75 | 1 | 62 | 75 | 100.00 | 0.00 | 0.00 |
| 3 | 0.87 | 0.87 | 0.87 | 0.87 | 0.870000 | 1050 | 2 | 913 | 1050 | 100.00 | 0.00 | 0.00 |
| 4 | 0.91 | 0.91 | 0.91 | 0.91 | 0.910000 | 400 | 1 | 364 | 400 | 100.00 | 0.00 | 0.00 |
| 5 | 0.91 | 0.91 | 0.90 | 0.91 | 0.910000 | 3632 | 6 | 3303 | 3632 | 100.00 | 0.01 | 0.00 |
| 6 | 0.91 | 0.91 | 0.91 | 0.91 | 0.910000 | 55 | 1 | 50 | 55 | 100.00 | 0.00 | 0.00 |
| 7 | 0.93 | 0.93 | 0.92 | 0.92 | 0.920000 | 74 | 2 | 68 | 74 | 100.00 | 0.01 | -0.01 |
| 8 | 0.89 | 0.89 | 0.89 | 0.89 | 0.880000 | 77 | 2 | 68 | 77 | 100.00 | 0.00 | 0.00 |
| 9 | 0.88 | 0.88 | 0.85 | 0.85 | 0.870000 | 84 | 2 | 73 | 84 | 100.00 | 0.03 | -0.03 |
| 10 | 0.84 | 0.88 | 0.84 | 0.86 | 0.840000 | 8100 | 10 | 6806 | 8100 | 100.00 | 0.04 | 0.02 |

IMAGE 3

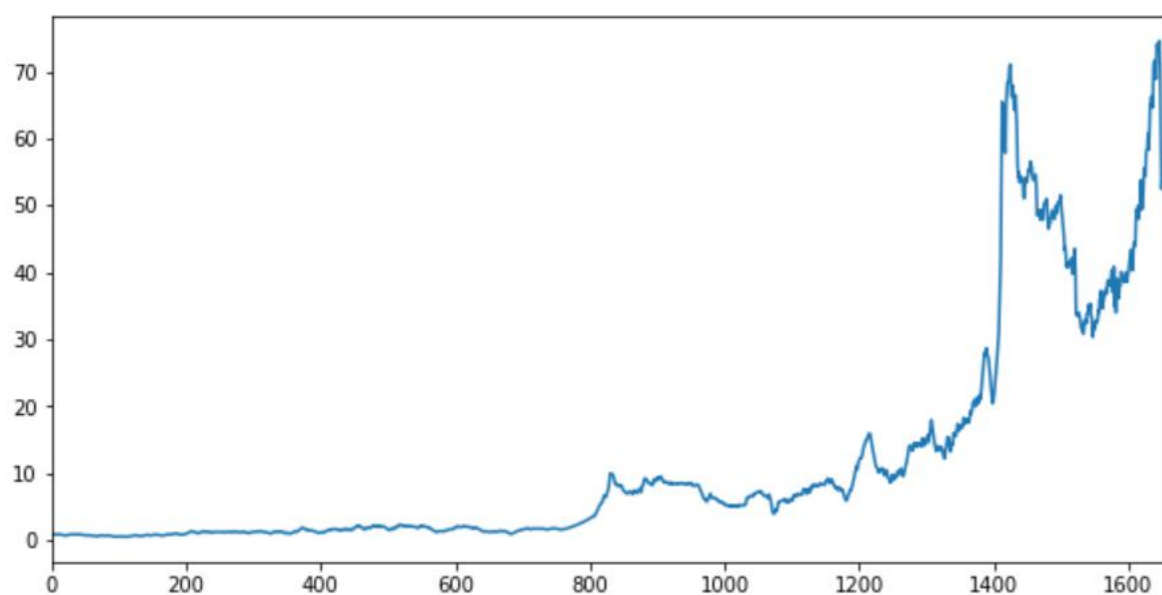


IMAGE 4

A plot showing the increase and decrease of **GOLD PRICE** .

```
t=0.8
t=int(t*len(df))
X_train=X[:t]
y_train=y[:t]
X_test=X[t:]
y_test=y[t:]
```

```
linear=LinearRegression().fit(X_train,y_train)
```

```
X_test.fillna(X_train.mean(),inplace=True)
```

C:\Users\tanay_paul\Anaconda3\lib\site-packages\pandas\core\generic.py:6130: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>
self._update_inplace(new_data)

```
predict_price = linear.predict(X_test)
```

```
predict_price=pd.DataFrame(predict_price,index=y_test.index,columns=['Total Turnover'])
```

```
predict_price.plot(figsize=(10,5))
y_test.plot()
plt.legend(['predict_price', 'actual_price'])
plt.ylabel("Gold ETF Price")
plt.show()
```

IMAGE 5

Usage of the **LINEAR REGRESSION MODEL** which has been applied here to find the prediction.

```
btn=widgets.Button(description="Press ME to know the Accuracy")
display(btn)
btn1=widgets.Button(description="Press to find GOLD ETF PRICE")
display(btn1)

def my_event_handler(btn_object):
    r2_score=linear.score(X[t:],y[t:])*100
    z=float("{0:.2f}".format(r2_score))
    print(z)
def my_event_handler1(btn_object):
    print ("Gold ETF Price = ", round(linear.coef_[0],2), " 3 Days Moving Average", round(linear.coef_[1],2), " 9 Days Moving Average")

btn.on_click(my_event_handler)
btn1.on_click(my_event_handler1)
```

Press ME to know t...

Press to find GOLD...

Gold ETF Price = 1.24 3 Days Moving Average -0.27 9 Days Moving Average + 1.27
96.58

IMAGE 6

The Predicted Price of the GOLD has been showed also along with 3 days moving average to 9 days moving average where its is shown as “Gold ETF Price = 1.24 * 3 Days Moving Average – 0.2 * 9 Days Moving Average + 0.39”

With accuracy= **96.58%**.

CHAPTER 5

5.1 IMPLEMENTATION AND RESULTS

```
btn=widgets.Button(description="Press ME to know the Accuracy")
display(btn)
btn1=widgets.Button(description="Press to find GOLD ETF PRICE")
display(btn1)

def my_event_handler(btn_object):
    r2_score=linear.score(X[t:],y[t:])*100
    z=float("{0:.2f}".format(r2_score))
    print(z)
def my_event_handler1(btn_object):
    print ("Gold ETF Price =", round(linear.coef_[0],2)," 3 Days Moving Average", round(linear.coef_[1],2)," 9 Days Moving Average")

btn.on_click(my_event_handler)
btn1.on_click(my_event_handler1)
```

Press ME to know t...

Press to find GOLD...

Gold ETF Price = 1.24 3 Days Moving Average -0.27 9 Days Moving Average + 1.27
96.58

The Predicted Price of the GOLD has been showed also along with 3 days moving average to 9 days moving average where its is shown as “Gold ETF Price = 1.24 * 3 Days Moving Average – 0.2 * 9 Days Moving Average + 0.39”

With accuracy= 96.58%

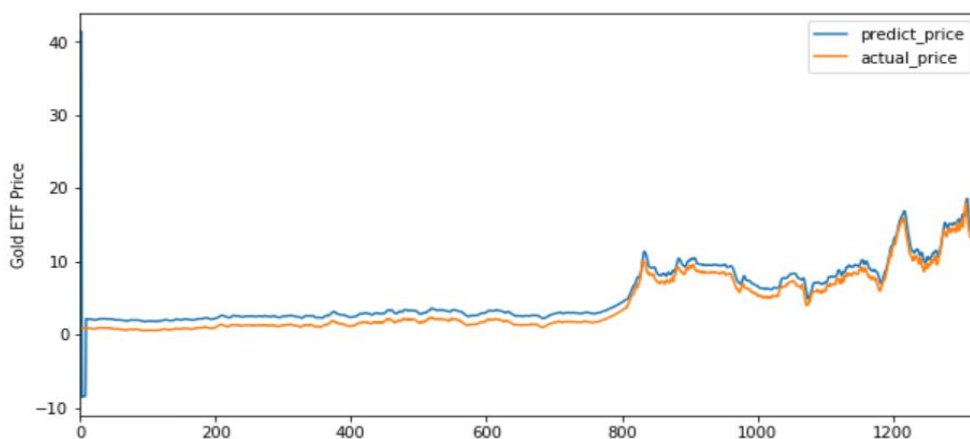


IMAGE 7

The plot showing a minor difference in the **PREDICTED PRICE** and **ACTUAL PRICE**.

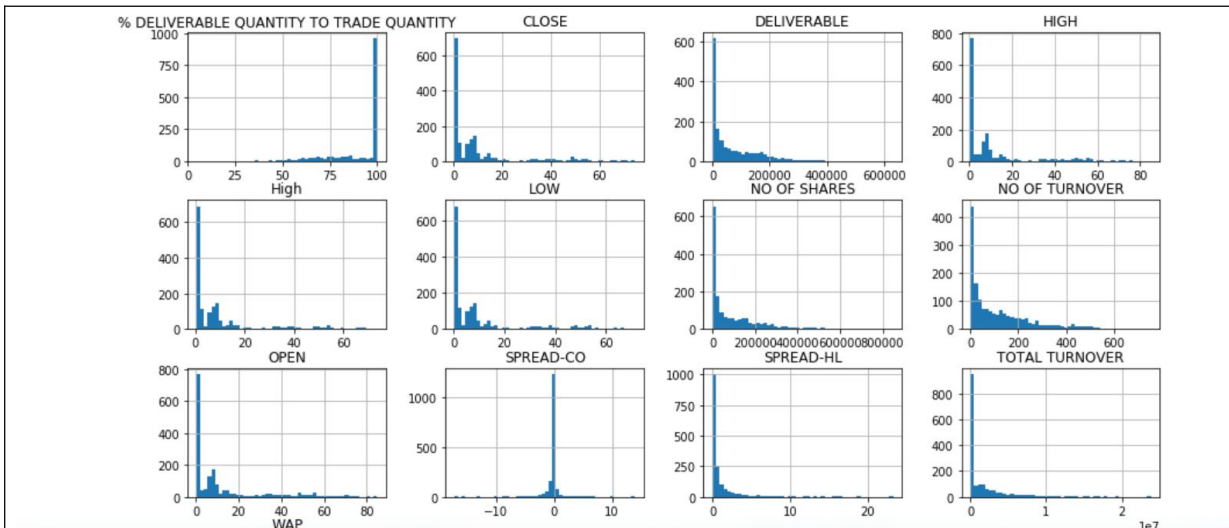


IMAGE 8

HISTOGRAM PLOT FOR EVERY VALUE IN DATASET

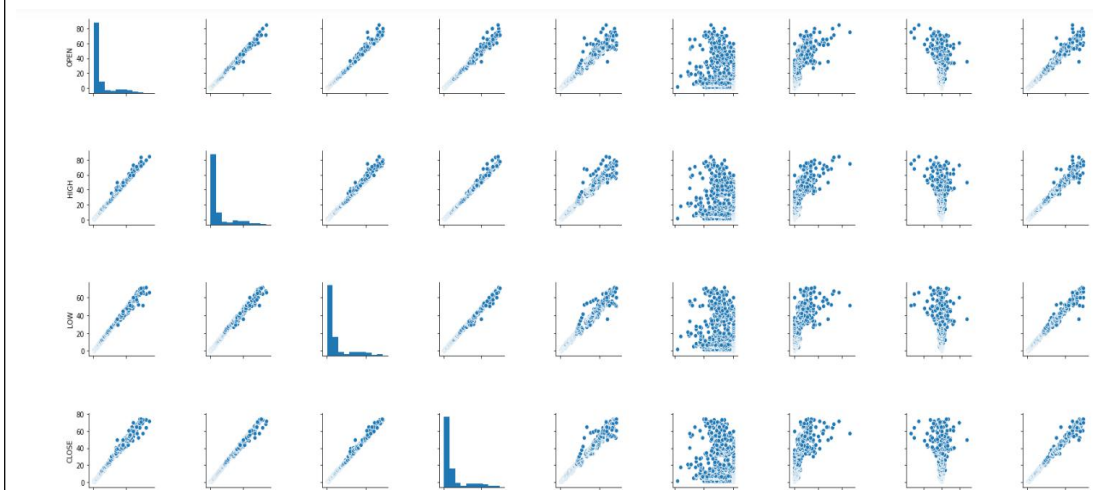


IMAGE 9

PAIR PLOT FOR EVERY VALUE IN THE DATA SET

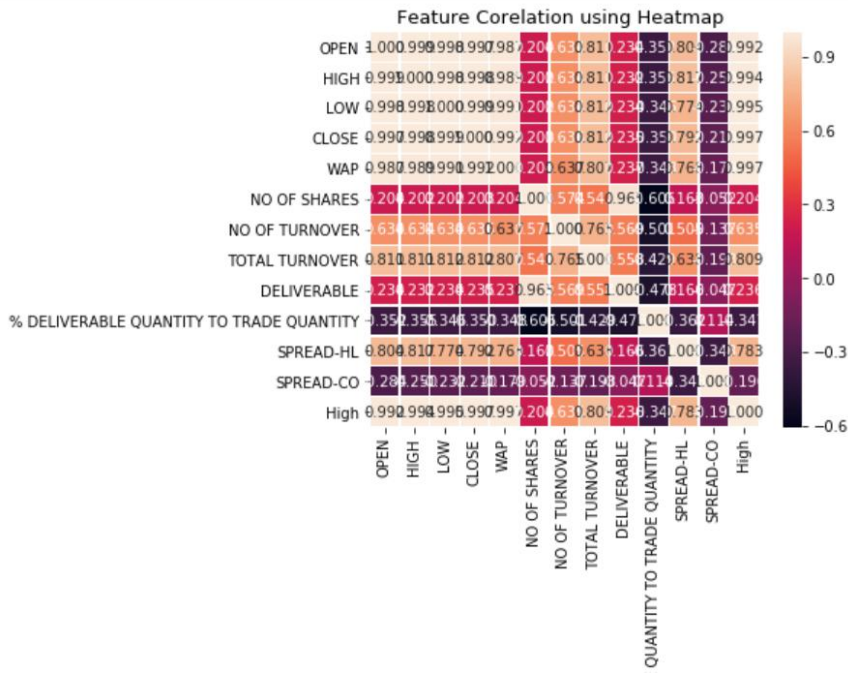


IMAGE 10

HEAT MAP DIAGRAM FOR THE GIVEN DATASET

CHAPTER 6

6.1 CONCLUSION

Gold has been one of the most important commodities throughout history. Maintaining gold reserves by central banks is crucial to support the current economic structure of the world. Some major companies and investors also invest a huge amount of money in gold. Although not easy, predicting the rate of gold would help investors and central banks to better decide when to sell and buy it, thus maximizing their profits. In this study, we used machine learning algorithms to predict the gold rates very accurately. Our study is also the most comprehensive to date, thus taking into consideration various economic indicators of various countries and companies. It is very helpful specially for common people and also for the investors where and when to actually invest in gold shares in future.

REFERENCES

- 1 M. C. & C.-C. C. Hsin-Hung Chen, "The Integration of Artificial Neural Networks and Text Mining to Forecast Gold Futures Prices," *Communications in Statistics - Simulation and Computation*, pp. 1532- 414, 2016.
- 2 D. G. Navin, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector," *International Journal of Science and Research*, pp. 2026-2030, 2015.
- 3 V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," *American International Journal of Contemporary Research*, pp. Vol 4, No 3, 2014.
- 4 M. S. S. S. S. Kumar Chandar, "Forecasting Gold Prices Based on Extreme Learning Machine," *International Journal of Computers Communications & Control*, pp. 372-380, 2016.
- 5 A. Y. A. S. Z. Ismail, "Forecasting Gold Prices Using Multiple Linear Regression Method," *American Journal of Applied Sciences*, pp. 1509- 1514, 2009.
- 6 Hassani, H., Silva, E. S., Gupta, R., & Segnon, M. K. (2015). Forecasting the price of gold. *Applied Economics*, 47(39), 4141-4152.
- 7 I. S. G. a. A. Mody, "International Gold Price Movements, 1972-1982," *Economic and Political Weekly*, vol. Vol. 17, no. No. 46/47, pp. 1861- 1870, 1982.
- 8 H. Long, "<http://money.cnn.com>," *cnn.com*, 10 february 2016. [Online]. Available: <http://money.cnn.com/2016/02/10/investing/china-gold/>. [Accessed 24 december 2016].
- 9 P. V. M. V. P. G. M. P. Sima P Patel, "Gold Market Analyzer using Selection based Algorithm," *International Journal of Advanced Engineering Research and Science (IJAERS)*, vol. 3, no. 4, pp. 55-102, 2016.
- 10 A. Y.-C. Hossein Mombeini, "Modelling Gold Price via Artificial Neural Network," *Journal of Economics, Business and Mangement*, vol. 3, no. 7, pp. 699-703, 2015.
- 11 Jan Ivar Larsen. *Predicting stock prices using technical analysis and machine learning*. NTNU, 2010.
- 12 Vatsal H. Shah. *Machine learning techniques for stock prediction*. NYU, 2007.
- 13 Shunrong Shen, Haomiao Jiang, and Tongda Zhang. *Stock market forecasting using machine learning algorithms*. Stanford University, 2012.

INDIVIDUAL CONTRIBUTION REPORT

“GOLD PRICE PREDICTION”

TANAY PAUL
1605326

Abstract: Given historical price data up to and including a given day, we attempt to predict whether the Indian PM price fix of gold the following day will increase or decrease relative to the current day's price fix. Our predictions are based primarily on technical indicators calculated from historical price data for gold as well as for a variety of financial variables. Using logistic regression for binary classification, we have attained over 94% accuracy using only features calculated from the gold price, and over 95.8% accuracy after adding certain inter-market variable features that display a high correlation with gold prices. I also simulate executing 1-day trades based on this prediction, where we can either buy or sell 1 ounce of gold each day, and achieve an average daily profit exceeding RS30000.

Individual contribution and findings: Basic idea of the project and all the necessary data sets for correction and updation and also the implementation of various technologies for the coding(in PYTHON) part of the project.

Individual contribution to project report preparation: In the writing of the abstract , various technologies used in the project and also about finding the information from various research and thesis papers.

Individual contribution for project presentation and demonstration: Introduction to the project ,abstract and implementation and conclusion .

Full Signature of Supervisor:
student:

.....

Full signature of the

.....

INDIVIDUAL CONTRIBUTION REPORT

“GOLD PRICE PREDICTION”

SHUBHAM KUMAR

1605231

Abstract: Given historical price data up to and including a given day, we attempt to predict whether the Indian PM price fix of gold the following day will increase or decrease relative to the current day’s price fix. Our predictions are based primarily on technical indicators calculated from historical price data for gold as well as for a variety of financial variables. Using logistic regression for binary classification, we have attained over 94% accuracy using only features calculated from the gold price, and over 95.8% accuracy after adding certain inter-market variable features that display a high correlation with gold prices. I also simulate executing 1-day trades based on this prediction, where we can either buy or sell 1 ounce of gold each day, and achieve an average daily profit exceeding RS30000.

Individual contribution and findings: In the coding part where all the plots of various data are shown and also in the coding in **PYTHON** and implementation.

Individual contribution to project report preparation: Basic gathering of information regarding the technologies used.

Individual contribution for project presentation and demonstration:

Responsible for various images in the presentation

Full Signature of Supervisor:
student:

.....

Full signature of the

.....

INDIVIDUAL CONTRIBUTION REPORT

“GOLD PRICE PREDICTION”

KOUSTUV SARKAR
1605281

Abstract: Given historical price data up to and including a given day, we attempt to predict whether the Indian PM price fix of gold the following day will increase or decrease relative to the current day’s price fix. Our predictions are based primarily on technical indicators calculated from historical price data for gold as well as for a variety of financial variables. Using logistic regression for binary classification, we have attained over 94% accuracy using only features calculated from the gold price, and over 95.8% accuracy after adding certain inter-market variable features that display a high correlation with gold prices. I also simulate executing 1-day trades based on this prediction, where we can either buy or sell 1 ounce of gold each day, and achieve an average daily profit exceeding RS30000.

Individual contribution and findings: Basic idea of the project and all the necessary coding in PYTHON and implementation of the idea.

Individual contribution to project report preparation: In the writing of the various technologies used in the project and also about finding the datasets.

Individual contribution for project presentation and demonstration: Advantages and disadvantages and the latest implementation in the market and conclusion.

Full Signature of Supervisor:
student:

.....

Full signature of the

.....

INDIVIDUAL CONTRIBUTION REPORT

“GOLD PRICE PREDICTION”

SAURAV DAS
1605304

Abstract: Given historical price data up to and including a given day, we attempt to predict whether the Indian PM price fix of gold the following day will increase or decrease relative to the current day’s price fix. Our predictions are based primarily on technical indicators calculated from historical price data for gold as well as for a variety of financial variables. Using logistic regression for binary classification, we have attained over 94% accuracy using only features calculated from the gold price, and over 95.8% accuracy after adding certain inter-market variable features that display a high correlation with gold prices. I also simulate executing 1-day trades based on this prediction, where we can either buy or sell 1 ounce of gold each day, and achieve an average daily profit exceeding RS30000.

Individual contribution and findings: Responsible in the designing of widgets.

Individual contribution to project report preparation: Conclusin and references writing

Individual contribution for project presentation and demonstration: Revising the presentation and necessary correction.

Full Signature of Supervisor:
student:

.....

Full signature of the

.....

