

FACULTY: DR.SIVAKUMAR.S

NAME & REG.NO:

PRATHIBAN.V 19MID0010

ARUN SREERAM.R 19MID0041

ANMOL 19BCE0891

*Data Analysis on inflation in India by CPI index and
Finding Value of money in 2025*

ABSTRACT:-

A Consumer Price Index (CPI) is designed to measure the changes over time in general level of retail prices of selected goods and services that households purchase for the purpose of consumption. The CPI measures price changes by comparing, through time, the cost of a fixed basket of commodities. The basket is based on the expenditures of a target population in a certain reference period. Since the basket contains commodities of unchanging or equivalent quantity and quality, the index reflects only pure price. This index is an important economic indicator and is widely considered as a barometer of inflation, a tool for monitoring price stability and as a deflator in national accounts. The dearness allowance (DA) of Government employees and wage contracts between labour and employer is based on this index. We are going to do Exploratory Data Analysis on Consumer Price Index in India and we are going to use Machine learning to find the value of money for future years.

KEYWORDS: CPI, Machine Learning, Exploratory Data Analysis

INTRODUCTION:-

Inflation is defined as the gradual loss of buying power of a fixed currency over time. It is crucial in the development of central banks' macroeconomic policies around the world. Inflation rises during periods of strong growth rates, while deflation occurs during periods of low or negative growth rates. We conducted research to determine the relationship between inflation and the Consumer Price Index (CPI) (Consumer Price Index). Consumer Price Indices (CPIs) track variations in the general level of prices of goods and services purchased by households for consumption across time. CPI statistics are frequently used as a macroeconomic indicator of inflation, as a tool by governments and central banks for inflation targeting and price stability monitoring, and as national account deflators. The CPI is also used to index employees' dearness allowances in response to price increases. As a result, the Consumer Price Index (CPI) is regarded as one of the most important economic indicators. Weighting diagrams (consumption patterns) and pricing data collected at regular intervals are required for the calculation of CPI values.

The formula for calculating Consumer Price Index is Laspeyre's index which is measured as follows : [Total cost of a fixed basket of goods and services in the current period * 100] divided by Total cost of the same basket in the base period

EDA, With the use of summary statistics and graphical representations, exploratory data analysis refers to the crucial process of doing first investigations on data in order to uncover patterns, spot anomalies, test hypotheses, and check assumptions.

Machine learning (ML) is the study of computer algorithms that may improve themselves over time based on experience and data. Artificial intelligence is considered to be an aspect of it. Machine learning algorithms create a model based on training data to make predictions or judgments without being specifically trained to do so.

LITERATURE REVIEW:

Title	Name and Year of journal	Findings
UNDERSTANDING INFLATION IN INDIA	NBER WORKING PAPER SERIES 2016	<ul style="list-style-type: none">• Leading macroeconomics textbooks explain inflation using a Phillips curve, in which the rate of inflation is affected by expected inflation, output relative to trend, and supply shocks.• Models assume that lags in the inflation rate determine expected inflation and that supply shocks are primarily changes in the relative prices of food and energy.

Modeling and Forecasting Inflation in India	IMF Working Paper,1999	<ul style="list-style-type: none"> • It concludes that, while the broad money target has been 's importance, changes in monetary aggregates continue to be an important predictor of future inflation. • The exchange rate and import prices are also important, particularly for manufacturing inflation.
Artificial Neural Network Based Model for Forecasting of Inflation in India	Sciencedirect,2016	<ul style="list-style-type: none"> • Economists attempt to predict future inflation rates using various theories and practises; however, no theory or practise can predict inflation rates with 100% accuracy. • That is because there are numerous other factors, such as natural disasters, sudden changes in weather, government policies, changes in taxation regulations, and qualitative political decisions, that cause the inflation rate to rise and fall.
Forecasting Costa Rican inflation with machine learning methods	Sciencedirect,2020	<ul style="list-style-type: none"> • Forecasts are computed using two k-nearest neighbours variants, random forests, extreme gradient boosting, and a long short-term memory (LSTM) network. • A combination outperforms the individual forecasts that comprise it as well as the average of the univariate forecasts.

What does machine learning say about the drivers of inflation?	Research Gate,2022	<ul style="list-style-type: none"> • Overall, the study confirms the utility of random forests in explaining inflation patterns in a diverse set of advanced economies. The method produces much better in-sample description as well as more accurate inflation forecasts. • This disciplined and intuitive computational method is capable of better capturing nonlinearities and potentially complex interactions between factors in the real world.
Effects of inflation and the time value of money on order quantity and allowable shortage	Research Gate, 1994	<ul style="list-style-type: none"> • According to this study, the optimal order quantity and maximum allowable shortage vary with the difference between inflation and time discount. • Computational results show that as the difference between the inflation rate and the time discount rate increases, so does the effect on the system.
Measures of Inflation in India: Issues and Perspectives	IARNIW Conference, 2010	<ul style="list-style-type: none"> • It is critical to have a reliable primary measure of inflation at the national level when formulating monetary policy. In this regard, the compilation of CPI (Urban) and CPI (Rural) data could pave the way for a country-wide representative CPI.

Is money supply the cause of inflation in India? An alternative postulate to understand inflation	Science Direct, 2013	<ul style="list-style-type: none"> The quantity theory of money provides the identity $MV = PT$, which is correct, but the functional relationship between M and P, namely $P = f(M)$, does not hold in India because it is based on the constancy of V and T.
Exchange rate pass-through in India	Research Gate, 2008	<ul style="list-style-type: none"> This paper investigates the relationship between inflation and the exchange rate in India during the post-reform period. It discovers that India's exchange rate pass-through into domestic prices is moderate. CPI has a long run pass through elasticity of 3.7-17 percent, while WPI has a pass through elasticity of 28.6 percent.
Utilizing Machine Learning in the Consumer Price Index	Springer, 2018	<ul style="list-style-type: none"> Machine learning has improved the quality of our CPI while also significantly reducing the time spent on a time-consuming task during a critical stage of the production cycle.

EXPLORATORY DATA ANALYSIS:-

Through the use of summary statistics and graphical representations, exploratory data analysis refers to the crucial process of doing first investigations on data in order to uncover patterns, spot anomalies, test hypotheses, and check assumptions.

While the objectives are distinct, projects analyse the urban CPI, rural CPI, and KMEANS datasets. The initiative focuses on the impact and relationship between the CPI index and products. Because the focus has shifted, raw data will be replaced with demeaned seasonal adjusted data.

➔ Using visualization we have found out the factors affecting inflation rate in Urban sector and rural sector

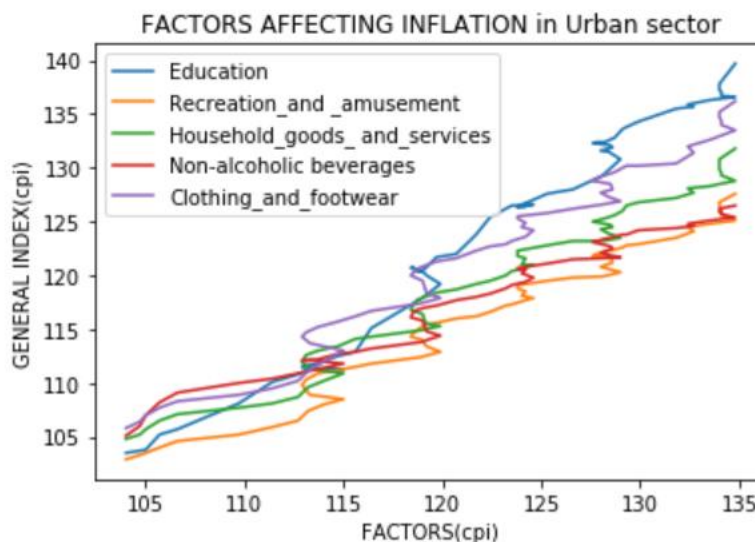
CODE:

```
x = df.General_index
y1=df[factors[0]]
y2=df[factors[1]]
y3=df[factors[2]]
y4=df[factors[3]]
y5=df[factors[4]]
plt.plot(x, y1, label=factors[0])
plt.plot(x, y2, label=factors[1])
plt.plot(x, y3, label=factors[2])
plt.plot(x, y4, label=factors[3])
plt.plot(x, y5, label=factors[4])
plt.xlabel('FACTORS(cpi)')
plt.ylabel('GENERAL INDEX(cpi)')

plt.title("FACTORS AFFECTING INFLATION in Urban sector")

plt.legend()

plt.show()
```



by using Karl Pearson's coefficient, we have been successful in finding out top 5 factors which has affected inflation.

Education plays a main role in affecting inflation in India compared to another factor.

CODE:

```
x = df1.General_index
y1=df1[factors1[0]]
y2=df1[factors1[1]]
```

```

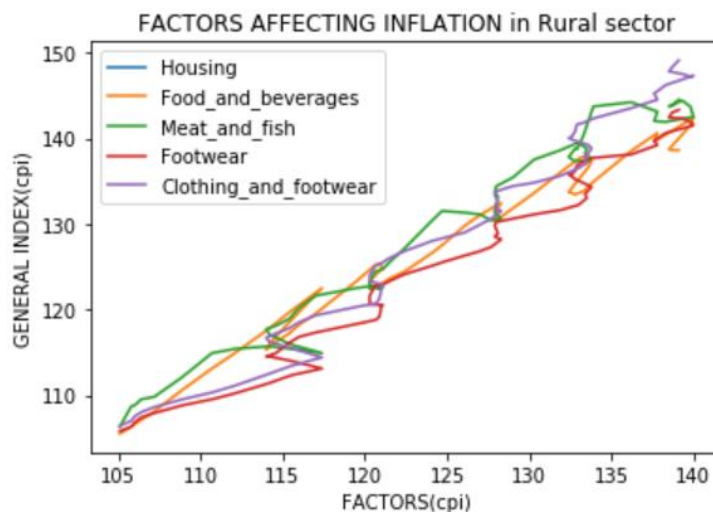
y3=df1[factors 1[2]]
y4=df1[factors 1[3]]
y5=df1[factors 1[4]]
plt.plot(x, y1, label=factors 1[0])
plt.plot(x, y2, label=factors 1[1])
plt.plot(x, y3, label=factors 1[2])
plt.plot(x, y4, label=factors 1[3])
plt.plot(x, y5, label=factors 1[4])
plt.xlabel('FACTORS(cpi)')
plt.ylabel('GENERAL INDEX(cpi)')

plt.title("FACTORS AFFECTING INFLATION in Rural sector")

plt.legend()

plt.show()

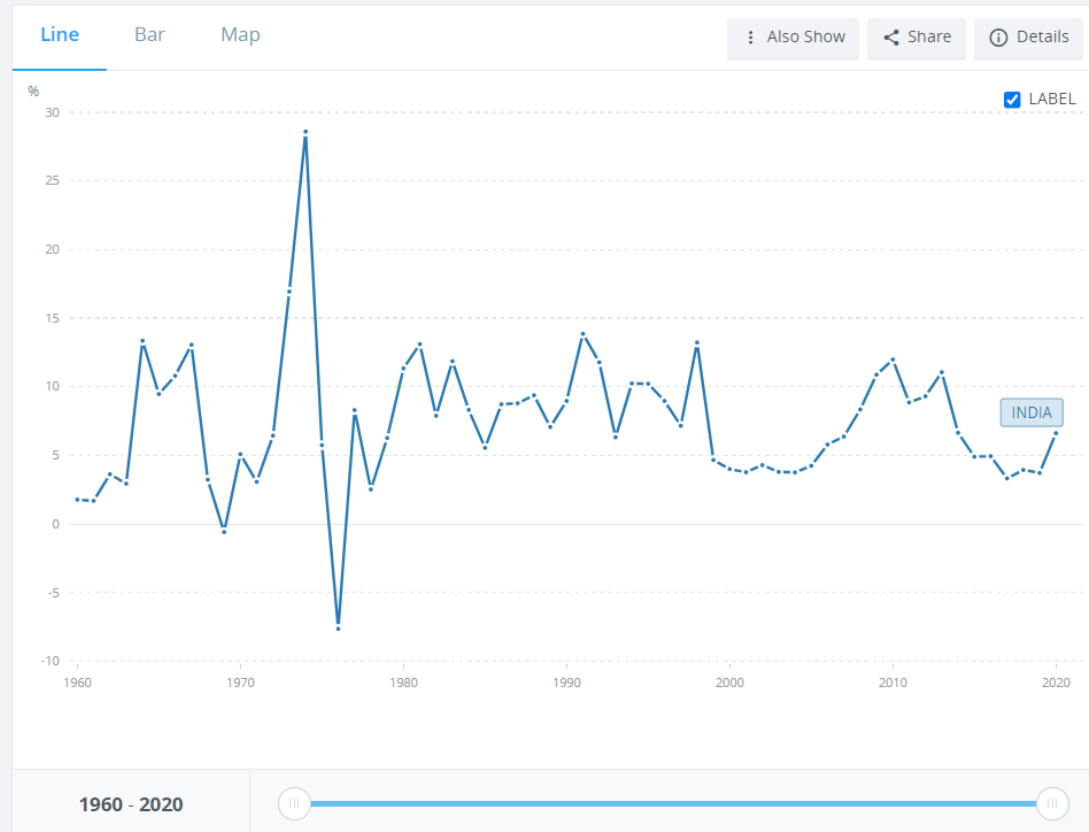
```



From this we can see that there is a steady increase in rate due to clothing and footwear when compared to other factors in the rural area

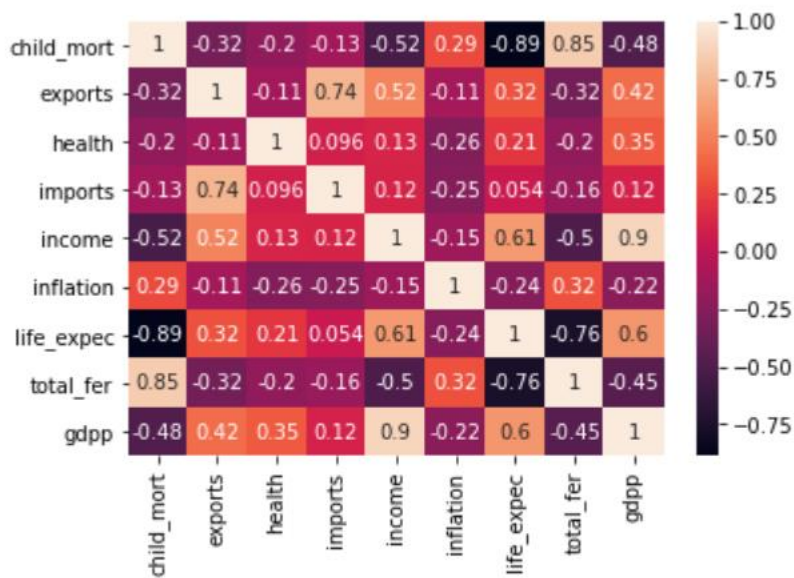
Inflation, consumer prices (annual %) - India

International Monetary Fund, International Financial Statistics and data files.



This graph shows us the percentage of consumer prices from 1960 to 2020.

From this we can very well understand that there has been a high rise and very low in consumer prices during 1970s to 1980s



This is a correlation map

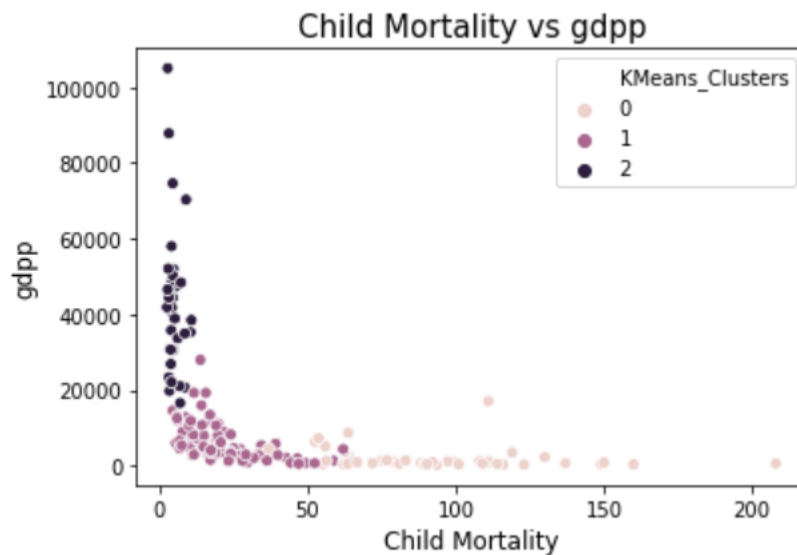
This shows:

1. GDP and income, imports and exports, child mortality and total fertility are highly positive correlated
2. whereas life expectancy and child mortality are highly negative correlated

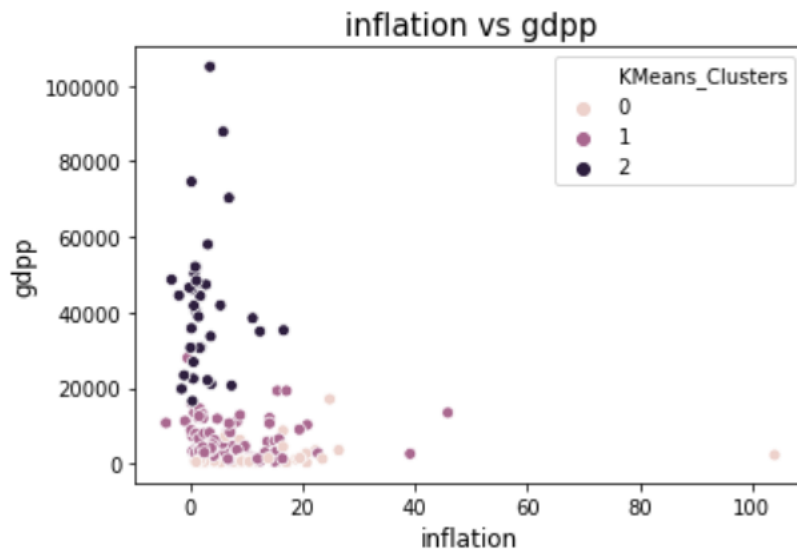
With the help of visualisation tools we have derived the scatter plots:

```
#visulazing hcluster results
#child mortality vs exports
```

```
sns.scatterplot(hcluster_df['child_mort'],hcluster_df['gdpp'],hue='hcluster',data=hcluster_df)
plt.title("Child Mortality vs gdpp", fontsize=15)
plt.xlabel("Child Mortality", fontsize=12)
plt.ylabel("gdpp", fontsize=12)
plt.show()
```



```
#visulazing hcluster results
sns.scatterplot(hcluster_df['inflation'],hcluster_df['gdpp'],hue='hcluster',data=hcluster_df)
plt.title("Inflation vs gdpp", fontsize=15)
plt.xlabel("Inflation", fontsize=12)
plt.ylabel("gdpp", fontsize=12)
plt.show()
```



From above two clusters graph we can conclude that

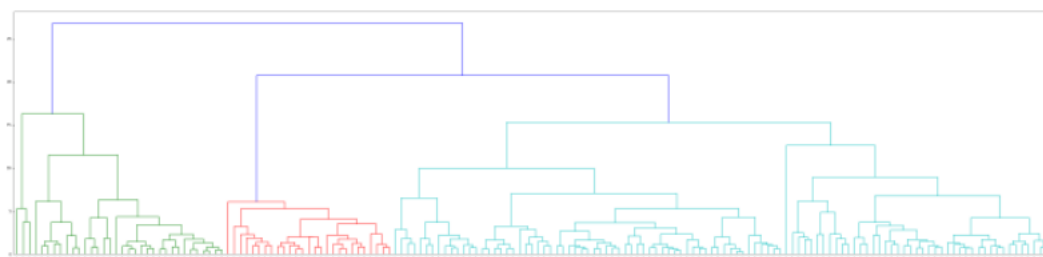
1. Country having high child-mortality, low GDP per capita and low inflation (The measurement of the annual growth rate of the Total GDP) is a **under-developing country**
2. Country having low child-mortality, high gdpp and high infaltion is the **developed country**

So here we conclude that

0 = **under-developing country**

1 = **developing country**

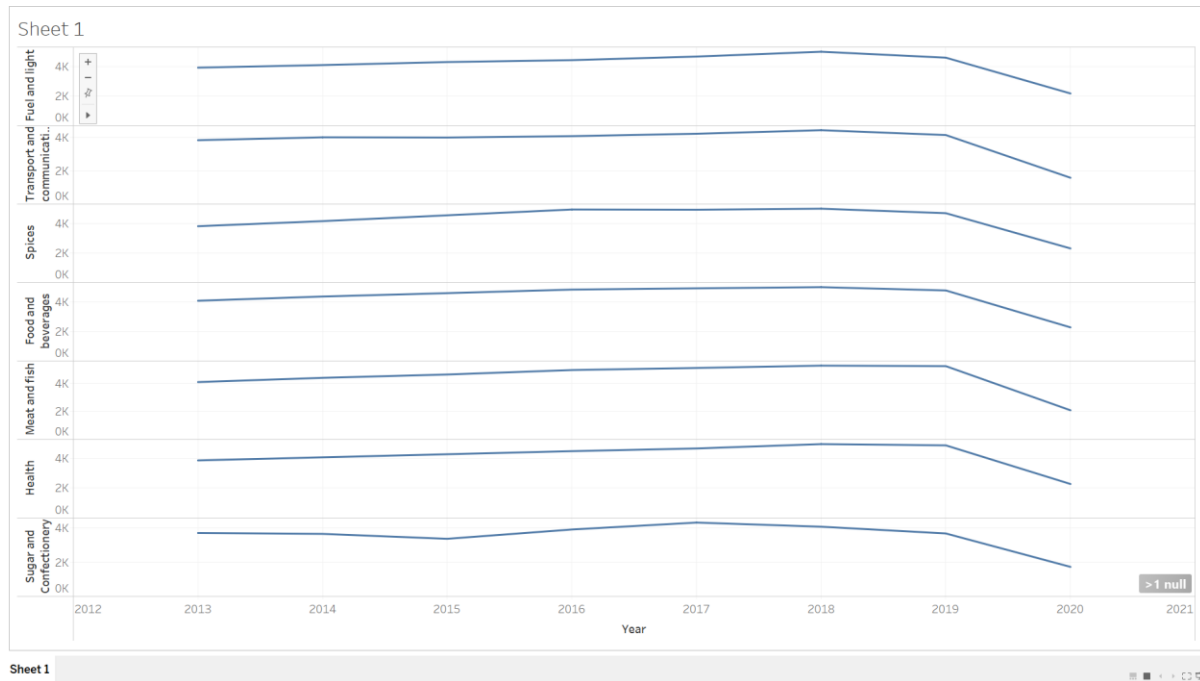
2 = **developed country**



From above dendrogram we can take minimum no of clusters as 2 and maximum number of cluster as 5. As we can see for dendrogram 3 in the right no of clusters, so we are going to take 3 no of clusters

COMPARISON OF INFLATION RATES FOR ESSENTIALS YEAR BY YEAR:

Using tableau we have derived this line chart



From this chart we can see that:

- Same trend pattern of inflation has been followed year by year for different essentials

Value of Money:

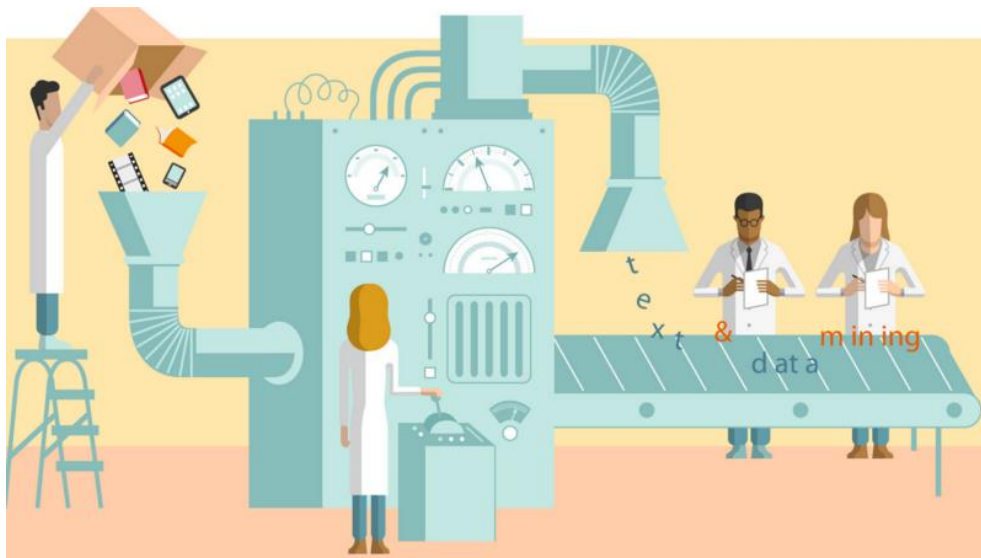
The quantity of things that will be exchanged for one unit of money, then, is the value of money. Money's purchasing power, or the amount of goods and services it can buy, determines its value. We will use VALUE_OF_MONEY dataset which contains value of money in each year we are going to apply machine learning to find value of money in 2025.

Let's see the steps involved in the process

MACHINE LEARNING:-

Machine learning is an AI application that allows systems to learn and improve based on their own experiences without being explicitly programmed. Machine learning is concerned with creating computer programs that can access data and use it to learn on their own.

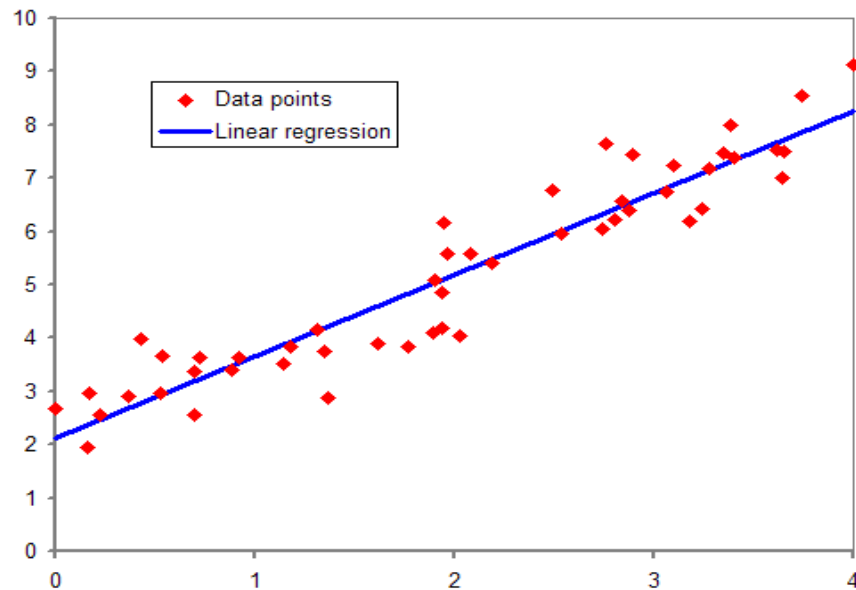
Machine learning can speed up data processing and analysis, making it a useful technology for predictive analytics programs. With minor changes in deployment, predictive analytics algorithms can train on even larger data sets and perform deeper analysis on multiple variables using machine learning.



REGRESSION:-

Regression analysis is a method for determining the relationship between independent variables or features and a dependent variable or outcome. It is used in machine learning as a method for predictive modelling, in which an algorithm is used to predict continuous outcomes.

To predict a continuous value, regression models are used. One common example of Regression is predicting the price of a house based on its features such as size, price, and so on. It is a technique that must be supervised. We are going to use this technique to find value of money in 2025.



STEPS IN REGRESSION:-

- I. Get data
- II. Clean the data
- III. Split the data into training and testing
- IV. Apply various ML algorithm on the data
- V. Select the ML algorithm with best accuracy
- VI. Find the value of money in 2025 with the selected algorithm

STEP(I):

Data collection is the process of acquiring, collecting, extracting, and storing large amounts of data, which can be structured or unstructured, such as text, video, audio, XML files, records, or other image files that will be used in later stages of data analysis.

"Data collection" is the first step in analyzing patterns or useful information in data. The data to be analyzed must be gathered from a variety of reliable sources. The data source is from government of India. The dataset contains year and value of money as its attributes. It is a verified dataset and data source.

```
import pandas as pd
from sklearn.model_selection import train_test_split
#import os
#print(os.getcwd())
val_money=pd.read_csv("VALUE_OF_MONEY.csv")
print(val_money.head())
```

	YEAR	VALUE_OF_MONEY
0	1958	100.00
1	1959	105.15
2	1960	109.60
3	1961	109.60
4	1962	113.11

(49, 1) (49,) (13, 1) (13,)

STEP(II):

The process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted is known as data cleaning. When it comes to data analysis, this data is usually not required or helpful because it can hinder the process or produce inaccurate results. We can clean the data using NULL option in python. We can also perform data cleaning with help of WEKA tool. There is no NULL value or outliers in the dataset we can proceed forward.

STEP(III):

Train/Test is a technique for determining the accuracy of your model. It is called Train/Test because the data set is divided into two parts: a training set and a testing set. Training accounts for 80% of the budget, while testing accounts for 20%. The training set is used to train the model. You use the testing set to put the model through its paces. Training the model entails creating the model. Testing the model entails determining the model's accuracy.

The reason for splitting data into training and testing set is that when the dataset is divided into train and test sets, the training dataset will not contain enough data for the model to learn an effective mapping of inputs to outputs. There will also be insufficient data in the test set to evaluate the model's performance effectively. We use in build function in python to split the data into training and testing.

```
X_train, X_test, y_train, y_test =
train_test_split(val_money[['YEAR']], val_money.VALUE_OF_MONEY, test_size=0.20,
random_state=42)
print(X_train.shape, y_train.shape, X_test.shape, y_test.shape)
```

STEP(IV):

We apply various ML algorithms on our dataset to find the algorithm with best accuracy and minimum error. Different algorithms have their own method for solving the problem so applying different algorithm helps us understand the problem to find the model with best accuracy.

Algorithm we are going to use are

- Ridge Regression
- Linear Regression
- XGBoost
- RandomForest
- Lasso Regression

Ridge:

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. Ridge regression uses a type of shrinkage estimator called a ridge estimator. Shrinkage estimators theoretically produce new estimators that are shrunk closer to the “true” population parameters. The ridge estimator is especially good at improving the least-squares estimate when multicollinearity is present.

FORMULA: $H_{\text{ridge}} = X(X'X + \lambda I)^{-1}X'$

```
from sklearn.linear_model import Ridge
clf = Ridge(alpha=1.0)
clf.fit(X_train, y_train)
predictions= clf.predict(X_test)
predictions
```

Linear Regression:

Linear Regression is a supervised learning-based machine learning algorithm. It carries out a regression task. Based on independent variables, regression models a target prediction value. It is primarily used to determine the relationship between variables and forecasting. Linear Regression is the process of determining a line that best fits the data points on the plot so that we can use it to predict output values for inputs that are not present in the data set we have, with the expectation that those outputs will fall on the line. FORMULA:

$$Y_i = f(X_i, \beta) + e_i$$

Y_i = dependent variable

f = function

X_i = independent variable

β = unknown parameters

e_i = error terms

```
from sklearn.linear_model import LinearRegression
model=reg = LinearRegression()
reg.fit(X_train, y_train)
pred=reg.predict(X_test)
```

XGBoost:

XGBoost is a gradient boosting ensemble Machine Learning algorithm that is decision-tree based. Artificial neural networks outperform all other algorithms or frameworks in prediction problems involving unstructured data (images, text, etc.). FORMULA

$F_2(x) = \sigma(0+1 \cdot h_1(x)+1 \cdot h_2(x))$ where the resulting value of $F_2(x)$ is considered as the prediction from XgBoost model

```
#gradient boosting regression
from sklearn import ensemble
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
```



```

from sklearn.metrics import r2_score
params = {'n_estimators': 500, 'max_depth': 4, 'min_samples_split': 2,
          'learning_rate': 0.01, 'loss': 'ls'}
clfg = ensemble.GradientBoostingRegressor(**params)
clfg.fit(X_train, y_train)
predg=clfg.predict(X_test)

```

Random Forest:

It constructs decision trees from various samples and uses their majority vote for classification and average for regression. When using the Random Forest Algorithm to solve regression problems, you are using the mean squared error (MSE) to how your data branches from each node. This formula calculates the distance of each node from the predicted actual value, helping to decide which branch is the better decision for your forest.

```

from sklearn.ensemble import RandomForestRegressor
from sklearn.datasets import make_regression
RandomForestRegressor()
regr = RandomForestRegressor(max_depth=2, random_state=0,n_estimators=100)
modelr=regr.fit(X_train, y_train)
print(regr.predict(X_test))

```

Lasso Regression:

Lasso regression is a type of linear regression that uses shrinkage. The lasso regression method allows you to shrink or regularize these coefficients in order to avoid overfitting and improve their performance on different datasets. When the dataset exhibits high multicollinearity or when you want to automate variable elimination and feature selection, this type of regression is used. Performs L1 regularization, i.e. adds penalty equivalent to absolute value of the magnitude of coefficients. Minimization objective = LS Obj + α * (sum of absolute value of coefficients)

```

from sklearn import linear_model
clf1 = linear_model.Lasso(alpha=0.1)
clf1.fit(X_train, y_train)
print(clf1.predict(X_test))

```

STEP(V):

Select the ML algorithm with best accuracy, Choosing the best algorithm from list of available of list of algorithm is a essence of machine learning. Various machine learning algorithms look for various trends and patterns. For all data sets and application circumstances, one algorithm isn't the best. You'll need to run a lot of tests, analyze machine learning algorithms, and fine-tune their hyperparameters to find the optimum answer. To analyze your model and justify the choice of an algorithm, you must first identify, justify, and use a model performance indicator. The F1 score, true positive rate, and within cluster sum of squared error are examples of model performance measures. At least one deep-learning and one non-deep-learning algorithm should be used to implement your approach. Compare and document model performance after that. Apply at least one further iteration to the process model, at least one of which contains the feature generation

task. Data normalization and principal component analysis, for example, have an impact on model performance (PCA). You may use specific technologies or frameworks to address your challenge depending on the algorithm class and data set size.

	Ridge	Linear	XGBoost	RandomForest	Lasso
TRAIN	0.777600	0.999954	0.999954	0.980355	0.777600
TEST	0.793255	0.793263	0.993353	0.981103	0.793263

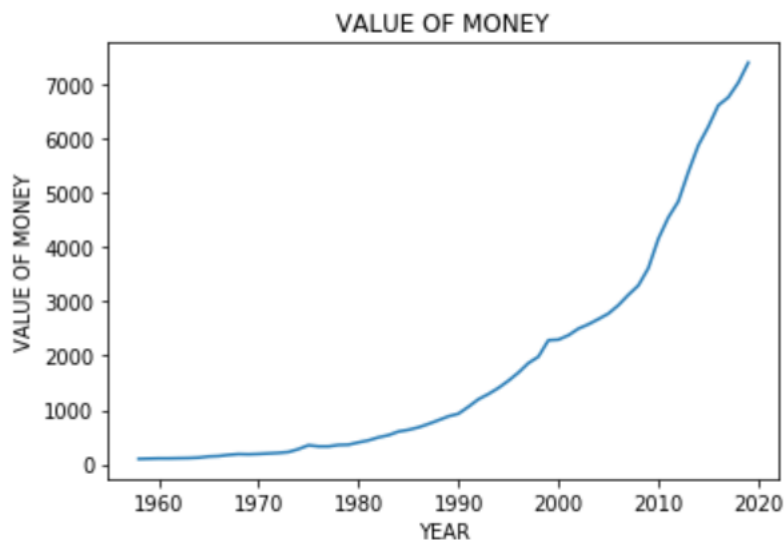
Gradient boosting has the best Accuracy we will choose Gradient boosting to find the value of money.

STEP(VI):

Find the value of money in 2025 with the selected algorithm, get the algorithm and put year 2025 in the algorithm to find the value of money.

```
clf.predict(2025)
```

The value of Money in 2025 in India is 7365.28.



VALUE OF MONEY OVER THE YEARS

CONCLUSION:

Inflation is defined as the gradual loss of buying power of a fixed currency over time. It is crucial in the development of central banks' macroeconomic policies around the world. Inflation rises during periods of strong growth rates, while deflation occurs during periods of low or negative growth rates. We conducted research to determine the relationship between inflation and the Consumer Price Index (CPI) (Consumer Price Index). We used a variety of machine learning techniques to figure out how the rate of inflation has changed over time.

In October 2019, India's CPI hit a 16-month high of 4.7 percent YoY, breaking the RBI's medium-term inflation goal of 4% for the first time in 15 months. Inflation in the CPI was at 4% YoY in the prior month. The increase in retail inflation was mostly driven by a rise in food prices.

In October, food inflation, as assessed by the consumer food price index (CFPI), reached a three-year high of 7.9% YoY, owing to rising costs of vegetables, fruits, and pulses. Food price inflation has been on the rise for the previous eight months, owing to supply disruptions caused by the country's delayed and uneven monsoon.

The rural economy contributed the majority of the increase in food costs, as inflation in that sector doubled in a month, from 3.1 percent YoY in September to 6.3 percent YoY in October. Core inflation, on the other hand, moderated to 3.6 percent YoY from 4.2 percent YoY in the previous month due to inflation in miscellaneous, services, pan, tobacco, and intoxicants. As a result, the economy's current consumer demand is restrained. Given that the CPI has surpassed the RBI's medium-term target of 4% and is in stark contrast to the economy's declining industrial activity, the monetary policy decision in December may be complicated.

The likelihood of a rate drop will be determined by the RBI's flexibility in dealing with retail inflation and the July-September quarter GDP number, which will be revealed later this month. *With the help of machine learning we have predicted that the value of Money in 2025 in India is 7365.28.*

REFERENCES:

1. https://www.nber.org/system/files/working_papers/w22948/w22948.pdf
2. <https://www.imf.org/external/pubs/ft/wp/1999/wp99119.pdf>
3. <https://www.sciencedirect.com/science/article/pii/S1616865816300061>
4. <https://www.sciencedirect.com/science/article/pii/S2666143820300120>
5. <https://www.bis.org/publ/work980.htm#:~:text=The%20machine%20learning%20model%20predicts,firmly%20grounded%20on%20economic%20theory.>
6. <https://www.sciencedirect.com/science/article/abs/pii/S0925527394900477>
7. <https://www.rba.gov.au/publications/bulletin/2014/sep/pdf/bu-0914-5.pdf>
8. <https://www.sciencedirect.com/science/article/pii/S1877042814031140>
9. http://www.igidr.ac.in/conf/money1/mfc-12/Pyne_Pradyut.pdf
10. https://www.stat.fi/static/media/uploads/ajk_en/Events/nsm2019/myklatur_-_utilizing_machine_learning_in_the_consumer_price_index.docx