

Predicting the GDP of India using Machine Learning

Navyasri S¹, Nayala Hafsa¹, Nayana S¹, Nireeksha Bilimale¹, Asha Rani M²

¹Student, Department of Computer Science Engineering, GSSSIETW, Mysuru, Karnataka, India.

²Assistant Professor, Department of Computer Science Engineering, GSSSIETW, Mysuru, Karnataka, India.

Corresponding author: navyashreenavi26@gmail.com

Abstract: - Gross domestic products (GDP) is a monetary measure of the market value of all the final goods and services produced in a specific time period, often annually. The GDP of India is depending on several factors like agriculture, land, manufacturing, electricity and gas etc. There is always changes in the GDP, it increases or decreases based on the above mentioned factors for every year. In this project we developed a web based application which helps to predict the next year GDP based on the several factors. This application helps the government to analyse over the predicted data and take several action so that we can stop the GDP decreasing rate

Key Words— GDP(Gross domestic product), k-means, clustering, Regression.

I. INTRODUCTION

The economy of India is characterized as a developing market economy. It is the world's fifth- largest economy by nominal GDP and the third-largest by purchasing power parity(PPP). According to the IMF, on a per capita income basis, India ranked 142nd by GDP (nominal) and 119th by GDP (PPP) per capita in 2018. Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing on business information. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades. Prediction algorithm is method we use to analyze the GDP data and helps in predicting the future result based on the previous data sets. Similarly, the K-means clustering algorithm is use to find out which states are similar to each other considering GDP rate, and which states are safer or riskier for GDP, clustering algorithm was performed on the all the previous gdp dataset. We used the GDP datasets for our study. The datasets are downloaded from the website data.gov.in. We collected data sets from 2011-2018. The analysis part involves the collecting the data and performing the statistics operation on the previous datasets and predicting the result of 2019, with the help of this analysis result we have predicted 2020 GDP.

II. METHODOLOGY

This project is implemented using java programming language. Both servlet and JSP technologies are used to create a web application. Servlet are java programs are precompiled which can create dynamic web contents. There are many interfaces and class in the servlet API such as http servlet, servlet request, servlet response etc. JSP is used to create a web application just as servlet.it can be thought of as a extension to servlet because it provides more functionality than servlet. MySQL server is used as a backend.

A. Regression algorithm

Regression algorithm is designed to find the historical relationship between an independent and a dependent variable to predict the future values of the dependent variable. A regression models the past relationship between variables to predict their future behaviour Steps Involved:

1. Scan the transaction database and perform the operation on the missing values.
2. Calculate Mean and Variance.
3. Calculate Covariance.
4. Estimate Coefficients.
5. Make Predictions.
6. Predict GDP.

B. K-Means Clustering

K-means clustering algorithm was used to investigate the high and low-frequency GDP locations. Further, they have been used association rule mining to recognize the association between the various factors related to road traffic GDPs at various places with changeable GDP occurrences.

Steps Involved:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

In the development of the predictive model the data sets were collected internally in secondary form. Secondary data imply statistical materials or information not originated or obtained by the investigator himself, but obtain from someone's record or published source such as the central government agencies i.e., data.gov.in.

C. Analysis

- The Clean Data was stored in the CSV format; the CSV data is imported into MySQL with help of JSON. Now data is ready to analyzed by the data analyzing tools.
- The datasets for 2013-2019 are collected from the data.gov.in which consists of different parameters the GDP and NSDP.
- Overall report with prediction 2020 is shown
- Comparison study based on state wise
- Classification based on high and low GDP frequency rates

Example:

Input: Consider the data sets of GDP from 2010 to 2017

Table.1.

X(year)	Y(value)
2008	3496
2010	3500
2010	3987
2011	2987
2012	3019
2013	3999
2014	4015
2015	4786
2016	4018
2017	4445

Table.2.

X (year)	Y(value)	A1=(x-mean of x)	B1=(y-mean of y)	A1*B1	(A1) ²	(b1) ²
2008	3496	-4	-329	1316	16	108241
2010	3500	-3	-325	975	9	105625
2010	3987	-2	162	-324	4	26244
2011	2987	-1	-838	838	1	702244
2012	3019	0	-806	0	0	649636
2013	3999	1	174	174	1	30276
2014	4015	2	190	380	4	36100
2015	4786	3	961	2883	9	923521
2016	4018	4	193	772	16	37249
2017	4445	5	620	3100	25	384400

$Y=b_0+b_1(x)$ // x is prediction value or independent value

$$b_1=(A1*B1)/(A1)^2$$

$$10114/85 \ 118.98$$

$b_0= \text{mean of } y - b_0(\text{mean of } x)$

$$=3825.5-(118.98*2012) = -235600.76$$

$Y=b_0 + b_1(X)$

$$Y=-235600.76+ 118.8(2019)$$

$Y=4500$ /// This is the predict value of 2019

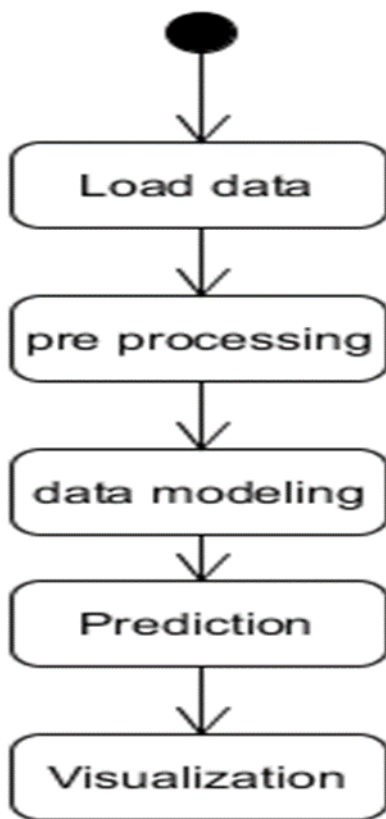


Fig.1. Data flow diagram

Datasets are loaded from data.goc.in website and data mining steps will be done i.e., pre-processing later data modelling have to be done. finally, prediction is done using regression

algorithm applied on those data sets. Predicted results are displayed using data analytics tool High charts.

III. RESULTS AND DISCUSSION

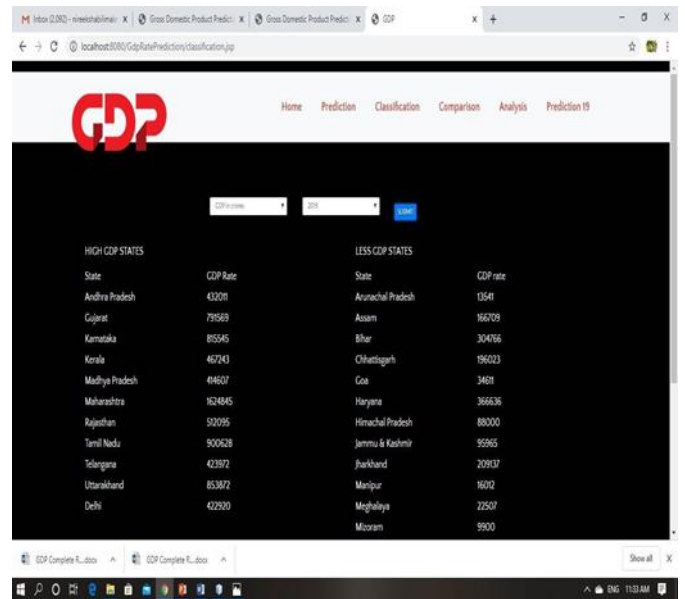


Fig.2. GDP of all the states in that selected year.

Screenshot1 shows Classification where user selects GDP in crores or percentage and year then this page displays classified high and low GDP of all the states in that selected year.

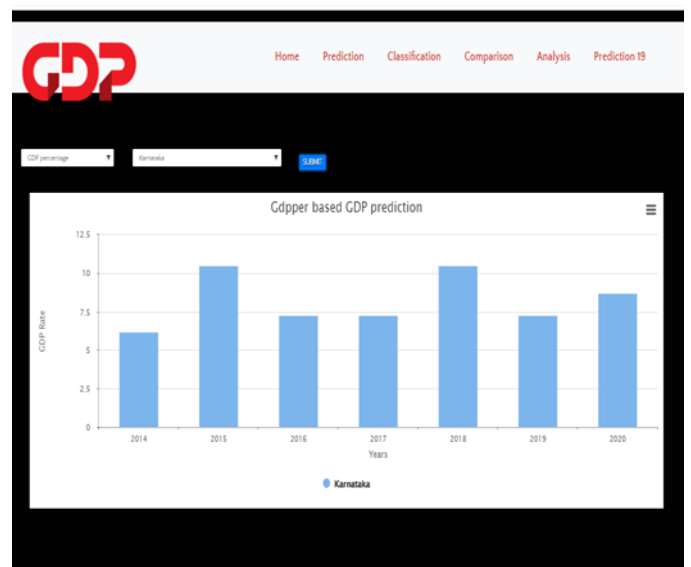


Fig.3. Predicted and displays 2020 GDP of that state

Snapshot2 shows Prediction Page where user selects GDP in Crores or percentage and state to which GDP of 2020 should be predicted and displays 2020 GDP of that state.

IV. CONCLUSION

In this project we developed a web based application which helps to predict the 2020 GDP based on the previous year GDP. This application helps the government to analyse over the predicted data and take several action so that we can stop the GDP decreasing rate.

Prediction algorithm is method we use to analyze the GDP data and helps in predicting the future result based on the previous data sets. Similarly, the K-means clustering algorithm is use to find out which states are similar to each other considering GDP rate, and which states are safer or riskier for GDP, clustering algorithm was performed on the all the previous GDP dataset.

We used the GDP datasets for our study. The datasets are downloaded from the website data.gov.in. we collected data sets from 2011-2018. Classification part involves grouping high and low GDP states. Comparison Part compares each state GDP of that selected year and shows all states GDP in pie chart form using High charts. The analysis part involves the collecting the data and performing the statistics operation on the previous datasets and predicting the result of 2019. We have done analysis between predicted and actual GDP of 2019. Based on this result we have predicted the 202 GDP of various states of India.

REFERENCES

- [1]. Amith K Shukla, Sandeep Kumar, "Interval Type-2 Fuzzy Weighted Extreme Learning Machine for GDP prediction", IEEE-2018.
- [2]. Zhao Rong, "Research on Expression and update method of GDP data", -2016
- [3]. Alfian presekala, "Predicting Recessions: Forecasting US GDP Growth through Supervised Learning", IEEE 2014
- [4]. Lia Jhang, "Forecasting GDP: A Linear Regression Model" IEEE-2012
- [5]. Yukari Shriota, "Predicting the impact of the GDP measure on the forest cover using forecasting and regression", IEEE-2011