CHAPTER 1

1 INTRODUCTION

1.1 OVERVIEW

Data analysis, also called as data analysis or analytics, is the act of scrutinising, cleansing, translating, or modelling data in order to unearth key data, offer ideas, and help decision-making. A significant source of funding for India's development is its mineral industry. To take advantage of reduced stipends and the shifting economic environment in India, foreign firms engage directly in expanding private Indian businesses. After the 1991 profitable extreme, India began its lucrative liberalisation, and since then, the country's mineral production has continuously expanded, leading to the creation of more than one crore jobs. The Financial Times reports that India overtook Arab nations as the main destination for such material in 2015. India attracted investment worth $ 31 billion in the first half of 2015, compared to $ 28 and $ 27 billion for China and the US separately.

Data mining is a specific data analysis system that focuses on models and knowledge discovery for predicted rather than merely descriptive reasons. Business intelligence includes data analysis that primarily depends on aggregation and is based on business information. Data booby-trapping uses various methods and technologies to make connections among enormous volumes of data. It is regarded to be one of the most important technological advancements in recent years. We use the purpose is to take method to analyse current mineral data and make predictions about the future based on sets of historical data. The K-means cluster method is also utilised to identify which Sections are safer while analyzing lower and higher growth rates, as well as whether Areas are comparable with each other when taking Nutrient into account.

The MINERAL databases are being used for our study. The figures are downloaded from the website data.gov.in. from 2001 until 2021 We gathered several data sets. The analytic procedure includes data collecting, statistics computing on previous information, and prediction for 2022.

1.2 PROBLEM STATEMENT

- There's some MINERAL product which makes profit and loss to the Government.
- The Government can not prognosticate the unborn profit from the MINERAL product

- One would suppose that acts of terrorism would have a negative impact on MINERAL product flows to affected Sectors.

1.3 EXISTING SYSTEM

- There's always dropping of the frugality rate or MINERAL product rate.
- There are number of investors in India but delicate to choose and to prognosticate downfall it has loss or profit.
- Developed husbandry saw the biggest drop in MINERAL product flows in 2021 — at 37 per cent, to USD 712 billion.
- The people of India don't have the factual idea of MINERAL PRODUCTION and the different sources.

1.4 PROPOSED WORK

In this design, we examine recent MINERAL trends in various nations and examine Nutrient in Indian in more depth. We also build a knowledge of MINERAL generalities.

1.4.1 Benefits inferable from the work

- vaticination of Result For 2022 for all the Above Parameters
- Comparing all time’s data for different parameters.
- Helping the Govt to assay the low and high profit mining
- Generating the time wise report of all parameters
- Developing the Dashboard for all the Above

1.4.2 Defense for concluding a particular methodology for completing the tasks

Sources of data collection

The data sets were internally gathered in secondary form to create the prophetic model. Statistics or information obtained from sources other than the investigator's own records or published sources, such as central government organisations, are referred to as secondary data.

Data Sources

https//data.gov.in/ major- index/ foreign-direct- investment- mineral- husbandry

Volume of data	No of tables	No of Records
100MB	1	16000

CHAPTER 2

2 LITERATURE REVIEW

[1]Analysis of a Retrogression Algorithm for Estimating Administration, Production, and Transportation to Quality of Delivery of Products in the Cosmetic Industry STMIK Nusa Mandiri, Jakarta, Indonesia W Gata1, H B Novitasari2, R Nurfalah3, R Hernawati4, M J Shidiq5

The study analyses the system of Gross Domestic Product computation in INDIA. Statistical styles have been used to assay the data handed by the Reserve Bank of India. Graphs and maps are used to represent the calculated results in R. This paper seeks to prognosticate unborn values of the Gross Domestic Product for the country, the way that Reserve Bank of India and government must take in order to increase this value. The direct retrogression algorithm show the topmost delicacy

[2]Analysis of an Accidentally removed Algorithm to Anticipate Administration, Production, and Transport to Precision of Delivery of Cosmetic

In the financial industry, where a plethora of training images and high frequency slice rates provide datasets that are suitable for efficient retrogression, machine learning methods have become increasingly popular. Learning algorithms haven't really been thoroughly investigated in macroeconomics, where data is poor and slice rates are much lower, and success has been constrained in the sparse literature. In this paper, I investigate whether linear retrogression (SVR) and feedforward neural networks, trained exclusively on macroeconomic indicators from 1978 to 2005, are capable of predicting the 2008 Great Recession. Their effectiveness is contrasted with an average ARIMA model drawn from recent successful literature.

[3] A comparison of manual vector regression and linear regression IEEE Paper by Kavitha Ramya R, Kavitha S, and Varuna S

Time series data are usually used in forecasting and prediction. For greater accuracy, most applications including weather forecasting, finance, and stock market mix historical data with more recent streaming data. Regression models are used to examine the time series data, though. In order to choose the best model for improved prediction and accuracy, the training data set is used in this study to evaluate the linear regression or support vector regression. The

authors of the fifth edition of Introduction in Linear Regression Analysis are Douglas C. Montgomery, Arlen Peck, and G. Geoffrey Vining.

One of the most popular methods for analysing multi-factor data is regression analysis. Its popularisation and use come from the theoretically sound

The method of expressing the connection here between variable of interest (the result) and a group of connected predictors using an equation. Because of the sophisticated statistical theory and elegant underlying mathematics, linear regression is also intriguing theoretically. Understanding both the theoretical and the practical concerns that frequently arise when using the techniques with real-world data is necessary for effective usage of regression.

CHAPTER 3

3 SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

3.1.1 The data sets were internally gathered in secondary form to create the prophetic model. Statistics or information obtained from sources other than the investigator's own records or published sources, such as central government organisations, are referred to as secondary data.

3.2 PURPOSE

3.2.1 The Goal of software Requirement Gathering is to give the specialised, Functional and non-beneficial qualities, recommended to accumulate an online software App. The entire piece of software was designed to provide users flexibility in finding the route that's the shortest and greenest. To positioned it bluntly, the incentive in the rear of this SRS study is to supply an itemised definition of our product objective, its bounds and aims. This repository shows the task's alleged hobbies firm as well as its UI, hardware, and programming requirements. It describes how our client, the organisation, and the public view the thing and its value.

3.3 SCOPE

This framework's purpose is to provide a survey of records-digging methods used for MPI's anticipated lifespan. The framework makes it clear that records mining method, like grouping, is incredibly effective in anticipation of MPI.

3.3.1 3.4 SOFTWARE ARCHITECTURE

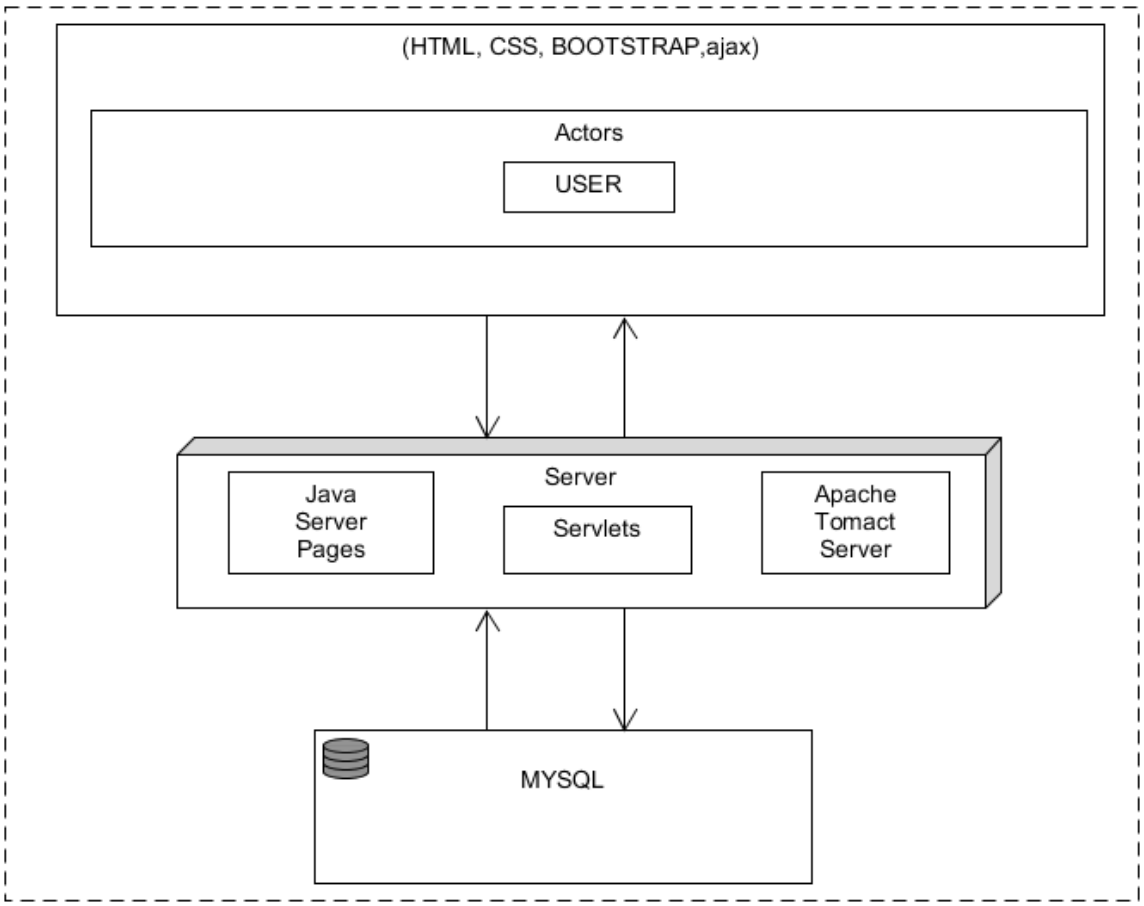


Fig.3.4 System Architecture

3.4 BACKGROUND

REGRESSION TECHNIQUE

- In statistics, linear regression regression is a method for simulating the relationship between a rate of 5.3 variable (y) and one or more explanatory factors (or neutral variables), denoted by the letter X. When there is only one explanatory variable, smooth regression model is employed. The approach is referred to as multiple linear regression when there are multiple explanatory variables. The purpose of the collection of rules in a regression problem is to anticipate output with actual value. Regression techniques' key lingo includes
- **Independent Variables (abilities)**: An impartial variable is a variable that is manipulated to determine the fee of a primarily based totally variable. Simply, they may

be the abilities which we want to use to are awaiting some given fee of Y. It can be moreover known as an explanatory variable

- **Dependent Variable(target):** The primarily based totally variable is based upon on the values of the impartial variable. Simply put, it's far the feature which we are searching for to are awaiting. This additionally may be normally known as a response variable.

K-MEANS CLUSTERING

K-technique Clustering is a technique for vector quantization that originated in signal processing and is widely used for cluster evaluation in data mining. n observations are divided into good enough clusters using the good enough-technique clustering approach, where each claim belongs to the cluster that has the closest mean, acting as the cluster's prototype.

The method uses an efficient kind of clusters (expect good sufficient groups) steady a priori to categorize a given record was set in a simple and straightforward manner. The key concept is to specify sufficient clusters, one for each grouping. Due to incredible strong motivations and incredible give-up outcomes, these centroids should really be positioned deftly. The preferable option is to place them as far apart from one another as you can. The following step is to associate each aspect from a given records broken with the closest centroid. The first phase is finished, and an earlier company age is carried out when there are no outstanding aspects. We need recalculate in this area to get a good enough result.

ACCURACY MEASURES

The give up final result for each regulations is compared to the corresponding give up outcome to determine accuracy. Each set of regulations' surrender final result is displayed in a table in an arranged manner according to the resource of the usage of their rank. With the aid of MPI, the genuine failure outcome is accomplished by maintaining parameters consistent and displaying them inside a format akin to algorithmic failure outcomes.

CHAPTER 4

4 SYSTEM DESIGN

4.1 INTRODUCTION TO DESIGN DOCUMENT

4.1.1 With the aid of the usage of providing the facts for how the software must be made, the software design can be applied to help with software programme application software programme improvement for android software. The use case models, series diagrams, and excellent helping requirement records that make up the software design specs are literary and components of information of the project's software programme program software programme layout.

4.2 SCOPE

This software programme design phase is for a basic level contraption that can be used to validate an idea for building a contrivance that provides a basic level of capacity to determine whether this is feasible for large-scale manufacturing use. The Application Software Design Document places emphasis on the length of a lines and their change. The device will be utilised with various pre-existing structure and could, in large part, correspond to an interaction in records that defies the objectification of record relations and management of the record objects. The MPI Design Specs are provided in this document.

4.3 DESIGN:

4.3.1 DFD:

A data flow diagram (DFD) depicts the "aqueduct" of information passing through a data system graphically. The figure of an information aqueduct can also be used to represent data processing. It's customary for a developer to start by creating a setup position DFD that demonstrates how the frame and external rudiments work together. To show more specifics of the frames being displayed, this setting location DFD is additionally "exploded."

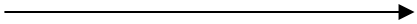
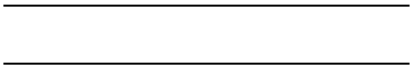
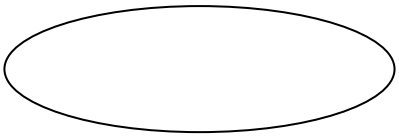
SYMBOLS	DESCRIPTION
	Represents Data Flow
	Represents Database
	Represents the External Entity
	shows the process through which incoming data flows are converted into outbound data flows.

Table.4.3 DFD Notations

Level 0

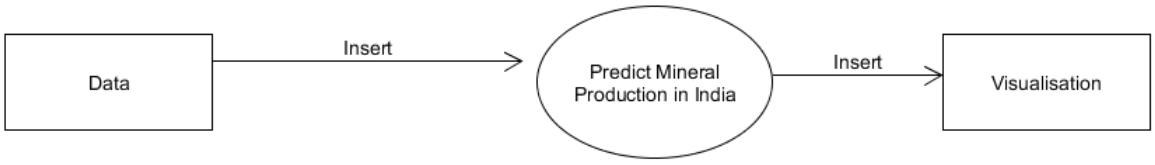


Fig.4.3.1 Data Flow Diagram

Level 1

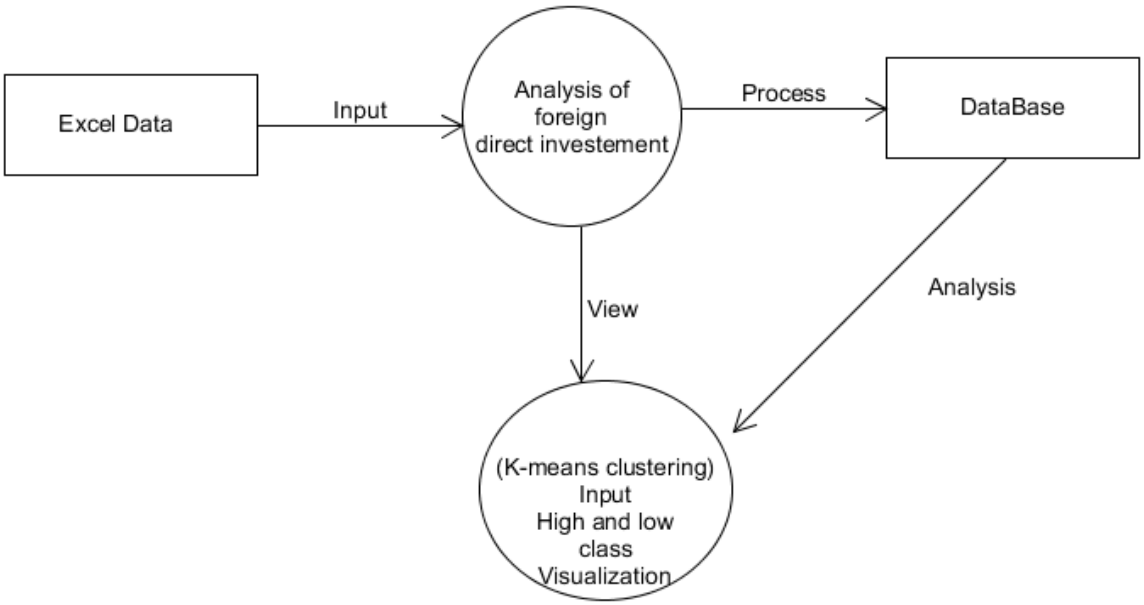
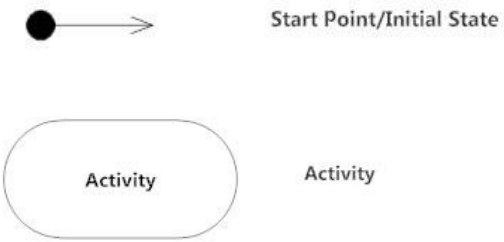


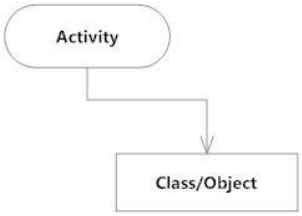
Fig.4.3.2 USER DFD

4.3.2 Activity Diagram:

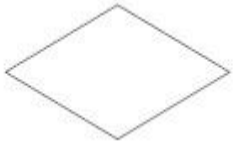
An information system's "aqueduct" of information is depicted graphically in a data flow diagram, or DFD. For the display of data processing, an information aqueduct diagram may also be used. Drawing a setting position DFD—which demonstrates how the frame and the outside rudiments work together—first is customary practise for artists. A further level of detail of the displayed frame is shown by "exploding" this setting position DFD.



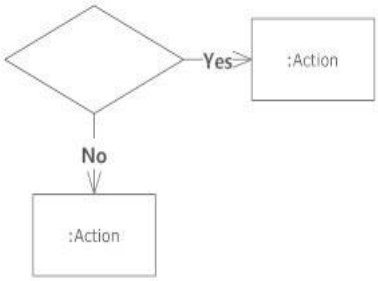
Action Flow



Object Flow



Decision Symbol



Guard Symbols

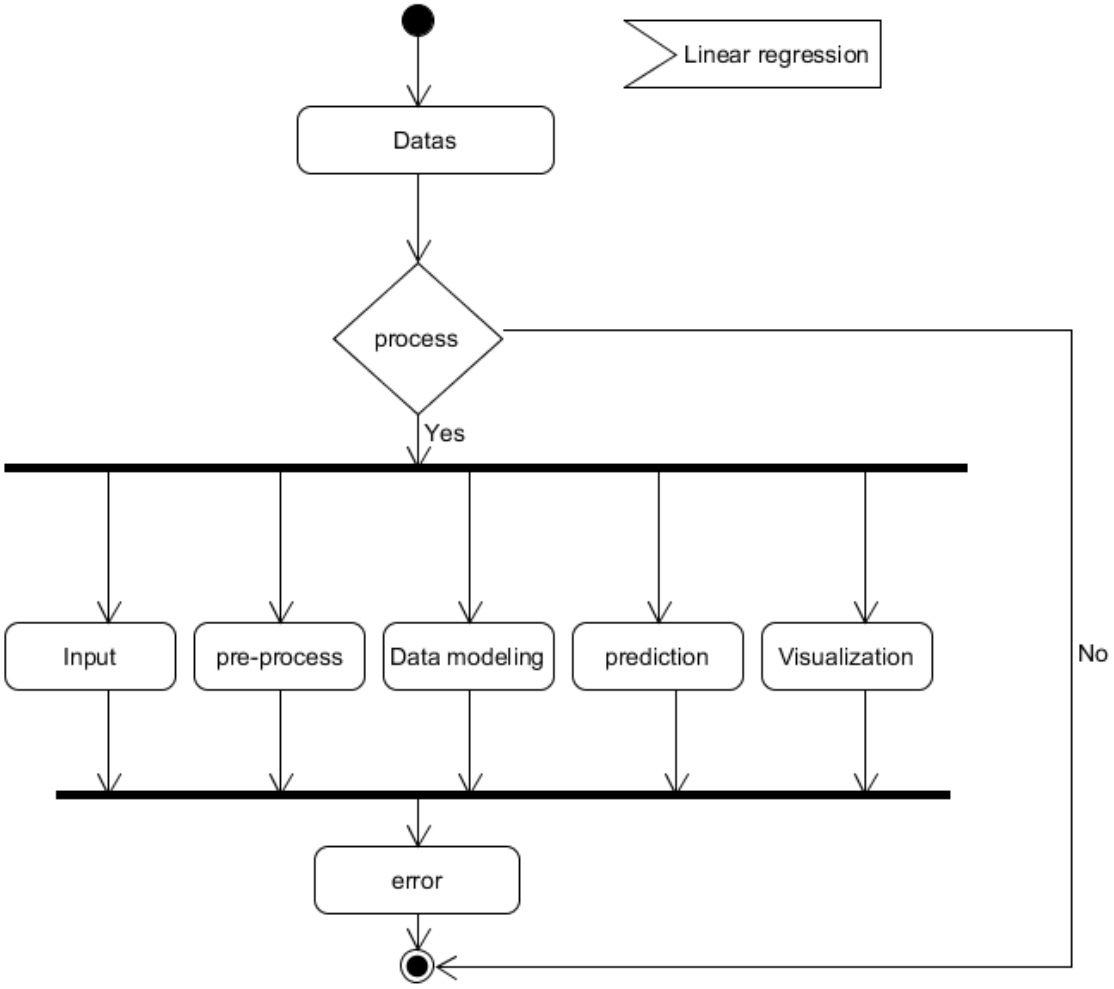


Fig.4.3.3 Activity Diagram

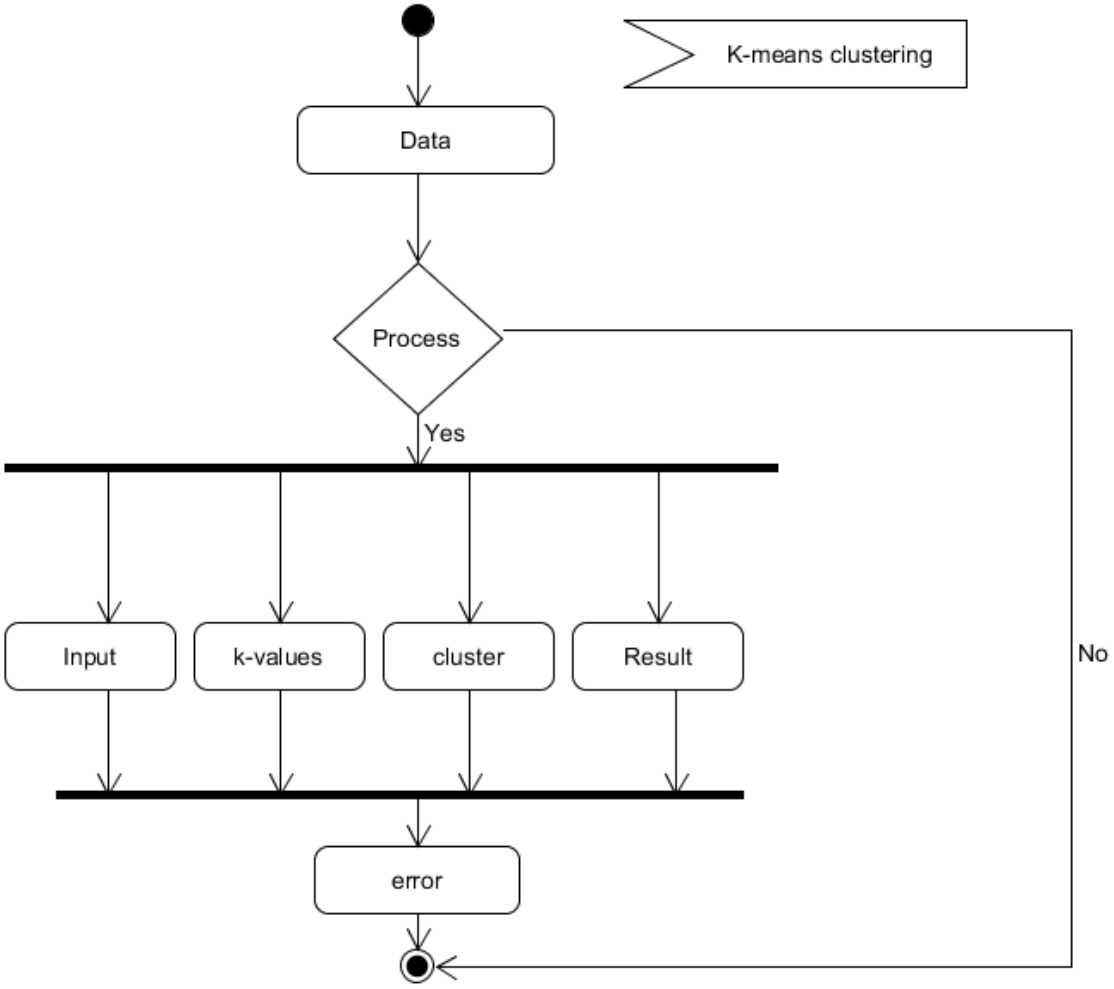


Fig.4.3.4 Activity Diagram of user

4.3.3 Usecase Diagram:

Use case illustration's challenge is to capture a frame's dynamic portion. Whatever the case, this description is too general to actually speculate on explaining the cause, given the other four outlines (activity, group, cooperation, and Sectorchart) also share a comparable explanation. We will look into a specific factor that will be apparent from the other four charts.

Use case charts are used to compile the requirements of a frame that includes external and external implications. These conditions are typically seen in plans. As a result, use cases are prepared and impersonators are identified when a frame is mined to gather its functionality.

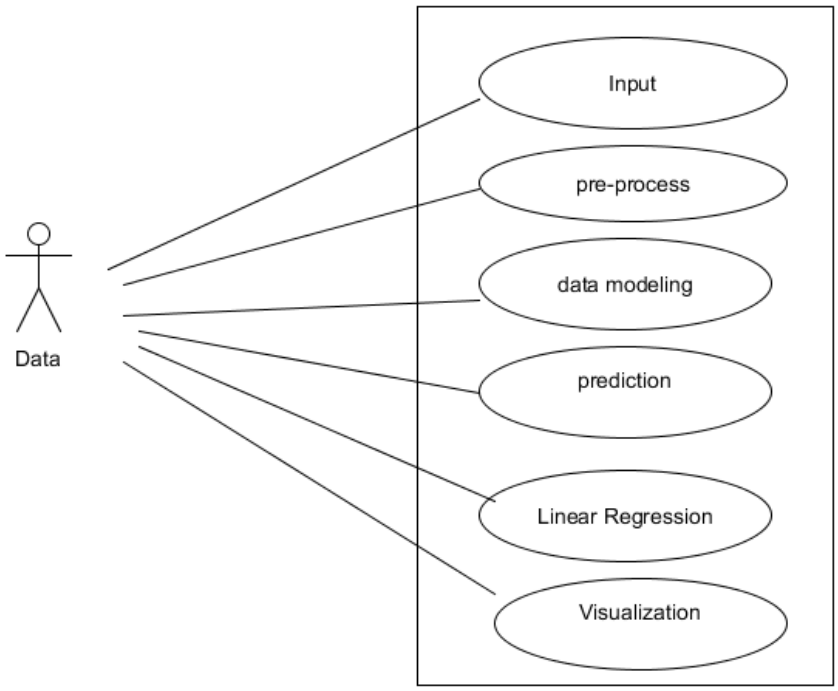


Fig.4.3.5 Use Case Diagram

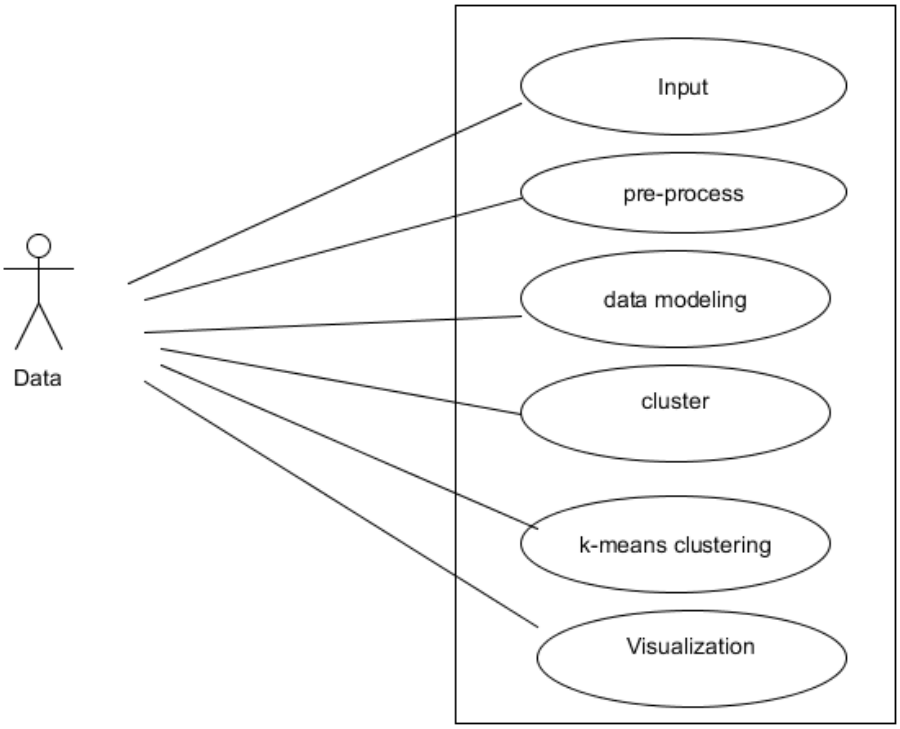


Fig.4.3.6 Use Case Diagram

4.4 SEQUENCE DIAGRAM:

Sequence illustrations show collaborations between tasks as a means of exchanging dispatches throughout time. They are additionally referred to as event maps. A grouping map is an excellent tool for examining and approving unique runtime scenarios. These can be useful for anticipating how a frame will behave and for demonstrating any outstanding orders during the time spent presenting other frames.

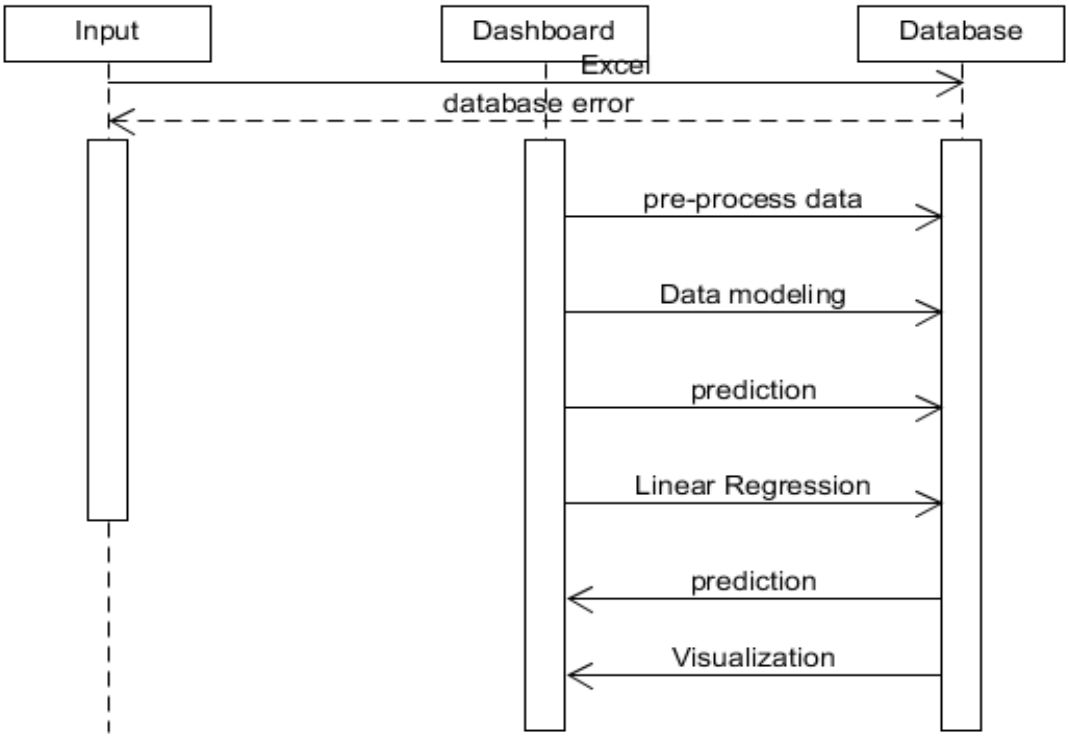


Fig.4.3.7 Sequence Diagram

Hardware Requirements:

Ram: 512 Mb

Hard disk: 50GB

Software Requirements:

Language: Java

Tools : JSP and Servlets

Front end :HTML,CSS

Back end :MySQL

CHAPTER 5

5 IMPLEMENTATION

5.1 PROJECT SCOPE

5.1.1 Work Flow Diagram

The strategy we used for this study adheres to the conventional data analysis steps.

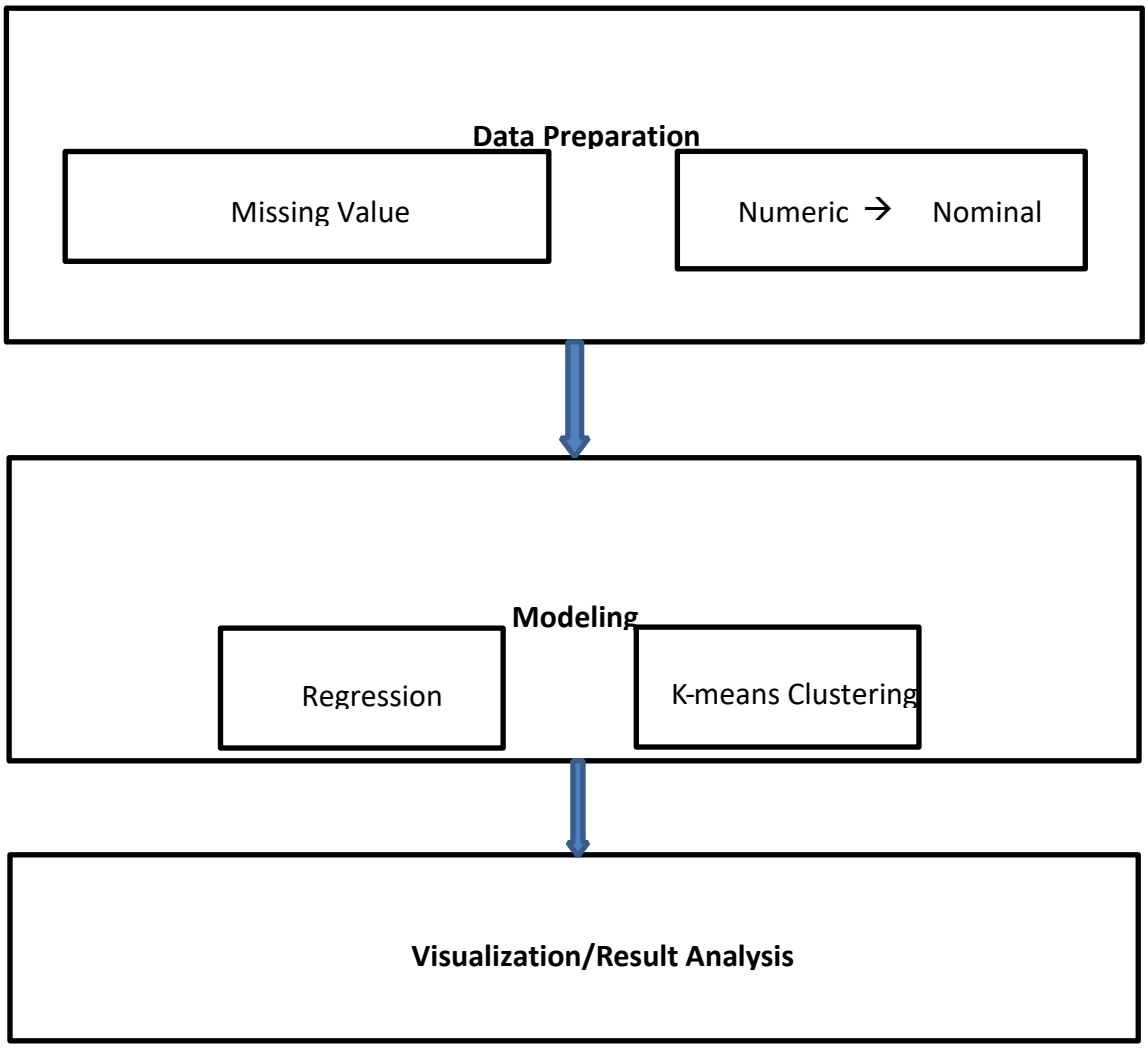


Fig.5.1 System Architecture

5.2 DATA PREPARATION

- Data instruction is completed before every version is built. Data with missing values, which are typically represented by using zero inside the dataset, were removed from all NAMED properties. As per the statistics dictionary, all numerical data had been Translated to nominal prices.
- **Missing values**

Occurs while the no statistics price is saved for the observation

Example:

States	TJunction	YJunction	FourArm	Round	RailCross	Year
West Bengal	3709	1240	4384	2265	935	2009
A & N Islands	3	0	2	12	2	2009
Chandigarh	67	14	68	18	0	2009
D & N Haveli	0	0	0	0	0	2009
Daman & Diu	0	0	0	0	0	2009
Delhi	0	0	0	0	0	2009
Lakshadweep	0	0	0	0	0	2009

Fig.5.1.1 Missing Value

Modeling

In order to identify the basic characteristics of the dangerous s, we first generate a large amount of data from the dataset. We also imposed Regression and grouping connections for a number of the qualities and patterns.

Result Analysis

The grouping of important industries within INDIA's mineral output, the classification of locations as high- or low-risk, and proposals for collaborative research all reflect the findings of our analysis. We used the statistics application Highcharts to do the review.

INPUTS

Parameters

- Bobby
- Garnet
- Vermiculite
- Magnesite
- Lime monument
- Kya nite
- Zinc
- Iron ore
- Chromite
- bauxite

DATA MAINTENANCE

gathering, processing, filtering, sorting, storing, reclaiming, issuing, transmitting, and simplifying database entries are all done by the data base service provider.

The information gathered from the gov.in website is abridged information that is saved in CSV format.

With JSON's aid, the MySQL database receives the imported CSV data (JavaScript Object memorandum).

Implementation of algorithm

Regression Algorithm

In order to predict the Future values of the PREDICTED variable, a retrogression regulations is created to DESCRY any MAJOR courting between a UNBIASED and PREDICTED variable. In order to predict how variables will behave in the future, a retrogression designs the HEREAFTER courtship among them. The algorithm mainly PREDICATED Totally at the statistical set gathered for the project using direct regression methodologies. The direct regression process makes it possible to predict the mpi's financial behaviour with the aid of statistical models. The set of rules uses the Variables y = b0 b1 * x to predict the FORTUNE conduct and describes the propose and DISUNION cost of the PREDICATED variables.

Way involved

1. Ignore the trade data and execute the operation using insufficient values.

2. Determine the mean and disunion.

3. Calculate covariance.

4. Measure pieces out.

5. Create predictions.

6. Calculate insurance.

K-Means Clustering

The high- and low-frequency powers were investigated using the K-way clustered set of rules. The set of rules uses a predetermined range of clusters (based on good clusters) to describe a given set of facts in a simple and fluid manner. One ok centroid for each cluster is the main objective. M Due to unique energy-related outcomes, these centroids should be positioned foxyly. The better option is to place them as far apart from one another as you possibly can. The next step is to pair each element from the given facts set with the nearest centroid. When there are no pending factors, step one is complete, and an early institute age is implemented. Recalculate the correct new centroids at this point to account for the bary facilities added to the clusters in the previous phase.

After we've got those ok new centroids, a brand new binding must be carried out among the equal facts set factors and the closest new centroid. A loop has been generated. As a end result of this loop we can also additionally note that the ok centroids alternate their energy grade by grade till no greater modifications are carried out.

Steps Involved:

1. Using the elements that can be clustered, add K factors to the distance that is being represented. These elements serve as initial institution centroids.

2. Distribute each item to the organisation with the closest centroid.

3. When all items were assigned, recalculate the positions of the K centroids.

4. Continue performing Steps 2 and 3 until the centroids stop moving. As a result, the objects are divided up into corporations, where the metric that needs to be decreased can be calculated.

Analysis

• The Clean Data was saved in CSV format, and with the aid of JSON, the CSV data was imported into MySQL. The tools for reading facts are now ready to be used for analysis.

• Analysis of MPI takes into account industries including metallurgy, energy, transportation, marine transportation, agriculture, and media, among others.

Based on Analysis of 2010-2021 facts units every of the above parameters are analyzed the use of the Appropriate Algorithm

VISUALIZATION

• Prediction of Dashboard for all of the above.

• The year sensible datasets from 2010-2021 are covered withinside the prediction dashboard as bar graphs, pie charts, etc.

• The dashboard is greater use complete for the respective government that assist them to recognize approximately the that goings to manifest with inside the present day years, in an effort to take a few preventive measures.

• The Application ought to produce the every year record approximately the MPI .

• Year sensible evaluation of MPI .

Training set

In our assignment anything facts we've got accrued from the year 2011 to 2021 are given enter to the linear regression set of rules, the enter incorporates the year, price and the parameters whichever we're going to use primarily based totally on that it's miles going to go back the end result for 2022. In this assignment we aren't schooling dataset, we're fetching datasets and returning end result for 2022.

Project Scope:

The project is finished the usage of java it truly is an object oriented programming language. Object orientated programming is an technique that gives a manner of modularizing software with the aid of using growing partitioned reminiscence place of each facts and feature that may be used as a template for growing copies of such module on demand.

Java programming language is used to complete this assignment. An online software is made using both JSP and servlet technology. Servlets are precompiled Java applications that may provide dynamic web content. The servlet API has a wide range of interfaces and sophistication, including Httpservlet, servlet query, servlet response, etc. JSP has been used to develop web applications that are known as servlets. Given that it offers more functionality than servlet, it might be considered a servlet extender. As a backend, MySQL server is utilised.

OVERVIEW

In order to predict the future readings of the structured variable, a regression set of rules seeks to identify the historical relationship between an unstructured and a structured variable. A regression creates a more intimate relationship between variables to predict their future behaviour.

The high- and low-frequency powers were investigated using the K-way grouping set of rules. Additionally, they employed affiliate rule mining to comprehend the relationships between the many parameters related with MPI visits at various sites with variable occurrences.

RESULT BASED ON PARAMETERS

The collection of rules is created to obtain the desired outcome depending on the criteria, which include energy, media, transport, agriculture, etc.

The multivariate set of policies makes it possible to forecast MPI's future behaviour while taking a number of factors into account. The dashboard offers a variety of Sectors and Types; based entirely on the choice, a set of rules applies statistical methodology to the prior data sets (2011-2021).

And count on the give up end result of 2022.

Algorithm to depict ordinary steps:

Algorithm for Regression Techniques

Input: MPI facts set from 2001-2021

Parameters

• Copper

• Garnet

• Vermiculite

• Magnesite

• Lime stone

• Kya nite

• Zinc

• Iron ore

• Chromite

• bauxite

Output: Predicting the Result of 2022 shown with Bar graph.

Procedure:

Step 1: scan the transaction dataset.

Step 2: Managing missing information (all documents with data point is symbolised as 0 with in set of data)

Step 3: Calculate the training data's mean and variance for the two inputs.

Step 4: Save the x-mean axis's and variance results.

Step 5: Store the y-axis result in a same manner.

Step 6: Compute the covariance using the mean values obtained from the x and y axes.

Step 7: Using the covariance, get the coefficient values for variables b1 and b0.

Step 8: After the coefficients have been calculated, make the forecast.

$$Y=b0+b1(x).$$

K-means clustering

Input: MPI data set from 2010-2021

Parameters

- Copper
- Garnet
- Vermiculite
- Magnesite
- Lime stone
- Kya nite
- Zinc
- Iron ore
- Chromite
- bauxite

Output: cluster of high and low frequency according to Sector wise.

Procedure:

First, go over the transaction data set.

Step 2: Know the average number based on the overall number of Sectors and s that occur.

Step 3: Pick the "C" cluster centre, assuming that the average value equals centroid "c."

Calculate the separation between every data point and the cluster centres .

step 4:The data point should be assigned to the cluster centre that is the closest to it among all the cluster centres .

step 5:Recalculate the initial cluster centre in Step 6 using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, *'c_i'* represents the number of data points in i^{th} cluster.

Step 7: Recalculate the distance between each data point and new obtained cluster centers.

Step 8: If no data point was reassigned then stop, otherwise repeat from step 5).

5.3 PSEUDO CODE FOR ALGORITHM USED

5.3.1 Regression Algorithm

X [] □ x-axis values

Y[] □ y-axis values

//Initialization

N □ length of x-axis

//First pass

//Initialization

Sumx □ empty

Sumy □ empty

Sumx2 □ empty

// first pass

For loop Begins

Initialise i □0 to until i less than n

Sumx □ sumx+ x[]

Sumy □ Sumy+ y[]

Sumx2 □ x[]*y[]

For loop ends

//second pass

```
// initialize          xxbar □ 0          yybar
□ 0          xybar□ 0.0;

for loop begins

 initialize  i □ 0 to  i  less than  n

  xxbar  =  xxbar+   (x[] - xbar) * (x[] -
xbar);                    yybar =  yybar+(y[]
- ybar) * (y[] - ybar);                    xybar
=  xybar+  (x[i]  -  xbar)  *  (y[i]  -  ybar);
for loop ends

  slope  = xybar / xxbar;

        intercept = ybar - slope * xbar;



        // more statistical analysis

     double rss = 0.0;     // residual sum
of squares         double ssr = 0.0;      //
regression  sum  of  squares            for
loop begins

double fit = slope*x[i] + intercept;

    rss += (fit - y[i]) * (fit - y[i]);
```

ssr += (fit - ybar) * (fit - ybar);

for loop end end

Example

Input:

Consider the data sets of Alcohol from 2009 to 2021

Table:

X(year)	Y(value)
2008	3496
2009	3500
2010	3987
2011	2987
2012	3019
2013	3999
2014	4015
2015	4786
2016	4018
2021	4445

//calculate the mean of X and Y

Sum of x=20125

Sum of y=38252

Mean of X=20125/10 2012.5

Mean of Y=38252/10 3825.2

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
2009	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
2021	4445	5	620	3100	25	384400
Total		5	2	10114	85	3003536

Fig 5.1.Stastical Table

Y=bO + b1(x) // x is prediction value or independent

value b1= (A1 *B1)/(A1)2

10114/85 □ 118.98

bo= mean of y –bo(mean of x)

=3825.5-(118.98*2012)

-235600.76

Y=b0 + b1(X)

Y=-235600.76+ 118.8(2022)

Y=4500 /// This is the predict value of 2022

Pseudo code for k-means clustering

//initialize

X □ 0 //count the

number of Sectors A[] □

load sub categories a□

A[0] b□ A[1] do

Cluster1[] □ length

Cluster2[] □ length

For loop begin

Initialize i□ 0 to i less than length

If A[i] <= A[i+1]

Cluster1[] □ A[]

Else

Cluster2[]□ A[]

For loop end

// Initialize

For loop begin

Initialize i □ 0; i less than length sum1 □sum1 +cluster1[i];

For loop end

for loop begin initialize i □ 0 to i les than length sum2 □ sum2 + cluster2[i];

for loop end

,mean1 □ sum1/k //printing centroid

Mean2 □ sum2/k While

end

add[cluster1] //high frequency Sectors add[cluster2] //low frequency Sectors end

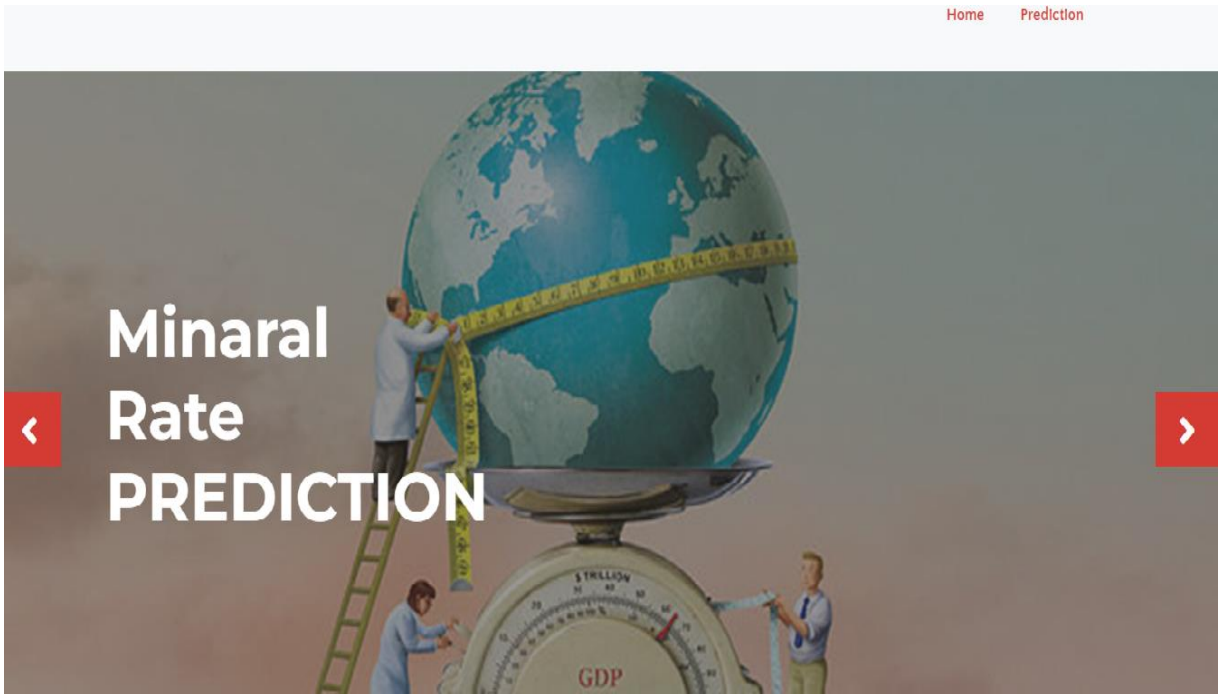


Fig.5.2.1:HOME SCREEN

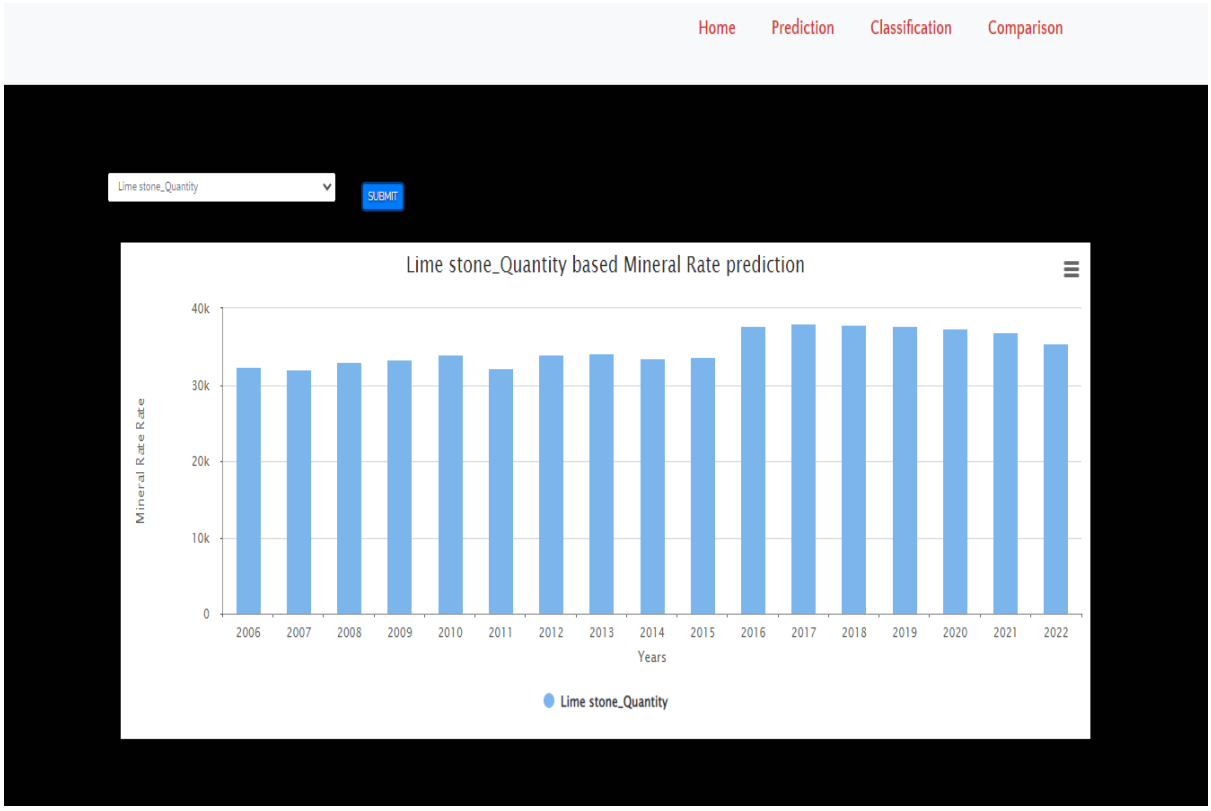


Fig 5.2.2 LIMESTONE_QUANTITY BASED MINERAL RATE PREDICTION

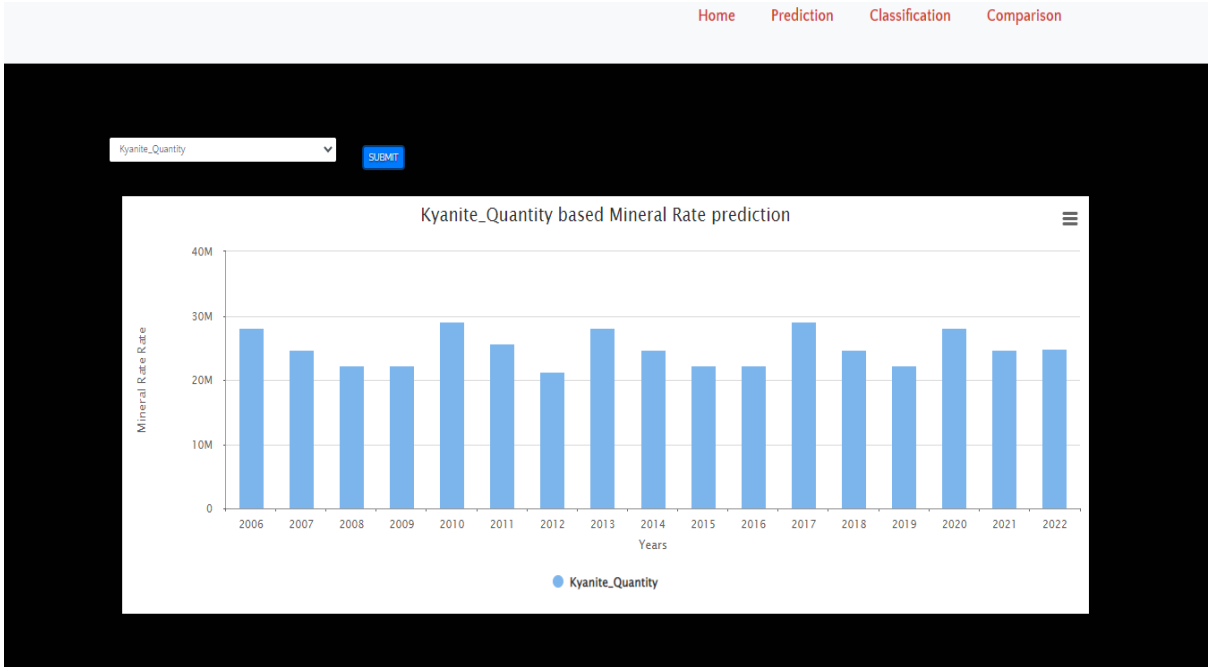


Fig.5.3.3 KYANITE_QUANTITY BASED MINERAL RATE PREDICTION

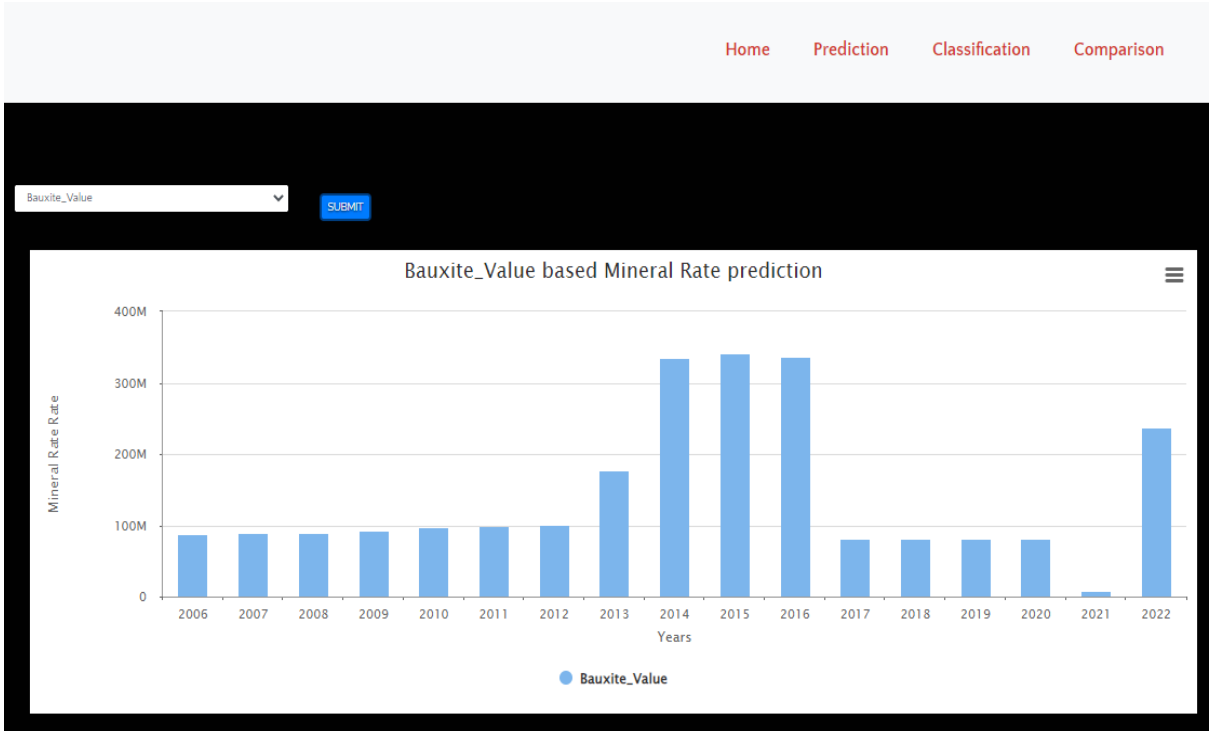


Fig.5.3.4 BAUXITE_VALUE BASED MINERAL RATE PREDICTION

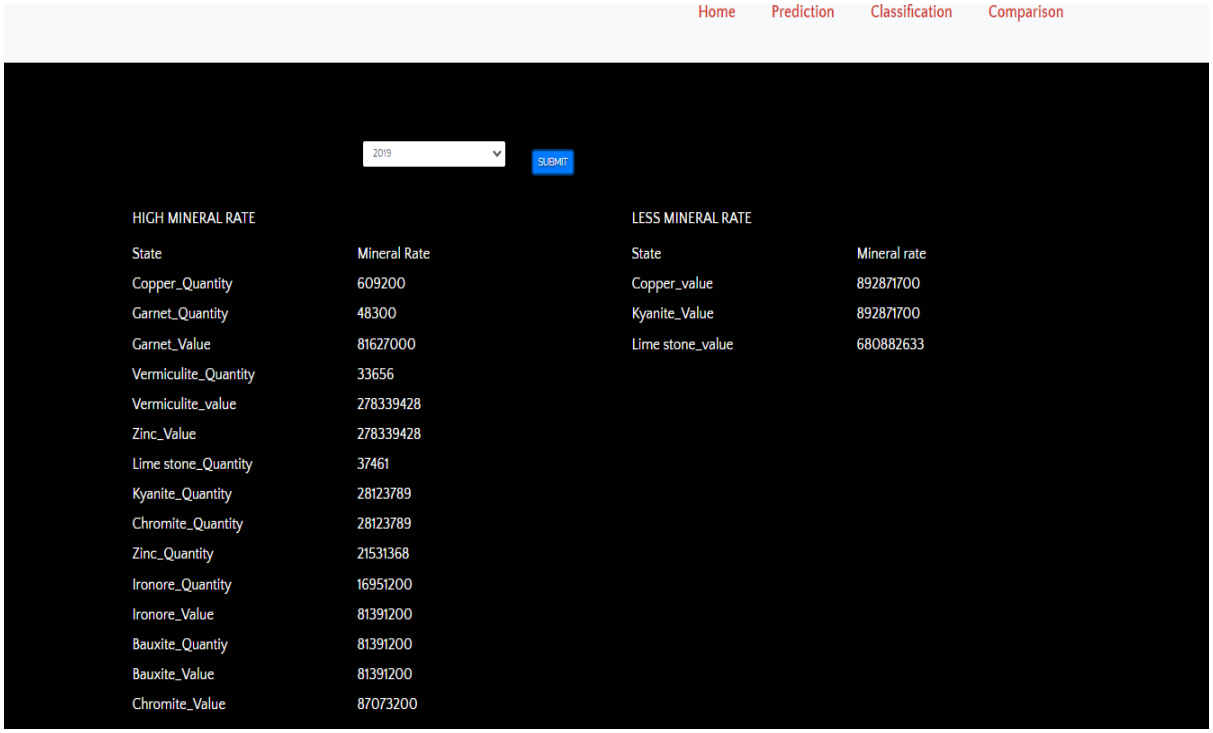


Fig.5.3.5 CLASSIFICATION OF HIGH AND LESS MINERAL RATE IN 2008

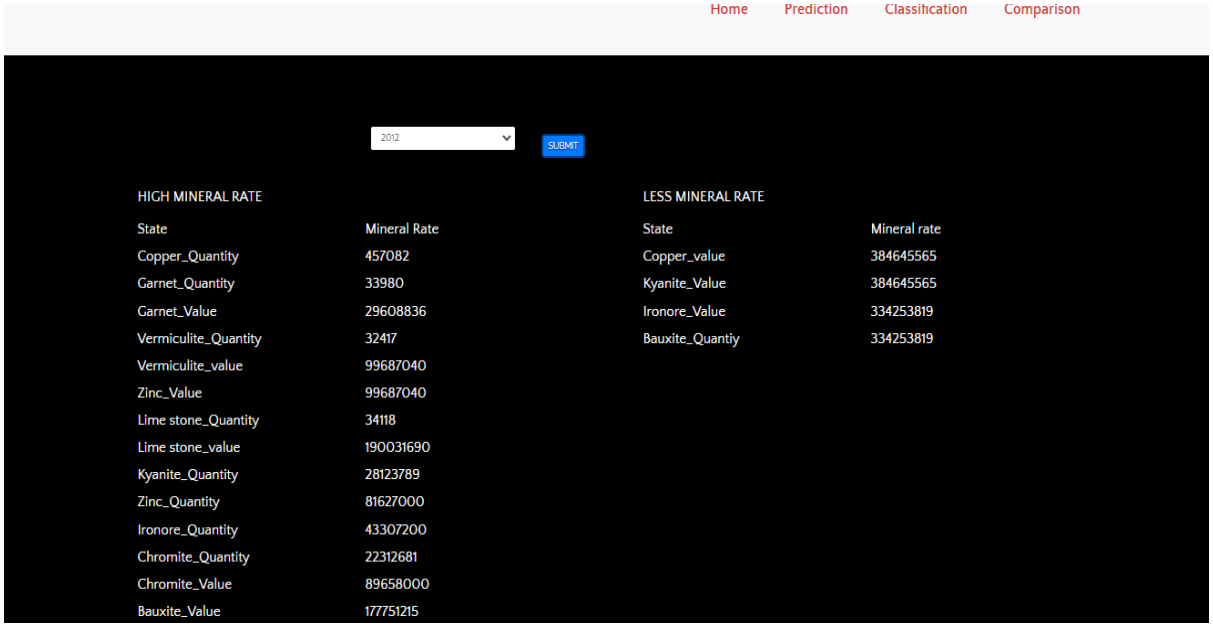


Fig.5.3.6 CLASSIFICATION OF HIGH AND LESS MINERAL RATE IN 2010

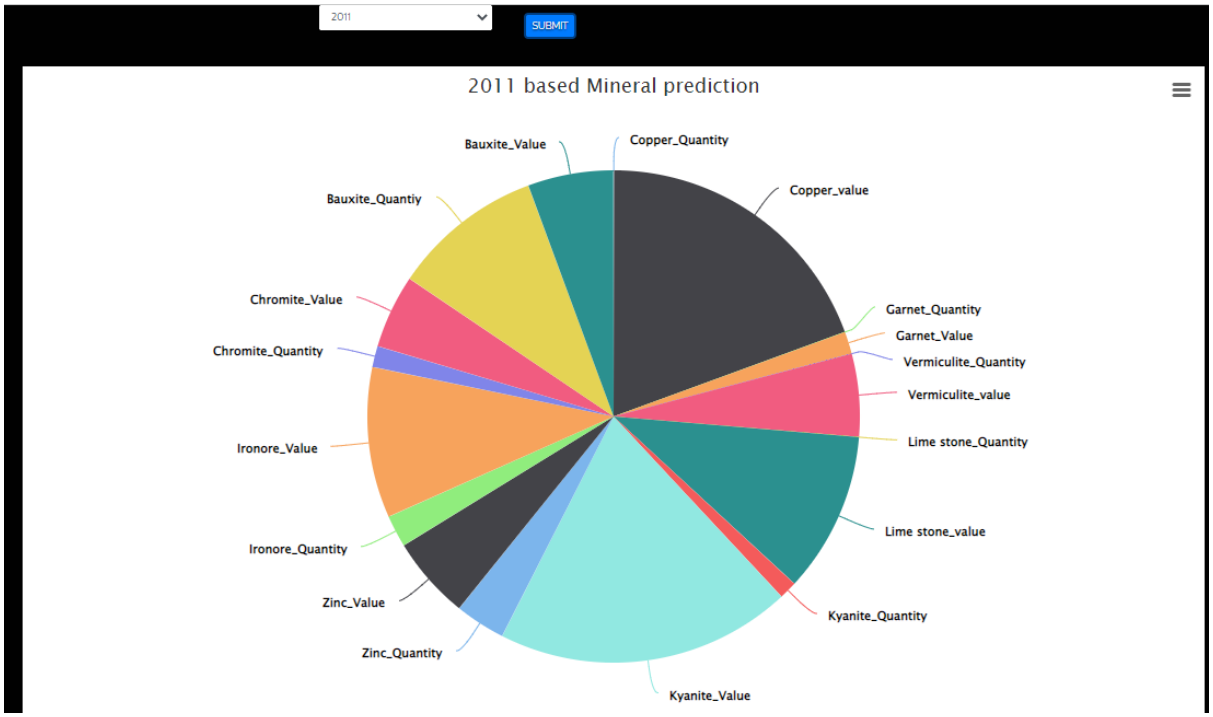


Fig.5.3.7 2011 BASED MINERAL PREDICTION

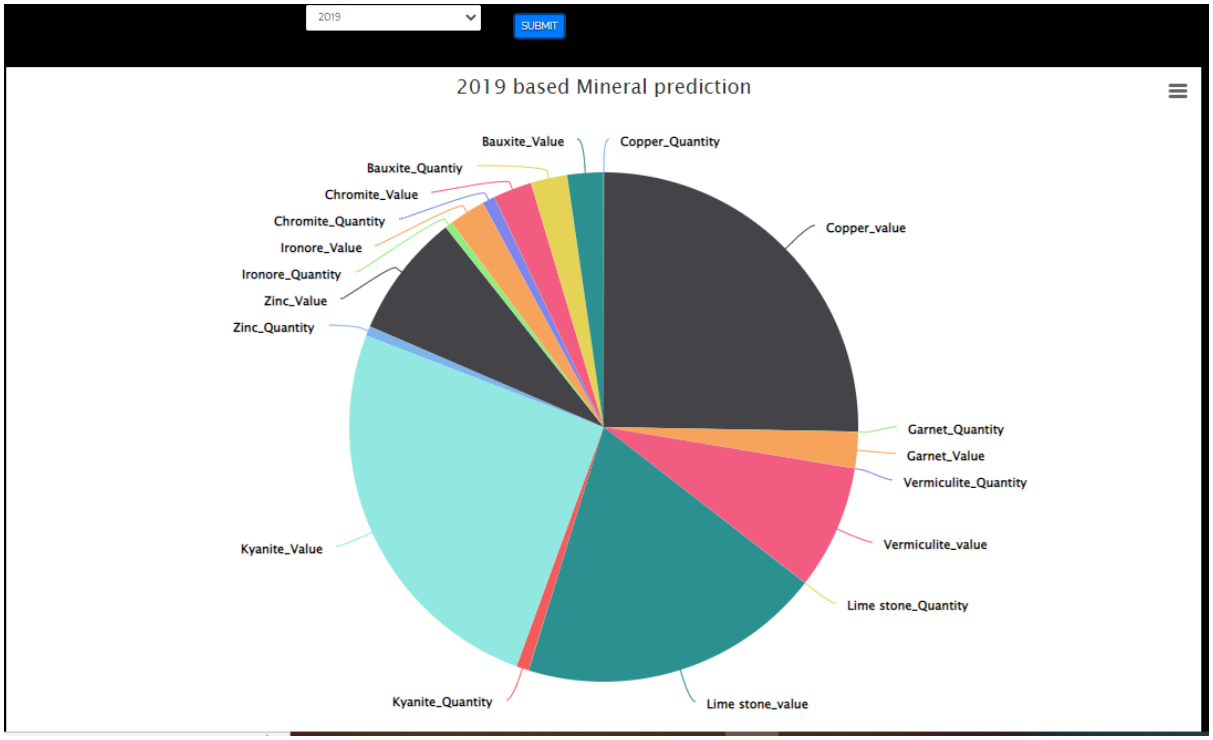


Fig.5.3.8 2019 BASED MINERAL PREDICTION

CHAPTER 6

6 EXPERIMENTAL RESULT AND ANALYSIS

6.1 EXPERIMENTAL SETUP

To calculate the performance of the proposed algorithms, experiments are conducted against five variants data sets(**Parameters**

- Copper
- Garnet
- Vermiculite
- Magnesite
- Lime stone
- Kya nite
- Zinc
- Iron ore
- Chromite
- bauxite

based). shows the characteristics of data set. Experiments were performed on a computer with 2.20 GHz Intel Core processor and 4gb of memory, running windows 8.1.All the Algorithms implemented in java.

6.2 RESULT BASED ON ALGORITHMS

Experiments conducted in this section shows the different output of each algorithm for different input parameters. Experiments are conducted based on the two Algorithms those are Linear Regression and K-means clustering Algorithms.

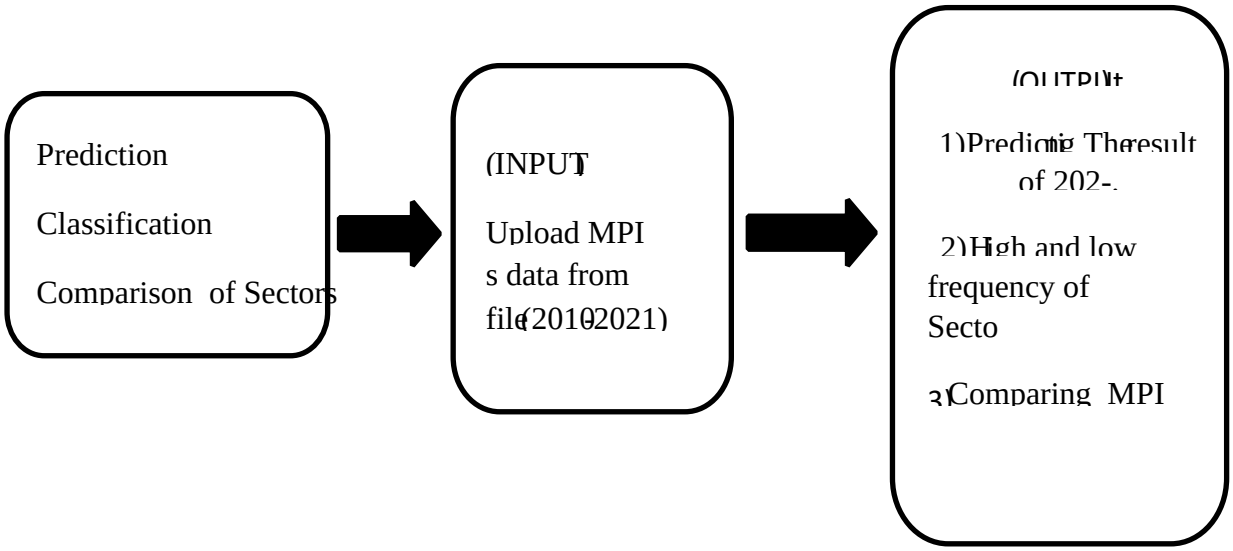


Fig.6.2 OVERALL DIAGRAMMATIC REPRESENTATION OF ALGORITHMS.

CHAPTER 7

7 TESTING

7.1 INTRODUCTION

Testing is the manner closer to strolling a framework with the expectancy of coming across mistakes. Testing enhancements the uprightness of the framework with the aid of using distinguishing the deviations in plans and mistakes withinside the framework. Testing objectives distinguishing mistakes – promenade zones. This aides withinside the avoidance of errors withinside the framework. Testing moreover provides esteems to the object with the aid of using asserting the patron's necessity.

The number one purpose is to differentiate mistakes and error get-promenade zones in a framework. Testing should be in depth and all round arranged. A truly attempted framework is as horrible as an untested framework. Furthermore, the value of an untested and under-attempted framework is high. The execution is the ultimate and full-size level. It consists of patron preparation, framework trying out that allows you to assure the powerful strolling of the proposed framework. The patron exams the framework and modifications are made with the aid of using their requirements. The trying out consists of the trying out of the created framework utilising exclusive kinds of statistics. While trying out, mistakes are stated and rightness is the mode.

7.2 OBJECTIVES OF TESTING

• Testing in a cycle of executing a software with the expectancy of coming across errors.

• A powerful test is one which exhibits an up 'til now unusual blunder.

Framework trying out is a section of usage, that is pointed closer to ensuring that the framework works correctly and productively in step with the patron's want earlier than the stay pastime initiates. As expressed previously, trying out is necessary to the fulfillment of a framework. Framework trying out makes the coherent presumption that if all of the framework is right, the goal might be efficiently accomplished. A development of exams are finished earlier than the framework is ready for the patron acknowledgment take a look at.

7.3 TESTING METHODS

System trying out is a level of implementation. This facilitates the climate device works correctly and efficaciously earlier than stay operation commences. Testing is important to the

achievement of the device. The candidate device is concern to a whole lot of exams: on line response, volume, stress, recovery, security, and usefulness exams collection of exams are finished for the proposed device are prepared for consumer attractiveness trying out.

7.3.1 White Box Testing

The take a look at is performed at some stage in the code technology section itself. All the mistakes have been rectified in the interim of its discovery. During this trying out, it's miles ensured that

• All impartial module were exercised at the least one

• Exercise all logical selections on their proper or fake side.

• Execute all loops at their boundaries.

7.3.2 Black Box Testing

It is focused on the product's reasonable requirements. It is not, however, an alternative to white box testing; rather, it is a reciprocal method that may reveal a greater magnitude of errors than White Box procedures. The following classes have errors that are being looked for.• Incorrect or lacking capacities

• interface errors

• Errors in an statistics shape or out of doors statistics base access

7.4 UNIT TESTING

Unit trying out mainly facilities across the littlest unit of programming plan. This is referred to as module trying out. The modules are attempted independently. The take a look at is completed at some stage in the programming level itself. In this development, each module is observed to be operating acceptably as respects the ordinary yield from the module

7.5 INTEGRATION TESTING

Mix trying out is a good technique for growing this system shape, whilst concurrently main exams to show mistakes associated with the interface. The intention is to take unit attempted modules and manufacture a software shape. All the modules are joined and attempted in general.

7.6 OUTPUT TESTING

Subsequent to appearing approval trying out, the subsequent level is yield attempting of the proposed framework, considering no framework may be treasured at the off hazard that it does not create the important yield in a specific configuration. The yield layout at the display is observed to be right. The agency turned into deliberate withinside the framework configuration time as indicated with the aid of using the patron needs. For the published reproduction likewise, the yield comes in step with the predefined conditions with the aid of using the patron. Subsequently yield trying out failed to result in any change for the framework.

7.6.1 User Acceptance Testing

Client acknowledgment of a framework is the important issue for the fulfillment of any framework. The framework feasible is attempted for patron acknowledgment with the aid of using usually staying in touch with the approaching framework customers on the hour of making and making modifications at something factor required.

7.7 VALIDATION

Toward the consummation of the reconciliation trying out, the product is completely collected as package deal interfacing mistakes were found out and changed and a ultimate association of programming exams begins offevolved in approval trying out. Approval trying out may be characterised from more than one factors of view, but a sincere definition is that the approval succeeds while the product paintings in a manner this is ordinary with the aid of using the patron. After approval take a look at has been directed as follows:

• The ability or execution characteristics modify to element and are acknowledged.

• A deviation from the precise is found out and a scarcity listing is made.

• Proposed framework feasible has been attempted with the aid of using utilising an approval take a look at and observed to be operating acceptably.

The customers take a look at the advanced device while modifications are made in step with the needs. The trying out section includes the trying out of the advanced device the use of diverse sorts of records. An tricky trying out of records is ready and device is examined the use of the take a look at records. Test instances are used to test for outputs with exclusive set of inputs.

Test cases with positive scenarios:

TC No	scenario	Required Input	Expected output	Actual output	Test Result
1	Enter Prediction values	Enter a valid values	Should predicted successfully	predicted successfully	Pass
2	Enter clustering values	Sectr,year, Type	Should cluster successfully	cluster successfully	Pass
3	Enter Input fr comparison values	Enter a valid values	Pie chart is shwn successfully	predicted successfully	Pass
4	Enter Input fr comparison values	Enter a valid values	Pie chart is shwn successfully	Nt displaying graph , library issue	Pass
5	Enter Prediction values	Enter a valid values	Should predicted successfully	Database error	fail
6	Enter clustering values	Sector,year, Type	Should cluster successfully	Database error	Fail

Table.7.4 Test cases

CHAPTER 8

8 CONCLUSION AND FUTURE ENHANCEMENTS

8.1 CONCLUSION

8.1.1 According to statistics, regression analysis, and classification, environmental elements such the coal industry, power, media, and transportation have produced results that are inconsistent with predictions, with some results showing low MPI rates and others showing high media rates.

8.2 SCOPE FOR FUTURE WORK

The web Application can be deployed into android application so it can be available for mobile platform and the application can be used by all the user.

9 BIBILOGRAPHY

[1]. An analysis of the regression algorithm to forecast admin, production, and delivery accuracy of products in the cosmetic industry Work will look Nusa Mandiri, Jakarta, Australia W Gata1, H B Novitasari2, R Nurfalah3, R Hernawati4, and M J Shidiq5

[2]. An analysis of the regression algorithm to forecast administration, production, and time - effective of products in the cosmetic industry Investors in investing Nusa Mandiri, Jakarta, Australia W Gata1, H B Novitasari2, R Nurfalah3, R Hernawati4, and M J Shidiq5

[3].Comparison of support vector regression versus linear regression IEEE Paper by Kavitha Ramya R, Kavitha S, and Varuna S

[4].Douglas C. Monty, Elizabeth A. Peck, and G. Geoffrey Vining's Introduction to Linear Regression, Fifth Edition