CRIME ANALYSIS SYSTEM

ASWATHA NARAYANAN S.(2015103546) **JAI SURYA R.K.**(2015103511) **HARISH.S**(2015103553)

ABSTRACT

In today's world, Security is an aspect that is given high priority by Government to reduce Crime Incidence. As data mining is an appropriate field to apply on high volume crime dataset and knowledge gained through this will be useful for Police Force.

OBJECTIVE

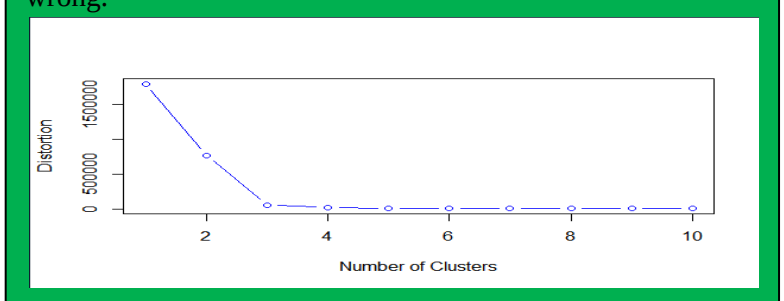
The objective of this application is to help the law enforcement to analyze on the Crime frequency in a given region and also gives information about the Highest Criticality Levels of the same, thereby deploying resources in an effective manner.

RESULT ANALYSIS

	A	B	C	D	E	F	G	H	I	J	K
1	STATES	ANTI_SOC	BICYCLE_T	BURGLARY	CRIMINAL	DRUGS	OTHER_CR	OTHER_TH	POSSESSIC	PUBLIC_O	ROBBERY
2	AMBER VA	300	2	68	65	19	5	44	3	8	2
3	ASHFIELD	4	0	1	0	1	0	0	0	0	0
4	BARNSLEY	0	0	0	0	0	1	0	0	0	0
5	BASSETLAV	0	0	0	0	0	1	0	0	0	0
6	BLACKPOO	0	0	0	0	0	0	0	0	0	0
7	BOLSOVER	175	2	32	51	13	4	43	4	8	0
8	BRADFORD	0	0	0	0	0	0	0	0	0	0
9	BROXTOW	5	0	0	0	1	0	0	0	0	0
10	CHESTERF	324	0	45	83	14	6	62	6	16	2

The data transformation is helpful in converting the raw dataset into suitable form for analysis of crime analysis.

The data preprocessing module is useful for detecting and eliminating outliers, so that the data mining tasks doesn't go wrong.



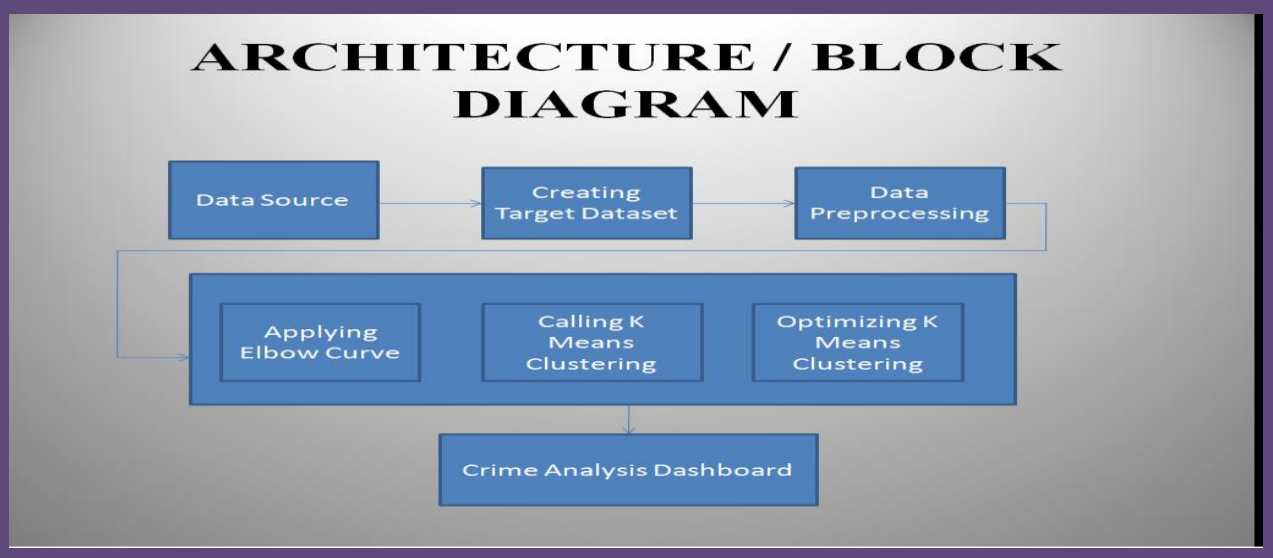
The crime data mining tasks helps to choose the optimal value of K from the elbow curve.

MAJOR REFERENCES

[1] J. Agarwal ,R . Nagpal and R. Sehgal, "Crime analysis using k-means Clustering", *International Journal of Computer Applications*, vol. 83, no. 4, pp.1-4,2013

[2] S.Sivaranjani,S.Sivakumari,S.Maragatham, "GIS based serial crime analysis using data mining techniques",vol.153,no. 8,pp. 19-23,2016.

SYSTEM DESIGN



ALGORITHM

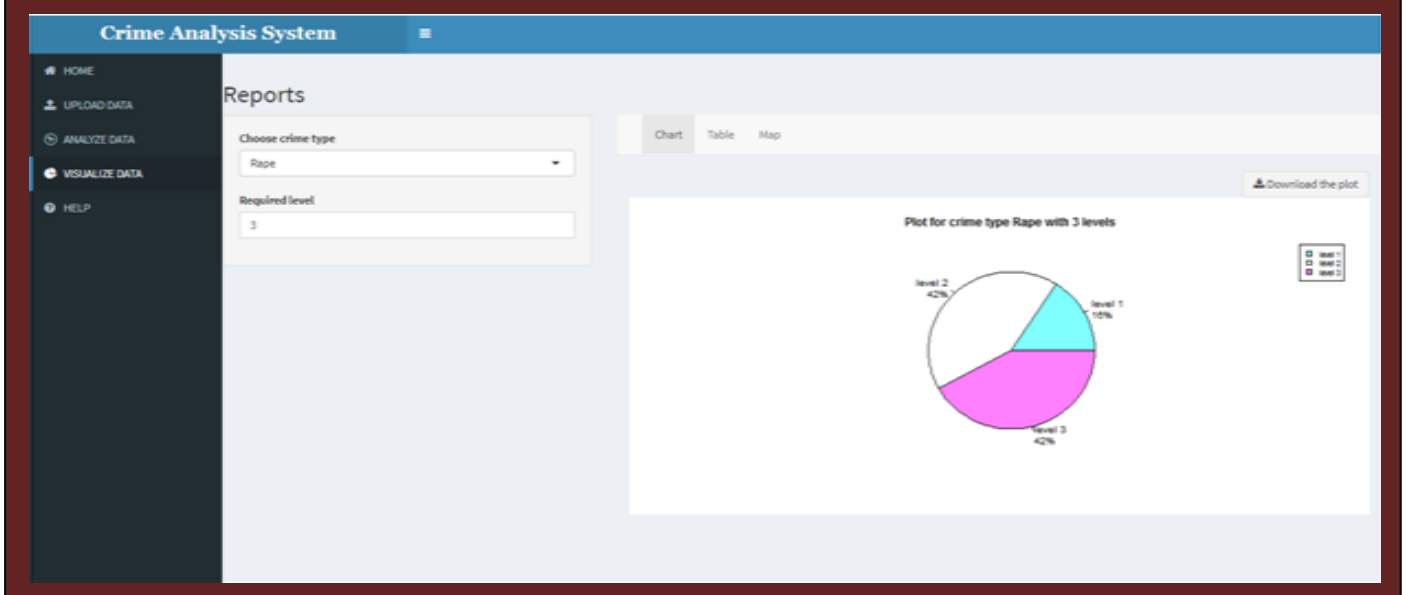
- Randomly choose 'K' examples as initial centroids
- While true:
  - Create 'K' clusters by assigning each example to closest centroid
  - Compute K new centroids by averaging examples in each cluster
  - If centroids don't change:
    - break

LIST OF MODULES

- DATA TRANSFORMATION
- DATA PREPROCESSING
- CRIME DATA MINING TASKS
- CRIME ANALYSIS DASHBOARD

```
K-means clustering with 4 clusters of sizes 23, 1, 4, 3

Cluster means:
  ANTI_SOCIAL_BEHAVIOUR BICYCLE_THEFT    BURGLARY CRIMINAL_DAMAGE_AND_ARSON      DRUGS
1              3.478261     0.04347826   1.478261                   1.391304  0.2173913
2           1052.000000    56.00000000 182.000000                 247.000000 77.0000000
3            183.250000     1.75000000  39.500000                  41.750000  8.2500000
4            303.000000     3.33333333  65.333333                  86.333333 16.0000000
  OTHER_CRIME OTHER_THEFT POSSESSION_OF_WEAPONS PUBLIC_ORDER   ROBBERY SHOPLIFTING
1   0.6956522    0.826087                0.000000    0.08695652  0.000000   0.5652174
2  27.0000000  227.000000               15.000000   50.00000000 38.000000 214.0000000
3   8.2500000   31.750000                2.250000    6.25000000  1.750000  24.5000000
4   6.6666667   55.666667                4.666667   12.00000000  2.333333  65.0000000
  THEFT_FROM_PERSON VEHICLE_CRIME VIOLENCE_AND_SEXUAL_OFFENCES
1        0.04347826      3.652174                    1.565217
2       36.00000000    245.000000                  430.000000
3        1.75000000     50.000000                   71.500000
4        6.33333333     58.000000                  112.000000

Clustering vector:
 [1] 4 1 1 1 1 3 1 1 4 2 1 1 4 1 3 1 1 1 1 1 3 1 1 1 1 1 3 1 1 1 1

Within cluster sum of squares by cluster:
[1] 12025.304     0.000  4730.500  4215.333
 (between_SS / total_SS =  98.8 %)

Available components:
 [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
 [6] "betweenss"    "size"         "iter"         "ifault"       "inicial.centers"
```

The Crime Data Mining tasks module describes the cluster means of each type of crime in each cluster and the cluster assignment of each observation.Also, it gives the ratio between between_ss and total_ss.



The Crime Analysis Dashboard provides the dashboard for the Crime analysis system, with which the user can interact. The above plot gives the information about highest critical rate level of given crime type, meaning level 1 is the highest frequency of crime and subsequent levels represent next higher level of crime criticality rate.

CONCLUSION AND FUTURE WORK

The user is provided with an Application portal for the Crime Analysis. It includes interactive features, user-friendly interface, that user can play around. This portal provides a dynamic clustering of the crime data points on the interface, and also visualizes the result in two forms namely, Pie Chart and Tables for easy understanding. This project can be further extended by using some advanced clustering algorithms to increase crime analysis accuracy and to enhance overall performance.

Submitted for CS7711- Creative & Innovative Project Laboratory (Group 1) Course Instructor: V Suganya