PHASE – 4

(CAPSTONE PROJECT)

4.4

Capstone project overall story in your own words. Min 1000 words.

Title : Crimes in India

Intoduction:

Crimes, in all forms, presents multiple challenges to societies around the world. In the context of India, a country with rich culture and diversity, understanding the dynamics of crime is essential for ensuring justice for all. This capstone project which was given to us made us analyze, how crime is being executed in different states and districts across various years so that we can explore it’s main causes, manifestations and implications for society.

PHASE 1: Data Collection/Preparation Part (State wise)

This Phase is the most important part of the project, as it consisted of data collection and also data preparation part. We were given files that contained some information about each state in India and along with this we were also given the task to extract additional data if required for our analysis. In this phase, I extracted the data for population and literacy rate of each state from 2001-2010, so that we could analyze the relation between crime in each state vs the population and literacy rate of each state (further continued in next phase).

PHASE 2: STATE/UT wise Analysis

In this Phase, we have done analysis on all of the States and Union Territories of India based on the data collected above. Some of the analysis done are:

Analysis of Literacy Rate vs Total Crimes.

Analysis of the type of crime vs each state vs Literacy rate.

Analysis of year-on-year total crime rate.

Analysis of Population vs overall Crime.

Each State wise crime report.

PHASE 3: Analysis of data using SQL

In this Phase, we have mostly analyzed on different districts of each state on how crime affects these districts and so on.

Now in this phase, we have found out which districts or areas have the highest count of crimes (such as rape, kidnapping, murders etc.) and also find out the various crimes committed against women in each district so that we can further action on these districts to make it safer in the long run.

Also we have analyzed which District in each state/UT has the highest number of murders year wise and use this data to analyze districts that appear 3 or more than 3 years, so that necessary action can be taken in these districts.

PHASE – 4 (UNSUPERVISED LEARNING – CLUSTERING)

In this final and last phase, we collect data (district wise) from various files given and divide the data into various clusters such as:

Peaceful Area, Moderate Area and Sensitive Area

Peaceful Area – Areas with less crime rate and with more safety.

Moderate Area – Areas with neither less nor more crime rate (somewhere in the middle).

Sensitive Area – Areas with high crime rate and with less safety.

After dividing into clusters, we make observations using these each clusters . After making observations, we give conclusions for each cluster, For example, if the data belongs to ‘Peaceful area’ cluster we make conclusions saying we do not need to deploy much forces to take action to control the crime level, instead we use some of these forces to deploy them to ‘Sensitive Area’ districts, so that crime level can be controlled in these districts.