CS ASSIGNMENT REPORT

Performing LDA on estimated number of homicides committed in India (2000-2019) based on gender

Name: Saloni Sivakumar

Roll no:44

Batch 2

Btech. CSBS

- Introduction-

LDA is a statistical method used to determine relationships between a set of categorical dependent variables (Y) and continuous independent variables (X). It allows us to understand how well the independent variables discriminate between the dependent categories.

I used real time data from WHO website about the estimated number of Homicides committed by male, female or both in India in the past 19 years.

By performing LDA, I aimed to understand if the number of homicides could aid in predicting the sex of an individual.

Hence I chose sex as the Y variable and the estimated number of homicides committed and in which years 2000-2019,as the X variables.

- Procedure-

After importing various libraries in Python I uploaded the csv file containing the tabular data. I dropped columns like “Location type” which was irrelevant in my analysis. Furthermore I dropped the rows that contained NULL values.

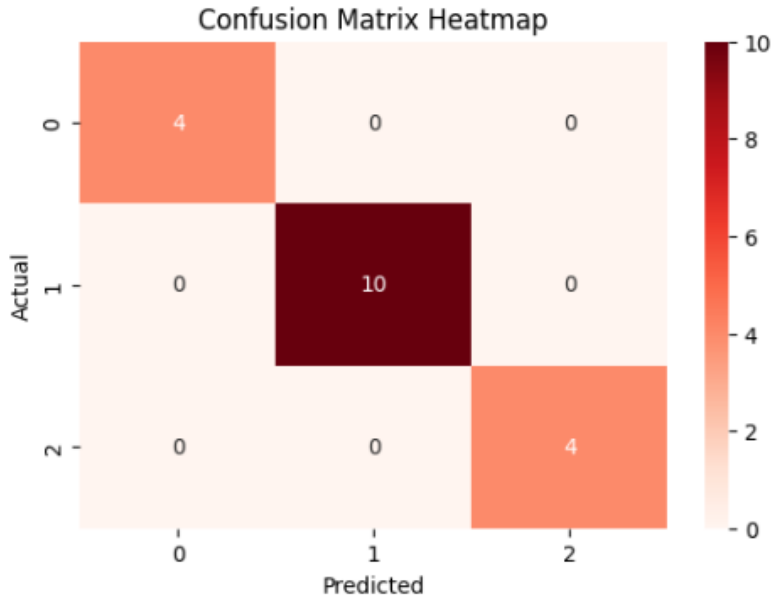
As our Y variable needs to be categorical in in LDA, I assigned 0 to males,1 to females and 2 to both sexes. I then separated the X and Y values and performed LDA on them.

- Conclusion of LDA-

Accuracy refers to the extent to which predicted classifications of a data set align with the actual or true classifications. I obtained an accuracy of 100 percent indicating that all the predictions made by the LDA model correctly matched the actual values.

A confusion matrix represents the accuracy of LDA. The confusion matrix I got was ([[4, 0, 0],[0, 10, 0],[0, 0, 4]]) which also showed that that the model has made accurate predictions for all the classes. The matrix indicates that there are 4 observations in the first class, 10 in the second class, and 4 in the third class, and all of them are correctly classified.

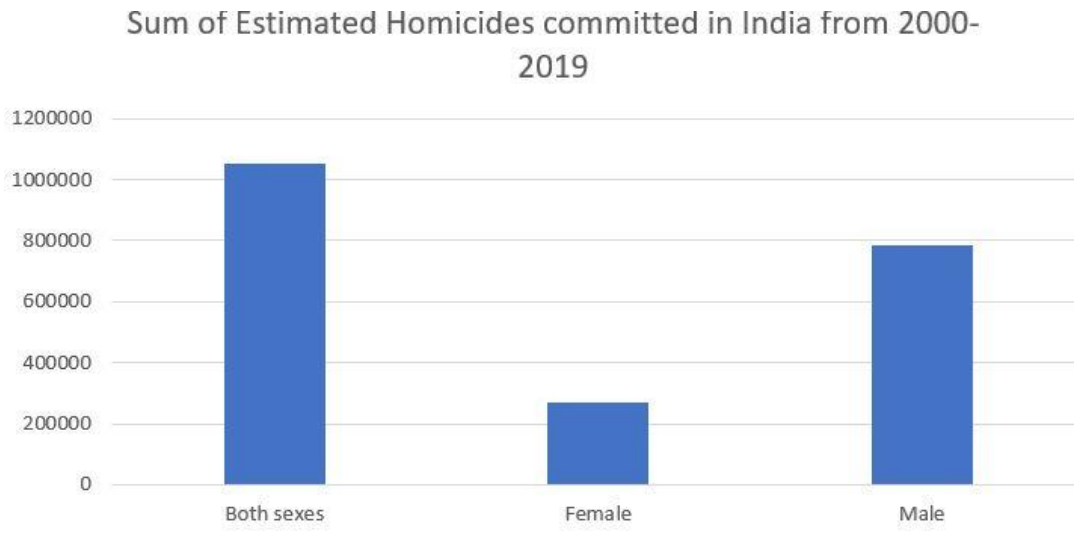
Hence the LDA model has performed perfectly for the dataset.



- Interpretation of Data set-

In India the data from recent years like 2019 and 2018 suggests that the majority of homicides were committed by men. In 2019, out of a total of 51,767 estimated homicides (excluding homicide committed by both sexes), approximately 76% (39,258) were committed by males, while females were responsible for about 24% (12,509). A similar pattern is observed in 2018, with males accounting for approximately 76% (39,502) and females for about 24% (12,477) out of a total of 51,979 homicides. A similar pattern is observed in the previous years.

This might be due to various factors such as sex ratio which is prevalent in India.



- Plagiarism Report-