UNDERSTANDING MAIN STREAM NEWS

Problem statement:

To visualize and classify the mainstream news into different categories using different Machine Learning models

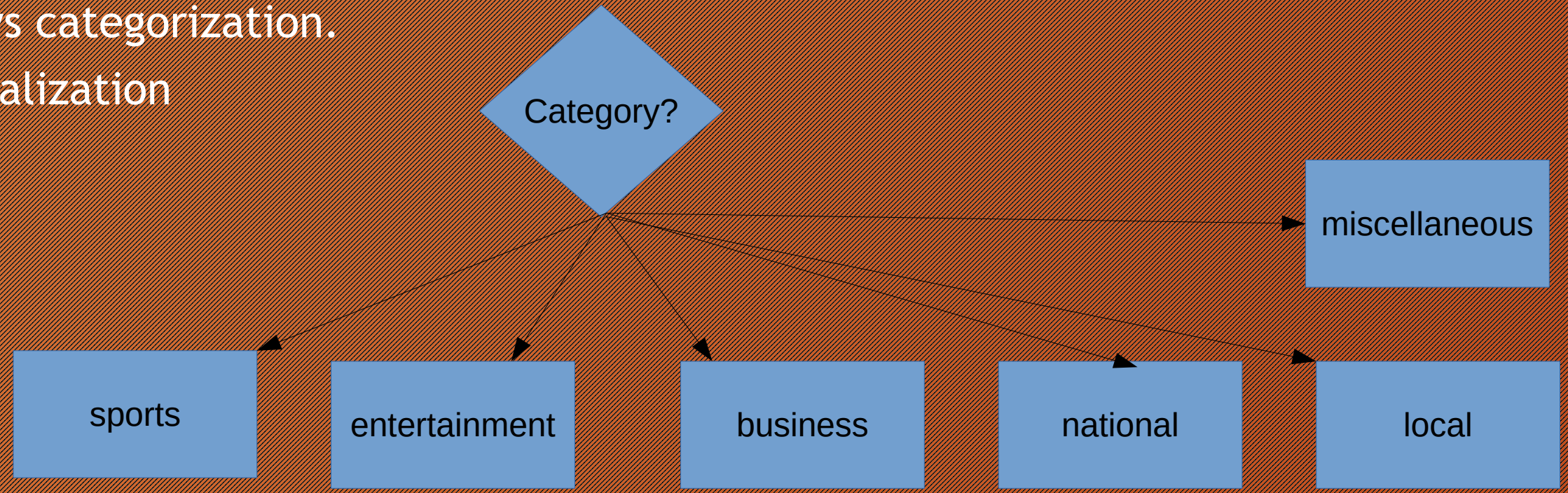
State of the art and limitations:

Various techniques have been proposed to classify the text:

Neural Network , Decision Trees , Support Vector Machines, Naive Bayes and Random forest. The basic concept of these techniques is the classification of news type using the trained classifier that can automatically predict an incoming news type to some of the predefined classes.

Objectives:

- News categorization.
- Visualization

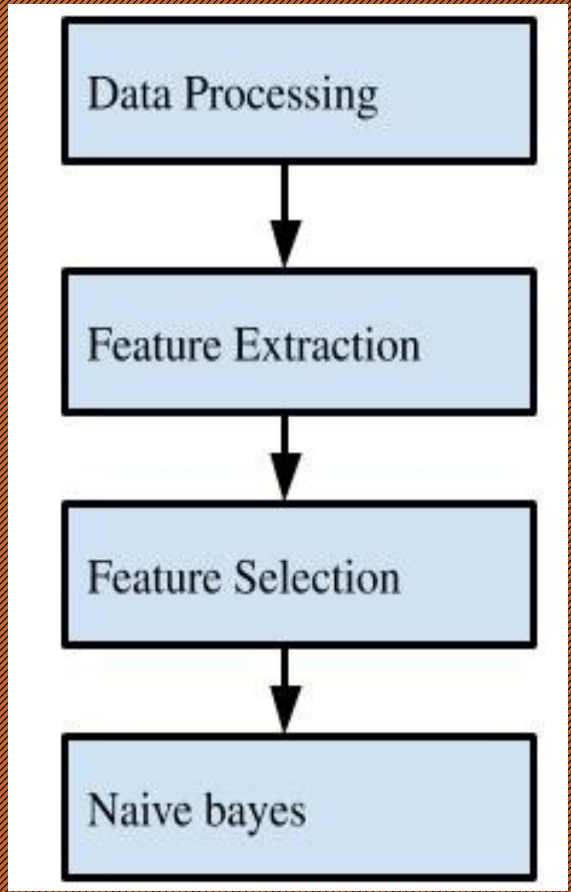


Work Distribution

- Selection of data set on news from online platform.
- Background research and relevant study for the project.
- Implementation of the model.
- SRS report and PPT.

  With 100% contribution from each member.

Proposed Design



Conclusion of the report:-

The main purpose of this project is to map news using appropriate machine learning algorithms and find accurate probability of particular headlines from news dataset for accurate categorization of news in five six categories. So on the basis of various functionalities that we tried and tested we were able to conclude that

- The probabilities for classifying news correctly are 0 for national, 0.24 for local, 0.83 for miscellaneous, 0.56 for entertainment, 0.52 for sports and 0 for business.
- News concerning India is the most published.
- Mumbai and Delhi lead the list as the most covered cities. Following these two cities, the decrease is fairly uniform.

Proposed work plan for remaining period :-

- Compare different ML models based on accuracy in news classification process, Sentiment Analysis- positive or negative ,Context analysis, Trends in local news- political , social etc