DCIT 316 HACKATHON GROUP 5 PROJECT PROPOSAL

Title: Sentiment Analysis using NLTK and TextBlob

Problem Definition:

The project aims to solve the challenge of accurately identifying and categorizing sentiments (positive, negative, or neutral) expressed in marginal volumes of text data, such as customer reviews, social media posts, and news articles.

Motivation:

With the exponential growth of user-generated content on social media and other platforms, understanding public sentiment is increasingly important for businesses, policymakers, and researchers. Effective sentiment analysis can provide valuable insights into consumer opinions, social trends, and public mood, enabling better decision-making and strategic planning.

Approach

Data Collection:

Gather a diverse dataset containing text data relevant to the project goals (e.g., customer reviews, tweets, news articles).

Data Preprocessing:

Tokenization: Split the text into individual words or tokens using NLTK.

Stopwords Removal: Remove common words that do not contribute to sentiment analysis using NLTK's stopwords list.

Lemmatization: Reduce words to their base or root form using NLTK's WordNetLemmatizer.

Sentiment Analysis:

Utilize TextBlob's sentiment analysis capabilities to calculate the polarity and subjectivity of each text entry.

Evaluation:

Compare the results from NLTK and TextBlob to assess their performance and accuracy.

Analyse the overall sentiment distribution and identify patterns or trends in the dataset.

Interpret the results to provide insights into the sentiments expressed in the text data.

References:

Abubakar, B. & Uppin, Chandrashekhar. (2021). A NATURAL LANGUAGE PROCESSING APPROACH TO DETERMINE THE POLARITY AND SUBJECTIVITY OF IPHONE 12 TWITTER FEEDS USING TEXTBLOB. Open Journal of Physical Science (ISSN: 2734-2123). 2. 10-17. 10.52417/ojps.v2i2.276.