INTRODUCTION

PROJECT DESCRIPTION :

Sentiment Analysis on Social Media Data Description: Analyze sentiment from social media posts (e.g., Twitter) regarding a specific topic or product. Use natural language processing techniques to preprocess text data, extract features, and classify sentiments as positive, negative, or neutral .

A Twitter sentiment analysis determines negative, positive, or neutral emotions within the text of a tweet using NLP and ML models. Sentiment analysis or opinion mining refers to identifying as well as classifying the sentiments that are expressed in the text source. Tweet are often useful in generating a vast amount of sentiment data upon analysis. These data are useful in understanding the opinion of people on social media for a variety of topics.

What is Twitter Sentiment Analysis?

Twitter sentiment analysis analyses the sentiment or emotion of tweets. It uses natural language processing and machine learning algorithms to classify tweets automatically as positive, negative, or neutral based on their content. It can be done for individual tweets or larger dataset related to a particular topic or event.

Twitter sentiment analysis refers to the process of analyzing the sentiment or opinion in tweets posted on the Twitter platform. Tweets are short text messages limited to 280 characters, making them concise expressions of opinions, emotions, or thoughts on various topics. Sentiment analysis on Twitter involves determining whether a tweet expresses a positive, negative, or neutral sentiment towards a particular subject, event, product, etc.

Overall, Twitter sentiment analysis enables organizations, researchers, and analysts to gain valuable insights into public opinion, consumer feedback, market trends, and social dynamics by analyzing the sentiments expressed in tweets on the platform.

LITERATURE SURVEY :

Sr. No	Title of Paper	Name of Authors	Published Year	Remarks / Findings
1	Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis	1.Rui Xia Jie Jiang Huihui He	2017	Traditional learning algorithms based on single tasks can often exhibit good performance in small-scale sentiment analysis
2	Study on Machine learning based Social Media and Sentiment analysis for medical data applications	1.R. Meena 2.Dr. V. Thulasi Bai	2019	Sentiment analysis and text based mining can be used as an ultimate tool for finding the user perception and public health intervention
3	Aspect-level Sentiment Analysis for Social Media Data in the Political Domain using Hierarchical Attention and Position Embeddings	1.Renny Pradina Kusumawardani 2. Muhammad Wildan Maulidani 3. Surabaya	2020	we have studied the use of hierarchical attention with position embeddings as in the HAPN for aspectbased sentiment analysis on social media data, specifically on the politic domain
4	Multilingual Sentiment Analysis on Social Media Disaster Data	1.Muhammad Jauharul Fuady 2. Roliana Ibrahim1	2018	In this paper, it has been proposed that a combined dataset and word embeddings during learning will help address the problem of multilingual sentiment classification.

OBJECTIVES

The main objective of the project is to help businesses identify when and how to engage with the customers directly and also to get valuable insights that helps to improve overall business performance.

From this project we will get to know more about the applications of machine learning as well as natural language processing. Here we use natural language processing to determine if the data is positive ,negative or neutral.

Machine learning projects play a crucial role in driving innovation, efficiency, and progress across diverse domains, ultimately leading to positive impacts on businesses, societies, and individuals.

1. **Understanding Customer Feedback:** By analyzing the sentiment of customer feedback, companies can identify areas where they need to improve their products or services.
2. **Reputation Management**: Sentiment analysis can help companies monitor their brand reputation online and quickly respond to negative comments or reviews.
3. **Political Analysis**: Sentiment analysis can help political campaigns understand public opinion and tailor their messaging accordingly.
4. **Crisis Management:** In the event of a crisis, sentiment analysis can help organizations monitor social media and news outlets for negative sentiment and respond appropriately.
5. **Marketing Research:** Sentiment analysis can help marketers understand consumer behaviour and preferences, and develop targeted advertising campaigns.

REQUIREMENTS:

1. Machine Learning Concepts
2. Python Language for Coding and their Libraries
3. Natural Language Processing
4. Data sets

METHODOLOGY:

Here's an overview of the steps involved in Twitter sentiment analysis:

1. Data Collection: The first step is to collect Twitter data relevant to the analysis. This can be done by accessing the Twitter API or using third-party tools that provide access to Twitter's data stream. Researchers or analysts typically specify keywords, hashtags, or user handles to filter tweets relevant to their analysis.

2. Preprocessing: Once the tweets are collected, they undergo preprocessing steps to clean and prepare the text for analysis. This may include tasks such as removing special characters, URLs, hashtags, mentions, and punctuation, as well as tokenization (splitting text into individual words or tokens) and lowercasing.

3. Feature Extraction: In this step, features are extracted from the pre-processed text to represent the content of the tweets. Commonly used features include word frequencies, n-grams (sequences of adjacent words), and syntactic or semantic features. Feature extraction techniques may vary depending on the specific sentiment analysis approach used.

4. Sentiment Classification: After feature extraction, sentiment classification algorithms are applied to classify the tweets into positive, negative, or neutral categories based on the extracted features. Various machine learning and deep learning techniques can be employed for sentiment classification, including but not limited to:

- Lexicon-based Methods: These methods rely on sentiment lexicons or dictionaries containing lists of words annotated with their sentiment polarity (positive, negative, neutral). The sentiment of a tweet is determined based on the presence and polarity of words in the lexicon.

- Machine Learning Models: Supervised machine learning models, such as Support Vector Machines (SVM), Naive Bayes, Logistic Regression, and Random Forests, are trained on labelled tweet datasets to learn patterns and associations between features and sentiment labels.

- Deep Learning Models: Deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer models, are capable of learning complex patterns in text data and have shown state-of-the-art performance in sentiment analysis tasks.

5. Evaluation: The performance of the sentiment analysis model is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, or area under the Receiver Operating Characteristic (ROC) curve, depending on the task and requirements.

6. Visualization and Interpretation: Finally, the results of the sentiment analysis can be visualized using charts, graphs, or dashboards to provide insights into the overall sentiment trends or patterns observed in the Twitter data. Analysts may also interpret the results to draw conclusions or make informed decisions based on the sentiment analysis findings.

- Dependencies:
  - re
  - pandas
  - numpy
  - seaborn
  - wordcloud
  - train test split
  - confusion matrix
  - classification report

- Logistic Regression
- Bernoulli Naive Bayes
- SVM
- TF-IDF
- Accuracy and F1 score
- NLP using nltk library
- Stemming
- Lemmatization
- Regexp Tokenizer
- ROC AUC Curve

ADVANTAGES

- Social Media data have several advantages for Sentiment Analysis , Such as its abundance and diversity .
- Millions of users post and share content on various topics, platforms and languages every day, creating a wealth of data .
- Additionally , Social media data is timely and relevant , reflecting the current and emerging opinions and sentiments of users .
- This allows businesses to respond quickly to customer feedback , market changes or crisis.
- Furthermore , more social media platforms provide API's or tools to collect and analyse data inexpensively and easily, reducing the cost and complexity of data acquisition and processing .
- We can get granular market analysis of customer likes and dislikes about products ,brands , advertising content and more .
- The benefits of sentiment analysis pervade across industries such as healthcare , aviation , banking , retail etc, and leads to insights that help companies better services and products for an enhances market share .

REFERENCE

[1] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” CS224N Project Rep., Stanford University, 2009.

[2] Z. Chen and B. Liu, “Lifelong machine learning,” Synthesis Lectures Artif. Intell. Mach. Learn., vol. 10, no. 3, pp. 1–145, 2016.

[3] E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research [review article],” IEEE Comput. Intell. Mag., vol. 9, no. 2, pp. 48–57, May 2014.

[4] R. Xia, F. Xu, J. Yu, Y. Qi, and E. Cambria, “Polarity shift detection, elimination and ensemble: A three-stage model for documentlevel sentiment analysis,” Inf. Process. Manag., vol. 52, no. 1, pp. 36–45, 2016.

[5] https://www.geeksforgeeks.org/links

[6] https://github.com