# Sentimental of Analysis of Text using Naive Bayes

Logesh kumar M – 727722EUIT096 III – IT -B

Pranav A – 727722EUIT131 III – IT -B

Nithish kumar P – 727722EUIT123 III – IT -B

# Introduction

**What is Sentiment Analysis?**

- Sentiment analysis is a method used to extract emotions or opinions expressed in text.
- It helps in understanding customer sentiment in reviews, feedback, social media posts, etc.

**Why Use Naive Bayes for Sentiment Analysis?**

- Naive Bayes is a simple, yet powerful algorithm based on probability, widely used for text classification tasks.
- Efficient in handling large datasets.

# Abstract

**Objective**:

To build a sentiment analysis system that classifies text into positive, negative, or neutral sentiments using the **Naive Bayes algorithm** and explore advanced approaches using **BERT (Bidirectional Encoder Representations from Transformers)** for enhanced context understanding.

**Method**:

- Preprocessing text (tokenization, stopword removal).
- Extracting features (Bag of Words or TF-IDF for Naive Bayes).
- Training a Naive Bayes model on labeled data.
- Utilizing **BERT** to capture deeper, bidirectional context for better sentiment prediction, especially in cases involving complex language such as sarcasm or ambiguous statements.

**Expected Outcome**:

- A **Naive Bayes classifier** capable of efficiently predicting sentiment from text.
- Enhanced accuracy and context understanding with **BERT**, particularly in nuanced scenarios where traditional models might misinterpret the sentiment.

# Existing system

- **Current Sentiment Analysis Solutions**:

  Lexicon-based models (e.g., VADER).

  Machine learning models like Support Vector Machines (SVM), decision trees, and deep learning models.

- **Challenges**:

  Complexity in training for large datasets.

  High computational requirements for advanced models.

  Poor handling of sarcasm, context, and mixed sentiments.

# Disadvantages

- **Sarcasm and Context Misinterpretation**: Difficulty in identifying sarcastic or ironic statements.
- **Ambiguity**: Mixed sentiments in one sentence are not handled well.
- **Domain-Specific Challenges**: General models do not perform well in specific domains like healthcare or finance.
- **Handling of Slang and Negations**: Difficulty in processing informal language and negation.
- **Imbalanced Datasets**: Skewed distribution of positive, negative, and neutral classes in datasets.

# Literature Survey

- **Naive Bayes for Sentiment Analysis**: Studies show Naive Bayes works efficiently for text classification.

- **Comparison of Naive Bayes with Other Algorithms**: Naive Bayes outperforms SVM in terms of efficiency and simplicity but struggles with deep contextual understanding.

- **Use of Transformer Models in Sentiment Analysis**: Research indicates models like BERT can handle complex context and sarcasm but are resource-heavy.

# Proposed System

- **System Overview**:
  - **Step 1**: Text Preprocessing (Tokenization, Stopword Removal, Stemming/Lemmatization).
  - **Step 2**: Feature Extraction (Bag of Words, TF-IDF).
  - **Step 3**: Naive Bayes Model Training on labeled data.
  - **Step 4**: Sentiment Classification (Positive, Negative, Neutral).

- **Architecture**:
- Input Data → Preprocessing → Feature Extraction → Naive Bayes Classifier → Sentir

# Bayes' Theorem

$$P(c|x)=P(x|c)\cdot P(c)/P(x)$$

Where:

- $P(c|x)$is the **posterior probability** that the class label ccc (e.g., positive or negative sentiment) is true given the input data xxx (i.e., the text).
- $P(x|c)$ is the **likelihood** of the data xxx given the class ccc.
- $P(c)$ is the **prior probability** of the class ccc.
- $P(x)$ is the **marginal likelihood**, or the probability of the data xxx, which acts as a normalizing factor.

# Disadvantages of Naive Bayes and How to Overcome Them

- **Sarcasm and Context Misinterpretation**:

  Naive Bayes fails to detect sarcasm because it does not account for complex sentence structures.

- **How to Overcome**:

  1.Use **Advanced NLP Models**: Leverage transformers like **BERT** to handle complex contexts and sarcasm better.

  **2.Sarcasm Detection**: Train a sarcasm detection model to recognize sarcasm before sentiment classification.

  3.Include **Contextual Features**: Use cues such as emoticons, punctuation, or metadata like past user behavior to understand sarcasm.

  **4.Use of Specialized Datasets**: Train on datasets labeled with sarcasm and irony.

# Advantages of Naive Bayes

- **Efficiency**:
- Naive Bayes is computationally light and can handle large datasets with low training

- **Simplicity**:
- Easy to implement with fewer computational resources compared to deep learning m

- **Strong Performance in Text-Based Applications**:
- Performs well on straightforward text classification tasks.

- **Interpretability**:
- Provides a probabilistic output that is easy to interpret and explain.