# NEWS ARTICLE CATEGORIZATION USING

# NAIVE BAYES MODEL

A project report submitted

in partial fulfilment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE & ENGINEERING**

By


| | |
|---|---|
| **KATURI ARAVIND** | **(2203A51140)** |
| **POTHANA NAGAVISHNU** | **(2203A51189)** |
| **SOMARTHI NIMESH** | **(2203A51446)** |
| **PANCHAKAM DAVAN VIKAS** | **(2203A51380)** |


Under the guidance of

**Mr. Dr. N. Venkatesh**

Associate Professor, CS & AI.

# SR UNIVERSITY

Ananthasagar, Warangal.


## CERTIFICATE

This is to certify that this project entitled **"NEWS ARTICLE CATEGORIZATION USING NAIVES BAYES MODEL"** is the bonafied work carried out by **KATURI ARAVIND, POTHANA NAGAVISHNU, SOMARTHI NIMESH, DAVAN VIKAS** as a Capstone Phase-II project for the partial fulfilment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING** during the academic year 2024-2025 under our guidance and Supervision.

**Dr. N. Venkatesh**
Assoc. Prof. CS & AI
S R University,
Ananthasagar, Warangal.

**Dr. M. Sheshikala**
Assoc. Prof. & HOD(CSE),
S R University,
Ananthasagar, Warangal.

**External Examiner**

# ACKNOWLEDGEMENT

We owe an enormous debt of gratitude to our project guide **Mr. Dr. N. Venkatesh, Assoc. Prof. CS and AI** as well as Head of the CSE Department **Dr. M. Sheshikala, Associate Professor** for guiding us fromthe beginning through the end of the Capstone Phase-II project with their intellectual advices and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction..

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

# 1. ABSTRACT

The Naive Bayes classifier, a probabilistic machine learning technique, is useful for classification tasks. It is based on the Bayes theorem, which states that the likelihood of an event occurring given some observed evidence is equal to the prior probability of the event occurring. The Naive Bayes classifier can be trained on a dataset of labelled news articles, each of which is associated with a particular class or category, for the purpose of classifying news articles. The features of the articles, such as the words used and the length of the article, can then be used by the classifier to predict the class of an unseen article. The "naive" assumption, which is one of the key assumptions of the Naive Bayes classifier, is that the articles' features are independent of one another. The classifier is able to predict outcomes without taking into account how features interact with one another because of this assumption. The Naive Bayes classifier can still perform well on many classification tasks, including the classification of news articles, despite this assumption.

## KEYWORDS

Natural language toolkit ,python ,machine learning algorithm.

# Table of contents

## List of Figure

## 2. INTRODUCTION:

The process of classifying a news article according to its content is known as news article classification. This is a common issue in information retrieval and natural language processing, and it can be useful for organizing and searching through large collections of news articles.

One approach to categorising news stories is to use a machine learning algorithm like the Naive Bayes classifier. A probabilistic model known as the Naive Bayes classifier makes predictions based on the likelihood that particular occurrences will occur. The events are the classes or categories to which news articles can be classified, and the features are the words or other characteristics of the articles.

A dataset of labelled news articles, each of which is associated with a distinct class, is required to train a Naïve Bayes classifier for news article classification. The classifier would then learn the probability distribution of the characteristics for each class. which would then use this information to predict articles that had not been seen before. The Naive Bayes classifier is able to efficiently simplify calculations and make predictions because it assumes that the articles' features are independent of one another.

A lot of classification tasks benefit from the Naive Bayes classifier's relative simplicity and ease of use, which is one of its advantages. It can also do well on a variety of classification problems, such as classifying news articles. However, in order to ensure that the classifier is effective, it is essential to evaluate its performance on your particular dataset and problem.

# 3. PROBLEM STATEMENT

In the digital age, with an exponential increase in online content, efficient organization and categorization of news articles have become imperative. News websites and aggregators face the challenge of quickly and accurately categorizing articles into relevant topics to enhance user experience and facilitate targeted content delivery.

Effective categorization is not just a matter of convenience; it's integral to enhancing user experience and ensuring that readers can quickly access the content most relevant to their interests. Moreover, targeted content delivery has emerged as a key strategy for retaining users and driving engagement in a competitive online environment.

By developing a Naive Bayes classification model, we aim to harness the power of machine learning to automate the process of assigning articles to predefined topics or categories. Leveraging the textual content of news articles, this model will learn to identify patterns and associations that distinguish one topic from another.

# 4. MOTIVATION AND SCOPE OF WORK

The motivation behind implementing a Naive Bayes classification model for news article categorization stems from the pressing need to streamline content management processes and improve content discoverability for users. By automating the categorization process, we aim to alleviate the burden on human moderators and enable news platforms to efficiently classify articles into relevant topics or categories. This not only enhances the overall browsing experience for users but also enables targeted content delivery, thereby maximizing user engagement and satisfaction.

The scope of this project encompasses the development and implementation of a Naive Bayes classification model for news article categorization, focusing on the following key aspects:

**Data Collection and Preparation:** Gather a diverse dataset of news articles spanning different categories such as politics, sports, entertainment, technology, etc. Preprocess the raw text data to remove noise, perform tokenization, and apply techniques like stemming or lemmatization to normalize the text.

**Feature Engineering:** Utilize techniques such as Bag-of-Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF) to transform the preprocessed text data into numerical feature vectors. Additionally, explore advanced feature extraction methods to capture semantic information and improve model performance.

**Model Development:** Implement the Naive Bayes algorithm, a probabilistic classifier based on Bayes' theorem, to build a classification model. Experiment with different variants of Naive Bayes, such as Multinomial Naive Bayes or Bernoulli Naive Bayes, to identify the most suitable approach for the task at hand.

**Model Evaluation and Optimization:** Assess the performance of the Naive Bayes model using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Employ techniques like cross-validation and hyperparameter tuning to optimize the model's performance and ensure robustness.

**Integration and Deployment:** Once the model achieves satisfactory performance, deploy it into a production environment where it can automatically categorize incoming news articles in real-time. Integrate the model with existing news platforms or APIs to facilitate seamless classification and content delivery.

By delineating the motivation and scope of the work, we aim to establish a clear roadmap for developing and implementing a Naive Bayes classification model for news article categorization, with the overarching goal of enhancing user experience and engagement in the digital news ecosystem.

# 5. LITERATIVE REVIEW

## 5.1 Related Work

R. Siva Subhramanian and D. Prabha [22] contributed their paper in In February 2020 on research of This research seeks to identify potential customers.
They used the SBC method to modify the NB model with the goal of enhancing prediction by removing unnecessary dataset features.

According to the experimental findings, the WSNB running time is 0.03 seconds for WSNB at depth 1, 0.06 seconds for WSNB at depth 2, and 0.15 seconds for WSNB at depth 3. Running time for Standard Naive Bayes was 0.16 seconds. Which was unmistakably demonstrating that WSNB shortens the model's running time as compared to traditional Naive Bayes.

Faculty of Agriculture, University of Novi Sad [23] published their article in 2022. The effectiveness of the Naive Bayes approach for predicting water quality was studied by the author. Nine water quality factors were examined, including temperature, oxygen saturation values, and others. Five locations and 68 samples of data were used to assess the water quality using the Naive Bayes model. The testing report ranked each parameter as very good, excellent, good, or bad; after analysing the report and using the method, the author came to the conclusion that the model correctly identified water class in 64 out of 68 instances.

Disha Sharma and Sumit Chaudhary [24] They studied various sources of stress which includes 1) The surrounding Environment 2) Social Stress 3) Physiological 4) Thoughts

Authors applied four machine learning technics that are logistic Regression, Naïve Bayes, Multilayer perceptron ,Bayer's Net.

Parameters like False Positive rate, True Positive Rate , precision, Recall considered for the performance. After comparing all the results of four methods they concluded that Baye's Net classifiers gives longest accuracy of 88 percentage and Naive Bayes gives accuracy of 86 percentage.
Mamata Thakur and team [25] by concerning the problem of huge growth of internet and difficulty in getting relevant topic according to search. Authors chose some news websites after that the important attributes from these

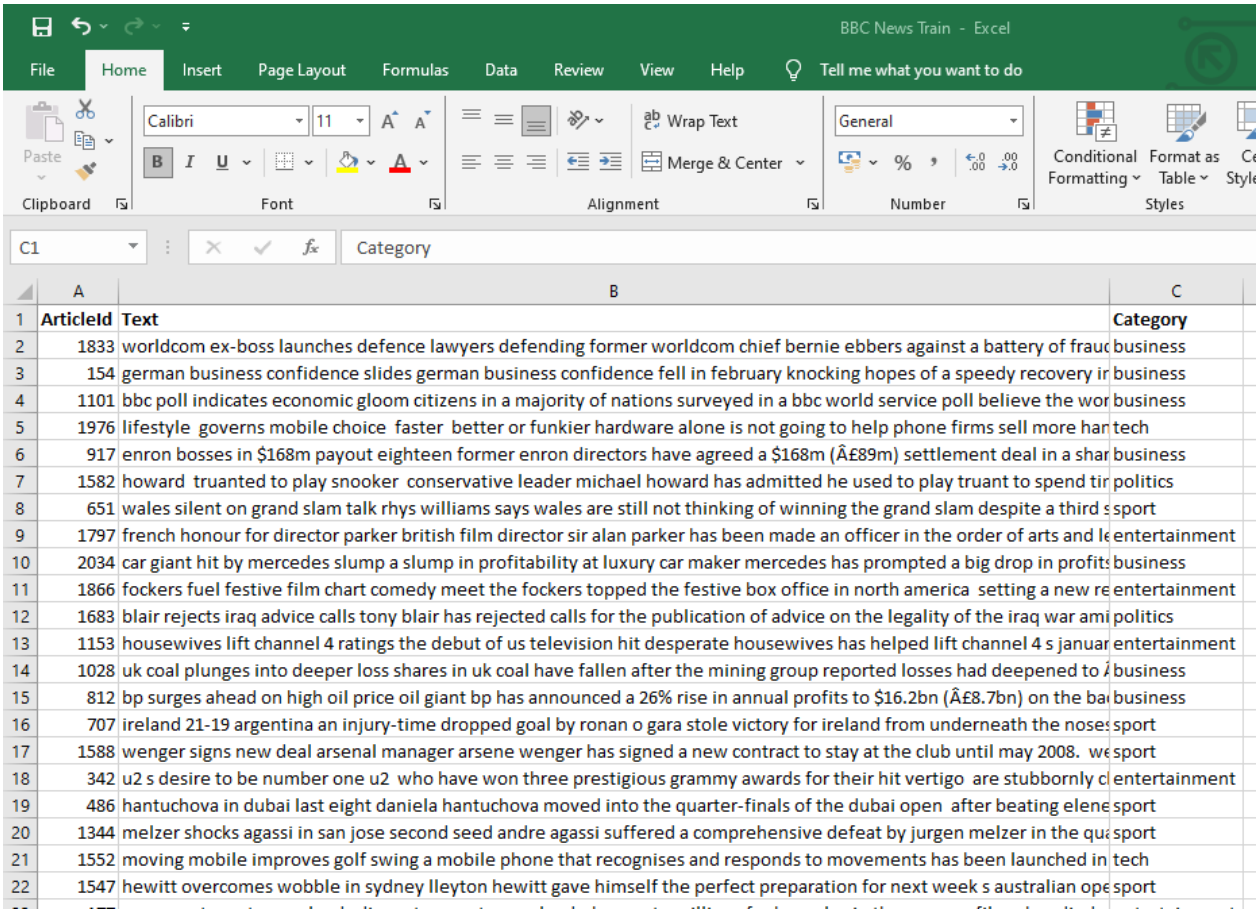
The Nave Bayes algorithm was used by the authors to classify data from 10 different websites, and the results of comparative studies with other current algorithms on the same dataset demonstrate that Nave Bayes outperforms them.

Yi Ying [26] The author of this study employed a variety of news stories to research and used news categories including sports, politics, business, etc. The Confusion Matrix results show that the Sarcasm model developed using the Naive Bayes approach above achieved an accuracy level of 66%, 70% withdrawal, and 68% precision.

By summarising the literature review we can understand sometimes Navies Bayes gives good results but not able to give 100% correct results and some of other machine learning algorithms are more effective than NB, so more researches can be done increase efficiency of NB

## 6. DATASET

Dataset consists of Article Id, Article Description and Article category which it belongs to



BBC News Train - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Calibri 11 Wrap Text General Merge & Center Conditional Formatting Format as Table

Clipboard Font Alignment Number Styles

C1 Category

| | A | B | C |
|---|---|---|---|
| 1 | ArticleId | Text | Category |
| 2 | 1833 | worldcom ex-boss launches defence lawyers defending former worldcom chief bernie ebbers against a battery of fraud | business |
| 3 | 154 | german business confidence slides german business confidence fell in february knocking hopes of a speedy recovery in | business |
| 4 | 1101 | bbc poll indicates economic gloom citizens in a majority of nations surveyed in a bbc world service poll believe the wor | business |
| 5 | 1976 | lifestyle governs mobile choice faster better or funkier hardware alone is not going to help phone firms sell more han | tech |
| 6 | 917 | enron bosses in $168m payout eighteen former enron directors have agreed a $168m (Â£89m) settlement deal in a shar | business |
| 7 | 1582 | howard truanted to play snooker conservative leader michael howard has admitted he used to play truant to spend tir | politics |
| 8 | 651 | wales silent on grand slam talk rhys williams says wales are still not thinking of winning the grand slam despite a third s | sport |
| 9 | 1797 | french honour for director parker british film director sir alan parker has been made an officer in the order of arts and le | entertainment |
| 10 | 2034 | car giant hit by mercedes slump a slump in profitability at luxury car maker mercedes has prompted a big drop in profits | business |
| 11 | 1866 | fockers fuel festive film chart comedy meet the fockers topped the festive box office in north america setting a new re | entertainment |
| 12 | 1683 | blair rejects iraq advice calls tony blair has rejected calls for the publication of advice on the legality of the iraq war ami | politics |
| 13 | 1153 | housewives lift channel 4 ratings the debut of us television hit desperate housewives has helped lift channel 4 s januar | entertainment |
| 14 | 1028 | uk coal plunges into deeper loss shares in uk coal have fallen after the mining group reported losses had deepened to Á | business |
| 15 | 812 | bp surges ahead on high oil price oil giant bp has announced a 26% rise in annual profits to $16.2bn (Â£8.7bn) on the ba | business |
| 16 | 707 | ireland 21-19 argentina an injury-time dropped goal by ronan o gara stole victory for ireland from underneath the noses | sport |
| 17 | 1588 | wenger signs new deal arsenal manager arsene wenger has signed a new contract to stay at the club until may 2008. we | sport |
| 18 | 342 | u2 s desire to be number one u2 who have won three prestigious grammy awards for their hit vertigo are stubbornly cl | entertainment |
| 19 | 486 | hantuchova in dubai last eight daniela hantuchova moved into the quarter-finals of the dubai open after beating elene | sport |
| 20 | 1344 | melzer shocks agassi in san jose second seed andre agassi suffered a comprehensive defeat by jurgen melzer in the qua | sport |
| 21 | 1552 | moving mobile improves golf swing a mobile phone that recognises and responds to movements has been launched in | tech |
| 22 | 1547 | hewitt overcomes wobble in sydney lleyton hewitt gave himself the perfect preparation for next week s australian ope | sport |

Fig 1. Dataset

# 7. PROPOSED METHODOLOGY

Our proposed methodology for news article categorization using a Naive Bayes model encompasses several key steps to ensure the development of an accurate and robust classification system. The first phase involves data collection and preprocessing. In this step, we gather a diverse dataset of news articles spanning different categories such as politics, sports, entertainment, technology, and more. We then preprocess the text data by removing noise, including HTML tags, punctuation, and stop words, and perform tasks like tokenization and stemming to normalize the text. Additionally, we may employ techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) or Bag-of-Words to convert the preprocessed text data into numerical feature vectors, which will serve as input to the Naive Bayes classifier.

Following data preprocessing, the next phase entails model development and training. Here, we implement the Naive Bayes algorithm, which assumes independence between features, to build the classification model. We split the dataset into training and testing sets to train the model on labeled data and evaluate its performance. During training, the model learns the statistical relationships between features and categories, enabling it to make accurate predictions. We fine-tune model parameters and optimize performance using techniques like cross-validation. Once the model achieves satisfactory performance metrics, we proceed to deploy it into a production environment, where it can automatically categorize incoming news articles in real-time. Continuous monitoring and evaluation ensure the model's effectiveness and adaptability to evolving news trends and topics. Through this methodology, we aim to deliver a robust and efficient news article categorization system that enhances user experience and facilitates targeted content delivery.

Here is a general approach to utilising the provided code to categorise news articles:

1. Assemble and classify a dataset of news stories, each with a category tagged (e.g., sports, tech, business, entertainment). The classifier will be trained and tested using this dataset.
2. Remove all stop words from the data and lowercase each word in each article as part of the pre-processing.
3. To turn the text input into numerical feature vectors, create a TfidfVectorizer.
4. The training and test data should be converted into feature vectors using the TfidfVectorizer.
5. Making use of the training data, create a Multinomial Naive Bayes classifier.
6. Calculate the classifier's accuracy by evaluating it against the test data.
7. Make predictions for fresh, unlabelled news articles using the classifier by converting them into feature vectors and passing them into the classifiers predict method

## 8. Flow Chart

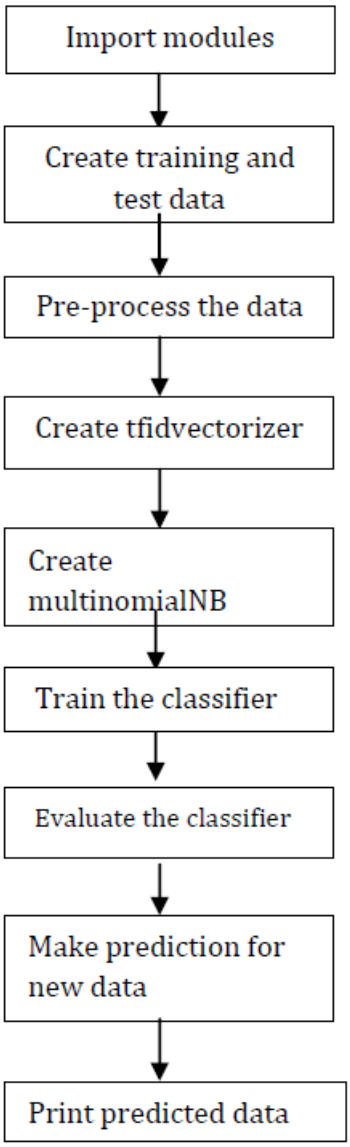



Fig .2 Flow chart

## 9.1 COMPARED ALGORITHMS

### 9.1.1 KNN

Twitter Data Sensory Analysis using KNN Editing. Emotional analysis refers to the use of natural language processing, text analysis, and computer languages to systematically identify, extract, evaluate, and learn practical situations and independent information. Sentiment Analysis is the most widely used method of quoting a text. Twitter Sentiment Analysis, therefore, means using advanced text-cutting techniques to analyze text emotions (here, tweet) in a positive, negative and neutral way.

### 9.1.2 SVM

Sentiment Analysis is an NLP method that works in a text to determine the author's intentions for a particular topic, product, etc. positive, negative, or neutral. The supporting vector machine analyzes the data, defines the decision parameters and uses the calculations to calculate what is done in the input field data entry for two sets of vectors of each size. Then all the data represented as vector is categorized. Next, we find the margin between the two sections away from any document. The distance defines the divider margin, increasing the limit reduces the final decisions. SVM also supports the subdivision and decline in practical mathematical learning theory and also helps to identify specific factors, which need to be considered, in order to understand them effectively.

### 9.1.3 DECISION TREE

Decision trees are the most common way to say something. They are strong on sound data and learn divisive sayings. Decision tree is a k-array tree where each internal node displays an experiment in a few elements from a set of input element that communicates with the data. Every branch from a node is related to the unimaginable feature values determined for that node. Also, all test results in branches, refer to changed test results. The basic algorithm for decision tree imports algorithm is the decision-making tree algorithm in the form of repeated downward divisions and conquests.

## 9.2 HARDWARE AND SOFTWARE TOOLS

### HARDWARE TOOLS

- System
- Hard Disk
- Ram-8 GB
- Processor

### SOFTWARE TOOLS

- Operating System-Windows 11
- Google Colab Notebook
- Python IDLE
- Pandas
- NumPy
- TensorFlow
- TPU
- NLTK

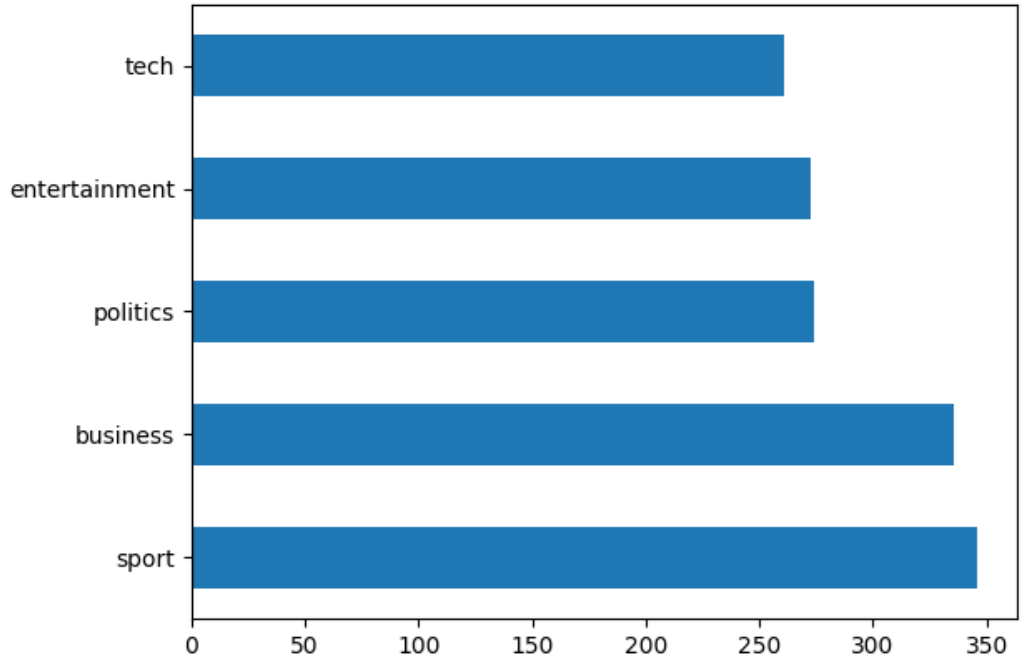
## 10.RESULTS & DISCUSSION

### Code:

```
# Import libraries

import pandas as pd
import re

import nltk
from nltk.tokenize import RegexpTokenizer
from collections import Counter
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords

import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

from sklearn.feature_extraction.text import TfidfVectorizer, TfidfTransformer
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import AdaBoostClassifier, GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
import xgboost
from sklearn.metrics  import classification_report
from sklearn import metrics
import time
```

### Finding Categories in dataset

```
df1['Category'].value_counts().plot(kind='barh')
plt.show()
```




### Checking Null values

```
# Check null values
df1.isnull().sum()
```

ArticleId 0
Text 0
Category 0
dtype: int64

### Text Pre Processing

```
# Text preprocessing
def preprocess(text):

    """
    Function: split text into words and return the root form of the words
    Args:
      text(str): the article
    Return:
      lem(list of str): a list of the root form of the article words
    """

    # Normalize text
    text = re.sub(r"[^a-zA-Z]", " ", str(text).lower())

    # Tokenize text
    token = word_tokenize(text)

    # Remove stop words
    stop = stopwords.words("english")
```

```
    words = [t for t in token if t not in stop]

    # Lemmatization
    lem = [WordNetLemmatizer().lemmatize(w) for w in words]

    return lem
```

**Output:**

```
df1["Preprocessed_Text"] = df1['Text'].apply(lambda x: preprocess(x))
df1.head(10)
```

[13] Python

| | ArticleId | Text | Category | Preprocessed_Text |
|---|---|---|---|---|
| 0 | 1833 | worldcom ex-boss launches defence lawyers defe... | business | [worldcom, ex, bos, launch, defence, lawyer, d... |
| 1 | 154 | german business confidence slides german busin... | business | [german, business, confidence, slide, german, ... |
| 2 | 1101 | bbc poll indicates economic gloom citizens in ... | business | [bbc, poll, indicates, economic, gloom, citize... |
| 3 | 1976 | lifestyle governs mobile choice faster bett... | tech | [lifestyle, governs, mobile, choice, faster, b... |
| 4 | 917 | enron bosses in $168m payout eighteen former e... | business | [enron, boss, payout, eighteen, former, enron,... |
| 5 | 1582 | howard truanted to play snooker conservative... | politics | [howard, truanted, play, snooker, conservative... |
| 6 | 651 | wales silent on grand slam talk rhys williams ... | sport | [wale, silent, grand, slam, talk, rhys, willia... |
| 7 | 1797 | french honour for director parker british film... | entertainment | [french, honour, director, parker, british, fi... |
| 8 | 2034 | car giant hit by mercedes slump a slump in pro... | business | [car, giant, hit, mercedes, slump, slump, prof... |
| 9 | 1866 | fockers fuel festive film chart comedy meet th... | entertainment | [fockers, fuel, festive, film, chart, comedy, ... |

## Finding Common Words

```
# Find the common words in each category
def find_common_words(df, category):

    """
    Function: find the most frequent words in the category and return the them
    Args:
      df(dataframe): the dataframe of articles
      category(str): the category name
    Return:
      the most frequant words in the category
    """

    # Create dataframes for the category
    cat_df = df[df["Category"]==category]

    # Initialize words list for the category
    words = [word for tokens in cat_df["Preprocessed_Text"] for word in tokens]

    # Count words frequency
    words_counter = Counter(words)

    return words_counter.most_common(10)
```

```
print("Most common words in each category")
for c in category:
    print(c, " News")
    print(find_common_words(df1, c))
    print()
```

[15] Python

```
Most common words in each category
business  News
[('said', 1100), ('year', 618), ('bn', 535), ('u', 523), ('mr', 394), ('company', 393), ('firm', 374), ('market',

tech  News
[('said', 1064), ('people', 647), ('mobile', 437), ('phone', 396), ('game', 396), ('technology', 380), ('service',

politics  News
[('said', 1445), ('mr', 1100), ('would', 712), ('labour', 494), ('election', 479), ('government', 471), ('party',

sport  News
[('said', 636), ('game', 487), ('year', 448), ('first', 350), ('win', 337), ('time', 336), ('england', 329), ('pla

entertainment  News
```

## Training And Evaluating Model

```
# Train and evaluate model
def fit_eval_model(model, train_features, y_train, test_features, y_test):

    """
    Function: train and evaluate a machine learning classifier.
    Args:
      model: machine learning classifier
      train_features: train data extracted features
      y_train: train data lables
      test_features: train data extracted features
      y_test: train data lables
    Return:
      results(dictionary): a dictionary of the model training time and
classification report
    """
    results ={}

    # Start time
    start = time.time()
    # Train the model
    model.fit(train_features, y_train)
    # End time
    end = time.time()
    # Calculate the training time
    results['train_time'] = end - start

    # Test the model
    train_predicted = model.predict(train_features)
```

```
    test_predicted = model.predict(test_features)

     # Classification report
    results['classification_report'] = classification_report(y_test,
test_predicted)

    return results
```

**Initializing the model**

```
# Initialize the models
nb = MultinomialNB()


# Fit and evaluate models
results = {}
for cls in [nb]:
    cls_name = cls.__class__.__name__
    results[cls_name] = {}
    results[cls_name] = fit_eval_model(cls, train_features, y_train, test_features,
y_test)
```

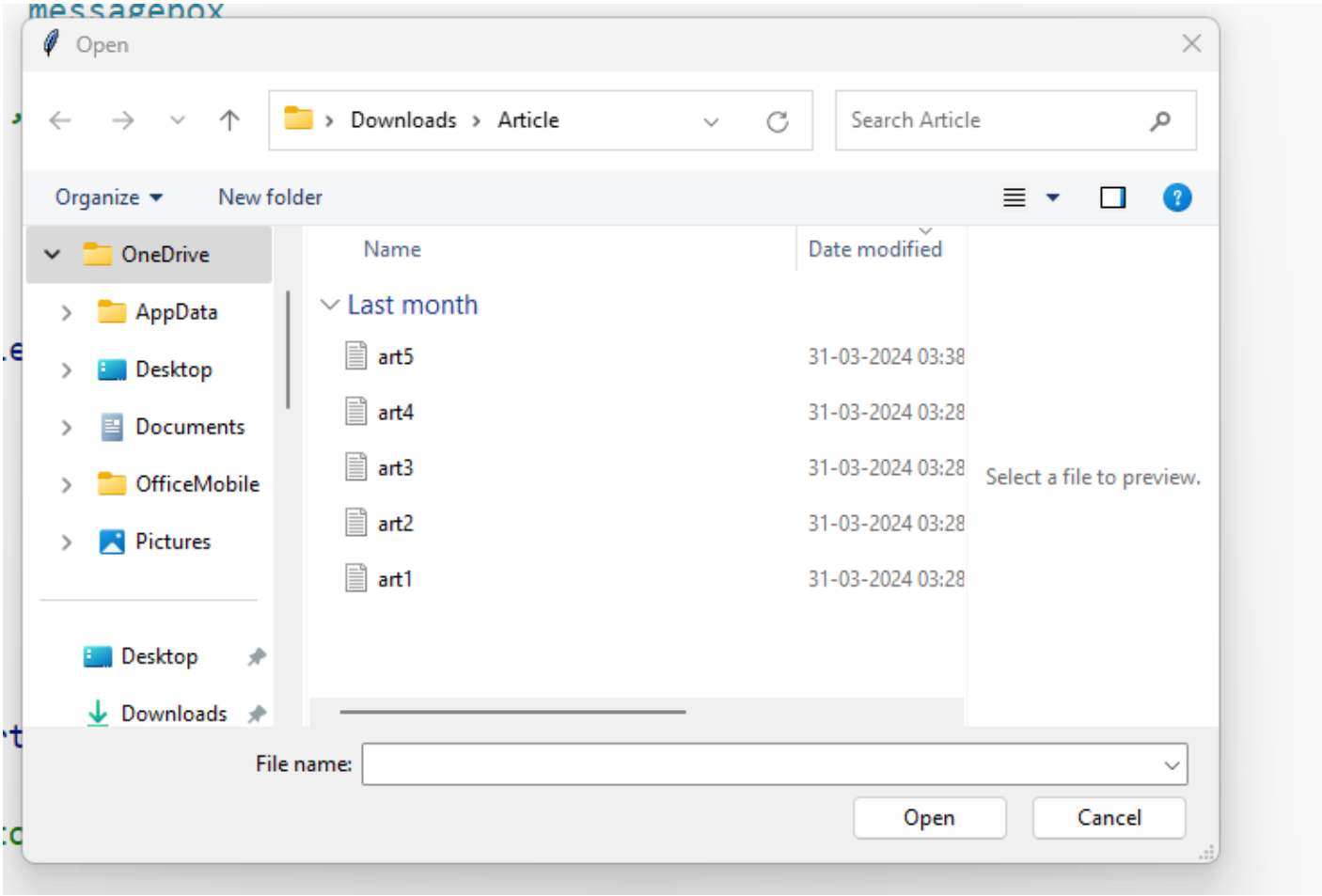
```
# Print classifiers results
for res in results:
    print (res)
    print()
    for i in results[res]:
        print (i, ':')
        print(results[res][i])
        print()
    print ('-----')
    print()
```

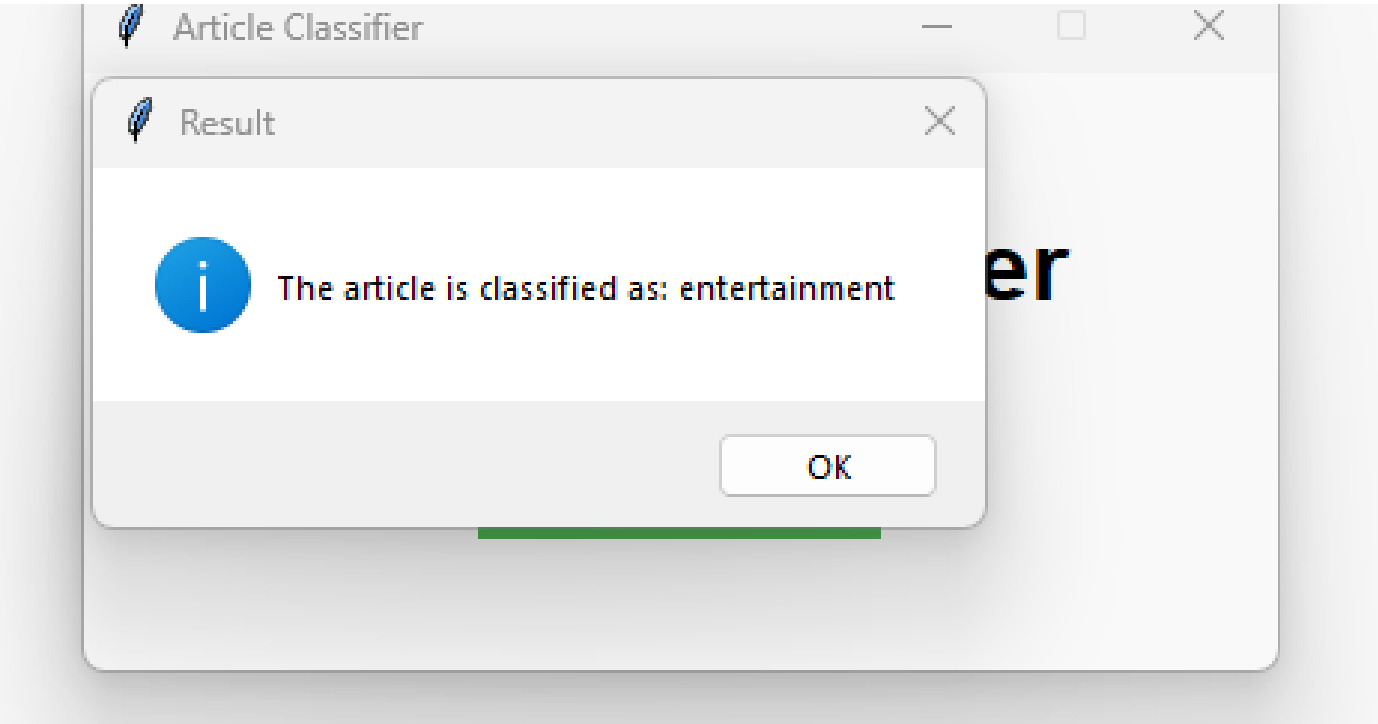
```
MultinomialNB

train_time :
0.013199567794799805

classification_report :
               precision    recall  f1-score   support

     business       0.95      0.94      0.94        64
entertainment       1.00      0.94      0.97        63
     politics       0.96      0.96      0.96        45
        sport       0.97      1.00      0.99        78
         tech       0.92      0.98      0.95        48

     accuracy                           0.96       298
    macro avg       0.96      0.96      0.96       298
 weighted avg       0.96      0.96      0.96       298
```

## Applying a Graphical User Interface (GUI)

```
import tkinter as tk
from tkinter import filedialog, messagebox

# Assume 'preprocess', 'tf_vec', and 'nb' are defined elsewhere in your code
# ...

def classify_article(path):
    # Read file
    with open(path, 'r') as file:
        artcl = file.read()

    # Text preprocessing
    artcl = preprocess(artcl)
    artcl = ' '.join(artcl)

    # Use TF_IDF
    test = tf_vec.transform([artcl])

    # Use MultinomialNB model to classify the article
    predict = nb.predict(test)
    category = predict[0]
```

## Output of the Project


Open
Downloads › Article
Search Article
Organize
New folder
OneDrive
AppData
Desktop
Documents
OfficeMobile
Pictures
Desktop
Downloads
Name
Date modified
Last month
art5
31-03-2024 03:38
art4
31-03-2024 03:28
art3
31-03-2024 03:28
art2
31-03-2024 03:28
art1
31-03-2024 03:28
Select a file to preview.
File name:
Open
Cancel


Article Classifier
Result
The article is classified as: entertainment
OK

# 11. CONCLUSION:

In conclusion, the versatility and simplicity of Naive Bayes models make them a compelling choice for news article categorization tasks in today's digital landscape. By leveraging probabilistic classification, Naive Bayes models can efficiently handle large volumes of textual data, providing a practical solution for organizing and categorizing news content across diverse topics. Their effectiveness lies in their ability to capture essential features from text data while making the assumption of feature independence, which often proves to be reasonable in practice.

Furthermore, as the digital landscape continues to evolve, Naive Bayes models can benefit from integration with complementary techniques, such as deep learning approaches, to capture more nuanced semantic relationships within news articles. By embracing interdisciplinary approaches and collaborating across fields such as natural language processing, machine learning, and information retrieval, researchers can unlock new opportunities to enhance the effectiveness of Naive Bayes models in facilitating efficient organization and categorization of news content in the digital age.

The Naive Bayes classifier is a popular machine learning method that can be used to categorise news stories. It has a lot going for it, like being easy to use, working well, and doing well on a lot of classification tasks. However, its performance on a particular dataset must be evaluated. The classifier's performance can be improved, more complex problems can be handled, and the classifier can be applied to new domains can all be developed further in this area. Natural language processing and information retrieval could benefit greatly from using the Naive Bayes classifier.

# 12. FUTURE WORK:

In the field of news article classification using Naive Bayes classifiers, there are numerous potential future directions for research and development. These are some: expanding the application of the classifier to new domains and languages, enhancing the classifier's performance, and incorporating it into news analysis systems. The Naive Bayes classifier offers a lot of potential for solving a variety of real-world issues, and more research may be done on its capabilities and restrictions when it comes to categorising news items.

# REFERENCES

[1] Kuldeep Vayadande, Aditya Bodhankar, Ajinkya Mahajan, Diksha Prasad, Shivani Mahajan, Aishwarya Pujari and Riya Dhakalkar, “Classification of Depression on social media using Distant Supervision”, ITM Web Conf. Volume 50, 2022.

[2] Kuldeep Vayadande, Rahebar Shaikh, Suraj Rothe, Sangam Patil, Tanuj Baware and Sameer Naik,” Blockchain-Based Land Record System”, ITM Web Conf. Volume 50, 2022.

[3] Kuldeep Vayadande, Kirti Agarwal, Aadesh Kabra, Ketan Gangwal and Atharv Kinage,” Cryptography using Automata Theory”, ITM Web Conf. Volume 50, 2022

[4] Samruddhi Mumbare, Kunal Shivam, Priyanka Lokhande, Samruddhi Zaware, Varad Deshpande and Kuldeep Vayadande,”Software.

[5] Preetham, H. D., and Kuldeep Baban Vayadande. "Online Crime Reporting System Using Python Django."

[6] Vayadande, Kuldeep B., et al. "Simulation and Testing of Deterministic Finite Automata Machine." International Journal of Computer Sciences and Engineering 10.1 (2022): 13-17.

[7] Vayadande, Kuldeep, et al. "Modulo Calculator Using Tkinter Library." EasyChair Preprint 7578 (2022).

[8] VAYADANDE, KULDEEP. "Simulating Derivations of Context-Free Grammar." (2022).

[9] Vayadande, Kuldeep, Ram Mandhana, Kaustubh Paralkar, Dhananjay Pawal, Siddhant Deshpande, and Vishal Sonkusale. "Pattern Matching in File System." International Journal of Computer Applications 975: 8887.