undefined
undefined

CAPSTONE PROJECT

SENTIMENTAL ANALYSIS

Presented By:

1. **Student Name :-** Prasad Gudivada
2. **College Name :-** SATYA INSTITUTE OF TECHNOLOGY AND MANAGEMENT
3. **Department :-** COMPUTER SCIENCE AND ENGINEERING

AGENDA

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Design implement a robust sentiment analysis model specifically tailored for restaurant reviews. The goal is to accurately classify these reviews as either positive or negative based on the expressed sentiments. The model should handle various linguistic nuances, account for context, and provide reliable predictions.

Description: AS online platforms for sharing opinions and reviews continue to grow rapidly, restaurants increasingly rely on customer feedback to enhance their services and attract new customers.

PROPOSED SOLUTION

- The challenge is to predict the required number of restaurant reviews at each hour to maintain a stable supply of restaurant food tastes. Accuracy forecasts are essential for optimizing food preparation and meeting customer expectations.
- Data Collection:
  - Gather historical data on restaurant reviews, including timestamps, meal types, customer sentiments.
  - Collect the information on restaurant menus, special dishes, and seasonal variations.
- Data Preprocessing:
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Create relevant features such as meal type (breakfast, lunch, dinner), day of the week, and special events.
  - Explore sentiment features (positive, negative).
- Machine Learning Algorithm:
  - Navie Bayes: A probabilistic algorithm that assigns probabilities to words or phrases being positive or negative.
  - Linear Regression: A statistical algorithm that predicts a Y value (output) based on X features (input).
  - Support vector machines, Random forests, Deep learning Algorithms and Logistic Regression.
  - Weather impacts dining experiences. A cozy café on a rainy day might receive positive reviews, while an outdoor patio during a heatwave could lead to negative sentiments.
- Deployment:
  - Develop an intuitive interface for restaurant managers.
  - Provide real-time predictions of customer demand based on historical data, time of day, and special events.

- Ensure ease of use and accessibility.
- By using NLTK for performing various Natural Language Processing tasks.

- Evaluation:
  - Monitor system performance: Track prediction accuracy and responsiveness.
  - Gather feedback from restaurant staff: Assess usability and effectiveness.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.
  - Accuracy: It measures the overall performance of the model by comparing the predicted sentiments with the actual sentiments present in a labeled dataset.
  - Precision: It measures the percentage of correctly predicted positive or negative sentiments out of all the positive or negative predictions made by the model.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the Restaurant reviews prediction system. Here's a suggested structure for this section:

Data Collection: The system should be able to collect restaurant reviews from various sources such as online review platforms, social media, and customer feedback forms.

Data Preprocessing: The system should have preprocessing capabilities to clean the raw text data, handle punctuation, remove stop words, and perform stemming or lemmatization to normalize the text.

Sentiment Analysis Algorithms: The system should have the ability to implement various sentiment analysis algorithms such as rule-based approaches, machine learning models (e.g., Naive Bayes, Support Vector Machines), or deep learning models (e.g., Recurrent Neural Networks, Convolutional Neural Networks).

Feature Extraction: The system should be able to extract relevant features from the text, such as n-grams, word frequency, or TF-IDF (term frequency-inverse document frequency).

Training Data: The system should have a dataset of pre-labeled reviews for training the sentiment analysis model. This dataset should include positive, negative, and neutral reviews to ensure a balanced and accurate model.

Library required to build the model:

NLTK (Natural Language Toolkit):
NLTK is a widely used library for NLP tasks in Python. It provides various tools and modules for preprocessing text data, feature extraction, and building machine learning models.

```
Pip install nltk
Import nltk
Nltk.download('all')

#Tokenization using NLTK
From nltk import word tokenize, sent tokenize
Sent = "I am the king of the world. \ I am the king of the world.'
Print(word tokenize(sent))
Print(sent tokenize(sent))
```

scikit-learn: scikit-learn is a popular machine learning library in Python that offers a wide range of algorithms and tools for data preprocessing, feature extraction, model training, and evaluation. It includes algorithms like Naive Bayes, Support Vector Machines (SVM), and more.

TensorFlow and Keras: TensorFlow is an open-source deep learning framework, and Keras is a high-level API built on top of TensorFlow. They can be used for building and training deep learning models, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), which have shown good performance in sentiment analysis tasks.

Stemming

Stemming generates the base word from the inflected word by removing the affixes of the word. It has a set of pre-defined rules That govern the dropping of these always result in semantically meaningful base words. Stemmers are faster and computationally less expensive than lemmatizers.

Lemmatization:

Lemmatization involves grouping together the inflected forms of the same word. This way, we can reach out to the base form of any word which will be meaningful in nature. The base from here is called the Lemma.

ALGORITHM & DEPLOYMENT

- In the Algorithm section, describe the machine learning algorithm chosen for predicting bike counts. Here's an example structure for this section:
- Algorithm Selection:
  - One algorithm that is commonly used for sentiment analysis is the Support Vector Machine (SVM). SVM is a binary classification algorithm that aims to find a hyperplane that separates the positive and negative sentiment classes. It works well for sentiment analysis tasks with smaller datasets and when the data is linearly separable.
  - Another popular algorithm for sentiment analysis is the Naive Bayes classifier. Naive Bayes is a probabilistic algorithm that uses Bayes' theorem to calculate the probability of a document belonging to a specific sentiment class. It assumes that the features are independent, which may not always hold true for sentiment analysis, but it still performs well and is computationally efficient.

Data Input:

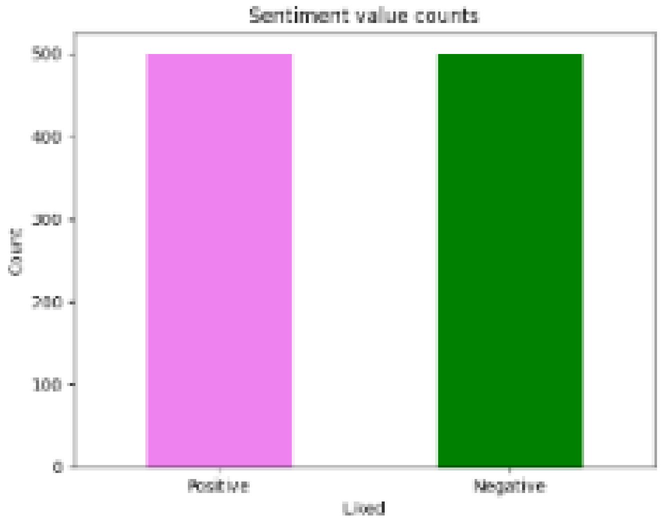
  - Bag-of-Words (BoW): This representation considers the frequency or presence of individual words in the text. Each word in the document is treated as an independent feature.
  - N-grams: N-grams are sequences of N words within the text. By considering not only individual words but also pairs or triplets of consecutive words, the model can capture some contextual information.
- Training Process:
  - Data collection: Historical restaurant reviews are collected, consisting of both the textual content of the reviews and their corresponding sentiment labels (positive or negative).

- Training the algorithm: A machine learning algorithm, such as a classification model (e.g., logistic regression, support vector machines, or neural networks), is trained on the preprocessed and feature-engineered data.
- Hyperparameter tuning: Machine learning algorithms often have hyperparameters that can be tuned to optimize their performance. Hyperparameters control the behavior and complexity of the algorithm, and finding the right values can significantly impact the model's performance.

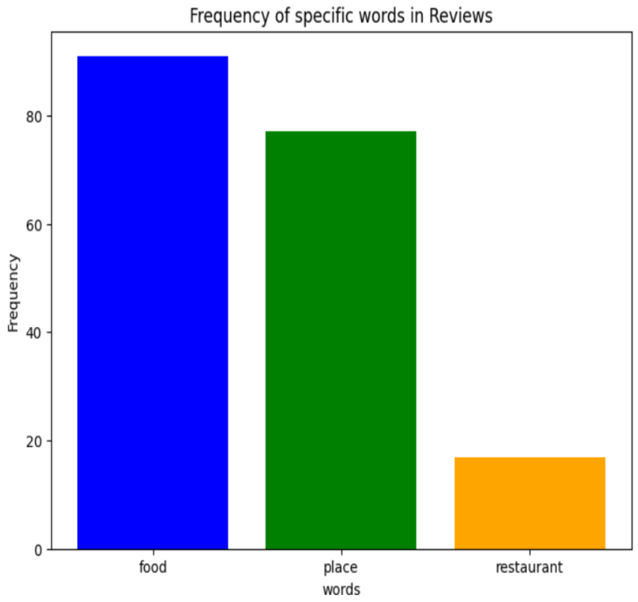
- Prediction Process:
  - Data preprocessing: The collected data is preprocessed to remove any irrelevant information, such as numbers or special characters. Text normalization techniques like lowercasing, stemming, or lemmatization may also be applied to reduce the vocabulary size and remove variations in word forms.

Post-processing and interpretation: After obtaining the sentiment predictions, any necessary post-processing steps can be applied to make the results more interpretable or useful. For example, setting a threshold on the probability scores to classify them into positive or negative categories or converting the scores into a sentiment intensity scale.

RESULT


Sentiment value counts
500
400
300
200
100
0
Count
Positive
Negative
Liked


Word Cloud of Reviews
food
place
good
service
back
will
time
great
really
restaurant
best
friendly
pretty
delicious
love
Vegas
go
onenice
amazing
came
staff
chicken
experience
eat


Frequency of specific words in Reviews
80
60
40
20
0
Frequency
food
place
restaurant
words

CONCLUSION

Overall, sentiment analysis is a valuable tool for analyzing restaurant reviews and predicting the sentiment associated with them. By using machine learning algorithms trained on labeled datasets, it is possible to accurately determine whether a restaurant review is positive or negative.

FUTURE SCOPE

- There will be a better understanding in the restaurant reviews.
- There is a lot of potential for the future scope of sentimental analysis using datasets in restaurant reviews.

Fine – grained sentimental analysis:

There is room for more fine-grained analysis that considers a wider range of sentiments, such as neutral, mixed, or specific emotions like happiness or anger.

REFERENCES

The reference of the Restaurant reviews by using the Educate skill development mentor reference data set.
I have some details about the data in github by using APSSDC

THANK YOU