Sentiment Analysis on Text Reviews Using Python and Pandas

Author

Rahul Gautam

MCA Student, DY Patil Institute of MCA, Akurdi, Pune

1. Introduction

This project involves performing sentiment analysis on a collection of user reviews. The goal is to clean the raw textual data and use it to classify the sentiment (positive/negative/neutral) for insights or downstream applications.

2. Problem Statement

Given a dataset containing user reviews and associated sentiments, clean the text and prepare it for analysis and machine learning tasks.

3. Tools & Technologies Used

- Python
- Pandas
- Regular Expressions (re)
- Google Colab / Jupyter Notebook
- Matplotlib / Seaborn (for visualization)
- Optional: Scikit-learn (for ML models)

4. Dataset Description

- Columns:

Sentiment Analysis on Text Reviews Using Python and Pandas

- review Raw text review provided by users.

- sentiment Label indicating sentiment.

- Source: (Mention if public or self-created)

5. Text Cleaning Process

Cleaning involved converting text to lowercase, removing punctuation, and handling missing values.

Code Used:

```
def clean_text(text):
    if not isinstance(text, str):
        return ""
    text = text.lower()
    text = re.sub(r'[^\w\s]', '', text)  # Remove punctuation
    return text

df['cleaned_review'] = df['review'].apply(clean_text)
```

6. Results Snapshot

Original Review	Cleaned Review
"Great product! Will buy again."	great product will buy again
"Terrible experience!!"	terrible experience

7. Conclusion

Sentiment Analysis on Text Reviews Using Python and Pandas

Text cleaning is a critical preprocessing step in NLP tasks. The cleaned dataset is now ready for further analysis, such as visualizations or model training for sentiment prediction.

8. Future Scope

Implement ML classifiers (e.g., Logistic Regression, Naive Bayes) to predict sentiment. Extend the dataset or deploy the model as a web application.

9. References

- https://pandas.pydata.org/

- https://docs.python.org/3/library/re.html