

Sentiment Analysis of Movie Reviews

Introduction

This project involves developing a machine learning model to analyze emotions expressed in movie reviews. By using natural language processing (NLP) techniques, we can understand how audiences feel about their favorite movies. The project utilizes Python, NLTK, Flask, and Matplotlib to achieve this objective.

Data Collection

The data was collected from IMDb by scraping movie reviews. The BeautifulSoup library was used to extract the reviews from the website. For the sake of simplicity, a predefined dataset was used in this example.

Data Preparation

The collected reviews were cleaned by removing special characters and extra spaces. The cleaned data was then used for further processing.

Data Labeling

The dataset was labeled using an existing corpus from the NLTK library. Emotions such as Joy, Sadness, Anger, Fear, Surprise, and Disgust were considered.

Data Visualization

Three types of visualizations were created to present the results of the sentiment analysis:

1. **Word Cloud:** Displays the most common words used in the reviews.
2. **Bar Chart:** Shows the distribution of different emotions.
3. **Pie Chart:** Illustrates the percentage distribution of emotions.

Code :

```
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import pandas as pd

# Sample text for the word cloud
sample_text = """
passionné extraction connaissances grandes quantités données visualisation faciliter prise
décisions stratégiques formation expériences professionnelles expertise conception
industrialisation rapports analytiques Power BI compétences avancées qualité données
diverses sources mise en place processus gouvernance fiabilité cohérence réalisations
tableaux de bord interactifs rapports dynamiques indicateurs clés performance décisions
éclairées analyse rigoureuse données interprétation résultats leviers performances
significatifs mise en œuvre plans d'action efficaces collaboration équipes techniques non
techniques promotion partage meilleures pratiques data analytics reconnaissance pairs
supérieurs contribution significative équipe opportunités collaborer entretien convenance
candidature attention salutations distinguées
"""

# Create a word cloud
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(sample_text)

# Display the word cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud - Movie Reviews Sentiment Analysis', fontsize=20)
plt.savefig('wordcloud.png')
plt.show()
```

```
# Example data for a bar chart
```

```
data = {'Emotions': ['Joy', 'Sadness', 'Anger', 'Fear', 'Surprise', 'Disgust'],  
        'Count': [50, 30, 10, 5, 3, 2]}
```

```
df = pd.DataFrame(data)
```

```
# Create a bar chart
```

```
plt.figure(figsize=(10, 5))  
plt.bar(df['Emotions'], df['Count'], color='skyblue')  
plt.xlabel('Emotions')  
plt.ylabel('Count')  
plt.title('Distribution of Emotions in Movie Reviews')  
plt.savefig('bar_chart.png')  
plt.show()
```

```
# Create a pie chart
```

```
plt.figure(figsize=(8, 8))  
plt.pie(df['Count'], labels=df['Emotions'], autopct='%1.1f%%',  
        colors=plt.cm.Paired.colors)  
plt.title('Percentage Distribution of Emotions in Movie Reviews')  
plt.savefig('pie_chart.png')  
plt.show()
```

Importance

This project demonstrates several key skills in data science and machine learning, including data collection, data cleaning, natural language processing, and data visualization. By sharing this project on LinkedIn or GitHub, you can showcase your ability to handle real-world data and extract meaningful insights, which is highly valuable to recruiters and employers.