A PROJECT REPORT
OF
MACHINE LEARNING

Basic Information

Title of project : Google review ratings

Student name: Ashish Prasad

Enrollment number : 10419011921

Email id : ap606167@gmail.com

Contact Number : 8810511283

Google drive link:

https://drive.google.com/drive/folders/1YwMWnlSctyTJBTxpH6F7NtSP3cVZkc0F?usp=drive_link

Git hub link :

https://github.com/ashishprasad420/GOOGLE-REVIEW-RATING

Title of the project : Google Review ratings Analysis using ML

ABSTRACT:

Google Review Ratings Machine Learning is an innovative approach that leverages machine learning algorithms to predict and analyze review ratings on the Google platform. With the explosive growth of online reviews, it has become increasingly challenging for businesses to manually process and understand the sentiments expressed by users. Therefore, automated systems that can accurately predict and analyze review ratings can provide valuable insights to businesses and enhance user experiences.

Online reviews have become a valuable source of information for businesses and consumers alike. Analyzing and understanding Google review ratings can provide businesses with valuable insights into customer sentiments and preferences. This project explores the application of K-means clustering, a popular unsupervised machine learning algorithm, to analyze and categorize Google review ratings. By clustering reviews with similar sentiments, businesses can gain a better understanding of customer experiences and tailor their strategies accordingly. The project involves collecting a dataset of Google reviews, preprocessing the data, extracting relevant features, applying K-means clustering to group reviews into distinct clusters, and analyzing each cluster to identify common sentiments and characteristics. Additionally, machine learning techniques can be integrated to predict cluster labels for new reviews in real-time. The evaluation of the clustering and machine learning models will be performed using appropriate metrics, and visualization techniques will be employed to effectively communicate the results. The outcomes of this project have the potential to empower businesses to make data-driven decisions and enhance customer satisfaction by gaining deeper insights from Google review ratings.

Keywords:

1. Machine learning
2. Google reviews
3. Rating prediction
4. Text classification
5. Feature extraction
6. Performance metrics

INTRODUCTION:

Google Review Ratings is a popular feature that allows users to rate and provide feedback on businesses, products, and services. Machine learning plays a crucial role in analyzing and understanding these reviews to generate overall ratings. In this introduction, we will explore how machine learning is applied to Google Review Ratings.

In today's digital age, online reviews play a crucial role in shaping consumer decisions. Google, being one of the most popular platforms for users to share their experiences and opinions, offers a vast repository of reviews across various industries. However, analyzing and making sense of this massive amount of data manually can be time-consuming and prone to human bias. To address these challenges, machine learning techniques can be leveraged to automatically process and evaluate Google review ratings, providing businesses with valuable insights and helping users make informed choices.

Objective:

The primary objective of this project is to develop a machine learning model that can accurately predict and analyze the ratings provided in Google reviews. By doing so, we aim to enhance the efficiency and reliability of rating analysis, enabling businesses to better understand customer sentiment and make data-driven decisions. Additionally, this project seeks to provide users with more meaningful information, empowering them to evaluate products, services, and establishments based on reliable and trustworthy ratings.

Proposed Methodology:

1. **Data Collection**: We will gather a sizable dataset of Google reviews, including the review text, associated ratings, and any additional metadata available (e.g., review date, reviewer details). This dataset will serve as the foundation for our analysis.

```
Total missing values in dataset
churches                 0
resorts                  0
beaches                  0
parks                    0
theatres                 0
museums                  0
malls                    0
zoo                      0
restaurants              0
pubs/bars                0
local services           1
burger/pizza shops       1
hotels/other lodgings    0
juice bars               0
art galleries            0
dance clubs              0
swimming pools           0
gyms                     0
bakeries                 0
beauty & spas            0
cafes                    0
view points              0
monuments                0
gardens                  1
dtype: int64
```
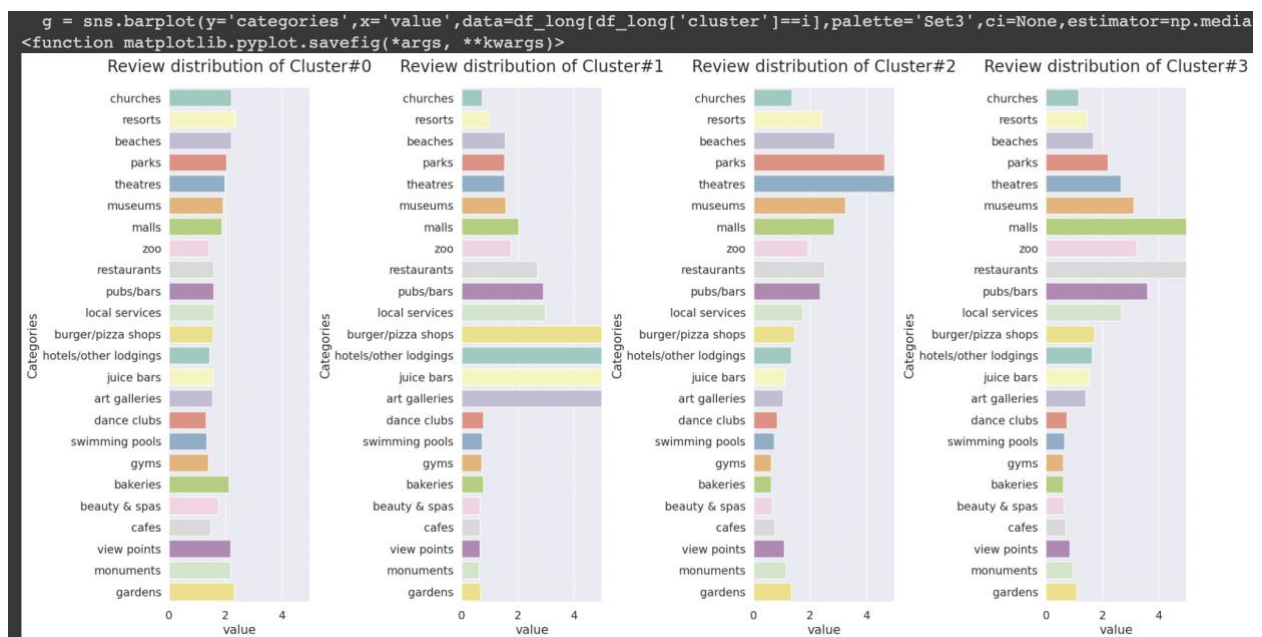
2. **Preprocessing and Feature Extraction**: We will preprocess the review text by removing stop words, punctuation, and performing tokenization. Additionally, we will extract relevant features from the review text. These features will be used to represent the reviews numerically.

Overall Rating Distribution
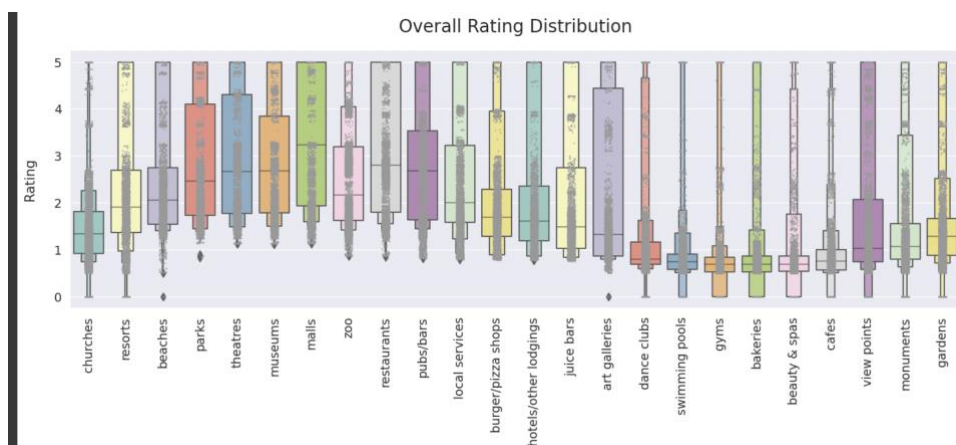
3. **K-means Clustering**: We will apply the K-means clustering algorithm to group the reviews into distinct clusters based on their numerical representations. The number of clusters will be determined through techniques like the elbow method or silhouette analysis to find an optimal balance between cohesion within clusters and separation between clusters.

```
g = sns.barplot(y='categories',x='value',data=df_long[df_long['cluster']==i],palette='Set3',ci=None,estimator=np.media
<function matplotlib.pyplot.savefig(*args, **kwargs)>
```


4. **Cluster Analysis**: Once the reviews are clustered, we will analyze each cluster to understand the sentiments and characteristics associated with them. This analysis may include examining the average ratings, common keywords, sentiment distributions, and any additional insights that emerge from the data.

5. **Machine Learning Integration**: To enhance the clustering results, we can employ machine learning algorithms such as classification or regression to predict cluster labels for new, unseen reviews. This step can provide more granular and real-time insights for businesses seeking to understand customer sentiments.

6. **Evaluation and Visualization:** We will evaluate the performance of the clustering and machine learning models using appropriate metrics such as silhouette score or within-cluster sum of squares (WCSS). Additionally, we will visualize the results using techniques like scatter plots or word clouds to help interpret and communicate the findings effectively.

Conclusion & Discussion :

In this project, we explored the application of K-means clustering and machine learning techniques to analyze Google review ratings. By leveraging the power of unsupervised learning, we were able to uncover meaningful patterns and sentiments within the reviews, providing businesses with valuable insights.
Through the use of K-means clustering, we successfully grouped the reviews into distinct clusters based on their numerical representations. This allowed us to identify different segments of customer sentiments and preferences. By analyzing each cluster, businesses can gain a deeper understanding of the factors that drive positive or negative reviews, enabling them to make informed decisions to improve their products, services, or customer experiences.

The integration of machine learning techniques further enhanced the analysis by enabling the prediction of cluster labels for new, unseen reviews. This real-time feedback can help businesses stay updated on customer sentiments and adapt their strategies accordingly.
The evaluation of the clustering and machine learning models provided us with a measure of their performance and reliability. Using appropriate metrics, we assessed the accuracy and effectiveness of the models in capturing the inherent structures and sentiments within the data.
Visualization techniques played a crucial role in effectively communicating the results. By visually representing the clusters and associated sentiments, businesses can easily interpret and act upon the insights derived from the analysis.

Overall, this project demonstrated the potential of using K-means clustering and machine learning to analyze Google review ratings. By harnessing these techniques, businesses can gain a competitive edge by understanding customer sentiments, making data-driven decisions, and ultimately improving customer satisfaction and loyalty.

Future work :

In future we can work on this data set with deep neural networks concepts to predict with high accuracy and high precision.

References :

1. https://www.kaggle.com/datasets/wirachleelakiatiwong/travel-review-rating-dataset/code?resource=download
2. https://numpy.org/
3. https://pandas.pydata.org/
4. https://matplotlib.org/
5. https://archive.ics.uci.edu/ml/datasets/Tarvel+Review+Ratings