Prepared by Team Zodiac

KRACK THE HACK

by consulting and analytics club

8th September 2024

CUSTOMER SENTIMENT ANALYSIS

1. Data Preprocessing

Loading hotel review data from CSV file
Cleaning and preprocessing text data (remove HTML tags, special characters, etc.)
Convert text to lowercase and remove stop words

2. Sentiment Analysis

Using **NLTK's SentimentIntensityAnalyzer** to assign sentiment scores to reviews
Categorizing reviews as positive, negative, or neutral based on compound sentiment score

3. Topic Modeling

Applying **Latent Dirichlet Allocation (LDA)** to identify main topics in reviews
Extract top words for each topic
Assign topics to individual reviews

4. Statistical Analysis

Performing **ANOVA** to identify significant relationships between hotel names, ratings, and sentiment scores
Identifying hotels with statistically significant effects on ratings

5. Aspect-based Sentiment Analysis

Defining key aspects of hotel reviews (e.g., Cleanliness, Staff, Amenities, Price)
Analyzing sentiment for each aspect across different hotels
Counting positive and negative mentions of each aspect

6. Results Compilation

Creating a CSV file summarizing sentiment analysis results for each hotel and aspect

PERSONALISED CUSTOMER SUPPORT

Data Preparation

Loading hotel booking data from dataset

Selecting relevant features (e.g., search parameters, hotel info)

Cleaning and preprocessing data if necessary

Feature Engineering

Identifying key features for prediction

Encoding categorical variables if needed

Handling missing data or outliers

Model Selection

Choosing **Random Forest Classifier**

Setting initial parameters (n_estimators=100, random_state=42)

Data Splitting

Dividing data into features (X) and target variable (y)

Splitting into training and testing sets (80% train, 20% test)

Model Training

Fitting Random Forest model on training data

Optional: Performing cross-validation or hyperparameter tuning

Prediction

Using trained model to predict hotel clusters on test set

Model Evaluation

Calculating accuracy score

Analyzing model performance

Optional: Generating additional metrics or visualizations

Results Compilation

Summarizing model accuracy and key findings

Identifying most important features for prediction

Preparing insights for presentation or further analysis

Team Members

Saanchi
Member

Vedant
Member

Surbhi
Leader

Thank you