Sentiment Analysis and Rating Prediction on Hotel Reviews

Milestone - 2
Raj Gandhi
Advisor: Dr. Rajendra Raj
Topics to cover

- Summary of Project
- Milestone - 1 review
- Milestone - 2 progress
- Next steps - Milestone - 3
Summary of the Project

- Project focuses on doing sentiment analysis on the hotel reviews.
- Getting the customer’s interest
- Implementing Random Forest classifier for sentiment analysis
- Creating a web based UI for visualizing recommendation
Milestone - 1 review

- Literature review studies on 3 research papers
- Literature review - 1 - tripadvisor dataset -> Naive Bayes classifier -> 72% accuracy
- Literature review - 2 - amazon review dataset -> SVM, Naive Bayes and Random Forest -> 94% accuracy while others below 90%
- Literature review - 3 - product reviews dataset -> Similar results as above
- Clean dataset
Define Categories

- Four main categories - Landscape, Business, Food, and Luxury.
- WordNet® is a lexical database of English. Developed by Princeton University
- Links words to its semantic relations.
- Get hyponyms of all 4 categories
Data preparation

- Tokenize words
- Remove stop words (e.g: “is”, “are”, “was”). NLTK library used.
- Remove single character words (e.g: full stop, comma, etc).
- Parts of Speech tagging.
- Lemmatize words -> convert words to root.
Classification

- Label 70% data using textblob sentiment analyzer and user rating.
- Using NLTK sentiment analyzer with Random Forest classifier.
- Accuracy: 94%
Milestone 3 roadmap

- Create a UI for data visualization.
- Compare results with other classifiers.
- Identifying improvements and future works.
- Final Report and Poster.
Thank you