Sentiment Analysis & Neural Networks

An insight into hotel reviews and predictions

Nawaz A.Rahman

Thesis Report 30p

Simon Celinder

Data Scientist

Göteborg | 2 May 2023

# Table Of Contents

Sentiment Analys with RNN 1

Table Of Contents 2

List of abbreviations 3

Background 4

Boundaries 5

Methods 5

# List of abbreviations

NLTK - Natural language toolkit

EDA - Exploratory Data Analysis

N\_grams -

LSTM - Long short term memory.

# Background

Sentiment analysis is a technique that has found greater use in a variety of industries, including hotel management. This project's goal is to conduct a sentiment analysis of hotel reviews using NLP tools like LSTM and NLTK. The project will investigate how effectively these tools function in order to categorize hotel reviews as either positively or negatively based on the information provided.

One of the major benefits of using sentiment analysis is that it allows for a quick and efficient analysis of large amounts of data, which may be quite helpful for hoteliers. This project will contribute to a better understanding of how the hotel experience may be improved by examining what factors influence the classification of reviews as positive or negative.

The outcome of the sentiment analysis can also be used to enhance the guest experience at hotels. The most common word combinations and phrases used in negative reviews can be analyzed to help hoteliers pinpoint the most pressing and pressing problem areas. On this basis, hoteliers can enhance their services and, as a result, improve customer satisfaction and community.

In conjunction with this project, LSTM and NLTK will be improved to increase their effectiveness for sentiment analysis of hotel reviews and, as a result, make it possible for hoteliers to quickly and easily identify and fix issues that impair guests' experiences.

# Boundaries

The project's primary goal is to analyze the sentiments of English-language hotel reviews that have been collected from online platforms. Only text data will be used to train and test models with the aid of NLP tools like LSTM and NLTK. The project won't include implementing the results; rather, it will just provide guidance on how to use them to improve the guest experience. The project's timeframe is restricted to the class term, and work will be done on regular workdays.

# Methods

The following steps will be taken to complete the project:

1. Data Collection: Collect hotel reviews from various online platforms.
2. Data Preprocessing: Preprocess the text data by removing stop words and applying tokenization, stemming, and lemmatization using NLTK.
3. Exploratory Data Analysis (EDA): Analyze the distribution of positive and negative reviews, and identify the most common words and n-grams in the reviews.
4. Data Preparation: Create a dataset with a positive column and a word count column. Then, preprocess the text data by applying tokenization and stemming using NLTK, and generate n-grams for all frequent words and for negative words.
5. Model Building: Build a recurrent neural network (RNN) model using LSTM layers. Split the dataset into training and testing sets, determine the number of words, perform visual inspection, and use the interquartile range (IQR) method to remove outliers. Tokenize the text and create a word-to-index dictionary, pad the text, and perform one-hot encoding on the target variable. Reshape the data from 2D to 3D for model training.
6. Model Training: Train the RNN model using the training dataset and save the weights using ModelCheckpoint.
7. Model Evaluation: Evaluate the performance of the model on the testing dataset using accuracy, precision, recall, and F1-score. Generate a confusion matrix to visualize the model's performance.
8. Making Predictions: Use the trained model to make predictions on new hotel reviews and analyze the results.
9. Conclusion and Recommendations: Draw conclusions from the results of the sentiment analysis and provide recommendations for improving the hotel experience based on the findings.