SENTIMENT ANALYSIS

Introduction

Sentiment analysis is a type of classification where data is classified into different classes which are binary in nature or can have multiple classes. It is a process of extracting opinions that have different polarities, that is, positive, negative, or neutral. This is also known as opinion mining and polarity detection. The aim of the project was to build a sentiment analysis model that would allow us to categorize words based on their sentiments, that is, whether they are positive, negative, and the magnitude of it. The dataset used was provided by the R package 'janeaustenR'. This package provides the textual data in the form of books authored by the novelist Jane Austen.

Analysis

The three general purpose lexicons used include AFINN (scores words from a range of -5to5), bing (classifies the sentiment into a binary category of negative or positive), and Loughran (performs analysis of the shareholder's reports). Words present in the book 'Emma' were visualized based on their corresponding positive or negative scores. Sentiment scores were also visualized and plotted along the axis that is labeled with positive and negative words. Finally, a wordcloud was created that would delineate the most recurring positive and negative words.

