Sentiment Analysis using Social Media

Grand Text Auto

1. INTRODUCTION

Social Media has become a ubiquitous source of information in the modern world. Many people use multiple platforms for social media, consuming for both entertainment and professional purposes. From a marketing perspective, it has become impossible to ignore the influence that Social Media has on consumers. Social Media also provides companies with the ability to directly interact with their consumers, gauging their interest in products and services. As automated techniques for text analysis grow more popular, it is now possible to understand the underlying sentiments of consumers. Latent thoughts and perspectives can be detected using techniques such as Sentiment Analysis.

In this project, we will explore the effect of news and communications from brands on the overall sentiment about the respective products[2]. Specifically, we will focus on Sony Playstation and Microsoft Xbox brands with both planning to release their computing products in November 2020. We plan to use sentiment analysis to find out what terms were positively associated with each product and what terms were negatively associated. This type of analysis can be useful for a company to find out where they are trailing their competition and use that knowledge to plan future communications.

To achieve our objective, we will train a model based on previously labeled data and use it to predict sentiment on new tweet data related to Xbox and Playstation. We will use these predicted labels to analyze the change in consumer sentiment towards these products as they approach their release.

2. DATASET DESCRIPTION

For Training our models, we will be using the sentiment140 dataset from Kaggle[3]. It contains 1.6m tweets that have been labeled for the sentiment. The labels range between 0 and 4, with 0 being negative sentiment, 2 being neutral sentiment, and 4 being positive sentiment.

The dataset contains the following attributes for each tweet:

- The **target** sentiment
- The **id** of the tweet
- The **date** in which the tweet was published
- The **user** that sent the tweet
- The **content** of the tweet

Using the Twitter API Tweepy[1], we will scrape tweets related to the Xbox and Playstation as the test dataset, which are all unlabeled texts. We will use this data to evaluate the change in sentiment over time for the two products.

3. LIST OF DELIVERABLES

- Historical and Newest dataset collected via Twitter API
- Train models based on Sentiment140 dataset
- Overall brand sentiment for both brands and its change over time
- Data Visualization
- Project Presentation
- Any source code, file, and resource used for the project

4. REFERENCES

[1] Tweepy api, Sep 2020.

[2] S. Guest. Sap brandvoice: Ps4 vs. xbox one: Winner emerges via social media analysis, Nov 2013.

[3] KazAnova. Sentiment140 dataset with 1.6 million tweets, Sep 2017.