Mohan Rao B C

mohanraobc@gmail.com

Abstract

Amazon Reviews and Ratings – What are people talking about? How good/bad is it? How are the ratings related to Sentiments of the people?

Sentiment Analysis

Edvancer – Certified Text Analytics Expert Certification Project Report

## Objective

Any online seller as big as amazon rely heavily on online product reviews to collect feedback regarding the products which are listed on the website. This is an important part of keeping quality check on non-amazon sellers listed on the market place.

These reviews in the form of text are hard to analyse because of their unstructured nature. Sentiment analysis one quick tool to give amazon analyst an idea about how the seller's products are viewed in favourable/unfavourable manner by the customer buying the products.

The input data has 31000 reviews and ratings and following are the questions to be answered

* What is the product that people have reviewed and rated?
* How are the emotions distributed across the reviews?
* Which is the strongest emotion and what the elements that people have the strongest emotion for?
* What is the sentiment of people?
* What are people talking about negatively/positively?
* How are ratings related to sentiments and emotions?

# What are the people talking about? What is the product that people have reviewed and rated?

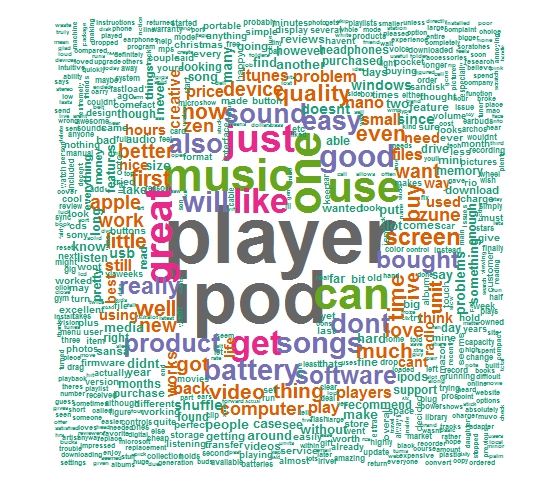


Figure 1: Word Cloud of the Total Corpora of the Reviews

Well, the answer is right at the middle the word cloud. The reviews and talking about the music player. On deeper study of the word cloud, we can observe the terms like “ipod”, “apple”, “itunes”, “battery”, “quality”, “scandisc”, “download” etc. These terms indicate that the object of the reviews in generic sense is MP3 Music Player. The reviews are for the different brands of the Mp3 music players, options, product features, ease of use and pricing.

One more undeniable insight is that, most of the reviewers are talking about apple iPod. Which is a testimony to the extreme popularity of the apple brand in the music player market currently

# How are the emotions distributed across the reviews?

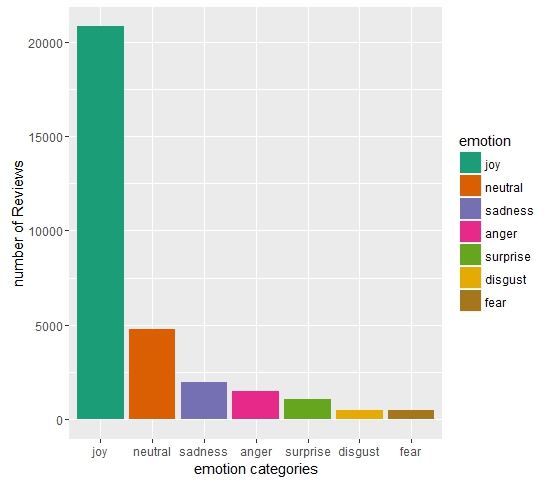


Figure 2: Distribution of “Emotions” expressed in the reviews

Based the “emotions” attached to the reviews using Baye’s algorithm, most the reviews are expressing joy as the emotion. In fact, almost 70% of the reviews are expressing joyful emotions. It may be indicative of the hypothesis that almost 70% of the reviewers are happy with their music players and are describing the features they like

Of course it might be worthwhile to explore the reviews associated with “sadness”, “anger”, “disgust” and “fear”, as they might throw light on which are the brands of music players that people are not happy with? What the features that people are sad about? Which features are making them angry?

# Which is the strongest emotion and what the elements that people have the strongest emotion for?

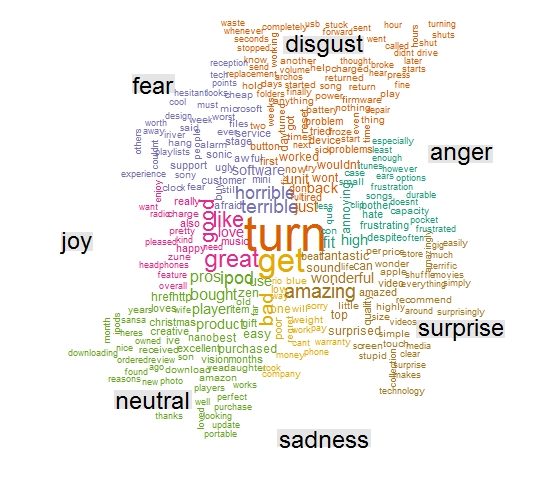


Figure 3: “Emotions” related to topics and features related to music players

## Joy

People are expressing joy about

* Radio feature
* Charge of the player
* Microsoft Zune
* Feature of the player
* Head phones etc

## Neutral

People are neutral emotions about

* Sansa Brand
* iPod, Nano
* Download feature
* Photos feature
* Amazon (probably delivery)

## Surprise

People seem to be surprised about

* Ipod Shuffle
* Price
* Video feature
* Screen
* Touch feature
* Movies feature

## Anger

People seem to be angry about

* Capacity (possibly storage)
* iTunes
* Clip (possibly)
* Durability

## Disgust

People seem to be expressing disgust about

* Firmware
* Battery
* Tuning
* Hours (Song Play Time)
* Freezing

## Fear

People seem to be expressing disgust about

* Reception (probably radio)
* Software
* Customer Service
* Sony
* Sonic

# What is the sentiment of people?

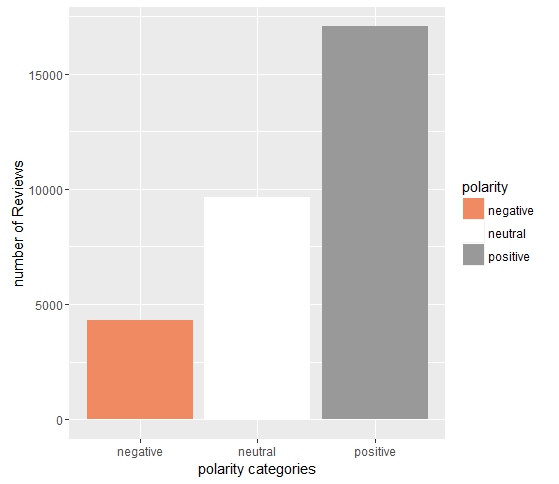


Figure 4: Distribution of “Sentiment” expressed in the reviews

As expected based on the analysis of the emotions in the reviews, most the reviews are “Positive” in the nature. Almost 55% of the reviews are positive in nature and only 13% of the review are negative

With both the “emotions” and “Sentiment” analysis, it is reasonable to conclude that most of the people are happy and upbeat about the mp3 music players in general and are liking most the product features across brands.

Extending the hypothesis, it is also safe to assume that apple is the most popular brand and most of the reviewers are happy about the iPod and are also spreading positive word about the apple iPod

# How are ratings distributed and how are they related to sentiments ?

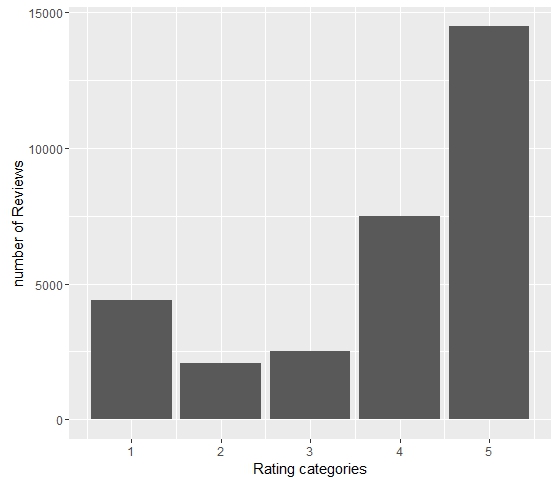


Figure 5: Distribution of “Ratings” given in the reviews

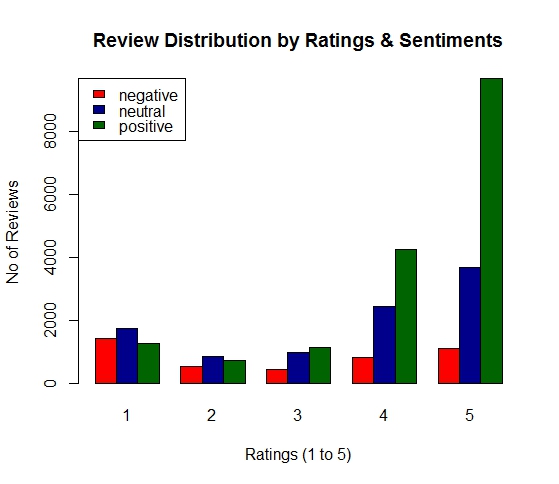
About 69% of the reviews have the rating of more than 4. The conclusion that the reviews are generally favourable to the music players is true as most of the reviews are positive in nature and also express joy as the emotion. Based on the analysis of sentiments, emotions and ratings, people are happy with most of the brands of the music player and also are happy about most of the features. But there are few areas of improvements like battery life, firmware, software freezing etc which needs attention

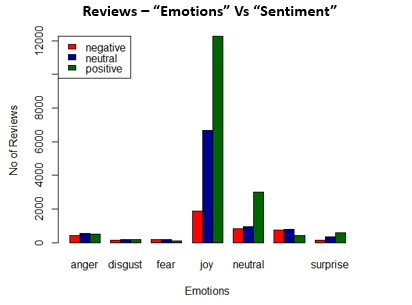
Figure 6: Distribution of “Ratings” associated with “Sentiment” in the reviews

The major chunk of negative reviews is associated with rating value of 1. This is expected and the area for further exploration to understand the improvement areas.

On the other hand, the major chunk of the positive reviews is associated with ratings more than 4.

There are few issues like finding some of the positive reviews with ratings 1 and 2. The reason might be that the model lacks accuracy or the people have behaved that way.

# Are “Sentiment” and “Emotion” related? How?

Figure 7: Distribution of “Sentiment” associated with “Emotions” in the reviews

Conforming to the insights derived earlier, the major chunk of positive reviews is also expressing “Joy” or “Neutral” emotions.

This shows that model is performing well for major part of the review. But there is scope for improvements like negative reviews associated with “joy”. The reason may that people have written sarcastic review in a joyful way or the model might be classifying them wrongly. It would need bit more exploration to confirm the hypothesis