**Amazon Fine Food Reviews**

**Sentiment Analysis**

**Asif Tahashildar**

**KLE College Of Engineering and Technology,Chikodi**

[**asiftahashildar17@gmail.com**](mailto:asiftahashildar17@gmail.com)

**T Aswin Barath**

**Jain University**

[**Aswin2001barath@gmail.com**](mailto:Aswin2001barath@gmail.com)

**Sai Varshini PV**

**Acharaya Institute Of Technology**

[**pv.beis.16@acharya.ac.in**](mailto:pv.beis.16@acharya.ac.in)

**Abstract-Sentiment Analysis is a technique widely used in text mining. Sentiment Analysis, also known as Opinion Mining, is primarily for analysing conversations, opinions, and sharing of views for deciding business strategy, political analysis and also for assessing public actions. As consumers move towards social media platforms like Twitter, Facebook and Amazon to air their views about a variety of products, performing sentiment analysis on their responses becomes a desirable activity that can return a wealth of information about public perception. Thus most work in this field has always concentrated on polarity detection of the opinion into two broad fields of positive or negative. The development of the Internet changed the way people eat and responding for food. Amazon is a biggest website where users can easily purchase all kind of food they need. In our project we aim at reviews given on food products. Reviews on Amazon are not only related to the product but also the service given to the customers. If users get clear bifurcation about product reviews and service reviews it will be easier for them to take the decision, in this paper we propose a system that performs the classification of customer reviews followed by finding sentiment of the reviews.**

**Keywords— Sentiment analysis, Amazon food reviews, classification.**

I. INTRODUCTION

The age of Internet has changed the way people express their views. It is now done through blog posts, online discussion forums, product review websites etc. People depend upon this user generated content to a great extent. When someone wants to buy a product, they will look up its reviews online before taking a decision. The amount of user generated content is too large for a normal user to analyze. So to automate this, various sentiment analysis techniques are used. Sentiment Analysis in Amazon is quite difficult due to its short length. Presence of emoticons, slang words and misspellings in food reviews forced to have a preprocessing step before feature extraction. There are different feature extraction methods for collecting relevant features from text which can be applied to food reviews also. But the feature extraction is to be done in two phases to extract relevant features. In the first phase, Amazon specific features are extracted. Then these features are removed from the food reviews to create normal text. After that, again feature extraction is done to get more features. This is the idea used in this paper to generate an efficient feature vector for analyzing Amazon sentiment.

By doing sentiment analysis on a specific domain, it is possible to identify the influence of domain information in choosing a feature vector. Different classifiers are used to do the classification to find out their influence in this particular domain with this particular feature vector.

II. RELATED WORKS

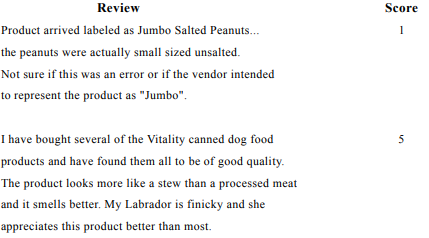
Concerning sentiment classification, Pang et al, incorporated a standard bag-of-features framework to predict the sentiment class of movie reviews. Their results showed that machine learning techniques using bag-of-words features outperformed simple decision-making models which used hand-picked feature words for sentiment classification

Traditional approaches on sentiment analysis use word count or frequencies in the text which are assigned sentiment value by expert. These approaches disregard the order of words. A recurrent neural network (RNN) can be used for sequence labelling on sequential data of variable length, which is natural for sentiment analysis tasks where the input sentence is viewed as a sequence of tokens.

Recent works explore the Gated Recurrent Units neural network(GRU) on the task of sentiment classification. GRUs are a special case of the Long Short-Term(LSTM) neural network architecture. GRUs are effective in this task because of their ability to remember long time dependencies. Furthermore, GRUs are faster to train and converge than LSTM networks.

III. DATA

We work on the Amazon Fine Food Reviews dataset which contains 568,454 reviews. The dataset consists of a single CSV file, which includes the ids of the products, ids of the reviewers, the scores(rating between 1 and 5) given by the reviewers, the timestamp for each review, a brief summary for each review, and the text of the reviews. We extract the columns of scores and review texts as our labels and raw inputs. Sample reviews with different scores are shown below:



IV. ARCHITECTURE AND METHODOLOGY

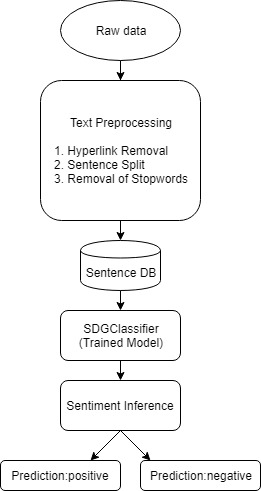


Fig 4.1. Block Diagram

Step 1- Gathering data:Amazon Food Reviews was used as a source of text in this project, we gathered reviews for sentiment analysis to analyse if it’s a positive or a negative review. Amazon Food Reviews(although not being the most efficient way )was used to gather text data.

Step 2-Text Preparation:Text preparation is nothing but filtering the extracted data before analysis.  It includes identifying and eliminating non-textual content, stopwords removal, punctuation removal and stemming.

Step 3-Sentiment Detection: We use SDGClassifier to classify the food reviews. We also consider the Score attribute to separate reviews which have score of 1 and 2 from 4 and 5.

SGDClassifier: SGD stands for Stochastic Gradient Descent: the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing strength schedule.

class sklearn.linear\_model.SGDClassifier(loss='hinge', penalty='l2', alpha=0.0001, rho=0.85, fit\_intercept=True, n\_iter=5, shuffle=False, verbose=0, n\_jobs=1, seed=0, learning\_rate='optimal', eta0=0.0, power\_t=0.5, class\_weight=None, warm\_start=False)[¶](https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/modules/generated/sklearn.linear_model.SGDClassifier.html#sklearn.linear_model.SGDClassifier)

Step 4-Sentiment Classification: Sentiments can be broadly classified into two groups, positive and negative.

V. WORKING OF WEBPAGE

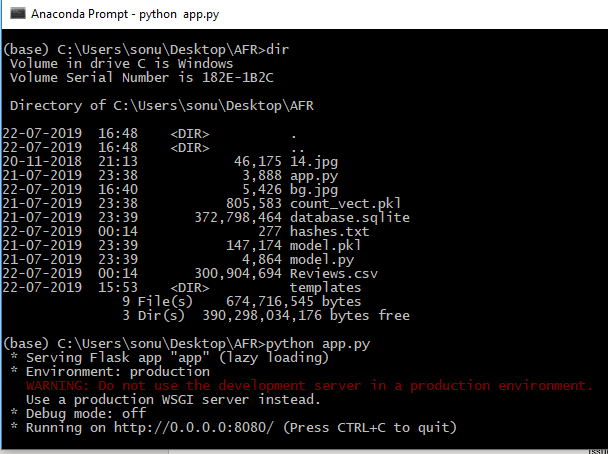


Fig 5.1. Run app.py on Anconda Prompt

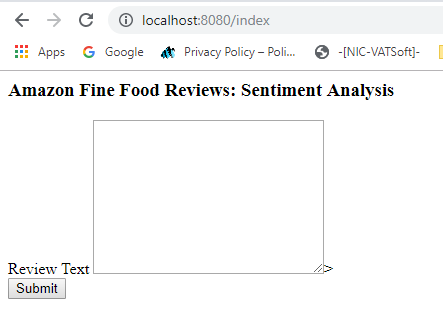


Fig 5.2. Open browser with port no. 8080



Fig 5.3. Place any text from the review database

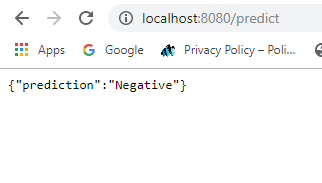


Fig 5.4. Outputs the sentiment of the review

VI. RESULTS

In this way , we can predict if the given review is a positive or a negative review. This sentiment analysis is done as social media is increasingly becoming the platform of communication for every means. Businesses can effectively utilize this by carefully listening and monitoring consumers. Companies can look into these sentiment analysis result instead of reading every review as it is not time consuming and more efficient. . The program is using a machine-based learning approach which is more accurate for analyzing a sentiment, together with natural language processing techniques will be used.As a result, program will be categorized sentiment into positive and negative

VII. CONCLUSION

Social media has not only become an integral part of our daily lives but also has become a valuable source of resource for businesses. With low costs, businesses are now able to gather actionable insights easily. From customer service to marketing, Sentiment analysis can improve everything. However, it should be kept in mind that no machine learning algorithm can provide an accuracy of 100% due to complexity and limitations of the technology. So, despite some of its limitations, sentiment analysis has proven to be a valuable opportunity for businesses to grow.

VIII. REFERENCES

[1] V. Nareyko, “Why python is perfect for startups,” Retrieved 01 10, 2014 from: <http://opensource.com/business/13/12/why-python-perfect-startups>

[2] A. Sweigart, “Invent your own computer games with Python. 2nd edition, 2012. Retrieved from <http://inventwithpython.com/>

[3]Internet World Start, “Usage and Population Statistic”,Retrieved,10,15,2013 from:<http://www.internetworldstats.com/stats.htm>

[4] P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.

[5]AppliedAI Course

[6] R. Xia, C. Zong, and S. Li, “Ensemble of feature sets and classification algorithms for sentiment classification,” Information Sciences: an International Journal, vol. 181, no. 6, pp. 1138–1152, 2011.