

Sentiment Analysis using Microsoft Azure Machine Learning and Python

Rohan Singh

Department of Computer Science and Engineering
Sachdeva Institute of Technology and Management
Mathura (U.P), India

Pankaj Sharma

Department of Computer Science and Engineering
Sachdeva Institute of Technology and Management
Mathura (U.P), India

Abstract— Sentiment analysis is a task to find out the polarity of a particular sentence. A particular sentence could be positive, negative or neutral. The polarity of that sentence is basically his/her frame of mind, feelings and opinion of that person about a particular product or service. Amazon is among the most popular platforms of online shopping currently. It is the biggest platform for selling and purchasing of products. As Amazon has products from millions of sellers that give a particular purchaser huge scope and variety of products. Amazon allows users to share reviews about a particular product and what they experienced after buying it. In this study using Microsoft Azure Machine Learning and Python a study is conducted to find that a particular review is positive, negative or neutral. This was demonstrated by performing a sentiment analysis on reviews of particular products. Accordingly, data was sourced from Amazon's review section. Data were obtained in the form of sentences. Azure cloud-based sentiment analytics and python was applied to conduct this sentiment analysis. The outcomes of this study confirmed that Microsoft Azure Machine Learning platform and python programming can be used for sentiment analysis on various data models to perform data analytics.

Index Terms- Sentiment analysis, opinion mining, Microsoft azure, Python, Amazon, reviews.

I. INTRODUCTION

The field of data analysis implies dissecting data in manners that uncover connections, designs, patterns found inside it. It basically gives us the result to what in particular level we can believe about appropriate responses we are getting by contrasting our data with others to find out needed result from the information. Basically there are two kind of data qualitative data that alludes to the information which is gathered or can be converted into numbers that can be represented numerically. And other one is qualitative data that is gathered in a format as it is descriptive, sentiment, quotes, translations, and so on. The famous and most basic tool for data analysis is MS Excel in which data can be arranged in the form to be in ascending and descending order, channel the information that meet specific models, display charts, perform conditional formatting on cells, get required information out from an enormous and itemized data collection utilizing pivot tables, formatting of tables to dissect our information rapidly and effectively. In this research work, an Add-in of MS Excel named Microsoft Azure Machine Learning that is basically used to do sentiment analysis was used. Many other tools like Analysis ToolPak which is an include program that gives solution for data analysis to do analysis of monetary, factual and designing engineering data analysis. Text data contains sentiment and feelings.

Sentiments are elements, occasions and their properties likewise called as target articulations. Conclusions depict an individual's sentiments, opinions, evaluation's towards elements, occasions and their properties.

Regardless of whether a particular decision is made or yet to be made is coordinating with the sentiments of the others to stay away from self-misfortune. One reason for the absence of sentiments is the way that there was minimal stubborn message accessible before the Internet. Prior to the web, when an individual expected to settle on a choice, the person in question requests feelings from loved ones. In any case, the web changed the way that individuals express their perspectives and sentiments by posting surveys of items at Ecommerce platforms, comments on social media platforms and websites which are on the whole called client created content. This internet based sentiment conduct addresses new and quantifiable wellsprings of data with numerous reasonable applications. Sentiment analysis is performed by utilizing the approach which can review about opinions that is also known as opinion mining, is intended to dissect the assessment of individuals towards anything like items, associations, and other related traits. In this current day, social media assumes a huge part in giving quality data about any topic going from various surveys websites, and remarks. Today, people, organizations and establishments are progressively utilizing the data available over web to advise their choices. At the point when an individual needs to purchase an item, he presently not just asks individuals around him however gets a ton of remarks, conversations, and other data about the item from the Web. For an association, the evaluation of these items and service can likewise be very important as it represents the image through the Web. Also, it is simple these days for companies to get public input on their services, just as to find out with regards to significant occasions in different areas. Be that as it may, because of the multiplication of different sites, online evaluation and checking of assessment sites and removing data from them is as yet a troublesome task. The normal human will struggle distinguishing applicable destinations, extricating and summing up remarks from them. Along these lines, evaluation framework is required.

II. LITERATURE REVIEW

A. Sentiment Analysis:

The field of sentiment analysis is a part from the field of opinion mining. Many researchers worked for this field to draw out the best result. An overview has been done to investigation about the procedure and devices accessible in sentiment analysis.

J.M. Weibe [1], has worked to find out various algorithm in best recognizable way to carry out sentiment analysis.

The research work of M. A. Hearst [2] was based upon including intelligence for sentiment analysis.

The research work of V. Suresh [3] was based upon a methodology that pre-owned stop words and space between stop words as the feature that helps sentiment analysis.

The study of Murthy G. et al. [4] was based upon a report on sentences and web based feelings.

During the study Dave et al. [5] utilized a tool to integrate audits.

Matsumoto et al. [6] in the study investigated about report level syntactic over the words that was found by them.

During the study Liu et al. [7] represented about multilevel classification that was based upon sentiment analysis.

The research work of Harvinder Jeet Kaur et al. [8] was related with investigation on various techniques to perform automatic polarity classification of textual information.

Emilio et al. [9] basically presented that an incredible general software subsystems will empower numerous different applications that need mixture of streaming and cluster information investigation.

J. Prabhu et al. [10] examined about the utilization of Rapid clustering Technique to dissect the attributes in social networks.

Xin Chen et al. [11] utilized an exceptional framework called Social Web Examination Buddy to dissect understudy posted substance via Social Media destinations to work with the comprehension of human practices and social propensities.

B. Microsoft Azure Machine Learning:

Microsoft Azure Machine Learning is a service provided by Microsoft that envelops cloud benefits which empowers the formation, arrangement, and data maintenance by means of a framework which is established based upon world wide networking of data centers for Microsoft. It is a cloud based model which contains the cloud facilities that helps in separating highlights that are based upon specific adaptability, readiness, and versatility. Presently, Azure computes the contribution score of the client dependability by means of social media metrics. This can be considered as the simple evaluation of the worth of clients of Microsoft add to its cloud business via social media to empower it and can offer many types of assistance. Azure ML additionally upholds numerous Machine Learning Algorithms that are classification, regression and clustering.

As per the study of Qasem et al. [12] Azure ML model is the customization of models utilizing python and R.

As per the research work of Ericson et al. [13] Azure ML studio takes into account the moving of Modules and datasets like ML algorithm, highlight choice, and pre-handling and connections them together. This test can be prepared and changed into a predictive analysis. This prescient test permits clients to fabricate their models.

Microsoft Azure is intended to set a playground for both experienced and fresher's who intend to perform data analysis. It provides an assortment of algorithm with just a solitary clustering algorithm. Azure ML is regularly described by the Cortana Intelligence Gallery, which is an assortment of ML platform which is reused and investigated by information researchers. Azure services can be ordered into two:

1. Azure Bot Service

2. Azure Machine Learning Studio.

Azure ML studio requires users to complete all the operations manually. This includes, data preprocessing, exploration, validating modeling results, and choosing methods. It supports about 100 techniques that address regression, anomaly detection, classification (binary and multiclass), text analysis, and recommendation.

Azure ML studio expects user to finish every one of the activities by manual method. This incorporates, information preprocessing, investigation, approving displaying results, and picking strategies. It upholds around more than hundreds of strategies that address anomaly detection, regression, order, text analysis and prediction.

C. Sentiment Analysis Using Python:

During the study Siddharth et al. [14] worked upon the data collected from platforms Twitter and Youtube to investigate individuals' perspectives utilizing ML tools and furthermore utilizing the python programming language for the coding part. During the study of Turney [15] the sack of-words approach was utilized for investigating the opinions.

Liu [7] focus on analyzing the sentiments and subjectivity of tweets. Bouazizi et al. [10] collected the Twitter dataset from social media and analyzing the sentiments as positive, negative, and neutral.

The work of Liu [7] centered on breaking down the opinions and subjectivity of tweets.

Bouazizi et al. [16] gathered the Twitter dataset and investigated the opinions in the form to be positive, negative, or neutral.

However, sentiment analysis is a highly restricted NLP problem because NLP can only understand some aspects of it, such as positive or negative emotion. [7] NLTK is the natural language processing toolkit for Python, one of the most commonly used Python libraries in the NLP world. This paper will use NLTK to do the research.

But, the limitation of sentiment analysis is that is an exceptionally confined NLP issue since NLP can just see a few parts of it if some sentence is positive or negative. [7] The sentiment analyzer used in this study is the basically language preparing tool designing using "TextBlob" library of Python perhaps the most regularly utilized Python library in the NLP world. This paper will utilize both Azure Machine Learning Add-in from Microsoft and Python for sentiment analysis.

III. METHODOLOGY

A. Data Collection

For the following study customer reviews of Xbox on Amazon has been used. So the data has been copied in to MS office Excel from Amazon itself. Before diving into data analysis, it is necessary to take a look at the dataset and clean data if necessary. For sentimental analysis, it needs more steps to prepare the text for later on analysis.

B. Review Length

Most of the reviews length are under 500 words. The mean length of the reviews is about 128.

It says the people who gave rating 1 will describe the facts and the feelings together. And the world could be just a big

picture of the words, we need to do more work to find more valuable things.

C. Methodology:

For doing sentiment analysis using Microsoft Azure Machine Learning following steps are needed:

1. First there must be a genuine copy of MS office 13 or above is installed in system.
2. Then after opening MS Excel install Azure Machine Learning Add-ins in Excel.
3. Then Install Web Service named Text Sentiment Analysis Excel Add-in sample.
4. After that the heading of column must be changed to “tweet_text” as it is Input in schema.
5. After that Input and Output area in sheet is defined.
6. After that we just have to click Predict.

For doing sentiment analysis using python:

1. First task is to import a python library “textblob”.
2. Then next through a simple 10 line code a sentiment analyzer is designed that helps in understanding polarity of sentence.
3. After that one by one reviews are entered on it and results again mentioned in excel sheet in which result of polarity of every sentence measured using sentiment analyzer is entered in the sheet.

IV. RESULT AND ANALYSIS

A. Sentiment Analysis Using Microsoft Azure Machine Learning:

To support work screenshots are mentioned below:

1. After importing reviews of Xbox available on Amazon in Excel shown in figure 1.

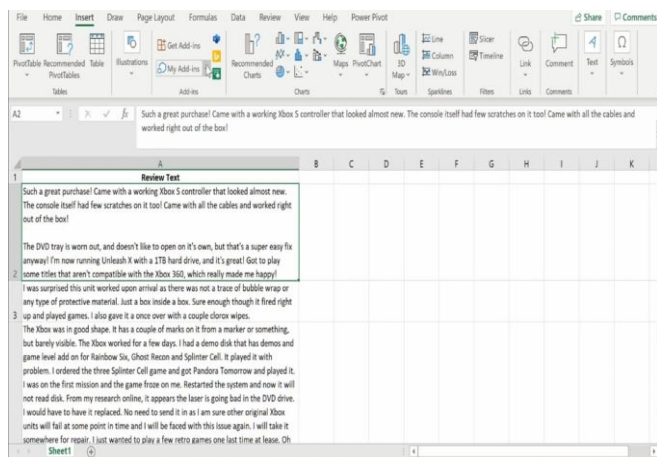


Figure 1.

2. Then, the heading of column is needed to be changed from Review text to “tweet_text” because the schema given in it takes that string in only “tweet_text” this form presented in figure 2.

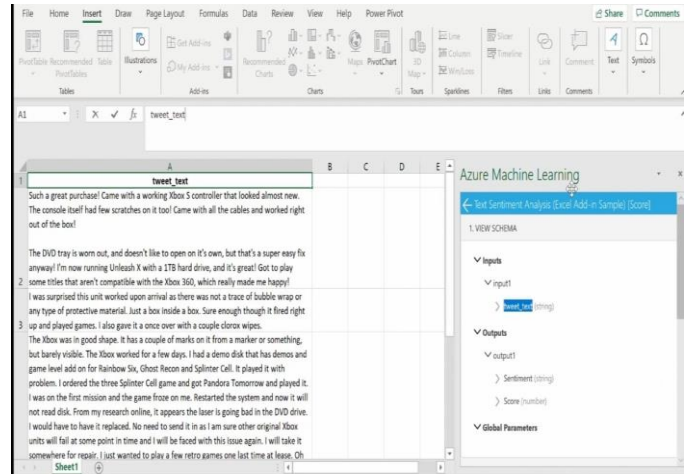


Figure 2.

3. Next, the section is needed to be selected as input and output area for presenting data after the prediction is done presented in figure 3.

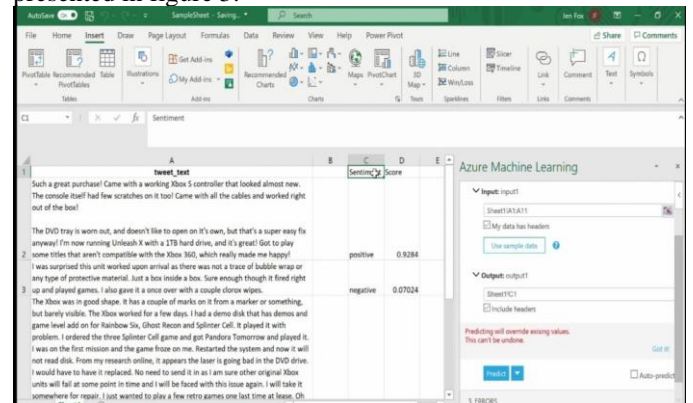


Figure 3.

B. Sentiment Analysis using Python:

To support the work screenshots of how the task was performed is mentioned below:

1. To perform the task first pip library of python and then text TextBlob is needed to install using command prompt terminal of windows system screenshot is mentioned in Figure 4 and 5.

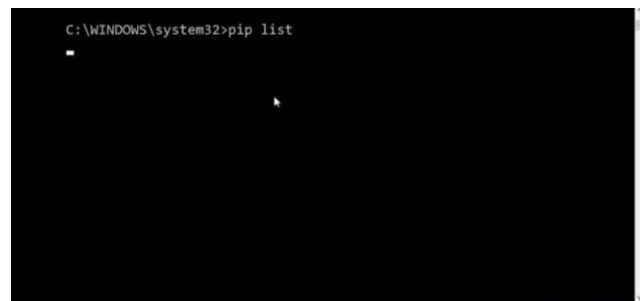


Figure 4.

```

PySocks      1.6.8
python-dateutil  2.8.0
requests     2.21.0
requests-oauthlib 1.2.0
scipy        1.2.1
setuptools    40.6.2
singledispatch 3.4.0.3
six          1.12.0
tweepy       3.7.0
urllib3      1.24.1

C:\WINDOWS\system32>pip install textblob
Collecting textblob
    
```

Figure 5.

- Next Using VS code coding part is performed in which (y) is termed as input for the sentiments and (x) is basically output that will define whether a review is positive, negative or neutral code part is Figure 6.

```

SentimentAnalysis.py 2 X
SentimentAnalysis.py > ...
1 from textblob import TextBlob
2 y=input("Type your sentence:")
3 edu=TextBlob(y)
4 x=edu.sentiment.polarity
5 # Negative = x<0 and Neutral = 0 and positive x>0 && x<=1
6 if x<0:
7     print("Negative")
8 elif x==0:
9     print("neutral")
10 elif x>0 and x<=1:
11     print("Positive")
    
```

Figure 6.

- Next after running the code how the program responded and outcomes are mentioned in MS Excel one by one after getting results mentioned in Figure 7, 8.

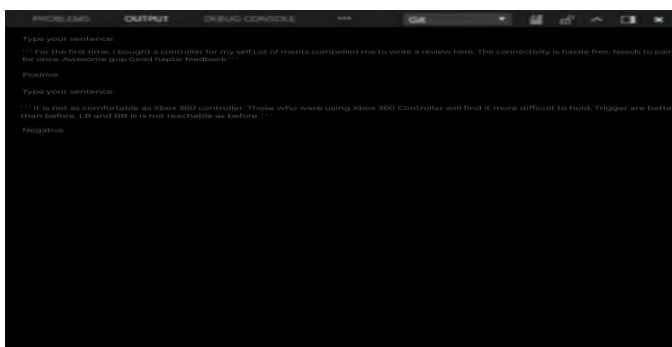


Figure 7.

tweet_text	Sentiment
Such a great purchase! Came with a working Xbox S controller that looked almost new. The console itself had few scratches on it too! Came with all the cables and worked right out of the box!	positive
The DVD tray is worn out, and doesn't like to open on its own, but that's a super easy fix anyway! I'm now running Unleash X with a 1TB hard drive, and it's great! Got to play some titles that aren't compatible with the Xbox 360, which really made me happy! I was surprised this unit worked upon arrival as there was not a trace of bubble wrap or any type of protective material. Just a box inside a box. Sure enough though it fired right up and played games. I also gave it a once over with a couple clorex wipes.	negative
The Xbox was in good shape. It has a couple of marks on it from a marker or something, but barely visible. The Xbox worked for a few days. I had a demo disk that has demos and game level add on for Rainbow Six, Ghost Recon and Splinter Cell. It played it with problem. I ordered the three Splinter Cell game and got Pandora Tomorrow and played it.	negative

Figure 8.

V. CONCLUSION

Sentiment analysis is very important task for organizations and businesses. As the organizations frequently need to know in time what customers and the public think about their product and services that are currently in market. However, it isn't reasonable to physically peruse each post on the site and concentrate helpful perspective data from it. In the event it is performed physically, there is an excessive amount of information so therefore sentiment analysis permits huge scope handling of information in a proficient and practical way. To find out about sentiment investigation, the work during this research work to provide easy methods for sentiment analysis on businesses to comprehend its qualities and limits. This paper has done a sentimental analysis over Xbox reviews that were taken from Amazon, and built a model to predict the sentiment of the comment given the review text. The sentiment analysis is certifiably not a simple errand in any event, for people, which implies sentiment analysis performed may not be pretty much as exact as we thought. All things considered, Sentiment analysis can naturally and straightaway perceive and investigate texts, it is helpful for checking popular assessment different preferences in a dramatically developing age.

VI. REFERENCES

- Wiebe Janyce "Identifying subjective characters in narrative, Proceedings of the International Conference on Computational Linguistics (COLING-1990)".
- Hearst M., 1992, Direction-based text interpretation as an information access refinement in TextBased Intelligent Systems, P. Jacobs, Editor 1992, Lawrence Erlbaum Associates.
- V. Suresh 2011, A Non-syntactic Approach for Text Sentiment Classification with Stopwords, 2011.
- Murthy G. and Bing Liu, 2008, Mining opinions in comparative sentences, Proceedings of the 22nd international conference on computational linguistics 2008.
- Dave, Lawrence & Pennock, "Opinion extraction and semantic classification of product reviews", 2003.
- Matsumoto, Takamura and Okumura "Advances in knowledge discovery and data mining", 2005.
- Liu, Chen "Multi label classification based approach for sentiment classification", 2015.
- Harvinder Jeet Kaur and Rajiv Kumar, Sentiment Analysis from Social Media in Crisis Situations, International Conference on Computing, Communication and Automation, 2015.
- Xiaoming Gao, Emilio Ferrara, and Judy Qiu, Parallel Clustering of High-Dimensional Social Media Data Streams, 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing
- J. Prabhu and M. Sudharshan and M. Saravanan and G. Prasad, Augmenting Rapid Clustering Method for Social Network Analysis, International Conference on Advances in Social Networks Analysis and Mining, 2010.
- Xin Chen, Krishna Madhavan and Mihaela Vorvoreanu, A Web-Based Tool for Collaborative Social Media Data Analysis, Third International Conference on Cloud and Green Computing, 2013.
- Qasem, M., Thulasiram, R., & Thulasiram, P.. Twitter sentiment classification using machine learning techniques for stock markets. Advances in Computing, Communications and Informatics (ICACCI), 2015.
- Ericson, G., Franks, L., & McKay, P. What is Azure Machine Learning Studio?, 2016.
- Siddharth, Darsini, and Sujithra, "Sentiment Analysis on Youtube & Twitter Data using Machine Learning," Int. J. Res. Appl. Sci. Eng. Technol., 2020.
- P. D. Turney, "Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Annu. Meet. Assoc. Comput. Linguist., 2002.
- M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", 2017.