

SENTIMENTAL ANALYSIS OF CUSTOMER PRODUCT REVIEWS USING MACHINE LEARNING

A. DURGA DEVI, KURRA.DEVA SRI DIVYA

¹ Assistant Professor MCA DEPT, Dantuluri Narayana Raju College ,Bhimavaram, AndharapradeshEmail id: - adurgadevi760@gmail.com² PG Student, , D.N.R. COLLEGE, P.G. COURSES (AUTONOMOUS), BHIMAVARAM-534202.Email id: - divyadevasri123@gmail.com**ABSTRACT**

Today, digital reviews play a pivotal role in enhancing global communications among consumers and influencing consumer buying patterns. E-commerce giants like Amazon, Flipkart, etc. provide a platform to consumers to share their experience and provide real insights about the performance of the product to future buyers. In order to extract valuable insights from a large set of reviews, classification of reviews into positive and negative sentiment is required. Sentiment Analysis is a computational study to extract subjective information from the text. In the proposed work, over 4,000,00 reviews have been classified into positive and negative sentiments using Sentiment Analysis. Out of the various classification models, Naïve Bayes, Support Vector Machine (SVM) and Decision Tree have been employed for classification of reviews. The evaluation of models is done using 10-Fold Cross Validation.

1. INTRODUCTION

With an ever-increasing demand of smart phones, the mobile phone market is expanding at an exponential pace. With such a boom in the smart-phone industry, there is a need to realize the holistic review of the brand and the model of phone. There are numerous brands present in the market, out of which some are dominant and occupy quite a big part of the industry. For instance, Samsung, Apple, etc. are names associated with brands which are famous throughout the world. Electronic commerce plays a vital role in increasing the sales of the mobile phones and influencing consumer buying patterns. Reviews available on such e-commerce platforms act as a guiding tool for the consumers to make informed decisions. Retail websites like Amazon.com offer different options to the reviewers for writing their reviews. For instance, the consumer can provide numerical rating from 1 to 5 or write comments about the product.

As there are innumerable products manufactured by many different brands, so providing relevant reviews to the consumers is the need of hour. Number of reviews associated with a product or a brand is increasing at an alarming rate, which is no less than handling the big data. Classifying the reviews on the basis of sentiment of customers into positive and negative sentiment provides sentiment orientation of the review, hence results in better judgement. Segregation of reviews on the basis of their sentiment can help future buyers to evaluate positive and negative feedback constructively and reach at better decisions as per their requirements. This evaluation acts as a testimony to the users who are looking to know the details and specifications of the smartphones; thereby increasing user credibility

In this research, unstructured data of Mobile Phone Reviews have been extracted from Amazon.com. It has been filtered to remove noisy data and has been pre-processed to evaluate sentiment of the reviews using supervised learning. The reviews have been classified using machine learning classification models like Naïve Bayes, Support Vector Machine (SVM) and Decision Tree and have been cross validated to find the best classifier for this purpose.

2. LITERATURE SURVEY AND RELATED WORK

1.TITLE: Big data consumer analytics and the transformation of marketing.

AUTHOR: S. Erevelles, N. Fukawa, and L. Swayne

The data explosion over the last couple of years combined with technological advances has opened up a new way of analysing and gaining information from the data. The evolution of Big Data has led to the rise of Big Data Consumer Analytics wherein the large resources of data are used to extract information and then exploited to create value. This has transformed the way marketing is done and the interaction between the firm and its customers. In this research, we explore the properties of Big Data and how they can influence marketing campaigns by providing the right kind of information a marketer is looking for, that will help the firm to move ahead of its competitors. We shall also explore how the traditional knowledge-based view has limitations when implemented with Big Data and why it should be combined with an ignorance-based view to make the most of the data. The importance and impact of using Big Data Consumer Analytics in the marketing strategies and how certain firms have created advantages for themselves with the help of the data has been discussed. Some of the challenges and limitations have been mentioned in the end.

2.TITLE: Predicting the semantic orientation of adjectives

AUTHOR: V. Hatzivassiloglou and K. R. McKeown,

We identify and validate from a large corpus constraint from conjunctions on the positive or negative semantic orientation of the conjoined adjectives. A log-linear regression model uses these constraints to predict whether conjoined adjectives are of same or different orientations, achieving 82% accuracy in this task when each adjective, a clustering algorithm separates the adjectives into groups of different orientations, and finally, adjectives are labelled positive or negative. Evaluations on real data and simulation experiments indicate high levels of performance: classification precision is more than 90% for adjectives that occur in a modest number of conjunctions in the corpus.

3.TITLE: Recognizing contextual polarity in phrase-level sentiment analysis

AUTHOR: T. Wilson, J. Wiebe, and P. Hoffmann,

This paper presents a new approach to phrase-level sentiment analysis that first determines whether an expression is neutral or polar and then disambiguates the polarity of the polar expressions. With this approach, the system is able to automatically identify the contextual polarity for a large subset of sentiment expressions, achieving results that are significantly better than baseline.

3. EXISTING SYSTEM

Number of reviews associated with a product or a brand is increasing at an alarming rate, which is no less than handling the big data. Classifying the reviews on the basis of sentiment of customers into positive and negative sentiment provides sentiment orientation of the review, hence results in better judgement.

4. PROPOSED SYSTEM

Sentiment analysis is not only confined to the English language but has been implemented for various languages. Sentiment analysis of Chinese text by implementing four feature selection methods and five classifiers viz. Centroid classifier, K-nearest neighbour, Window classifier, Naïve Bayes and SVM has been done [11]. Through this learning paradigm it was concluded that SVM outperforms all the other learning methods in terms of sentiment classification. Sentiment analysis on travel reviews using three machine learning models namely, Naïve Bayes, SVM and character-based N-gram model has been performed in which SVM and N-gram approaches have better performance than Naïve Bayes [12]. It has been observed that in maximum number of cases SVM showcases best performance in comparison to other classification models.

5. METHODOLOGIES

MODULE

Tensor flow:

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of

databases.

Pandas :-

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyse. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib :-

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

Scikit – learn :-

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

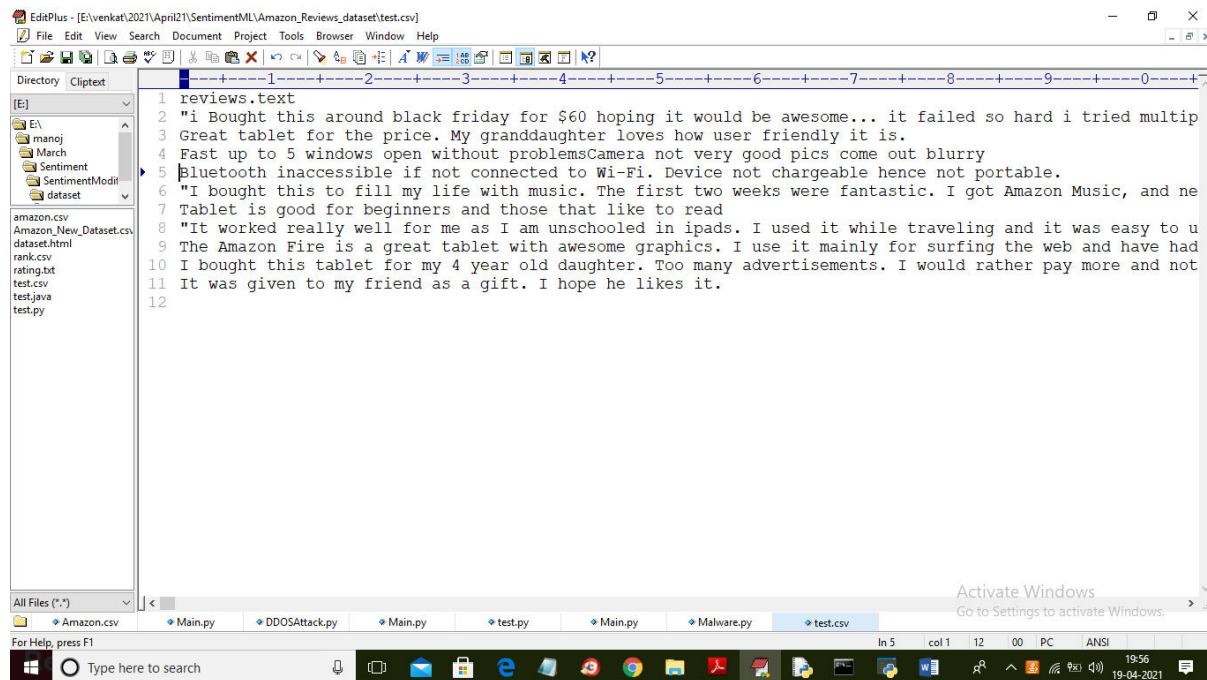
Python is Interpreted – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive – you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak

behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

6. RESULTS AND DISCUSSION SCREEN SHOTS



```

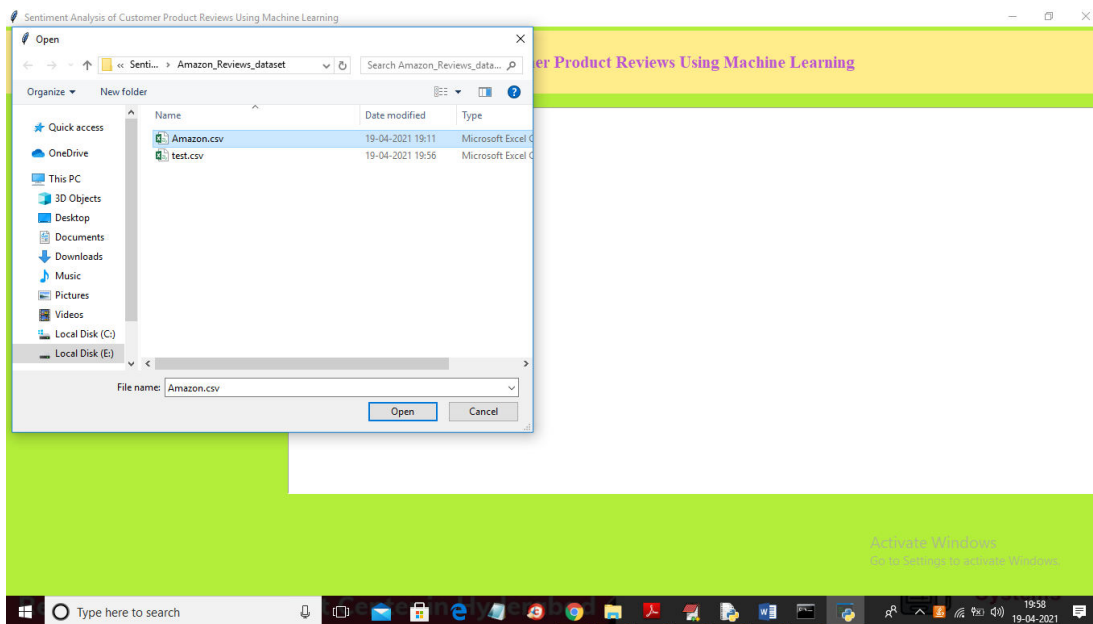
1 reviews.text
2 "i Bought this around black friday for $60 hoping it would be awesome... it failed so hard i tried multip
3 Great tablet for the price. My granddaughter loves how user friendly it is.
4 Fast up to 5 windows open without problemsCamera not very good pics come out blurry
5 Bluetooth inaccessible if not connected to Wi-Fi. Device not chargeable hence not portable.
6 "I bought this to fill my life with music. The first two weeks were fantastic. I got Amazon Music, and ne
7 Tablet is good for beginners and those that like to read
8 "It worked really well for me as I am unschooled in ipads. I used it while traveling and it was easy to u
9 The Amazon Fire is a great tablet with awesome graphics. I use it mainly for surfing the web and have had
10 I bought this tablet for my 4 year old daughter. Too many advertisements. I would rather pay more and not
11 It was given to my friend as a gift. I hope he likes it.
12

```

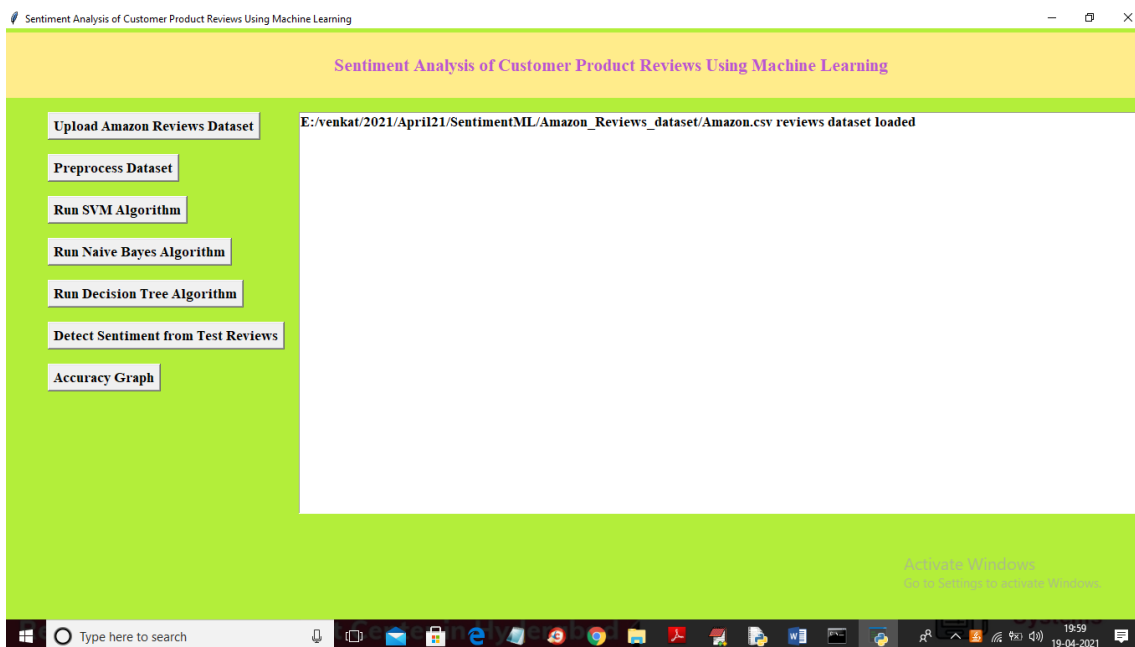
In above test data we have only test reviews and by applying ML trained model on above test data we can predict sentiment label.

To run project double click on 'run.bat' file to get below screen

In above screen click on 'Upload Amazon Reviews Dataset' button to upload dataset



In above screen we are selecting and uploading 'Amazon.csv' file and then click on 'Open' button to load dataset and to get below screen



In above screen dataset loaded and now click on 'Preprocess Dataset' button to read all reviews from dataset and then apply Preprocess steps to get below screen

```

C:\Windows\system32\cmd.exe
526 == 5 year old mom loves something waiting doctors offices loves read crosswords etc
527 == 5 absolutely love kindle great picture storage sound cant live without kindle
528 == 5 love mine works great prime best movies
529 == 5 got two moms love many great features
530 == 5 tablet awesome fact use micro sd card memory listen content card great additional screen size wanted disappointed great purchase
531 == 5 im happy kindle fire beyond able read books sorts things able download huge variety apps also holds charge long size perfect portable yet screen much larger sm
arphone love would definitely recommend tablet others
532 == 5 amazon fire hd awesome device price tablet mind boggling tablet everything would expect want tablet amazon store amazon prime preloaded much fun easy use perfo
rmance tablet unbelievable never stop thinking modest price paid fantastic tablet feels comfortable hands design nice size adequate tablet use expand memory storage dev
ice additional believe outstanding
533 == 5 bought christmas present son uses watch youtube netflix works great
534 == 5 got fire tablet youngest daughter christmas loves also love option restrict apps content age
535 == 5 bought two amazon fire hd tablets december christmas gifts recipients happy gifts plan purchase one near future
536 == 5 biting gift concern size amazon interface easy enough download google play store familiar user size initial concern right watch movies without cumbersome
537 == 5 son really likes larger screen play games
538 == 5 easy use much offer low price
539 == 5 great product easy use love lighted pad ease loading books
540 == 5 using christmas morning didnt take long getting best tablets
541 == 5 love amazon nice read book play games
542 == 5 bought kindle gift mother could access facebook account easily loves ease using kindle tablets home used daily mother uses email light gaming watching netflix
well satisfied
543 == 5 wanted low cost internet device feed internet radio choices stereo fit bill well use headphone jack connect stereo get great sound dont use apps come device
544 == 5 kindle fire wonderful small device use multitude tasks use mostly ereader simple use happy chose one lightweight convenient take
545 == 5 bought temporary replacemente wives kindle fire hd service liked model much continued using one returned
546 == 5 easy download read enjoy reading take anywhere without heaviness book
547 == 5 bought subscribing amazon prime membership great little tablet low price good deal pass sale ability talk alexa surf web listen music read booksmagazines lots
free music books prime members
548 == 5 second fire im upgrading generation model love screen resolution movies videos sharp pixelating camera bonus first version together great buy sale thank best b
uy
549 == 5 tablet amazon offers tons free apps hundreds thousands free ebooks tons free games tablet intuitive intelligent seamlessly enjoy streaming favorite shows readi
ng articles books periodicals love tablet doesnt place apps device uses users miro sd card instead additional storage dizzying amount storage games apps books movies do
unloaded enjoyed also allows owner share device family members delegating member account personalized tailored individual even option create childrens accountextra bon
us able minron content tablet onto samsung smart television definitely consider amazon first foremost making future tablet purchases maintains charge good deal time all
eviating need near outlet enjoy using charge evening use several hours next day sometimes two days one even create tablet considerate ownerenduser
550 == 5 love read tablet makes easy carry entire library everywhere also become addicted free underground games
551 == 5 bought model use addition current year old kindle fire hd inch model far pleased quality unit slightly larger screen makes watching movies much better didnt wa
nt larger screen yet
552 == 5 purchased black friday could pass price tablet works great basic internet needs great shopping amazon purchase another android tablet look something little pow
er since tablet slows multitasking
553 == 5 bought son hes years old could play basic toddler games found tablet extremely useful school large screen super long battery life in waiting surface pro tablet
got cheap making second guess paying much basic lecture taking notes watching netflix hulu kid games cant beat
554 == 5 best tablet price use everyday watch youtube videos
555 == 5 great purchase granddaughter christmas
556 == 5 tablet perfect anything could want tablet

```

In above black console we can see application read all reviews from dataset and then generate below TF-IDF vector

Sentiment Analysis of Customer Product Reviews Using Machine Learning

Sentiment Analysis of Customer Product Reviews Using Machine Learning

Upload Amazon Reviews Dataset

Preprocess Dataset

Run SVM Algorithm

Run Naive Bayes Algorithm

Run Decision Tree Algorithm

Detect Sentiment from Test Reviews

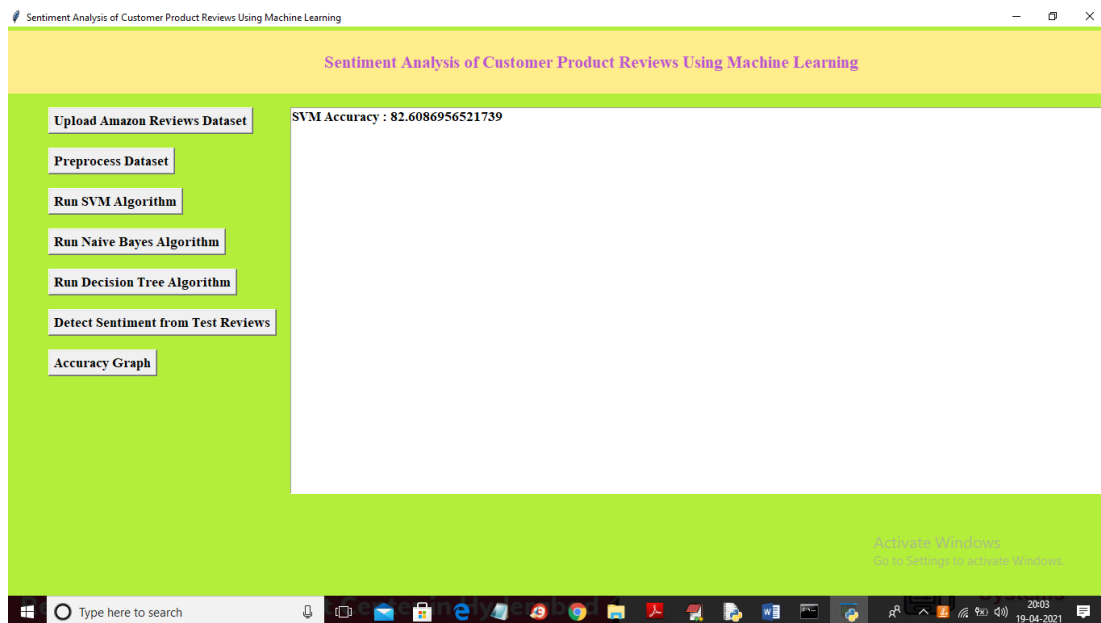
Accuracy Graph

	ability	able	absolutely	ac	accent	accept	acceptable	...	young	younger	youngest	youre	youtube	yr	yrs
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
568	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
569	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
570	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
571	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
572	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

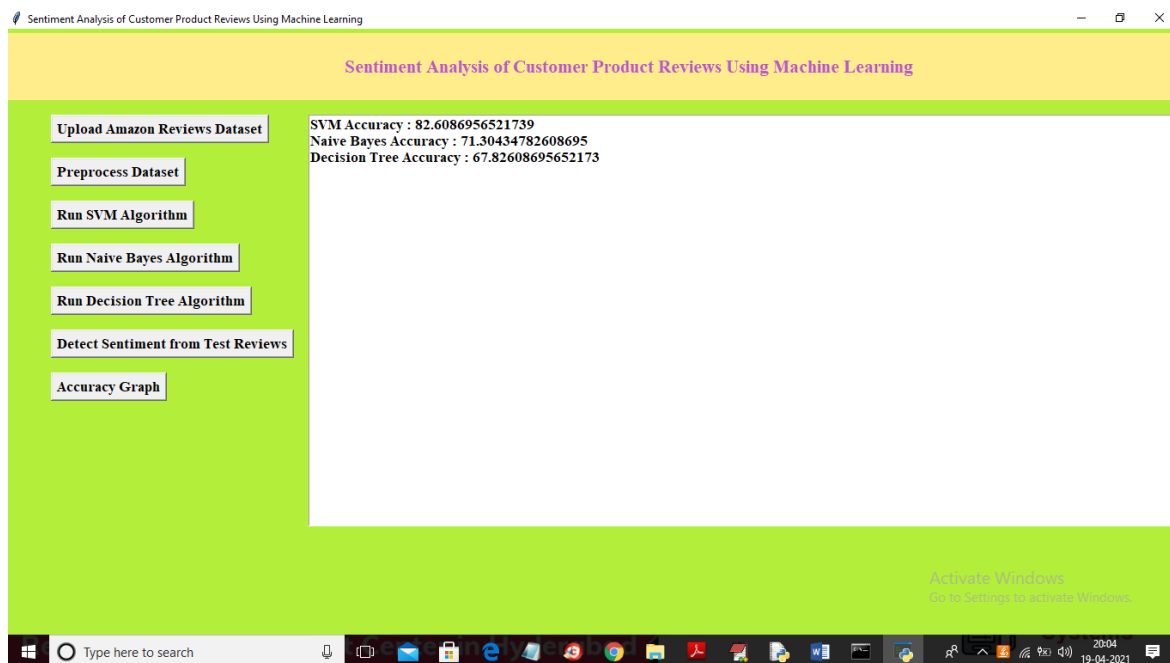
[573 rows x 2361 columns]Total reviews found in dataset : 573
Total words found in dataset : 2361

In above screen in text area, we can see application extract all words from reviews and then put in top line of above test area and in remaining rows if that word appears then it put average count value of that word and if word not appear then 0 will put. In above screen vector generated and I am showing few records from that vector. In that vector total reviews are 573 and all reviews contains total 2361 unique words. Now vector is ready and now click on 'Run SVM Algorithm' button to train SVM with above

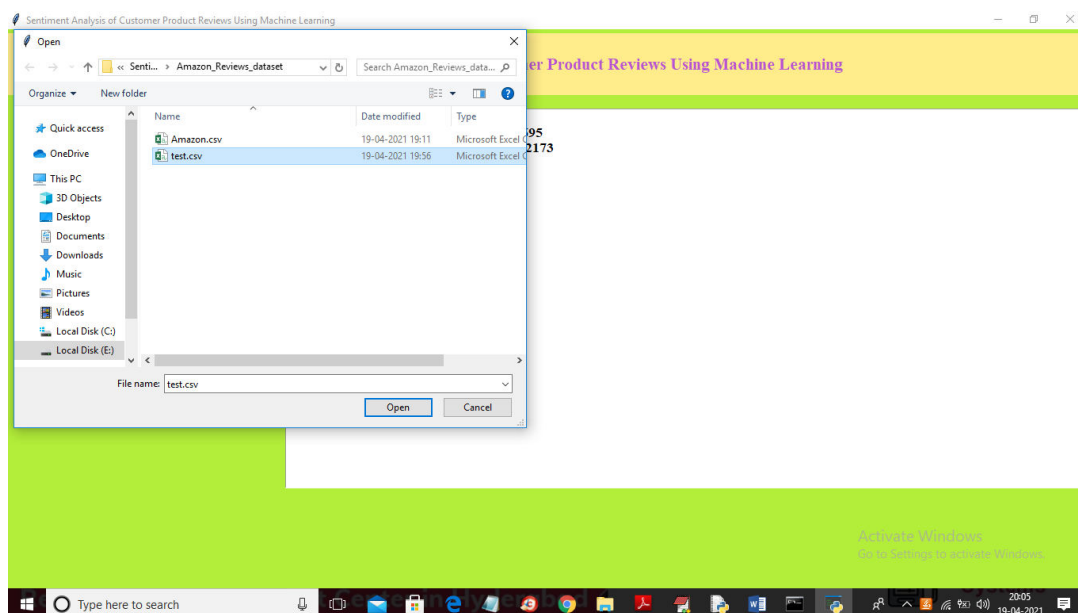
vector



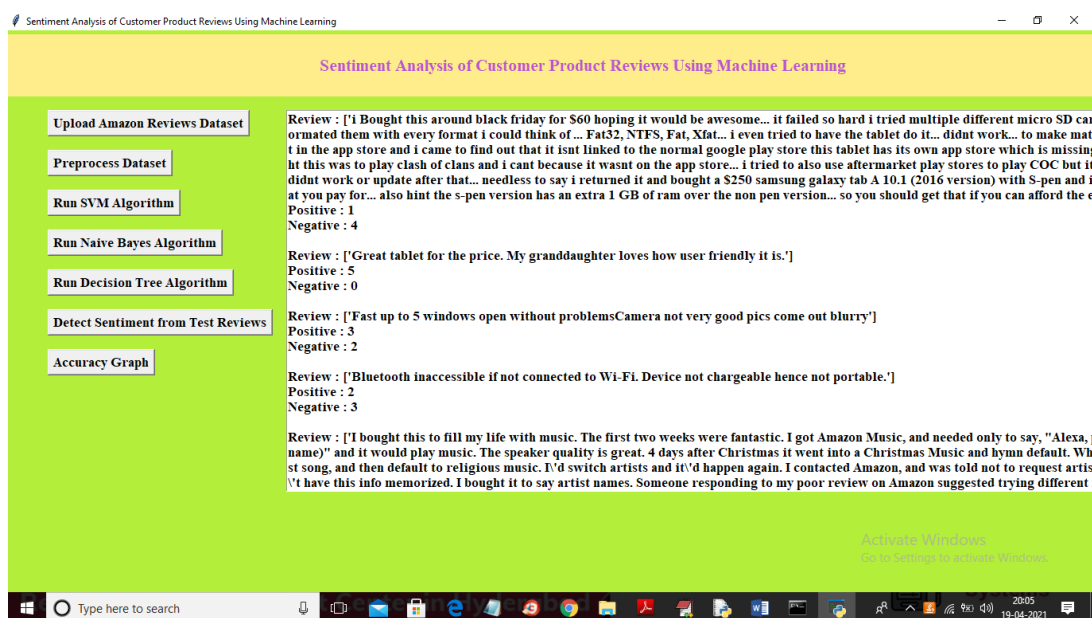
In above screen with SVM we got 82% accuracy and now click on Naïve Bayes and Decision tree button to get their accuracy



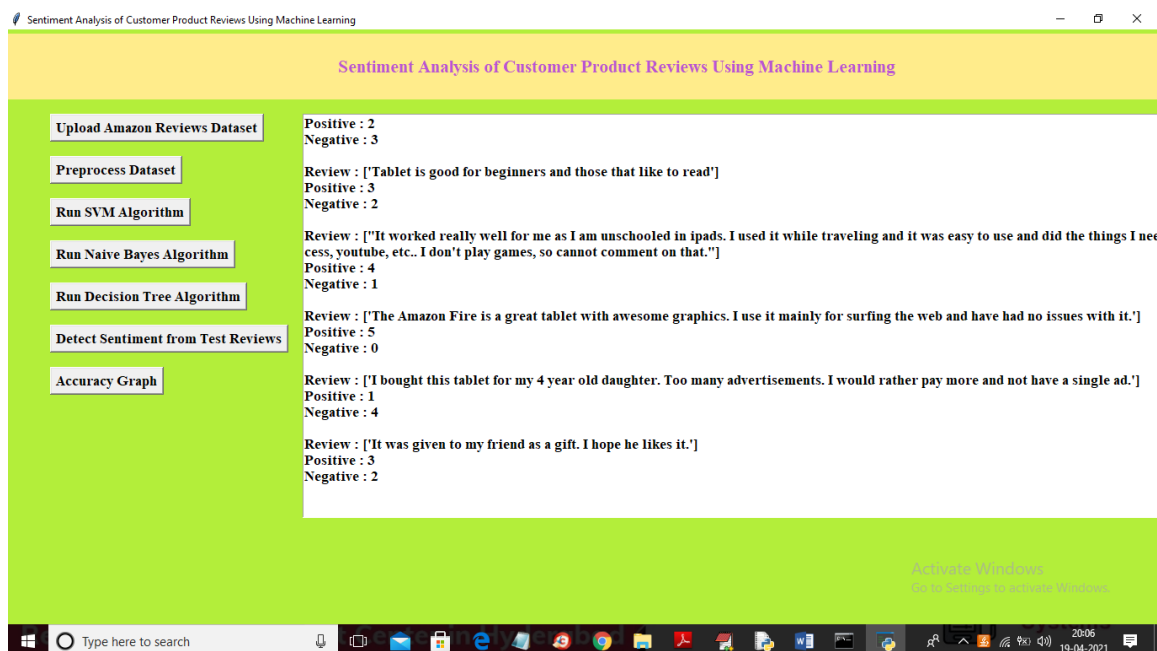
In above screen with all 3 algorithms SVM gave better prediction accuracy and now click on 'Detect Sentiment from Test Reviews' button to upload test reviews



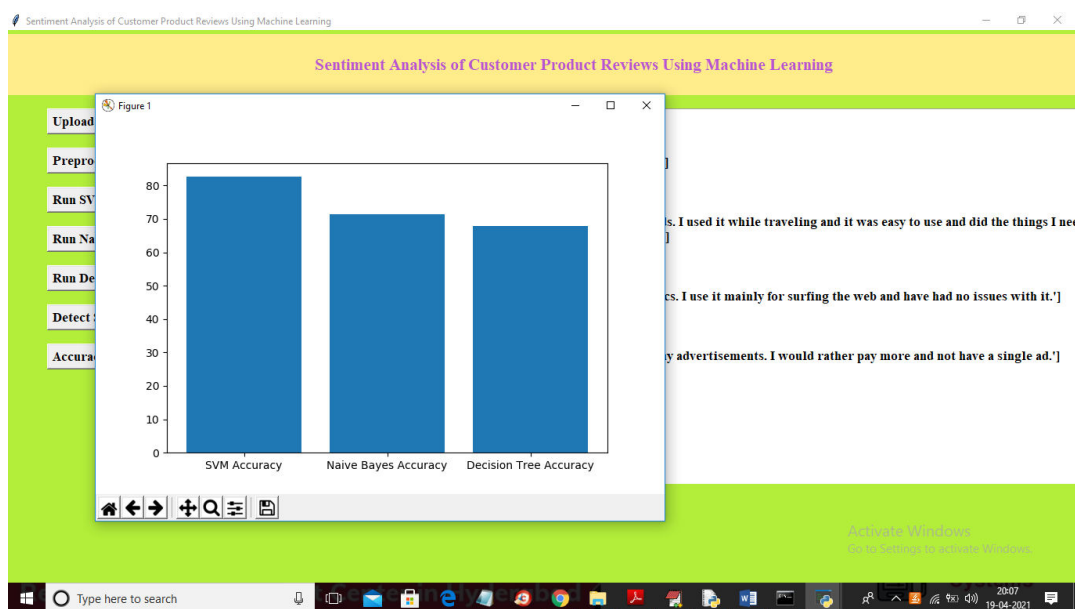
In above screen selecting and uploading 'test.csv' file and then click on 'Open' button to get below prediction result



In above screen first I am displaying reviews from uploaded test file and then predicting positive and negative sentiment for each review and you can scroll down above text area to get all outputs



In above screen we can see sentiment prediction result for all reviews and now click on 'Accuracy Graph' button to get below graph



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and in all 3 algorithms SVM got higher accuracy

7. CONCLUSION AND FUTURE SCOPE

CONCLUSION

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. Our research is aiming to achieve this by conducting sentiment analysis of mobile phone reviews and classifying the reviews into positive and negative sentiment. After balancing the data with almost equal ratio of positive and negative reviews, three classification models have been used to classify reviews. Out of the three classifiers, i.e., Naïve Bayes, SVM and Decision Tree, predictive accuracy of SVM is found to be the best. The accuracy results have been cross validated and the highest value of accuracy achieved was 81.75% for SVM among the three models.

FUTURE SCOPE

In future, the work can be extended to perform multiclass classification of reviews which will provide delineated nature of review to the consumer, hence better judgement of the product. It can also be used to predict rating of a product from the review. This will provide users with reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

8. REFERENCES

- [1] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," *Journal of Business Research*, vol. 69, no. 2, pp. 897–904, 2016.
- [2] P. Russom et al., "Big data analytics," TDWI best practices report, fourth quarter, pp. 1–35, 2011.
- [3] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," *Journal of Business Research*, vol. 69, no. 2, pp. 897–904, 2016.
- [4] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997, pp. 174–181.
- [5] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, pp. 347–354.
- [6] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004, p. 271.
- [7] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *LREc*, vol. 10, no. 2010, 2010.
- [8] M. WAHYUDI and D. A. KRISTIYANTI, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." *Journal of Theoretical & Applied Information Technology*, vol. 91, no. 1
- [9] D. N. Devi, C. K. Kumar, and S. Prasad, "A feature-based approach for sentiment analysis by using support vector machine,"

in Advanced Computing (IACC), 2016 IEEE 6th International Conference on. IEEE, 2016, pp. 3–8.

[10] V. Narayanan, I. Arora, and A. Bhatia, “Fast and accurate sentiment classification using an enhanced naive bayes model,” in International Conference on Intelligent Data Engineering and Automated Learning. Springer, 2013, pp. 194–201.

.