

## Analysis of product usage based on Twitter data

P.Pushpa Rani ,M.Tech<sup>1</sup> , G.Rachana<sup>2</sup> ,SK.Sushma<sup>3</sup> ,V.Sravani<sup>4</sup>,G. Ramanjaneyulu<sup>5</sup>,  
V.Anantha Venkata Sai Kumar<sup>6</sup>

<sup>1</sup> Asst. Professor, Department of CSE, Qis College Of Engineering and Technology, Ongole  
<sup>2,3,4,5,6</sup> Asst. Professor, Department of CSE, Qis College Of Engineering and Technology, Ongole

**Abstract-** This study further improves chronological approaches and procedures in order to conduct modification and analysis of the corresponding forecast. We have a real-time dataset based on the comments sections of the twitter user. The singularity of the dataset is that only certain commentary have been retrieved which synthesises a specific product based term. Then repeated extraction is done so that the dataset is completed. Our dataset has three ideological columns that improve the details on the Product that we see with the individual user or our main subject. In this exact dataset we examined the use of the goods by the consumers produced by the Google and Apple firms. Both technological titans in this age have knowledge of their technology and will be more focused on future cyber-projects and goods. They are interested in further optimization and improvement of their gadgets. Without input, advances and reverses of their prior initiatives, their project would be a difficult endeavor. You may extract this from the twitter api dataset and improve your product further. They may assess the disadvantage by finding out if it has reached the market and succeeded in obtaining all kinds of information from social media rather than doing a poll. Such a procedure would be more chaotic and no reliable results could be predicted. We thus continue with the process of social media data mining.

**Keywords**—dataset, data mining, api, analyze.

### I. INTRODUCTION

The key criteria for a person to categorise the unique user's characteristics. Social media has become one of the major issues for information and news transmission. The major aim of Internet development is to convey important information from one side of the site to another.

Data transmission improves information and reaches individuals in a distinctive way. But marketing and corporations utilise the data and provide their advertising goods at their destination. And utilized for the users' dataset. The first thing is to extract the dataset of the individual user or extract many data sets from the huge users. Then for a constant interval, the data set should be extracted. Initially, the framework would be unstructured information. Since unstructured data are kept in a large database, they need be analysed and transformed into structured data. It includes several classes and characteristics.

One of the essential phantoms to consider in the data extraction process is classes and characteristics. The data may be broken into the chronological sequence after the structured data have been received. Using certain machine learning techniques and statistical analysis, the structuring data gathered are manipulated in the end with regard to login, logging, posting photographs, or tagging various characteristics, the evaluation of the individual person may be achieved. [1] "The Opinion Mining and Sentiment Analysis research challenge provides a conceptual insight into privacy as determined. We have just recently derived from this article some theories of social media trafficking and the psychology of users[2]. We did not adopt the method the same, but developed a real-time model, providing just theoretical analysis[3-5]. We have taken theoretical principles and also give background information on how users' privacy is used as data. [8] Paper focuses on Facebook handling concepts and how users have access to them. The experience of its consumers is obvious.

## II. RELATEDWORKS

The main aim of this specific study is to forecast the users' purpose for the product being monitored. Different machine learning methods may be employed to conduct the handling but in order to perform that work we approach it in a very specific way. Thus, for the regression analysis of our product, just one categorization approach is employed. The outcome of our work might be to forecast the product's use. However, by dissolving comments and closed observations, we may conduct different manipulations such as enhanced gadget specifications, product market reach, camera performance or product performance. Each of the comments is examined by the specific keywords. Each score is provided whether it is good or negative. This procedure is performed repeatedly and we have built our particular data collection. In the module developed using python packages, the data set is currently available as unstructured data. The result must only be examined and focused on a specified regression analysis.

## III. PROPOSED SYSTEM

The problem was divided into two divisions, whether a good or a negative emotion. Thus, this dataset is classified as the approach and analysis of machine learning. In classification techniques there are few techniques.

### A. Techniques involved for Manipulation Process

We participate solely according to two criteria, either good or negative. Therefore, the complexity of the Decision Tree method reduces. This was one of the greatest approaches that many data scientist firms have still utilised to forecast the product result. Banks are using this analytical approach to ensure that their customers are either capable or in financial crisis and that their companies are at risk of not being able to pay their leases. The high-end banks were using this approach to forecast their companies and clients' happiness using the dataset. For example, even an ordinary individual can use this approach to purchase his essential necessities, such home. In the single product he can't invest very much in that fast way. Small calculations and analyses required to take place, for example the land information, its selling rate, water facilities, transit facilities, whether they rural or urban, either close to shopping. These are the characteristics in the tree of decision. If either yes or no response occurs one by one at a certain point these questions cannot be dissolved. That's why we have a result.

**Decision tree:** We must be extremely clear about numerous words involved in the decision tree process before we engage in this specific issue.

**Entropy:** They are the fundamental complete thing with which we will participate. In our scenario we will employ a twitter data set which will either provide the users with misinformation or positive intention for the product. The complete system with which we participate is defined. The categorization procedure is followed by this entropy.

**Information gain:** This part offers the data gathered using the dataset. The comment area of our relevant dataset in particular. The primary information of the module is included here.

**Leaf node:** The entropy is broken into subsequent classes and the entropy cannot be broken further by the end. This case is stated as the Leaf node for the final scenario.

**Root Node:** The root node shows the gain of information that has been fragmented to define how many people have been kept away from the negative emotion and good emotion.

**Model fine tuning:** More than enough information has been retrieved from our csv file for data analysis. However, our data to input into our model should be further simplified. Therefore we disregarded the comment section column because either a good remark or negative comment on the specific product is our class variable. We have carried out about eight analytical products. The commentary detailing the product is removed and then analysed and considered to be either good or negative. We have dissolved the module to further simplify the floating variable parts. All products have been specified with a float number and provided to the model. They are then done in order to enable prediction status by supplying the relevant python code.

**Preprocessing:** the remark area is deleted because the csv file contains our sustained product information. Moreover, zero commentaries such as Cant state that there were no optimal feelings for the product since they include undesired product information. The entire csv file contains around 3000 data, but is further optimised to 450-600 data, which is included in the pre-processed module.

### Algorithm

```
x = np.array(train.drop(["Users intention about the product"], 1))
y = np.array(train["Users intention about the product"])
x_train, x_test, y_train, y_test = train_test_split(
x, y, test_size = 0.9, random_state = 100)
model=model.fit(x_train,y_train)
y_pred = model.predict(x_test)
print("accuracy:",accuracy_score(y_test,y_pred)*100)
```

This algorithm represents the simpler way to feed the dataset into the model and to determine the accuracy based on various test size.

### Further Methodologies

#### K Nearest neighbour Classification

Following the same process, several categories might be worked on to obtain precision. It is not a question of improving accuracy following a single categorization. Many categories for analysis can be explored. The output is a class membership in k-NN classification. An item is categorised by a plurality of votes of its neighbours, with the object attributed to the most frequent class of its k neighbours. If  $k = 1$ , the object is just assigned to that neighbor's class. KNearest Neighbors is one of Machine Learning's most basic yet crucial classification algorithms. It is within the controlled field of learning and is applied intensively in the identification of pattern, data and intrusion. It is commonly available in real-life settings since it is not parametric, which means that it makes no underlying premises regarding the distribution of data (in opposition to other algorithms like as GMM, which assume that the provided data are distributed Gaussian). We have some previous data (often referred to as training data) that classifies co-ordinates in attribute defined groups.

**Naïve Bayes Classification**

Naive Bayes is a basic approach for classifying: class-label modelling, expressed as feature-value vectors, where class-labeling is chosen from a limited set. There is not a single classification method, but a family of algorithms based on a similar principle: all Bayes classifiers assume that the value of a specific feature, given the class vario, is independent of the value of any other feature.

**Field description:**

**Twitters users comment:** This option defines the randomly selected comment section of different users. The twitter search engine selects this exact field. The search engine searches for each of the user's keywords and displays the results one by one. For example, if the company wants statistics or commentary on its particular product, the name of the product might be specified and the data selected for further arithmetic regression and correlation.

**Product description:**

This parameter defines which product is specified by users. Thus, in current criteria we may represent the use of the product and the market reach. By extracting the name of the product to which consumers remark, the data set is already pre-processed. These approaches are pre-processed by the preprogrammed engine of extracting the product description to be remarked on by the consumers. The keywords are called essential and distinctive terms and the device should be identified, for example if a user portrays @kellymat on my new iPhone 7 for a summer holiday. The preset search engine underlines the terms that companies want. In this situation, we are seeking the product dataset use of iphone7, the engine recognises these remarks with iphone 7 and manufactures them to the user for building this data set. Every comment is then split into many algorithms, and the keyword is then found, and the whole data set is defined in each cell.

**Users intention about the product:**

This column is specified by this data set as the principal class variable in which the procedure is performed. By analysing this column, the primary result will be retrieved. In order to get the appropriate outcome, data and information in this column are provided in the pre-processing method module and regression analysis is not available. Consider a single row from our dataset and how the following columns are built: The next row shows a user remark @sxsw I hope the festival this year is not as collapsive as the Iphone app this year. This phrase is now broken down into several fragments:

@sxsw I hope this  
 year's festival isn't as crashy as  
 this year's Iphone app

Now it is important to determine the aim column and the keyword specified in this particular remark

@sxsw I hope this  
 year's festival isn't as crashy  
 as this year's Iphone app.

The words are recognised and separated into each area. Blue shows the username and neutral words considered to be null points. The good words highlighted by green, Red identifies negative words and thus possibility to reverse the entire phrase into a negative phrase. A red mark is therefore recognised in the phrase that gives it a negative phrase. The product description to be stated is provided in yellow.

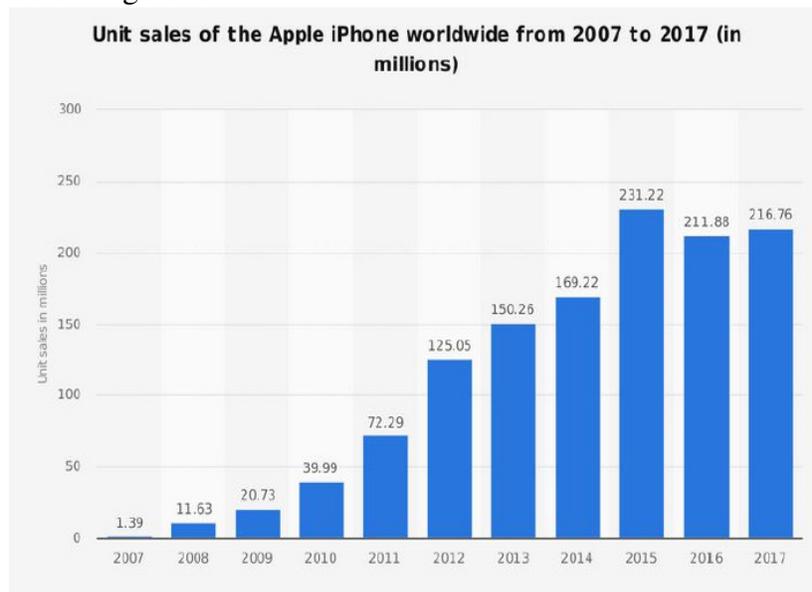
**Methodology:**

**How the problem will be approached:**

Whether a pleasant or a negative emotion, two classes. This dataset is therefore part of the Machine Learning and Analysis Classification Technique. In classification techniques there are few techniques.

**Technique involved for manipulation process:**

We are solely interested on the basis of two criteria, either good or negative. This lowers the complexity of the decision-making process. This was recognised by several datology companies as one of the finest strategies for predicting the result of the product. Banks employ this analytical approach to ensure their clients are either in a financial crisis or able to forecast that their businesses are in risk of not paying their rental. The high-end banks utilised this method to forecast their companies and customers' happiness using the dataset. Even an ordinary guy can utilise this approach to acquire a basic necessity like a house, for example. In the single product, it cannot invest too much in such a fast way. Small calculations and analyses, such as land information, sales rates, water supply, transit facilities, either rural or urban locations, were required, either close to the retail area. These are the characteristics in the tree of decision. When either yes or no response is answered one by one at a certain time, it cannot be dissolved. That's why we have a result in figure 1.

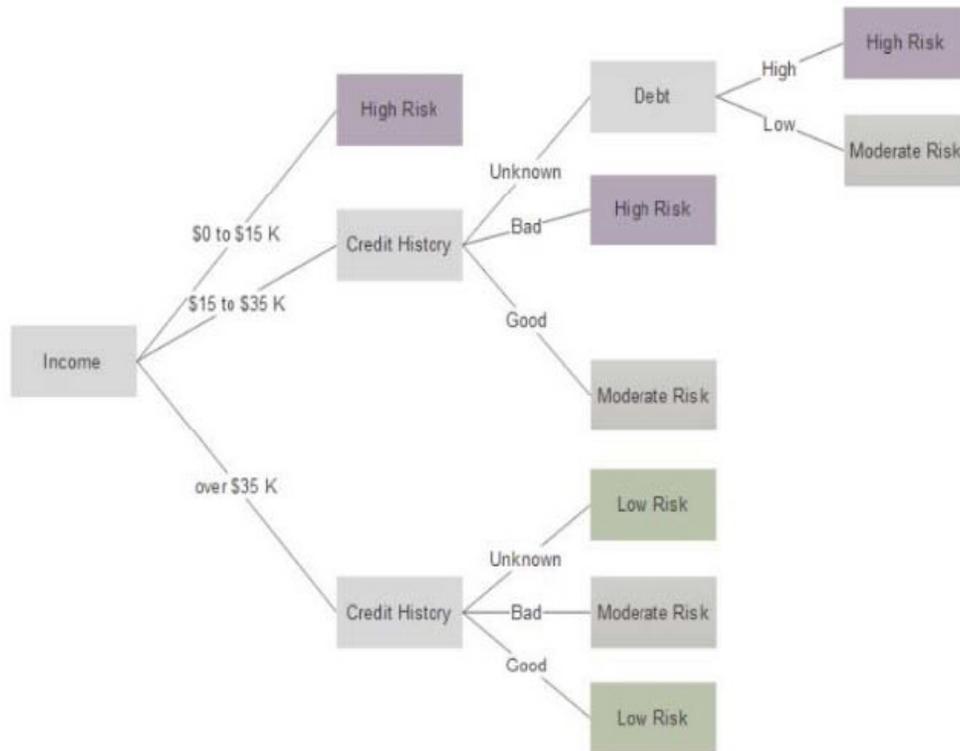


**Fig 1.** Unit sales of the Apple iPhone from 2007 to 2017

The major aim of the Decision Tree Technical Tree was derived from the manipulation neural networking method. The complete entropy has been produced in the main branch when people purchase their basic necessities like housing, for instance. In the single product he can't invest very much in that fast way. They had to participate in little calculations and analyses such

as land information, sales rate, water supply, transit facilities, rural or urban settlement, either close to shopping. These are the characteristics in the tree of decision. When either Yes or no response occurs one by one at a given moment, these questions cannot be broken down. That's why we have a result. The root node is progressively dissolved till a further fragmentation is not possible. The methodology of this sort of study is called decision-making tree. The Root node mentioned here is Income, it is further fragmented and the pre-processed results may be achieved on the basis of many criteria. The final result is generated via a regression analysis utilising Decision Tree methods with highly precise details in Fig 2.

**Example:**



**Fig 2:** Decision tree example

**IV. FUTURE SCOPE AND CONCLUSION**

Many techniques are used by e-mail clients to filter spam. They are driven by machine learning to ensure that these spam filters are updated continually. When spam filters are done on a rule basis, spammers cannot monitor the latest tactics. C 4.5 Decision Tree Induction Multi layer perceptron are some of the approaches for the spam filtering that ML supports. There are more than 325,000 viruses per day, and every piece of code resembles its prior versions by 90–98%. The coding pattern is understood by system safety algorithms driven by machine education Therefore, they quickly identify and guard against new viruses with 2–10% variation. Many websites now provide the opportunity to speak with representatives of customer care when exploring the site. All websites don't have a live manager to address your questions. You speak with a chatbot in most situations. These bots tend to retrieve and deliver information to the client from the website. The chatbots are moving on over time, meanwhile. They tend to better comprehend and provide better responses to customer inquiries, which may be done through

their machine learning algorithms. Thus, we may improve the relevant real-time applications with the greatest precision by considering the best categorization technique.

## REFERENCES

1. David Osimo and Francesco Mureddu, "Research challenge on Opinion Mining and Sentiment Analysis"
2. Maura Conway, Lisa McInerney, Neil O'Hare, Alan F. Smeaton, Adam Bermingham, "Combining Social Network Analysis and Sentiment to Explore the Potential for Online Radicalisation, Centre for Sensor Web Technologies and School of Law and Government.
3. A.Cutillo,R.Molva and T.Strufe.Safebook:A privacy preserving online social network Leveraging on real life.
4. S.Buchegger,D.Schioberg,L.Vu p2p social networking –early experience and insightsSNS2009.
5. Lucas C., "Sentiment Analysis a Multimodal Approach," Department of Computing, Imperial College London\_, September 2011.Celli, F., Pianesi, F., Stillwell, D. S., and Kosinski, M. 2013.Workshop on Computational Personality Recognition
6. Mining Facebook Data for Predictive Personality ModelingDejanMarkovikj, Sonja Gievska ,MichalKosinski, David Stillwell
7. Center of Attention: How Facebook Users Allocate Attention across Friends Lars Backstrom, EytanBakshy, Jon Kleinberg ,Thomas M. Lento, ItamarRosenn
8. Sentiment Analysis for Social Media R. A. S. C. Jayasanka, M. D. T. Madhushani, E. R. Marcus, I. A. A. U.Aberathne ,and S. C. Premaratne
9. Pang, B., Lee, L., Vaithyanathan, S.: "Thumbs up? sentiment classification using machine learning techniques", in Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Volume 10, July 2002, pp. 79-86.
10. Socher, R., et al.: "Semi-supervised recursive autoencoders for predicting sentiment distributions" in Proceedings of EMNLP '11 - the Conference on Empirical Methods in Natural Language Processing, ISBN: 978-1-937284-11-4, pp. 151-161
11. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: "Lexicon-Based Methods for Sentiment Analysis", in "Computational Linguistics", June 2011, Vol. 37, No. 2, pp. 267-307.
12. Chaovalit, P., Zhou, L.: "Movie review mining: A comparison between supervised and unsupervised classification approaches", in Proceedings of the Hawaii International Conference on System Sciences (HICSS), 2005.
13. Esuli, A. Sebastiani, F.: "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", in Proceedings of LREC-2006, Genova, Italy.
14. Baldini, N., Neri, F., Pettoni, M.: "A Multilanguage platform for Open Source Intelligence", Data Mining and Information Engineering 2007, 8th International Conference on Data, Text and Web Mining and their Business Applications, The New Forest (UK), Proceedings, ISBN: 978- 184564-081-1, WIT Transactions on Information and Communication Technologies, Vol 38, June 2007, 18-20.
15. Neri, F., Pettoni, M.: "Stalker, A Multilanguage platform for Open Source Intelligence", Open Source Intelligence and Web Mining Symposium, 12<sup>th</sup> International Conference on Information Visualization, Proceedings, ISBN:978-0-7695-3268-4, pp. 314-320, IEEE Computer Society, LSBU, London (UK), July 2008, 8-11.

16. Neri, F., Geraci, P.: "Mining Textual Data to boost Information Access in OSINT", Open Source Intelligence and Web Mining Symposium, 13th International Conference on Information Visualization, IV09, Proceedings, ISBN: 978-0-7695-3733-7, pp. 427-432, IEEE Computer Society, Barcelona (ES), 16-17/07/2009.
17. Y. Yang and M. O. Shafiq, "Identifying High Value Users in Twitter Based on Text Mining Approaches," 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2019, pp. 634-641, doi: 10.1109/TrustCom/BigDataSE.2019.00090.
18. S. Umamaheswari and K. Harikumar, "Analyzing Product Usage Based on Twitter Users Based on Datamining Process," 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2020, pp. 426-430, doi: 10.1109/ICCAKM46823.2020.9051488.
19. S. Hegde, S. G, S. M, S. R and S. L. N, "User Interest Prediction based on Social Network Profile with Machine Learning," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-6, doi: 10.1109/I2CT51068.2021.9418126.
20. M. I. Eshak, R. Ahmad and A. Sarlan, "A preliminary study on hybrid sentiment model for customer purchase intention analysis in socialcommerce," 2017 IEEE Conference on Big Data and Analytics (ICBDA), 2017, pp. 61-66, doi: 10.1109/ICBDAA.2017.8284108.