Analyzing Product Usage Based on Twitter Users Based on Data mining Process

E.Pavan Kumar[1], A.Rohith[2], P.Arun[3], B.Naresh[4]

[1, 2, 3]Scholar, Department of CSE in Jayamukhi institute of technological sciences, Narsampet, Warangal, Telangana, India

[4]Assistant professor, Department of CSE in Jayamukhi institute of technological sciences, Narsampet, Warangal, Telangana, India

ABSTRACT

In order to carry out the necessary modification and subsequent prediction, this study further develops the procedures and chronological approaches. analysis. We have obtained a dataset that is in real time and is based on the comment's areas for users of Twitter. The one-of-a-kind qualities of the dataset is that we have omitted everything except the specific fields. comments that create a new term based on the components of the existing word product. After that, more iterative extractions are performed in order to: in order to finish the dataset. Our data collection has three columns that are organised as follows: ideology held by the specific user or the issue that we are concentrating on improved the quality of any aspect of the product that we have been currently selling observing. In this specific collection of data, we have chosen to focus on the topic. On the use of the items by the customers who have been created by Apple or Google, two of the most successful firms in the world. Both of these colossal technological companies are quite knowledgeable about they're This age is dominated by technology, and those advances are becoming increasingly concentrated in the next digital initiatives and the things they produce will be more developed in the years to come. Due to the fact that they are now engaged in additional optimization in their products and growth in their market share a list of requirements to achieve their goals might be a difficult undertaking. project without the comments, suggestions, and criticisms of its users. projects that came before it, they are able to glean such characteristics from. dataset provided by Twitter's API, and they might improve it even more product. They might investigate the disadvantage, determine if the product has made its way into the market and has been received favourably there All of these various sources of information may be gleaned through social media. media as an alternative to carrying out a poll. A procedure like this one would be categorization becomes more difficult, and we cannot anticipate any outcomes accurate outcomes. Therefore, we will move on to the data from the social media mining Methods Used.

I. INTRODUCTION

When attempting to determine the characteristics of a specific user, one of the most crucial criteria and critical factors to consider is the data that the user has. Social Networking sites has become one of the main relevance for the transmission of information and news. The transmission of vital information from one part of the world to another has been the driving force behind the expansion and growth of the internet as a global communications network.

The transmission of data enables information to be improved in a unique manner and sent to target audiences. However, marketing and firm companies make advantage of the data in order to

provide their advertising goods while passengers are still in transit to their destination.

Further Data mining techniques were adjusted \sand employed in the dataset of the users. Extraction of the data set associated with a certain user or numerous data sets associated with a large number of users comes first. The next step is to extract the data set for a consistent interval of period. In its early stages, the system would consist of unstructured data. Since the unstructured data has been saved in a large database, it needs to be evaluated and then turned into structured data. It comprises of several properties and \sclasses. In the method of data collection, one of the important developments that has to be taken into consideration is the existence of classes and characteristics. After having received the structured data, the data may then be decomposed into a chronological order after it has been received. In the final stage, the extracted structured data is manipulated with the assistance of some machine learning techniques and statistical analysis regarding logging in, logging out, posting photos, or being tagged as various criteria. This allows for the personality assessment of a specific user to be obtained. [1] The mental view of the attitude toward privacy is provided by the "Study Challenging task on Recommender Systems and Sentiment Classification." Only a few theories and insights on the psychology of users' attitudes toward illegal activity on social media have been gleaned from this research. [2] We have used the method, but it is not exactly the same. Instead, we have created a real-time model, which simply gives the theoretical analysis. [3-\s5] We had embraced the notions that are theory building and also provided the backend method of how the users' privacy is utilised as data. [8] This presentation will mostly cover issues about manipulation on Facebook and how it encourages people to visit it. The experience that its consumers have is articulated really well. The primary conceptual paper that we have taken for the purpose of serving as a reference is located at [18-20].

II. PROBLEM ANALYSIS

Some social media platforms secure the privacy of their customers in accordance with additional confidentiality standards, but other social media platforms expose the information of their customers as open - source software so that it may be manipulated in further ways. Facebook protects the data of its users thanks to the privacy agreement status granted by those users; nonetheless, some commercial enterprises acquire the data of Users on facebook by paying Facebook with a sizeable fee in exchange for access to the Facebook domain.

Due to the fact that the data contains a great number of users' private communications, Facebook security has a quality vault. However, beyond this, they have the ability to sell the data. Twitter, on the other hand, generates comments and makes the data it collects available in an open source format, which enables data scientists to construct the project and conduct accurate research. Each remark left by a different user is given a random weight depending on whether or not it was left on an iPad or via a Google service. The primary purpose of this project is to determine, based on the feedback generated by participants of Twitter, the percentage of people who are reached by each of the following goods. The primary purpose of this specific piece of research is to speculate on the intentions of customers towards a product that is being monitored and maintained under surveillance. There

are many other approaches to machine learning that may be utilised to conduct the modification; however, in order for us to be successful in completing this work, we are using an extremely targeted approach. As a result, we will only use a single categorization method for the purpose of conducting the regressive analysis of our product. It's possible that the outcome of our work will be a prediction about how the product will be used. By deconstructing the comments and closed observations, however, we are able to carry out a variety of manipulations, such as improving the specifications of the gadget, the market reach of the product, the performance of the camera, or the performance of the product itself. Each and every remark and its associated keywords are examined, and based on the attribute data, each comment and its associated keywords are assigned a score indicating whether it is favourable or negative. This procedure is carried out in an iterative manner, and the corresponding data set has been compiled. The dataset is now located at an unstructured data is now being fed within the module that was constructed with the assistance of python packages There will only be one specific regression analysis that will be evaluated and targeted to process the outcome.

Proposed system

This specific issue may be broken down into just two categories, namely, whether it is a pleasant feeling or a bad emotion. As a result, the Classification approach of the supervised learning methodology and analysis applies to this dataset. There aren't very many different approaches that go into making up categorization techniques.

Methods that are used during the Manipulation Process

We only become engaged based on two criteria, and they are either happy emotions or bad emotions. The use of the Decision Tree approach therefore results in a reduction in the level of complexity. It has been determined that this is one of the most effective methods, which is why it is still used by a great number of data scientist companies to make predictions about product outcomes. Banks utilise this method of analysis to ensure that their customers should either be able to or in the event of a financial crisis, they could easily predict that their companies are in danger that they could not pay their lease. This protects the banks from losing customers who are unable to pay their bills. High-end banks have made use of this method in order to anticipate the level of pleasure felt by their client corporations and individual customers based on the dataset. For instance, even an average Joe may utilise this strategy to purchase something as essential as a home for himself if he so chooses. He is unable to invest a very substantial sum of money on a single product in such a short length of time. They were required to include in a number of variables in their calculations and analyses, such as the location of the property, its selling price, the availability of water and transportation, whether the area was rural or urban, and whether or not it was close to stores for shopping. And those are the characteristics that are being considered by the logistic regression. When these questions are addressed one at a time, either a yes or a no response will emerge at a certain moment; hence, they cannot be dissolved. Hence our result occurs.

Then we can go into this specific subject, we need to have a good understanding of the numerous terminologies that are associated with the decision tree method. These terms are as follows:

Entropy: They are the fundamental components of the overall set with which we are going to be concerned. In this particular instance, we are going to be working with a Twitter dataset that compiles information on how users interact with the service, regardless of whether they have negative or positive intentions about the product. The whole of the system in which we are participating may be characterised by it. The method of categorization is based on the level of entropy in the system. Obtaining New Information: This part offers the required information that has been collected by the dataset and is presented here for your convenience.

Particularly the remark part that each of our separate datasets provides. This keeps the primary pieces of information that are being provided into the module in place.

The entropy is broken down into further classes at the leaf node, and at this point, the entropy can no longer be fragmented any further.

In this particular case, the culmination of events is referred to as the Leaf node.

Root Node: The root node illustrates the information gain that has been fragmented in order to describe the number of users who have provided a bad sentiment, while the number of users who have provided a favourable emotion will be set aside.

Adjusting the details of the model: The primary extracted portion of our CSV file contains an excessive amount of information that has been supplied to us for the purpose of data analytics. However, the data that we are going to enter into our model need to be simplified even more. Since our class variable is either a favourable remark or a bad comment about the specific product, we have chosen to ignore the comment section column as a result. We have around eight items that are now being evaluated. The remarks identifying such a product are removed, followed by an analysis in which the comments are classified as either good or unfavourable.

We have disassembled the module into even more simple components, each of which is a float variable. We have assigned a float number to each and every product, and then we have entered that information into the model. The relevant Python code is then specified in order to activate the prediction state and turn them into actual work.

During the preprocessing step, the remark area was eliminated since the csv file includes information that is continually updated on our goods. In addition, the "null" comments, which include things like "Can't tell" and "no feelings toward the goods," have been optimised since they provide unwelcome information about the product. There are around 3000 data included inside the CSV file as a whole, however they have been reduced to between 450 and 600 data so that they may be fed into the module and preprocessed.

Algorithm

x = np.array(train.drop(["Users intention regarding the product", 1)), where x represents the user's intention.

y = np.array (train["Users intention about the product"])

x train, x test, and y train, y test are the results of the equation train test split(x, y, test size = 0.9, and random state = 100).

model=model.fit(x train,y train)

y pred is equal to the model's prediction of x test.

print("accuracy:",accuracy score(y test,y pred)*100)

• This approach is the simplest technique to assess the correctness of the model based on a variety of test sizes, and it is used in the process of feeding the dataset into the model.

This specific issue may be broken down into just two categories, namely, whether it is a pleasant feeling or a bad emotion. As a result, the Classification approach of the machine learning methodology and analysis applies to this dataset. There aren't very many different approaches that go into making up categorization techniques.

The process of manipulation involves the following techniques:

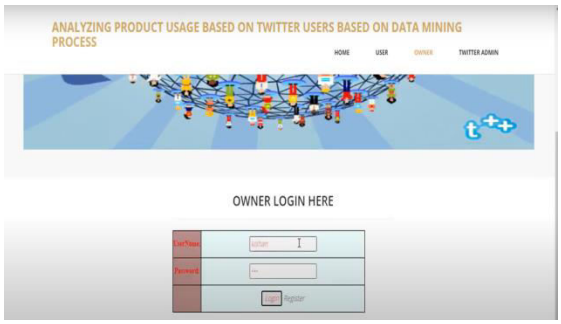
We are only interested based on two criteria, and they are either happy emotions or bad emotions. The use of the Decision Tree approach therefore results in a reduction in the level of complexity. It has been determined that this is one of the most effective methods, which is why it is still used by a great number of data scientist companies to make predictions about product outcomes. Banks employ this method of analysis to ensure that their customers will either be able to or in the event of a financial crisis, they could easily predict that their companies are in danger that they could not pay their lease. This protects the banks from losing customers who are unable to pay their bills.

High-end banks have made use of this method in order to anticipate the level of pleasure felt by their client corporations and individual customers based on the dataset.

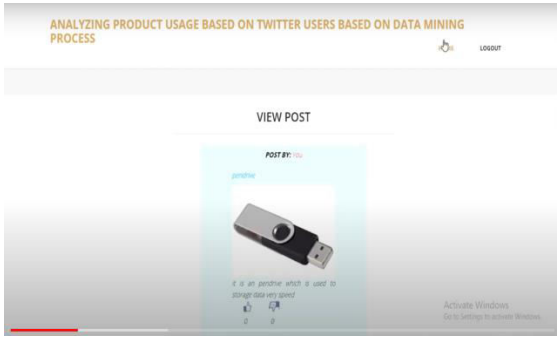
For instance, even an average Joe may utilise this strategy to purchase something as essential as a home for himself if he so chooses. He is unable to invest a very substantial sum of money on a single product in such a short length of time. They were required to factor in a variety of factors when performing their calculations and analysis, including information regarding land, the selling price of that land, water facilities, transportation facilities, whether the location was rural or urban, and whether or not it was close to shopping areas. These are the characteristics that are being considered by the decision tree. When each of these inquiries is addressed individually, either a yes or a no response will result at a given moment, making it impossible to disassemble them.

III RESULTS

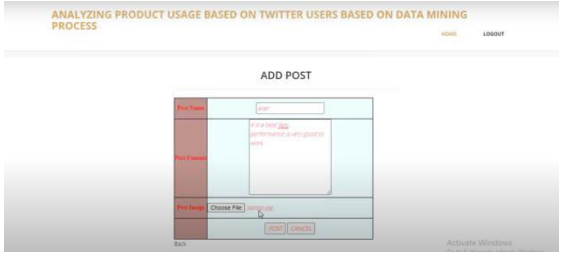
Home Page



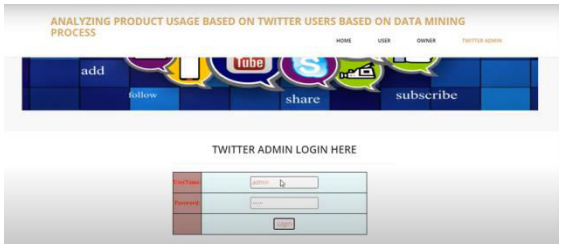
Owner login Page



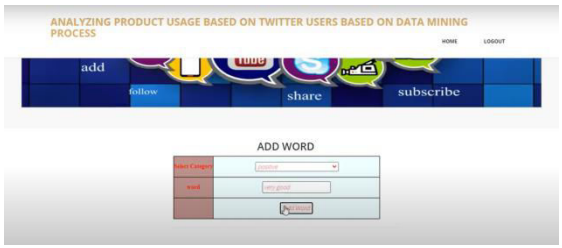
View Post



Add post

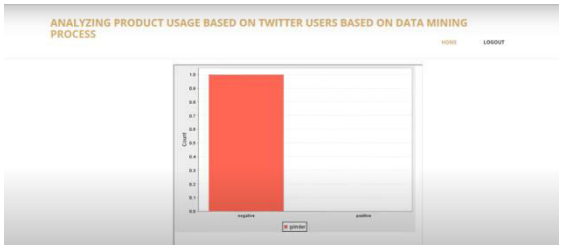


Admin login



Add Word

Add Comment



IV CONCLUSION

Email clients may choose from a variety of spam filtering strategies to combat unwanted messages. Machine learning is used to fuel these spam filters so that they can ensure they are always receiving the most recent updates. Rule-based spam filtering is ineffective because it does not keep up with the most recent strategies used by spammers. ML serves as the driving force behind a number of different spam filtering strategies, including Multi-Layer Perceptron and Decision Tree Induction.

Every day, about 325,000 new malwares are discovered, and each new piece of code is between 90 and 98 percent comparable to its earlier variants.

The machine learning-based security systems on the system are able to recognise the coding pattern.

As a result, they are able to identify new malware with a variation of between 2 and 10 percent quickly and provide protection against it.

These days, a growing number of websites provide visitors with the opportunity to carry on a conversation with a customer service agent even while they navigate other parts of the website. However, not all websites have a real executive available to respond to your questions and concerns. The majority of the time, you will converse with a chatbot. The majority of the time, these bots will harvest information from the website and then provide it to the users. In the meanwhile, chatbots continue to improve with time.

Because of the machine learning algorithms that it employs, they have a greater capacity to comprehend the questions asked by users and provide superior responses to those questions.

Therefore, by taking into account the most effective classification approach, we are able to improve the particular real-time application with the highest possible rate of accuracy.

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 Berminghan, "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, 12th 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.