BRAND SENTIMENT ANALYSIS USING TWITTER

Sakshi Agarwal   Prof.Prasanna M

Vellore Institute of Technology, Department of Computer Applications,
School of Information Technology and Engineering.

Abstract-Our goal in this paper is to try and analyze tweets. Social media is a place where people can express themselves about any particular product. And Twitter is a platform where people are encouraged to express their opinions on each product. Brands can use these platforms to their advantage since it gives them a free product review service with thousands of testers. We extract the tweets from Twitter and classify each tweet into a "positive" or "negative" sentiment. We are using a number of algorithms to predict the sentiment for our dataset. In our case we go through each tweet and determine whether it affects the brand or the product positively or negatively. This way, companies can get a general idea of whether the attitude towards the product is good or bad upon which they can act accordingly. It should be mentioned that the timing of such analysis is relevant to the discussion. For example, running such a model when a new product releases is the best way to use or implement this idea.

Keywords-sentiment Analysis, opinion mining, machine learning algorithms, tokenization, tweet

I. INTRODUCTION

We believe that we can harness the reach and inclusivity of social media to find out the response of the general public towards a certain product. This way we can accelerate the process of reviewing a product and also the correction and rerelease. Both the producer and the consumer benefit from this sort of analysis. In this paper we look particularly at twitter mentions since we believe that it is one of the best platforms to ascertain how people feel about different products. The nature of the medium makes it so. We shall use a data set from kaggle and then apply various machine learning algorithms to predict the sentiment of the tweets and would find which algorithm provide us with most accurate resultOpinion mining which is also known as the sentiment analysis, is a sort of characteristiclanguage preparing concept for following and analyzing the mind-set of people in general abouta specific item or product. Assessment mining, which is likewise called assumption investigation, includes building a framework to gather also, arrange conclusions about an item. Robotized assessment mining takes usage of machines. Conclusion mining can be helpful in a few different ways, it can assist advertisers with the achievement of a promotion or new item preparation, figure out which adaptations of an item or administration are well known. For instance, a survey on a site may be very sure or certain views about an advanced camera, in any case, be clearly negative about how substantial it is. Being ready to distinguish this sort of data in this manner gives the merchant a much clearer image of popular conclusion than overviews what center gatherings do, in perspective of the fact that the information is made by the client. There are a few difficulties in opinion mining. The first is that a word that is viewed as positive in onecircumstance might be viewed as negative in another circumstance. Take "long" for example. Ina scenario let's take, that a client said a PC's battery life was long, that would be a positive recommendation. On the other hand, the other client said that the PC's fire up time was long, inany case, that would be is a negative conclusion. These distinctions imply that a feelingframework prepared to accumulate feelings on one sort of item or items highlight may notperform very well on another. A second test is that individuals don't in every case expresssentiments a similar way. Most of the content that is prepared, depends on the way that the little contrasts between two pieces of text don't change the significance without a doubt. In conclusionmining, "the film was incredible" isaltogether different from "the film was not extraordinary".At last, individuals can be conflicting in their announcements. Most surveys will have both positive and negative remarks, which is fairly reasonable by both the sentences. In any case, the more casual the medium, the more probable individuals are to combine various assumptions.

II. PROBLEM STATEMENT

Twitter is a popular social networking website where members build and interact with messages known as "tweets". This helps people to convey their sentiments about various topics. Several people did sentiment analysis on such tweets to assemble information about products or to conduct market analysis. Furthermore, with the recent developments in machine learning algorithms, we are able provide more accurate results for sentiment analysis predictions. In this paper, we will perform sentiment analysis on "tweets" using various different machine learning algorithms. We will differentiate the polarity of the tweet whether it is positive or negative. If the tweet contains both positive and negative word, the more superior sentiment should be chosen as the final sentiment. We have used the dataset from Kagglewhich has positive and negative words contained in it. The datasetcontains emoticons, usernames and hashtags which are required to be cleaned and transformed into a form which is suitable to use. We will extract useful features from the text such as unigrams and bigrams which will represent the "tweet". We will apply several machine learning algorithms to conduct sentiment analysis using the extracted features. Finally, we will get our results at the end as which algorithm is giving us more accurate result and further show the graphical representation of the results so obtained.

III. RELATED WORK

[1] This paper analyzes the use of different methods utilizing Naive Bayes Classifier for more accurate sentiment analysis. It observes that a combination of methods like effective negation handling, word n-grams and feature selection by mutual information results in a significant improvement in accuracy. It tries to build a simple Naive Bayes model that has linear training and testing time complexities. The method can be generalized to a number of text categorization problems for improving speed and accuracy.

[2] This paper aims to filter the data collected every day from various social media sources to obtain useful information. It classifies data collected from users as positive, negative and neutral. It tried to separate objective and subjective information i.e. facts from opinions using Naive Bayes Classifier. The method used filters noise from useful data so surveyors and data analyzers can utilize the information effectively.

[3] Various Machine Learning algorithms are currently employed in classification of tweets into positive and negative classes supported their sentiments, like Baseline, Naive Bayes Classifier, Support Vector Machine etc. This paper contains implementation of Naive Bayes using sentiment140 training data using Twitter database and proposes a method to improve classification. Use of SentiWordNet alongside Naive Bayes can improve accuracy of classification of tweets, by providing positivity, negativity and objectivity score of words present in tweets. For actual implementation of this technique python with NLTK and Python-Twitter APIs are used.

[4] In this paper sentiment analysis is applied to urge information from digital news articles about its positive or negative sentiment regarding a specific politician. This paper suggests an easy model to research digital newspaper sentiment polarity using naive Bayes classifier method. The model uses a group of initial data to start with which can be updated when new information appears.

[5] This paper focuses on finding the solution for the domain transfer problem by using both old-domain data and unlabeled new-domain data. For the old domain data, an effective measure called Frequently Co-occurring Entropy (FCE) is used. To gain knowledge from the new-domain data, Adapted Naïve Bayes, a new version of Naïve based classifier is proposed. The experimental results achieved shows that the proposed approach provides a better performance than the Naïve Bayes Transfer Classifier (NTBC).

[6] This paper focuses to build an algorithm that can perfectly classify tweets on Twitter as positive or negative, taking into account a query word. High accuracy is acquired on classifying Twitter data sentiment using machine learning techniques like the Naïve based approach.

[7] This paper tries to find tweets to predict data and sentiments from Twitter with more accuracy. The information from tweets are drawn out using keyword based knowledge extraction. The knowledge drawn is further strengthened using domain specific seed based enrichment technique. The proposed methodology facilitates the extraction of keywords, synonyms and parts of speech from

tweets which are then used for tweets classification and sentiment analysis. The proposed system is tested on a collection of 40,000 tweets. The proposed methodology has performed better than the existing system in terms of tweets classification and sentiment analysis.

[8] This paper describes how mobile providers acquire new customer through advertisement and campaign on social media platforms. Now, the problem arrives to the mobile operator is to measure brand prestige based on customer contentment through customer sentiment analysis from Twitter. This paper uses Naive Bayes, support vector machine and decision tree classifier method for data cleaning, data balancing and sentiment analysis. This paper measures customer satisfaction on 5 products that is 3G, 4G, short messaging, voice and internet services.

[9] This paper describes how internet has become a basic necessity for people and how social networking applications and platform give people a voice to raise their opinions regarding daily issues. Twitter provides a great platform where people can convey their emotions about a product, public services and many other things. This paper aims to discover the sentiment behind the opinion of people in the format of text on various topics. This paper also shows us how scientist is analyzing opinion of people on different topics such as movies, commercial products, and social issues. Opinion investigation of Twitter data helps to dissect tweets or comment into positive and negative so that a person can analyze about a product or a service. This paper also explores various sentiment analysis which is applied to the data.

[10] This paper describes how opinion and reviews plays an important factor in preparing views and determining the success of a brand, product or a service. This paper also tells us that as there is hike of social media in the world the stakeholders of one try to use social media platform to know opinion about their product from public. Twitter data is very informative and is easy to extract as it is in text format. This paper helps the brand by using the concept of sentiment analysis of people's opinion. The author of this paper has used a predefined data set which can be used for prepossessing. This paper has used probabilistic model based on Bayes

theorem which was used for grammatical corrections and research studies. This paper also compares Naive Bayes and support vector machine and an artificial neural network model which is called as Muti-layer perceptron.

[11] This paper describes how social media is a type of repository where one can find information of around the world. This paper also so says that Reddit it is a rising social media platform which contains very useful information. Reddit platform can be used by the people to share their opinion on a topic. Therefore, read it can be used as a platform where we can achieve large set of data on a particular product or topic. This paper is an analysis of human to stick to a topic while commenting on social networking site in this case Reddit.

[12] This paper describes us how social media helps to generate content and offers a unique opportunity to understand how individual or group of people interact. This paper focuses more on Reddit which has valuable information that can be processed and analyzed. The main goal of this paper is to investigate using sentiment analysis and activity analysis of behavioral and content it of suspicious Reddit account. This will help to detect suspicious account and it will be easier to ban them from that social media platform.

[13] This paper tells us about the sentiment analysis that analyses the opinions, evaluations, attitudes, emotions and sentiments of the people any written language. The increasing significance of sentiment analysis coincides with the growing significance of social media including their reviews, discussions, blogs, etc.

[14] This paper tells that how sentiment analysis helps to find out solutions regarding the post what people share regarding the problems they encounter on the web. Sentiment analysis is becoming so much popular due to the abundant opinions present in the social media. It segregates the content in the positive and negative manner on the bases of sentiments. This paper naive Bayes classifier and support vector machine for finding out how individuals feel whenever they encounter something new on the web.

[15] This paper tells us that how the text mining has become so much popular due to its characteristic of finding out valuable information from unstructured

text. Since Unstructured text can't be used by the computers for processing so various algorithms and techniques are being used to fetch out the important data.

[16] This paper tells us that using sentiment analysis technique how we can easily find out the expressions and emotions from the comments, discussions and blogs written by the user on the social media either it is of regarding friendship, support, anger, happiness etc.

[17] This paper tells us about the consequences made after using the sentiment analysis technique in social media comments of the users knowing about the sentiments and loyalty regarding the brand or products. This knowledge correlates knowledge consequences with experimental results made by the Osservatorio di Pavia, which is an Italian institute of research specialized in media analysis at theoretical and empirical level, involved in the analysis of political communication in the mass media.

[18] Internet is a huge area where people can express their feelings. this paper shows that social media is a very important aspect in marketing area as it helps to get opinion from online users about the product or services. This this paper suggests to have knowledge of some confirmed and some rejected parts of the city being reviewed from the concept of Bologna on the internet. With the help of sentiment analysis, the idea of perception of city brand can be a interesting way to investigate with the more qualitative and quantitative technique.

[19] According to this paper treat Twitter is providing and area so that people can express their opinion and can discuss about an event services or brand. a particular company or brand use this platform to get feedback about their services and improve the quality and quantity. but the main problem over here is that the bulk amount of data is very difficult to detect the consumer's opinion. so to eliminate this problem sentiment analysis using naive Bayes and lexical dictionary is used so that a machine learning technique can be e developed for sentiment analysis on processed Twitter data can be used. Positive and negative sentiments can be separated which can be used by the brands so that they can do development on the product or service.

[20] This paper focuses on sentiment analysis in social media. It gives a brief description as how social media can be used and what are its. it is about how social media contains large amount of raw data and with proper sentiment analysis it can be used to separate positive and negative comment which can be processed for the development of product or brand. The result of the people considers most article uses of opinion lexical method to analyze text sentiment in social media. This paper also shows us that Twitter with sentiment analysis can be used in many world, health care, politics and businesses.

[21] This paper analyzes the use of different methods utilizing Naïve Bayes Classifier for more accurate sentiment analysis. It observes that a bundle of methods like effective negation handling, word n-grams and feature selection by mutual information helps in a large upgraded accuracy. It tries to build a simple Naïve Bayes model that has linear training and testing time complexities. The method can be concluded to a number of text categorization issue for enhancing speed and accuracy.

[22] This paper aims to filter the data collected every day from various social media sources to obtain useful information. It classifies data collected from users as positive, negative and neutral. It tried to separate objective and subjective information i.e. facts from opinions using Naïve Bayes Classifier. The method used filters noise from useful data so surveyors and data analyzers can utilize the information effectively.

[23] Various Machine Learning algorithms are currently employed in classification of tweets into positive and negative classes based on their sentiments, such as Baseline, Naïve Bayes Classifier, Support Vector Machine etc. This paper contains implementation of Naïve Bayes using sentiment140 training data using Twitter database and proposes a method to improve classification. Use of SentiWordNet along with Naïve Bayes can improve accuracy of classification of tweets, by providing positivity, negativity and objectivity score of words present in tweets. For actual implementation of this system python with NLTK and Python-Twitter APIs are used.

[24] In this paper sentiment analysis is used to get data from digital news articles about its positive or negative sentiment regarding a particular politician. This paper provides a simple model to analyze digital newspaper sentiment polarity using naive Bayes classifier method. The model uses a set of

initial data to start with which will be having the new information when new information appears.

[25] This paper aims at solving the domain transfer idea by making use of both old-domain data and unlabeled new-domain data. For the old domain data, an effective measure called Frequently Co-occurring Entropy (FCE) is used. To achieve data from the new-domain data, Adapted Naïve Bayes, a new concept of Naïve based classifier is proposed. The experimental results obtained shows that the proposed approach provides a better solution than the Naïve Bayes Transfer Classifier (NTBC).

[26] This paper aims to construct an algorithm that can precisely classify tweets on Twitter as positive or negative, depending on a query word. More accurate results are obtained on classifying Twitter data sentiment using machine learning techniques like the Naïve based approach.

[27] This paper aims to find tweets to validate data and sentiments from Twitter specifically. The data from tweets are obtained using keyword based knowledge extraction. The knowledge obtained is further increased using domain specific seed based enrichment technique. The proposed methodology facilitates the extraction of keywords, synonyms and parts of speech from tweets which are then used for tweets classification and sentiment analysis. The proposed system is tested on a collection of 40,000 tweets. The proposed methodology has performed better than the existing system in terms of tweets classification and sentiment analysis.

IV. PROPOSED SYSTEM

In this paper we will first preprocess the data and then run every algorithm one by one to get the accuracy of the model and would finally conclude the best model with best accuracy to be used for sentiment analysis. The following steps have been performed for sentiment analysis:
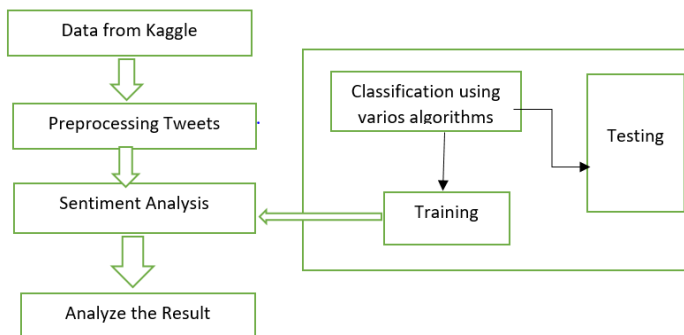

Fig: Twitter Sentiment Analysis Overall Process

- Collected data from kaggle
- Split data into training and testing data
- Preprocessing the data
- Applying various machine learning algorithms
- Visualize the result
- Give graphical representation for every result

We have used kaggle to get the data set which have train and test data. Train data is a csv file which has 31962 rows and 3 columns with attributes as tweet_id, sentiment, tweets and another file which has 17198 rows and 2 columns with attributes as tweet_id, tweets. The data have tweets from customer and that data contains various symbols, urls, hashtags and various other irrelevant things which we need to clean to get the appropriate results so for this the first step we need to perform is preprocessing.

In the preprocessing phase we have removed all the urls, symbols, hashtags and every irrelevant thing to make the data clean and also performed tokenization lemmatization and normalization for making the model easily understand the data after preprocessing we will also get csv files with frequency distribution of unigrams and bigrams respectively.

After the data is cleaned we have then applied each algorithm one by one which includes:

- Naïve Bayes
- Maximum Entropy
- Decision Tree
- Random Forest
- SVM
- Multi-Layer Perception
- Recurrent Neural Networks
- Convolutional Neural Networks

In every algorithm we have shown the unigram and bigram features for once for frequency and once for presence for which we observed that using bigram features are giving more accurate results then unigram. At last we concluded that Convolutional Neural Networks gives most accurate result which is 93% and below is the graphical representation for comparison of all the algorithms.
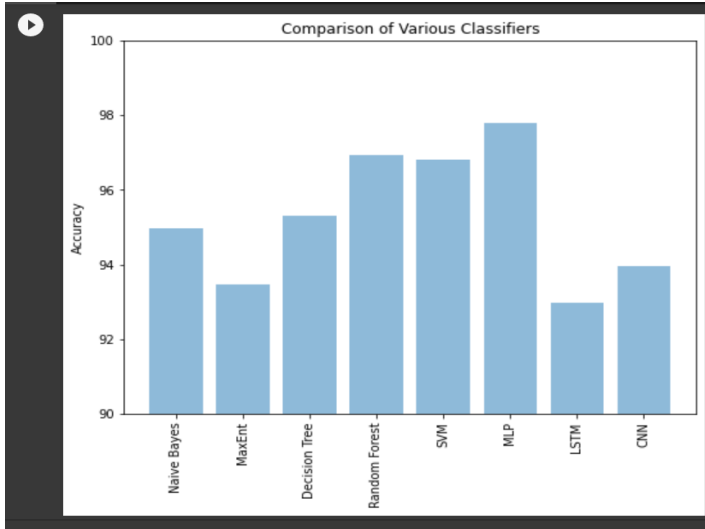
Fig: graphical representation of algorithms performance

FIG: Majority Voting Result

CONCLUSION

In this paper the brand sentiment analysis was done using twitter for which initially the data was cleaned all the hashtags, urls, symbols present in the data were removed and made suitable to be used for our model after that several machine learning algorithms like Naïve Bayes, Maximum Entropy, Decision Tree, Random Forest, SVM, Multi-Layer Perceptron, Recurrent Neural networks and Convolutional Neural Networks were used to find whether the comments used in tweets are positive or negative which will be beneficial for the company to know about the products review as received from the users through tweets and the most accurate algorithm among them was analyzed for reproducing better results for sentiment analysis further to improve the accuracy the two different features was used which are unigrams and bigrams and noticed that the building the feature vector with bigrams have enhanced the accuracy as compared to the unigrams and then the five best algorithms were merged to obtain the most accurate result.
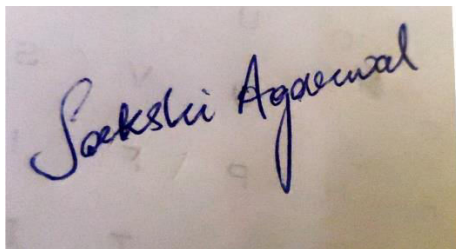
V.FUTURE SCOPE

We can further improve the accuracy of the model by helping the model to identify whether the tweets are good better or best instead of just identifying whether it is good or bad which will be very useful for analyzing the sentiment with more accurate result and would help to better understand the sentiments of the customer for the respective products we can also improve our model by adding punctuation marks whichwe discarded during preprocessing phase as a sentence with punctuation mark depicts a completely new emotion and can make a sentence different from one another.

REFERENCES

[1] Songbo Tan, Xueqi Cheng, Yuefen Wang, Hongbo Xu (2009). Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. Lecture notes in Computer Science, 5478, pp. 337-349.

[2] M. Venugopalan, D. Gupta (2015). Exploring sentiment analysis on twitter data. Eighth International Conference on Contemporary Computing (IC3), pp. 241-247.

[3] Narayanan V., A. I. (2013). Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model. Intelligent Data Engineering and Automated Learning, vol. 8206.

[4] R. Batool, A. M. Khattak, J. Maqbool, S. Lee (2013). Precise tweet classification and sentiment analysis. IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS), pp. 461-466.

[5] C. Troussas, M. V. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. IISA Piraeus, 1-6.

[6] LopamudraDey, S. C. (2016). Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier. arXiv:1610.09982.

[7] YustinusEkoSoelistio, M. R. (2015). Simple Text Mining for Sentiment Analysis of Political Figure Using Naive Bayes Classifier Method. arXiv, arXiv:1508.05163 .

[8] Vidya, N. A., Fanany, M. I., & Budi, I. (2015). Twitter sentiment to analyze net brand reputation of mobile phone providers. Procedia Computer Science, 72, 519-526.

[9] Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. International Journal of Advanced Computer Science and Applications, *10*(2), 361-374.

[10] Firoz Khan, A. M., Meghana, M., Pavan Kumar, P. S., & Rakshanda, B. K. Sentiment Analysis of Twitter Data.

[11] Puri, S., &Romanowski, C. J. Statistical Analysis of Comments on Reddit.

[12] Lee, E., Ashford, J., Turalska, M., Turner, L., Liao, V., Bellamy, R., ... & Whitaker, R. An Exploratory Analysis of Suspicious Reddit User Accounts based on Sentiment and Interactions.

[13] Chimaobiya, O. J., &Priya, M. H. Sentiment Analysis and Opinion Mining.

[14] Srividya, K., &Sowjanya, A. M. Sentiment Analysis of Face Book Statuses.

[15] Dang, S., & Ahmad, P. H. (2014). Text mining: Techniques and its application. *International Journal of Engineering & Technology Innovations*, *1*(4), 866-2348.

[16] Ahkter, J. K., & Soria, S. (2010). Sentiment analysis: Facebook status messages. *Unpublished master's thesis, Stanford, CA*.

[17] Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012). Sentiment Analysis on Social Media. *ASONAM*, *12*, 919-926.

[18] Grandi, R., &Neri, F. (2014). Sentiment analysis and city branding. In New Trends in Databases and Information Systems (pp. 339-349). Springer, Cham.

[19] Rasool, A., Tao, R., Marjan, K., & Naveed, T. (2019, March). Twitter Sentiment Analysis: A Case Study for Apparel Brands. In Journal of Physics: Conference Series (Vol. 1176, No. 2, p. 022015). IOP Publishing.

[20] Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. Procedia Computer Science, 161, 707-714.

[21] Kaur, A., &Baghla, S. (2018). Sentiment Analysis of English Tweets Using Data Mining.

[22] Norman, J., Mangayarkarasi, R., Vanitha, M., Praveen Kumar, T., &UmaMaheswari, G. (2017). A Naive-Bayes strategy sentiment for sentiment analysis on demonetization and Indian budget 2017-case-study. Int J Pure Appl Math, 117(17), 23-31.

[23] Dinu, L. P., &Iuga, I. (2012, March). The Naive Bayes classifier in opinion mining: in search of the best feature set. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 556-567). Springer, Berlin, Heidelberg.

[24] Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.

[25] Mansour, A. M. (2018). Texture classification using Naïve Bayes classifier. IJCSNS Int. J. Comput. Sci. Netw. Secur, 18(1), 112-121.

[26] Suppala, K., & Rao, N. (2019). Sentiment analysis using naïve bayes classifier. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8, 264-269.

[27] Kaur, S. Naive Bayes Classification Technique for Opinion Mining in Data Mining.

[28] Kohavi, R. (1996, August). Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In Kdd (Vol. 96, pp. 202-207).

[29] Kaviani, P., &Dhotre, S. (2017). Short survey on naive bayes algorithm. International Journal of Advance Engineering and Research Development, 4(11), 607-611.

[30] Kaviani, P., &Dhotre, S. (2017). Short survey on naive bayes algorithm. International Journal of Advance Engineering and Research Development, 4(11), 607-611.

[31] Baquillas, J., &Gozun, B. Twittersphere's reaction to Archer's graduation: A case of an unofficial brand ambassador.

[32] Preety, S. D., &Dahiya, S. (2015). Sentiment analysis using SVM and Naive Bayes algorithm. International Journal of Computer Science and Mobile Computing, 4(9), 212-219

Student Digital Signature   **Guide Digital Signature**