

EMOTION ANALYSIS OF TWEETS

D.Durga Prasad

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
durgaprasad.d@vishnu.edu.in

Pardhu Guttula

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
19pa1a04c7@vishnu.edu.in

Tangellapalli Sai Ramya Manasa

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
19pa1a04f5@vishnu.edu.in

Vemulapalli Bavya Sri

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
19pa1a04h3@vishnu.edu.in

Thote Prashanth

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
19pa1a04g1@vishnu.edu.in

Pitani Niranjan Sai

Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, INDIA
19pa1a04d2@vishnu.edu.in

Abstract—As technology is developing day by day messages, social media platforms became a platform of representing person's life, thoughts and their behaviour. We all know this developing technology has more freedom to share their thoughts, personal feelings, what they are feeling at the current moment within a single second, within a single sentence on social media. There will be plenty of data available to know about a person behavior and this data can be used for detecting various emotions. Our project Emotion Analysis of tweets itself says that to detect a particular emotion held by a tweet. That is done by analyzing expressed opinions, images, sentiments, and other activities. In our case we took mainly posted opinions or tweets as input to implement our NLP (Natural Language Processing) model. This project is about constructing machine learning models to find which model is predicting emotion of tweets with highest accuracy.

Index Terms—Twitter, Sentiment, Social media, natural language processing

I. INTRODUCTION

Twitter, is undoubtedly one of the most popular social media platforms available today, which was launched in 2006, There are nearly 100 million daily active users and 500 million tweets sent daily. This twitter platform is very easy to use and even a normal person who is not aware of using any technology can use this platform easily. The main aim of using this platform is to receive world-wide news, follow any celebrities they want, or can stay in-touch with their friends. Initially Twitter has begun as an SMS-based platform, where a person can send a message with approximate length of 140 characters. However, as Twitter platform had grown to unimaginable level and become a very important and necessary web platform to share any updates regarding the things that are happening around the world. The increasing exponential graph of Twitter

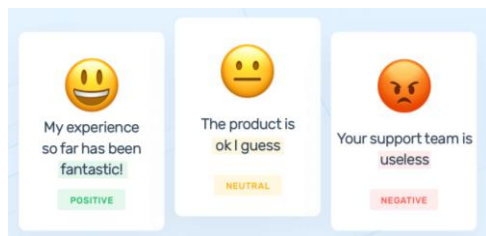
is inspiring and very impressive which obtained within 10+ years. As the main purpose of this platform is to spread information at faster rate irrespective of seriousness of that information.

Actually, the short posts which are posted by users are known as tweets. These posts can be any form such as text messages, videos, links, photos. All users can have freedom of choosing what they want to see on Twitter platform by simply following other users or companies or any searching topics. Usually, the Twitter timeline reflects the users preferences, but the re-tweets can be seen to the followers of a particular person, these can be paid advertisements also. However, there are chances of having fake accounts also, for this reasons Twitter platform created the verified account symbol, which tells that the account belongs to the particular high-profile person or a company. If this symbol has to be achieved, Twitters has to confirm whether identity of a person is true or not. Due to this process Twitters gains trust from people and try to maintain it with users. To create a Twitter account a person should have to provide their credentials such as their email address and needed to be verified.

Twitter decides what is trending at that present moment and try to filter out with person's interests, locations and preferences based on an algorithm and users' preferences, locations and interests. It works based on an algorithm. This platform has a search bar which enables user to search a particular topic. Twitter has a special function known as #Explore function. It helps in searching for trending topics. When a post is tweeted by a user on their profile, those posts will appear in the feeds of that persons' followers. These tweets can include news, random thoughts and can be sharing articles. But there is a restriction on length of the message. At the start of the Twitter platform, it has limited characters of tweets to 140. Now, this limit was increased to 280 characters, which may contain spaces and punctuation. The hash tag, which is unique feature for twitter platform acts like a meta tag, used to express as #keyword. This helps the user can searchable the tweet under that keyword.

II. EMOTION ANALYSIS OF TWEETS

This project Emotion analysis of tweets is the process which helps in identifying and analyzing the type of emotions which were expressed in the provided tweets. This analysis helps in extracting the textual information from the available sources to have detailed analysis of the main subjective information and try to understand the various emotions expressed behind the textual information. The emotions expressed by a person in his posted tweets can be identified and analyzed with the help of advanced machine learning techniques. Recently, emotional marketing is the key factor of success for many companies, especially companies which are known to be global ones. Because emotional marketing works with basic goal of convincing customers. This marketing strategy involves a start with an in-depth understanding of customers' emotional queries. This opinion analysis consists in "classifying a piece of text into positive versus negative classes," while emotion analysis consists in "multi classifying" text into different types of categories (anger, disgust, fear, happiness, sadness, surprise, etc.)



1. INTRODUCTION TO PROBLEM

Though many software techniques available to extract information about a particular person's emotional opinion on a product through the feedback provided by them, still many organizations and data analysts facing issues for extracting the data from the provided feedback by people. If we want to detect or find a particular emotion is holding for a particular tweet it is very easy for a tweet but with the increasing graph in the World Wide Web usage, public are habituated to use social-media platforms such as Twitter and other platforms results in generating large volumes of data and various people opinions in form of tweets or comments then in this case it will be very difficult to know a particular emotion that a particular tweet holding. This problem requires an organized manner solution.

So, our project converts this huge amount of data of tweets into organized manner that a machine can easily understand by using Natural Language Processing techniques and finally we are going to develop a machine learning model with best accuracy which is useful for calculating emotions of tweets.

2. TOOLS AND ALGORITHMS

A) Natural Language Processing(NLP)

NLP methods mainly depend on machine learning and particularly realization which uses a general learning calculation with an understandable example, a collection of

information to get familiar with the standards.

B) Support Vector Machine

Support vector machine is a directed learning framework and utilized for characterization and relapse issues. Support vector machine is very preferred by a lot of people as it produces prominent rightness with less calculation power. It is generally utilized in order issues. As we all know that there are mainly three kinds of machine learning algorithms. They are supervised algorithm, unsupervised algorithm and reinforcement learning algorithm. This Support Vector machine algorithm comes under supervised machine learning model. A support vector machine is a particular classifier officially characterized by partitioning the hyperplane. The aim of this algorithm is it tries to fit the hyperplane between different classes that maximizes the distance from hyperplane line to points of classes.

C) Random Forest

In our project we use another machine learning technique is random forest. Random forest algorithm follows voting procedure. Multiple decision trees combine to form tree type structure and this is known as Random Forest. The main aim of this technique is to predict which type of data gets highest voting. The highest voted data is shown to us as an output. It uses outfit realizing, which is a strategy that joins numerous classifiers to give answers for complex issues. This calculation comprises of numerous choices of decision trees.

This calculation lays out the result in view of the forecasts of the choice of decision trees. It predicts by taking the normal or mean of the result from different trees. Expanding the quantity of decision trees builds the accuracy of the result in efficient manner.

III. CASE STUDY

The main aim of our project is to find the emotion of particular tweet and then we develop the best suitable machine learning algorithms test the module based on the accuracy. The flowchart is as follows:

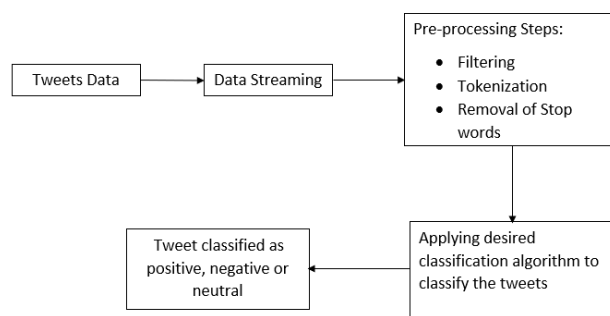


Figure.1 Flowchart for classification of tweets

1. Collection of Data

Our project requires huge amount of data consists of tweets. It will be difficult for us to collect or gather data consists of tweets through manual work. It may be found as time consuming process and also it may be found as collection of data by our own is treated as not that much important in this present generation. As the present world have lots of technologies which are developing or improving day by day it

doesn't make sense to collect data by our own waste our time on it. So to make our work easier this data is taken from available online resources and tried to observe the above few tweets corresponding to each labels, it is safe to conclude that label 0 is for non-offensive tweets and label 1 is for offensive tweets. After observing data we applied data preprocessing and cleaning techniques used in NLP. This data is raw data and it not in organized manner. We apply some data filtering techniques to delete the unnecessary content from our data in the following steps.

2. Data Streaming

It is the process making data in form of organized manner which means in this process the derived words or sentences would diminish to their root forms. For example, a sentence having a word giving now this stemmer will reduce the phrases "give", "given", "giving" to root word "give". The main benefit of stemming is that it makes examination between the words which are less complex because we don't have to manage complex syntactic changes of the word. For our project we make use of Porter Stemmer algorithm on our data consists of tweets. In this data streaming processing of data takes place. This processing of data includes tokenization which is the method involved with dividing tweets into individual words which are known as tokens. These tokens are partitioned when a sentence contains a whitespace or any punctuations. For example, the tweet is like "Such an amazing climax!!" is split into different words from each whitespace as follows:

```
{
Such
an
amazing
climax

!!
}
```

3. Filtering of Data

The tweets obtained after processing of data still the data contains information which is not in an organized manner or it may contain noise in the data i.e raw data. This data found to be not useful to our project. So to remove the unnecessary information from our tweets our data need to be operated by further more improving techniques to obtain the data in desired and simple manner by processing it. In this filtering process the tweets are filtered for the removal of noise in it by using techniques of removing of stop words, characters which does not make any value to our data and any other punctuations, numbers which are not necessary.

This filtering process contains removal of any special characters, words which does not add any value to sentence etc.

Stop words: After tokenization and stemming processes tweets may contain some stop words which looks like incredibly

familiar words like "the", "is", "a", etc and these words does not hold any extra data. So these words are treated as unimportant and finalize them as they are not serving any purpose and this component is carries out utilizing a rundown put away in stopfile.dat. Then we contrast each and every word in tweets and this rundown can erase the words which are matching the list of words which does not add any value to a tweet.

Removing non-alphabetical characters: The symbols for example "!#^&*\$%@" and numbers which does not holds any kind of information are treated as useless and we try to remove these symbols from our data so it makes our work easier to perform further more operations on data to find emotion of tweets. Ordinary expressions like we use in our daily life and add some importance to the sentence due to their presence will be remained without removing and remaining characters or numbers or symbols are removed. This assists with diminishing the messiness in our tweets data.

Stemming: It is the process making data in form of organized manner which means in this process the derived words or sentences would diminish to their root forms. For example, a sentence having a word "fishing" now this stemmer will reduce the phrases "fishes", "fishing" to root word "fish". The main benefit of stemming is that it makes examination between the words which are less complex because we don't have to manage complex syntactic changes of the word. For our project we make use of Porter Stemmer algorithm on our data consists of tweets. In this data streaming processing of data takes place

4. Applying Algorithms

After analyzing data we will implement some machine learning algorithms such as Support Vector Machines (SVM) and Random Forest algorithms. By using these techniques we would attempt to fit our model in the best suitable algorithm. This best algorithm can be identified by observing F1 score of each algorithm. We will come to conclusion by observing the maximum F1 score of a particular algorithm means if an algorithm has higher F1 score compared to another algorithm, we can say that algorithm provide very accurate result.

The data is in need of adding labels to unlabeled information for the detailed study of emotion of tweets, this can be done by using "vader lexicon" model. This model treated as one of the best approaches for the detailed analysis of emotions of tweets. The vader lexicon can be accessed using Natural Language Toolkit library in python. The necessary libraries in python and the unlabeled data which is in need of adding labels imported to our working environment. Since our taken dataset contains two columns, we try to add labels for the unlabeled data. After adding labels to our tweet information then we try to add four more new columns to the labeled data. These four columns are "Positive", "Negative", "Neutral" and "Compound" by ascertaining sentimental scores of particular tweet. The below figure shows the code for adding 4 columns of Positive, Negative, Neutral and Compound to the dataset.

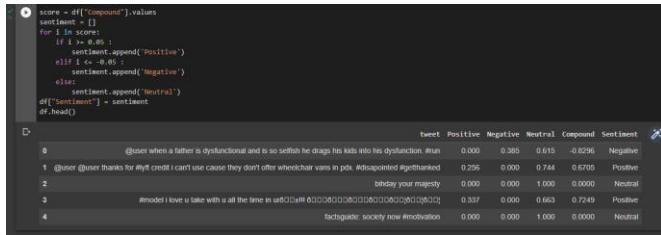


Figure. 2 Code snippet for calculating Compound score

As indicated by the business principles, in the event that the compound score of emotion of a sentence is more than 0.05, it is treated as Good, and if the score is lesser than -0.05, it is classified as Bad, if none of the above cases are satisfied then, it's unbiased. With the help of above information, another section is added in our labelled dataset which consists of emotion scores. The below figure shows classification of emotion of tweet by comparing compound score.



Figure. 3 Code snippet to classify emotion of tweet

From the above values we will draw a pie chart to have a better visualization of data. The below figure shows the count of “Positive tweets”, “Negative tweets” and “Neutral tweets” present in the given dataset.

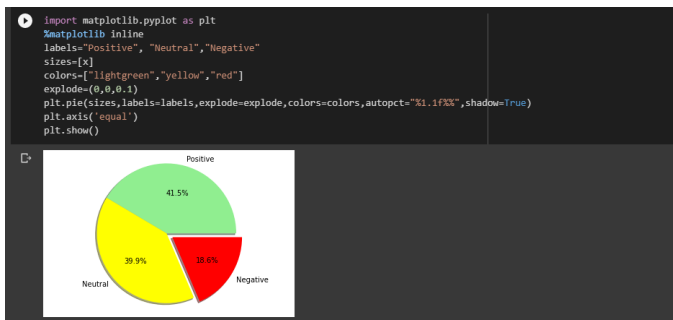


Figure. 4 Code snippet for pie chart visualization of count of tweets based on emotion

IV. RESULTS – MODELS ACCURACY

After cleaning and clear analysis of data, we will implement machine learning algorithms such as Logistic Regression, Support vector machine, Random Forest etc. After implementing algorithms, we will try to find out F1 score of each model to check which model is giving best accuracy

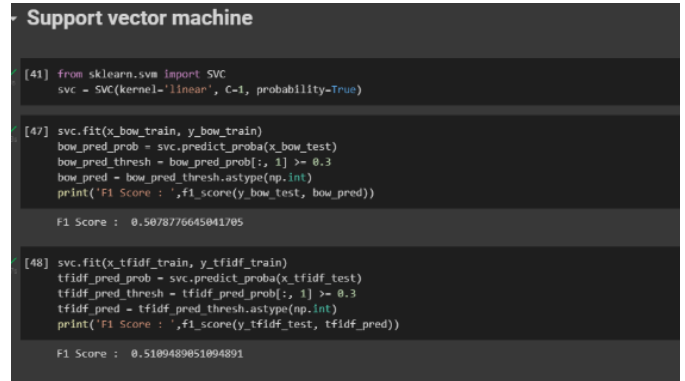


Figure.5 Code snippet for finding SVM algorithm accuracy.

Next, we will implement Random Forest model and we will find F1 score of this model. The below figure explains the code for Random Forest model and finding the F1 score of this model.

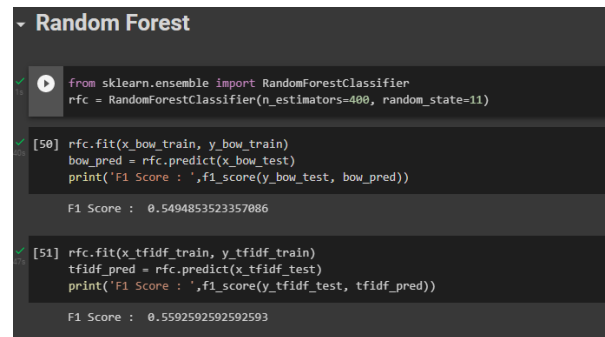


Figure.6 Code snippet for finding Random Forest model accuracy

From the above code we observe that F1 score of Random Forest algorithm is 0.55 which is near to SVM model. So we can conclude saying that this two works with the same accuracy and we can use any of these two models as F1 score is same.

V. CONCLUSION

One of the wellspring platform like twitter consists of huge amount information in which some of the information is treated as important and some of the information may contain noisy and unwanted information. For our project this unwanted information may lead machine learning algorithms to work with less efficiency. So to eliminate this extra information which does not add or create any importance to the sentences, we use few data filtering techniques. These filtering methods are also used in providing data which we want and remove unnecessary information. Generally, tweets represent feedback or person's individual opinion on particular thing or an event. These tweets may be offensive or non-offensive, good or bad. Irrespective of it we took the whole tweet dataset as input and try to detect emotion of a particular sentence. This can be done by using Natural language toolkit. After cleaning the data, we develop machine learning algorithms such as Support Vector Machine and Random Forest. The best model fit is calculated by finding F1 score of these two models and based on the F1 scores we took maximum value of F1 score algorithm and we treat that algorithm as our main priority model. By doing this project we have

learnt how to find and calculate a particular tweet emotional analysis score for a large dataset within few minutes and divide tweets respective to their condition of feeling of a person in an organized manner.

REFERENCES

- [1] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30-38
- [2] Pang, B. and Lee, L. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278.
- [3] Neethu M, S and Rajashree R, "Sentiment Analysis in Twitter using Machine Learning Techniques" 4th ICCCNT 2013, at Tiruchengode, India. IEEE – 31661
- [4] A. Pak, and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Special Issue of International Journal of Computer Application, France: Universitede Paris-Sud, 2010
- [5] Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacharjee, S. K., & Viswanathan, A. (2015, October). Sentiment analysis of Facebook data using Hadoop based open source technologies. In Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on (pp. 1-3). IEEE.
- [6] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, "Election result prediction using Twitter sentiment analysis," in Inventive Computation Technologies (ICICT), International Conference on, 2016, vol. 1, pp. 1–5.
- [7] V. Sahayak, V. Shete, and A. Pathan, "Sentiment analysis on twitter data," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, no. 1, pp. 178–183, 2015.
- [8] Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, 2010, pp. 1-15. Germany: Springer, 2010, pp. 1-15.
- [9] Dmitry Davidov, Ari Rappoport. "Enhanced Sentiment Learning Using Twitter Hashtags and Smiles". Coling 2010: Poster Volume pages 241{249, Beijing, August 2010.
- [10] V. M. K. Peddinti and P. Chintalapoodi, "Domain adaptation in sentiment analysis of twitter," in Analyzing Microtext Workshop, AAAI, 2011.
- [11] Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541.
- [12] Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. In Recent developments in intelligent information and database systems (pp. 263-276). Springer International Publishing.
- [13] G. P. Kumar, G. S. Priya, M. Dileep, B. E. Raju, A. R. Shaik and K. V. S. H. G. Sarman, "Image Deconvolution using Deep Learning-based Adam Optimizer," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 901-904, doi: 10.1109/ICECA55336.2022.10009073.
- [14] Sarman, K. G., Madhu, T., & Prasad, M. (2022). Fault Diagnosis in the Brushless Direct Current Drive Using Hybrid Machine Learning Models. ECTI Transactions on Electrical Engineering, Electronics, and Communications, 20(3), 414-426. <https://doi.org/10.37936/ecti-ec.202203.247517>
- [15] N. Gupta, H. K. Gupta, V. Pathak and P. Pareek, "Implementation of Deep Learning Based Compression Technique and Comparative Analysis With Conventional Methodologies," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984363.
- [16] A. Jain, K. K. Ramachandran, S. Sharma, T. Sharma, P. Pareek and B. Pant, "Detailed Investigation of Influence of Machine learning (ML) and Big Data on Digital Transformation in Marketing," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 1061-1065, doi: 10.1109/ICACITE53722.2022.9823810.
- [17] Padmanandam, K., M. V., R., N. Upadhyaya, A., Chandra, K. R., B., C., & Sah, S. (2022). Artificial Intelligence Biosensing System on Hand Gesture Recognition for the Hearing Impaired. International Journal of Operations Research and Information Systems (IJORIS), 13(2), 1-13. <http://doi.org/10.4018/IJORIS.306194>
- [18] Srinivas, C. V. V. S., Gubbala, S., Lakshmi, N. D. N. (2022). Human Emotion Classification Based on Speech Enhancement Using Neural Networks. In: Gupta, D., Sambyo, K., Prasad, M., Agarwal, S. (eds) Advanced Machine Intelligence and Signal Processing. Lecture Notes in Electrical Engineering, vol 858. Springer, Singapore. https://doi.org/10.1007/978-981-19-0840-8_3
- [19] Mulagala Sandhya, Ujas Rudani, Shubham Rathod, Dilip Kumar Vallabhadas, Mulagala Dileep, Sriramulu Bojjagani, Sravya Pallantla and P. d. s. s Lakshmi Kumari, Deep Neural Networks with Multi-Class SVM for Recognition of Cross-Spectral Iris Images, Symposium on Machine Learning and Metaheuristics Algorithms, and Applicati (SoMMA'20), pp.29-41, Springer, https://doi.org/10.1007/978-981-16-0419-5_3, 2021
- [20] G. P. Kumar, M. Dileep, B. V. V. Satyanarayana and S. P. Manikanta, "Modified Channel Estimation Techniques for DCT based OFDM Systems," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-6, doi: 10.1109/ICCCI54379.2022.9740842.
- [21] Saride Jagan Mohan Rao, Piyush C. Dalsania, Sudharani Chidurala, Ch Murali Krishna, Puneet Narayan, D. Durga Prasad, Fractal segmented lotus shape planar monopole antenna for multiband applications, Materials Today: Proceedings, Volume 66, Part 8, 2022, 3450-3456, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2022.06.145>.
- [22] C. M. Krishna, N. V. K. Maha Lakshmi, N. Suguna, J. P. Kumar, E. J. Satapathy and D. D. Prasad, "Design of Wheel shaped Millimetre Wave MIMO Antenna for N257band (26.5 – 29.5GHz) Applications," 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 2022, pp. 1-6, doi: 10.1109/ICACCM56405.2022.10009320.
- [23] C. M. Krishna, D. D. Prasad, C. R. Krishna, R. Orugu, P. K. K. Varma and J. P. Kumar, "Design and Analysis of Tortoise Shaped Quadband Antenna For C & Ku – Band Applications," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402698.
- [24] C. M. Krishna, D. D. Prasad, C. R. Krishna and M. K. V. Subbareddy, "Design and Analysis of Bow-Tie Antenna for Sub — 6GHz Applications," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402455.
- [25] G. Shine Let, G. Josemin Bala, C. Benin Pratap, Doondi Kumar Janapala & Talluri Kezia, Concentric semi-circular ring loaded multi-band antenna for wireless applications. Analog Integrated Circuits and Signal Processing (2022). <https://doi.org/10.1007/s10470-022-0212>

