



SENTIMENT CLASSIFICATION AND OPINION MINING ON AIRLINE REVIEWS

Sanjana Rajesh Revankar¹, C V Yashashwini², Spandana K³, N S Shravani⁴, Neha S⁵, Raghu B R⁶
^{1,2,3,4,5,6}Computer Science and Engineering

Bapuji Institute of Engineering and Technology, Davangere, India

sanjanarrdvg@gmail.com¹, yashaswinicv2000@gmail.com², spandanak96@gmail.com³
nsathyashravani122@gmail.com⁴, nehasatish15@gmail.com⁵, raghubr585@gmail.com⁶

Abstract—Opinion Mining (OM) or Sentiment Analysis (SA) can be characterized as the task of recognizing, extracting and ordering the opinions on something. It is a kind of processing of the natural language (NLP) to follow the public state of mind to a specific law, marketing, or policy, etc. It includes a way that development for the assortment and assessment of the comments and opinions about enactment, policies, laws, and so on, which are posted on online media. Air travel is additionally one of these intriguing issues that are generally spread on Twitter. Air travellers generally share their travel experience on Twitter. This data can be useful if evaluated using machine learning techniques and can provide acuteness that is useful to understand the comfort level of the traveller in the flight. The current techniques for sentiment analysis consist of machine learning (supervised and unsupervised), and lexical-based approaches. Consequently, the main goal of this project presents study of sentiment analysis (SA) and opinion mining (OM) approaches, various techniques utilized that related in this field.

Keywords—NLP, Airline Reviews, Sentiment Analysis, Tweets, Opinion Mining.

I. INTRODUCTION

Nowadays people check for the review of the airline before they go for the boarding process so that they could get know about the quality of service like food quality, flight delay, cabin crew behaviour, Loss of baggage, seat comfort, airfare etc. in the airline industry. Manually analysing and classifying these reviews takes more time. In this paper we are going to discuss about how efficiently the reviews of the

passengers are taken and analysed based on the sentiments of the passengers i.e. POSITIVE review, NEGATIVE review and NEUTRAL review. So to classify these model based on these three categories i.e. positive, negative, neutral we are using ARTIFICIAL NEURAL NETWORK Where neural network possess any number of neurons in each layer can have one or more neurons or units. Each of the neurons is interconnected with each and every other neuron. As ARTIFICIAL NEURAL NETWORK Accept only the numeric and structured data as input to train the model. However, some passengers will refuse to give feedback .so the most convenient way for the passenger to share their review is the social media which provides a platform where user can freely express his/her review during flight. Twitter is one of the most popular platforms worldwide.so from these tweets we frame the dataset and train the model based on ANN.

II. RELATED WORK

A. Artificial Neural networks (ANN)

An Artificial Neural Network is an information processing paradigm that is inspired from the brain. The brain is the fundamental part in the human body. The fundamental unit of the brain is the neuron.It is the biological neural network which receives the input inform of signals and processes it and send out the output signals. Similar to the brain the Artificial Neural Network, imitates the biological network of human body.ANN is a deep learning algorithm that emerged & evolved from the idea of biological human brain. The ANN formed from the artificial neurons made up of Silicon and wires which imitates the neurons and the interconnection which are formed from coefficients(weights). Synaptic weights are

inter-neuron connection strengths which stores the knowledge learned by ANNs. The ANN is strongly interconnected with each other to make solutions for specific problems and complications and it is highly useful in the sectors of Pattern recognition, data classification, clustering etc. They can find hidden patterns and information from complicated sources which are much difficult for the machines and human. The major Artificial Neural Network Applications are:

Image Processing and Character recognition: ANN's are playing a big role in Image Processing and Character recognition. Image recognition is an ever growing field with widespread applications from facial recognition in social media, cancer detection in medicine and defence usage.

Stock Exchange Prediction – Stock market investments and rise and dip in markets are complicated. The rise and dip depends on various factors any given day. Thus neural networks can examine a lot of information in a fast manner and sort it all out. So we can use them to predict stock prices.

Travelling salesman problem - Neural network can also solve the travelling salesman problem. But this is to a certain degree of approximation only.

Handwriting Recognition – The idea of Handwriting Recognition has become very important. This is because handheld devices like palm pilot are becoming very popular. Hence we can use neural networks to recognize handwritten characters.

Image Compression – Vast amounts of information is received and processed at once by neural networks. This makes them useful in image compression. With the internet explosion and more images on their sites, using neural networks for Image Compression is worth a look.

B. Airline Industry and Sentiment Analysis

Airlines are easier, safer and more profitable mode of transportation in recent times than any time in history, but the industry must innovate much more rapidly in order to secure its environmental and financial viability & Safety and Convenience in the future. Since, cost of tickets is high everyone can't afford to travel in flights. There are advantages and disadvantages of travelling in flights.

Advantages:

- We can reach to our destination within mean time compare to any other transport.
- Easy to carry people and goods to different areas.
- Reduces traffic.

Disadvantages:

- High cost
- Carries a smaller number of people.
- More risks due to hijacks, breakdowns and accidents.
- Large amount of investment.

No doubt that the Airline industry is a million and billion-dollar industry and large number of passengers use this service daily for transportation from place to place. Increasing competition has made different to a lot of start-ups as well as well-established airline companies to survive in the current market. These days a great challenge for these companies is to build customer loyalty and it is also the major concern among companies. For these reasons, it is very essential to provide good service, value passengers' feedbacks and satisfy their needs. In the past, reviews from passenger were read manually, analyzed and categorized into their specific sentiment which in turn consumed a lot of time.

Sentiment analysis or opinion mining refers to the use of natural language processing (NLP), text analysis and computational linguistics to identify and extract subjective information from the source materials. Business enterprises can leverage opinion polarity and sentiment topic recognition to gain a deeper understanding of the drivers and the overall scope. Sentiment analysis is the task of identifying positive and negative opinions, emotions, and evaluations and deriving actionable insights. These insights become useful in planning and execution of customer service initiatives, media relationships etc. The marketing and sales teams can also devise their strategies based on these insights.

III. METHEDODOLOGY

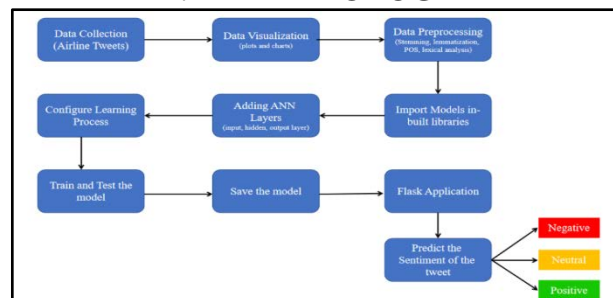


Fig. 1. Block Diagram

One of the best proposed solutions is: In order to provide a more accurate picture of the user's opinions, it is simply not enough to classify them by polarity. More in-depth knowledge of their expression is needed, thus in the projects, you need to aim to extract human emotions conveyed in the tweet with the help of existing emotion models from the domain of psychology. We deployed a Sentiment Analyzer for Airline Tweets using Artificial Neural Networks. This model can detect positive, negative as well as neutral tweets on respective Airlines.

A. Importing libraries and Data Collection

We first imported libraries like pandas for data visualization, numpy for dealing with arrays, matplotlib and seaborn for plotting the graphs. The dataset contains data of tweets on various airline's twitter handles. It contains a total of 15 columns in which 3 columns have been dropped, out of which one column specifies the sentiment of the tweet. All other columns provide various information related to what was the tweet, where was it posted from, when was it posted etc. We read the dataset using pandas. The dataset we considered contains around 14.5k samples with 15 columns.

B. Data Visualization

We checked for null values and dropped the unwanted columns. Splitted the dataset for plotting scatter plot to know the distribution of neutral, positive and negative tweets. We plotted count plot to count the total tweets in each category. Then we plotted some other plots like box, violin, pairplots. By plotting factor or cat plot we visualized the total count of positive,neutral and negative tweet count in all the airline mentioned and also the frequency of tweets like bad flight, can't tell etc. We used pie-chart to know the distribution of negative, neutral and positive tweets over all. Using word count, we displayed the most used words in both tweets and negative sentiment. At last, we analyzed the kernel density by plotting bivariate distribution between negativereason_confidence v/s airline_sentiment_confidence.

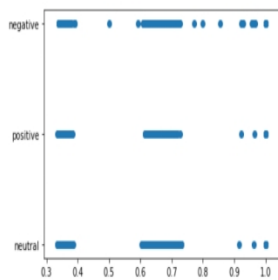


Fig. 2. Scatter Plot

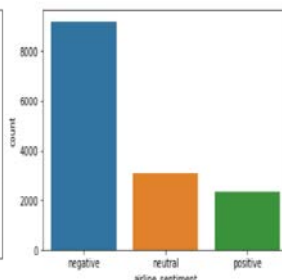


Fig. 3. Count Plot

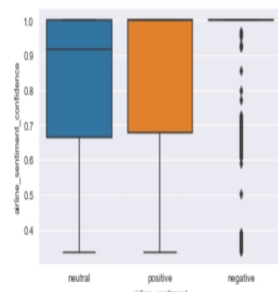


Fig. 4. Box Plot

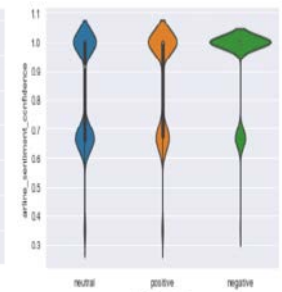


Fig. 5. Violin Plot

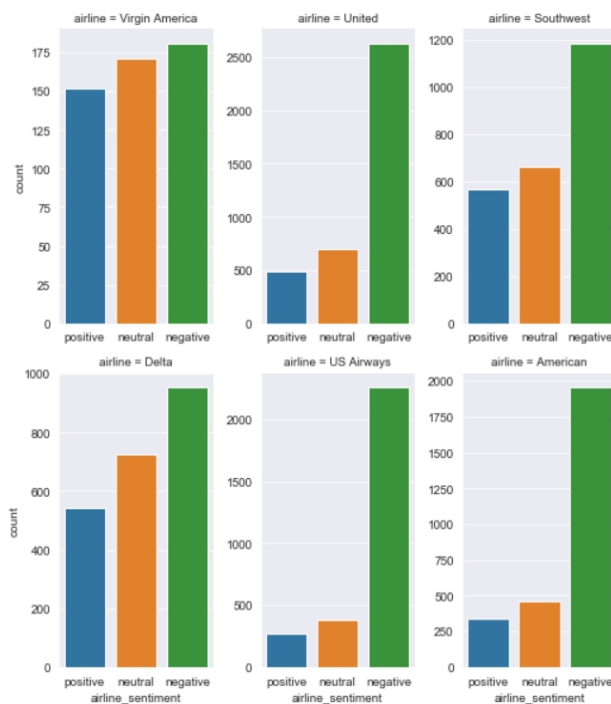


Fig. 6. Factor Plot

C. Data Preprocessing

Import necessary libraries for nlp tasks like stopwords for removing words like a the etc, porterstemmer for stemming. We then replace negative with -1, positive with 1 and neutral string with 0. Then we dropped unwanted columns and kept only text that is tweets and its corresponding sentiment. We then tokenize, apply lemmatization on the tweet. Then we changed the case of the tweet to lower case and removed the stopwords. Then transform the data into array using CountVectorizer and save that file for application. We label encoded the sentiment column.

D. Model Building

Splitting the dataset into train and test in ratio 80% and 20%. We initialized the model, added the input, hidden and output layers then compiled the model by optimizing it using adam optimizer and sparse_categorical_crossentropy as our loss function and used Accuracy as the matrices to get the accuracy of the model. After compiling, we trained the model on 70 epochs with batch size as 10. Our model can successfully predict negative, positive and neutral sentiment in Airline tweets.

E. Prediction

Given a message, our system classifies whether the message is positive, negative, or neutral sentiment. The training data was provided by the task organizers (No other tweet data are used). We have created our model using Artificial Neural Networks and Natural Language Processing. From the given data we “text” as the input variable and “airline_sentiment” as the output variable. Depending on the text we give our model predicts the sentiment.

IV. RESULTS AND DISCUSSION

We trained and tested our algorithms on the complete data set to start with. Later we randomly separated the data set into training data and test data so that we had samples from each class. 80% of data is used for training data and 20% is used for test data. The dataset consists of 14640 reviews on different airline companies, and was used as a development set for ANN. The model was able to classify the sentiment with about 96% accuracy with 71% test accuracy. The following images show the screenshots of our application of Sentiment Classification and Opinion Mining on Airline.



Fig. 7. Home Page



Fig. 8. Output of Positive Review



Fig. 9. Output for Negative Review



Fig. 10. Output for Neural Review

Some of the advantages and disadvantages of our system are:

Sentiment analysis is a useful tool for any organization or group for which public sentiment or attitude towards them is important for their success. Our web-app allows you to keep track of what's being said about your service during travelling and can help you detect angry customers or negative mentions before they turn into a major crisis. This app will predict the type of tweet more accurately.

The human language can be complex for machine-based learning systems to interpret. For example, opinions can be expressed with sarcasm or irony, and the order of words can add even more confusion.

For example: 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase “I

wasn't disappointed", it should be classified as positive. "It is not so bad" this statement is actually a positive tweet but the words in it will be.

V. CONCLUSION

The ANN which is short for Artificial neural Networks, is the very useful model, and it can be applied in problem-solving and machine learning. Neural Networks can be very beneficial to this computing world. Their ability to learn by example makes them very flexible and powerful. To get best out of ANNs, it is important to understand their strengths as well as their weaknesses. In this project, we tried to show the basic way of classifying tweets into positive, negative or neutral category using ANN as baseline and how language models are related to the ANN and can produce better results. Our algorithm achieves a good accuracy. To make the model give best results, we can try adding more hidden layers, training with larger datasets and increasing the number of epochs for better learning of the model. NLP performs best and yields good accuracy in sentiment analysis task. We also integrated our deep learning model with a user-friendly UI which was built using flask web form and it was successfully able to identify the sentiment hidden in the airline tweets. Airline companies can make use of this model in order to understand the feedback of their customers and improvise themselves.

REFERENCES

- [1] Dataset - <https://www.kaggle.com/welkin10/airline-sentiment>
- [2] Wang, Lei, Jianwei Niu, and Shui Yu. "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis." *IEEE Transactions on Knowledge and Data Engineering* (2019).
- [3] Joshi, Sunil. "Sentiment Analysis on WhatsApp Group Chat Using R. (2019) " *Data, Engineering and Applications. Springer*, Singapore, 47.
- [4] Ohana, B., & Tierney, B. (2009). "Sentiment classification of reviews using SentiWordNet". *In 9th. it & t conference*, 13: 18-30.K. Elissa, "Title of paper if known,"
- [5] Younis, Eman MG. (2015) "Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study." *International Journal of Computer Applications*, 112(5): 0975 – 8887.
- [6] Mishra, Prerna, Ranjana Rajnish, and Pankaj Kumar. (2016) "Sentiment analysis of Twitter data: Case study on digital India." *International Conference on Information Technology (InCITE)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds. IEEE*, 148 – 153.
- [7] Kamal, Rashid, et al. (2019) "Emotion Classification and Crowd Source Sensing; A Lexicon based Approach." *IEEE Access*, 27124-27134.
- [8] Geetha, R., Pasupuleti Rekha, and S. Karthika. (2018) "Twitter Opinion Mining and Boosting Using Sentiment Analysis." *International Conference on Computer, Communication, and Signal Processing (ICCCSP). IEEE*, 1 – 4