Tweet Classifier: Advancements in Multi-Label Analysis

Swastik Anupam

Amity University Kolkata, Newtown, Kadampukur, Kolkata, West Bengal, 700135

Abstract

Tweet Classifier is my submitted work to AISoMe FIRE 2023. In this research, I propose a text classification model for multi-label classification tasks using a domain-specific model to classify tweets as Unnecessary, Mandatory, Pharma, Conspiracy, Political, Country, Rushed, Ingredients, Side-effect,Ineffective and Religious.The vaccination process was on- going worldwide to fight against the novel coronavirus disease(COVID-19), and the sentiment analysis of tweets is expected to provide helpful insights regarding the stance of people about the vaccines.I employed a deep neural network architecture implemented using TensorFlow, with TF-IDF vectorization as a feature engineering technique.The model is trained on a labeled dataset and evaluated on a test dataset, achieving competitive macro F1 scores.This approach provides a robust framework for automated text classification tasks. The evaluation score of our submitted run is reported in terms of accuracy and macro-F1 score.We achieved an accuracy of 0.4975, a macro-F1 score of 0.25, the 41th rank among other submissions.

Keywords

Sentiment Analysis, COVID-19 Vaccine Tweets, Tweet Analysis, Text Analysis

1. Introduction

Amidst the relentless battle against the COVID-19 pandemic, vaccines have emerged as a crucial lifeline, with their proven safety and effectiveness in combating infectious diseases.The rapid development and distribution of COVID-19 vaccines have ignited a global conversation on platforms like Twitter.These discussions span vaccine progress,accessibility, efficacy, and potential side effects, reflecting a spectrum of public opinions.This research introduces a cutting-edge machine learning model designed to classify COVID-19 vaccine-related tweets on Twitter.Manual classification of these tweets is both time-consuming and error-prone, necessitating an automated solution. Leveraging the latest advancements in Natural Language Processing (NLP) and deep learning techniques, my model demonstrates exceptional accuracy in categorizing tweets into relevant classes. This model aligns with Twitter's ongoing efforts to combat vaccine-related misinformation, contributing to the identification and management of vaccine discussions. By providing a robust tool for monitoring public sentiment regarding COVID-19 vaccines, my research empowers health authorities to make informed decisions and combat the infodemic

Forum for Information Retrieval Evaluation, December 15-18, 2023, India
swastik.anupam@s.amity.edu (S. Anupam)
https://github.com/SwastikAnupam (S. Anupam)
0009-0002-5368-9027 (S. Anupam)

surrounding the pandemic effectively. This work underscores the potential of machine learning in addressing real-world challenges, especially within the context of a global health crisis.

2. Task

For task, "Building an effective multi label classifier to label a social media post (tweets) according to the specific concern(s) towards vaccines".
Note: a tweet can have more than one label (concern).
Our objective is to construct a robust multi-label classifier for social media posts, specifically tweets, aimed at categorizing them based on the distinct concerns expressed by authors regarding vaccines.
It is important to note that a single tweet can encompass multiple concerns, necessitating a multi- label approach. The concerns we consider as labels for classification encompass:
1. **Unnecessary**: Tweets suggesting vaccines are unnecessary or that alternative remedies are superior. *Example: "Why bother with vaccines when natural immunity is better?"*
2. **Mandatory**: Tweets opposing mandatory vaccination, asserting that vaccines should not be enforced. *Example: "Vaccination should always be a choice, never mandatory."*
3. **Pharma**: Tweets criticizing Big Pharmaceutical companies, alleging profit-driven motives, or expressing general distrust based on their history.*Example: "Big Pharma profits while we suffer."*
4. **Conspiracy**: Tweets delving into deeper conspiracies related to vaccines, extending beyond financial motivations (e.g., tracking people, COVID-19 being a hoax).*Example:"Vaccines are a tool for population control."*
5. **Political**: Tweets voicing concerns that governments or politicians are advancing their agendas through vaccines.*Example: "Politicians are exploiting vaccines for their own gain."*
6. **Country**: Tweets expressing objections to vaccines based on their country of origin.*Example:"I won't trust a vaccine made in that country."*
7. **Rushed**: Tweets expressing concerns about insufficient testing or inaccurate published data regarding vaccines.*Example:"These vaccines were rushed and not properly tested."*
8. **Ingredients**: Tweets raising concerns about vaccine ingredients (e.g., fetal cells, chemicals) or the technology used (e.g., mRNA vaccines).*Example:"I'm worried about what's in these vaccines."*
9. **Side-effect**: Tweets expressing concerns about vaccine side effects, including fatalities.*Example:"Too many people are experiencing severe side effects."*
10. **Ineffective**: Tweets doubting vaccine efficacy, asserting they are not effective and, thus, useless.*Example:"These vaccines don't work as advertised."*
11. **Religious**: Tweets opposing vaccines on religious grounds. *Example:"My faith prohibits me from getting vaccinated."*
12. **None**: Tweets with no specific reason stated or citing other reasons not covered above.*Example:"I haven't decided yet if I want to get vaccinated."*

3. Related Work

Users post content on microblogs like twitter for various purposes, including their sentiments about vaccines and vaccination drives.Data extraction from these textual tweets is very popular part of sentiment analysis. The traditional machine learning methods like Naive-Bayes classifier, Linear classifier,Support Vector Machines and Deep neural methods like Long Short-Term Memory (LSTMs) and Bidirectional RNN are very successful for text classification.More recent language models for natural language processing includes XGBoost models, KNN,KLNext, BERT (Bidirectional Encoder Representations from Transformers) and its domain-specific version CT BERT(COVIDTwitterBERT),TensorFlow(TF-IDF).The papers which i have used for citations are related to this research using tensorflow and deep neural network to classify text.All the citations which I have taken proved very informative as they provided the base for my research by giving info about the TD-IDF and other techniques with precision [1], [2], [3], [4].The overview paper,a comprehensive study gives the idea of the growing importance and integration of AI in online social platforms [5].

3.1. TensorFlow (TF-IDF)

TensorFlow with TF-IDF (Term Frequency-Inverse Document Frequency) is an approach in which text data is converted into numerical TF-IDF features and then processed using TensorFlow, a prominent deep learning framework.In the context of multi-label text classification, this method employs TensorFlow to construct neural networks that utilize TF-IDF as the input.It seamlessly integrates the powerful text representation capabilities of TF-IDF with the modeling strengths of neural networks to classify text into multiple labels. This code exemplifies multi-label text classification utilizing TensorFlow and scikit-learn. It involves data preprocessing, conversion of text into TF-IDF features, construction of a neural network with three layers, model training, prediction generation, and evaluation through metrics such as the macro F1 score and a comprehensive classification report.

4. Dataset

The training dataset given comprises 9,921 tweets expressing concerns about COVID vaccines, posted during 2020-21.The dataset includes two essential components: tweet IDs and corresponding labels. My approach used the dataset taken from updated version on Arxiv["CAVES: A dataset to facilitate explainable classification and summarization of concerns towards COVID vaccines"] [6].I augmented the dataset with the given tweet, tweet ID and labels by using K-means clustering and DBSCAN (Density-Based clustering algorithm) to observe the various trends in the given dataset.

4.1. Trends in the dataset

Based on the given information, the following trends were observed in the dataset.The dataset contains tweets with various labels. Here are some key findings:
• The label "side-effect" is the most common, with 2,883 occurrences, representing approximately

29.06 percent of the dataset.
• The label "ineffective" appears 1,204 times, accounting for about 12.14 percent of the dataset.
• Labels like "rushed," "pharma," and "none" also have significant counts.
There are a total of 288 unique label combinations in the dataset. Some labels are combined (e.g., "side-effect pharma ingredients"),indicating that a single tweet might be associated with multiple themes or topics.The test data is annotated by human annotators, where a label is assigned on the unanimous agreement or majority agreement from the given labels.

5. Pre-processing

I pre-processed the tweets in order to improve the quality of text produced by TF-IDF.Tweets generally contains like HASHTAGS, HTTP-URL and EMOJIS which without pre-processing, often reduce the performance of the model.Thus, i used the following data cleaning tasks as part of pre-processing the tweets in the dataset:
• **Stop Words Removal**: A stop word is a commonly used word such as "the", "a", "an", "in", which do not provide any valuable information. I removed the stop words in order to give more focus to the important information.
• **Text Standardization**: Tweets are written more casually, thus by lower casing every word, i am keeping only a single version of every word, enhancing the text analysis.
• **Emoji Conversion to words**: Emojis are extensively used on Twitter to express feelings and emotions. Completely removing them removes a lot of sentiment information; thus, I converted the emojis to text and retained their meaning using 'emoji' library available.
• **Contraction Expansion in text**: In order to standardize the text, each contraction is converted to its expanded, original form.
• **Non-Alphanumeric Characters Removal**: To ensure completely refined textual data, I removed all the non-letter characters like brackets, colon, semi-colon, @, etc.
• **URL Elimination**: URLs are not sufficient for sentiment analysis; I removed them with the help of regular expression from the text.

6. Methodology

6.1. Model

Term Frequency Inverse Document Frequency-TensorFlow (TF-IDF): TensorFlow with TF-IDF (Term Frequency-Inverse Document Frequency) is an approach in which text data is converted into numerical TF-IDF features and then processed using TensorFlow, a prominent deep learning framework.
Representation: Used tool scikit-learn to transform textual data into TF-IDF vectors and this representation emphasizes the importance of words that are frequent in a specific dataset. This representation can be particularly useful for tasks like text data classification or clustering.
Neural Network with TensorFlow: After TF-IDF representations, I used them as input features to a neural network model built using TensorFlow. This model is designed for classification and

sentiment analysis.Pre-training: In Pre-training, I pre-processed the dataset before fine-tuning it on a specific task.
However, in the case of a TF-IDF representation, pre-training is not a standard practice since TF-IDF vectors are task-specific. We typically train our TensorFlow model on our dataset with TF-IDF vectors without any pre-training.

6.2. Experimental Setup

My experimental framework was constructed using TensorFlow and scikit-learn.I have used dataset comprising tweets and their respective labels and once loaded, preprocessed the tweets and transformed the labels into a multi-label binary format using a MultiLabelBinarizer.Further to convert my textual data into a numerical format suitable for a neural network, i employed TfidfVectorizer. The neural architecture chosen for this task was a sequential model consisting of two hidden layers, with the first layer housing 512 neurons and the subsequent containing 256 neurons using Keras.I added dropout layers with a rate of 0.7 to prevent overfitting and adopted a batch training approach which is processing the dataset in chunks of 64 samples, and the training spanned over 100 epochs for a maximum of 1,000 iterations.I have attached the repository link also which can be refereed to see the the experimental setup. [click this] link to go the repository for this research project.

7. Prediction

For the prediction over the given test data, I harnessed the power of TFIDF-Neural Net Multi-Label Classifier. Instead of the conventional embeddings like CT-BERT,my approach transformed each tweet into a rich TF-IDF representation,capturing the essence of the content.This data was then passed through my finely-calibrated neural network to ascertain probability scores against all classes. The classes with the highest probability emerged as the predicted classes for a respective tweet. My submission, a prediction file (in CSV format) with both the Tweet ID's and its corresponding class prediction, marked my entry for the FIRE Track 2023 task.

8. Evaluation

Task - AISOME FIRE 2023 Track Results: Evaluations for the AISOME FIRE 2023 Track were conducted using two primary metrics: the Jaccard index and the macro-F1 score, both applied to the specified classes. The outcome of my submitted run for the Track is detailed in Table. Swastik Anupam (individual) secured the 41st rank among the various submissions, achieving a Jaccard index of 0.29 and a macro-F1 score of 0.25.

Sr No.	Team_Name	Jacc	macro-F1 score	Rank
41	Swastik Anupam (Individual)	0.29	0.25	41

Table 1
Table 1: Result of AISOME FIRE Track 2023

9. Conclusion and Future Work

This paper uses TensorFlow (TF-IDF), is an approach in which text data is converted into numerical TF-IDF features and then processed using TensorFlow, a prominent deep learning framework. I observed that the TensorFlow (TF-IDF) outperformed the traditional natural language processing classifier, namely Naive Bayes and Support Vector Machines, as text computed by the TensorFlow (TF-IDF) are more expressive and yield better results on the given task. I further propose for improving the performance of my model based on the code and several potential enhancements in the areas of Embeddings Layer,Hyperparameter Tuning, Batch Normalization, Custom Loss Functions, Complex Model Architecture (Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) for handling sequential data like text).To further enhance the model's accuracy, adversarial training techniques can be applied.

References

[1] M. Neethu, R. Rajasree, Sentiment analysis in twitter using machine learning techniques, in: 2013 fourth international conference on computing, communications and networking technologies (ICCCNT), IEEE, 2013, pp. 1–5.

[2] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (2018) e1253.

[3] A. M. Ramadhani, H. S. Goo, Twitter sentiment analysis using deep learning methods, in: 2017 7th International annual engineering seminar (InAES), IEEE, 2017, pp. 1–4.

[4] P. Semberecki, H. Maciejewski, Deep learning methods for subject text classification of articles, in: 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2017, pp. 357–360.

[5] S. Poddar, M. Basu, K. Ghosh, S. Ghosh, Overview of the fire 2023 track:artificial intelligence on social media (aisome), in: Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, 2023.

[6] S. Poddar, A. M. Samad, R. Mukherjee, N. Ganguly, S. Ghosh, Caves: A dataset to facilitate explainable classification and summarization of concerns towards covid vaccines, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3154–3164.