HaleLab_NITK@SMM4H'24: Binary Classification of English Tweets reporting Children's Medical Disorders

Ritik Mahajan and Sowmya Kamath S.
Healthcare Analytics and Language Engineering (HALE) Lab,
Department of Information Technology,
National Institute of Technology Karnataka, Surathkal, Mangalore 575025 INDIA
ritik.232it026@nitk.edu.in sowmyakamath@nitk.edu.in

Abstract

This paper describes the work undertaken as part of the SMM4H-2024 shared task, specifically Task 5, which involves the binary classification of English tweets reporting children's medical disorders. The primary objective is to develop a system capable of automatically identifying tweets from users who report their pregnancy and mention children with specific medical conditions, such as attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorders (ASD), delayed speech, or asthma, while distinguishing them from tweets that merely reference a disorder without much context. Our approach leverages advanced natural language processing techniques and machine learning algorithms to accurately classify the tweets. The system achieved an overall F1-score of 0.87, highlighting its robustness and effectiveness in addressing the classification challenge posed by this task.

1 Introduction

The proliferation of social media platforms such as Twitter (now known as X), Reddit, and Facebook has led to an unprecedented surge in user-generated content. Millions of individuals publicly share their thoughts, experiences, and health-related information online, which presents a unique opportunity to analyze and investigate public health trends and issues. Among these platforms, Twitter stands out as a particularly valuable source of rich information for both the general public and researchers. By analyzing tweets, researchers can gain insights into various health-related phenomena, track the spread of diseases, monitor public sentiment toward health policies, and identify emerging health concerns. The real-time, streaming nature of Twitter data makes it an indispensable tool for public health surveillance and research, facilitating a deeper understanding of health behaviors and outcomes on a global scale (Bachina et al., 2021).

The reporting of children's medical disorders stands out as a crucial area of study, given the importance of early detection, diagnosis, and treatment in pediatric healthcare. Many children are diagnosed with disorders that can profoundly impact their daily lives and may persist throughout their lifetimes. Conditions such as attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorders (ASD), delayed speech, and asthma are frequently mentioned by parents and caregivers on social media. Studying these discussions on Twitter provides valuable insights into the prevalence and public perception of these disorders (Saini and Yadav, 2022). It also helps in understanding the challenges faced by families and the effectiveness of various interventions, through analysis of patterns and trends in symptom reporting, treatment experiences, and support mechanisms, which are essential for improving healthcare strategies and policies (Unnikrishnan et al., 2023). Furthermore, by monitoring these conversations, researchers can identify gaps in awareness and education, potentially guiding more targeted and effective public health campaigns.

2 SMM4H'24 Task 5 - Description

Task 5 is a binary classification task that involves automatically distinguishing tweets posted by users who have reported their pregnancy on Twitter and specifically mention children with ADHD, ASD, delayed speech, or asthma (annotated as "1"), from tweets that merely mention a disorder (annotated as "0"). This task enables the large-scale utilization of Twitter, not only for epidemiologic studies but also to explore parents' experiences and directly target support interventions.

The dataset (Klein et al., 2023) consists of 7,398 English language tweets for training, 389 tweets for validation, and 10,000 tweets for testing purposes. Tweets in which parents explicitly mention that

Proceedings of The 9th Social Media Mining for Health Applications (#SMM4H) Workshop and Shared Tasks, pages 133–135

their child is suffering from ADHD, ASD, delayed speech, or asthma are annotated as '1'. In contrast, other tweets are annotated as '0', which may or may not have mention of a disorder. By applying natural language processing (NLP) techniques and machine learning algorithms to this dataset, we aim to develop a robust model capable of accurately identifying and categorizing tweets related to children's medical disorders.

3 Methodology

This binary classification task involves automatically distinguishing tweets posted by users who had reported their pregnancy on Twitter and mentioned that their children had ADHD, ASD, delayed speech, or asthma in other tweets. Various preprocessing techniques were employed on the tweets during the data processing phase to ensure they were standardized and prepared for analysis The names of disorders and digits were standardized. Additionally, terms referring to a child, such as, *son*, *child*, *daughter*, etc., were unified to the common term "*child*" since the focus is on identifying tweets about child disorders posted by their parents. Furthermore, URLs, usernames, hashtags, emojis, and smileys were eliminated using the tweet-preprocessor library in Python (Van Rossum and Drake Jr, 1995). Common abbreviations and contractions found in tweets, such as "*lol*", "*thx*", "*btw*", and "*we're*" were expanded to their full forms, while non-alphanumeric characters and extra white spaces were pruned. The text was converted to lowercase, elongated words were corrected by keeping the occurrence of repeated characters to two, and lemmatization was performed using the Spacy lemmatizer (Honnibal and Montani, 2017). These steps standardized the text format and improved its suitability for subsequent analysis and modeling.

Experiments were carried out using various Transformer-based models to model the tweets. RoBERTa-base (Liu et al., 2019) was chosen for its ability to capture contextual information effectively and was implemented using the Huggingface toolkit (Wolf et al., 2019) for classifying tweets mentioning children's medical disorders. After preprocessing the textual data as elaborated earlier, the text sequences were subjected to tokenization using the RoBERTa tokenizer, with the maximum text length set to 128. Model optimization was achieved using the Adam optimizer with a batch size of 8 and a learning rate of 1e-5. The training was carried out for up to 10 epochs, with early stopping triggered by validation set performance and a patience value set at 4 epochs. Moreover, a dropout rate of 0.3, determined iteratively, was applied to regularize the model. The model architecture is built on RoBERTa-base, incorporating an additional hidden dense layer with a Rectified Linear Unit (ReLU) activation function. A sigmoid activation function is used in the output layer. The experiments were conducted on Google Colab using a T4 GPU as the hardware accelerator.

4 Results and Discussion

The system's performance was evaluated on both the validation and test sets. Initially, the performance on the validation set was compared to assess the model's effectiveness before evaluating it on the test set for submission. Efforts towards hyperparameter tuning helped achieve optimal performance on the validation set. Key hyperparameters, such as dropout rates and the number of hidden layers, were varied systematically to enhance the model's performance. This method enabled systematic exploration of hyperparameter configurations to determine the most effective settings based on validation set performance. The evaluation of the model's performance was based on the F1-score metric. The F1-score is a critical metric for this binary classification task because it balances precision and recall, offering a unified measure that considers both false positives and false negatives. This is particularly vital for distinguishing tweets about children with specific medical disorders from general mentions of disorders, ensuring that the model accurately identifies relevant tweets and minimizes the misclassification of non-relevant ones. Precision and recall scores are reported alongside the F1-score to comprehensively evaluate the model's performance in terms of the accuracy of positive predictions and the model's ability to capture all relevant instances. The results obtained from evaluating the system on the validation set are reported in Table 1.

Table 1: System Performance on *test* and *val* datasets

Dataset	F1-score	Precision	Recall
validation	0.88	0.89	0.88
test	0.87	0.86	0.88

The system achieved an F1-score of 0.87 on the test set. There is a substantial difference in perfor-

mance between our RoBERTa classifier (0.868) and the RoBERTa baseline classifier (0.927) in (Klein et al., 2024). The baseline classifier is built on the RoBERTa-large pre-trained model and has been tested on a set of 1,947 tweets, while the proposed classifier leverages the RoBERTa-base pre-trained model and is tested on a set of 10,000 tweets. Achieving an F1-score of 0.87 on the test set demonstrates that the model generalizes well to unseen data, performing consistently with high accuracy. This suggests that the system effectively learns and captures the underlying patterns and features in the tweets related to children's medical disorders.

5 Concluding Remarks

In this article, an approach to accurately distinguishing between tweets that mention specific disorders in the context of parenting and those that merely reference a disorder, using advanced NLP techniques and Transformed-based models is presented. Evaluation on both the validation and test sets demonstrates the system's reliability, with consistent F1-scores indicating its effectiveness in generalizing to unseen data. Moving forward, we aim to explore further refinements to the model architecture, incorporating additional features, and expanding the training dataset to enhance the system's performance.

References

Sony Bachina, Spandana Balumuri, and Sowmya Kamath. 2021. Ensemble albert and roberta for span prediction in question answering. In *Proceedings of the 1st workshop on document-grounded dialogue and conversational question answering (DialDoc 2021)*, pages 63–68.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Ari Z Klein, José Agustín Gutiérrez Gómez, Lisa D Levine, and Graciela Gonzalez-Hernandez. 2024. Using longitudinal twitter data for digital epidemiology of childhood health outcomes: An annotated data set and deep neural network classifiers. *J Med Internet Res*, 26:e50652.

Ari Z Klein, Shriya Kunatharaju, Karen O'Connor, and Graciela Gonzalez-Hernandez. 2023. Pregex: rule-based detection and extraction of twitter data in pregnancy. *Journal of Medical Internet Research*, 25:e40569.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Gurdeep Saini and Naveen Yadav. 2022. Ensemble neural models for depressive tendency prediction based on social media activity of twitter users. In *Security, Privacy and Data Analytics: Select Proceedings of ISPDA 2021*, pages 211–226. Springer.

Reshma Unnikrishnan et al. 2023. Efficient parameter tuning of neural foundation models for drug perspective prediction from unstructured socio-medical data. *Engineering Applications of Artificial Intelligence*, 123:106214.

Guido Van Rossum and Fred L Drake Jr. 1995. Python tutorial.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.