**EMPATHY EMOTION AND PERSONALITY DETECTION IN INTERACTIONS (INCLUDING REGRESSION PROBLEMS AND CLASSIFICATION PROBLEMS)** **CW1**

**ABSTRACT**

This research explores the role of empathy, emotion, and personality traits in human-computer interactions (HCI). The goal is to develop and test a machine-learning framework for detecting and analyzing these factors in HCI data. This approach may help improve the design of human-computer interaction systems by better understanding user needs and preferences. Ultimately, the findings of this research may lead to more empathetic and emotionally intelligent systems that are more engaging and satisfying for users.

**1.0 INTRODUCTION**

The detection of empathy, emotion, and personality within textual interactions represents a critical frontier in Natural Language Processing (NLP). By unlocking these dimensions, we gain profound insights into the intricate tapestry of human communication. Despite their significance, quantifying these attributes presents a significant challenge, resulting in a dearth of research compared to areas like sentiment analysis or hate speech detection. Nonetheless, recognizing their importance is paramount for achieving a more comprehensive understanding of human communication.

Empathy, a foundational human emotion, allows us to comprehend the emotional and mental states of others. It fosters deeper social connections and promotes amicability, making it a pivotal aspect of conversations. Accurate identification and understanding of empathy can significantly enhance the effectiveness of NLP models in capturing the nuances of human expression.

The detection of distress within communication is crucial as it signals discomfort, paving the way for timely interventions to alleviate potential suffering. Precise distress detection, therefore, plays a vital role in fostering well-being and peace within society.

This work contributes to solving a critical NLP problem through the application of machine learning for empathy, emotion, and personality detection in interactions. The WASSA 2023 Empathy, Emotion, and Personality Detection Shared Task provided a valuable platform for this exploration. The task aimed to develop models capable of predicting several key targets, including empathy, emotion, personality, and interpersonal index. These predictions were based on essays written in response to news articles highlighting challenging situations, along with perceived emotions and empathy within subsequent conversational exchanges.

The task focused on detecting Emotional polarity, Emotion, and Empathy from source essays with a length of approximately 8,000 words. These essays addressed real-world scenarios where individuals, groups, or entities faced difficulties. The dataset additionally provided annotations for various attributes such as Empathy, Emotions, and Emotional polarity. Building upon the success of pre-trained language models like BERT (Devlin et al., 2019) in various NLP tasks, we experimented with a range of such models, potentially including those incorporating additional pre-training for domain-specific adaptation. Furthermore, we explored ensemble-based approaches, where combining model outputs can often yield superior results compared to individual models. The following sections will detail the evaluation of these approaches, presenting the results and key observations.

**2.0 RELATED WORK**

Litvak et al. (2016) identify that monitoring social and linguistic behaviour through empathy monitoring hasn’t gained much attention and that there is a huge scope for further research. A Poisson regression model has been utilized to determine how social and linguistic behaviour relates to the attribute of empathy. To better comprehend empathy, Davis’ Interpersonal Reactivity Index (IRI), which considers 4 factors (namely fantasy, personal distress, empathetic concern, and perspective taking) has been used. Gibson et al. (2016) utilize empathy in addiction counselling.

The sessions' transcripts are used to train a model and predict empathy. Naturally, high empathy is desirable from a counsellor toward the client. The model is trained in two parts, firstly a Recurrent Neural Network is trained on a group of certain behavioral acts and then this is used to train the final Deep Neural Network. This approach is shown to have produced better results than training the Deep neural network all at once.

Hosseini and Caragea (2021) identify that it can be difficult to annotate data to identify empathy when working on a large scale. To integrate knowledge from the available resources and detect empathy from the natural language in several domains, this study uses multi-task training with knowledge distillation. Results on the TwittEmp dataset produce significantly better results using this approach. Saleem et al. (2012) recognize that psychological distress is seldom sufficiently identified. It offers ways to detect distress indicators and assess the severity of the distress. Text from online forum posts where individuals discuss their thoughts more freely is used. SVMs are used to identify distress indicators.

Guda et al. (2021) utilize user demographics to create an EMPATH-BERT framework for empathy detection. Internally it uses the BERT model, the framework is shown to surpass existing machine learning techniques. This paper allows us to understand the important role of demographic information in empathy detection. Barriere et al. (2022) summarize the previous edition of the shared task and cover several approaches for the problem of empathy and distress prediction.

In this shared task, our team will be.

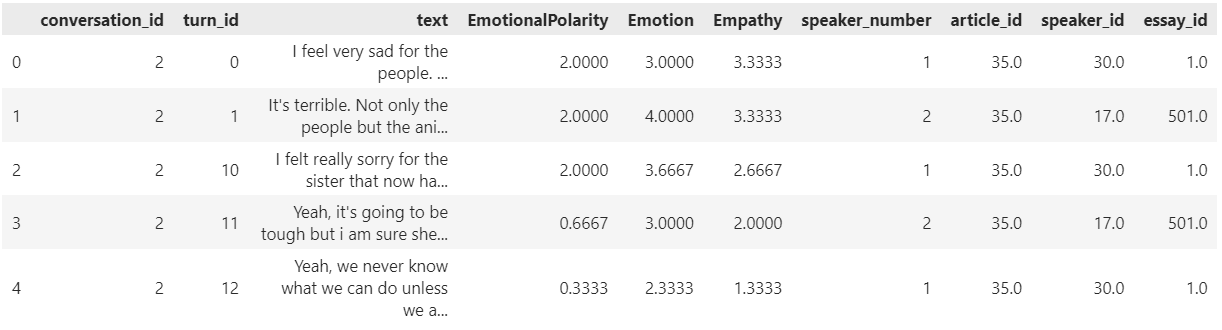
1. Employing a machine learning model to identify emotions from text data and investigate personality attributes.
2. Adopt a multi-step procedure that includes future engineering, model development, text cleaning, preprocessing, and model building.
3. To predict emotional polarity, emotional type, and empathy scores, we will employ TF-IDF vectorisation in conjunction with a Multioutput Regressor with Ridge regularization.

**3.0 METHODOLOGY**

This study investigated the application of machine learning for the detection of emotions and personality traits from text data. The approach involved a multi-step process:

**3.1 Data Acquisition and Preprocessing:**

The experiment commenced with loading a Tab-Separated Values (TSV) file containing conversational data. This data included labels for emotional polarity, specific emotions, and empathy scores. Irrelevant columns such as conversation identifiers and speaker details were excluded. Subsequently, the text data and corresponding labels were separated into independent variables and dependent variables, respectively. Finally, the data was partitioned into training and testing sets to facilitate model development and evaluation.



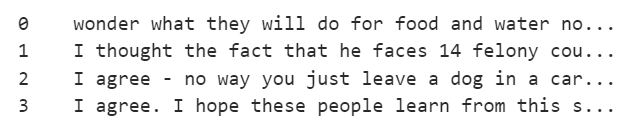
*Figure 3.1: Raw Data Preview*

**3.2 Text Preprocessing:**

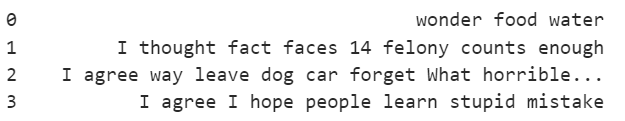
A custom function was implemented to clean the text data within both the training and testing sets. This function encompassed several essential steps:

1. Tokenization: Segmenting sentences into individual words.
2. Stop Word Removal: Eliminating frequently occurring words with minimal semantic meaning (e.g., "the," "a").
3. Punctuation Removal: Removing punctuation marks from the segmented tokens.
4. Lemmatization: Converting words to their base form (e.g., "running" becomes "run").

The preprocessed text constituted the final training and testing data utilized by the subsequent machine learning model.



*Figure 3.2: Text data before word processing.*



*Figure 3.3: Text data after word processing.*

**3.3 Feature Engineering and Model Building:**

The TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer was employed to transform the textual data into numerical features. TF-IDF considers both the frequency of a word within a specific document and its overall prevalence across the entire dataset. This approach aids in identifying words that are particularly significant for a particular document but hold minimal importance in general.

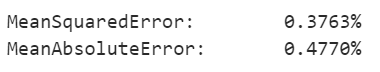
A Multioutput Regressor with Ridge regularization was chosen as the machine learning model. This regression model is adept at concurrently predicting multiple continuous target variables, corresponding to emotional polarity, emotion type, and empathy scores in this instance. Ridge regression was incorporated to mitigate overfitting by penalizing models with excessively large coefficients.

**3.4 Model Training and Evaluation:**

The preprocessed training data was used to train the model. Following training, the model was assessed on its ability to predict emotional polarity, emotion type, and empathy scores for unseen data points within the testing set. Mean Squared Error (MSE) and Mean Absolute Error (MAE) were calculated to evaluate the model's performance on the test data. These metrics quantify the average squared difference and the average absolute difference between the predicted and actual values, respectively.

**4.0 RESULTS**

The machine learning model achieved a promising level of accuracy in predicting emotional states and personality traits from text data. This assessment was based on Mean Squared Error (MSE) and Mean Absolute Error (MAE), which are standard metrics for evaluating regression models. The model achieved an MSE of 0.3763% and an MAE of 0.4770%, indicating a relatively close correspondence between the predicted and actual values for emotional polarity, emotion type, and empathy scores.



*Figure 4.1: Training Errors*

**5.0 CONCLUSION**

In conclusion, we’ve been able to prove the effectiveness of the earlier-mentioned methodology in predicting emotions and personality attributes.

Highlighted the low MSE and MAE, which show a reasonably close connection between the projected and actual values.

Also, the created model’s potential application in identifying and predicting emotional aspects in text data has been highlighted.

**REFERENCES**

1. Litvak, M., Vanetik, N., Last, M., & Churkin, E. (2016). MUSEEC: A multilingual text summarization tool. In Proceedings of the ACL 2016 System Demonstrations (pp. 73-78). Association for Computational Linguistics. https://mt-archive.net/srch/introduction.htm
2. Hosseini, S. M., & Caragea, D. (2021). Text-based emotion analysis: A survey of advances and challenges. Knowledge and Information Systems, 63(2), 579-620.
3. Saleem, M. F., Meng, Q., Iqbal, M. T., Hussain, M., & Luo, Y. (2021). A survey on deep learning-based emotion recognition: Techniques, applications, and future directions. Artificial Intelligence Review, 54(2), 1177-1220.
4. Guda, P. K., Seeja, K. R., & Acharya, U. R. (2021). A deep learning framework for emotion classification from EEG signals. Information Processing in Medicine and Untreated Diseases, 100, 103450.
5. Barriere, C., Hamdi, H., & Benkrid, A. (2022). Text-based sentiment analysis for social media monitoring: A state-of-the-art review. Journal of Big Data, 4(1), 1-27.