

Detecting Emotional Pressure by decoding social interactions using NLP models.

1st Nirnoy Chandra Sarker
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
nirnoy.chandra.sarker@g.bracu.ac.bd

2nd Jennifer Abedin
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
jennifer.abedin@g.bracu.ac.bd

3rd Ishrat Jahan
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
ishrat.jahan5@g.bracu.ac.bd

4th Adib Muhammad Amit
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
adib.muhammad.amit@g.bracu.ac.bd

6th Mehnaz Ara Fazal
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
mehnaz.ara.fazal@g.bracu.ac.bd

5th Annajiat Alim Rasel
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
annajiat.alim.rasel@g.bracu.ac.bd

Abstract—This paper explores the application of Natural Language Processing (NLP) models to decode emotional pressure in social interactions. Leveraging advanced language models, the study delves into the nuanced aspects of communication, aiming to identify and understand emotional pressure signals within textual data. The research methodology involves training models on diverse datasets to recognize patterns indicative of emotional pressure. The findings contribute to a deeper comprehension of emotional dynamics in social discourse, offering insights that can inform the development of more empathetic and context aware NLP applications. Decoding Emotional Pressure in Social Interactions via NLP Models” constitutes a groundbreaking study at the intersection of natural language processing (NLP) and emotional intelligence. This paper delves into the intricate landscape of social communication, employing advanced NLP models to decipher emotional pressure signals embedded in textual data. The research methodology involves the meticulous training of models on diverse datasets to discern patterns indicative of emotional pressure, thus contributing to a nuanced understanding of emotional dynamics within social discourse. The abstract opens with an exploration of the paper’s overarching goal – unravelling emotional pressure in social interactions through the application of NLP models. The study recognizes the pivotal role of understanding and decoding emotional undertones in human communication for the creation of empathetic and responsive artificial intelligence systems

Index Terms—Text-Based Emotion Detection (TBED), Machine learning (ML) models, Stochastic Gradient Descent (SGD), Human-Computer Interaction, Emotional Intelligence in AI.

I. INTRODUCTION

In the era of universal digital communication, a revolutionary change in documentation, expression and sharing of human emotions has been brought by the Social Web. Within Natural language programming (NLP), Text-Based Emotion

Detection (TBED) is relatively a new field which has emerged in response to this shift in thinking. [?] At the same time, it offers a great opportunity to learn about the complex range of human emotions expressed through written language. By using TBED, we can analyze massive volumes of digital exchanges for sophisticated emotional expressions using a wide range of machine learning (ML) models. Academics seeking to research and understand the complete spectrum of human emotions have a veritable resource of emotional data produced on the Social Web. Luckily, we have more than 4.3 billion active social media users across the world. [?]The intricate concept of TBED is conceptually illustrated by the profound effect of digital communication on emotional perception of human beings. We are exploring this field with the goal of developing the emotional intelligence (AI) systems by decoding the nuances of human expression concealed in textual data which are broadly applicable. For the first acknowledgement, we recognize that emotions are complicated and have many forms and different visions in digital platforms of interactions. These online platforms allow people to express quite a range of emotions such as happiness, sadness, sorrowness, fearness, anger issues and many more. [?] As a result, it has become difficult to recognize these signs of feelings and emotions of human beings. This research recognizes the need of TBED in preventing serious social issues like suicide and monitoring on community health. [?] Along with this, it also looks forward to a day when technology can interact with people’s emotions in a natural way and will improve mental health counseling and client services. In this research paper, we have approached through a wide range of ML models that reflect the complexity of the issue. As we examine the field of TBED, this research sorts valuable papers according to the emotional models along with methodological approaches

utilized. This thorough sorting does double duty. At first, it organizes the mountain of material and next points the way for future research. [?] It will lead to advances in AI systems with emotional intelligence. Our interrogation is predicted by the acknowledgement of the obstacles related to TBED. Although our study is not unconscious to the field's recognized limits, it seeks to expand emotion identification boundaries. Here, we have followed a systematic approach to ensure that our research is rooted in a practical understanding of current and future applications of TBED by conducting a realistic evaluation of its capabilities and potential. Our research visualizes a future when technology can detect and react to human emotions, re-shaping their interactions with computers in various ways. This path of prediction goes beyond the present moment. It will lead the transition of customer service to mental health aid to the emotionally sensitive AI systems that are crucial in various social contexts. When the research will come to an end, we will see how we have surveyed the current state of TBED in great depth by covering every area of machine learning techniques to the field as a whole. Other researchers and professionals will find this carefully builded classifications scheme useful for understanding the intricacies of emotional modeling. It will encourage them for further study and advancement in this field. To summarize, our research explores the intricacies of emotion detection that emphasizes its possible societal outcomes as we embark on this academic journey into the essence of TBED. [?] At first, we have laid the framework of our research for a comprehensive analysis of the methods, challenges and new directions. It indicates that our research offers for the development of AI systems with emotional intelligence. As we will examine TBED, we want this framework will be helpful for the future when technology will grasps human emotions more delicately and it will result in a deeper comprehension of the emotional threads interwoven throughout our digital manifestations.

II. LITERATURE REVIEW

In order to improve the accuracy of data and solve the present problem that is identifying fake news on the internet, this research presents a novel machine learning model [1]. By integrating advanced Natural Language Processing (NLP) techniques, the model strategically uses both content-based and social features of news articles, above the limits of standard method of detection above the limits of standard methods of detection. The model's efficiency is demonstrated by the provided findings and results, which shows 96% accuracy rate within the data set. The study shows the deep impact of fake news on society, focusing on the negative effects. The suggested approach distinguishes out for its accuracy and reliability, above the previous approaches and highlighting the importance of promoting social and content elements for accurate identification of the data set. The paper highlights various representations of research like application and regular updating of databases made feasible by a "news grabber" method which enables for quick recognition.

The paper investigates the effectiveness of ChatGPT, a large language model (LLM), in text-based mental health classification tasks. Focusing on stress, depression and suicidality detection using labeled data sets, the study employs a zero-shot suicidality classification approach and prompts the OpenAI ChatGPT API for evaluation. [2] Achieving F1 scores of 0.73 for stress, 0.86 for depression and 0.37 for suicidality, ChatGPT outperforms baseline models. While stress and depression detection align competitively with other models, suicidality detection, a more complex 5-class problem, yields a lower F1 score, possibly due to overlapping class boundaries. Acknowledging limitations, such as prompt settings and dataset size, the authors advocate for future work involving fine-tuning, diverse prompts and larger datasets. The paper underscores the potential of LLMs like ChatGPT in mental health applications, with ongoing efforts to refine models for nuanced tasks, including exploring the performance of the GPT-4 model.

The system is designed to support therapists in real-time during psychotherapy sessions. utilizing a turn-level rating mechanism, the system predicts therapeutic outcomes by assessing the similarity between deep embeddings of a scoring inventory and the patient's current sentence. [3] In response to the mental health practitioner shortage, worsened by COVID-19, the authors introduce SuperviseBot, an AI companion offering real-time feedback and treatment recommendations. Leveraging Working Alliance Inventory ratings, speaker diarization and deep embeddings, the system incorporates an Embedded Topic Model for topic modeling and employs reinforcement learning (DDPG, TD3, BCQ) to refine recommendations. Empirical results on a psychotherapy dataset demonstrate efficacy across various psychiatric conditions. SupervisorBot, presented as a web-based system, ensures ethical data processing. This study introduces a proof of concept for a real-time recommendation system, integrating NLP, reinforcement learning and topic modeling to aid therapists in psychotherapy sessions. Future work may expand system capabilities and explore ethical considerations in multi-participant interactions.

While traditional counterparts focus on information revival, these agents contradictically utilize natural language processing (NLP) and sentiment analysis to identify emotional nuances in user input. [?] To acknowledge sentiments like happiness or sadness, the agent tailors response by enhancing empathy and providing personalised assistance. The training on diverse emotional expressions allows the agent to involve in natural conversations with ongoing learning adapting to linguistic and cultural shifts. Interventions based on the detected distress, application ranges from customer service by addressing dissatisfaction proactively, to mental health support. Obstacles like ethical considerations and privacy that require a delicate balance for broad trust. To sum up, sentiment-aware agents represent a transformative AI phase by understanding both spoken words and conveyed emotions that reshaped human-computer interactions into a new era of emotionally intelligent AI systems

The research paper [?] explores the intersection point of healthcare and advanced AI communication. It examines the application of deep learning in social healthcare by concentrating on improving the understanding and categorization of textual data. As a subset of machine learning, deep learning reveals the complexities of healthcare discussions by going into various methodologies and models for text classification. Natural Language Processing (NLP) is fundamental that enables effective comprehension and processes of human language. This survey emphasises the pivotal role of NLP in connecting the gap between unstructured textual data along with intelligent algorithms in social healthcare networks. It addresses the application of deep learning in sentiment analysis, disease identification and information extraction by addressing obstacles such as interpretability, ethics and handling sensitive patient data. This paper outlines the evolution of deep learning methods in healthcare conversion and their likelihood impact on improving outcomes which facilitates early disease detection. It also enhanced patient care through data-driven insights. As a result, it serves as a roadmap for integrating advanced AI technologies into healthcare communication for more effective and informed practices.

This paper explores a critical area in natural language processing (NLP) research. [6] Utilizing NLP techniques, the research develops models for recognizing stress-inducing elements in both languages, employing machine learning algorithms and linguistic features to analyze textual data. The study emphasizes the significance of considering cultural and linguistic context in identifying stress agents, recognizing variations between Hindi and English speakers. The research envisions applications in mental health support, customer service and sentiment analysis, enhancing conversational agents and customer support systems across diverse language groups. Acknowledging challenges like limited labeled datasets for Hindi stress identification, linguistic variations and the subjective nature of stress perception, the study aims to address these for robust, generalizable models. In conclusion, this research pioneers the understanding of stress expression in multilingual settings, contributing insights into culturally aware and linguistically attuned NLP models, with practical applications in improving emotional intelligence in AI-driven systems across diverse languages and cultures.

This paper focusing on the creation of witty puns within ambiguous contexts. [7] The study leverages advanced NLP techniques to develop a model that excels in generating puns with double entendre and multiple interpretations. By intentionally introducing ambiguity, AmbiPun surprises users with unexpected wordplay, distinguishing it from traditional pun generation models reliant on clear linguistic contexts. The research delves into the nuances of pun creation, highlighting the importance of context and linguistic ambiguity in eliciting laughter. AmbiPun manipulates ambiguous scenarios to generate puns exploiting various word or phrase meanings, adding cleverness to the humor. AmbiPun's applications span entertainment, chatbots and creative writing, offering a novel way to engage users in human-computer interactions.

Challenges include balancing ambiguity and coherence in pun generation and aligning humor with user expectations. The research pioneers computational humor, opening avenues for entertaining AI systems that use linguistic ambiguity to evoke laughter and creativity. To conclude, AmbiPun introduces a pioneering approach to NLP-based pun generation, pushing the boundaries of computational humor for more engaging and entertaining AI systems.

The paper explains social inhibition, a personality trait marked by behavioral and interpersonal difficulties, has implications for emotion regulation, yet mechanisms remain unclear. [8] Socially inhibited individuals often struggle in social situations, potentially due to emotion regulation challenges. Suppression, a common strategy in socially inhibited individuals, may lead to increased stress and negative outcomes. Limited research explores the physiological consequences of suppression in socially inhibited individuals. In contrast, frequent use of reappraisal, an adaptive emotion regulation strategy, is associated with positive emotional outcomes. However, the relation between social inhibition and reappraisal remains unexplored. This study addresses this gap, examining emotional and physiological responses in socially inhibited individuals during sadness induction and instructed emotion regulation, shedding light on the psychosomatic implications of emotion regulation in social inhibition.

III. METHODOLOGY

In the figure one we have shown our work flow of this research. Initially we generated our data set and used a variety of preposition techniques. The preposition data was then divided in a 70:30 ratio.

A. Dataset

As we Previously stated, there isn't much research on emotion detection. After research we have chosen to use the data set that is accessible on kaggle. We selected this dataset as it is scrapping Facebook and tweeters emotion expression. The two main columns are description and emotion text preposition is a crucial step in NLP that involves transforming raw text data into a format suitable for analysis or ML models. In this dataset there are 16000 data. From there we have trained 70% data and 30% where used to access our model.

TABLE I
DETAILED EVALUATION OF DIFFERENT CLASSIFIER

Emotion	Count
joy	5362
sadness	4666
anger	2159
fear	1937
love	1304
surprise	572

B. Pre processing

Tokenization: Tokenization is the process of breaking down text data into individual words or tokens. Here we have used

spaCy library to tokenize the input text. Also we have to filter some common words like “I”, “am”, “is”. We have to remove those kinds of words as well. By using spaCy we have removed both stop words and punctuations. It would help us to specify the main words of our analysis.

Lemmatization: lemmatization explains removing words to their root form, aiding standardization and reducing influenced forms to a common base. [?] For this we also used the spaCy library. It is the simplest and easiest form to use. After those steps we have to join lemmatized tokens. The lemmatized tokens are our final output for the whole pre-processing techniques.

C. Machine Learning Models

An actual process that is learnt from data is represented mathematically by a machine learning model. It gains the ability to forecast or make judgments based on previously unknown data by discovering patterns and correlations within the data.

K-Nearest Neighbors (KNN) Classifier: For both regression and classification applications, the K-Nearest Neighbors (KNN) Classifier is an easy-to-understand technique. [?] Predictions using KNN are based on the average value (for regression) or the majority of classes (for classification) of the k-nearest data points in the feature space. The selection of 'k' indicates how many neighbors were taken into account for the forecast. KNN does not assume anything about the underlying data distribution because it is an instance-based, non-parametric learning technique. In our paper the accuracy is 78%.

Multinomial Naive Bayes (Multi NB) Classifier: A probabilistic machine learning approach called the Multinomial Naive Bayes Classifier [fig-4] works especially well for text classification applications like spam detection and document classification. [?] It is predicated on the independence of characteristics and the Bayes theorem. When dealing with discrete data, such word frequencies in text data, the "Multinomial" option is appropriate. Naive Bayes models are unexpectedly effective in a variety of real-world applications, despite their simplicity. In our paper the accuracy is 68%. It was 2nd lowest accuracy.

Multi-Layer Perceptron (MLP) Classifier: Under the broad category of deep learning, one kind of artificial neural network is the Multi-Layer Perceptron (MLP) Classifier [fig-5]. It is made up of an input layer, one or more hidden layers, and an output layer, which are all layers of nodes, or neurons. [?] During training, the model discovers the weights assigned to each link between nodes. MLP are useful for a variety of applications, such as voice and picture recognition, since they can identify intricate patterns in data. In our paper the accuracy is 61%. It was the lowest accuracy.

Support Vector Machine (SVM) / Support Vector Classifier (SVC): Two very comparable techniques for classification problems are the Support Vector Machine

(SVM) and Support Vector Classifier (SVC). In order to maximize the margin between classes, support vector machines (SVM) search for the best hyperplane in the feature space that divides various classes. A particular kind of SVM intended for two-class classification is called SVC. Because of their strength and efficiency in high-dimensional environments, support vector machines (SVM) are widely used in informatics, image classification, and handwriting recognition applications. In our paper the accuracy is 83%. It was 3rd highest accuracy.

Random Forest: During training, the Random Forest ensemble learning approach creates many decision trees, and during testing, it combines their predictions. A random portion of the training data and a random subset of the characteristics at each node are used to build each tree. This unpredictability lessens over-fitting and strengthens the resilience of the model. Because of their versatility, random forests may be used for problems solving both regression and classification. They are less susceptible to outliers and frequently offer great accuracy. In our paper the accuracy is 85%. It was 2nd highest accuracy.

Stochastic Gradient Descent (SGD) Classifier: One popular optimization technique for training machine learning models is the Stochastic Gradient Descent (SGD) classifier, especially for big datasets. It is computationally efficient since it updates model parameters iteratively using tiny batches of data. First, the model's parameters are initialized. Next, a loss function is defined to quantify the difference between the expected and actual values. Finally, the parameters are repeatedly updated for each batch using forward and backward passes. Its "stochastic" quality, which updates following each batch, sets it apart from conventional gradient descent. The algorithm's adaptability to a wide range of classification problems is demonstrated by its use of widely used loss functions, including cross-entropy. Because SGD uses stochastic updates, it might be vulnerable to local least convergence even if it is highly efficient. In our paper the accuracy is 87%. It was the most highest accuracy.

IV. EXPERIMENTS AND RESULTS

A large amount of methodical research and experimentation has been done which includes testing and modification in models which are considered as the pre-processing. The ultimate aim of this testing was to identify an optimal pipeline that comes with various and better results means accuracy with better performance. We methodically find different sets of pre-processing and model settings throughout an amount of systematic cycles. To enhance our Machine Learning (ML) models we used Tokenization technique to strip emojis, punctuation and irrelevant strings. After our prolonged testing and experimentation we find notable improvements in our results. The significant change in results provides ample proof that our pre-processing procedures

that we used were essential in improving the quality of the input data that our models for machine learning used. The outcomes of our model mentioned evaluation shows notable variations in performance amongst the models. With an 87% accuracy rate, the Stochastic Gradient Descent (SGD) Classifier demonstrated its strong efficacy in managing the provided data. This classifier also showed the highest accuracy. With an accuracy of 85%, the Random Forest model came second overall, displaying how well it could identify the patterns in the dataset. With a slightly lower accuracy of 78%, the K-Nearest Neighbors (KNN) model lagged behind. Additionally, the Multi-Layer Perceptron (MLP) Classifier shows lower accuracy of 57%, suggesting challenges in accurately capturing the dataset's fundamental trends. For a comprehensive breakdown of our model evaluation, please see the detailed table provided below

V. CONCLUSION

In this research, we have formed a unique method to detect emotional pressure through the use of Natural Language Processing (NLP) models to decode social interactions. The findings of the effective use of advanced Natural Language Processing techniques are beneficial, especially when compared to the positive impacts of better pre-processing on a balanced data set. Surprisingly, our research has shown that using existing embedding methods improves the overall performance of the model. Our findings point to a promising future study by implementing more Machine Learning Models. [9] Also, we suggest developing an exceptional data set that should be built in the starting for recognizing emotional strain using NLP model decoding of social interactions.

REFERENCES

- [1] Shubham Bauskar, Vijay Badole, Prajal Jain, Meenu Chawla, "Natural Language Processing based Hybrid Model for Detecting Fake News Using Content-Based Features and Social Features", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.11, No.4, pp. 1-10, 2019. DOI:10.5815/ijieeb.2019.04.01
- [2] B. Lamichhane, "Evaluation of CHATGPT for NLP-based mental health applications," *arXiv.org*, 28-Mar-2023. [Online]. Available: <https://arxiv.org/abs/2303.15727>.
- [3] Helping therapists with NLP-annotated recommendation - hal.science. [Online]. Available: <https://hal.science/hal-03999948/document>.
- [4] I. D. U. de Lisboa, I. Dias, U. de Lisboa, U. de L. Profile, R. R. U. de Lisboa, R. Rei, Universidade de Lisboa Search about this author, P. P. U. de Lisboa, P. Pereira, L. C. U. de Lisboa, L. Coheur, I. S. Técnico, U. do Algarve, U. of Twente, Contributor MetricsExpand All, Publication Years2022 - 2022Publication counts1Available for Download1Citation count1Downloads (cumulative)81Downloads (6 weeks)4Downloads (12 months)43Average Citation per Article1Average Downloads pe, "Towards a sentiment-aware conversational agent: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents," *ACM Conferences*, 01-Sep-2022. [Online]. Available: <https://dl.acm.org/doi/10.1145/3514197.3549692>.
- [5] P. Lavanya and E. Sasikala, "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey," *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, Coimbatore, India, 2021, pp. 603-609, doi: 10.1109/ICSPC51351.2021.9451752.
- [6] V. Gupta, A. Jain and N. Kumar, "Identification of emotion stress agents in Hindi and English sentences," *2009 International Conference on Intelligent Agent & Multi-Agent Systems*, Chennai, India, 2009, pp. 1-2, doi: 10.1109/IAMA.2009.5228090.
- [7] A. Mittal, Y. Tian, and N. Peng, "Ambipun: Generating humorous puns with ambiguous context," *arXiv.org*, 04-May-2022. [Online]. Available: <https://arxiv.org/abs/2205.01825>.
- [8] Author links open overlay panelStefanie Duijndam, I, "Emotion regulation in social interaction: Physiological and emotional responses associated with social inhibition," *International Journal of Psychophysiology*, 18-Oct-2020. [Online]. Available: <https://sciencedirect.com/science/article/pii/S0167876020302336>.
- [9] A. Gupta, A. Raut, R. Yadav, M. Kumar and V. K. Chaurasiya, "A Hybrid Approach based Stress Monitoring System for Office Environment using IoT," *2022 IEEE 19th India Council International Conference (INDICON)*, Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039725.
- [10] J. -L. Wu, Y. He, L. -C. Yu and K. R. Lai, "Identifying Emotion Labels From Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model," in *IEEE Access*, vol. 8, pp. 66638-66646, 2020, doi: 10.1109/ACCESS.2020.2985228.
- [11] A. Kumar, T. E. Trueman and E. Cambria, "Stress Identification in Online Social Networks," *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, Orlando, FL, USA, 2022, pp. 427-434, doi: 10.1109/ICDMW58026.2022.00063.
- [12] Author links open overlay panelNilesh Shelke a, a, b, c, d, e, f, and AbstractText devices are effectively and heavily used for interactions these days. Emotion extraction from the text has derived huge importance and is upcoming area of research in Natural Language Processing. Recognition of emotions from text has high pra, "An efficient way of text-based emotion analysis from social media using LRA-DNN," *Neuroscience Informatics*, 26-Jan-2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772528622000103>.