CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Research Outcomes

This project study about the use of data science techniques to detect early signs of mental health issues from social media post and comments. This project focus on posts from r/Malaysia subreddit on Reddit, where users often share personal experiences and concerns related to mental health.

To collect the data for Malaysian post, Python library called PRAW was used. 13,723 posts and comments were collected by searching for keywords such as “mental health”, “depression”, “anxiety”, and “stress”. The dataset included some details like post title, body text, comment, and timestamps.

Title, body text and comments of the data be combined into a new column, this is to make sure the emotion detection more accurate. Before analysis the data, the data cleaning was done. The text data preprocessing step included removing the URL links, non-English characters, unnecessary symbols.

In this project, a pre-trained DistilBERT model ‘bhadresh-savani/distilbert-base-uncased-emotion’ was used to classify each data into one of six emotional categories like anger, fear, joy, love, sadness, and surprise. From the project, there were 2 emotion which were sadness and fear commonly linked with high-risk mental health situations.

Out of all the posts, there were 6,035 posts have the highest probability with label sadness and fear. This mean that these posts have high probability to have mental health problem. In this project, Word Cloud was used to visualize the most common words associated with the emotions.

To better understand why the model made the predictions, explainable AI technique called LIME was used. These explanations showed which words influenced the model's decision the most, this helps to increase the trustworthy.

5.2 Contributions to Knowledge

This project makes meaningful contributions to both research and practical applications in the field of mental health analysis. One of the key contributions to knowledge is the development of an end-to-end pipeline for social media analysis. Start from data collection using Reddit's API to preprocessing, until emotion classification with DistilBERT, and model interpretation using XAI model's LIME. Additionally, the project demonstrates the value of explainable AI in emotion detection by using LIME to interpret DistilBERT predictions. This adds transparency to model decisions which is especially important when working on sensitive topics like mental health where trust and understanding are important. Moreover, this project contributions in identifying of high-risk language patterns. The word clouds and LIME explanations identified specific language patterns and words that linked with sadness and fear. These findings can help inform future monitoring system or early detection tools for mental health problem.

5.3 Future Works

While the current system shows promising results, there are several areas for improvement. One of the improvements is the fine-tuning of emotion detection model on domain-specific mental health datasets which could significantly enhance its accuracy and relevance for real-world applications. Another important area is the expansion of data sources other that Reddit which include platform such as Twitter/X, Facebook groups, or local Malaysian forums. This would provide a more diverse and representative dataset, which allowing for broader insights into public discussions around mental health.

References

(WHO), W. H. (2019). *Mental health*. Retrieved from World Health Organization (WHO): https://www.who.int/health-topics/mental-health#tab=tab_2

(WHO), W. H. (2021). *Suicide prevention*. Retrieved from World Health Organization (WHO): https://www.who.int/health-topics/suicide#tab=tab_1

(WHO), W. H. (2022). *Mental disorders*. Retrieved from World Health Organization (WHO): https://www.who.int/news-room/fact-sheets/detail/mental-disorders

(WHO), W. H. (n.d.). *Anxiety disorders*. Retrieved from World Health Organization (WHO): https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders

A.Naslund, J., Bondre, A., Torous, J., & A.Aschbremmer, K. (2020). Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice.

A.Vogels, E., Gelles-Watnick, R., & Massarat, N. (2022). Teens, Social Media and Technology 2022.

Acito, F. (2023). Naïve Bayes.

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review.

Angiulli, F., Fassetti, F., & Nisticò, S. (2021). Local Interpretable Classifier Explanations with Self-generated Semantic Features.

Antonio, V., Julio, T., & M., C.-M. J. (2020). Urbanization and emerging mental health issues.

Auxier, B., & Anderson, M. (2021). Social Media Use in 2021.

Banna, M. H., Ghosh, T., Md. Jaber Al Nahian, M. S., Mahmud, M., & Taher, K. A. (2023). A Hybrid Deep Learning Model to Predict the Impact of COVID-19 on Mental Health From Social Media Big Data.

Bao, E., Pérez, A., & Parapar, J. (2024). Explainable depression symptom detection in social media.

Bauer, B., Norel, R., Leow, A., Rached, Z. A., Wen, B., & Cecchi, G. (2024). Using Large Language Models to Understand Suicidality in a Social Media–Based Taxonomy of Mental Health Disorders: Linguistic Analysis of Reddit Posts.

Becker, M., & Correll, C. U. (2020). Suicidality in Childhood and Adolescence.

Braghieri, L., Levy, R., & Makarin, A. (2022). Social Media and Mental Health.

Carballo, J. J., Llorente, C., Kehrmann, L., Flamarique, I., Zuddas, A., Purper-Ouakil, D., . . . Arango, C. (2020). Psychosocial risk factors for suicidality in children and adolescents.

Chang, H.-J., Song, S., Kim, G., Kim, T.-Y., Shin, D., Kim, S.-H., . . . Sung, J.-H. (2023). BERT-based model to predict cardiovascular disease by analyzing healthcare utilization behavior of patients newly diagnosed with metabolic diseases.

Choi, J. E., Shin, J. W., & Shin, D. W. (2024). Vector SHAP Values for Machine Learning Time Series Forecasting.

Convolutional Neural Networks. (2022).

Das, A. K., & Ahmed, S. S. (2023). Convolutional neural networks.

Dobrek, L., & Glowacka, K. (2023). Depression and Its Phytopharmacotherapy—A Narrative Review.

Eleftheriades, R., Fiala, C., & Pasic, M. d. (2020). The challenges and mental health issues of academic trainees.

Eric, R., Punyanunt-Carter, N., R.LaFreniere, J., S.Norman, M., & G.Kimball, T. (2020). The serially mediated relationship between emerging adults' social media use and mental well-being.

Feizi, A., & Nazemi, A. (2022). Classifying random variables based on support vector machine and a neural network scheme.

Fusar-Poli, P., Gonzalo Salazar de Pablo, A. D., Nieman, S. H., Correll, C. U., Kessing, L. V., Pfenning, A., . . . Amelsvoort, T. c. (2020). What is good mental health? A scoping review.

Garg, K., Garg, S., Dixit, H., & Pandey, K. (2024). Machine Learning Driven Analysis of Mental Health Indicators in Social Media Posts.

Gerlings, J., Shollo, A., & Constantiou, I. (2021). Reviewing the Need for Explainable Artificial Intelligence (xAI).

Ha, H. T., Nguyen, D. T., & Stoeckel, T. (2024). What is the best predictor of word difficulty? A case of data mining using random forest.

Hasija, Y., & Chakraborty, R. (2021). Support Vector Machines.

Health, N. N. (n.d.). *Anxiety*. Retrieved from MedlinePlus: https://medlineplus.gov/anxiety.html

Herren, A., & Hahn, P. R. (2022). Statistical Aspects of SHAP: Functional ANOVA for Model Interpretation.

Hulsen, T. (2023). Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare.

IBM. (n.d.). *What is random forest?* Retrieved from IBM.

Jo, A. A., Raj, E. D., Vino, A. S., & Menon, P. V. (2024). Exploring Explainable AI for Enhanced Depression Prediction in Mental Health.

K, K., & Wong, L. (2023). Support Vector Machine.

Kabir, M. K., Islam, M., Kabir, A. N., Haque, A., & Rhaman, M. K. (2022). Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques.

Kamal, M., Khan, S. U., Hussain, S., Nasir, A., Aslam, K., Tariq, S., & Ullah, M. F. (2020). Predicting Mental Illness using Social Media Posts and Comments. *(IJACSA) International Journal of Advanced Computer Science and Applications*.

Kaushik, P., & Sharma, P. (2024). Integrating Machine Learning and Sentiment Analysis: A Comparative Study on Mental Health Classification from Social Media Data.

Khanduja, D. K., & Kaur, S. (2023). The Categorization of Documents Using Support Vector Machines.

Kim, J., Lee, J., Park, E., & Han, J. (2020). A deep learning model for detecting mental illness from user content on social media.

Kumar, R., Goswami, B., Mhatre, S. M., & Agrawal, S. (2024). Naive Bayes in Focus: A Thorough Examination of its Algorithmic Foundations and Use Cases.

Kumar, S., & Gota, V. (2023). Logistic regression in cancer research: A narrative review of the concept, analysis, and interpretation.

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools.

Lim, S. R., Kamarudin, N., Ismail, N. H., Ismail, N. A., & Kamal, N. A. (2023). Predicting Mental Health Disorder on Twitter Using Machine Learning Techniques.

Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., & Kitchen, G. B. (2021). Natural language processing in medicine: A review.

Marwaha, P. S., Palmer, P. E., Suppes, M. P., Cons, M. P., Young, M. P., & Upthegrove, M. P. (2023). Novel and emerging treatments for major depression.

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2021). Explainable artificial intelligence: a comprehensive review.

Odja, K. D., Widiarta, J., Purwanto, E. S., & Ario, M. K. (2024). Mental illness detection using sentiment analysis in social media.

Pinaya, W. H., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Convolutional neural networks.

Qorich, M., & Ouazzani, R. E. (2025). Lightweight advanced deep learning models for stress detection on social media.

Renuka, O., & Radhakrishnan, N. (2024). BERT for Twitter Sentiment Analysis: Achieving High Accuracy and Balanced Performance.

Richardson, R., Connell, T., Foster, M., Blamires, J., Keshoor, S., Moir, C., & Zeng, I. S. (2024). Risk and Protective Factors of Self-harm and Suicidality in Adolescents: An Umbrella Review with Meta-Analysis.

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2024). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME.

Sarno, E., Moeser, A. J., & Robison, A. J. (2021). Chapter Eight - Neuroimmunology of depression.

Scheda, R., & Diciotti, S. (2022). Explanations of Machine Learning Models in Repeated Nested Cross-Validation: An Application in Age Prediction Using Brain Complexity Features.

Schonlau, M. (2023). Logistic Regression.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning.

Stryker, C., & Holdsworth, J. (2024). *What id NLP (natural language processing)?* Retrieved from IBM: https://www.ibm.com/think/topics/natural-language-processing

Tahir, W. B., Khalid, S., Almutairi, S., Abohashrh, M., Memon, S. A., & Khan, J. (2025). Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques.

Tang, H., Rekavandi, A. M., Rooprai, D., Dwivedi, G., Sanfilippo, F. M., Boussaid, F., & Bennamoun, M. (2024). Analysis and evaluation of explainable artificial intelligence on suicide risk assessment.

Vishwakarma, S., & Ganguly, S. (2023). Optimal partition of feature using Bayesian classifier.

Walter, H. J., Bukstein, O. G., Abright, A. R., Keable, H., Ramtekkar, U., Ripperger-Suhler, J., & Rockhill, C. (2020). Clinical Practice Guideline for the Assessment and Treatment of Children and Adolescents With Anxiety Disorders.

WHO. (2025). *World Health Organization (WHO)*. Retrieved from Mental health: https://www.who.int/health-topics/mental-health#tab=tab_1

Wijaya, V., & Rachmat, N. (2024). Comparison of SVM, Random Forest, and Logistic Regression Performance in Student Mental Health Screening.

William, D., & Suhartono, D. (2020). Text-based Depression Detection on Social Media Posts: A.

Zafar, M. R., & Khan, N. M. (2021). Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability.

Zhang, T., Yang, K., Shaoxiong, & Ananiadou, S. (2023). Emotion fusion for mental illness detection from social media: A survey.

Zodage, P., Harianawala, H., Shaikh, H., & Kharodla, A. (2024). Explainable AI (XAI): History, Basic Ideas and Methods.

Zogan, H., Razzak, I., Wang, X., Jammel, S., & Xu, G. (2022). Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media.