Journal of Affective Disorders 369 (2025) 390–391

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Model-specific feature importances: Distinguishing true associations from target-feature relationships

Kusuma et al. (2024) developed logistic regression, random forests, and gradient boosting algorithms to predict suicide attempts at two distinct stages. They calculated feature importances to identify the most significant predictors of suicide attempts. However, many researchers, including Kusuma et al., fail to recognize the distinction between true associations and model-specific feature importances. Feature importances in machine learning are inherently model-specific due to biases and can be influenced by the bias and variance of the model, potentially leading to results that differ from true outcomes (Ribeiro et al., 2016; Pagano et al., 2023; Johnsen et al., 2023). Additionally, complex models can capture non-linear relationships that simpler models may miss, further complicating the interpretation of feature importances. Overfitting is another issue, where the model captures noise in the training data as if it were a true signal, leading to misleading importance scores. Therefore, this paper recommends using chi-squared tests and *p*-values instead of feature importances for identifying true associations between the target and features.

Feature importance and chi-squared tests serve different purposes in statistical analysis and machine learning, and they have distinct implications for understanding associations between features and the target variable. While feature importance measures indicate how much each feature contributes to the model's predictions, chi-squared tests and *p*-values are used in statistical hypothesis testing to determine whether there is a significant association between categorical variables, providing a more robust measure of true associations.

Feature importance measures in machine learning indicate how much each feature contributes to the model's predictions. However, they do not necessarily reflect true associations between the features and the target variable. Feature importance values are specific to the model used, meaning different models (e.g., decision trees, random forests, logistic regression) may assign different importance scores to the same features (Saarela and Jauhiainen, 2021). Additionally, feature importance can be influenced by the bias and variance of the model, with complex models capturing non-linear relationships that simpler models may miss (Saarela and Jauhiainen, 2021; Johnsen et al., 2023; Gichoya et al., 2023). Furthermore, feature importance measures can be affected by overfitting issues, where the model captures noise in the training data as if it were a true signal, leading to misleading importance scores (Alqahtani et al., 2024).

Chi-squared tests and *p*-values are used in statistical hypothesis testing to determine whether there is a significant association between categorical variables. They are considered to provide true associations for several reasons (Sharpe, 2015; Andrade, 2019). First, chi-squared tests assess whether the observed frequencies in a contingency table differ significantly from expected frequencies, with a low *p*-value indicating a significant association between the variables. Second, the chi-squared test of independence specifically tests whether two categorical variables are independent, and if the test rejects the null hypothesis, it suggests a true association between the variables. Lastly, *p*-values provide a measure of the probability that the observed association is due to random chance, allowing researchers to control for Type I errors (false positives).

Permutation feature importance (PFI) measures the performance degradation when features are randomly sorted in a model-independent manner (Breiman, 2001). PFI quantifies the relative importance of features in relation to the target variable within a specific model, while chi-squared tests identify true associations between categorical variables and the target. Although PFI does not measure true associations, it can be used alongside *P*-values to provide a comprehensive analysis (Breiman, 2001).

We emphasize the practical importance of complex predictive models, which are indispensable in real-world applications. Additionally, we acknowledge the limitations of basic statistical methods, such as *p*-values, particularly in the context of modern AI and machine learning. This paper underscores the necessity for caution in the contemporary AI landscape, where there is a tendency to overemphasize feature importance. By addressing these points, we aim to clarify that feature importance metrics do not necessarily indicate true associations. Furthermore, we highlight the role of Chi-squared tests and *p*-values in statistical methods, emphasizing their relevance and limitations in the analysis of predictive models.

CRediT authorship contribution statement

Yoshiyasu Takefuji: Writing – review & editing, Writing – original draft, Validation, Investigation, Conceptualization.

Funding

This research has no fund.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

https://doi.org/10.1016/j.jad.2024.10.019
Received 5 September 2024; Received in revised form 2 October 2024; Accepted 7 October 2024
Available online 9 October 2024

References

Alqahtani, A., Bhattacharjee, S., Almopti, A., Li, C., Nabi, G., 2024. Radiomics-based machine learning approach for the prediction of grade and stage in upper urinary tract urothelial carcinoma: a step towards virtual biopsy. Int. J. Surg. (London, England) 110 (6), 3258–3268. https://doi.org/10.1097/JS9.0000000000001483.

Andrade, C., 2019. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. Indian J. Psychol. Med. 41 (3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Gichoya, J.W., Thomas, K., Celi, L.A., Safdar, N., Banerjee, I., Banja, J.D., Seyyed-Kalantari, L., Trivedi, H., Purkayastha, S., 2023. AI pitfalls and what not to do: mitigating bias in AI. Br. J. Radiol. 96 (1150), 20230023. https://doi.org/10.1259/bjr.20230023.

Johnsen, P.V., Strümke, I., Langaas, M., DeWan, A.T., Riemer-Sørensen, S., 2023. Inferring feature importance with uncertainties with application to large genotype data. PLoS Comput. Biol. 19 (3), e1010963. https://doi.org/10.1371/journal.pcbi.1010963.

Kusuma, K., Larsen, M., Quiroz, J.C., Torok, M., 2024. Age-stratified predictions of suicide attempts using machine learning in middle and late adolescence. J. Affect. Disord. 365, 126–133. https://doi.org/10.1016/j.jad.2024.08.043.

Pagano, T.P., Loureiro, R.B., Lisboa, F.V.N., Peixoto, R.M., Guimarães, G.A.S., Cruz, G.O.R., Araujo, M.M., Santos, L.L., Cruz, M.A.S., Oliveira, E.L.S., Winkler, I., Nascimento, E.G.S., 2023. Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. Big Data Cognit. Comput. 7 (1), 15. https://doi.org/10.3390/bdcc7010015.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why you should trust your machine learning model: inherent stability of feature attributions. Adv. Neural Inf. Proces. Syst. 29, 3786–3796. https://doi.org/10.1145/2939672.2939778.

Saarela, M., Jauhiainen, S., 2021. Comparison of feature importance measures as explanations for classification models. SN. Appl. Sci. 3 (272). https://doi.org/10.1007/s42452-021-04148-9.

Sharpe, D., 2015. Your Chi-Square test is statistically significant: now what? Pract. Assess. Res. Eval. 20, 1–10.

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

E-mail address: takefuji@keio.jp.