Cultural-aware Machine Learning based Analysis of COVID-19 Vaccine Hesitancy

Raed Alharbi[1], Sylvia Chan-Olmsted[2], Huan Chen[3], and My T. Thai[1*]

[1]*Computer and Information Science and Engineering Department, University of Florida, Gainesville, FL, 32611, USA.*
[1]*{r.alharbi, mythai}@ufl.edu*
[2]*Department of Media Production, Management, and Technology, University of Florida, Gainesville, FL, 32611, USA.*
[2]*chanolmsted@jou.ufl.edu*
[3]*Department of Advertising, University of Florida, Gainesville, FL, 32611, USA.*
[3]*huanchen@jou.ufl.edu*

***Abstract*—Understanding the COVID-19 vaccine hesitancy, such as who and why, is very crucial since a large-scale vaccine adoption remains as one of the most efficient methods of controlling the pandemic. Such an understanding also provides insights into designing successful vaccination campaigns for future pandemics. Unfortunately, there are many factors involving in deciding whether to take the vaccine, especially from the cultural point of view. To obtain these goals, we design a novel culture-aware machine learning (ML) model, based on our new data collection, for predicting vaccination willingness. We further analyze the most important features which contribute to the ML model's predictions using advanced AI explainers such as the Probabilistic Graphical Model (PGM) and Shapley Additive Explanations (SHAP). These analyses reveal the key factors that most likely impact the vaccine adoption decisions. Our findings show that Hispanic and African American are most likely impacted by cultural characteristics such as religions and ethnic affiliation, whereas the vaccine trust and approval influence the Asian communities the most. Our results also show that cultural characteristics, rumors, and political affiliation are associated with increased vaccine rejection.**

***Index Terms*—COVID-19, Social Network**

I. Introduction

A better understanding of individuals and ethnic groups attitudes towards vaccination hesitancy is essential since it can provide insights into: (1) The distinct cultural behavior associated with vaccine-reluctant individuals and ethnic groups. (2) The impact of ethnic groups on individual members' decisions against vaccination, and (3) The characteristics shared by vaccine skeptics (when/why/how do they decide to accept/refuse vaccination). These insights, in turn, help us: (1) Design a model to predict the vaccine hesitancy for a given individual; and (2) develop efficient methodologies to vaccinate those who are reluctant and help in containing the current disease from spreading and combating future epidemics. Hence, it is very critical to have a full grasp of the reasons for vaccine hesitancy.

Unfortunately, vaccine hesitancy is a complicated and context-dependent phenomenon that is influenced by factors variable across time and location. Vaccine hesitancy can be associated with social-economic background (e.g., education, income, and occupation), perceptions of susceptibility (e.g., government, healthcare system, and doctors trust), and severity of disease (e.g., long-term side effects and health consequences). Furthermore, vaccination hesitation is not self-contained and can be influenced by a variety of variables such as the fast development and approval of COVID-19 vaccines, side effects shown in some individuals, and deaths of individuals after vaccination due to other complications. Some individuals use these factors to promote disinformation, adding to the long-standing distrust of public health institutions among various ethnic groups.

To tackle these challenges, we first focus on building a vaccine hesitancy prediction machine learning model called VacAdopt, with an emphasis on culture-awareness. In doing so, we conduct a systematic survey collecting 125 possible factors underlying vaccine hesitancy. This new data set consists of comprehensive features ranging from vaccination health concerns to culture-related characteristics. We next focus on analyzing the important factors that impact an individual's decision on taking vaccination. This is equivalent to finding important features of an input that constitute to the VacAdopt's prediction, which can be done using recent developments in explainable AI (XAI).

XAI research has been emerged recently to provide explanations for a given input, in a form of important features, for any given black-box ML model in a model-agnostic fashion. Two notable explainers which this paper use are SHAP [1] and PGM [2]. We perform an in-depth analysis using SHAP to identify key characteristics independently associated with COVID-19 vaccine acceptance or rejection. We next turn our attention to analyzing the composite features of vaccination willingness. To identify the correlation between important features, we adopt PGM explainer, which is developed for graph neural networks to work on our tabular dataset. Finally, we analyze the greater impact of particular features on specific ethnic groups, including Asian, African American, and Hispanic.

Organization. The remainder of the paper is structured as follows. Section II presents the background and related works. The proposed framework details are introduced in section III,

*Corresponding author.

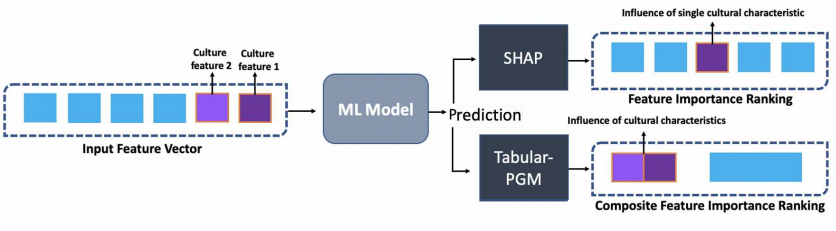


Fig. 1: VACADOPT: Cultural-awareness model architecture.

while the experimental analysis is discussed in section IV. Finally, section V concludes the paper.

II. BACKGROUND AND RELATED WORK

Vaccination Willingness Models for Tabular Data. The most frequent data format in real-world applications is tabular data, consisting of samples (rows) and a set of properties (columns). Tabular data is utilized across practical applications, including medical, banking research, and many other relational database-based applications. Due to their superior performance, conventional ML methods have remained dominant in tabular data applications over the past decade.

In particular, eXtreme Gradient Boosting (XGB) [3] is the state-of-the-art ML model for tabular data. XGB integrates numerous weak predictors, usually Decision Trees, to build an additive predictive model. Another technique is the decision tree [4], which presents feature-based split predictions in a flowchart-like tree. Other ML methods include k-Nearest Neighbors (KNN) [5], which classify data based on its surrounding data points, and random forest (RF) [6], an ensemble of classifiers built with randomization sources.

Our work is related to exploring the above methods for COVID-19 vaccination willingness detection. Yet, there needs to be more characterizing comprehensive features ranging from vaccination health concerns to culture-related characteristics. Therefore, we first conduct a systematic survey collecting 125 possible factors underlying vaccine hesitancy. We next focus on developing a vaccine hesitancy prediction ML model, with a focus on cultural awareness.

Explainable Machine Learning. Regardless of how widely ML models are adopted, they remain black boxes to users. Thus, ML explainers are employed to deduce (1) Why the selected models generated such predictions on a given input. (2) The model's behavior enables us to identify situations where the feature may significantly influence one ethnic group more than another. (3) The model's transparency increases model trust. ML explainable methods can be categorized into two types: post-hoc explainability and intrinsic explainability [7]. Post-hoc accomplishes interpretability using various statistical approaches to understand the model's behavior fully [1]. In contrast, intrinsic explainability is obtained by reducing the complexity of the model [8].

To understand the reasoning behind models' predictions, we select two well-known explainers, SHAP [1] and PGM-Explainer [2]. SHAP is designed for Conventional Neural Networks (CNNs), while PGM-Explainer is for Graph Neural Networks (GNNs).

In more details, SHAP is a game-theoretic technique to compute Shapley values to explain each feature's contribution to a prediction. The Shapley values of the model's conditional expectation function are approximated using additive feature attribution methods. And PGM-Explainer is an explanation approach that uses a graphical model to approximate the target prediction; PGM-Explainer demonstrates the effects of given features non-linearly to a prediction. However, due to its perturbation architecture, the current version of the PGM explanation is not directly usable on tabular data. Therefore, we propose a perturbation schema to use PGM-Explainer to handle our tabular dataset.

III. THE PROPOSED FRAMEWORK - VACADOPT

This section provides the implementation details of our proposed model, VACADOPT in Figure 1, where the key components are: (i) A new benchmark that collects and evaluates many aspects of vaccination reluctance in a systematic manner, with an emphasis on culture-awareness features. (ii) A proof-of-concept ML model for predicting a user's desire to receive COVID- 19 vaccination based on various baseline variables. (iii) And the adopted version of PGM (Tabular-PGM) to explain the influence of composite cultural features.

This section starts with a description of the dataset, followed by the experimental setup implementation details, including encoding schema, hyperparameter optimization. Next, we discuss in depth the adopted version of PGM.

A. Data preparation and collection.

Given the nature of this paper in investigating acceptance or refusal of COVID-19 vaccination among America's ethnic minorities, an online survey is the most suitable approach for collecting data. Prior to the primary data collection, a pretest is conducted using Amazon's MTurk to evaluate the scale items' reliability. MTurk is a robust crowdsourcing tool that allows academics to collect data more effectively while maintaining data quality [9]. Once the scale's reliability has been determined, a national online survey was distributed via Qualtric (an American experience management company that allows users to create professional surveys and generate reports) in the Fall of 2021.

Audience panels from Qualtrics provide researchers with access to a nationally representative sample that closely represents America's demographic makeup. According to recent US census data, Hispanics account for 18.1% of America's racial makeup, followed by African Americans at 13.4%, and Asians at 5.9% [10]. Thus, the ratio of three ethnic minority groups in the sample is maintained using a soft quota by Qualtrics to reflect America's racial demography best.

Information is collected include but not limit to general media consumption/use, special COVID and language media consumption factors, COVID-19 information seeking, COVID-19 information value, Trust in primary health information source/media, CDC, FDA, vaccine and vaccination perception. The final dataset's structure, characteristics, and statistics can be found in [11].

B. Experimental Setup.

The ML model in Figure 1 is implemented to construct a predictor of COVID-19 vaccination willingness: whether a person would accept or refuse the COVID-19 immunization (target variable). We utilize five well-known ML-based classification techniques: XGB [3], decision tree [4], k-KNN [5], RF [6], Support Vector Machine (SVM) [12]. For a fair comparison, we consider a deep learning method, TABNET [13], that is designed for tabular data.

Encoding Schema. Our structured dataset consists of many columns, including numerical and categorical features. Finding an appropriate encoding schema in which categorical features are converted to numbers is critical for developing an effective and successful machine learning model. In our experiment, we encode our categorical features by combining two well-known encoding techniques, one-hot [14] and label encoding [15]. Developing such a hybrid technique is not trivial, as one encoder can work better than the other according to specific features.

We first divide the non-numerical features in our prescribed dataset into (1) order-based and (2) unorder-based. The order-based refers to the features where the order of the values in the feature column is matter such as level of agreement/disagreement with the vaccine. In contrast, The unorder-based represents the values of the feature in an irrelevant order. For example, the gender feature column's should have equal encoding representation for the values on it (male and female). Hence, we use label encoding to encode the order-based features and the one-hot encoder to encode the unorder-based features.

Hyperparameter Optimization. In our experiments, we employ a random search technique [16] with 30 rounds of 5-fold cross-validation to identify the optimal hyperparameters for VACADOPT by generating random combinations of the hyperparameters. To investigate whether increasing the size of the k-folds used in cross-validation during randomized search influences predicted accuracy, we retrained the selected models using 10-fold and 15-fold cross-validation. The predictive performance of the 5-fold and both 10, 15-fold cross-validation models are only marginally different by around 1%.

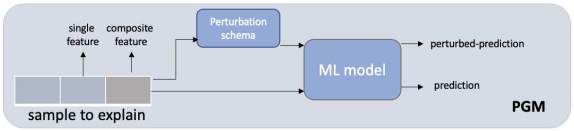


Fig. 2: Our Tabular-PGM explainer, adapting PGM-explainer [2] to work on tabular datasets.

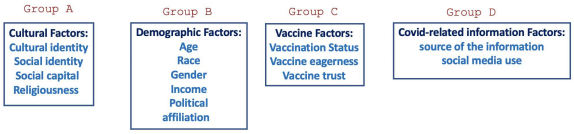


Fig. 3: The composite features are defined into distinct groups, including culture-based, demographic-based, vaccine-based, and COVID-19 related information factors. For simplicity, we have highlighted the key characteristics of each category.

Thus, we conduct our final analyses on the average of the three 5, 10, and 15 fold cross-validation models. The complete list of final hyperparameters for the six models is provided in [11].

C. Tabular-PGM explainer.

Methodologically speaking, the perturbation part is a crucial step in XAI approaches [17] to obtain successful explainers. Perturbations provide a window into a black-box of DNN models by examining the input-output connection and determining which portion of the input is given special weight by the model (importance value). The PGM-explainer is a perturbation-based technique that concentrates on the correlation between various perturbed inputs and model outputs in a model-agnostic fashion. However, the perturbation in PGM-explainer is not applicable to be used on tabular dataset due to the fact that the PGM is designed for GNNs.

Therefore, to explain our tabular datasets, we adapted PGM-Explainer to handle the tabular data, called Tabular-PGM explainer. Figure 2 shows our process to extract important feature values for a given sample, thereby giving an explanation in a form of which important features constitute a prediction of the model. The key difference between Tabular-PGM explainer and PGM-explainer lies in the perturbation scheme. In our Tabular-PGM explainer, we first set a variable $v \in (0, 1)$ representing a probability that the features in each record are perturbed. For each record, we introduce a random variable R indicating whether the features on v are perturbed. The perturbed sample is then passed to the ML model to generate **perturbed prediction**. The exact sample is also passed directly to the ML model to have the **prediction**. Finally, the important feature values are computed by observing the change in VACADOPT's output given the perturbed examples using the probabilistic model described in PGM [2]. The procedure is performed several times to

TABLE I: Performance Comparison.

Model \ Evaluation Metric	Accuracy %	F1-Score %	Precision %	Recall %
XGB Classifier - one hot encoding	87.3	92.2	90	94.3
XGB Classifier - label encoding	86.67	91.7	88.2	95
Decision Tree Classifier	82	88.5	90	87.3
KNN classifier	82.65	89	85	94.4
Random Forest Classifier - one hot encoding	86	92	88	96
Random Forest Classifier - (label encoding + one hot encoding)	**88.8**	**93.2**	88.5	**98**
SVM Classifier	87.13	91	**91**	92
TABNET	80	88	81.2	96

minimize the effect of random permutations, and the feature values are averaged across five runs.

Furthermore, since we are interested in knowing the influence of composite features on an individual's decision, we introduce a simple yet effective explanation method that adopts the PGM explainer to explain composite features. We merge multiple features into a single PGM node. In particular, we first define the composite features into distinct groups (Figure 3). Then, the features inside the groups are treated as sensitive features. In other words, the perturbation schema in Figure 2 is implemented based on the presence of sensitive features in a group. For example, if a group contains two sensitive features (composite features), we would perturb the two features together to learn about their representation and effects.

IV. Experimental Analysis

Our goal is to first evaluate ML models in terms of predictive accuracy (section IV-A). Next, we conduct an in-depth analysis (section IV-B) using SHAP on the top predictive model to investigate (i) the key features behind an individual decision about COVID-19 and (ii) the correlation between certain features and vaccination acceptance or rejection. Our attention in section IV-C is turned to focus on examining composite features using PGM-tabular to illustrate the influence of the related features, as the presence and effect of one feature could be derived from different factors. Section IV-D presents the details of identifying vaccination hesitancy characteristics of ethnic groups.

A. Predictive performance.

We begin by evaluating the performance of each detection model on the described dataset to determine its ability to predict user vaccination willingness using well-known evaluation metrics, including accuracy, precision, recall, and F1 score [8]. The evaluation and the comparison of the models are shown in Table I. TABNET performs the worst in comparison to the other approaches, as demonstrated in Table I (a higher number gives better results). This is because TABNET, a deep neural network, could not learn and represent characteristics that can aid in determining if a person was willing or reluctant to receive the vaccination. Deep neural networks, in general, do poorly with structured data. On the other hand, for the models designed to handle structured data (tabular dataset), we can see that the RF classifier (label+one hot encoding) outperforms XGB, the Decision Tree classifier, KNN, and SVM. For example, the RF classifier with hybrid encoding beats the second model in terms of the accuracy performance by about 1.7.

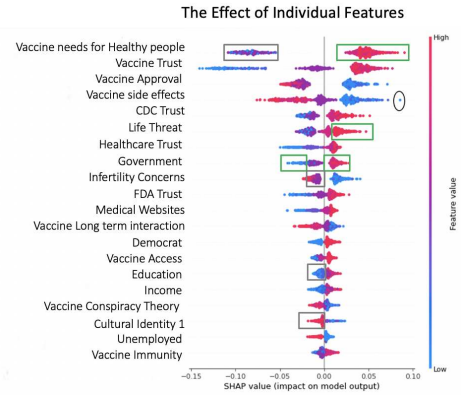
B. Features importance.

Providing a trustworthy and faithful explanation of a model's behavior requires an effective and successful predictive model. Therefore, we select the model with the top performance (RF model) to be explained. Next, we examine the key features behind an individual decision about the COVID vaccine adoption using SHAP (Figure 4a).

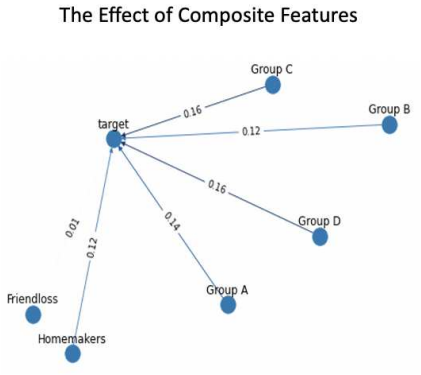
Figure 4a shows the top 20 significant factors that impact a person's decision, where the features are listed according to their total contribution (sum of SHAP values) to the final prediction, with the top feature being the most important. Each point corresponds to a person in the study. The position of each point on the x-axis shows the impact that feature has on the classifier's prediction for a given individual, with colors ranging from red (high feature values) to blue (low feature values). On the x-axis, positive SHAP values indicate a willingness to receive COVID-19 vaccination, whereas negative SHAP values indicate vaccine reluctance. For example, the individual surrounded by a circle in Figure 4a illustrates that the vaccine side effect is not a big concern for the person (low feature value (blue)). Hence, the person decides to receive the vaccination (positive SHAP values).

We highlight here some of the most important 20 features in Figure 4a (1-5 scale from strongly disagree to strongly agree): (1) Vaccine needs for healthy people: Do you think that COVID-19 vaccine is necessary for healthy people as well? (2) Vaccine Trust: Do you trust the COVID-19 vaccine? (3) Vaccine approval: Do not you trust the COVID-19 vaccine because of its current approval status? (4) Vaccine side effects: The possible side effects of the COVID-19 vaccine prevent me from receiving the vaccine. And (5) Conspiracy theory: an attempt to explain the COVID19 infection as the result of the actions of a minor, powerful group. Refer to [11] for more description of the features.

The vaccine-hesitant causes, using the SHAP explanations as shown in Figure 4a, can be divided into rational-based and irrational-based. The rational-based features are those that are debatable and understandable from the perspective of their health association. For example, the Food and Drug Administration (FDA) has authorized health organizations to

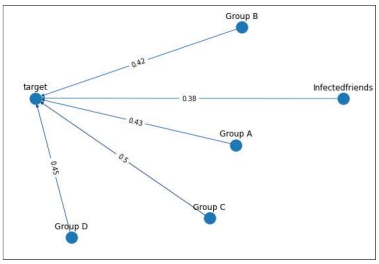


(a) SHAP interpretation of the RF model

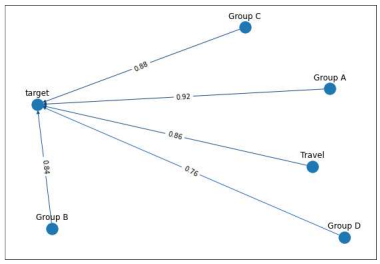


(b) The adopted PGM interpretation of the RF model.

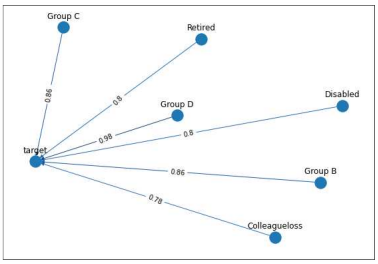
Fig. 4: **The top important features and composite features.**



(a) Asian Group



(b) African American



(c) Hispanic Group

Fig. 5: **The top important features/composite features for three ethnic groups, including Asian, African American, and Hispanic.**

use COVID-19 vaccines only in an emergency setting during the early phase of the pandemic. Unfortunately, such usage restrictions make individuals doubt the vaccine's effectiveness, thus refusing it (FDA trust in Figure 4a). The rapid development of the COVID-19 vaccine also raises key vaccine hesitancy issues such as vaccine trust, possible side effects, or long-term interaction that may pose a threat to people life, as illustrated in the RF interpretation in Figure 4a. If the right circumstances arise, rational-based cases can lead to vaccination. Several reasonable justifications are highlighted in green, shown in Figure 4a.

The irrational-based features, on the other hand, are those related to (1) cultural characteristics, (2) rumors, and (3) political affiliation. Although conspiracy theories (considered as a specific type of rumors) are logically and scientifically implausible, they significantly influence vaccine adoption, as shown in Figure 4a. Another key feature is an individual's political background, or belief in governments. The uncertainty of certain groups might express itself in vaccination reluctance. More details can be seen in Figure 4a.

Looking closer, we notice that irrational-based features are strongly associated to the vaccine rejection. For instance, the majority of people who feel negative about their ethnicity (low feature value) refuse to obtain the vaccine. Some of the irrational-based features are highlighted in gray in Figure 4a to offer more insights into vaccine rejection. Overall, cultural characteristics, rumors, and political affiliation are associated with increased vaccine rejection.

C. *Composite features importance.*

We now turn our attention to the composite factors instead of evaluating them independently. Grouping the characteristics can be used to illustrate the combined strength and influence of the related features, as the presence and effect of one main feature could be derived from different factors. For instance, cultural identity is developed and sustained via the exchange of common knowledge [18]. Hence, a person's cultural identity cannot be formed only based on a single variable; rather, it can be formed based on numerous characteristics. Along this direction, we concentrate on the four main groups, as shown in Figure 3. As long as the characteristics are relevant to one of the four major categories, they are assigned to that group.

Figure 4b shows the PGM explanation for individual/composite features in the form of a graph network. The weights on the edges reflect the strength of the dependency connection between the target (receiving/refusing the vaccine predictions) and other nodes. The dependency connection shows how strongly that feature contributes to a prediction.) A high value indicates strong connections (dark blue), whereas a low value represents low connections (light blue).

As shown in Figure 4b, the composite features in human health (represented by group C), such as COVID-related concerns, fear, vaccine trust, fertility problems, and side effects, remain the most prominent influences on vaccine adoption. However, the low vaccination rate has been motivated not only by significant healthcare concerns but also by the dissemination of various rumors and disinformation regarding the COVID-19 vaccine and its causes, as can be seen in Figure 4b (group D has the same strength as group C). Differentiating and mitigating other rumors (group D) that take advantage of the government's long history of injustice can aid in the development of particular recommendations to boost vaccination rate acceptability.

Such a rise of social media misinformation (group D) as one of the primary sources or cultural factors (group A) as the second main composite feature for determining vaccination willingness decisions is a significant concern. Further investigations are needed to limit the misinformation and approach individuals with cultural-based attitudes as they impact by the people and environment.

D. Ethnic groups features importance.

Recognizing vaccination hesitancy characteristics of ethnic groups is crucial to understanding each group's concerns and need to accept the vaccine. Thus, for each ethnic group, we extract all data from the trained RF model that belong to that group and use Tabular-PGM explainer to analyze the key influences. As shown in Figure 5a, the vaccine factors (represented by group C) significantly influence vaccination adoption for Asian people. In contrast, the African American group is primarily influenced by cultural factors (represented by group A in Figure 5b), such as cultural identity (e.g., group affiliation), social identity (e.g., socioeconomic status), and religiosity (e.g., following certain religions), whereas COVID-related information factors (represented by group D in Figure 5c), such as social media posts, are most important factors to the Hispanic group. We can also observe that the vaccination factors are the second most important composite feature for both the African American and Hispanic individuals (group C in Figures 5b and 5c). This indicates that the trust in the vaccine is still a crucial concern for both groups but not as much as the cultural factors for African Americans and COVID-related information factors for Hispanics.

These issues, when considered together, are well-established causes of mistrust aimed towards governments and the healthcare system, which impact vaccination uptake among various ethnic groups. The various key patterns among different ethnic groups provide insights into new strategies and methods for building trust, increasing collaboration, creating tools and resources to respond to the concerns and feedback from various ethnic groups. These efforts, in conjunction with scientifically backed information, can serve to boost COVID-19 vaccine acceptability.

V. Conclusion

This paper proposes a culture-aware machine learning (ML) model to predict vaccination willingness based on our new data collection. We adopt the advanced AI explainer (PGM) to work on a tabular dataset. We further analyze the most important features and composite features that contribute to the ML model's predictions using advanced AI explainers. Our findings show that Hispanic and African Americans are most likely impacted by cultural characteristics, whereas the vaccine factors influence the Asian communities the most.

Acknowledgment

This work was supported in part by the National Science Foundation Program under award No. 1939725

References

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *NIPS*, 2017.
[2] M. Vu and M. T. Thai, "Pgm-explainer: Probabilistic graphical model explanations for graph neural networks," *NIPS*, 2020.
[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.
[4] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, 1991.
[5] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, 2009.
[6] L. Breiman, "Random forests," *UC Berkeley TR567*, 1999.
[7] R. Alharbi, M. N. Vu, and M. T. Thai, "Evaluating fake news detection models from explainable machine learning perspectives," in *ICC 2021-IEEE International Conference on Communications*, 2021.
[8] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *SIGKDD*, 2019.
[9] J. Kees, C. Berry, S. Burton, and K. Sheehan, "An analysis of data quality: Professional panels, student subject pools, and amazon's mechanical turk," *Journal of Advertising*, 2017.
[10] Y. Kim, A.-F. Kassam, I. E. McElroy, S. Lee, A. Tanious, E. L. Chou, S. S. Patel, A. A. Pendleton, and A. Dua, "The current status of the diversity pipeline in surgical training," *The American Journal of Surgery*, 2021.
[11] R. Alharbi, S. Chan-Olmsted, H. Chen, and M. T. Thai, "Cultural-aware," https://github.com/raed19/Cultural-aware, 2022.
[12] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in *CVPR*, 2006.
[13] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *AAAI*, 2021.
[14] K. J. Berry, P. W. Mielke Jr, and H. K. Iyer, "Factorial designs and dummy coding," *Perceptual and motor skills*, 1998.
[15] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *Journal of Big Data*, vol. 7, no. 1, pp. 1–41, 2020.
[16] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization." *Journal of machine learning research*, 2012.
[17] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, 2018.
[18] A. J. Umaña-Taylor, A. Yazedjian, and M. Bámaca-Gómez, "Developing the ethnic identity scale using eriksonian and social identity perspectives," *Identity: An international journal of theory and research*, 2004.