Machiavellian Expression and Relational Classification

Final Project Report for DATA 1030, Fall 2021

https://github.com/ktoleary13/mach_rel_classification.git

Katie O'Leary

1. Introduction

Within personality studies, "The Dark Triad" describes sub-clinical yet offensive traits. These personality traits are non-pathological but colloquially negative descriptors: narcissism, psychopathy, and Machiavellianism. Machiavellianism, the manipulative personality, describes a person's perception and willingness to neglect others for personal gain. Here, a psychometric dataset containing Machiavellian, personality, and demographic surveys were used to build a classification model predicting relationship status where the target variable describes a participant as married, never married, or previously married. The ability of individuals to form meaningful relationships impacts their lifelong flourishing and can be influenced by negative personality traits like Machiavellianism [1]. Thus, with a classification model based on manipulative tendencies, relationship outcomes could be predicted and/or avoided.

Containing seventy thousand data points, this dataset houses one hundred and five original features. The features are Machiavellian test questions (MACH IV), a personality assessment (TIPI), a vocabulary test, and demographic variables collected from 2017-2019. Each of these is well documented in a text file provided by the "Open Psychometrics" website. Though the content is identical, the dataset was accessed through Kaggle instead [2], where a single public project had been published. This previous project attempted machine learning but is incomplete with no substantial outcomes. Additionally, this dataset was utilized in a thesis submitted to Australian National University [3]. While validating the data rigor and usability, that work does not inform the proposed machine learning model here but discusses factor analysis in personality studies.

2. Exploratory Data Analysis

During data exploration, the data was winnowed to suit project goals. For example, because marriage status is the target variable, datapoints with improper ages were immediately filtered out to twenty-six thousand points [4]. Additionally, raw questionnaires needed to be scored manually, and repetitive or irrelevant information was removed (e.g. screenwidth). After this preparation, forty-four variables remained and displayed several trends.

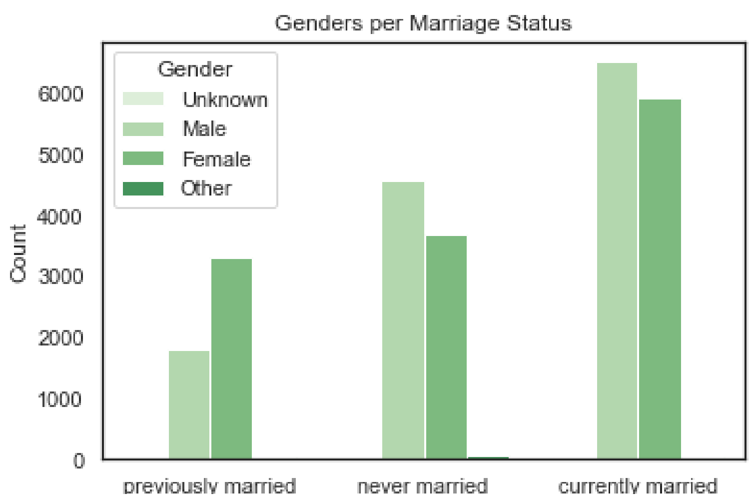

Figure 1. Bar chart of relationship status target variable. Indicates that currently married individuals are the most represented and that males are dominant in never married and curriently married classes.

First, the proportion of datapoints in the target variable indicate that currently married individuals are the most represented (Figure 1). As demonstrated in Figure 2, the never married category skews younger than the other categories despite controlling for average age of marriage. While it is unsurprising that increased age increases opportunity and likelihood for marriage/previous marriage, the influence of age should be noted. Also, consider the direct relationship between MACH IV test scores and relationship status is distinct, but not obviously so (Figure 3). Distributions reflect the population averages, but the means and quartile ranges differ more substantially (Figure 3). Further, these differences may expand in conjunction with personality test results [5]. For example, Figure 4 shows a distinction in MACH IV scores with signifiers of self-assurance.

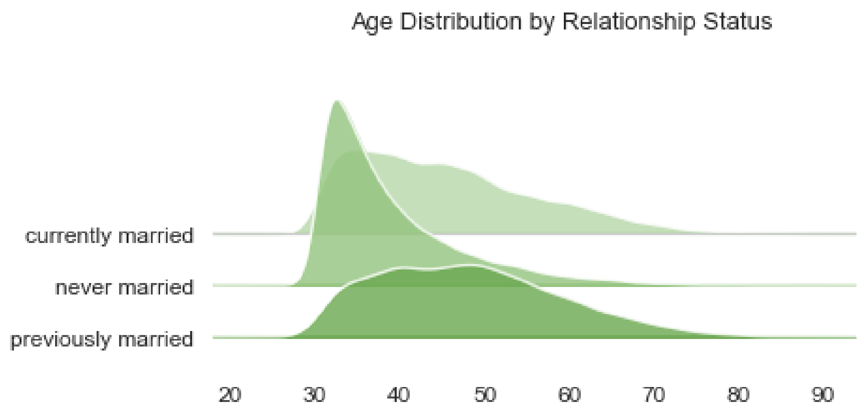

Figure 2. Distribution of each relationship versus age of those individuals. Currently married and never married individuals skew young, which reflects the role of age in life partnerships.

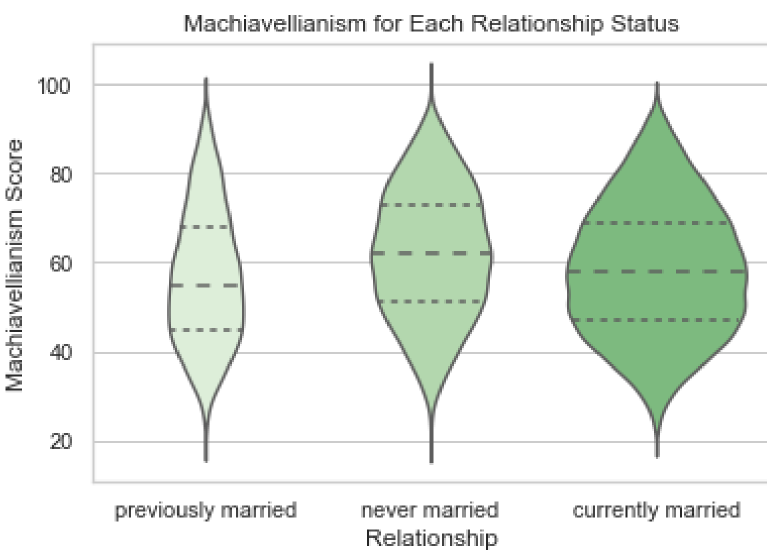

Figure 3. Violin plot with quartile ranges of MACH IV scores for each relationship status. Never married individuals show highest means and quartile ranges.

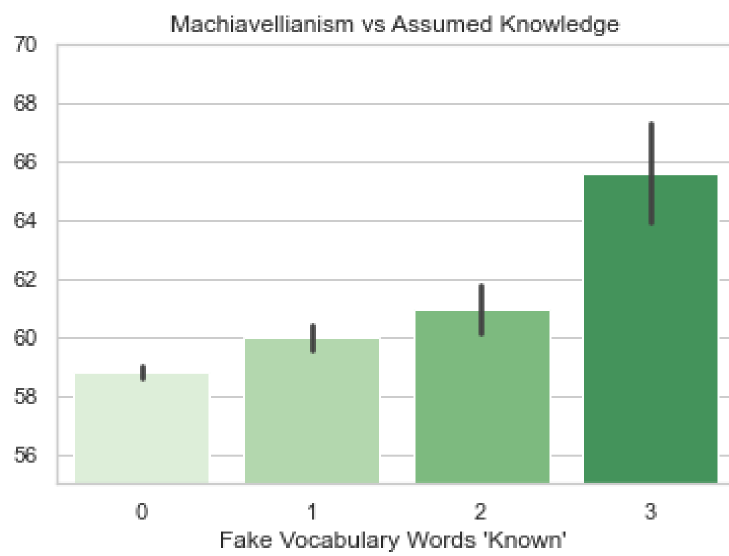

Figure 4 Bar chart of MACH IV scores for number of fake voacbulary words known. The more words 'known', the higher the MACH IV score.

3. Methods

3.1 Preprocessing

The dataset was split and processed as a non-IID dataset. The approach was based on the fraction of each represented class in the target variable. While the least represented class is still over fifteen percent, a stratified, shuffled, K-fold split was implemented because of the difference in predicted class proportions and associated ages. This accounts for the variation represented in the features of each class and minimizes the chance of overrepresenting young ages in any one split. The fractions of training, validation, and testing sets were 0.75, 0.19, and 0.06 respectively. Notably, all models were initially trained on a 500-point subset of the data to maximize the speed of implementing a grid search. However, the subset was evaluated for target class representation and baseline score to ensure the results were translatable to the whole.

Encoding and preprocessing the forty-four features was simplified by the nature of the dataset. Many of the personality assessments[1] (e.g., TIPI) and all the demographic variables were recorded on a categorical scale. All these[1,2] were encoded with OneHotEncoder. While time-oriented variables were encoded with StandardScalar, the MACH IV score (20-100 scale) and age of individuals were encoded with MinMaxEncoder. After preprocessing, there were 117 features.

3.2 Model Selection

Four machine learning algorithms were applied to the dataset: Logistic Regression, Random Forest, K-Nearest Neighbors, XGBoost. For Logistic Regression, the regularization parameter was only tuned according to L_2 regularization because of the large set of parameters that may have predictive power [6]. It was not considered desirable for feature coefficients to be set to zero. Even so, all models were tuned through a brute force grid search with the search spaces defined in Table 1. In addition to a grid search, XGBoost's early stopping rounds were

Table 1. Models tuned, the parameters tuned/selected, and the optimal value ascertained from the search space. Note that the optimal value was based on the most frequent best parameters

Model	Hyper-param	Search Space	Optimal Value
Logistic Regression	L_2 Regularization	$[10^3, 10^2]$	2.15
	Solver	saga	saga
Random Forest	Maximum Depth	[1,300]	100
	Maximum Features	[0.5,1]	0.6
K-Nearest Neighbors	Number of Neighbors	[1,80]	10
	Weights	[Uniform, Distance]	Uniform
XGBoost	Maximum Depth	[1,300]	20
	L_1 Regularization	[0,100]	0
	L_2 Regularization	[0,100]	0.1
	Early Stopping	60	60

[1] MACH IV Questions ('Q1A', … 'Q20A') , TIPI Scores, and Education)
[2] Demographic Variables: ('race', … 'religion'), Vocabulary responses: ('voc_fake', 'voc_conf')

determined by finding the minimum distance and minimum error between validation and training curves.

The models were evaluated on a balanced accuracy metric. Balanced accuracy disregards the proportion of classes and considers the macro average of recall scores per class. In this project, no target class is more important and there is no greater risk associated with false positives versus false negatives. Still, the dataset did present moderate imbalances. Thus, the priority was simply to mitigate these differences and prioritize correctness through balanced accuracy. However, while XGBoost reported test scoring in balanced accuracy, validation sets were evaluated using log-loss. During hyperparameter tuning, log-loss is standard for multiclassification problems in XGBoost [7,8] so this was retained to prevent errors and maximize performance.

The influences of randomness during splitting and non-deterministic methods were quantified by recording the mean and standard deviation of multiple model outputs. In Table 2, Logistic Regression, Random Forest, and K-Nearest Neighbors reported balanced accuracies above the baseline, which is always equal to the fraction of one over the target classes. However, these scores were far below any useful predictive power. Alternatively, XGBoost demonstrated substantial improvement from both these scores and the baseline. Because of this, XGBoost was selected as the final model and was retrained using the best hyperparameters on new random states with the same splitting strategy.

Table 2. Mean and standard deviation of metric. XGBoost: 4 random states, 16 outputs. Others: 5 random states, 5 outputs. Baseline has no standard deviation, but is always the fraction of classes in the dataset

Model	**Balanced Accuracy**	**Outputs**
Logistic Regression	0.64 ± 0.05	5
Random Forest	0.66 ± 0.18	5
K- Nearest Neighbors	0.69 ± 0.15	5
XGBoost	0.85 ± 0.08	16
Baseline Score	*0.333*	

4. Results

4.1 Performance Evaluation

All models exceeded the baseline balanced accuracy score average by at least 30% (Table 2). However, XGBoost’s average was 6.5 standard deviations from the baseline average, which made it the best by far. The final model performance is visualized in Figure 5 by compiling four test set predictions into a confusion matrix. Each box indicates the number of points represented and the percentage of those points to the whole. Notice that the most abundant class, “Currently Married,” has the highest accuracy and the most incorrect predictions while “Previously Married” is the least abundant and least accurate. This describes the unsurprising trend that the model performs the best on the most abundant class.

4.2 Model Interpretation

To interpret the model, four global feature importances and Shapely (SHAP) values were calculated. Yet, during this investigation, many demographic variables were the most predictive, not personality metrics. Further, it was noteworthy that of the four global importance measures implemented, they collectively ranked demographic variables higher than personality metrics (Figure 6). To compare the influence of demographics versus personality, the top fifteen features of permutation, gain, cover, and weight importances were collected and searched for uniqueness. Of a possible sixty, forty features were unique. Thus, there was substantial variation between importance types so that the model relied on most features for some predictive power. In fact, of the 117 features, only ten of them had no global importance to the model predictions. These included family sizes greater than twelve, "Unknown Gender," "Unknown Native Language," "Indigenous Australian," and "Sikh" religion.

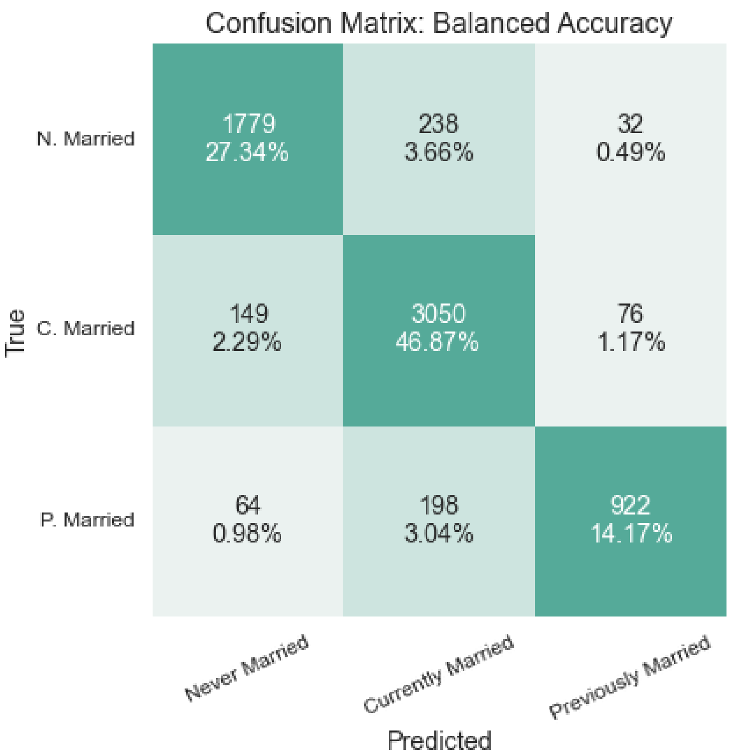

Figure 5. Confusion matrix of final model output using 4 iterations. The overall accuracy is 88.38%. Each independent class accuracy: "Currently Married" class average (93.1%) is above the the total average, "Never Married" (86.8%) is close to the total average, and "Previously Married" (83.7%) is below the total average.

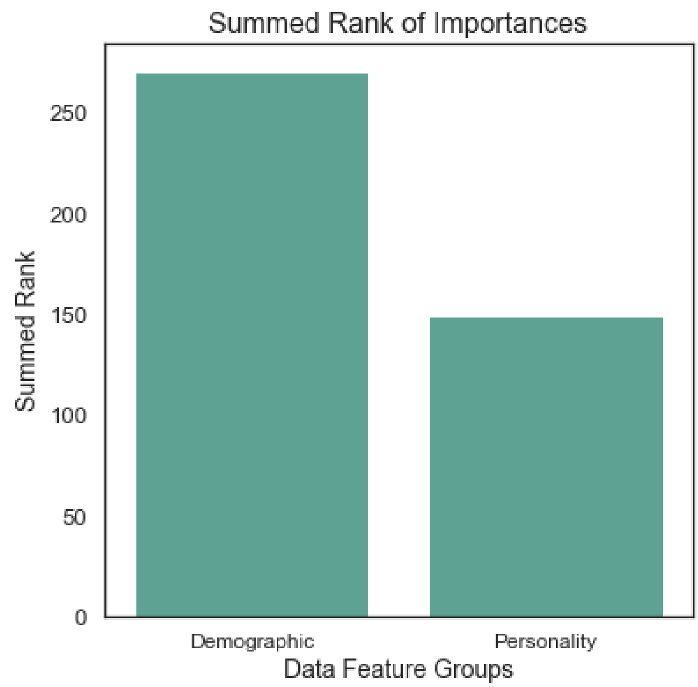

Figure 6. Bar chart of global feature improtance summed across weight, cover, gain, and permutation. The rank of each feature importance (15 – most important of top 15, 1 - least important of top 15) was collected, summed, and categorized by feature type. Demographic variables were more highly ranked summed across the four types of global importance.

Considering the context of the problem, these results were unsurprising. In Figures 7,8,9 the distribution of important features across target classes consistently shows that demographic variables are at least the top three most important features when using SHAP values. Considering that age, sexuality, and religion are heavy influences on behaviors, it is anticipated that they would influence marriage status. Still, one surprising feature importance is that being female was a predictor for *not* being currently married. While more women populated the previously married class, men were the majority in "Currently Married."

Regarding Machiavellianism, the MACH IV and time taken to answer questions were important for every target category, but none of these features tended to contribute extreme importance values. In fact, these features cluster at the boundary of predictive influence. In contrast, individual questions of the MACH IV were deemed less important (Figure 7,8,9), but were far more polarizing in how their answers influenced the ultimate prediction. All individual questions were from the MACH IV positive categories and were sometimes more important than the total test score. This is especially interesting, noting that not being positive consistently predicted no/previous marriage, but no other category of questions was as highly influential.

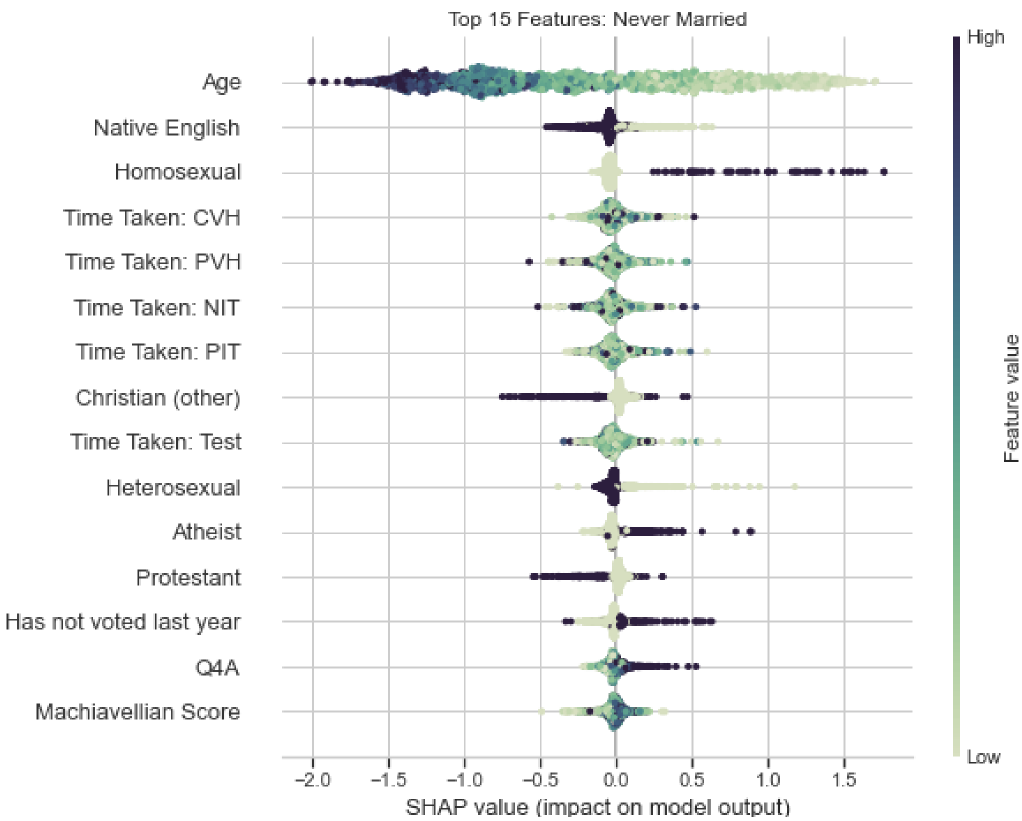

Figure 7. Shapely values for top 15 most important feautres for "Never Married" in target class.

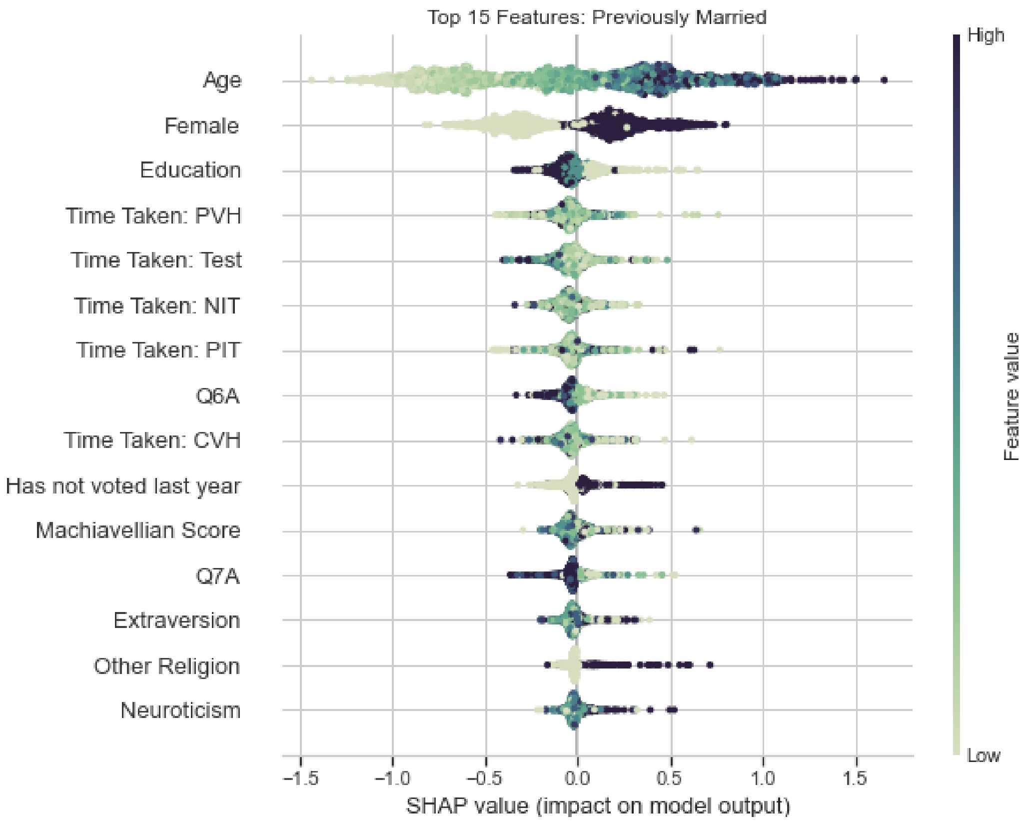

Figure 8. Shapely values for top 15 feautres for “Previously Married” in target class.

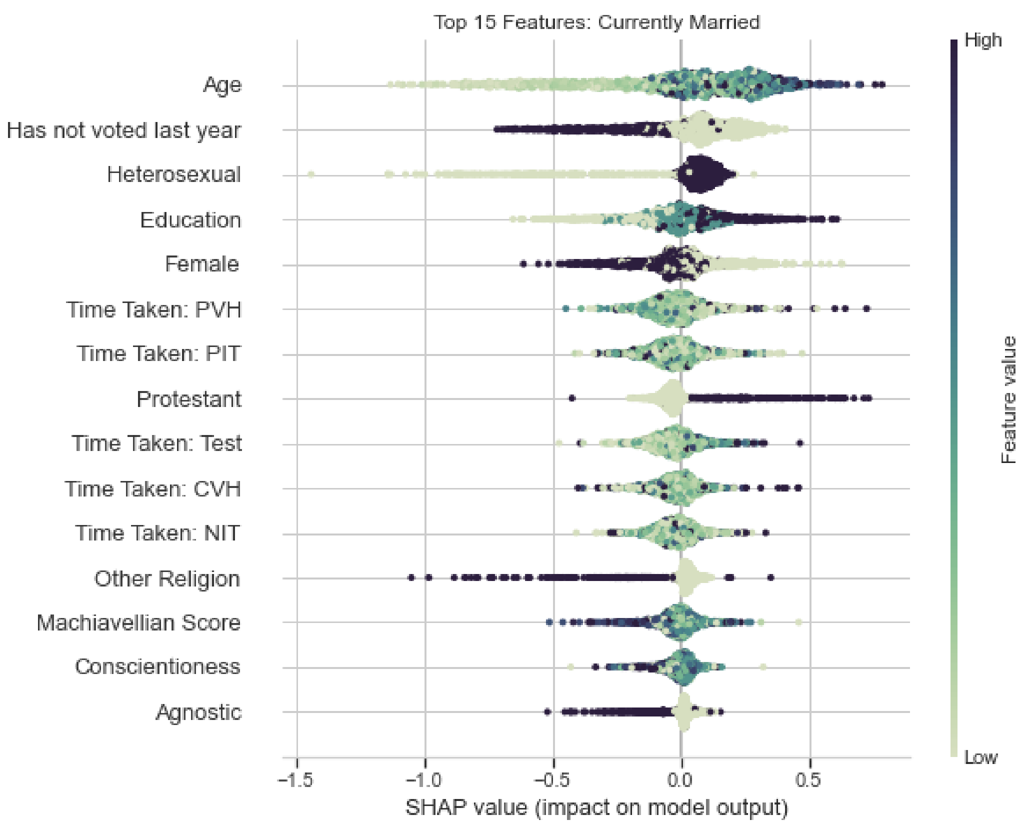

Figure 9. Shapely values for top 15 most important feautres for “Currently Married” in target class.

5. Outlook

Given this XGBoost model, the results are interpretable and accurate, but not entirely clear. The model incorporates sensible demographic and personality features into the final prediction. Both feature categories appear important in the final model and may be useful in making predictions of how the status of a person with some current disposition may evolve in the future. However, some amount of group structure seems present. The influence of demographics (e.g., orientation) across relationship status suggests that the model does not represent personality directly. Thus, to improve progress toward the goal, incorporating the most recent data and controlling for obvious groups would be best. However, considering demographic features' predictive influence, the overall accuracy may decrease. While the intent was to primarily use personality features to predict relationship, these were only moderately influential. Finally, because a person's personal life is both important and private, this model would have to be far more accurate before deploying to anything more serious than a leisure internet survey. Nevertheless, these results emphasize the validity in making connections between Machiavellianism and relationship by demonstrating some amount of connection.

IV. References

[1]:Kern, Margaret L., et al. “Lifelong Pathways to Longevity: Personality, Relationships, Flourishing, and Health.” *Wiley Online Library*, John Wiley & Sons, Ltd, 16 Oct. 2013,

[2] Greenwell, Lucas. “Machivallianism Test Responses.” *Kaggle*, 28 May 2020, https://www.kaggle.com/lucasgreenwell/machivallianism-test-responses.

[3] Monaghan, Conal, “Two-Dimensional Machiavellianism: Conceptualisation, Measurement, and Well-Being”. *Thesis to Australian National University* 2019/01/0 https://doi.org/10.25911/5d149b879161d

[4] “Median Age at First Marriage: 1890 to Present.” *Census.gov*, United States Census Bureau , 2020, https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/families-and-households/ms-2.pdf.

[5] Paulhus, Delroy L, and Kevin M Williams. “The Dark Triad of Personality: Narcissism, Machiavellianism, and Psychopathy.” *Journal of Research in Personality*, Academic Press, 19 Nov. 2002,

[6] Bhattacharyya, Saptashwa. “Ridge and Lasso Regression: L1 and L2 Regularization.” *Medium*, Towards Data Science, 28 Sept. 2020, https://towardsdatascience.com/ridge-and-lasso- regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b.

[7] “XGBoost Parameters.” *XGBoost Parameters - Xgboost 1.5.1 Documentation*, https://xgboost.readthedocs.io/en/stable/parameter.html.

[8]“Sklearn Metrics: log_loss.” *Scikit Learn*, https://scikitlearn.org/stable/modules/generated/sklearn.metrics.log_loss.html.