

Student Depression Classification using Machine Learning Methods and Analysis

By: Daniel Ginzburg, Isaac Berlin, Anthony Krenek
University of Minnesota - Twin Cities

ABSTRACT

This study investigates the potential gender-based differences in predicting student depression using machine learning models. A student depression dataset was analyzed, with the data split by gender to assess model performance variations and the importance of different features. We implemented three commonly used machine learning algorithms: Logistic Regression, Decision Tree, and Neural Network, to predict depression levels in students. The models were trained on features such as academic performance, social interaction, lifestyle factors, and mental health indicators. Performance metrics, including accuracy, precision, recall, and F1-score, were computed for each model to evaluate their effectiveness in predicting depression. Additionally, feature importance was assessed to determine if certain factors played a more significant role for different genders. The results did not display major differences in depression prediction or feature importance between the two genders.

Previous Work

Several studies have explored the use of machine learning algorithms to predict mental health conditions such as depression and anxiety. For instance, a study by Islam et al. (Nayan et al. 75) compared the performance of various machine learning models in predicting depression and anxiety among university students in Bangladesh. The authors used six different models and found that Random Forest performed the best with an accuracy of 89%. The study was also able to highlight gender differences with female students showing higher rates of depression and anxiety.

Nayan, Md. Iqbal Hossain, et al. "Comparison of the Performance of Machine Learning-Based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh: A Result of the First Wave of the COVID-19 Pandemic." Asian Journal of Social Health and Behavior, vol. 5, no. 2, 2022, pp. 75–84. https://doi.org/10.4103/shb.shb_38_22.

METHODOLOGY

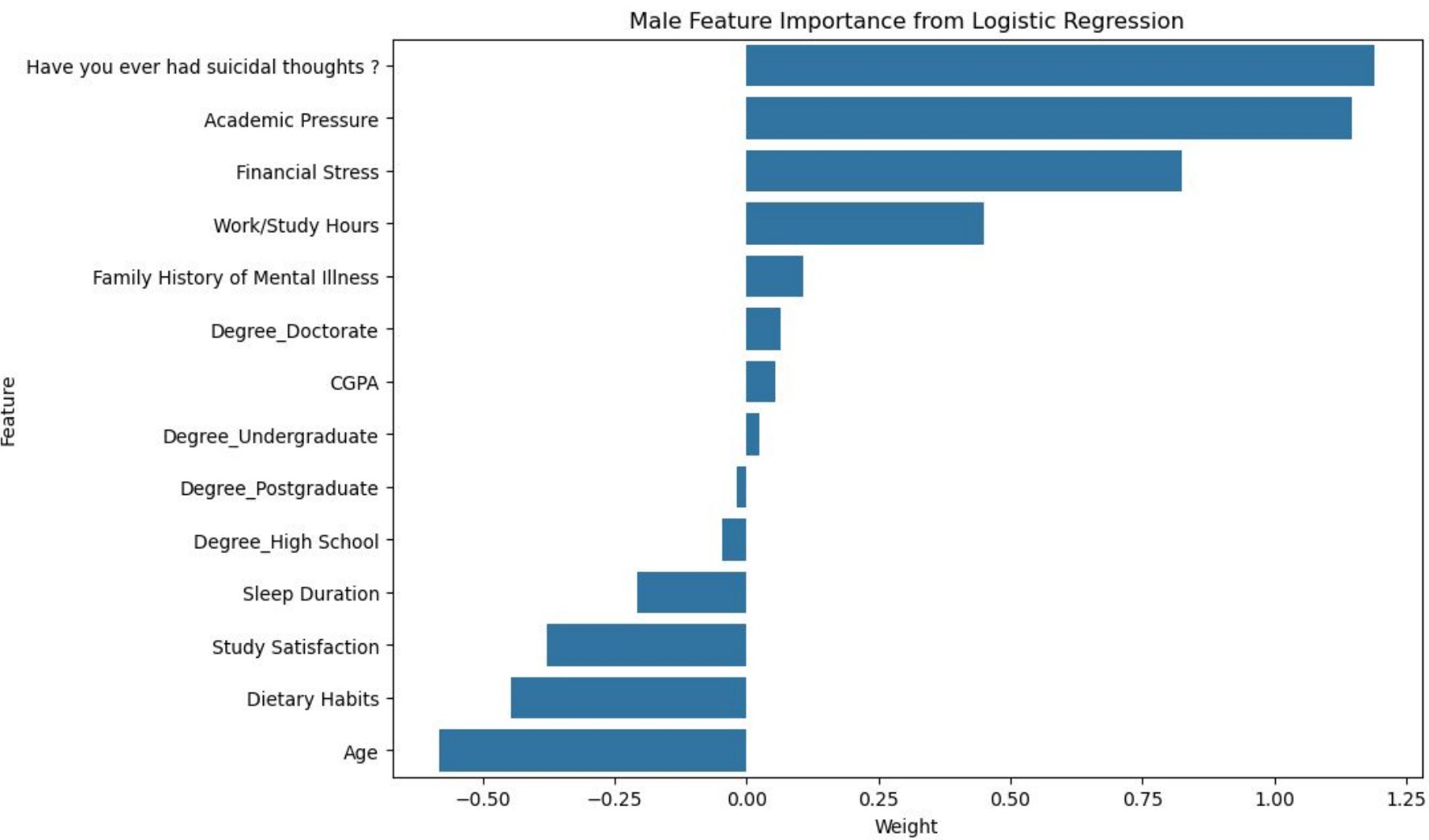
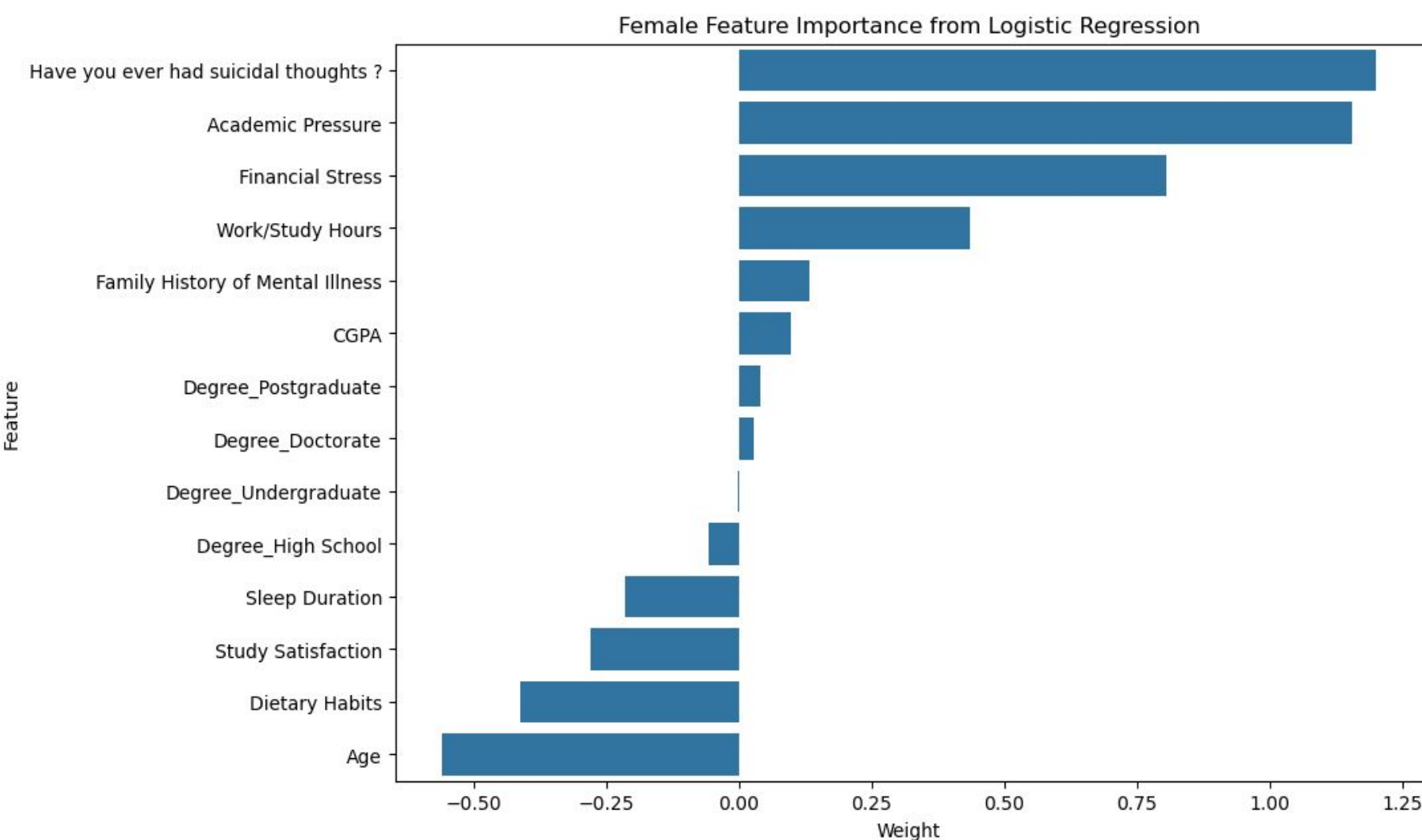
Logistic Regression was used as a baseline linear model to identify key predictors of depression. We used binary classification as well for Male and Female for our dataset. It allowed us to interpret feature weights directly, highlighting academic pressure, study satisfaction, and work/study hours as influential factors.

Decision Tree is a non-linear, rule-based approach. The decision tree offered intuitive paths that helped us understand how combinations of features like high academic pressure and low study satisfaction and how they increase depression risk.

MLP (Multi-Layer Perceptron) was used as a simple feedforward neural network to capture more complex, non-linear relationships between features. It consisted of one hidden layer with ReLU activation and was trained using binary cross-entropy loss.

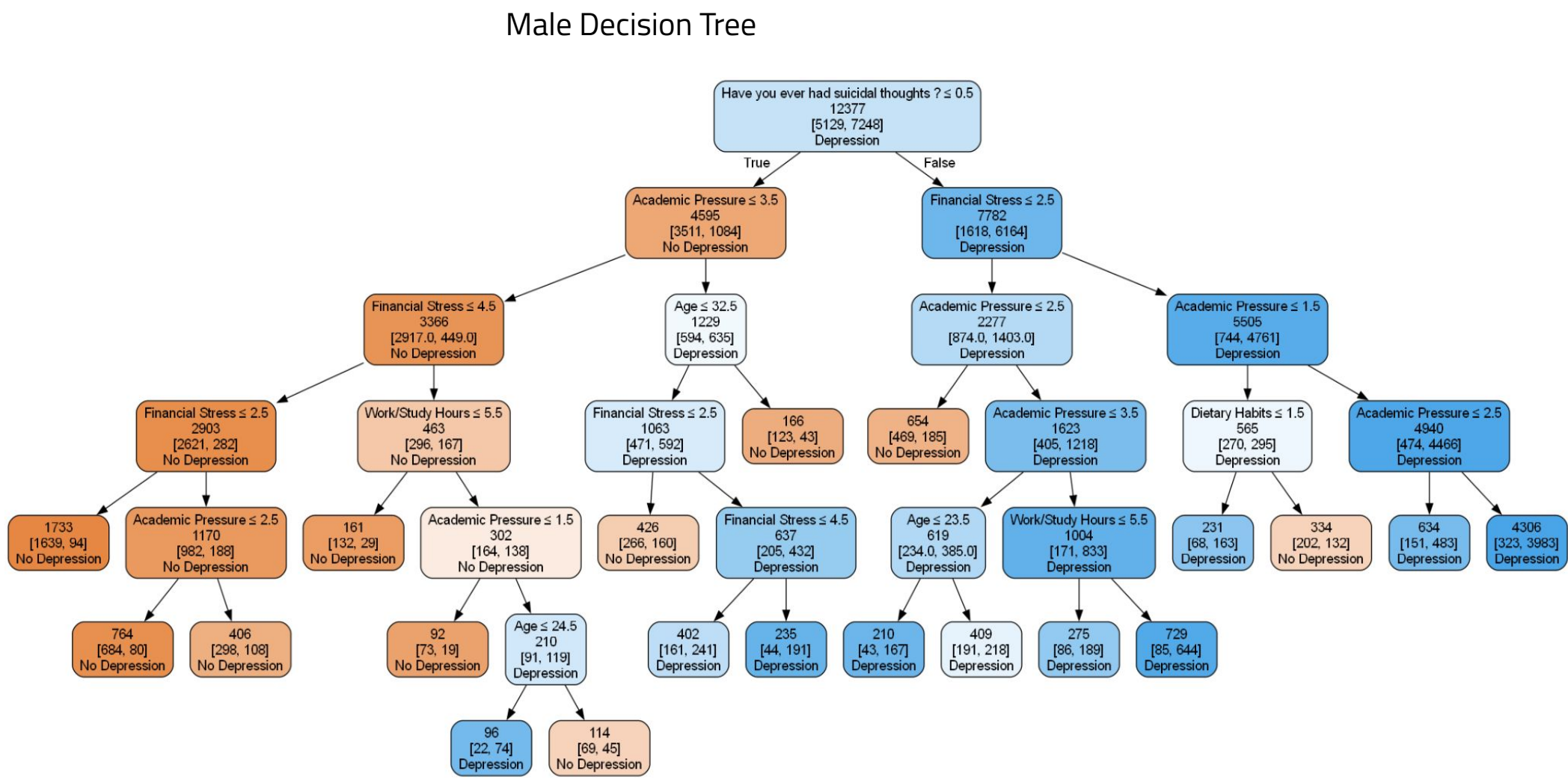
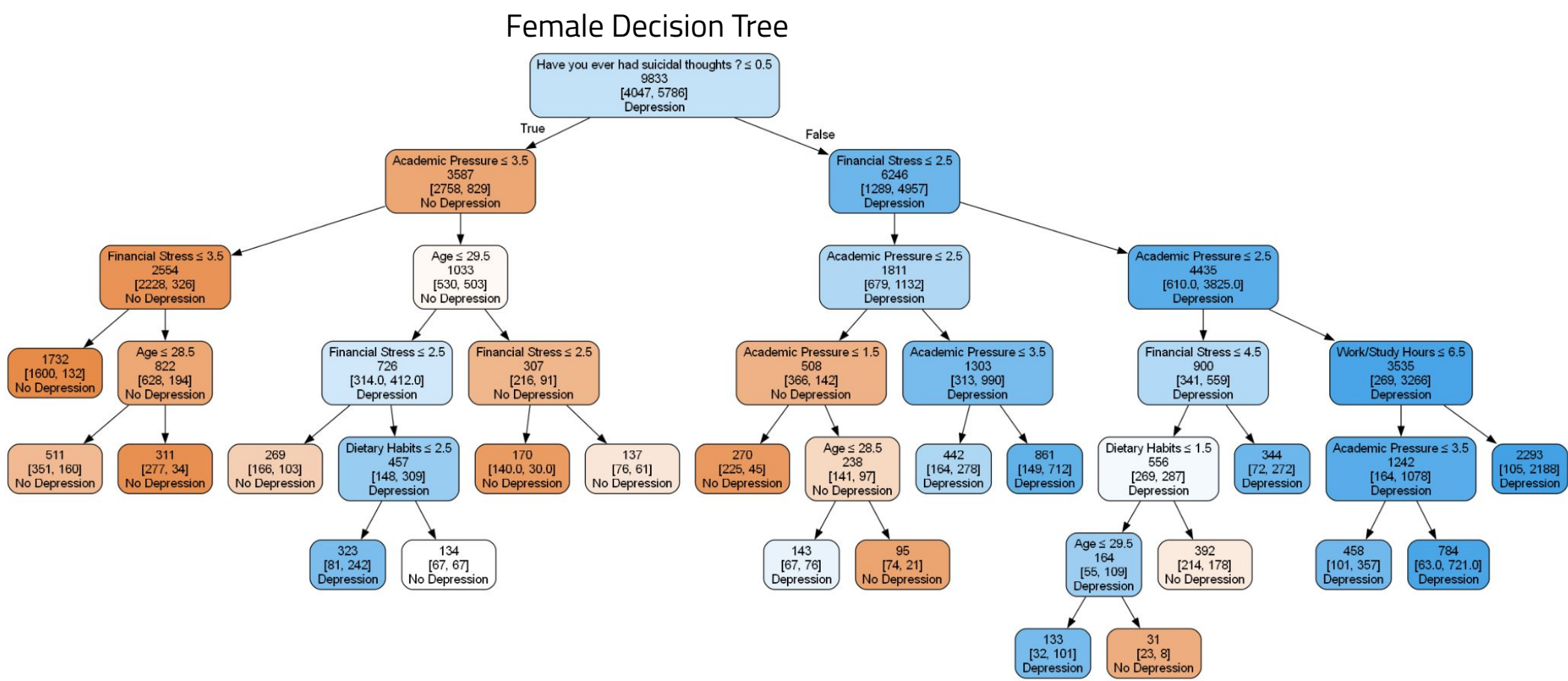
RESULTS

Below we display some of our initial findings. For our logistic regression model we display the relative feature importance of the weights. We notice that suicidal thoughts, academic pressure, and financial stress are the most strongly related to depression, while age and diet are negatively related.



	Model	Gender	Accuracy	Precision	Recall	F1	ROC AUC
	MLP	Female	0.850	0.851	0.890	0.870	0.924
	Logistic Regression	Female	0.848	0.852	0.887	0.869	0.926
	Decision Tree	Female	0.825	0.841	0.852	0.847	0.896
	MLP	Male	0.840	0.844	0.894	0.868	0.915
	Logistic Regression	Male	0.845	0.847	0.898	0.872	0.917
	Decision Tree	Male	0.826	0.833	0.879	0.855	0.878

Next, we visualize decision tree models. We once again notice suicidal thoughts, financial stress, and academic pressure are the driving cause of depression. We also see that work study hours plays a significant part.



CONCLUSION

Based on our results we were unable to conclude a significant difference between male and female depression indicators. Both the prediction accuracy as well as feature importance were approximately the same for the groups. Additionally, the performance of our three models was comparable, with each achieving similar results via evaluation metrics.