Applying ML - DSA 210 – Muhsin Efe Tarım – 32230

To evaluate different machine learning approaches on the depression dataset, we applied three models per task (regression and classification): Linear/KNN/Decision Tree for regression and Logistic/KNN/Decision Tree for classification. These models were selected for their balance between interpretability and performance.

**REGRESSION:**

In the regression task, the goal was to predict the continuous depression score (PHQ-9) of university students based on various academic, psychological, and lifestyle-related features. This allowed us to explore how different factors quantitatively contribute to the severity of depression symptoms, using models such as Linear Regression, KNN Regressor, and Decision Tree Regressor.

**linearRegression:**

As a baseline regression model, Linear Regression was employed to understand the linear relationships between various student-related features (e.g., academic pressure, financial stress, and sleep duration) and their corresponding depression scores. The model assumes additive and independent contributions of each predictor to the overall mental health outcome.

*Mean Squared Error (MSE): 0.1595276477424041*

*R-squared (R²): 0.34222943346651413*

A graph with blue rectangles

AI-generated content may be incorrect.

The model achieved an R² score of 0.342, indicating that approximately 34% of the variance in depression scores can be explained by the selected features. The coefficient plot shows that academic pressure and financial stress were the strongest positive contributors, while study satisfaction and sleep duration had the most notable negative effects.

**K Nearest Neighbours (KNN) Regressor:**

The K-Nearest Neighbors (KNN) Regressor was applied to explore whether students with similar characteristics (e.g., academic and psychological factors) tend to have similar depression scores. Unlike parametric models, KNN makes predictions based on local similarity without assuming a specific functional form.

*KNN Regression - Mean Squared Error (MSE): 0.18507890961262555*

*KNN Regression - R-squared (R²): 0.23687548238738998*

The model achieved a Mean Squared Error (MSE) of 0.1851 and an R² score of 0.237, indicating that it explained approximately 24% of the variance in depression scores. Compared to Linear Regression, KNN performed worse, suggesting that depression scores in this dataset are better captured by global linear relationships rather than local proximity patterns.

**Decision Tree Regressor:**

The Decision Tree Regressor was utilized to capture potential non-linear relationships and hierarchical decision patterns between student-related features and depression scores. This model recursively splits the data based on feature thresholds, enabling interpretable, rule-based predictions.

*Decision Tree Regression - Mean Squared Error (MSE): 0.1657206989803403*

*Decision Tree Regression - R-squared (R²): 0.31669400510035395*

After limiting the tree depth to reduce overfitting, the model yielded a Mean Squared Error (MSE) of 0.1657 and an R² score of 0.317, indicating that it explained approximately 32% of the variance in depression scores. Its performance was comparable to Linear Regression, making it a viable alternative with added interpretability through its decision paths.

**RESULT:**

Among the three regression models tested—Linear Regression, K-Nearest Neighbors (KNN) Regressor, and Decision Tree Regressor—Linear Regression demonstrated the best overall performance, achieving the lowest Mean Squared Error (MSE = 0.1595) and the highest R-squared value (R² = 0.3422). This indicates that a global linear model most effectively captured the relationship between the selected features and students’ depression scores.

While the Decision Tree Regressor also performed well (R² = 0.317), its performance was slightly lower and required regularization (limiting tree depth) to avoid overfitting. The KNN Regressor, on the other hand, underperformed compared to both, suggesting that local similarity-based prediction was less effective in this context.

Therefore, Linear Regression was selected as the preferred model for regression tasks due to its superior predictive accuracy, simplicity, and interpretability.

A screenshot of a computer

AI-generated content may be incorrect.

While regression models aimed to predict the depression \*\*score\*\* as a continuous outcome, classification models focused on predicting whether a student is depressed or not based on a binary label derived from the same score.

**CLASSIFICATION:**

In the classification task, the goal was to predict whether a student is experiencing depression or not, based on a binary label derived from their PHQ-9 score. This involved mapping various academic, psychological, and behavioral features to a categorical outcome (depressed vs. not depressed), using models such as Logistic Regression, K-Nearest Neighbors (KNN), and Decision Tree Classifier.

**logistic regression:**

Logistic Regression was selected as a baseline classification model due to its simplicity, interpretability, and ability to model the probability of a binary outcome—in this case, whether a student is depressed or not. The model estimates the likelihood of depression using a linear combination of the input features transformed by the logistic (sigmoid) function.

**A graph of confusion matrix

AI-generated content may be incorrect.**

*Classification Report:*

*precision recall f1-score support*

*0 0.75 0.69 0.72 2306*

*1 0.79 0.84 0.81 3270*

*accuracy 0.78 5576*

*macro avg 0.77 0.76 0.77 5576*

*weighted avg 0.78 0.78 0.78 5576*

A graph of a logistic regression

AI-generated content may be incorrect.

*ROC AUC Score: 0.8449139460680952*

The model achieved an overall accuracy of 78%, with a recall of 0.84 and an F1-score of 0.81 for the depressed class, indicating strong sensitivity and balanced performance. The ROC AUC score of 0.845 further confirms the model’s ability to effectively distinguish between depressed and non-depressed students. As shown in the confusion matrix), Logistic Regression correctly identified 2732 out of 3270 depressed individuals, making it the best-performing classification model in this study.

**K Nearest Neighbours (KNN) Classifier:**

The K-Nearest Neighbors (KNN) Classifier was used to evaluate how well local similarity patterns among students can predict their depression status. As a non-parametric model, KNN classifies a sample based on the majority class of its nearest neighbors in the feature space.

A graph of a number of blue squares

AI-generated content may be incorrect.

*Classification Report:*

*precision recall f1-score support*

*0 0.71 0.67 0.69 2306*

*1 0.78 0.80 0.79 3270*

*accuracy 0.75 5576*

*macro avg 0.74 0.74 0.74 5576*

*weighted avg 0.75 0.75 0.75 5576*

A graph of a positive rate

AI-generated content may be incorrect.*ROC AUC Score: 0.7962622436881848*

The model achieved an overall accuracy of 75%, with a recall of 0.80 and an F1-score of 0.79 for the depressed class. The ROC AUC score was 0.796, indicating decent discriminative power .As shown in the confusion matrix, the model correctly classified 2622 out of 3270 depressed individuals. However, compared to Logistic Regression, KNN resulted in a slightly higher false positive rate and lower overall performance, possibly due to its sensitivity to feature scale and neighborhood structure.

**Decision Tree Classifier:**

The Decision Tree Classifier was used to model non-linear and rule-based decision boundaries for predicting students’ depression status. By recursively splitting the feature space, the model learns interpretable conditions that lead to each classification outcome.

A diagram of a tree classifier

AI-generated content may be incorrect.

*Classification Report:*

*precision recall f1-score support*

*0 0.74 0.68 0.71 2306*

*1 0.79 0.83 0.81 3270*

*accuracy 0.77 5576*

*macro avg 0.76 0.75 0.76 5576*

*weighted avg 0.77 0.77 0.77 5576*

A graph of a positive tree classifier

AI-generated content may be incorrect.

*ROC AUC Score: 0.8309494444753881*

The model achieved an overall accuracy of 77%, with a recall of 0.83 and an F1-score of 0.81 for the depressed class. The ROC AUC score of 0.831 highlights strong discriminative ability. As seen in the confusion matrix, the model correctly classified 2713 out of 3270 depressed individuals. Its performance closely matched Logistic Regression, making it a strong alternative, particularly when interpretability through decision rules is desired.

**RESULT:**

All three classification models—Logistic Regression, K-Nearest Neighbors (KNN), and Decision Tree Classifier—performed well in identifying students experiencing depression.

Logistic Regression achieved the highest overall performance with an accuracy of 78%, recall of 0.84, and a ROC AUC of 0.845, making it the most balanced and effective model in distinguishing between depressed and non-depressed individuals.

The Decision Tree Classifier closely followed, with an accuracy of 77% and ROC AUC of 0.831, while also offering greater interpretability through rule-based decision paths.

KNN Classifier, though slightly behind (accuracy = 75%, ROC AUC = 0.796), still demonstrated competitive performance but was more sensitive to the structure and distribution of the data.

As a result, Logistic Regression was selected as the primary classification model due to its strong performance across all metrics and its robustness in generalizing to new data.

A screenshot of a graph

AI-generated content may be incorrect.

NOTE1:

*Note on Focused Evaluation Metrics (Recall & F1-score for Class 1 - Depressed):*

*In classification problems where the outcome represents a sensitive or high-risk condition—such as predicting depression—recall and F1-score for the positive class (i.e., depressed students) become especially important.*

*• Recall for the depressed class (class 1) measures how many truly depressed individuals were correctly identified by the model.*

*Missing these cases (false negatives) can lead to serious consequences, such as lack of support or intervention.*

*• F1-score balances recall and precision, providing a more comprehensive view of model performance, especially when there’s a trade-off between false positives and false negatives.*

*For this reason, while overall accuracy is considered, recall and F1-score for class 1 were prioritized in evaluating model effectiveness.*

NOTE2:

*Parts of this report were prepared with the assistance of OpenAI’s ChatGPT. Support was received particularly in the coding process, structuring the report, generating the generalTableForClassification and generalTableForRegression, understanding key concepts, and interpreting the results.*