Lab 1 analysis

Data Preparation:

1. The "studentInfo.csv" dataset has been read in.

2. The creation of the target variable pass\_status indicates whether a student passed (1) or failed (0).

3. It turns the variable disability into a factor.

4. The imd\_band variable undergoes feature engineering to extract its integer representation as well as a factor.

Model Building:

1. For classification, a logistic regression model is selected.

2. To specify the target variable (pass\_status) and predictor variables (disability\_status and imd\_band\_numeric), a recipe is created.

3. "Classification" is the mode and "glm" is the engine of the logistic regression model.

4. A process is created by integrating the model and recipe.

Model Fitting:

Fit() is used to fit the model to the training set of data.

Model Evaluation:

1. Training and testing sets of data are separated out.

2. Resampling the testing data is used to assess the model.

3. One looks at the finished fitted model.

4. From the final assessment, predictions are gathered.

5. Comparing the expected classes with the actual pass\_status allows one to determine accuracy. The number of accurate and inaccurate predictions is displayed in the table.

Interpretation:

1. The IMD band and disability status of a student were used to train the logistic regression model to determine whether or not the student would pass.

2. The accuracy of the model, which expresses the percentage of accurate predictions, was used to evaluate its performance.

3. The count of correct versus incorrect predictions in the accuracy table can be interpreted to gain insight into the predictive performance of the model.