# Data Mining Peer Review

## Code by: Abducadir Aligure

## Reviewed by: Artur Komissarenko

The code provided by my partner Abducadir aims to find whether K-Nearest Neighbour (KNN) model is sufficiently accurate at predicting the severity of a road accident in the UK (based on several different features such as road conditions, weather, time, and location type) compared to a Logistic Regression model and a Random Forest model.

Within the given code my partner has done an excellent job at preparing the data by first selecting only relevant columns to be used in the models, filling in missing values within numerical columns using the median (which is less susceptible to outliers), changing categorical values into numeric ones, and using SMOTE (Synthetic Minority Oversampling Technique) to account for the class imbalance, preventing the predictive accuracy being affected by it.

The models were not only compared by their prediction accuracy, but also through the use of the classification report to review their Precision, Recall and F1-score, some of which proved instrumental to the comparison of the models. My partner also used a Confusion Matrix Heatmap to help better visualize the results of the predictive models.

In the end, the results show that the KNN model had a better prediction accuracy overall (77% compared to the Logistic Regression model: 55%, and the Random Forest model: 61%), however the Random Forest model performed significantly better (more than double) at predicting the less occurring classes such as the “Fatal” and “Serious” accidents, which would be critical to be able to predict accurately.

Despite the data preparation that my partner has performed, the models still yield very low Precision and Recall across all models used. This could be the result of their low appearance within the original dataset, with SMOTE being unable to compensate for this.

Overall, I think my partner did an excellent job at using multiple different models to find which is best at predicting the accuracy of a road accident based on different features, while ensuring to the best of his ability that the models are accurate in drawing their results.