**3.2.4     Model Training and Evaluation (32%)**

Describe the training process, including the parameters involved and how they fit, concerns about underfitting and/or overfitting, and concerns about the convergence of the optimisation.

Describe the hyperparameter selection and tuning process, including the hyperparameters involved and how they were selected and tuned, the candidate values that were considered, and the performance metric that were used for optimisation.

Describe the evaluation metrics and explain why they are appropriate.

The classification task process following these steps:

1. Cleaning the raw data, remove outliers (Position Index the report)
2. Up sampling the minority data samples
3. Load the data into X(Features) and Y(Target)
4. Perform Feature Scaling on X
5. Divide the dataset into training set and test set
6. Train the model and predict using test set
7. Using chosen evaluation metrics to evaluate a model

* Samples

From earlier section, we specified the task as multiclass classification, and decided to up sample the imbalanced class. However, under different circumstances, depending on what results we want from the models, it might be best to remove the minority 3-other samples as a trade-off for a better prediction on 1-male and 2-female data. Therefore, we will try first up sampling the imbalanced dataset, in situations where we care 3-other equally to the other two classes or wanting to achieve a better prediction on 3-other. In the opposite situation, where a higher priority is put on the common appearing gender 1-male and 2-female, we will remove all other instances. Thus, two separate training process will be carried out on these two situations respectively, and both will be evaluated and compared in the end.

Other parameters and hyperparameters will also be adjusted in a similar manner. However, different evaluation metrics for each situation will be used to ensure they reached the best performance level of their purpose.

* Outliers

In the data processing section, erroneous data and statistical outliers from the data set were removed. In order to decide whether the outliers will hinder the three models’ performance, a separate training procedure is carried out and accuracy is used here as evaluation metrics. It is found that there aren’t any significant differences in micro accuracy. Still, we are going to remove the outliers, just to ensure the data is well defined.

* Up sample or not

By default, in order to counter imbalanced data set, we are going to use SMOTE technique. However, we are also going to try not using any of the up-sample techniques, but to train the original data and evaluate whether they have a better prediction under the two situations with different prediction objectives.

It is found that

* Up sample technique

In cases where up sample technique is applied, the default up sample technique is SMOTE. However, there are different techniques such as borderline-SMOTE, borderline-SMOTE (SVM) and adaptive synthetic sampling (ADASYN). The one scores the best on evaluation metrices will be used.

Hyperparameters

Parameters

Performance Metric

10-fold cross validation is applied to reduce model’s chance of overfitting, with 9 for training and 1 for testing.

Evaluation Metrics

Two types of evaluation metrics will be used on each algorithm, they are used to measure the performance of the algorithm under different circumstances as described in parameter-sample section.

1. ROC (Receiver Operator Characteristic) curve and ROC AUC (area undercurve) score

Measure a classifier’s ability to differentiate between each class in balanced classification

2. Precision, Recall and f1 score, using macro average

Focus on decreasing the false positives of a single class: **Precision** for that class

Focus on decreasing the false negatives of a single class: **Recall** for that class.