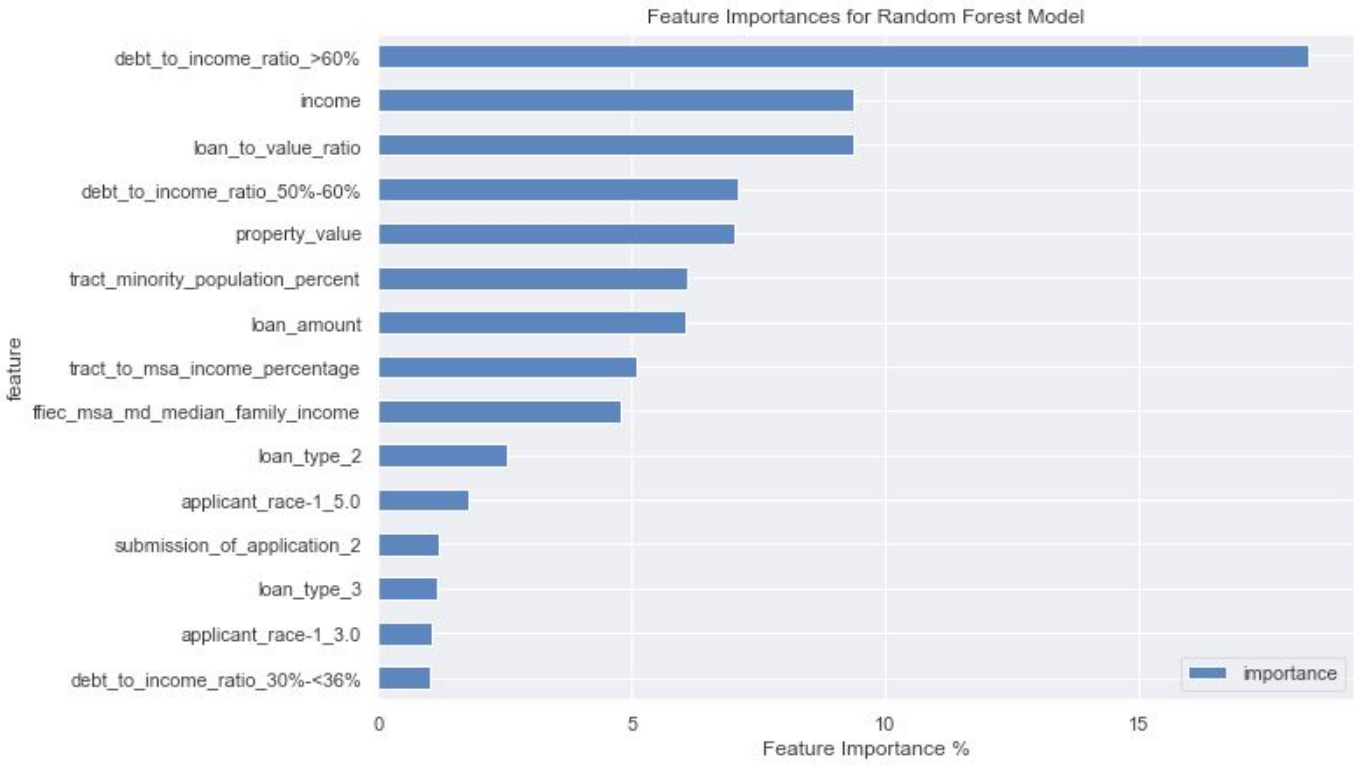
Model metrics

The model selected is the Random Forest Classifier model from the sklearn library, an ensemble method for classification problems. This brief document outlines the model's features, parameters, hyperparameters, and performance metrics.

A more detailed description of the model, in comparison with other tested models can be found in the Modeling notebook of the project.

Features

The features identified as the model as being the most important are summarized in this graph.



Parameters and hyperparameters

RandomizedSearchCV was applied on 35 iterations across 3 folds. Scoring was set to 'roc_auc' and the following hyperparameters were considered:

# Number of trees in random forest
n_estimators = [10, 50, 100, 300, 500]

Number of features to consider at every split

max_features = ['auto', 'log2']

# Maximum number of levels in tree
max_depth = [np.linspace(10, 110, num = 11)]
max_depth.append(None)

# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10, 20, 30]

# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4, 10, 20, 40]

# Method of selecting samples for training each tree
bootstrap = [True, False]

The best performing model's parameters are:

'n_estimators': 500, **'min_samples_split'**: 10, **'min_samples_leaf'**: 4, **'max_features'**: 'auto', **'max_depth'**: None, **'bootstrap'**: False

Metrics

The model's ROC_AUC score is: 0.7793627537619571

The model's prediction classification report is as follows:

	precision	recall	f1-score	support
0	0.74	0.63	0.68	178286
1	0.68	0.78	0.73	177643
accuracy			0.71	355929
macro avg	0.71	0.71	0.70	355929
weighted avg	0.71	0.71	0.70	355929