Duc Minh Nguyen – 288721

THE DATA MINING PROJECT

Bike Sharing Demand

Data Description:

The recipe ingredients of categorized cuisine are provided.

The aim is predicting the category of a dish's cuisine given a list of its ingredients.

The data is stored in JSON format.

Data Fields:

id – the recipe id

cuisine – the type of cuisine

ingredients – the list of ingredients of each recipe (of variable length)

Link to data: https://www.kaggle.com/c/whats-cooking/data

Used Methods:

Logistic Regression Model

One-Vs-Rest Logistic Regression

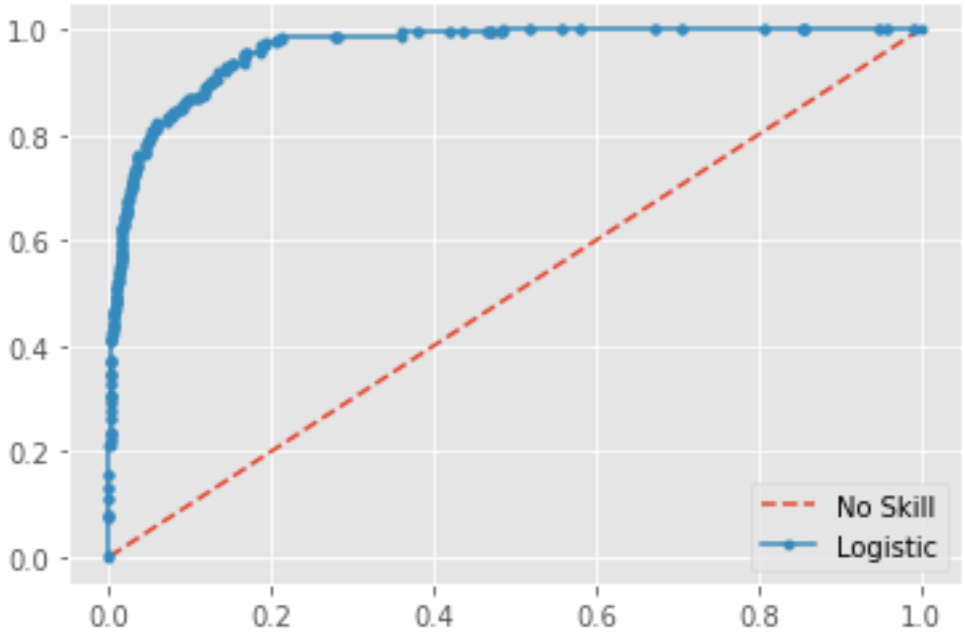
Random Forest Classifier

KNN Classifier

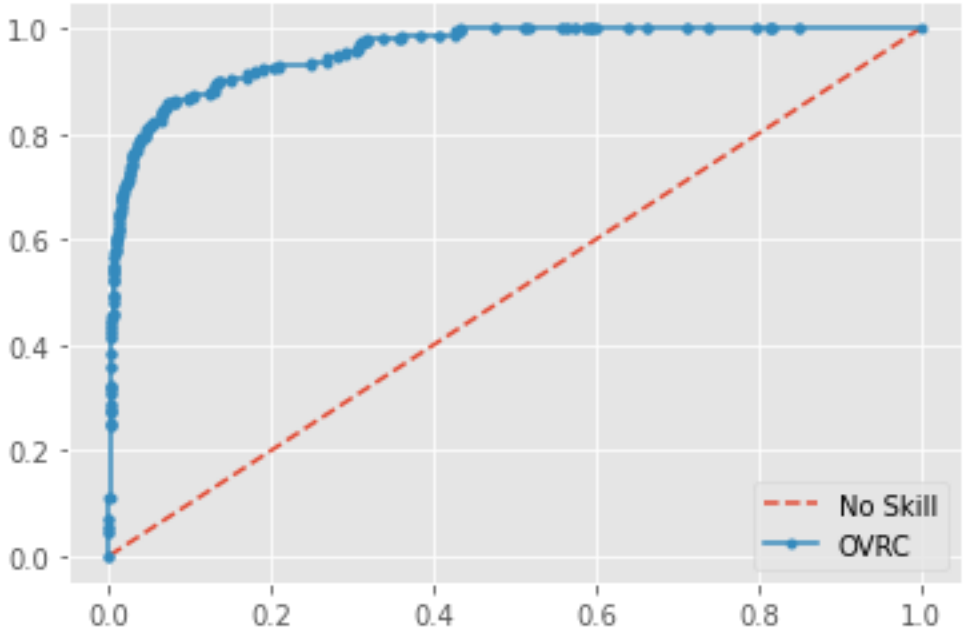
SVM Classifier

Evaluation Method: ROC

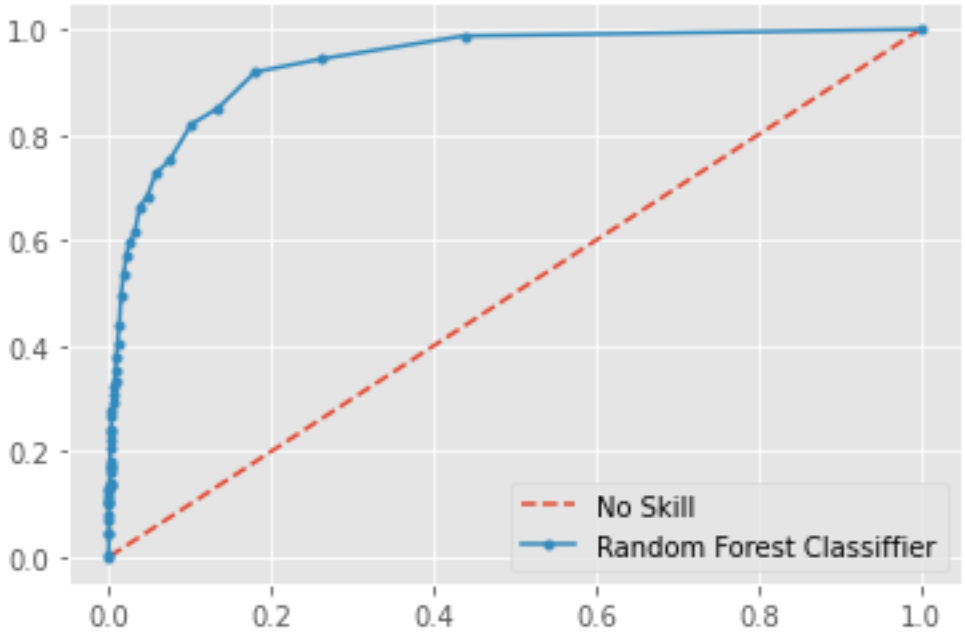
1) **Logistic Regression**



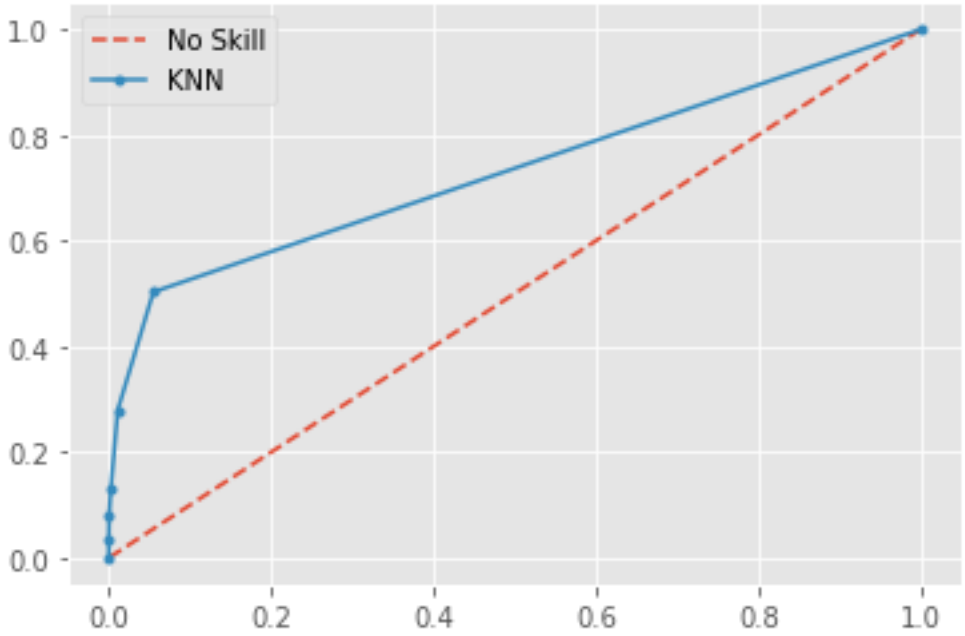
2) **One-Vs-Rest Logistic Regression**



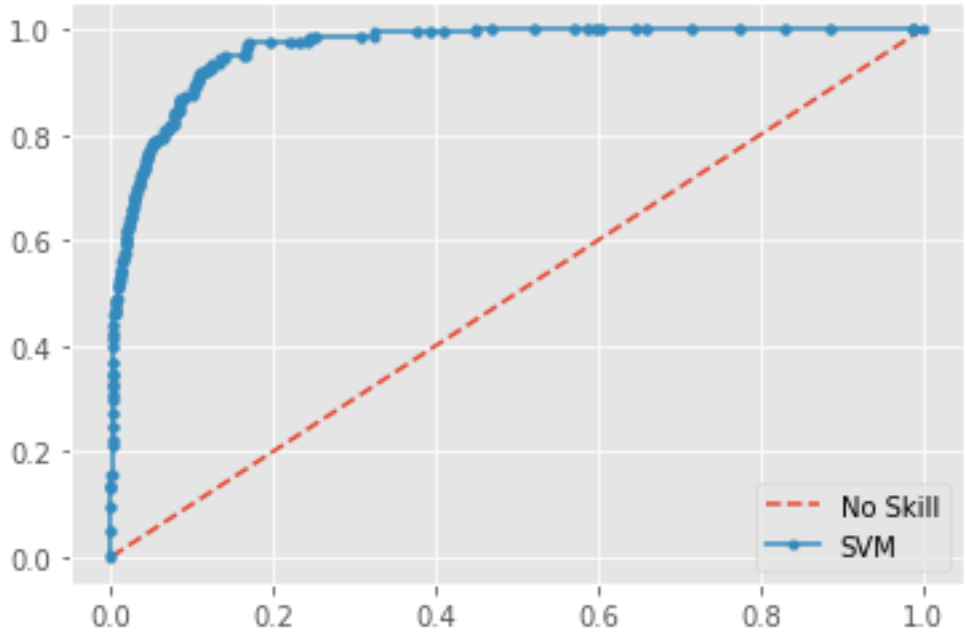
3) Random Forest Classifier



4) KNN Classifier



5) Calibrated SVM Classifier



As can be seen, the One-Vs-Rest Logistic Regression gave results performing better than the others.