Task 5 --- Analysis.

By Team 12

In the process of machine learning, what we pay attention is not only build models but also train and test them. According to results of test, we can know whether it is a good model. From the test cases we have, we can briefly know the situation of our model. Then we general descript the results of test.

The test sets' role is to fit the model and train the classification model by setting the parameters of the classifier. When the combination of the validation set is followed, different values of the same parameter are selected, and multiple classifiers are fitted. From the linear regression model, the train set's result is slightly better than test set. From the neural network model, the train set's result is far greater than test set. But there are many factors to influence it. From the random forest model, the train set's result is also slightly better than test set. When we deal with model, we split the data into training and test sets.

Based on the data obtained from our model, we draw conclusions that the client should not use our model to predict interest rates. There are still many problems with our data model which at the beginning stage. In practical considerations, there are some reasons to prove the model should not be used in practical applications.

The Analysis and processing of attributes is simple, and there are still many attributes that we do not analyze clearly. We extracted several important variables from a large number of attributes. But we still need to figure out what each variable represents and what proportion of weight it takes. Take the house_ownership variable as an example, the customer who has its own house would obtain more credits because he has fixed assets. On the other hand, what we need to analyze is not limited to interest rates, we can analyze the personal credit history to determine the level of credit of this person. There are complicit attributes analysis need to be finished to get better results.

Second, the process of feature processing needs to be further improved. In our model, what we can do is just using the interface for analysis, and the internal logic cannot be understood. There are tutorials in the websites to teach us how to implement functions. We should propose our own modifications to the model and try to apply it. The results of machine learning are not standard, and different analyses inferred a variety of results.