Jihong Yu
Professor Paul McNelis
Machine Learning in Finance
11/21/2021

The sensitivity test for titanic classification models

1) Introduction

The researcher conducted the initial analysis to solve the problem of classifying people who would survive when the Titanic was destroyed. The research uses methods DT(Decision Tree) and NN(Neural Network), based on the real Titanic passengers data with 1,045 passengers data with seven variables, class, sex, age, number of siblings, and number of parents passenger fare.

The accuracy of the DT model is 83.44% for the entire data set. While 'sex' is the most crucial variable for survival, both 'class' and 'age' are important, but the difference of importance between sex and the other variables is vital. [Figure1] Furthermore, the confusion matrix with this model shows that False positive is 2.5% and False-negative is 14%. [Figure2]

The accuracy of the NN model is 78.28% for the total data set. While 'sex' is the most important variable for survival too, 'class,' 'age,' 'the number of siblings,' and 'fare' are similarly important. However, the difference of importance between sex and the others is not significant. [Figure10] In addition, the confusion matrix with this model shows that False positive is 9.1% and False-negative is 13%. [Figure9]

2) Research purpose

a) DT model has some disadvantages, and I can consider RF (Random Forest) model to overcome those deficits. Furthermore, the RF model has a variety of hyper-parameters which users should manipulate to increase the performance, so I conducted a sensitivity test to improve the accuracy of the RF model.

the disadvantages of the DT model:

1. Because the DT model has a Hierarchical structure, if an error occurs in the middle of the tree, the error will spread at the end of the tree.
2. The minor change of training data can generate a significant influence on the results.
3. The small number of noises can have a massive influence on the results.
4. If the end of the tree increase, the risk of Low Bias and Large Variance can increase

b) The original NN model has seven hidden layers and 1,000 maximum iterations. I conducted a sensitivity test of how many hidden layers and iterations generate the most accurate classification.

3) Methods of analysis

A) I analyzed the sensitivity test to find the best hyper-parameters for the RF model to compare the accuracy of the DT model to that of the RF model.

i) I conducted the sensitivity test with the RF model, changing the number of trees. [Figure3] As a result of the test, the accuracy is the highest when the number of trees is 6. To conduct further sensitivity tests with other hyper-parameters, I choose four candidates of trees (6, 11, 8, and 7) based on the high accuracy. [Figure4]

ii) I choose hyper-parameters, the number of trees, Maximum dept, Maximum leaf nodes, and the Criterion with various numbers of candidates. [Figure5] Based on the highest accuracy, the best RF model has the 'Gini' criterion, 20 maximum depth, 35 maximum leaf nodes, and 8 trees based on the highest accuracy. I conducted the sensitivity test with the 'GridSearchCV' algorithm based on the mean value of the test score, and the highest test score is 81.68% [Figure6]

iii) The accuracy performance of the best RF model with total data set is 85.74%, about 2.3%p more than that of the DT model.

iv) Confusion matrix based on the best RF model with total data shows that False positive is 4.7% and False-negative is 9.6%. [Figure7]
v) The importance chart based on the best RF model shows that 'sex' is the most essential variable and both 'fare' and 'age' are also significant variables for survival. [Figure8]

B) I analyzed the sensitivity test to find the best hyper-parameters for the NN model to compare the accuracy of the original NN model to that of the best NN model.
   i) I conducted the sensitivity test with the NN model, changing the number of hidden layers and maximum iterations. [Figure11] When the 1,000 iterations, 128 hidden layers show 79.48% accuracy. [Figure12-a] With the 2,000 iterations, 102 hidden layers shows 79.35% accuracy. [Figure12-b] With the 5,000 iterations, 104 hidden layers shows 79.48% accuracy. [Figure12-c] With the 10,000 iterations, 103 hidden layers shows 79.62% accuracy. [Figure12-d]
   ii) Therefore, I confirm that the best NN model has 103 hidden layers with 10,000 maximum iterations.
   iii) The accuracy performance of the best NN model with total data set is 81.91%, about 1.53%p less than that of the DT model.
   iv) Confusion matrix based on the best NN model with total data shows that False positive is 5.1% and False-negative is 13%. [Figure13]
   v) The importance chart based on the best NN model does not present meaningful information about which variables influence survival because of the enormous numbers of iterations and hidden layers.

C) Comparison between the results of the best RF model and those of the best NN model.

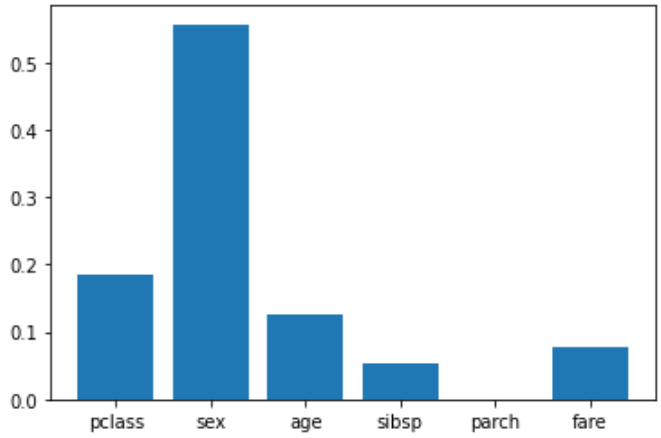
4) Result and inference

RF model is well known for overcoming the disadvantage of the DT model. Based on the accuracy of both models, because the accuracy of the best RF model is more 2.3%p than that of the original DT model, the RF model is preferred. Considering both the "women and children first" rule and first-class ticket holders' survival rate(62%) was much higher than that of third-class passengers(25%), the importance chart of the best RF model illustrates more accurate survival results based on the real history.

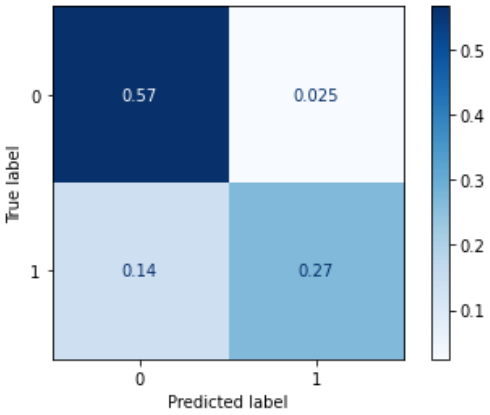
Furthermore, because the accuracy of the best NN model is 3.83%p less than that of the best RF model, the best RF model is more accurate than the best NN model. Moreover, I compare the importance charts of both the best NN model and the best RF model. Whereas the importance chart with the best NN model does not show any meaningful result, the importance chart with the best RF model presents sex, age, and fare highly influenced passengers' survival.

Therefore, to classify Titanic passengers' survival, the Random Forest model is preferred to other models based on my analysis.

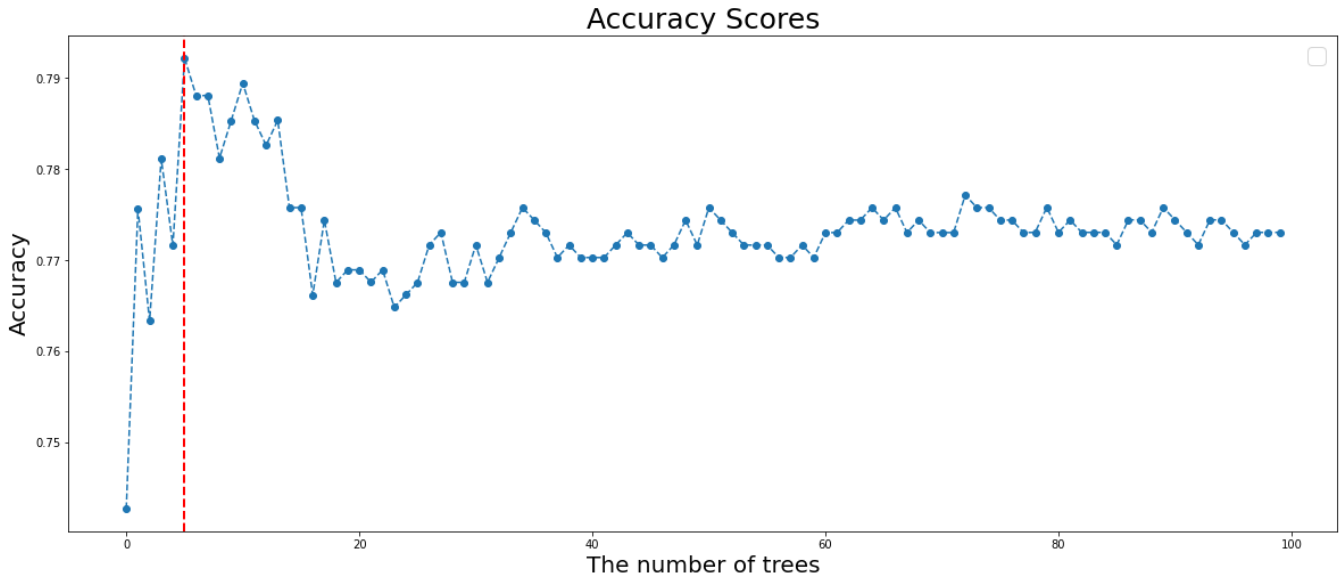
[**Appendix**]
[Figure1]
Importance chart based on the original decision tree model



[Figure2]
Confusion matrix based on original decision tree model with total data



[Figure3]
Accuracy score test based on Random Forest model



[Figure4]

Accuracy Rank	The number of trees	Accuracy
1	6	0.7922
2	11	0.7894
3	8	0.7881
4	7	0.7880

[Figure5]

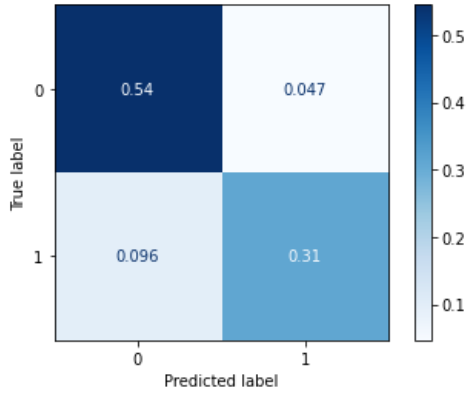
Hyper-parameters	Candidates for variables
The number of trees	6, 11, 8, 7
Maximum Depth	10, 15, 20, 25
Maximum leaf nodes	25, 30, 35
Criterion	‘Gini’, ‘Entropy’

[Figure6]

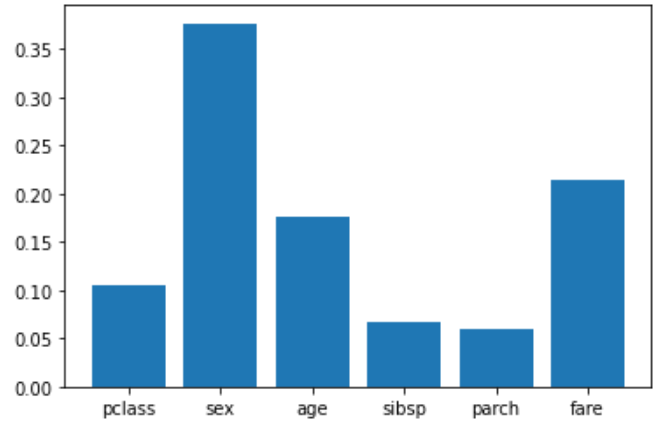
	params	mean_test_score
34	{'criterion': 'gini', 'max_depth': 20, 'max_le...	0.816790
25	{'criterion': 'gini', 'max_depth': 20, 'max_le...	0.812662
71	{'criterion': 'entropy', 'max_depth': 15, 'max...	0.812643
41	{'criterion': 'gini', 'max_depth': 25, 'max_le...	0.812643
77	{'criterion': 'entropy', 'max_depth': 20, 'max...	0.812643
..	...	...
93	{'criterion': 'entropy', 'max_depth': 25, 'max...	0.785320
46	{'criterion': 'gini', 'max_depth': 25, 'max_le...	0.785283
24	{'criterion': 'gini', 'max_depth': 20, 'max_le...	0.783950
50	{'criterion': 'entropy', 'max_depth': 10, 'max...	0.783932
7	{'criterion': 'gini', 'max_depth': 10, 'max_le...	0.777083

[96 rows x 2 columns]

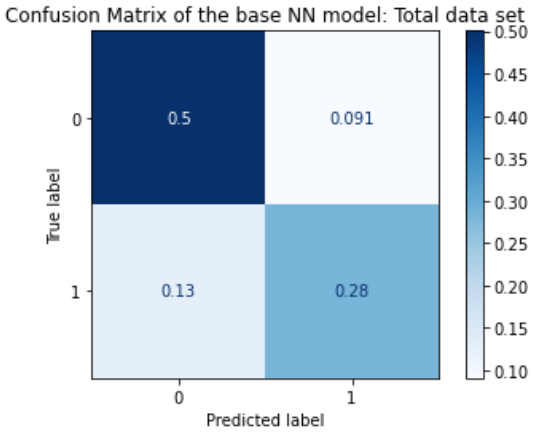
[Figure7]
Confusion matrix based on the best Random Forest model with total data



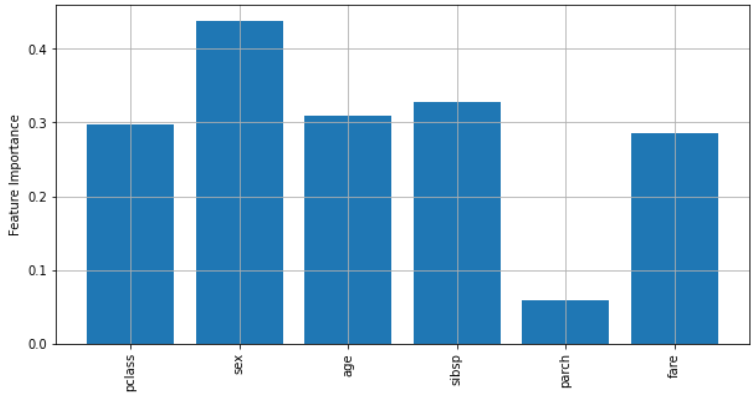
[Figure8]
Importance chart based on the best Random Forest model



[Figure9]



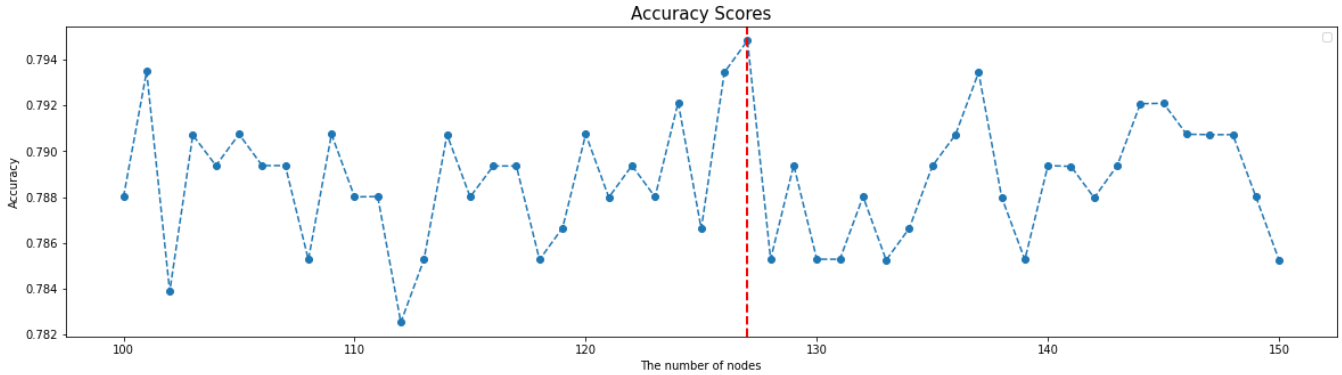
[Figure10]



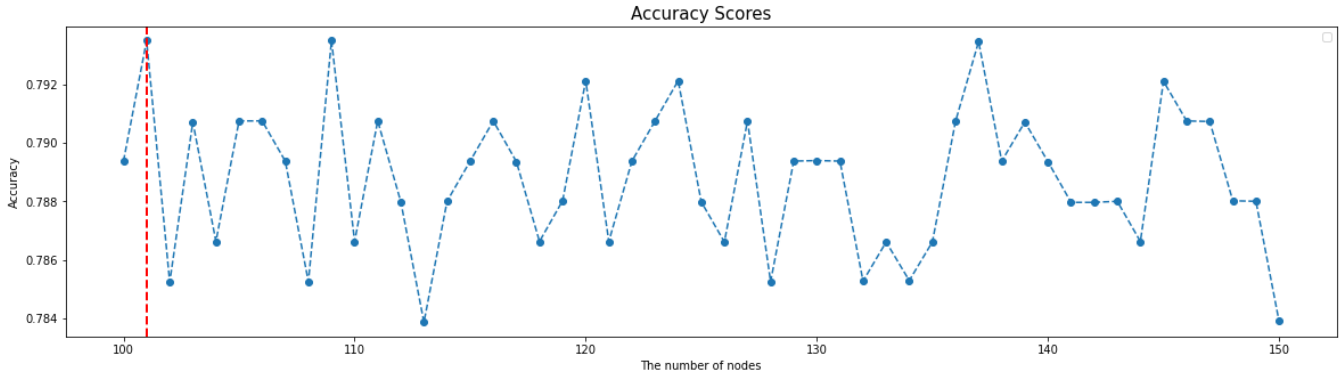
[Figure11]

Hyper-parameters	Candidates for variables
The number of hidden layers	100~150
Maximum iterations	1,000, 2,000, 5,000, 10,000

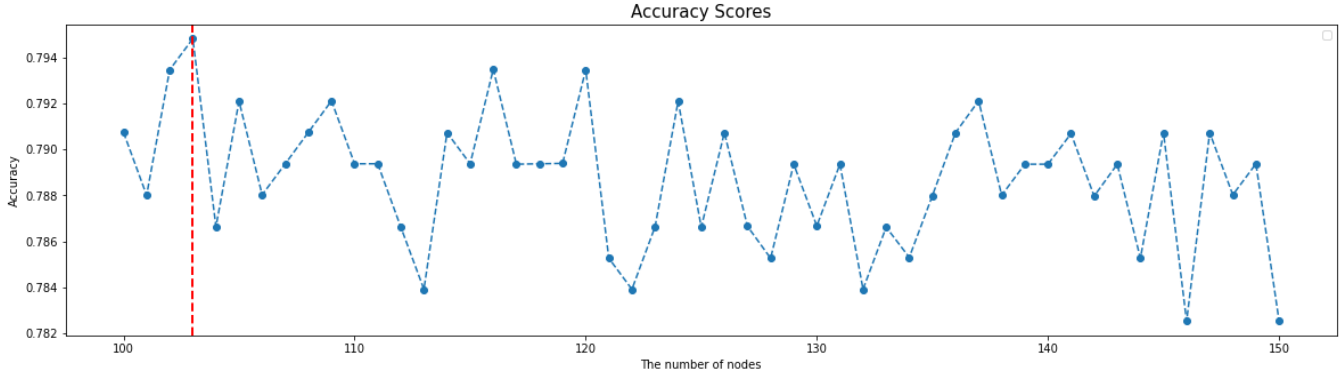
[Figure12-a]
Accuracy test based on Neural Network model with 1,000 iterations



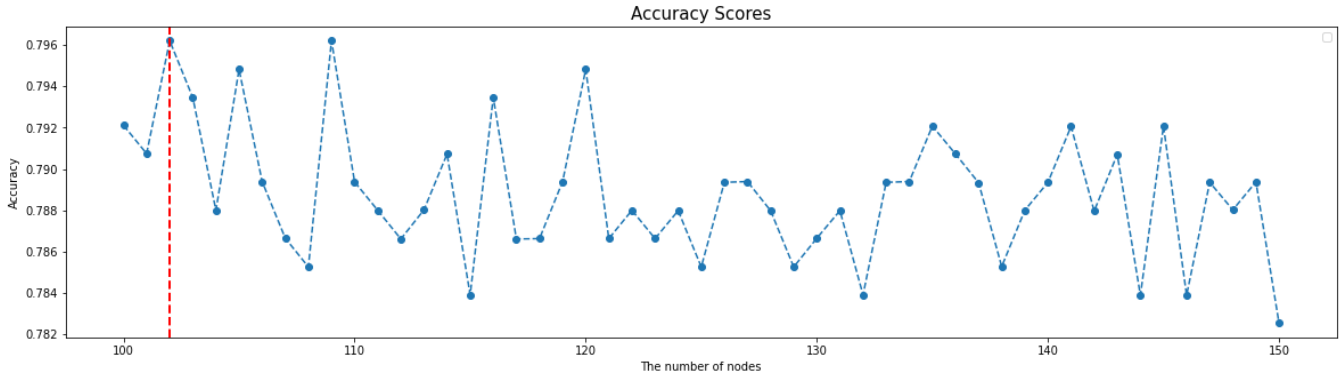
[Figure12-b]
Accuracy test based on Neural Network model with 2,000 iterations



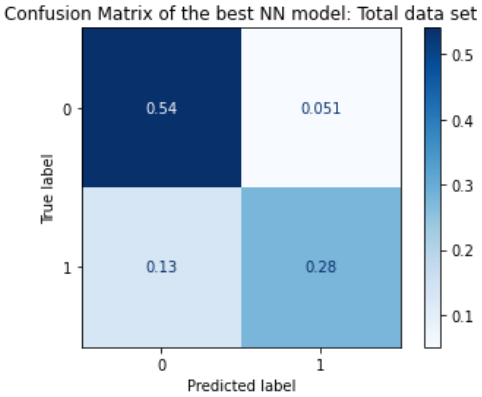
[Figure12-c]
Accuracy test based on Neural Network model with 5,000 iterations



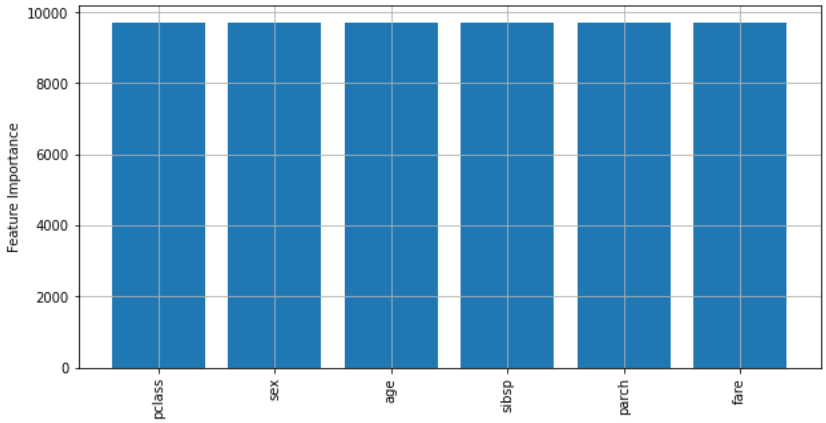
[Figure12-d]
Accuracy test based on Neural Network model with 10,000 iterations



[Figure13]


Confusion Matrix of the best NN model: Total data set
0.54
0.051
0.13
0.28
True label
Predicted label
0
1
0
1
0.5
0.4
0.3
0.2
0.1

[Figure14]


Feature Importance
10000
8000
6000
4000
2000
0
pclass
sex
age
sibsp
parch
fare