CASE STUDY 021
[Machine Learning: R]
Insurance claims

Difficulty Level: 3 of 3

Disclaimer: SuperDataScience has no affiliation with data sources. The scenario is made up for educational purposes.

You are a Data Analyst working for Insurance Claim Mangements who would like to predict whether someone will make a claim

1. Format all variables to be factors
2. Count the number of claims and no claims
3. Split the data into training and testing data sets (80% in training set)
4. Build a random forest using the training data set
5. Predict the results for the training data set
6. Calculate the confusion matrix for the training data set
7. Calculate the accuracy for the training data set
8. Predict the results for the testing data set
9. Calculate the confusion matrix for the testing data set
10. Calculate the accuracy for the testing data set

Good luck!

Difficulty note: this is a difficult assignment. Do not be surprised that there will be lots of nuances we have not covered off in the courses. But just like in the Real Life – there will be things training has not prepared you for and you will need to do research to find how to solve the problems at hand. If you get stuck, check the clues file.