JHU IDS Module 12 Lab

Audrey Long

08/15/2020

Purpose
To use RapidMiner to build and test a machine learning model to detect intrusions based on network traffic.

Assumptions
1. You have RapidMiner installed with the educational license.

2. You have watched the RapidMiner training videos identified in the Lectures for this module.

Procedure
1. Download the data file inside.labeled.csv from under the Assignment tab.

2. Open RapidMiner. Start with a blank new process.

3. Click Import Data and select the inside.labeled.csv file. Select next.

4. Confirm the data format is correct. (For example, the file is comma separated and that the first row is a header row. Click next.

5. Select the correct date format. Go to the Truth column and change its role to “label”. The Truth column should become highlighted in green. Click next.

6. Store the file in the data repository. The data will show in the Results view. Click on Design to switch to the Design view to start developing a process.

7. Create a process that reads in this data file and generates a Bayesian model. This process should also test your model using 10-fold cross-validation. The operators you will need for this are:

1. Read CSV,
2. Filter Examples,
3. Remove Duplicates,
4. Cross Validation,
5. Naïve Bayes
6. Apply Model,
7. and Performance.

Submit screenshots of your process, along with the confusion matrix it Generates.

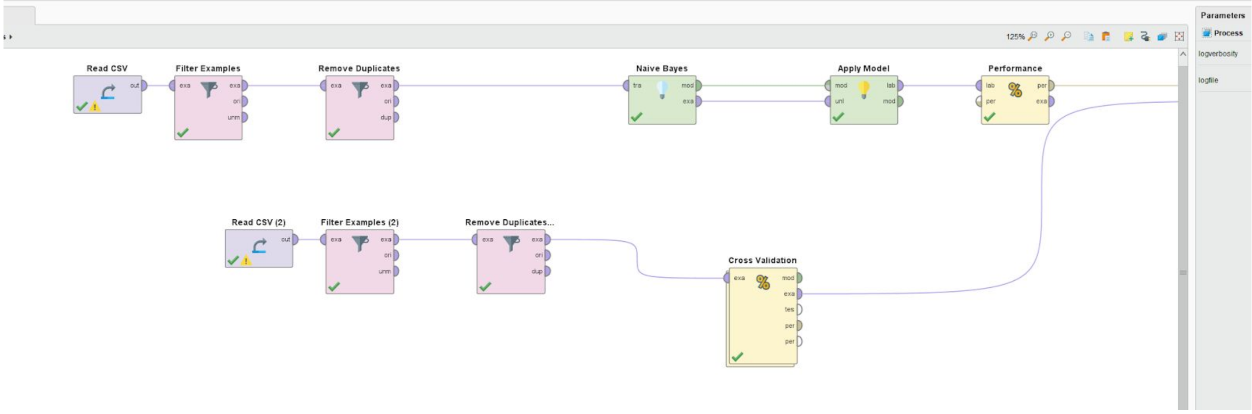


Figure 1: Design

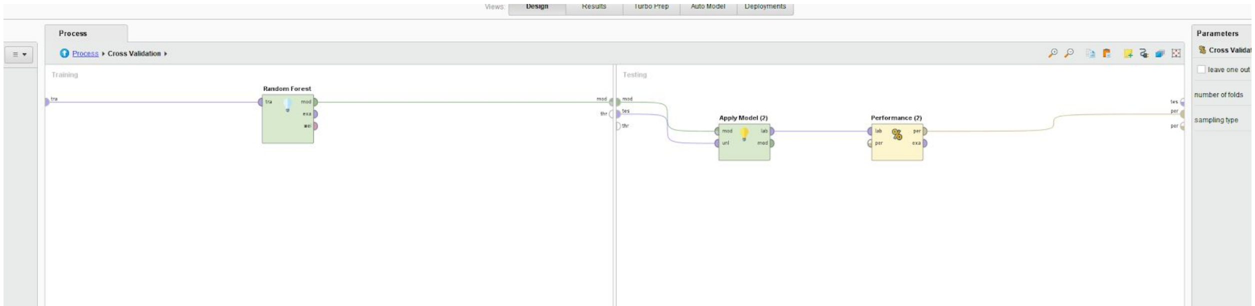
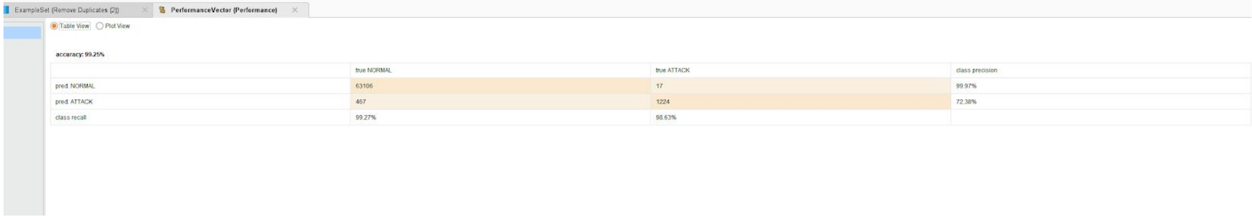


Figure 2: Cross Validation with random forest


ExampleSet (Remove Duplicates (2)) | PerformanceVector (Performance)

● Table View ○ Plot View

accuracy: 99.25%

	true NORMAL	true ATTACK	class precision
pred. NORMAL	63106	17	99.97%
pred. ATTACK	467	1224	72.38%
class recall	99.27%	98.63%	

Figure 3: Performance Vector

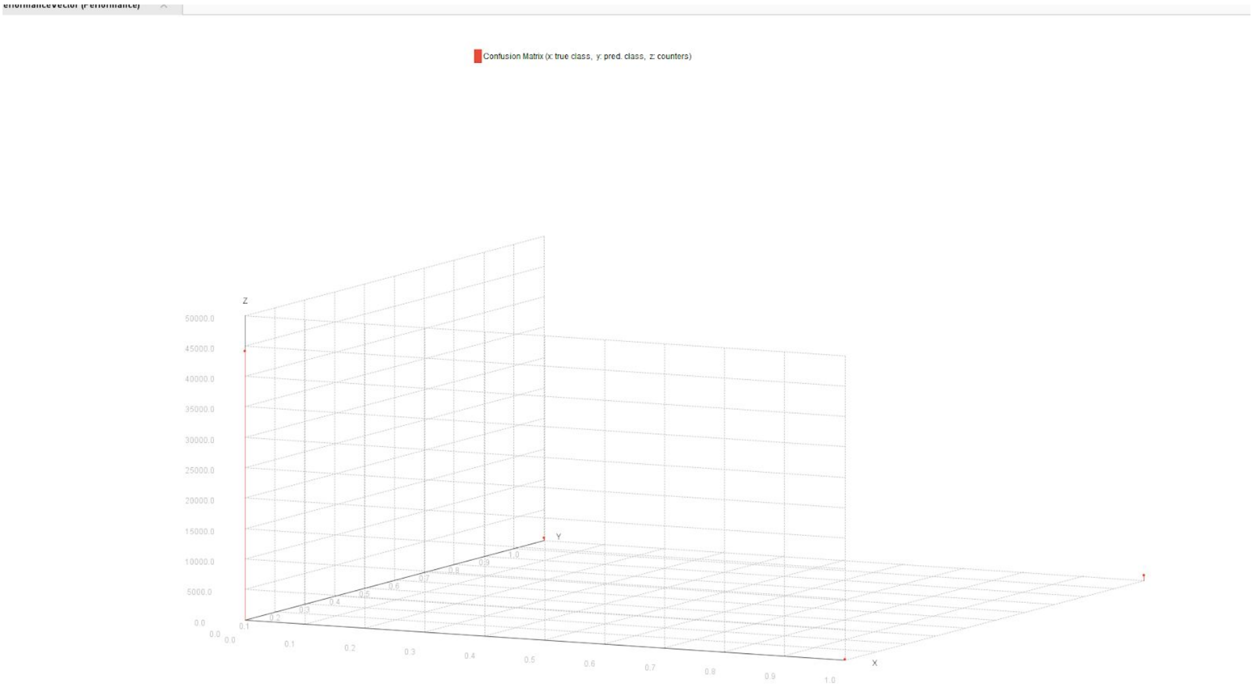
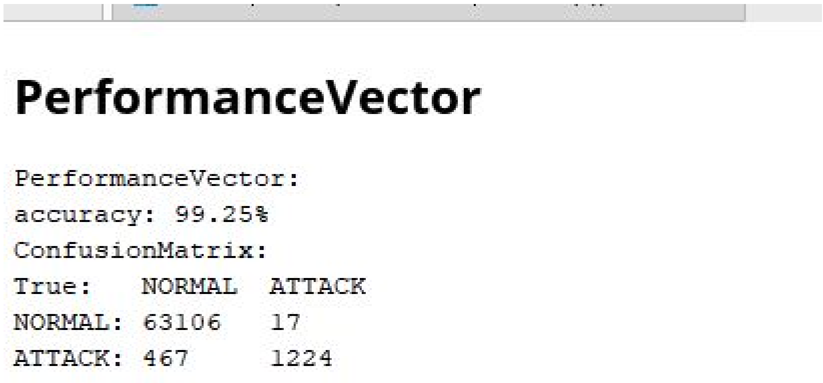


Figure 4: Confusion Matrix



PerformanceVector

```
PerformanceVector:
accuracy: 99.25%
ConfusionMatrix:
True:   NORMAL  ATTACK
NORMAL: 63106   17
ATTACK: 467     1224
```

Figure 5: Performance Description

8. Comment on your results. How do they compare to the results of the Naïve Bayes using automodel in your assignment?

The results of my Naive Bayes model seems to have a slightly higher accuracy than the automodel from the assignment. The model I created has a 99.25% accuracy compared to the automodel which came around 98.5% accuracy. I would assume a few factors to this come with the extra data parsing and cross validating my model did compared to the other one.

References

https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive_bayes.html