

PROJECT



Creditworthiness

A part of the Business Analyst Nanodegree Program

PROJECT REVIEW

ANNOTATIONS 1

NOTES

SHARE YOUR ACCOMPLISHMENT!  

Meets Specifications

This is a perfect submission. You have a very good understanding of underlying concepts. Congratulations on successfully completing the project.

Business and Data Understanding

The section is written clearly and is concise. The section is written in less than 250 words.

All following questions have been answered:

1. What decisions need to be made?

2. What data is needed to inform those decisions?

3. What kind of model (Continuous, Binary, Non-Binary, Time-Series) do we need to use to help make these decisions?

Well done!

Building the Training Set

The section is written clearly and is concise. The section is written in less than 100 words.

The following question has been answered:

1.In your cleanup process, which field(s) did you impute or remove?

Please justify why you imputed or removed these fields. Visualizations are encouraged.

The correct fields are removed or imputed.

Awesome: Data clean up has been done perfectly

Train your Classification Models

The section is written clearly and is concise. The section is written in less than 500 words.

All questions have been answered for each of the four models built: Logistic, Decision Tree, Forest Model, Boosted Model

1. Which predictor variables are significant or the most important? Please show the p-values or variable importance charts for all of your predictor variables.

2. Validate your model against the Validation set. What was the overall percent accuracy? Show the confusion matrix. Are there any bias seen in the model's predictions?

There should be 4 sets of questions answered.

This is perfect. I am very happy that you have correctly built the forest and boosted model now!

Writeup

The section is written clearly and is concise. The section is written in less than 250 words.

All questions have been answered:

1. Which model did you choose to use? Please justify your decision using only of the following techniques:

- Overall Accuracy against your Validation set
- Accuracies within "Creditworthy" and "Non-Creditworthy" segments
- ROC graph

Note: Your manager only cares about how accurate you can identify people who qualify and do not qualify for loans for this problem.

1. How many individuals are creditworthy?

Indeed forest model is the best choice. No. of creditworthy has been calculated correctly!

 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)

[Rate this review](#)

[Student FAQ](#)