UDACITY

Return to "Business Analyst" in the classroom

Creditworthiness

REVIEW

HISTORY

Requires Changes

3 SPECIFICATIONS REQUIRE CHANGES

Dear Student,

Very well done on your submission! 👍🏽

Please update your project based on my comments.

If you need any assistance, please contact your course's mentor or the community on StudentHub, they will be happy to help you!

Look forward to review your updated project soon!

Keep up the good work and take care,
Your Reviewer

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?

Great job answering all 3 questions! 👍

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.

Well done so far on this specification!

Good job on the Age variable!

You have selected correctly 6 out of 7 variables to be removed: Telephone, No-of-dependents, Foreign-Worker, Duration-in-Current-address, Concurrent-Credits and Guarantors.

Please note that "Type-of-apartment" should not be removed because it does not present low variability.

Please find the last variable to be removed following below guidelines provided:

- For numerical data fields, are there any fields that highly-correlate with each other? The correlation should be at least .70 to be considered “high”.
- Are there any missing data for each of the data fields? Fields with a lot of missing data should be removed
- Are there only a few values in a subset of your data field? Does the data field look very uniform (there is only one value for the entire field?). This is called “low variability” and you should remove fields that have low variability. Refer to the "Tips" section to find examples of data fields with low-variability.

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.

Well done so far on this specification!

Please run the calculations again taking in consideration the correct variables based on above comments.

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 all of the following techniques. Please only use these techniques to justify your decision:
    - Overall Accuracy against your Validation set
    - Accuracies within "Creditworthy" and "Non-Creditworthy" segments
    - ROC graph
    - Bias in the Confusion Matrices

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?

Well done so far on this specification!

Please run the calculations again taking in consideration the correct variables based on above comments.

RESUBMIT

DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

RETURN TO PATH