

[Return to "Business Analyst" in the classroom](#)

Creditworthiness

REVIEW

ANNOTATIONS 1

HISTORY

Meets Specifications

WELL DONE TRAINING A BINARY MODEL TO PREDICT CREDITWORTHINESS! 🎉

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?

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.

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.

Awesome:

- Confusions matrices are included.
- Good job identifying bias in the models predictions.

If there is a big difference between the accuracy of the two classes, it's an indication of biased predictions. E.g decision tree cannot predict the non-creditworthy class with the same accuracy as the creditworthy class.

As you point out, logistic regression misclassifies many creditworthy applicants. Decision tree has even more false negatives (14). Forest has only 3. So decision tree and logistic regression would deny loans to many creditworthy applicants.

Rate this review



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

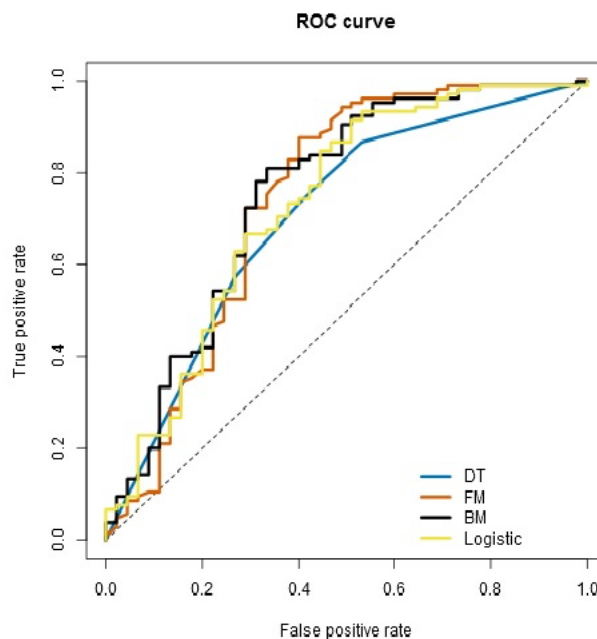
Note: Your manager only cares about how accurate you can identify people who qualify and do not qualify for loans for this problem.

1. How many individuals are creditworthy?

Suggestion:

- We could look closer at *Accuracies within "Creditworthy" and "Non-Creditworthy" segments*. In the creditworthy segment, all models have an accuracy of around 80%. When looking at the non-creditworthy accuracy, we can see that forest has a big advantage. Forest has a non-creditworthy accuracy well above 80%. E.g decision tree has only 60%
- By using the *union tool* in Alteryx, we can visualize all curves in the same graph. It makes the comparison a little easier.

EXAMPLE:



 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)