

## 10 Questions an audience would ask you

1. Q: Why choose a random forest model over a decision tree model?

A: While the decision trees and random forest models appear to be similar, a random forest is created to increase the accuracy of the model and makes it where the model isn't overfit. Decision trees do not provide this.

2. Q: Based on the feature importance, couldn't the model have been made using only the credit score, income amount, loan term, and loan amount?

A: Yes, using only these columns would help determine approval odds for a customer. These cases would be better suited for an individual that is not already with the bank as the bank may not have their financial assets. However, if we just eliminate all the other columns than we wouldn't capture outliers. Loans that have been approved with a low credit score due to the amount of assets.

3. Q: Which model would be your choice to go into production and why?

A: If we're working in a live environment, I would prefer the logistic regression model. This is because the model's results were still excellent but wouldn't have the performance issues that the random forest will have in a live environment.

4. Q: Please explain why you only used these two approaches for modeling the data.

A: I only used these two approaches because the number of features in the data wasn't large enough to bring in a more expensive model. Since the data was light it was better to use logistic regression, and I chose random forest for accuracy. However, we could have also used K-Means to determine approval odds.

5. Q: How will you handle model updates when receiving new data?

A: As there were transformations made to the data, we would need to ensure that the same transformations are done to the incoming data. We would also need to rebalance the number of approvals and rejections that are coming in. If We receive a significantly large number of approvals and this rebalance isn't done, the model would be bias towards approving loans.

6. Q: If the Random Forest model performs slower with a higher accuracy, should we move forward with the Logistic Regression approach?

A: If you're working in a live environment, it would be better to move forward with the logistic regression model. However, if the goal is to review for pre-approval

offers and the amount of data is fixed, the model with higher accuracy would be better.

7. Q: How does the model handle outliers?

A: The model does not handle outliers. These are left in because customers with a lot of assets can use these to secure the loan, increasing their chances of approvals. These instances would be considered outliers.

8. Q: You said that you wanted to add more risk characteristics to the model. Would this in-turn slow down its processing time?

A: It's possible that it would slow down the performance of the model, which is always the case with adding more features. However, once more risk characteristics are identified and added to the model, we can then run feature importance and remove additional features that do not contribute to the model.

9. Q: If there was no change in the model's accuracy before and after feature reduction, was the feature reduction needed? The additional features could provide new insights especially if they had outliers.

A: You could leave the features in, but once additional loans are added to the data set, we could experience a slow down to the model's performance. We could have a separate data set that keeps these features and revisit the model later once more loans have been added to see if the feature importance has changed or not.

10. Q: Would having a model that predicts loan approvals really help in sending out pre-approvals?

A: Yes, if we can identify what determines a loan approval, we can apply this information to other customers to determine if they would be approved or not. This would allow us to send out a pre-approval offer to customers that have not applied for a loan.