Slide 1: Good morning, executive management for The Bank of Metis. My name is Marcos Dominguez and I am here to present a data science solution to our credit card default problem

Slide 2: If you’ll recall the CFO’s financial report last week, credit card defaults are on the rise

Slide 3: And as you can see here, it’s causing considerable losses. These graphs are a painful but necessary reminder that an effective and rapid solution is necessary in order to improve the bottom line.

Slide 4: I am proposing to implement machine learning. Machine learning is, to put it simply, the process of teaching a computer how to make predictions. The way it works is: customer data goes into the model, the model is created based on certain user-defined parameters, and a decision is produced

Slide 5: In order to get the best results, I tuned the model to ensure it does not overlook the “No” predictions. It would be very costly if the model were to predict that a customer would not default and it turns out they did in fact default. This is known in data science as “Recall.” This cost is actually greater than overlooking a “Yes” prediction because missing a “Yes” prediction has no effect on the bottom line. It may affect customer service, but we can work with the relationship managers to remedy this effect.

Slide 6: There are many models to choose from, but the particular model I chose for the problem is called random forest. After comparing the accuracy of multiple models, I found random forest was the most accurate. I also compared numerous parameters to find those which make the best predictions. Random forest also takes a reasonable time to compute relative to other models.

Slide 7: I used a total of 26 input variables to make our model, some of which are credit limit, payment status, monthly payment, monthly balance…and many more. However, we found the most important feature to be monthly payment status, and as you can see by this graph, it has the highest correlation with default. This makes sense since a customer who has defaulted in the past is likely to default in the future.

Slide 8: Our strategy is quite simple. We would make our predictions and identify those customers who have a high-risk of default. The relationship manager will then act upon these results, providing options for customer to either establish a forbearance on their payment or consolidate all of their credit card debt. This gives the customer additional time to make payments or reduce their rates on all of their unsecured debt. This way, instead of taking a loss, the interest income is simply postponed to a later date.

Slide 9: Here’s the forecast of losses with and without machine learning implementation. As you can see, losses are significantly smaller, and they even appear to be getting smaller as time goes by.

Slide 10: And here we can see how much we can potentially save over the next 6 months. Without machine learning we would likely lose an additional $80 million, while with implementation that loss would be much smaller at about $41 million. That equates to a total savings of $39 million over the next 6 months. The model accuracy was 55% and almost 22,000 defaults were prevented

Slide 11: Here’s a demonstration of how the model works. You input the customer information, click “Predict” and out comes the likelihood of default. You’re going to see three examples: an extreme defaulter, a moderate defaulter, and a customer that always pays on time. Please notice how the likelihood of default decreases as payment status improves.

Slide 12: Thank you for your time and I’m happy to answer any questions.