# Slide 21

As we begin to discuss our model selections and results, I would like to remind everyone of what we are trying to accomplish. We would like to predict that a slide event will occur within the next 30 minutes. This will give ample time for drill operators to make adjustments and improve drilling efficiency.

# Slide 22

We decided to attempt 3 different model builds - Logistic Regression, Random Forest, and Support Vector Machine. We ran all three models against each other and looked at accuracy and f1 scores to determine the best model for our problem at hand.

# Slide 23/24

Ultimately, our problem is a classification problem. That’s why we went with Logistic Regression, Random Forest, and SVM as these are good classifier models. We created a predictive model to classify if the next data point would be a sliding or rotating state.

All 3 of our models performed well achieving greater than 97% accuracy. The Support Vector Machine Model performed the best at almost 99% accuracy.

# Slide 25/26

We decided to attempt re-sampling techniques on the raw data as processing the full dataset took several hours. We attempted resampling rates of 1 minute, 5 minutes, and 10 minutes. As we can see, the 1-minute resampling had very little drop in accuracy. Resampling of 5 minutes and 10 minutes created noticeable drops in accuracy so we decided to go with a 1-minute re-sampling rate. This dropped our data down to 120,000 rows and allowed us to run our models within a few minutes.

# Slide 27/28

Here you can see the classification results of our models. Along the top of the graph, you’ll see data points indicating rotating drill state which corresponds to a high drill RPM value. Along the bottom of the graph are data points representing a sliding drill state which corresponds to a low drill RPM. The circles are data points that were mis-classified by our model. As we can see, it appears our models are mis-classifying the beginning and end of a slide event. Fortunately, this is not a big issue as we are more concerned about the overall duration of a slide event.

# Slide 29/30

We next attempted to build out our future predictions. Ultimately, we would like to be able to predict an impending slide event within 30 minutes with high accuracy. This would give drill operators ample time to adjust operations in order to prevent the slide event. With the Logistic Regression and Support Vector Machine models, we were able to achieve almost 75% accuracy at the 30-mintue window and achieved almost 85% accuracy at the 10-minute window. The Random Forest modeled performed significantly poorer at all time windows.

# Slide 31

Ultimately we decided the Logistic Regression model was the best model for our data. It performed almost as well as the Support Vector Machine model, but ran significantly faster. Also, the Logistic Regression model is easier to explain to stakeholders.