This Lab is similar to QwikLab – Intro to Machine Learning. However, we will be using different data sources for Machine Learning.

Our Sample Data source will help identify potential customers for a targeted marketing campaign related to banking. This Data source is available publicly from the University of California at Irvine (UCI) Machine Learning Repository. These datasets contain general information about customers, and information about how they responded to previous marketing contacts. You will use this data to identify which customers are most likely to subscribe to your new product, a bank term deposit, also known as a certificate of deposit (CD).

To begin with our lab, we need to download the data set and prepare the data. We need to transform the data of the target value in terms of 1 and 0. Once done, upload the data to Amazon S3 bucket.

After you upload the dataset to your Amazon Simple Storage Service (Amazon S3) location, you use it to create a training datasource. In Amazon Machine Learning, reference the data source from S3 bucket where data set was uploaded. To get S3 access, give permission for Amazon Machine Learning to access S3 bucket. Verify the Schema. Select the target variable. Review and Continue.

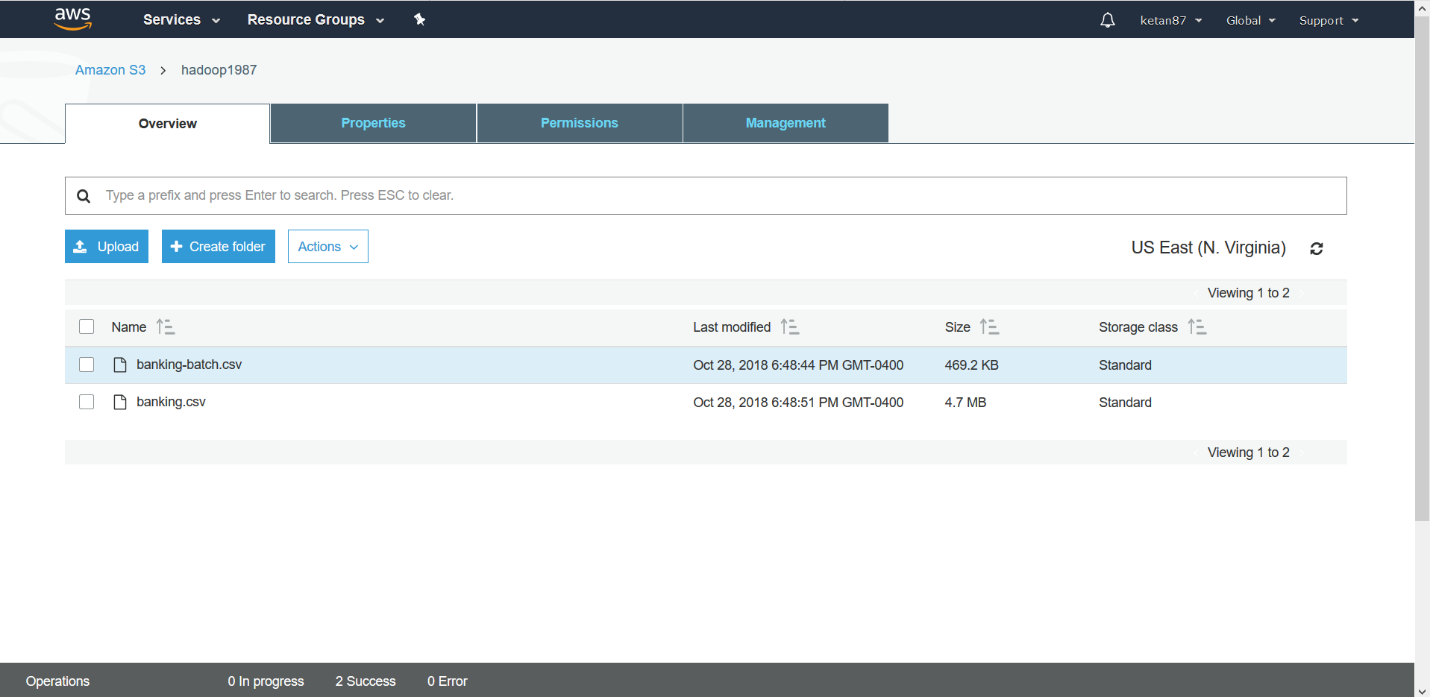
Once Training Data Source is complete, create a ML Model. Reference the earlier data source. Review and Finish. Post, Amazon ML adds your model to the processing queue. It splits the training datasource into two sections, one containing 70% of the data and one containing the remaining 30%. Then trains the ML model on the section that contains 70% of the input data. Finally evaluates the model using the remaining 30% of the input data.

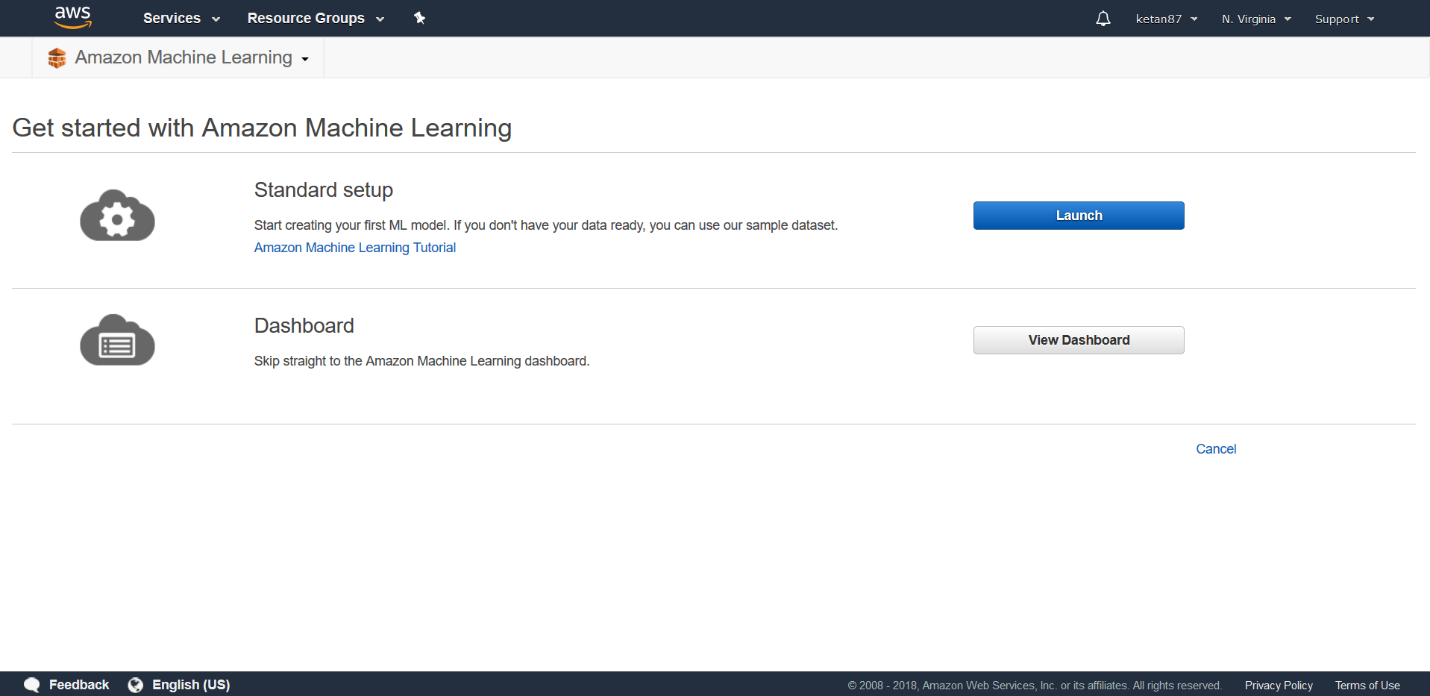
Amazon ML computed an industry-standard quality metric, called the Area Under a Curve (AUC) metric, that expresses the performance quality of your ML model. Amazon ML also interprets the AUC metric to tell you if the quality of the ML model is adequate for most machine learning applications. In the evaluation section, you can also adjust the score threshold.

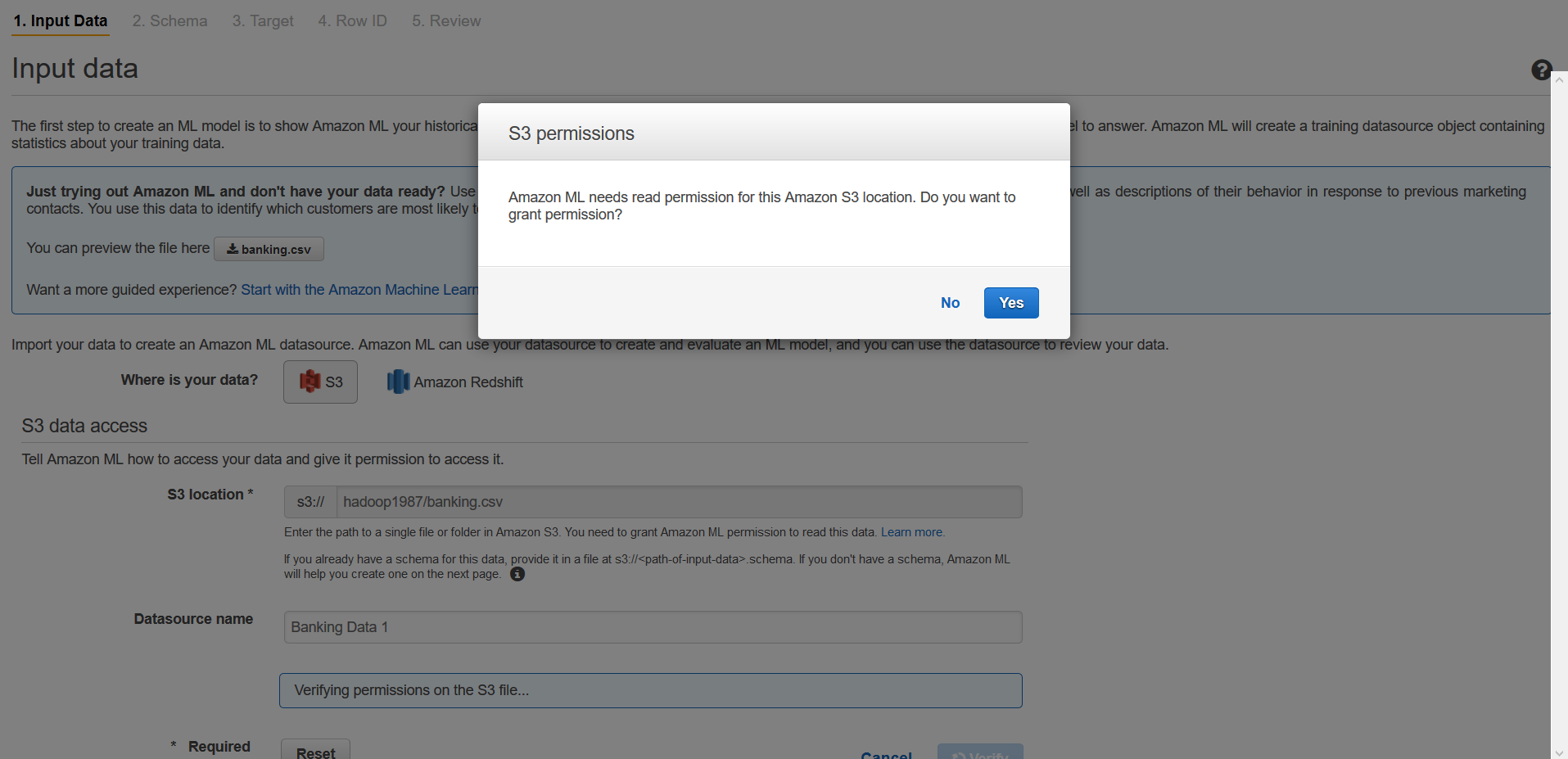
Once you have define your score threshold, you can generate two types of prediction - batch and real-time. A real-time prediction is a prediction for a single observation that Amazon ML generates on demand. Real-time predictions are ideal for mobile apps, websites, and other applications that need to use results interactively. A batch prediction is a set of predictions for a group of observations. Amazon ML processes the records in a batch prediction together, so processing can take some time. Use batch predictions for applications that require predictions for set of observations or predictions that don't use results interactively.

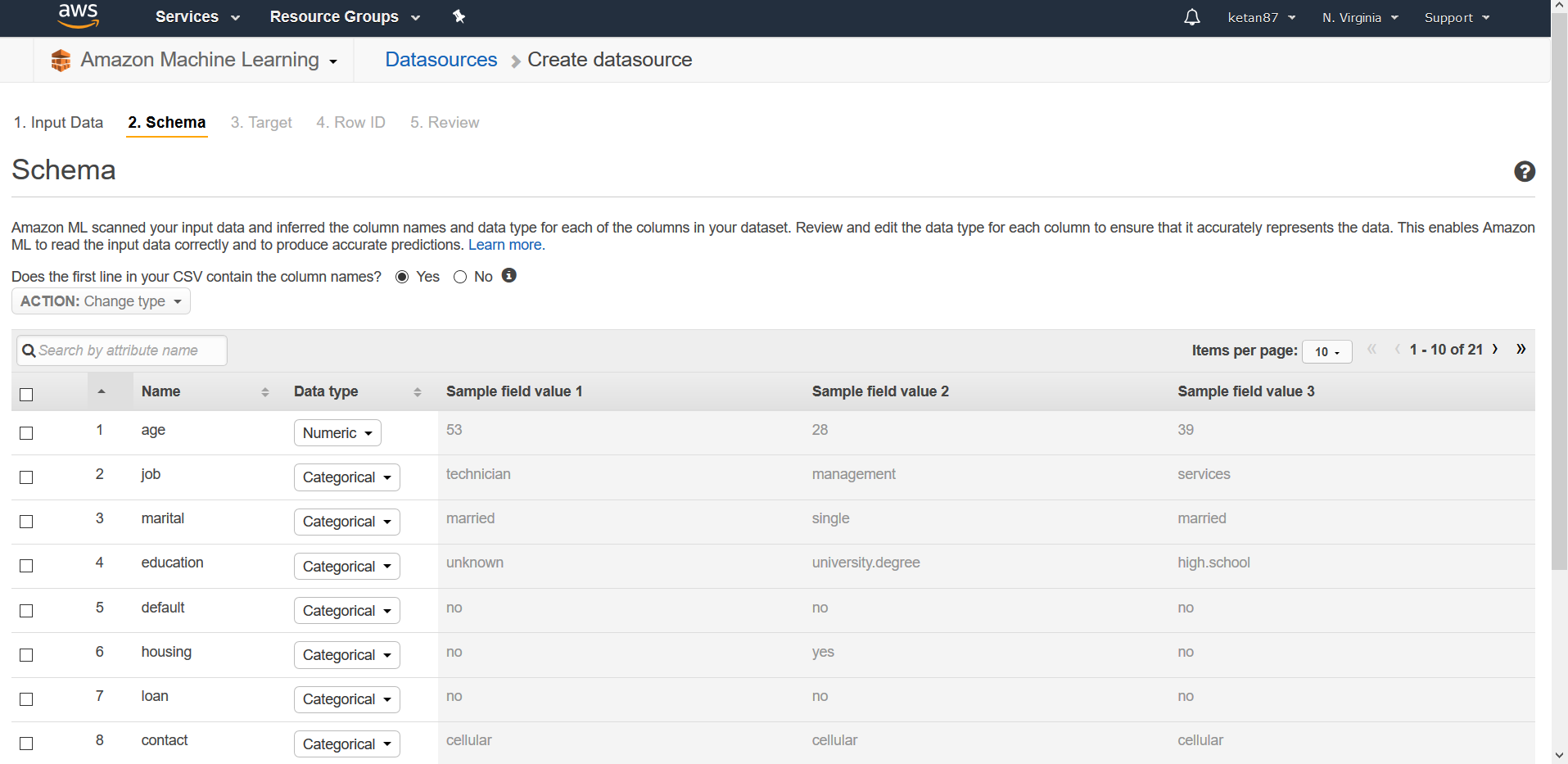
Thus we have created, reviewed, and used the ML Model to predict the data.

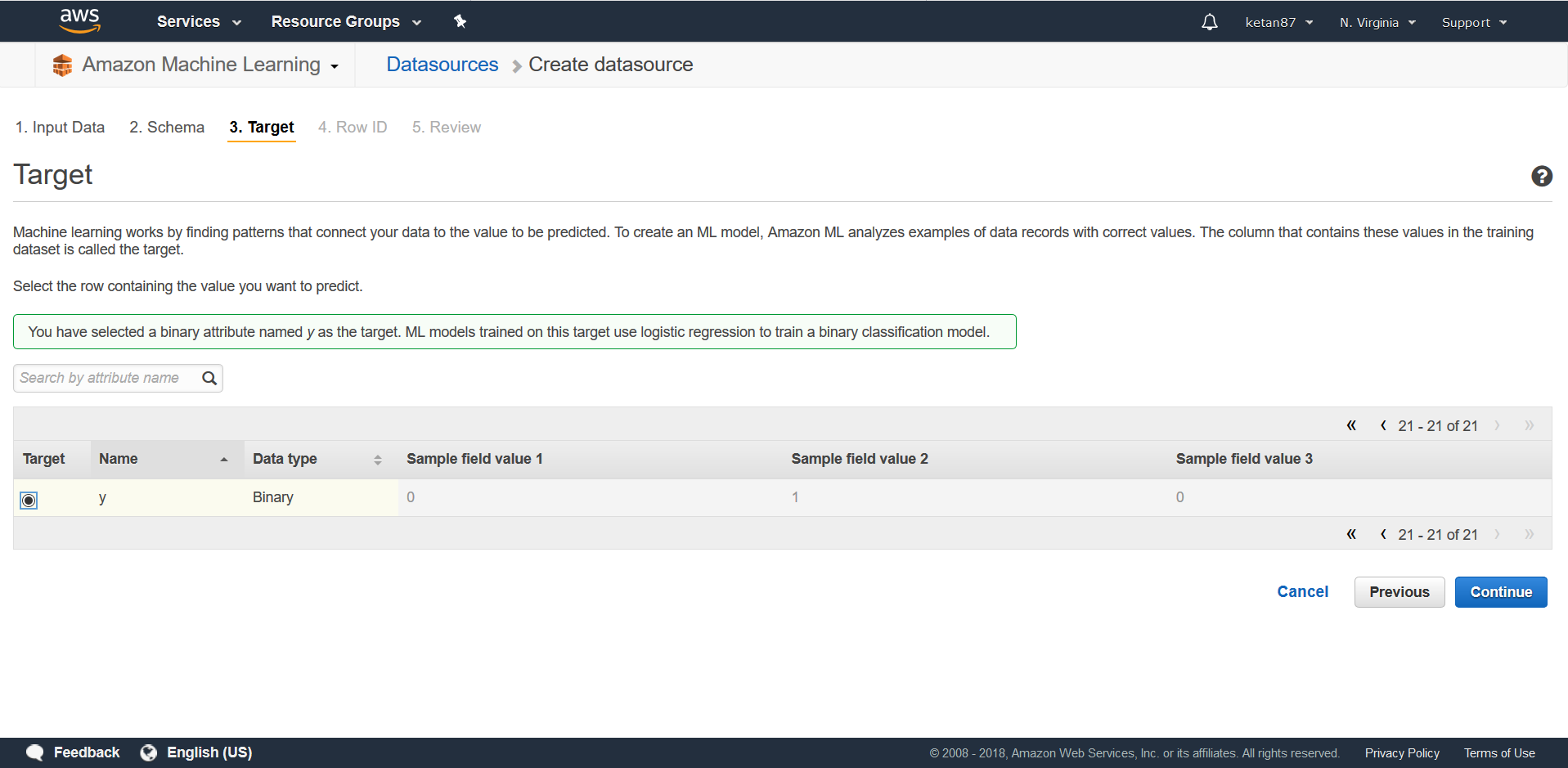
Screenshots:

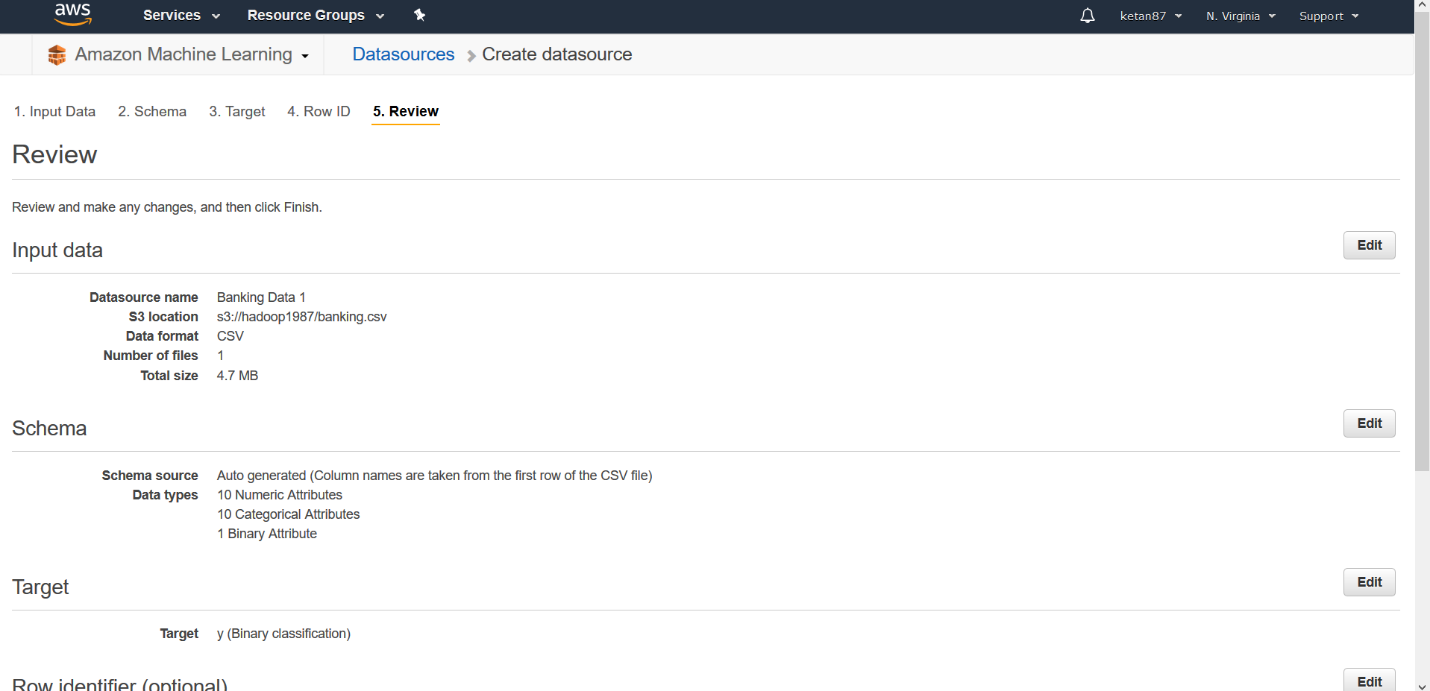


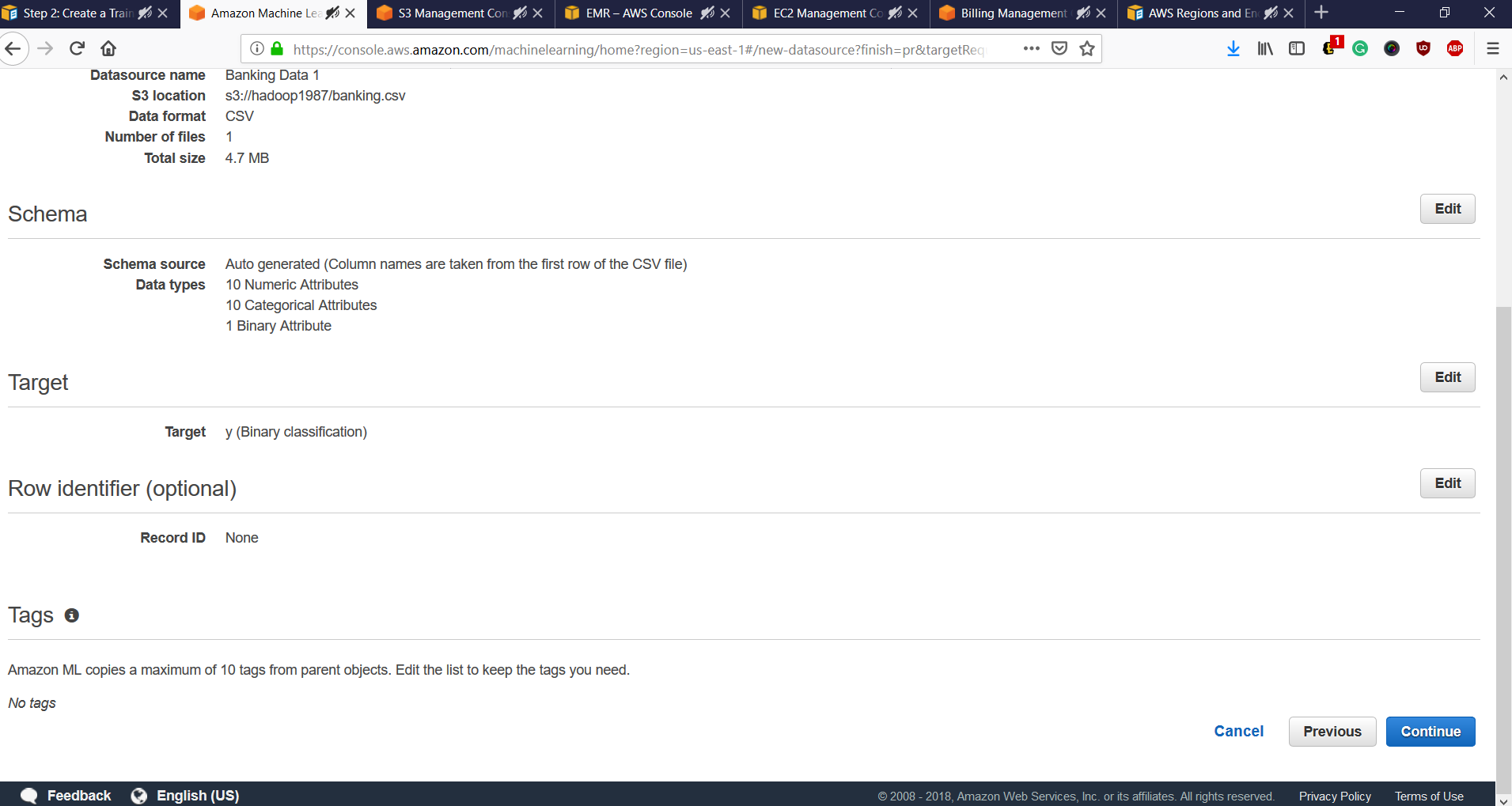


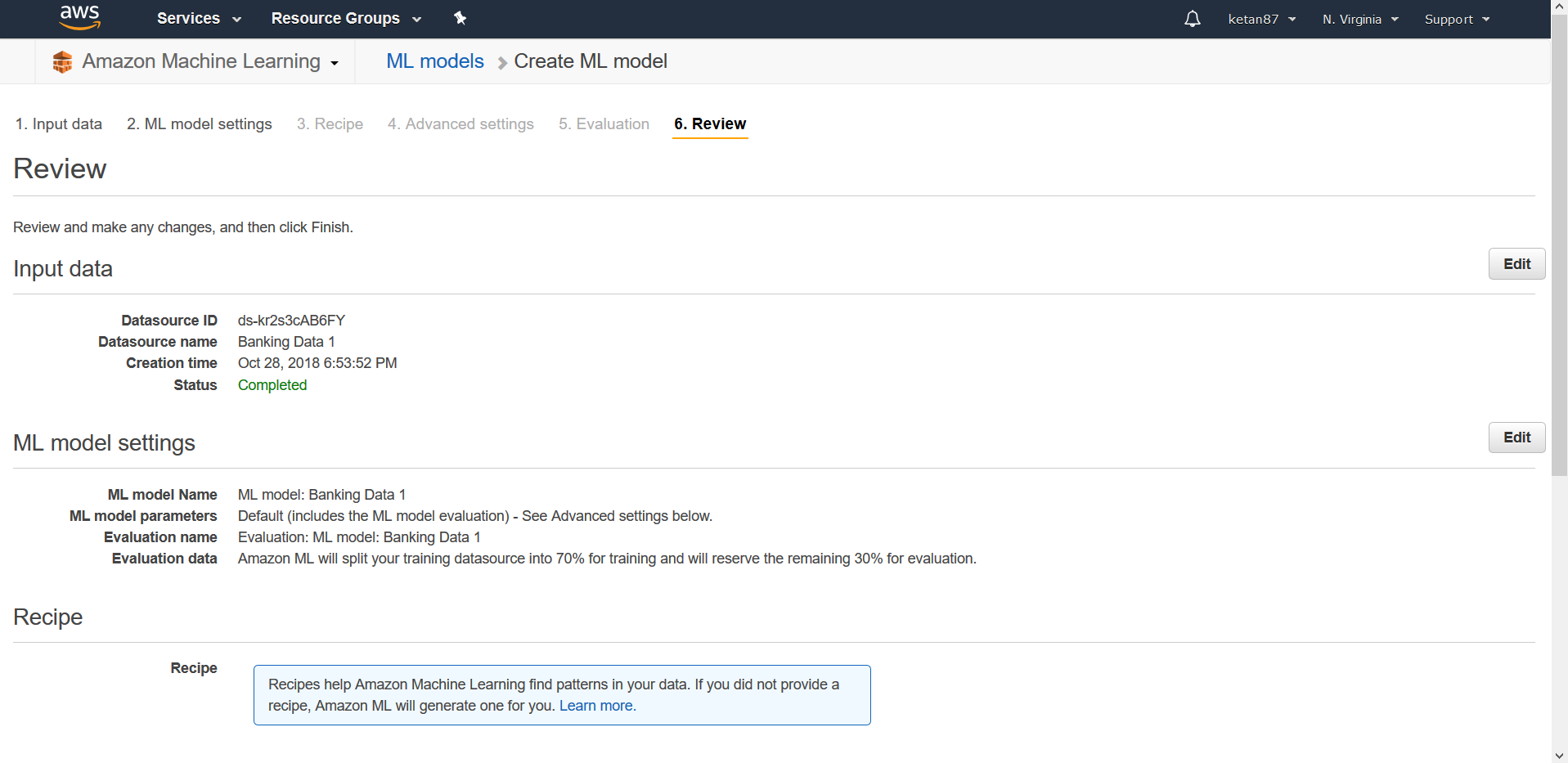


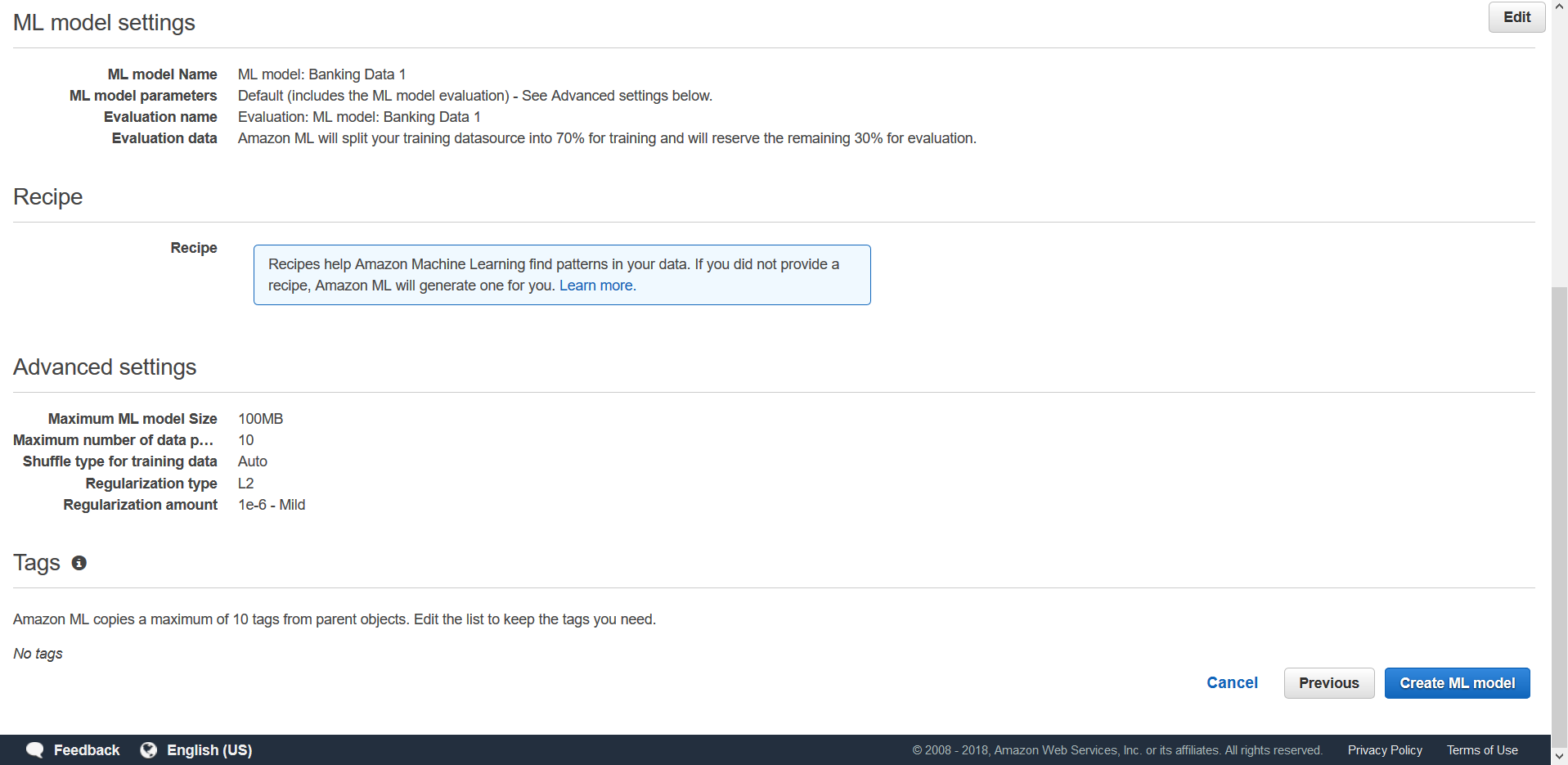


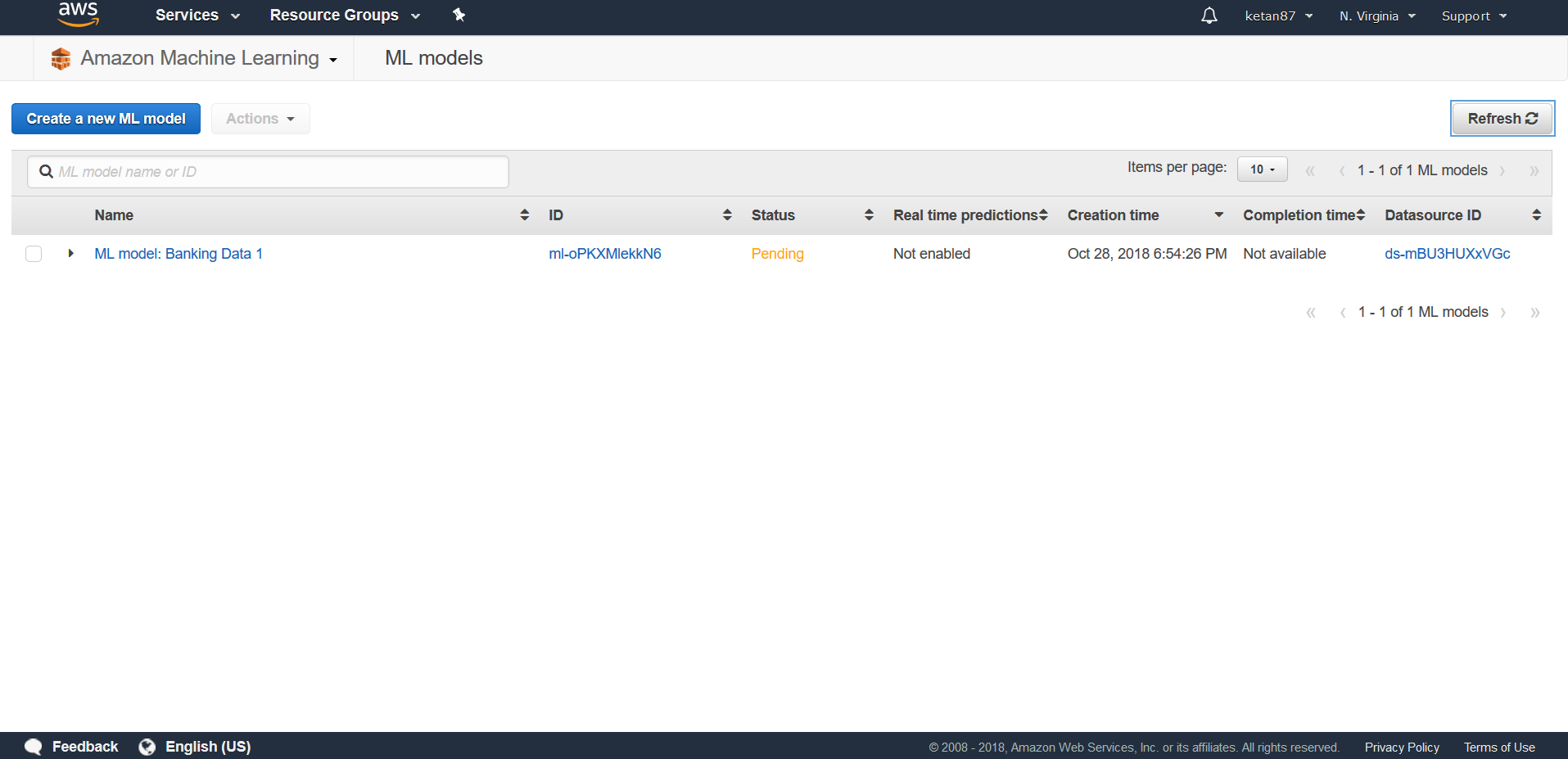


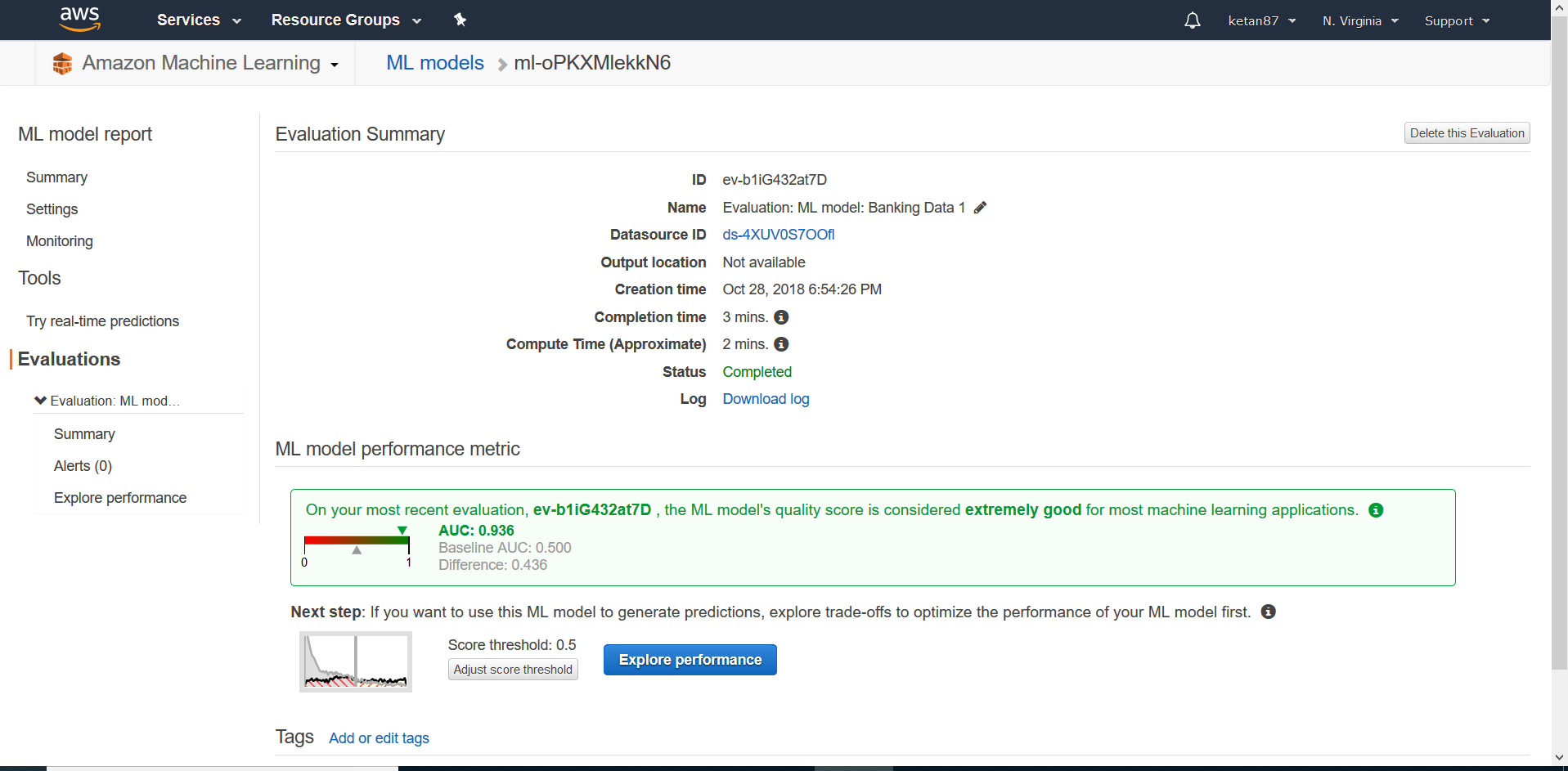


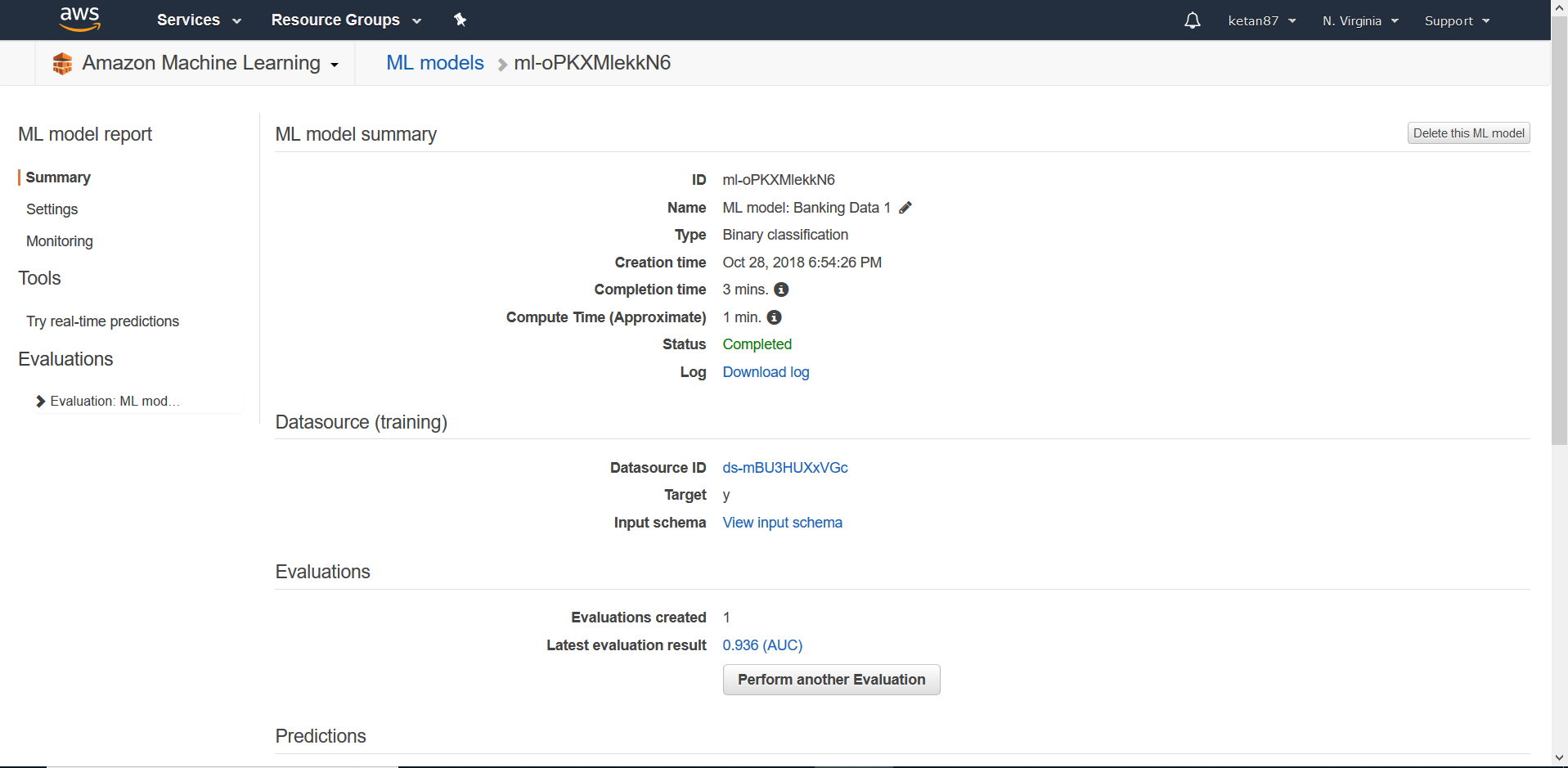


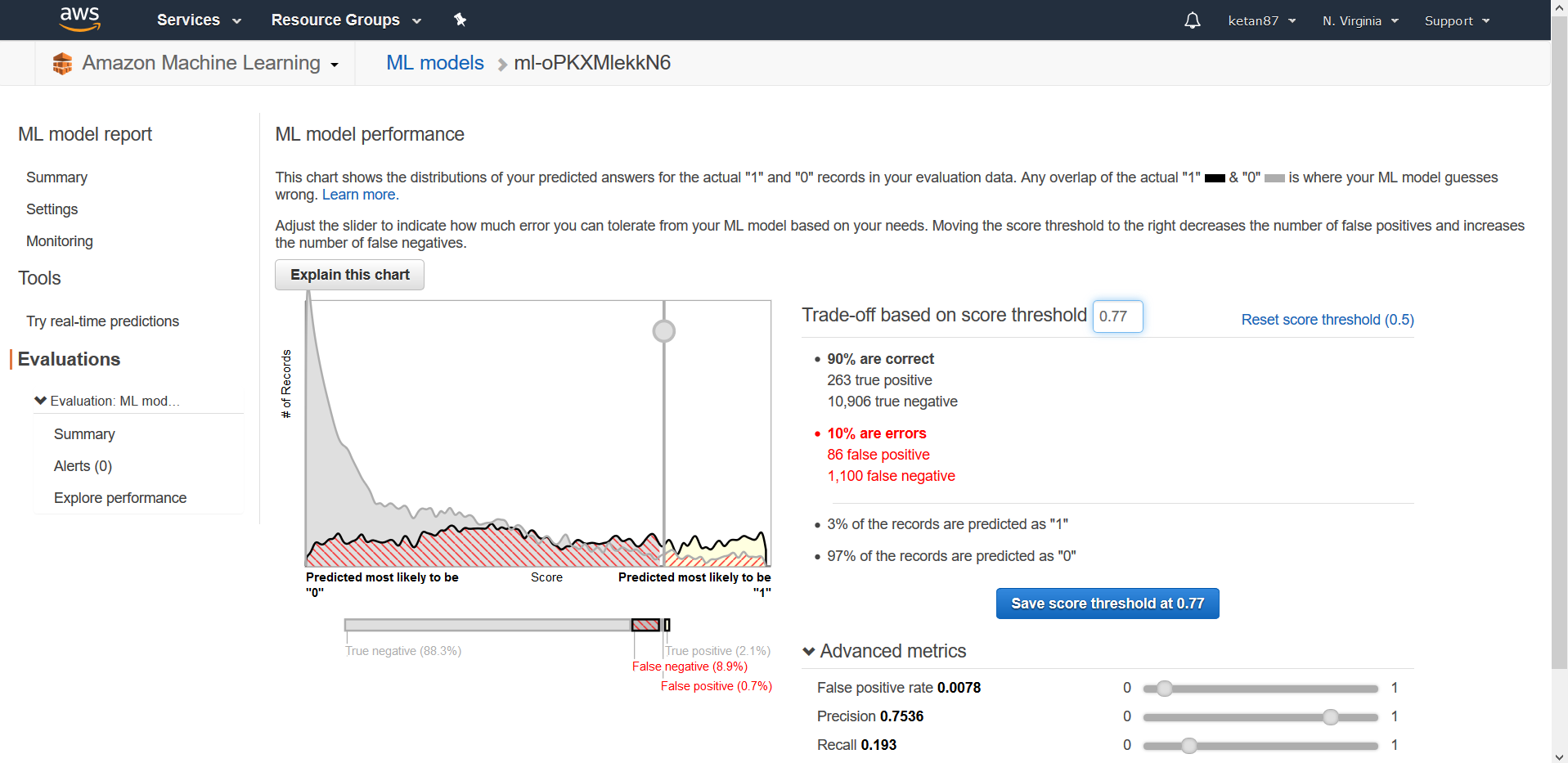


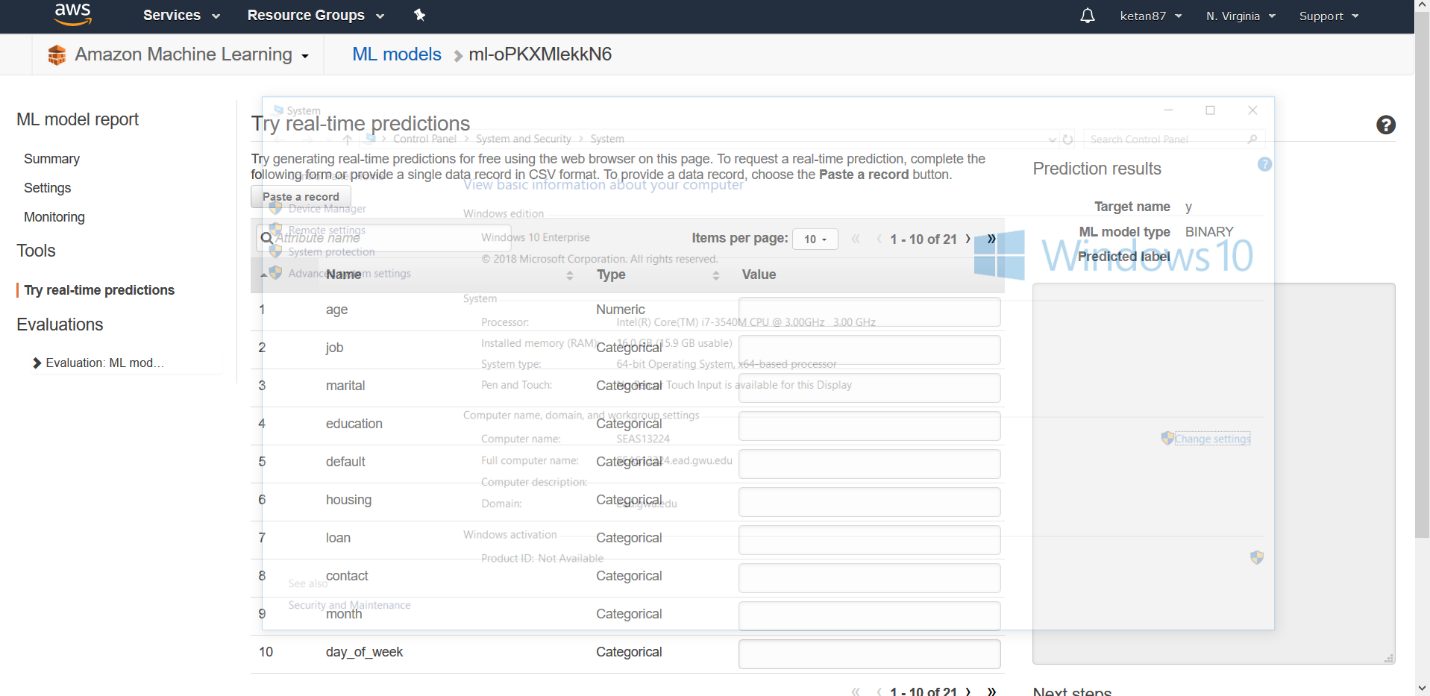


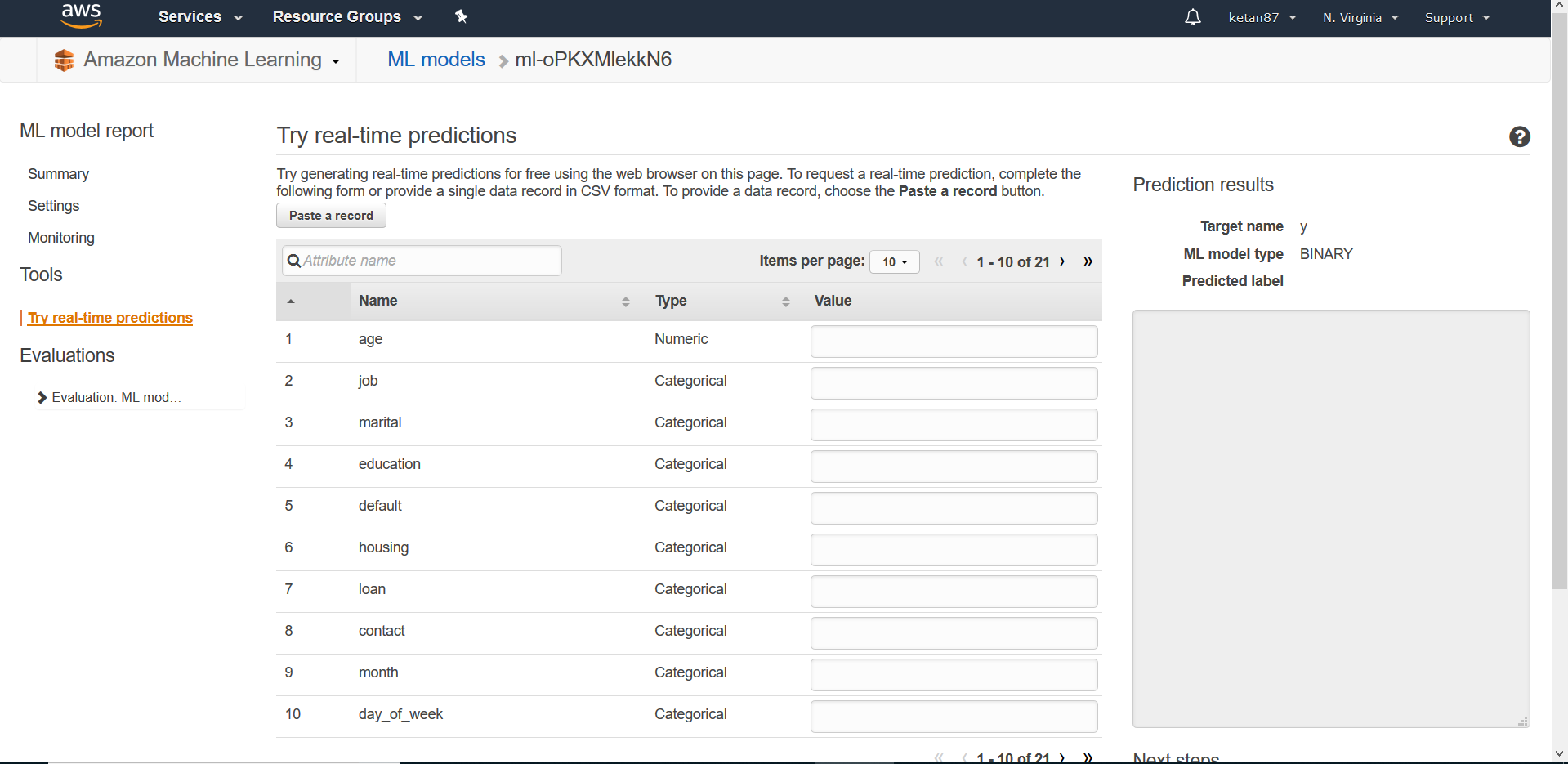


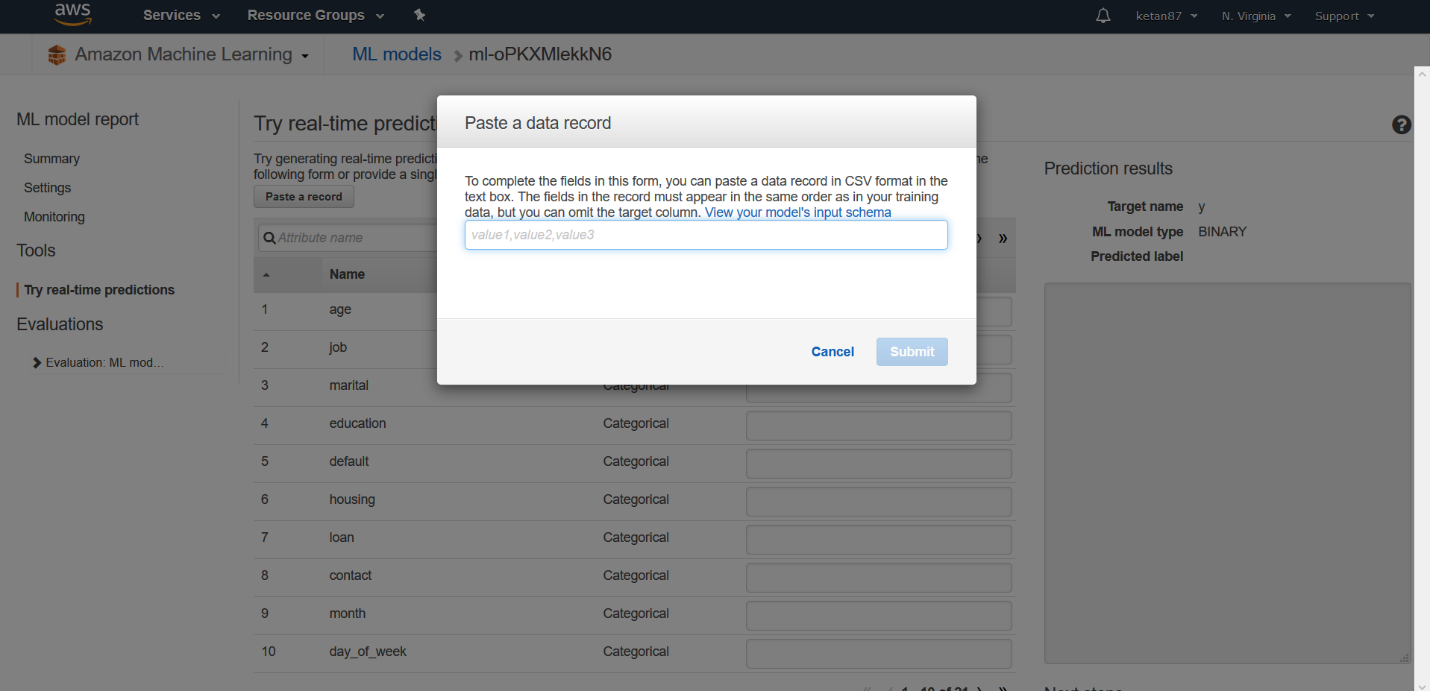


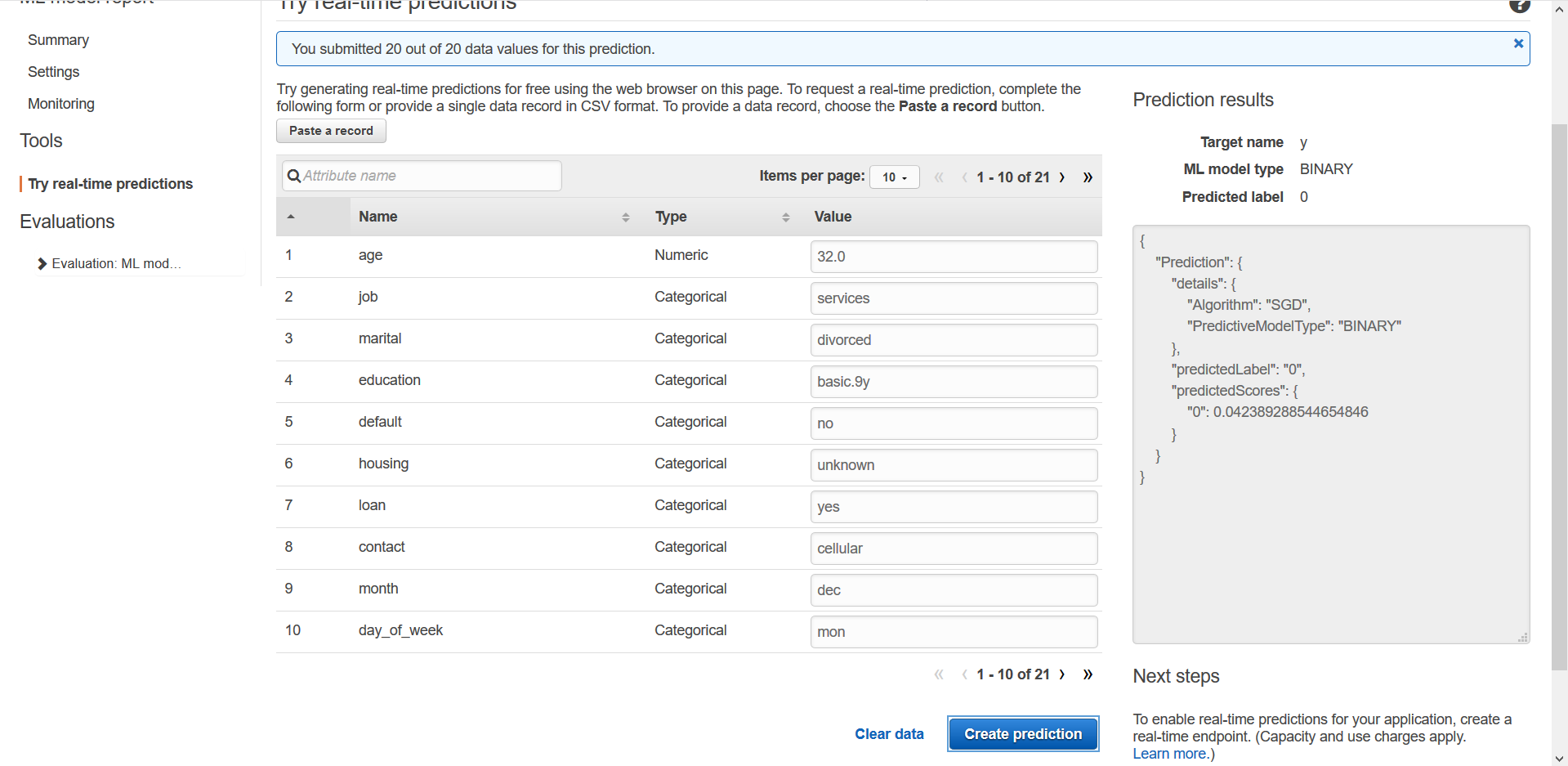


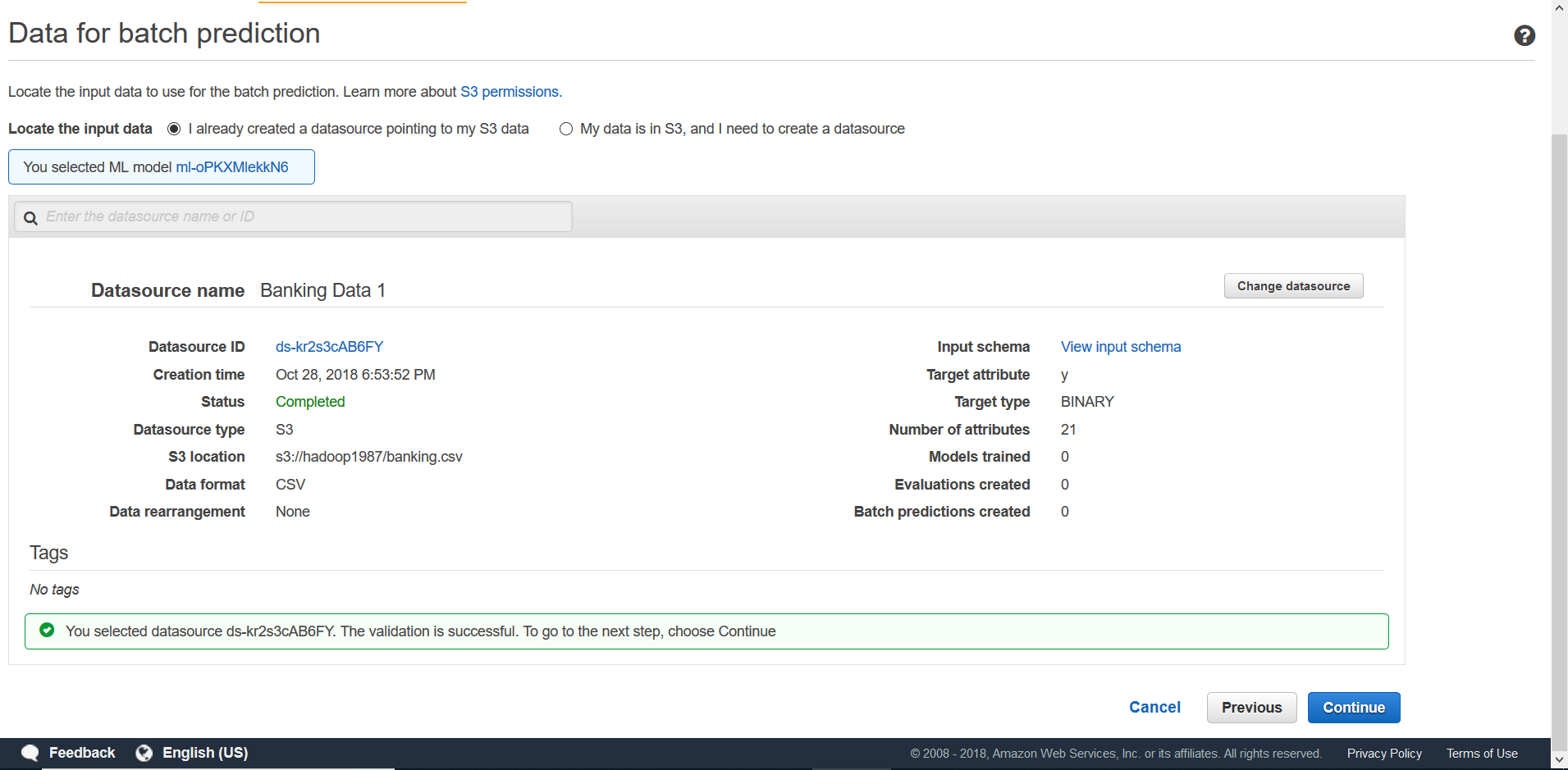


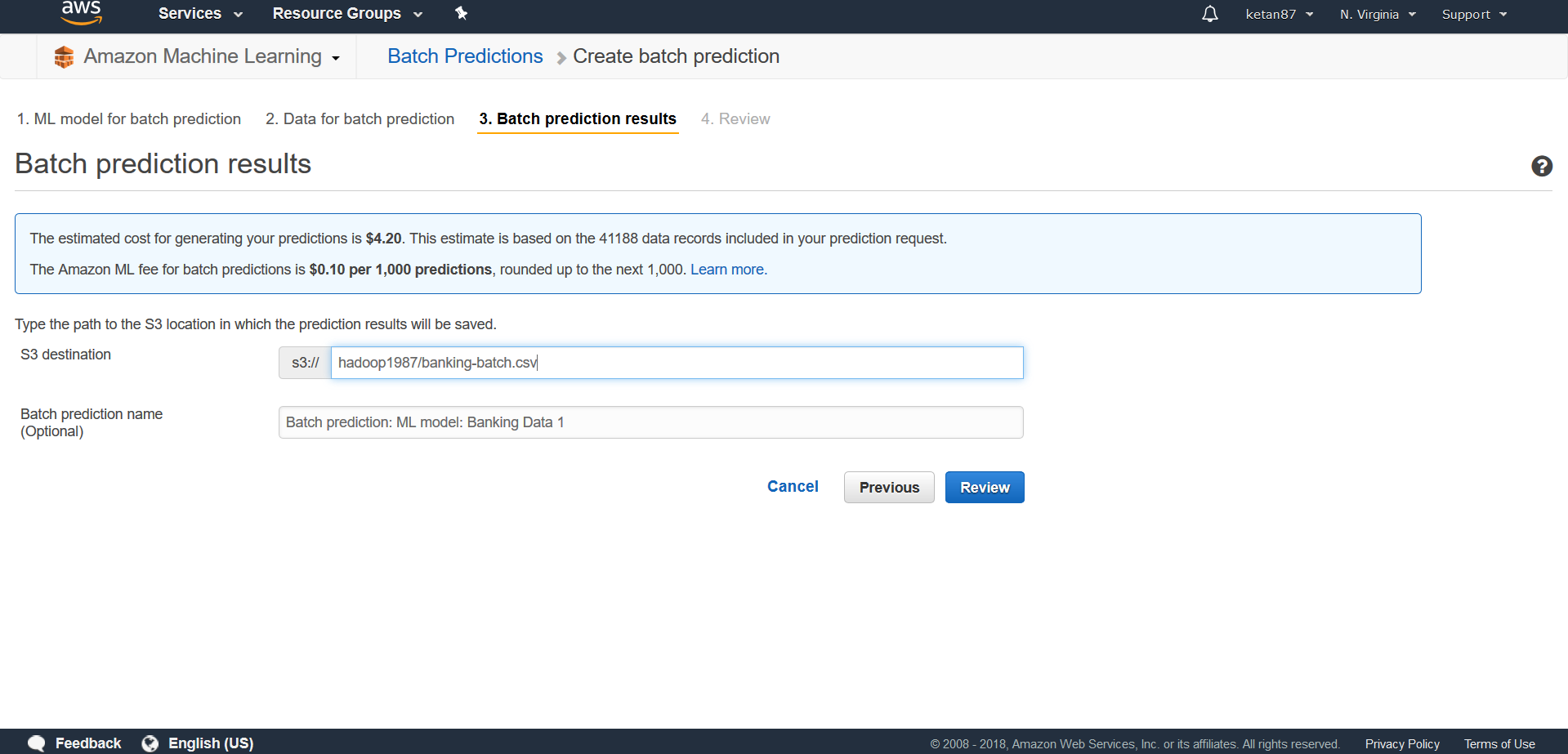


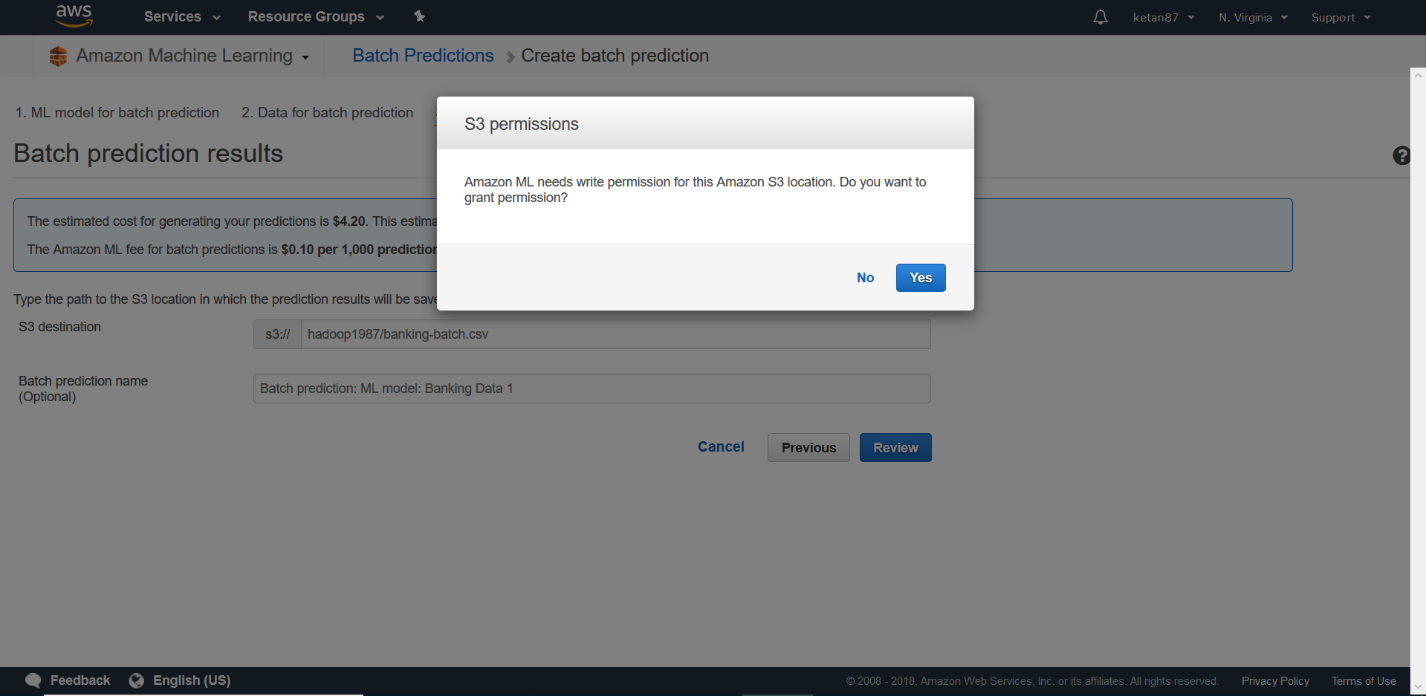


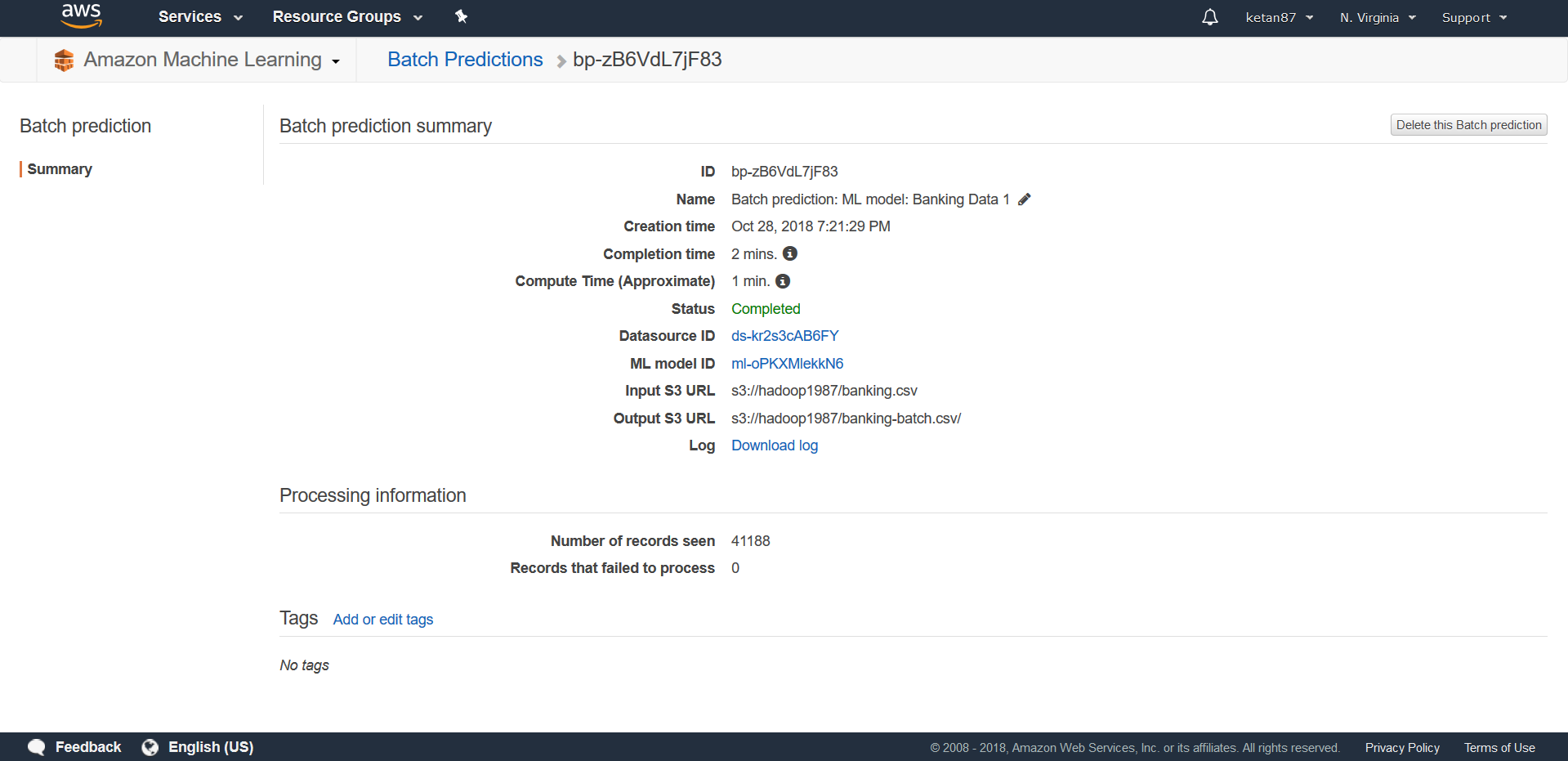


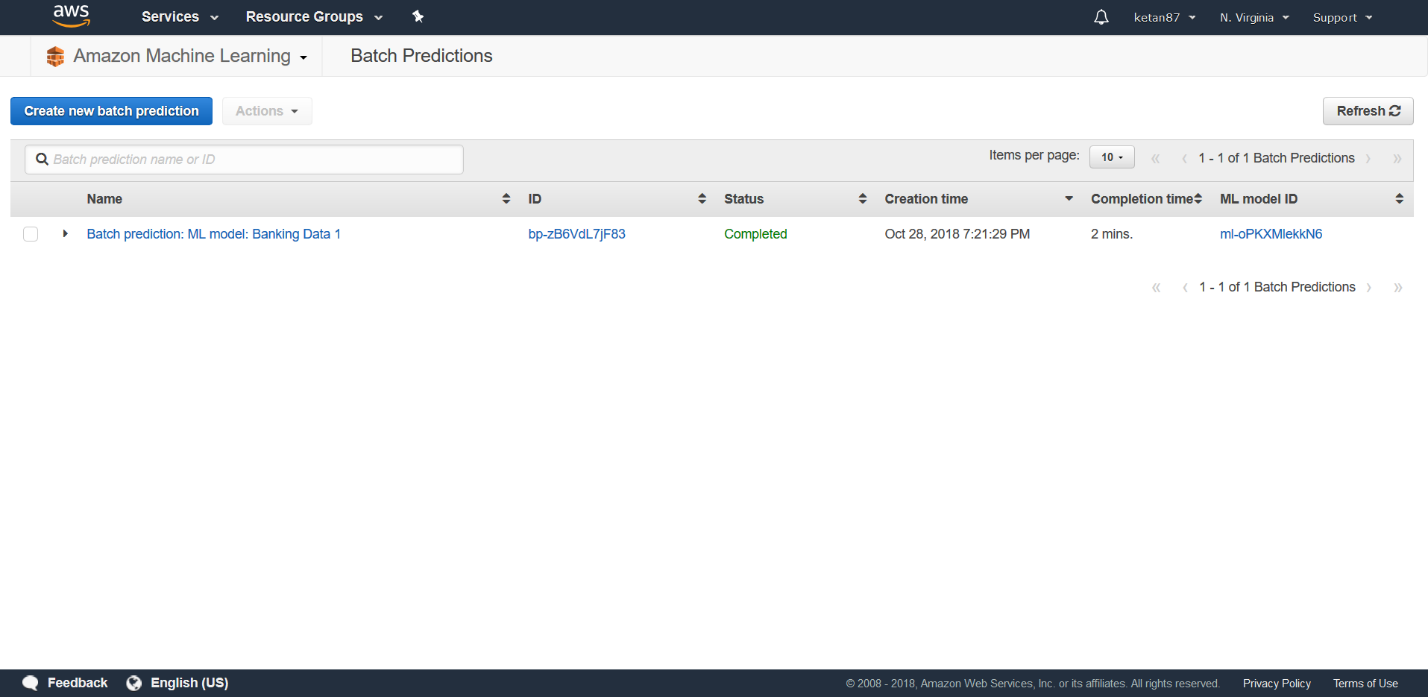


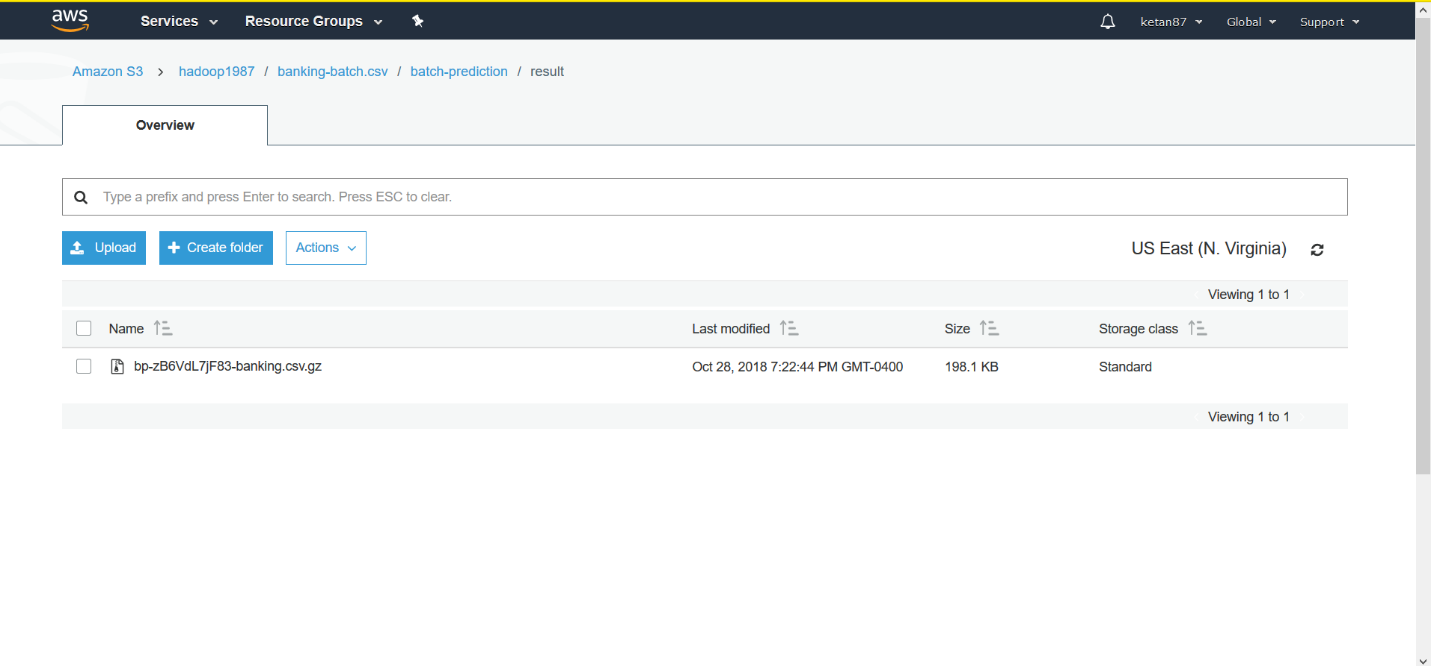












Total Time: 1 Hour