Paul Polsinelli

IFT 410 - 22485

3/14/2021

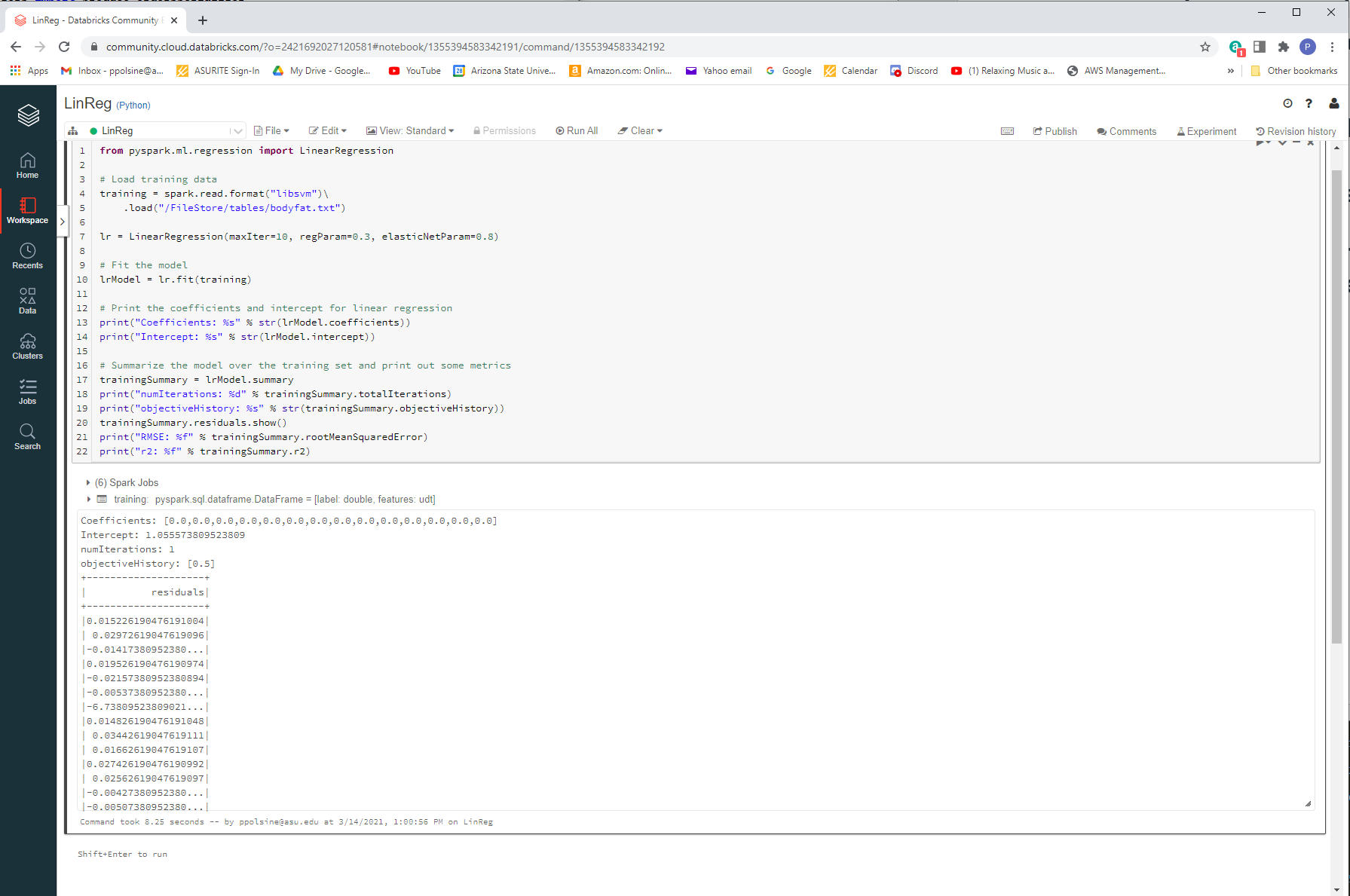
# Lab 7: Building a Regression Model

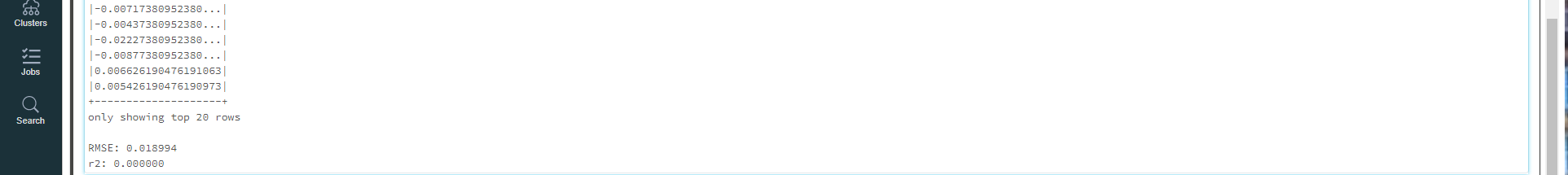
**Part I: Regression Models with Spark’s ML Library**

What is the RMSE of the regression model that the code built? 0.018994

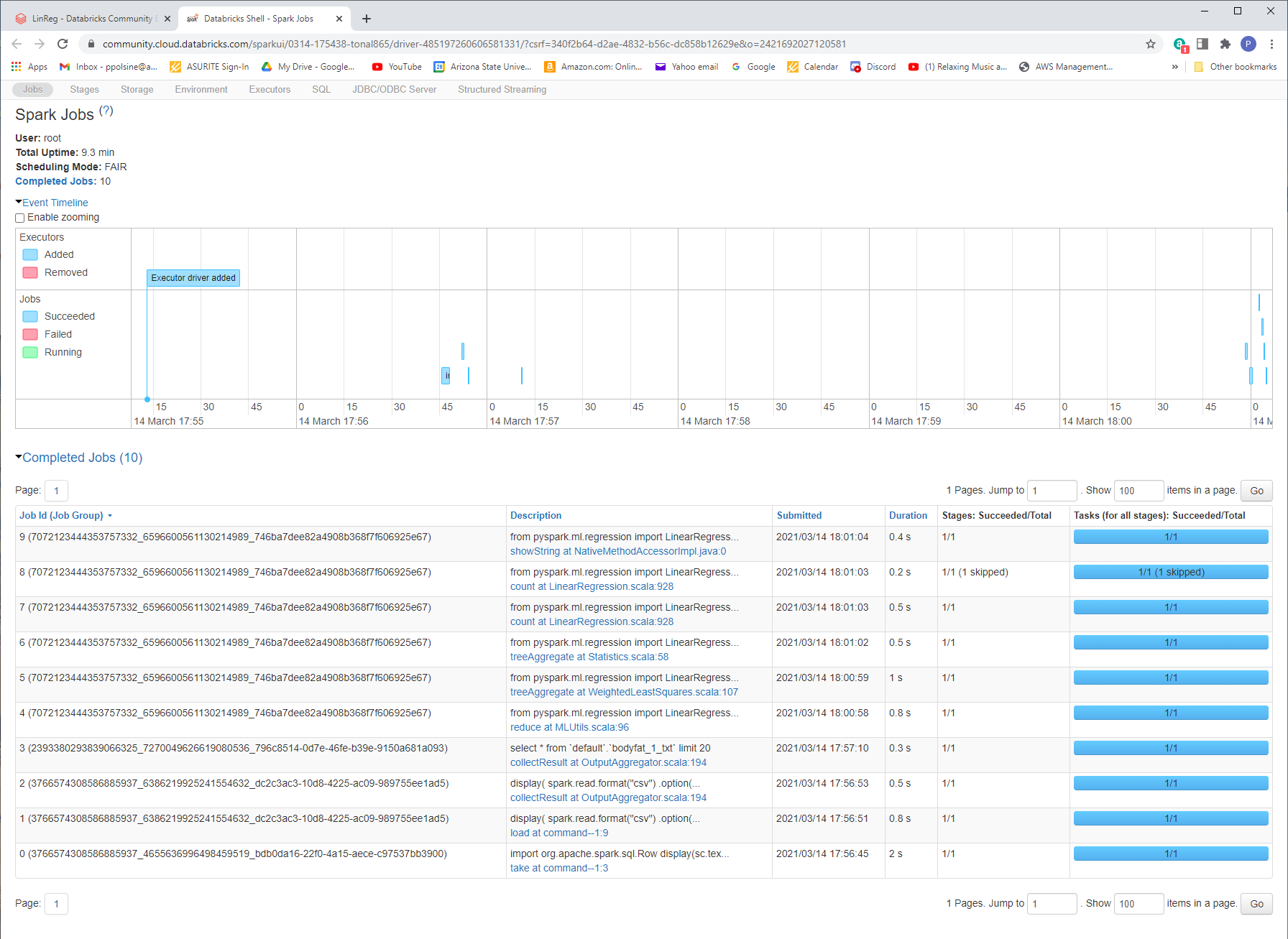
Is it high or low? As a value of standard deviation, it is very low

Research online for the meaning of RMSE, compare it to the SSE that was discussed in the lesson slides. Root Mean Squared Error (RMSE) is the standard deviation of the prediction errors (residuals), while the Sum of Squared Error (SSE) is the sum of all the Squared Errors (SE)s of each of the data points and is used to set the slope and intercept of the linear prediction. RMSE is used as a measure of concentration of data around the linear prediction.





How many jobs did Spark submit to run your code? 10



How many blocks was the RDD holding the dataset divided into? 16

