INSTRUCTIONS FOR EXECUTION:

There are three separate codes in the project.

1. One is for data pre-processing:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/5696471087116212/2890935191632296/7612795337787516/latest.html>

1. GBT code:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/5696471087116212/1334741296364335/7612795337787516/latest.html>

1. ALS code:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/5696471087116212/3934074860311288/7612795337787516/latest.html>

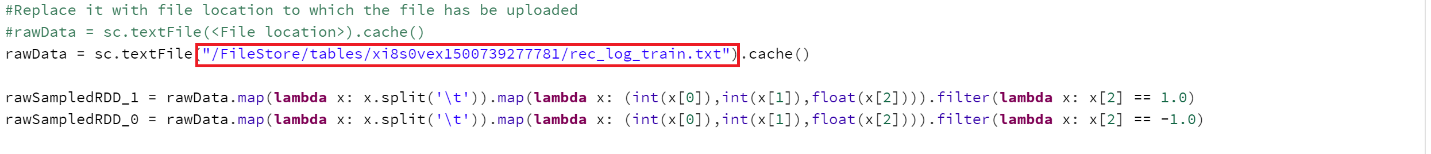
Execute preprocessing step first:

1. In the notebook, there is import notebook button on the top right. Import the project using this

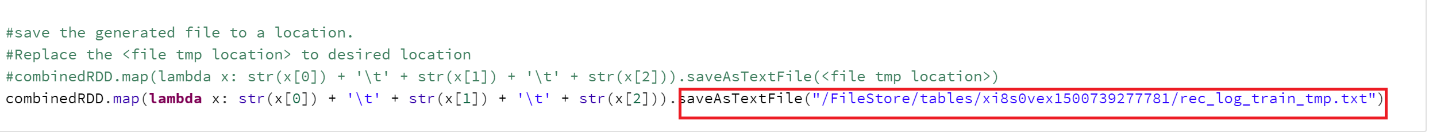


1. Upload the file rec\_log\_train.txt into cluster using upload option in the data bricks and copy the file location.

Replace the file location in the code with the new file location.



1. Specify the location of the output file to which the data must be saved and make note of the location specified.



1. Run the code.

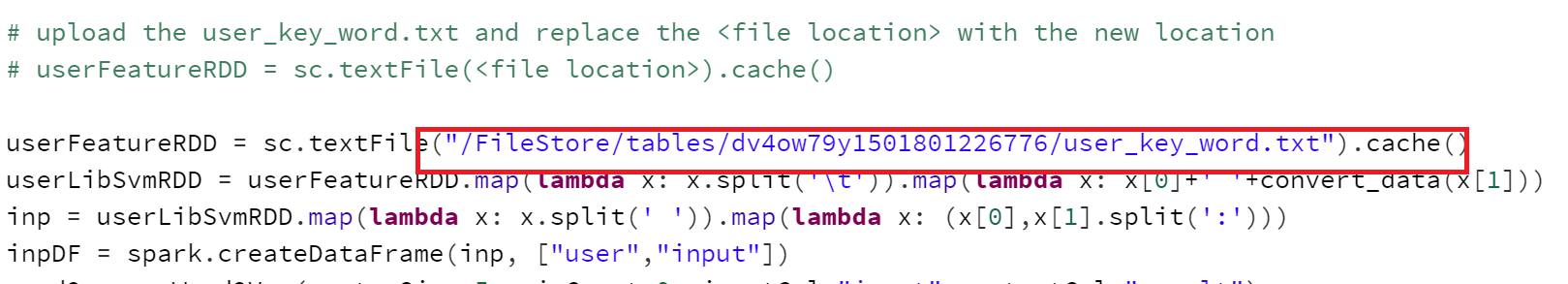
Executing the GBT code:

1. Import the notebook using the same method as in previous step.
2. Replace the file location in the first command with the temporary file location generated in the preprocessing step.

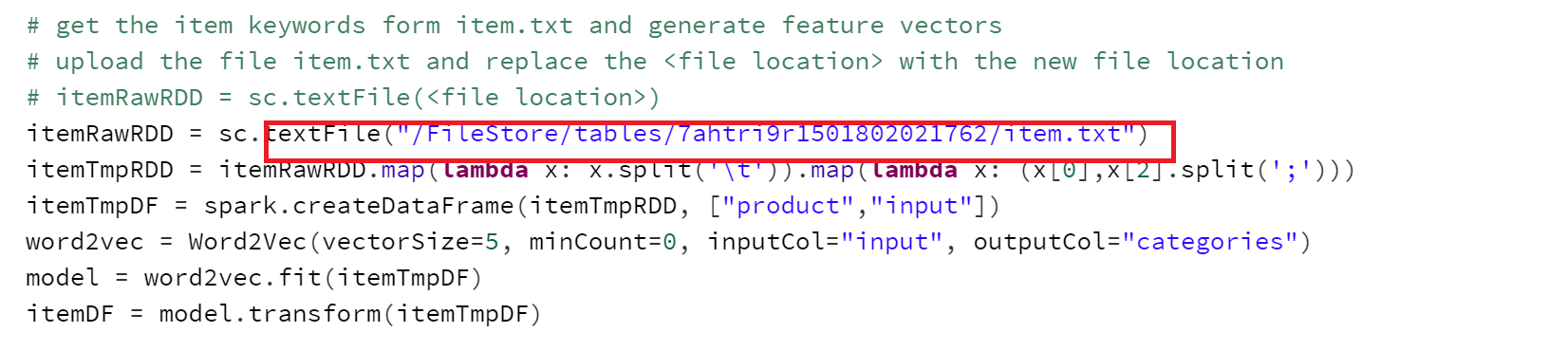


1. For this code Two additional files are required item.txt and user\_key\_word.txt. Upload both the files from the data directory. Make note of the file locations to which the files were uploaded.

* Replace user\_key\_word.txt file location with the new file location. The variable to which the data is stored is userFeatureRDD.



* Replace item.txt file location in the code with the new file location. The variable is itemRawRDD.

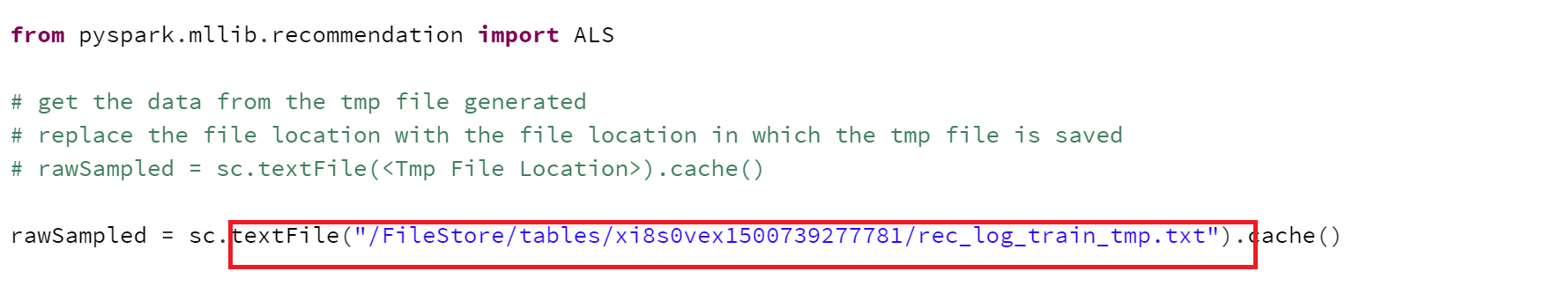


1. Run the code in order from the first command in the notebook.

Note: The training time for GBT model is approximately 1hr. Please be patient.

Executing the ALS code:

1. Import the notebook as in the previous notebooks.
2. Replace the input file location in the first command with the temporary file generated in the preprocessing step.



1. Run the code in order form top.