

Big Data Technologies

2023/2024

Lab 7

Run the virtual machine.

➤ Churn Prediction with SparkSQL and Spark MLlib

Perform the following tasks:

1. Download the files “churn-train-header.csv” and “churn-test-header.csv” from the Moodle website and upload them on HDFS in the directory “/lab7/”.
2. Download the Zeppelin notebook “churn_prediction_template.zpln” from the Moodle website.
3. Rename this notebook as “\$NAME_churn_prediction.zpln” where “\$SNAME” denotes your family name.
4. To start Zeppelin, run the command “zeppelin-daemon.sh start” in a terminal. Then, open the webpage “<http://localhost:8080/>”
5. Upload the notebook in Zeppelin.
6. Answer to the questions included in this notebook.