Big Data Technologies

2024/2025

Lab 7

➢ **Churn Prediction with SparkSQL and Spark MLlib**

Perform the following tasks:

1. Build the image "zeppelin:v1" after downloading the Dockerfile from the Moodle website
   ```
   docker build -t zeppelin:v1 .
   ```
2. Run a container from the image "zeppelin:v1"
   ```
   docker run -it -p 8090:8090 zeppelin:v1
   ```
3. Open the webpage http://localhost:8090/ to use Zeppelin
4. Download the Zeppelin notebook "churn_prediction_2024_template.zpln" from the Mooodle website.
5. Upload the notebook in Zeppelin (click on "Import note" on the Zeppelin welcome page). This notebook exploits the data files "churn-train-header.csv" and "churn-test-header.csv" that have been downloaded in the directory "/home/user/" when the image "zeppelin:v1" was built.
6. Rename this notebook as "$NAME_churn_prediction.zpln" where "$SNAME" denotes your family name.
7. Answer the questions included in this notebook.