This section has all the notebooks used for data preparation, for building machine learning models as well as Spark notebooks from Databricks.

Below is the information about the folder and the corresponding files.

**ApacheSparkStreamingPrototype**

StreamingPrototype.ipynb - Illustration of how apache spark can be used for real time analysis of HistoricCTR

GenerateStreamingFiles.ipynb contains sample code for breaking a huge excel file into many small files for simulating streaming purposes.

**Data Preparation Notebooks**

AWSS3Boto – code for putting up the data programmatically to S3 bucket

FetchData.ipynb – SQL queries used for data preparations

RFE.ipynb – Illustration of how sklearn rfe can be used for feature importance

DataJoins.ipynb – merging data from RDBMS into a flat file

**Deep Learning Notebook**

Deep Learning model using Keras

**Linear Regression Notebook**

LinearRegression.ipynb - Linear Regression model applied to the datasets

**Logistic Regression Notebook**

LogisticRegression.ipynb - Logistic Regression model applied to the datasets

**LogisticRegression\_ApacheSpark**

LogisticPythonNotebook.ipynb - MLLib model applied to the datasets

**XGBoost Notebook**

XGBoost.ipynb - XGBoost model applied to the datasets