

Module 11 – Apache Spark Streaming

Case Study II

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Case Study: Synthetic Financial Domain: Financial

There is a lack of publicly available datasets on financial services and specially in the emerging mobile money transactions domain. Financial datasets are important for many researchers at performing research in the domain of fraud detection.

Tasks:

Now with data pipeline ready, you are required to develop the model and predict the fraud using spark streaming.

1. Explore the dataset and develop a model to predict the fraud.
2. Develop the application to train the model and persist the model to disk.
3. Create a new spark streaming application for the same
4. Application will connect to the flume to retrieve the data
5. Load the model
6. Predict the fraud and print the result to the logs
7. Test the application by sending dummy data rows from the consumer

Data:

Paysim synthetic dataset of mobile money transactions. Each step represents an hour of simulation. This dataset is scaled down 1/4 of the original dataset which is presented in the paper "Paysim: A financial mobile money simulator for fraud detection".