Indium Software co-develops a real-time collaborative Fraud Analytics solution to combat Identity Theft.

Fraudsters have stolen $112 billion in the past six years. That equals $35,600 stolen per minute, or enough to pay for 4 years of college in just 4 minutes. (Al Pascual, 2016)

Situation

The client is a U.S based Fraud Analytics startup, who envisions to provide an end-to-end real-time fraud detection solution platform and analytical support. The platform would help customers to identify specific transaction or user behavior patterns through a combination of Machine learning and human ingenuity.

The worldwide losses on payment card frauds hit $16.31 billion on 2014 (PYMNTS, 2015), and it's expected to double by 2020. The existing fraud detection strategies are reactive and takes days to detect fraud. The best way to combat today's sophisticated fraudsters, at scale, is to leverage the insights of Fraud experts and the speed of Machine Learning.

The client chose Indium Software for its prior successful deliveries of similar high performance and scalable Big Data Analytics solutions, flexible engagement models, and on-demand access to a pool of Top Data Science, Big Data Infrastructure & Predictive Analytics talent. Indium Software was commissioned to design a high performance, scalable, interactive fraud analytics solution that offers the below:

- Precise identification of fraudulent customer/transaction.
- Automated variable and rule generation.
- Collaborative anomaly detection algorithms

Solution

Indium Software leveraged open source technologies from the Apache ecosystem for their high flexibility, scalability, throughput, and for the thriving community of supporters / developers. A sample model was built and tested using the test dataset provided by the client, before finalizing the architecture.

- Developed the core application in Java.
- Used Apache HDFS (Hadoop distributed file system) for data storage; due to its high throughput access to very large data sets.
- Integrated MLlib library from Apache Spark, to enable automated variable and rule generation.
- Model developed using Logistic Regression, to identify fraudulent transactions/customers.
- Decision tree was used to let users develop variables on the fly.
- Java was leveraged to generate user scores, update database and refresh variables, all at real-time.
- Apache Kafka queue was used to provide low-latency high throughput platform, and to send and receive data streams.

Technologies Used
Java, Apache HDFS, Apache Spark MLlib, Logistic Regression, Decision Tree, Apache Kafka Queue.

Duration
8 months (ongoing)

Resources
1 Data Scientist
1 Big Data Analytics Architect
2 Sr. Java developer

Engagement Model
Offshore

Result

Below are some of the key benefits of this architecture, identified based on similar past Big Data Analytics engagements and the prototype test results.

- Higher performance – real-time user profile scoring – at scale.
- Easily scalable to handle larger data volumes.
- Interactive fraud analytics.
- Higher accuracy – precise identification of specific transaction/customer.
- Lower cost; as solution is designed to run on commodity hardware.

For more information, contact: **info@indiumsoftware.com**