Credit Assessment Engine

- a case study

Objective -

DataToBiz consultation is required create a robust credit scoring engine using Machine Learning (ML), Natural Language Processing (NLP) and alternate means of data (does not contain traditional credit bureau data), helping SME Lenders and Other Financial Institutions to give hassle-free loans in real time with minimum default rate.

Client -

A Small and Medium Enterprise (SME), giving small loans to the population comprising majorly of unbanked people.

The Client is an African based SME, trying to capture business out of larger segment of unbanked population of Africa. Though, it is very easy to decide about loan approval amount based on Credit Bureau data, but public data infrastructure is not that much strong over there. So client felt a need of a credit engine based on alternate sources of data.

Approach -

Alternate data means the non-traditional data available. It does not include traditional Credit Bureau data. We have focused on one of the most prominent alternate data source i.e. mobile device data to build the model. Using NLP, we built an engine which extracts various features such as income, expense, investments from the SMS data obtained from the mobile device of the customer.

Along with this, other features like type of mobile, already installed apps helped us to understand more about interests.
Also, the contact list helped us to understand the affinity of customer to do fraud.

Once we had such a data, we assigned a credit score to every customer based on the above parameters. Initially, it was a rule based engine to assign the credit score, but eventually as more and more data got accumulated, an ML model was built on top above of data points to generate credit score automatically.

Method -

In order to achieve the above approach, following flow/method was chalked out -
1. Data Collection from Customer Phone Device
2. Data Warehousing
3. NLP Engine to fetch desired information from data.
4. Credit Scoring Engine
   A. **Data Collection -** Using a mobile SDK, all the SMS data from customer's phone was fetched on customer's consent and stored into the server database.
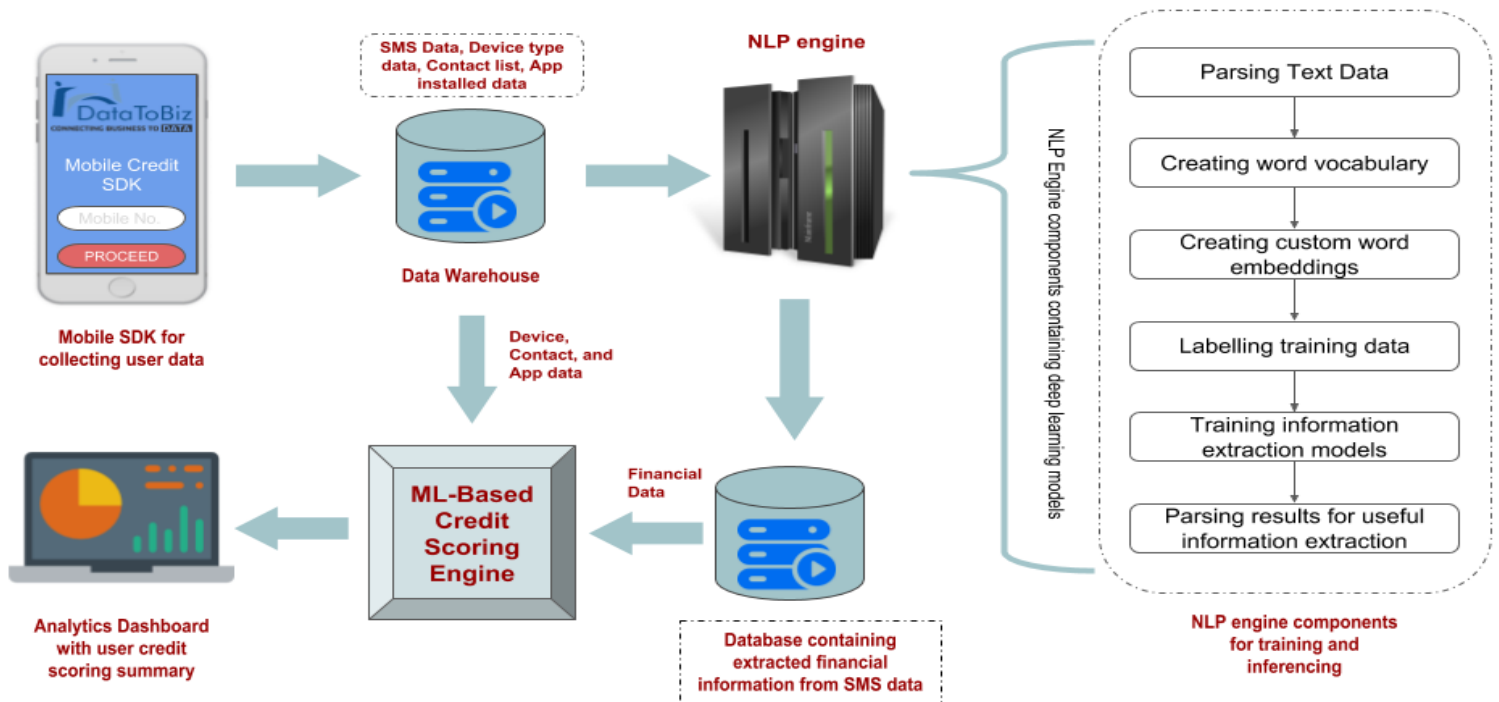
B. **Data Warehousing -** Once we had all the data points from different data sources, it was warehoused into a common data warehouse. Once warehousing was done, data was transformed in a way that is readily consumed by our NLP engine and Credit scoring Engine.

C. **NLP Engine/Text Analytics -** First of the two engines written was NLP engine. NLP engine was responsible for reading the raw SMS from the customer and fetching out various information like income, expense, investment amounts. Running this engine over all the sms of customer, gave us the financial strength of the customer and trend of this financial strength over the various months.

D. **Credit Scoring Engine -** Once we got the financial strength of the user, this data along with device type data, contact list data and type of apps installed data was fed into our credit scoring engine. Initially, to solve the cold start problem, this engine was rule based - having given different weights to different data points. Once we had enough data points (data of loans given along with frauds and non-frauds outcomes), we trained our Machine Learning model extensively and shifted to this ML based approach for calculating the credit score.

Result - Our ML engine based credit scoring engine showed a great result in place of traditional loan approval mechanism.

- **Loan Approval Time** was reduced from 3 days to 30 minutes.
- **Default Rate** was reduced from 17% to 7%.