# Introduction-Sell more

## Project Background

Nowadays, understanding customers’ decision behavior is significant in different applications. In many cases, customers might not even be aware of their motivations or their behaviors. Understanding consumer behavior is essential for the success of any organization. Organizations want to analyze and understand consumers’ behavior in order to generate products that will be embraced and purchased, so they can sell more products and earn more profits.

Thus, we aim to collect customers’ data, then analyze and interpret them by using the machine learning models and generate some recommendations for customers’ optimal decision making, finally we need to visualize the output and interpretation to help people understand it.

In order to improve the accuracy of the output from the machine learning model, we also use the influence function techniques. Because this technique can trace the machine learning model’s prediction and back to its training data, which can differentiate the data-set input and detect dataset errors. Thus, we can adjust and optimize the training data such as fix some mislabeled examples and debug some domain mismatch, which can optimize the machine learning model so that to improve the accuracy of predictions or recommendation of customers’ behaviors.

## Project Tasks

The main objective for us is to develop a system with

* Data input; (we can use some online data online as examples at first to testify system implementation);
* Do research and understand machine learning models and their selection (the models could be Random Forest, SVM, Naïve Bayes);
* Get the output from machine learning model, and the influence based model interpretation;
* Visualization of these steps (especially model output and model interpretation)， so that users can understand the model and output more easily;
* Integrate all these steps into a framework.

## User Roles Description

* Normal customer: people who want to get more product recommendations and related information.
* Organization: organizations which focus on selling more products to their customers.
* System users: people who would like to use our system to analyze the data from customers. System users can understand the output and make use of it.

## User Stories

* As a normal customer, I want to know what the products recommendations generated from the model, so that I can view the related product information to help me make right decisions.
* As an organization, we want to analyze consumers’ behaviors so that we can provide the products that are more likely to be purchased by corresponding customers.
* As an organization, we want to easily understand the output from the machine learning model, so that we can use the result more efficiently.
* As an organization, we want to differentiate the value of the data which are input in the machine learning model, so that we can optimize the input data.
* As an organization, we want to improve the accuracy of product recommendations, so that we can attractive more consumers and provide the most related products.
* As a system user, I want to view all the data-set which I can use in the shell so that I can choose a data-set as input data.
* As a system user, I want to fetch the data-set in the shell so that I can see the detailed information of this data-set.
* As a system user, I want to choose the features and labels of the data-set I fetched before so that I can do the machine learning for this data.
* As a system user, I want to train the data so that I can see the difference between train score and test score.
* As a system user, I want to view the visualization of the output from the model so that I can easily and directly understand this output.
* As a system user, I want to know about the influence of the output from the input data-set so that I can fix the features or labels of the input data-set.