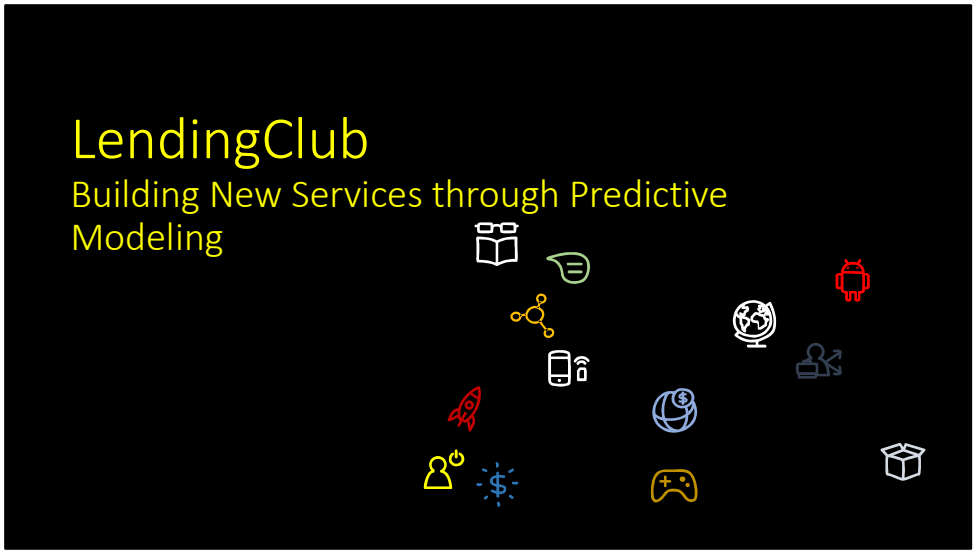
LendingClub

Building New Services through Predictive Modeling

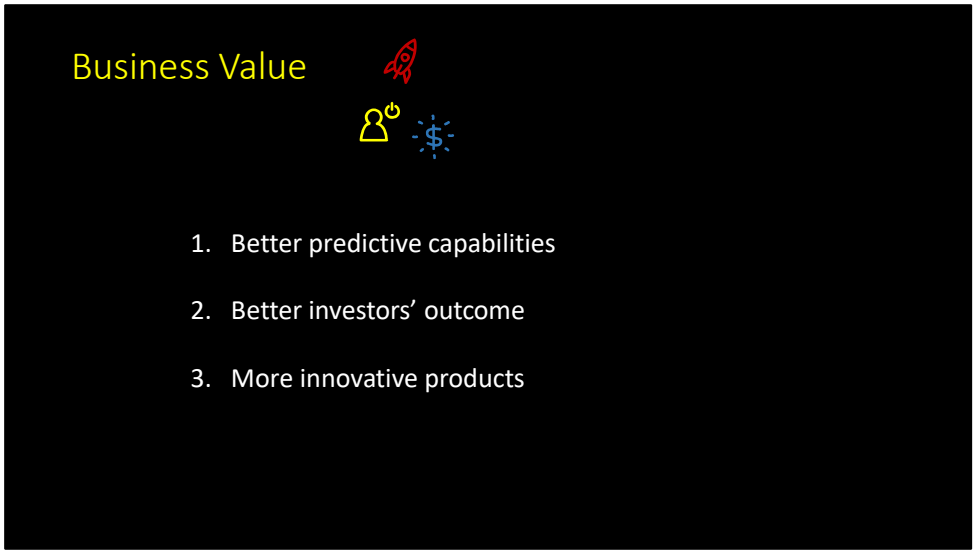


I will walk you through three core talking points.

Problem Statement

Determine a good and bad loan before and during the LendingClub loan approval process

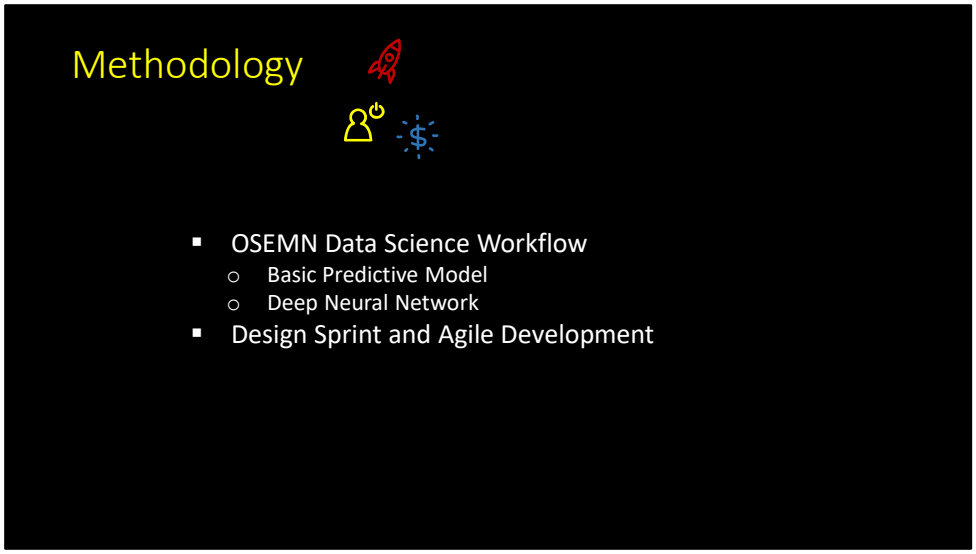
First stop – what is the problem that we are trying to solve today.



Second stop – let's talk about business value of the problem we are trying to solve.

Better Predictive and Investor Outcome
We are able to create a better predictive outcome of loan default.
As a result, we will be able to attract more investors interest because they have higher confidence to invest.

More Innovative Products
Through understanding LendingClub's domain problem, and with predictive modeling modeling, we can design forward new services for LendingClub.



I have breakdown my methodology into two parts. First, is the data science workflow where I leveraged on OSEMN as a guide post; and based on the workflow, I have started to obtain the data, scrub the data, explore the data, model the data, and interpret the data. In predictive modeling, I have used the basic predictive model, and deep neural network. Second, I have used the design sprint and agile development to attempt to build new services.

Results

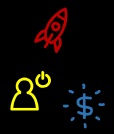
The results based on basic predictive deep neural network are almost identical – swinging between 78% and 79% test accuracy

Basic Predictive Modeling

1. Logistic Regression : 78.5%
2. Random Forest : 78.5%
3. XGBoost : 77.2%
4. SVM : 78.5%

Deep Neural Network Modeling

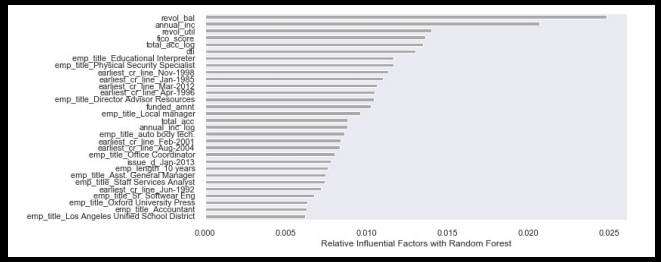
1. Layer 1: 79%
2. Layer 2: 78%
3. Layer 3: 79%



There is not much different between the basic, and more complicated deep neural network. If the basic predictive modeling have almost similar predictive power, then we should use the basic model to save effort and cost.

Results

Based on Random Forest – relative influential factors to predict bad and good loans.

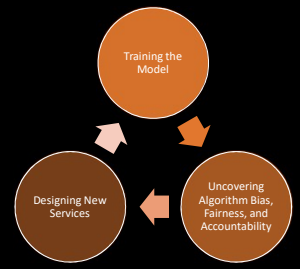


This is the relative importance / predictors for LendingClub to consider when building new algorithm, and new services

Conclusion

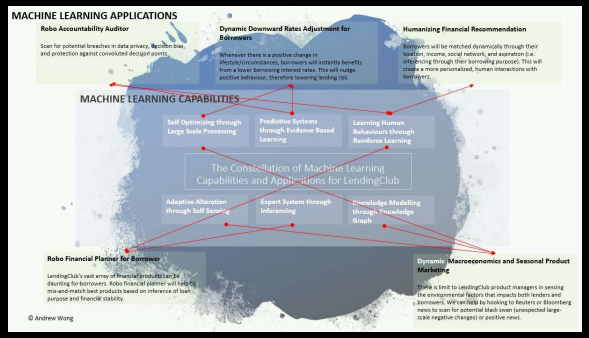
Product-Driven Machine Learning

The learning-feedback cycle
Constant feeding of updated datasets, as well as new type of datasets.



Conclusion

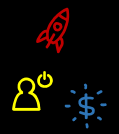
What Machine Learning capabilities that LendingClub can build based on the predictive modeling that I have built (no doubt it is not as simple as that!).



Based on the predictive modeling results, I have attempted to design forward new capabilities, and applications for LendingClub.

Future Work

1. Faster and more scalable processing in the cloud
2. Focus on predicting investors' risk behaviour



For future work. First, I'd like to run the dataset in the cloud to save time and effort. Second, the current problem statement focus on borrowers' behaviour. I'd like to shift my focus on investors' behaviour, especially on their risk management.


Q &A