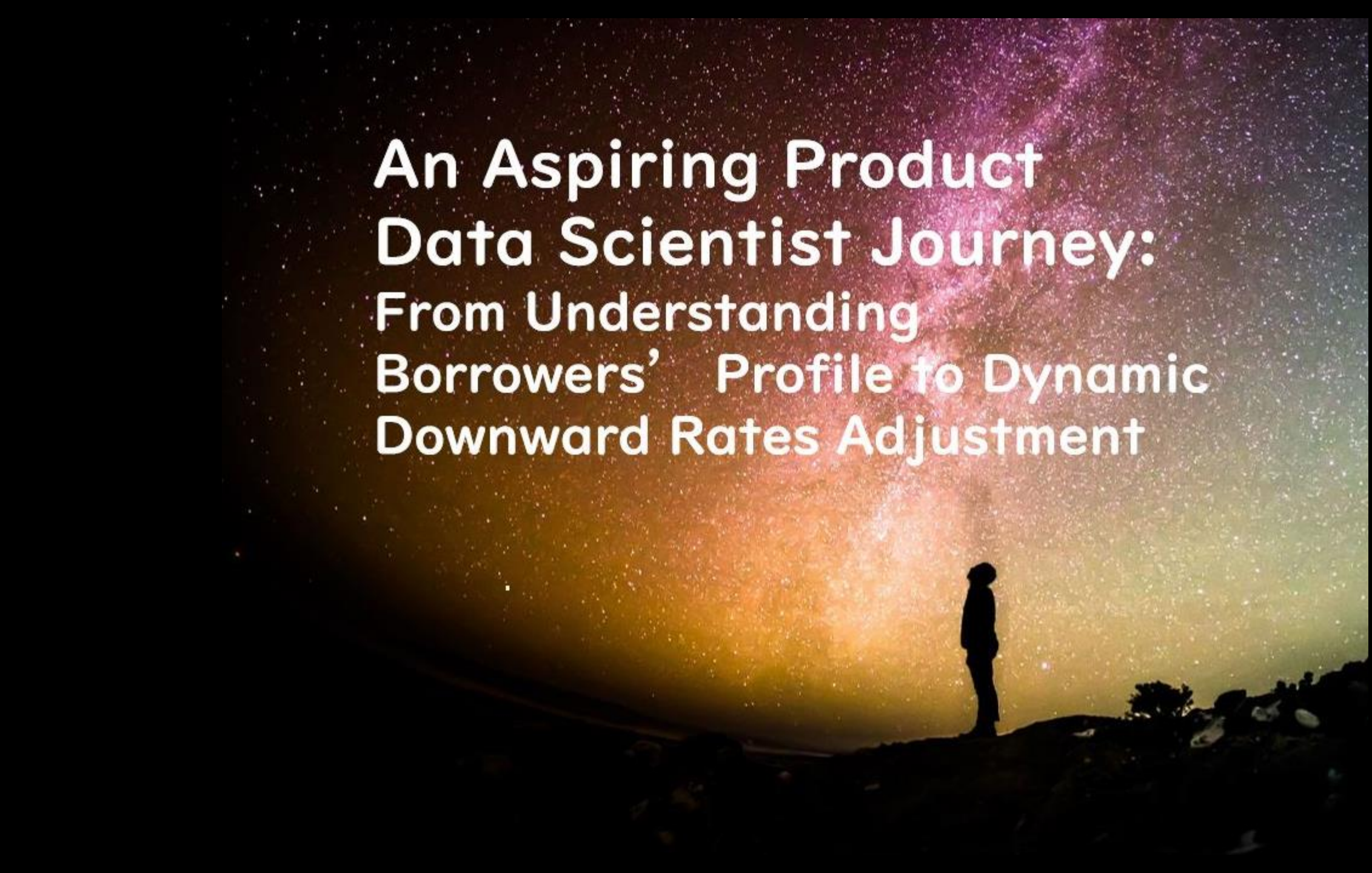

An Aspiring Product Data Scientist Journey:
From Understanding Borrowers' Profile to Dynamic Downward Rates Adjustment

Objectives

- Run an end-to-end data science project based on Lending Club dataset and its domain problems.
- Model/engineer a predictive model that could determine a good and bad loan before and during the loan approval process.
- Take you reader on a journey on how I will create AI/Machine Learning product/services.

Approach

1	2	3
START HERE WITH CUSTOMERS IN MIND THROUGH DESIGN SPRINTING	THEN GO THROUGH THE DATA SCIENCE JOURNEY GETTING TO PREDICTIVE MODELLING AND INTERPRETATION	WORK THROUGH MVP AND MMP IN PRODUCTION
The sprint is a five-day process for answering critical business questions through design, prototyping, and testing ideas with customers.	**OSEMN Data Science Workflow**	**Machine Learning MVP (Minimum Viable Product): Test different value propositions based on predicting modelling outcome and leveraging on human-in-the-loop**

DAY 1	DAY 2	DAY 3	DAY 4	DAY 5
Make a map and choose a target problem	Sketch competing solution	Decide on the best solution	Build a realistic predictive model or prototype	Test with target customers

1. Obtain	2. Scrub	3. Explore	4. Model	5. Interpret

Integration >	Deployment >	Post-Deployment

Machine Learning MMP (Minimum Marketable Product): Design the smallest feature/benefit set that addresses the customer needs and creates the right customer experience.

CYCLING ITERATIVELY THROUGH AGILE DEVELOPMENT AND DELIVERY TO BRING CONSUMABLE PRODUCT VALUE TO INTERNAL / EXTERNAL CUSTOMERS

Predictive Needs/ Requirements Backlog	Agile Data Science Team	Testing Working ML Model
	1. Iterative 2. Deliver small, but consumable, increments 3. Fast feedback cycles 4. Predictive needs/requirements are evaluated continuous as a natural mechanism for responding to change quickly 5. Continuous improving data and modelling	

© Andrew Wong

Results

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

Basic Predictive Modeling

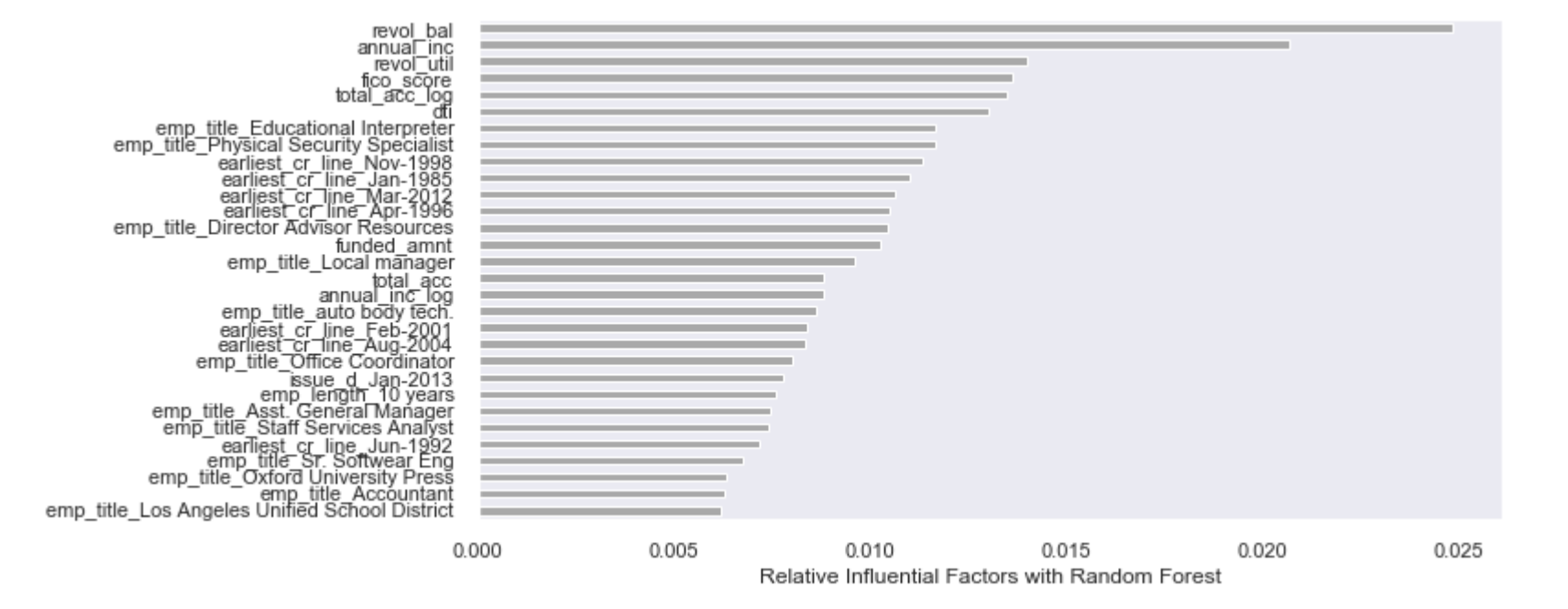
1. Logistic Regression : 78.5%
2. Random Forest : 78.5%
3. XGBoost : 77.2%
4. SVM : 78.5%

Deep Neural Network Modeling

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

Results

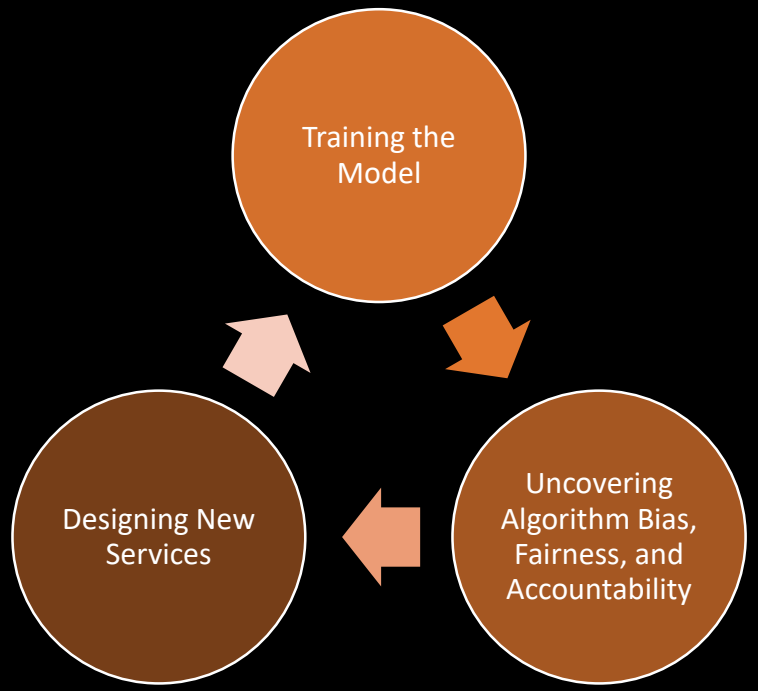
Based on Random Forest – relative influential factors to predict bad and good loans.



Conclusion

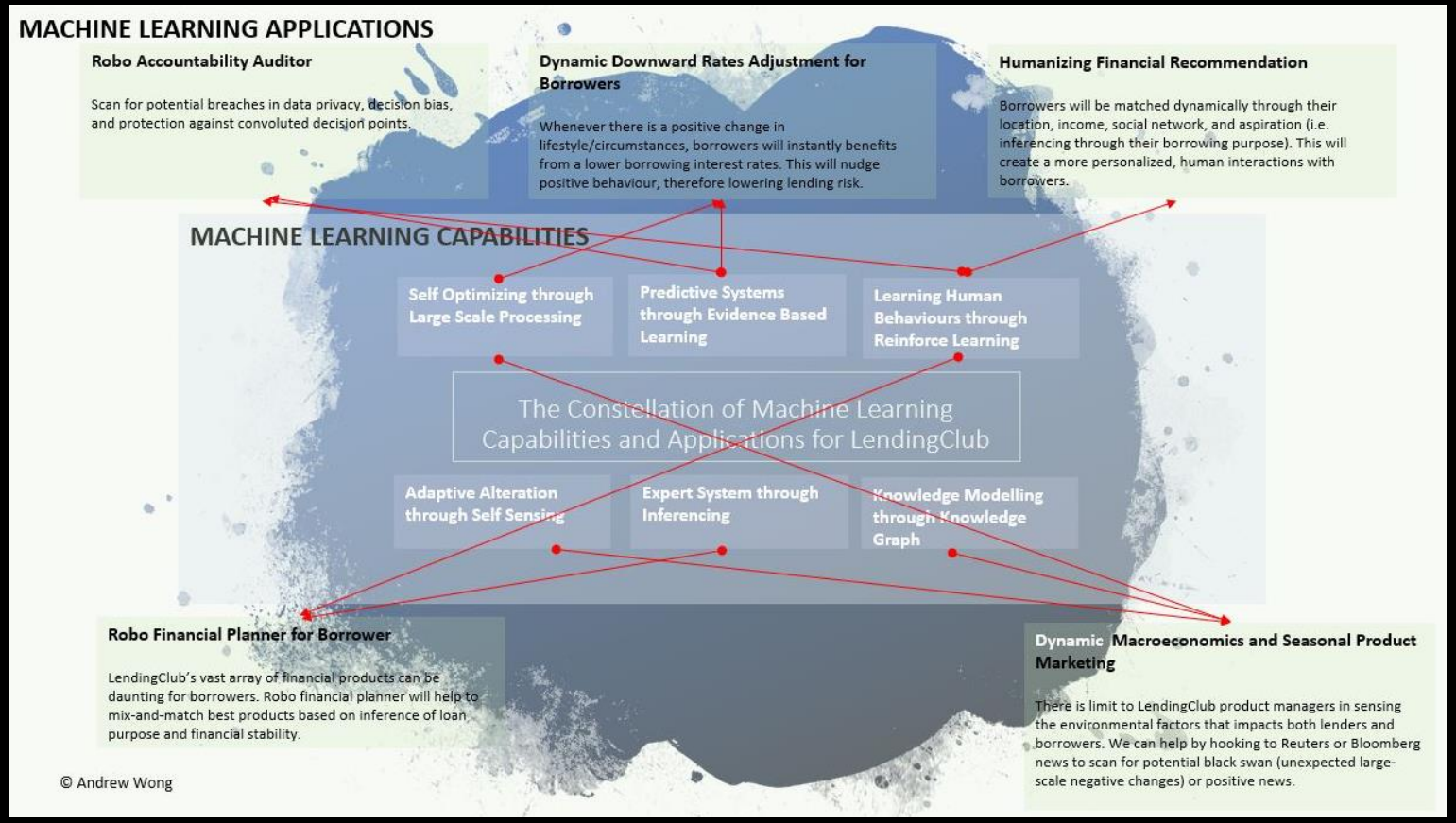
Product-Driven Machine Learning

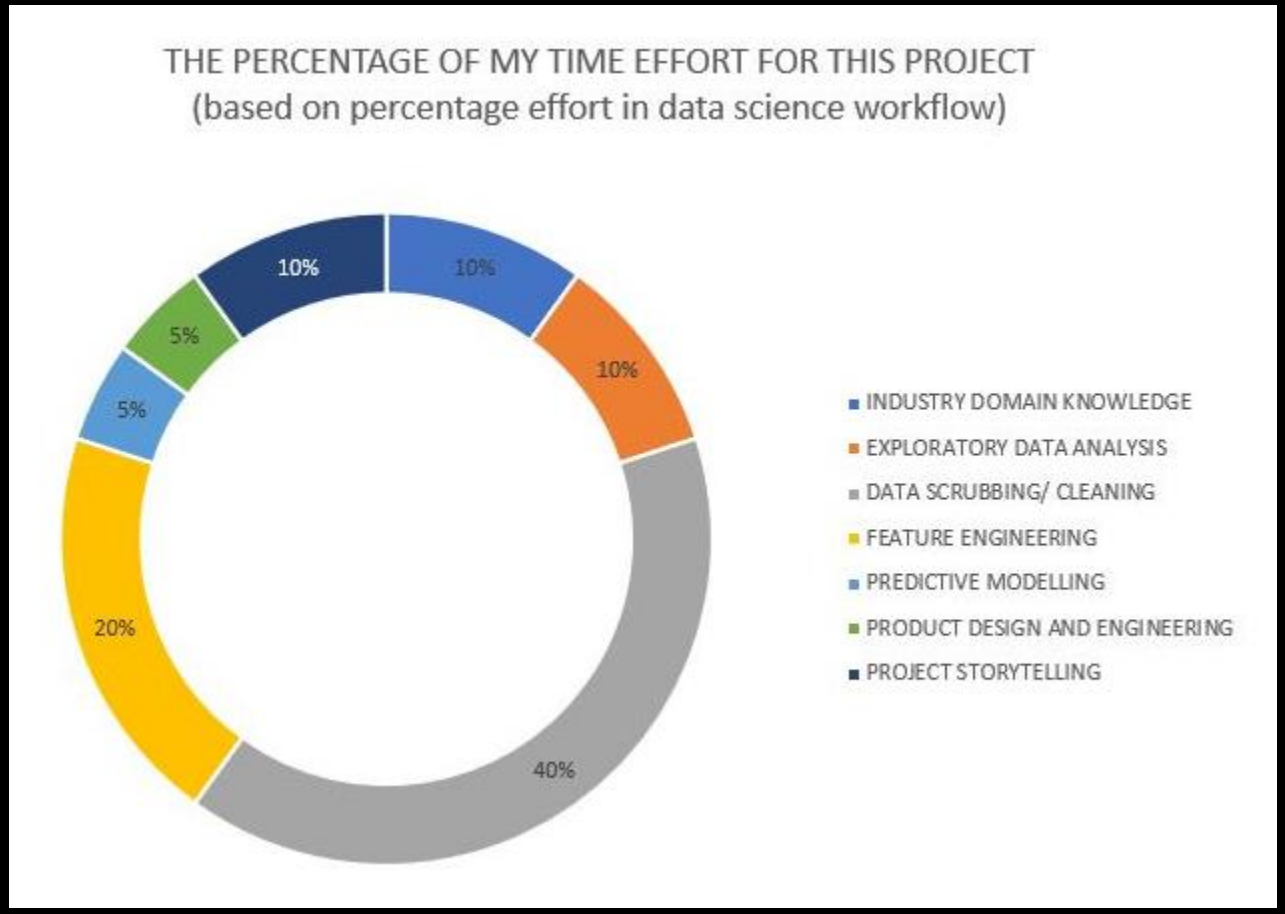
The learning-feedback cycle
Constant feeding of updated datasets, as well as new type of datasets.



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!).



Q &A

Appendix – How I Spent My Time and Effort