# From Unstructured Text Data To Interpretable Financial Prescriptions: An Optimization Approach

Pengcheng(Pace) Han<sup>1</sup>; Tiantian Ye<sup>1</sup>; Dimitris Bertsimas<sup>2</sup>; Ryan Cory – Wright<sup>3</sup>; Nadine Kawkabani<sup>4</sup>; Brian Shaw<sup>4</sup>; Brendan Mannix<sup>4</sup> <sup>1</sup>MIT MBAn 2019; <sup>2</sup>Faculty Advisor, MIT; <sup>3</sup>PhD Student – mentor, MIT; <sup>4</sup>Mentor, MFS Investment Management, Boston, US

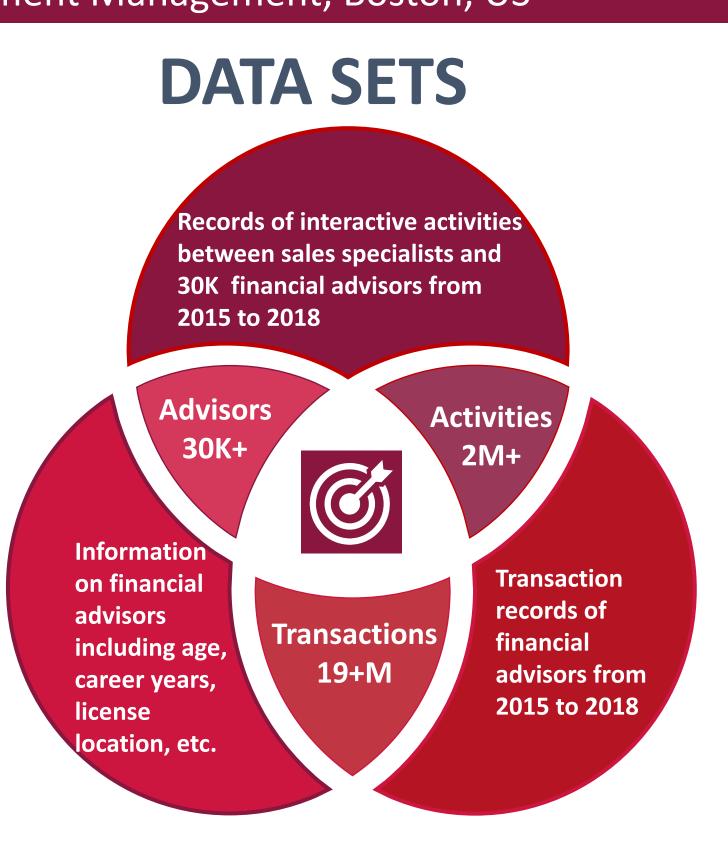
# **ABSTRACT**

For over 90 years, MFS has maintained its leading position in the financial services marketplace by being a customer centric company. To maintain its competitive position, MFS has enacted a policy where sales staff regularly interact with their clients on a personal basis, by regularly emailing, calling and meeting their clients. After each of these interactions, the data generated is recorded in MFS's internal system. In the age of big data, how to utilize those data to help MFS expand its market share becomes a new challenge MFS faces. In this project, we assist MFS in extracting valuable information from unstructured data on activities with one of the 300+ partner financial advising companies of MFS. This allows MFS to understand the influence of different activities on client relationships, and manage sales activities in a more efficient manner.

Which Advisor

What Format What Content How Frequent

#### **TIMELINE Receive Data** March Data Cleaning **Discussion With** April **Business Team About Project Goal** May Feature Engineering June **Prediction And** Prescription July **Optimization And** Final Deliverables August



### **Activities Categorized Into 4 Types Based On Content**

(6.8%)

(15.0%)

**Product Sale** 

Non Business Topic

Value

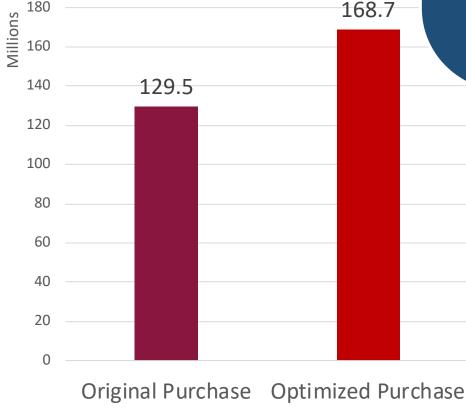
Added

(26.8%)

## **Identified 173 Keywords And Phrases** distributor estimate = TINNET TOWN Inbound Call<sup>©</sup> interest specialist value

Graph1: Word cloud of 173 key words we identified, the larger the word/phrase, the more important it is in activity content modeling 5.Optimization

#### **Optimization Model Lifts Total Future Purchase Prediction**

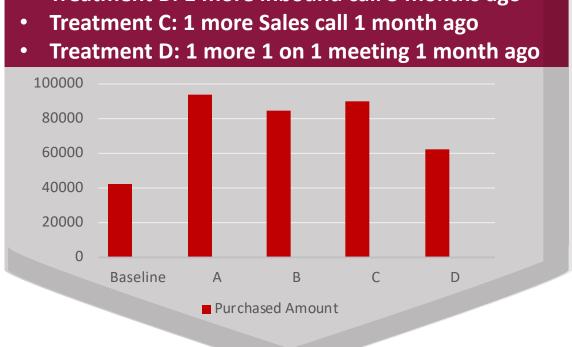


Graph5: Influence of optimizing current month sales resource allocation on the predicted purchase amount next month on 100 financial advisors. Note due to compliance we do not have information on actual sales capacity, there is about 30% increase for the total purchase amount compared to original without optimization

## **Better Understand Activity Influence**

The prediction model gives a better understanding of how different frequency and different types of activities influence future purchase amount. For example, we examine by adding a specific activity, what change will bring compared to baseline which includes 2 inbound call 3 months ago, 2 sales call one month ago and over 10 market fact content conversation over 6 months.

- **Treatment A: 1 more in bound call 1 month ago Treatment B: 1 more inbound call 3 months ago**



#### 1. Feature Engineering **Data:** Activity

Methods: Bag Of Words, **Rapid Automatic Keywords Extraction, N-Gram** Result: Detect key words, phrases and activity format

for future modeling

Data: Transactions, activities and

Tree, Linear Regression

with

**Methods:** Linear Optimization, Optimal

activity amount, optimize the recourse

allocation of sales representatives by

prescribing who and how to interact

Result: Given certain constraints on

advisor

#### 2. Activity Content Modeling Data: Activity with features constructed in

previous step Method: Random Forest, CART, Optimal **Result:** Classified activities into 4 categories based on content with

accuracy > 70%

#### 3.Prediction

Data: Activity, Transactions and Advisors Methods: Random Forest, **CART, Optimal Tree** Result: predicted which level of the purchase amount in any specific month of a financial advisors will fall into based on previous transactions and activities with accuracy around 50%

Graph2: Pie Chart on the distribution of the four activity content we identified in training set, the model has an over 70% accuracy on the test set.

Accuracy

> 0.7

Market

(51.4%)

Fact

*Graph3: The performance of different* methods in predicting different future transactions evaluation metrics measured in accuracy

#### **Prescription Model Lifts Total Future Purchase Prediction**

4.Prescription

**Methods**: Optimal

Advisors

Data: Activity, Transactions and

Result: Based on previous

activity and transactions,

Prescription Tree

prescribed which format and

what content of activity is the

purchase in the following

month

best choice in order to increase

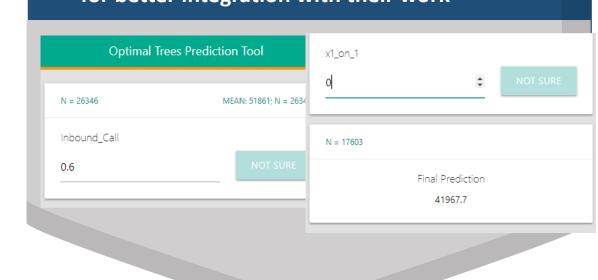
**Prescription Meeting Format Prescription Meeting Content Original Amount** 

9122.188827 8014.210884 1925.117845 2000 4000 6000 10000 Millions

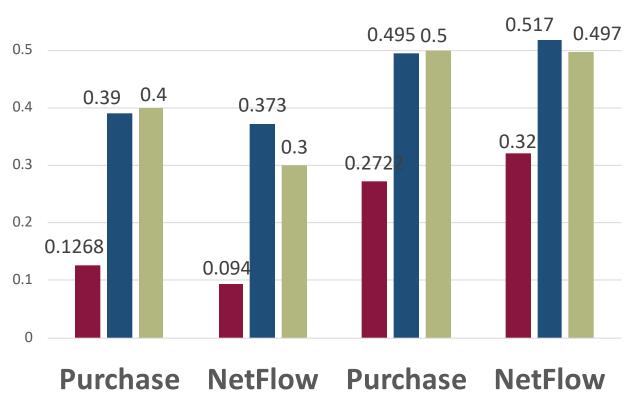
Graph4: Influence of prescribing current month meeting format and meeting content on predicted following month purchase amount based on optimal prescription tree

#### **Optimal Trees Helps To Make Decision**

- Majority of the sales team have no background in data science and rely on user friendly and easy interpretable tools to implement the models we built
- With the help of optimal trees and prescription trees, anyone interested in using data in the decision making process can utilize the models
- The tools we developed could be further adjusted based on feedbacks from sales team for better integration with their work



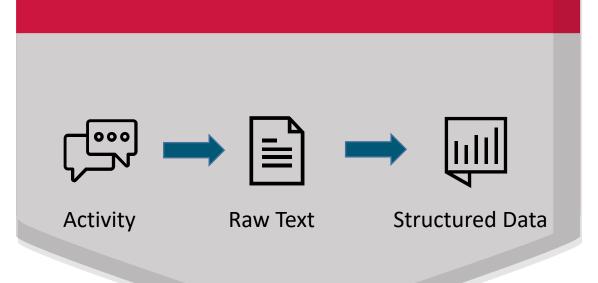
## **Prediction Model Performance**



Change Change ■ Random Forest ■ Optimal Tree

#### **Influence on Future Data Management**

- It is the first time unstructured data has played an important role in data science project in MFS.
- It provides a new perspective on how unstructured data could be stored and managed in the future
- By using the pipeline for processing unstructured data we developed, this project could be expanded to cover activities with the other 300+ partner financial advising companies



- D. Bertsimas, J. Dunn. "Optimal Classification Trees". Machine Learning, 2017
- D. Bertsimas, N. Kallus. "From Predictive to Prescriptive Analytics". POMS Applied Research Challenge, 2016.