MIT MBAn 2020
CAPSTONE PROJECT

**ADVISORS:** *Prof. Amr Farahat, Michelle Li*
**LOCATION:** *Philadelphia, PA*
**COMCAST:** *Trace Hawkins, Michael McLarnon, Kevin Benecchi, Linge Lass*

# OPTIMIZING CONTENT LIKELY PERSONALIZATION



Hamza Tazi Bouardi


Pierre-Henri Ramirez

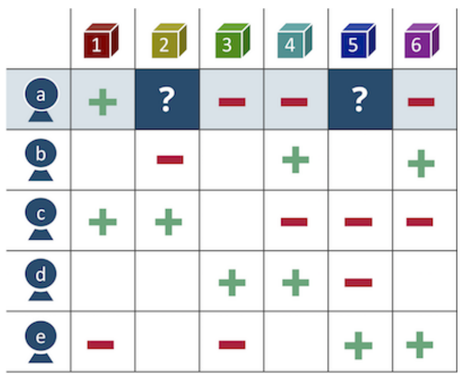
COMCAST
MIT MANAGEMENT BUSINESS ANALYTICS

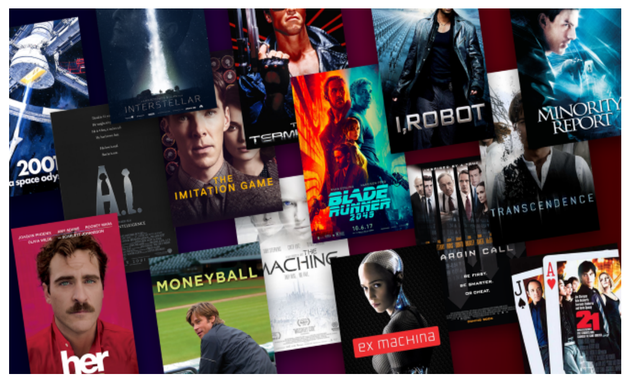

## INTRODUCTION

**Comcast** is the second-largest broadcasting and cable television company in the world by revenue and the largest home Internet service provided in the United States. Comcast's main streaming platform, **Xfinity X1**, is a gateway for multiple streaming content providers such as Netflix or Amazon Prime. To improve customer satisfaction after COVID-19 has put a significant emphasis on the streaming industry, Comcast has been trying to optimize their content recommendation routine.

**Problem Statement**: How can we recommend the right content at the right moment to the right user ? And how can we improve targeted marketing campaigns to boost ads placement and sales revenues using recommender systems ?

## DATA




We used **two data representations** 36K users and 34K items on October 2019:

- Sparse user-item matrix where a cell corresponds to the rating, here whether an item was liked by the user or not based on viewing percentage
- An Items Data Dictionary with relevant features such as genre, title, actors, synopsis, year, etc.

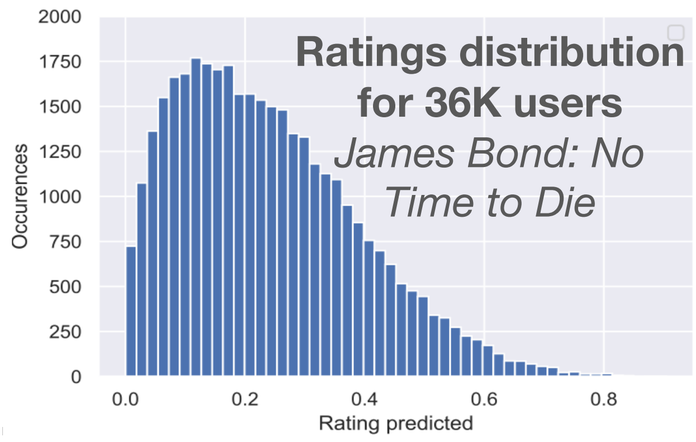
## KEY RESULTS

**Significant Improvement relative to Baseline**: According to our simulations results, we could improve the quality of good recommendations (**Precision**) by **3** points and the relevant recommendations actually retrieved (**Recall**) by **8** points.

**Well calibrated predictions and ratings distributions** using the Collaborative Filtering based methodology

**Our model could bring an uplift of at least $5M in sales revenues and $20-30K in additional revenue on Advertisements for each new show** compared to the current methodology

| Metric\Model | Our Model | Baseline |
|---|---|---|
| **Precision** | **0.64** | 0.61 |
| **Recall** | **0.61** | 0.53 |
| **Accuracy** | **0.67** | 0.63 |
| **F1-Score** | **0.62** | 0.57 |
| **Runtime (minutes)** | 0.3 | **0.1** |



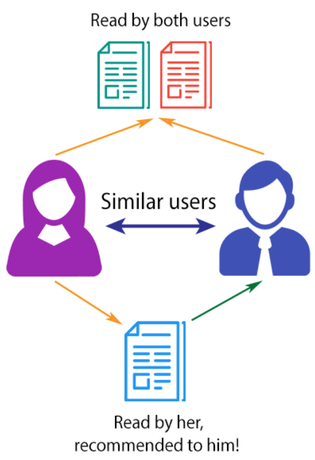

**+$5M**
Uplift in sales revenue assuming 1% better targeting

**+10%**
Better targeting on large movies compared to best current methodology

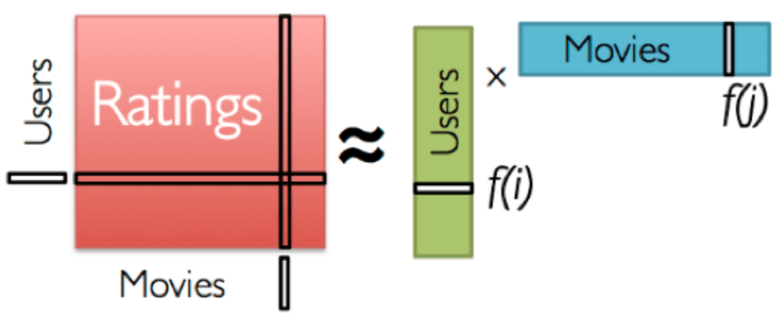
## METHODOLOGY

After a review of all available open-source libraries, we chose an Apple implemented library called *turicreate* to do the **ratings / probabilities predictions** using **Collaborative Filtering** methods as it is the most scalable and efficient one

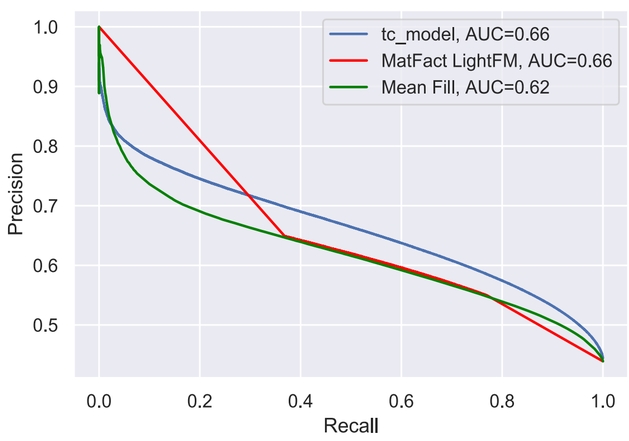
1 Paradigm we used **Collaborative Filtering**




2 Algorithm we used **Low-Rank Matrix Factorization**

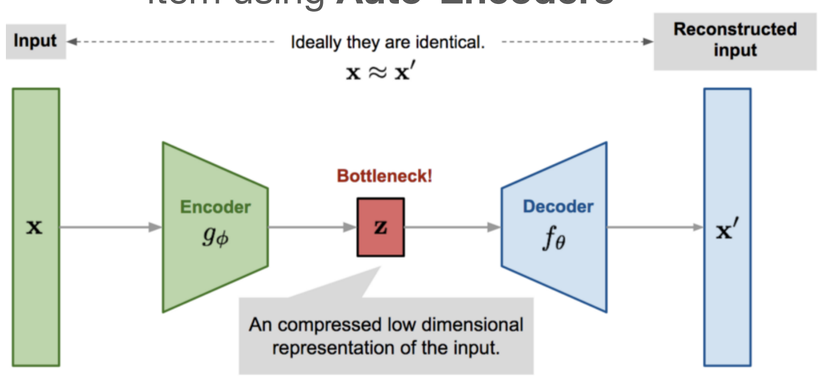



3 Results we obtained **Precision-Recall Curves for 3 models on 36K users and 34K items**




Despite a weak performance in pure prediction, **Content-Based Filtering** and **Auto-Encoders** can be used to find similar items to answer the **Cold Start Problem** in Collaborative Filtering

4 Obtain a similarity score between each item using **Auto-Encoders**




5 Find 2-5 similar items using this similarity score and use average rating as proxy for target show

*For a given user:*


## PROJECT TIMELINE

**JAN**
Matched with Comcast

**FEB**
Visa Procedure & Project Scoping

**MAR**
EDA & Literature Review

**APR**
First CF models evaluations

**MAY**
Refining computational understanding

**JUN**
Feature Engineering & Building CBF models

**JUL**
Establishing criteria and selecting models

**AUG**
Evaluating CBF and CF based recommenders