Forecasting The Shopping Cart

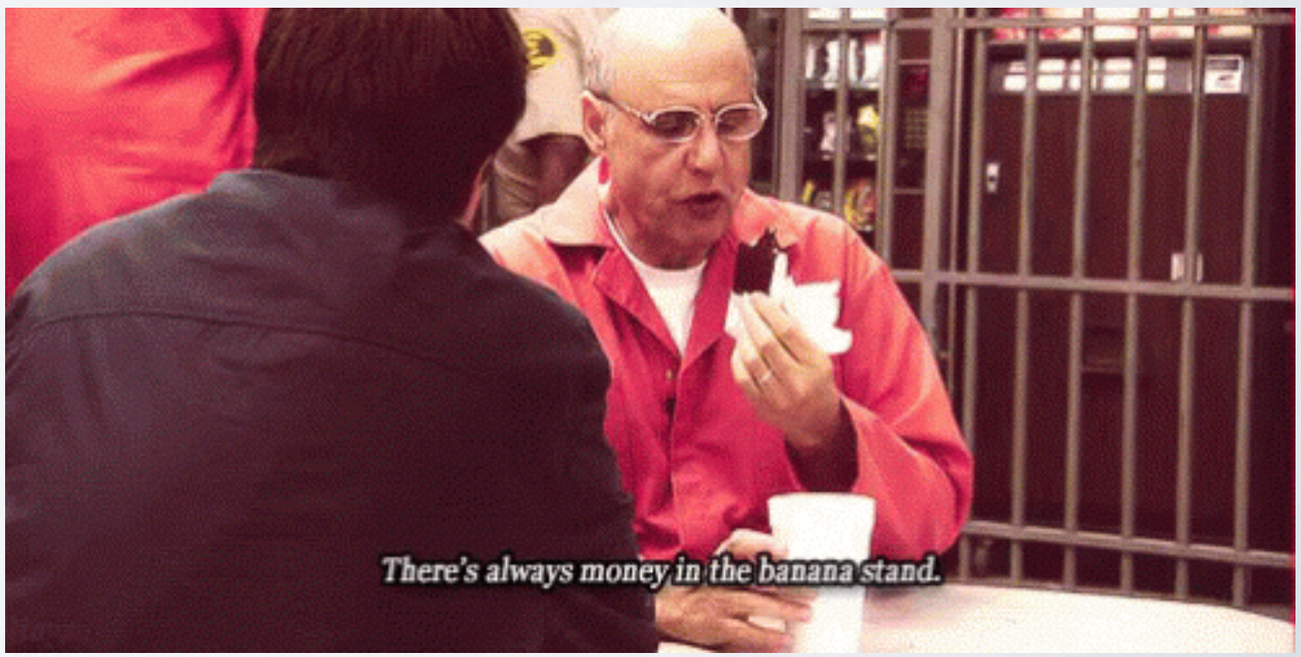
Predicting which products an instacart user will reorder

Introduction

- **Data**: 30 million+ instacart product orders

- **Goal**: create classification model to predict product reorders

- **Application**: recommendation system, inventory management



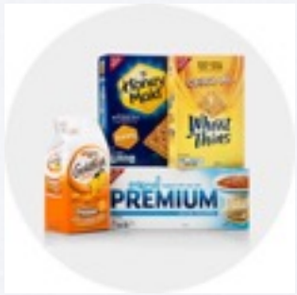
Features

- **Past Order Characteristics**: frequency, cart add priority, time/day, orders since, past reorders, recent streak

- **Order Patterns**: binary autocorrelations with lag 1, 2, 3

Lag 1 Autocorr = .33

- **Product Type**: department & aisle categories



Model Selection & Training

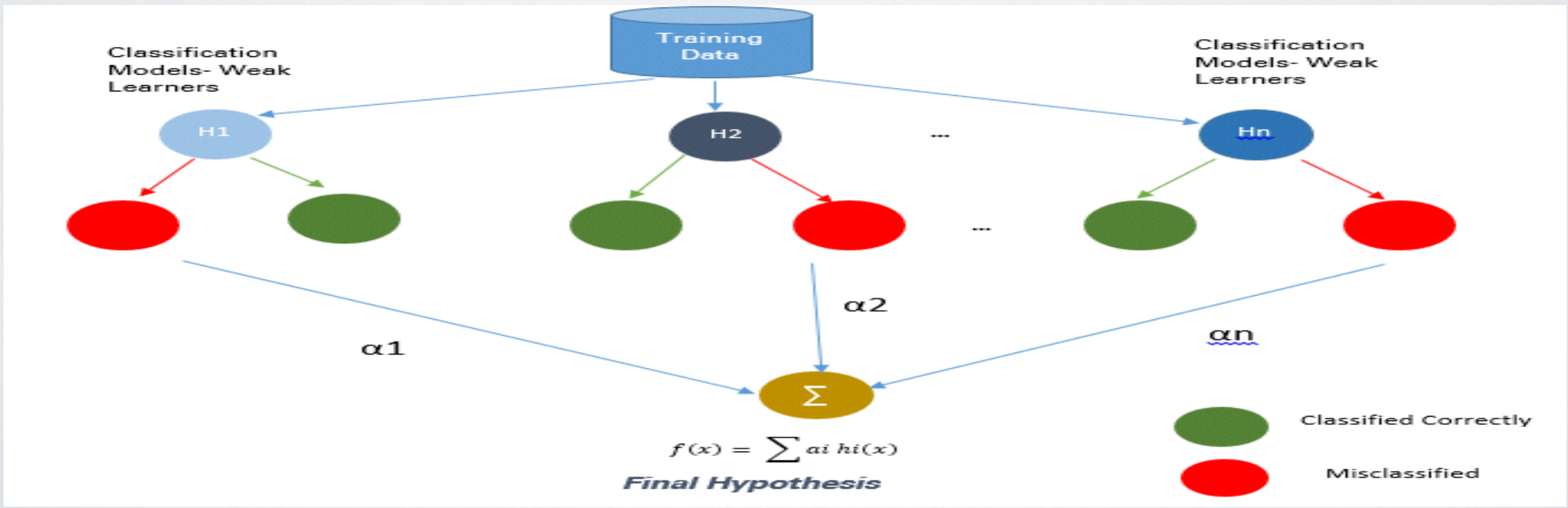
- **Local Testing**: Small user subset, quick modeling

- **F1 Optimization**: choose model with best precision/recall tradeoff)

- **Cloud Training**: Large user subset, many xgboost epochs

Model Details

- **Gradient Boosting**: Invert the random forest paradigm (reduce variance of strong learner). Reduce bias of weak learner by iteratively fitting residuals
- **Parameters**: include tree depth and count, learning rate, row/column subsampling (stochastic element)
- **Cloud Model**: ~1,700 trees of depth 6, no subsampling



Source: Analytics Vidhya

Results

- **Most important features (>.05)**: priority, orders since, time/day features, order streak

- **Also, but less important (.03 - .05)**: autocorrelations, frequency, past reorders

- **Most important category features (.002-.01)**: produce, pantry, fresh fruit & vegetables

Tuned GBM Model Metric	Test Set Score
Accuracy:	0.88
Precision:	0.39
Recall:	0.48
F1:	**0.43**

Thank You!!!

Github: JEddy92; **Kaggle**: Anatidaephobia