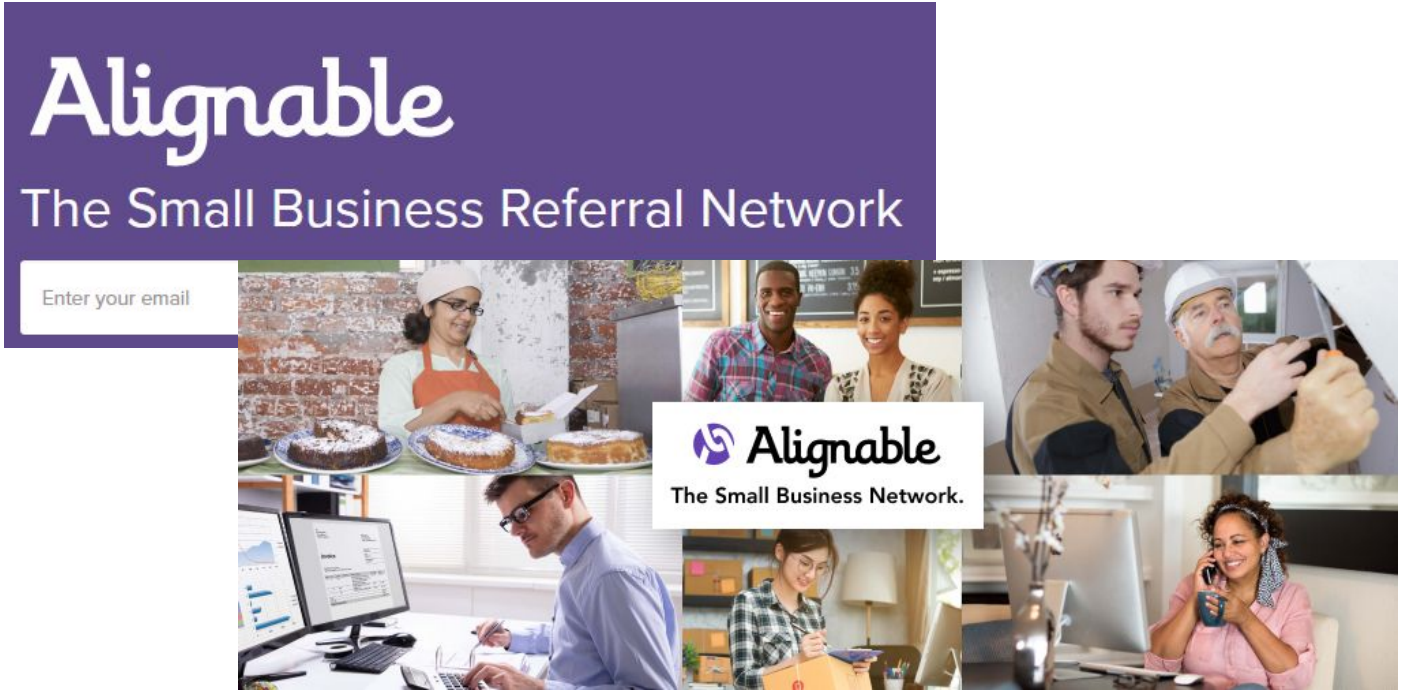
Network Recommendation
for
Small Businesses

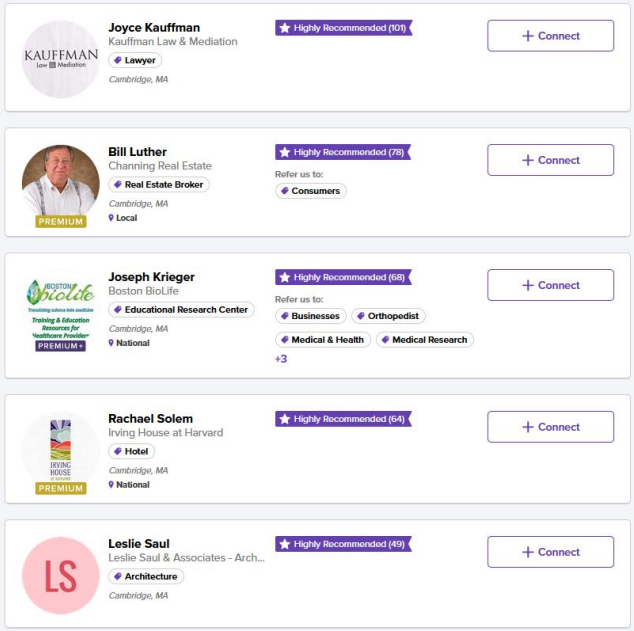
Presented by: Zhenru Wang, Catharine Wu, Royce Yap
Supervised by: Will Claybaugh, Micah Lanier
Video Walk-through: https://youtu.be/8lYnGJo8iiY

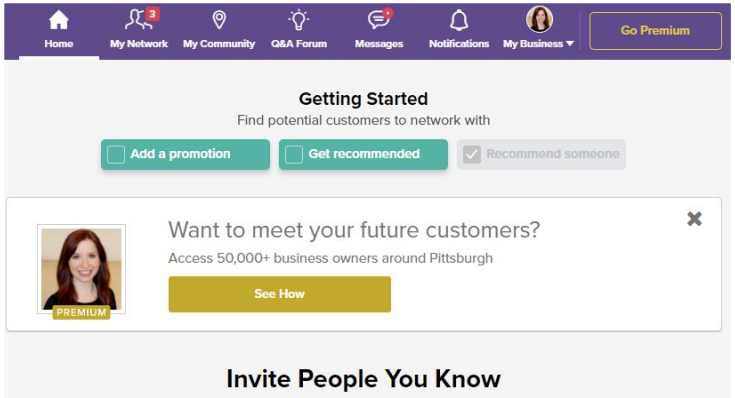
IACS Harvard School of Engineering and Applied Sciences
IACS Institute for Applied Computational Science
Alignable

Introduction



Problem Statement and Motivation





Who should connect with whom?

Outline

Data

- Unification
- Missing Values
- Transform Variables
- Train - Test Split
- Model Preparation

Model

- Baseline Model
- Predictive Models
- Collaborative Filtering

Result

- Evaluation
- Interpretation

Data sources

Business-level attributes (nodes)	Relationships between businesses (edges)
Businesses	Recommendations
Industries	Connections
Roles	Business Requests
Communities	
Users	
Discussion Posts	

Data points for Alignable businesses

Community

Longitude

State

Latitude

Industry / Role

Customer Type
(B2B / B2C)

Date of joining

Recommendations

Added logo

Customer Location
(Local/Regional/National)

Existing connections

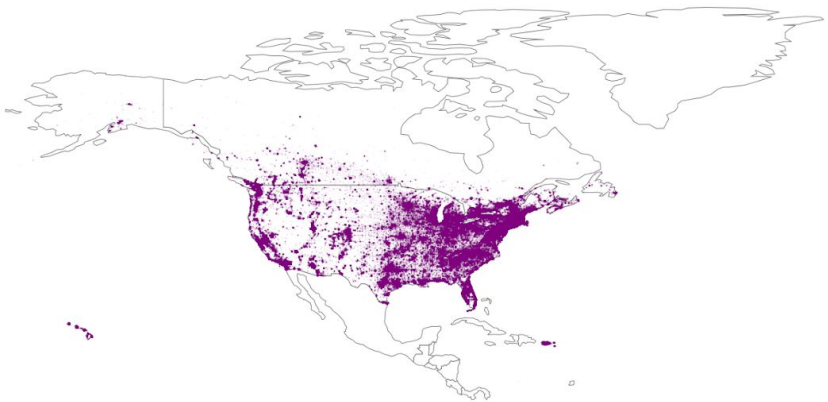
Most recent sign-in

Business Requests

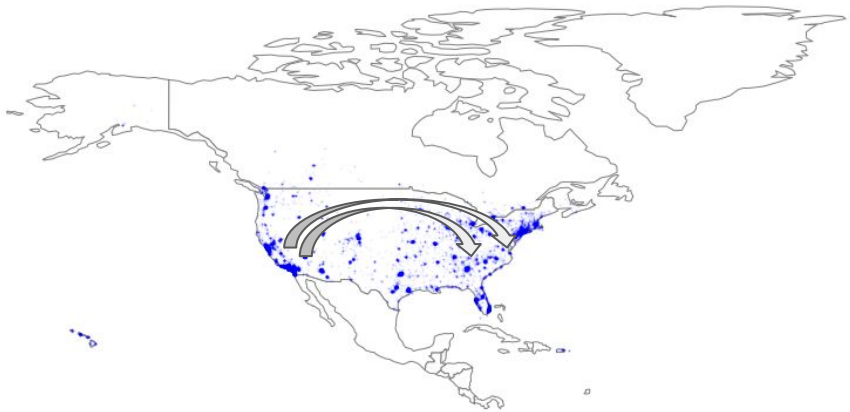
Added banner

Discussion Posts

Data Exploration - Geographical Distribution

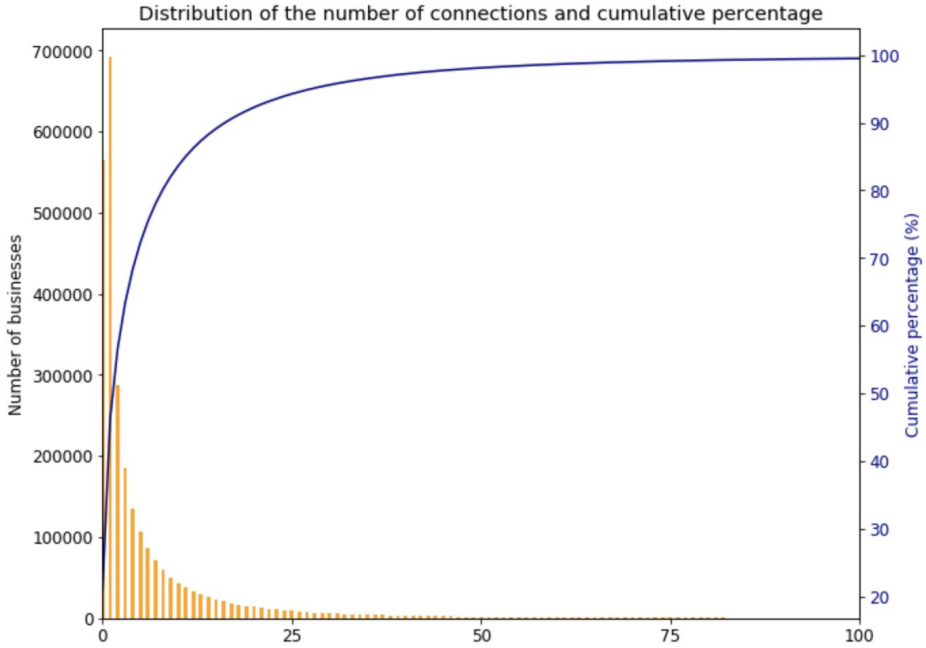


Users across the Platform

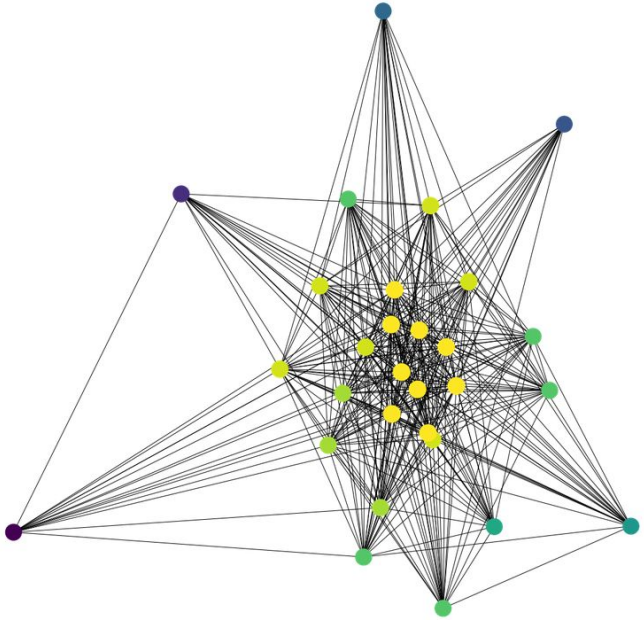


<u>Example:</u> Connections to Los Angeles Users

Data Exploration - Connections

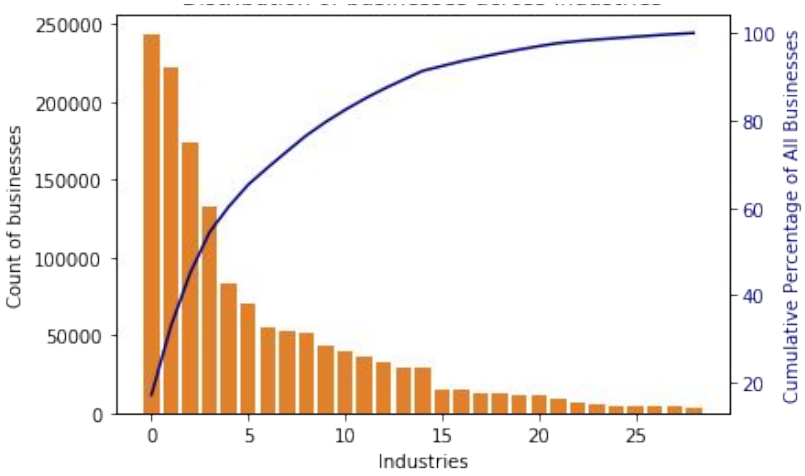


Number of connections per business

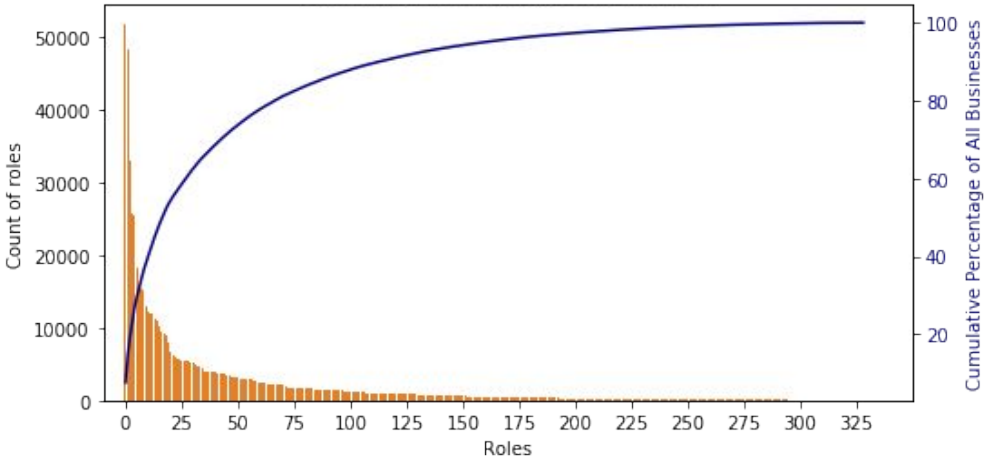


Connections across industries

Data Exploration - Industry / Role



Distribution across 29 industries



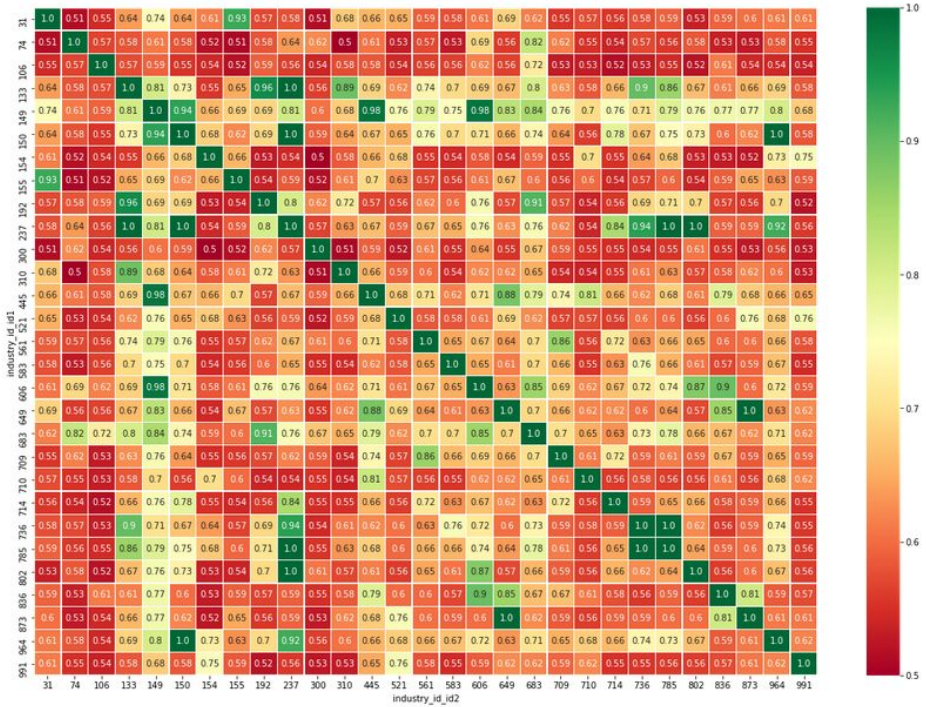
Distribution across 332 roles (sub-categories)

Data Wrangling / Preprocessing

- Unification
- Missing Values
- Transforming Variables

E.g. Industry/Role Affinity Scores

Industry Affinity Score



Train-test split

Connections for a given business

connected businesses	A	B	C	D	E	F	G	H	I	J
created_at (ascending)	0.12	0.2	0.27	0.29	0.32	0.48	0.57	0.68	0.79	0.98

Train

Test

Modeling Approach

Baseline

Randomly predict whether connected

Predictive Models

Linear & Logistic Regression, Decision Tree, Random Forest, AdaBoost, Neural Network

Collaborative Filtering

Singular Value Decomposition (SVD), Alternating Least Squares (ALS)

Modeling Approach

Baseline

Randomly predict whether connected

Predictive Models

Linear & Logistic Regression, Decision Tree, Random Forest, AdaBoost, Neural Network

Collaborative Filtering

Singular Value Decomposition (SVD), Alternating Least Squares (ALS)

Predictive Models

Input

Output

source_ business_id	target_ business_id	source_business _profile	target_business _profile	distance	# of mutual connections	connected
2422	7257010	...	...	14.072	2	1
2422	8550114	...	...	1.8673	8	0

Business Profile
type, scale, location, industry, role, profile completeness, time joined, latest activity

Modeling Approach

Baseline

Randomly predict whether connected

Predictive Models

Linear & Logistic Regression, Decision Tree, Random Forest, AdaBoost, Neural Network

Collaborative Filtering

Singular Value Decomposition (SVD), Alternating Least Squares (ALS)

Collaborative Filtering

Target Businesses

Source Businesses

	Target 1	Target 2	...	Target 24520
ID 1	0	1	0	0
ID 2	0	0	0	0
...	...	...	...	...
ID 3791	0	0	0	0

PROBLEM 1

Data Sparsity

- Community Clustering
- Data Balancing

SOLUTION

PROBLEM 2

Expensive Computation

Business Clustering

SOLUTION

Model Comparison

- Metric: Normalized Discounted Cumulative Gain (**NDCG**)

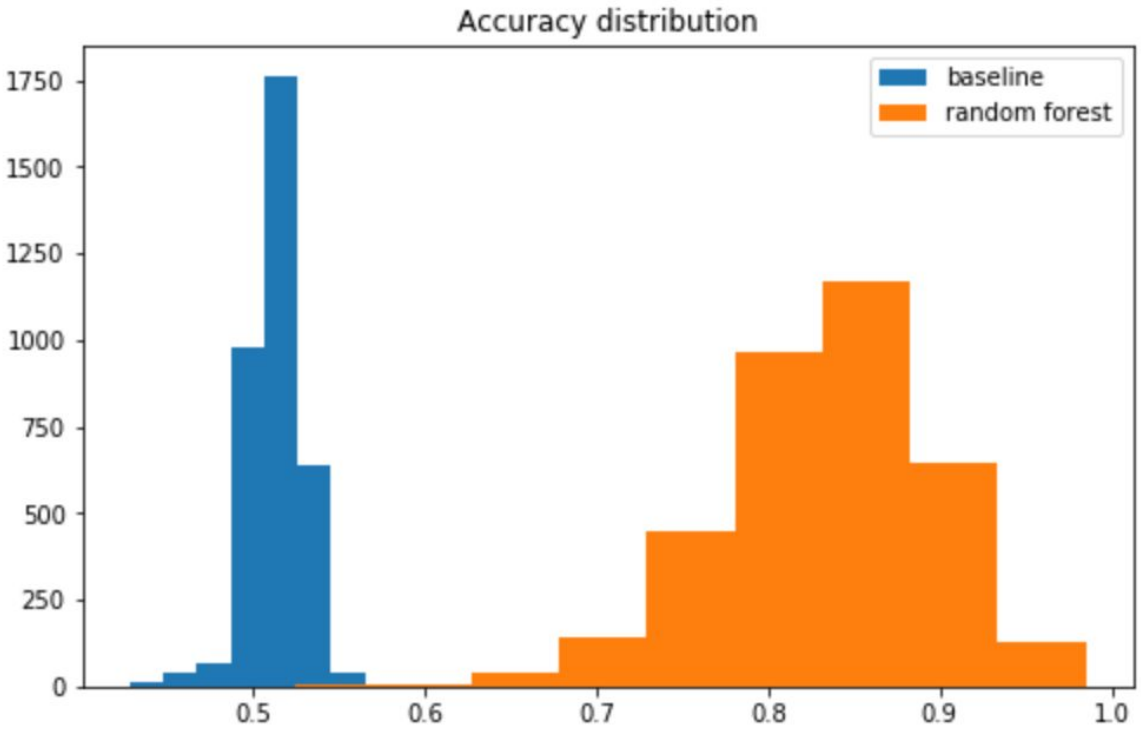
	Baseline	Random Forest	SVD
LA	0.0002	0.0031	0.0030
NYC	0.0003	0.0086	0.0073
Boston	0.0023	0.0322	0.0270

83.44%

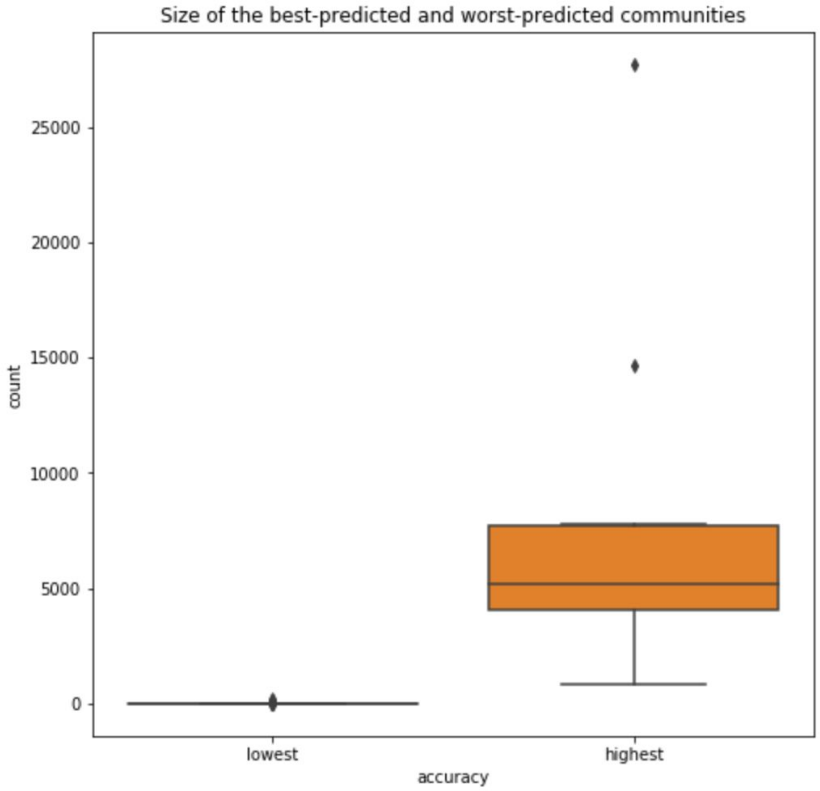
Averaged Overall Accuracy among all communities

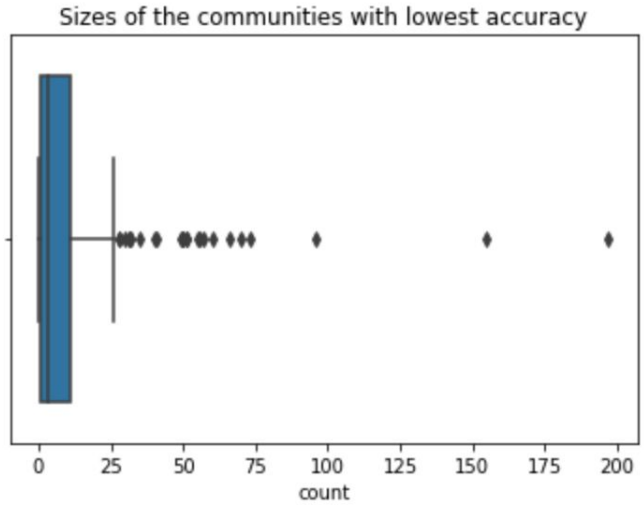
by Random Forest

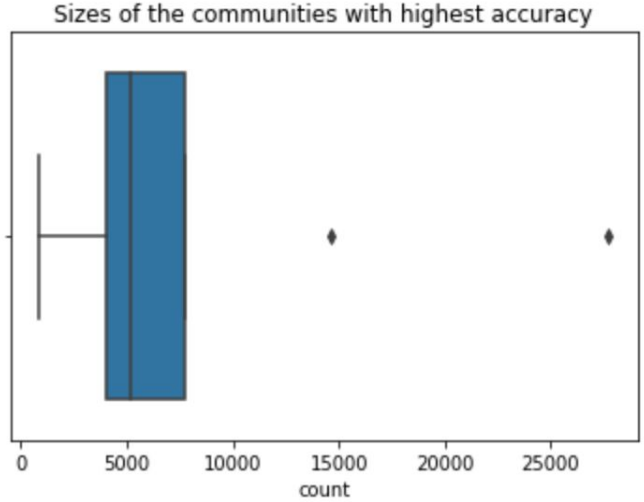
Evaluation


Accuracy distribution
baseline
random forest
1750
1500
1250
1000
750
500
250
0
0.5
0.6
0.7
0.8
0.9
1.0

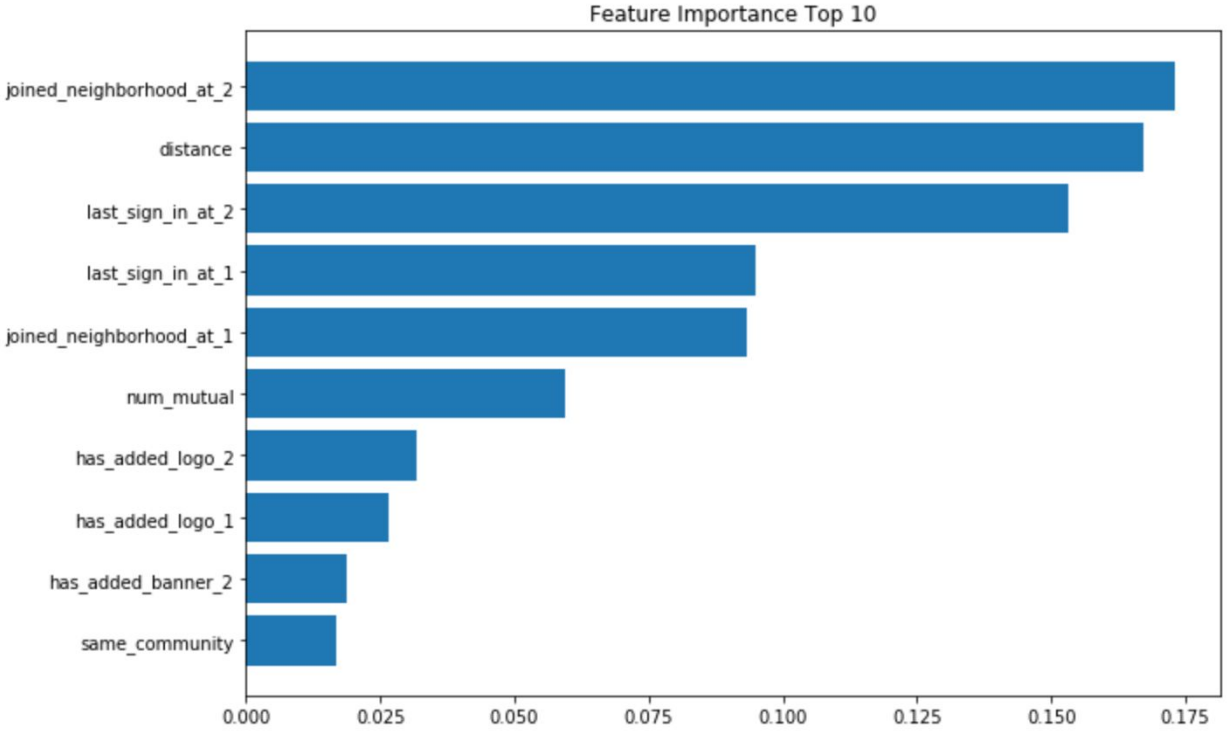
Model performs better on larger communities







Feature Importance



Deliverables

A **csv** file containing all Alignable's businesses and 5 ranked recommendations for each

A **python** file to produce the csv

Modeling approaches and analysis

Literature and Sources

Herlocker, Jonathan L., et al. "Evaluating collaborative filtering recommender systems." *ACM Transactions on Information Systems (TOIS)* 22.1 (2004): 5-53.

Wu, Lili, et al. "The Browsemaps: Collaborative Filtering at LinkedIn." *RSWeb@ RecSys*. 2014.

Liao, K. (2018, November 19). Prototyping a Recommender System Step by Step Part 2: Alternating Least Square (ALS) Matrix Factorization in Collaborative Filtering https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-2-alternating-least-square-als-matrix-4a76c58714a1

Review

Data	Model	Result
• Unification • Missing Values • Transform Variables • Train - Test Split • Model Preparation	• Baseline Model • Predictive Models • Collaborative Filtering	• Evaluation • Interpretation