Finding Your Dream Home: REX House Recommendations

Cecilia Yang, Nikhil Vanderklaauw, Weiru Chen, Selina Wu

Agenda

1. **Introduction and Preparation**
   a. Problem Statement and Motivation
   b. Datasets

2. **Modeling Approaches**

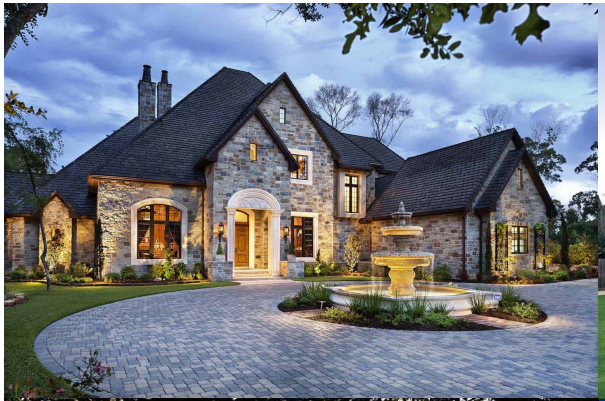
3. **Evaluation, Interpretation and User study**:
   What works? Why choose it?

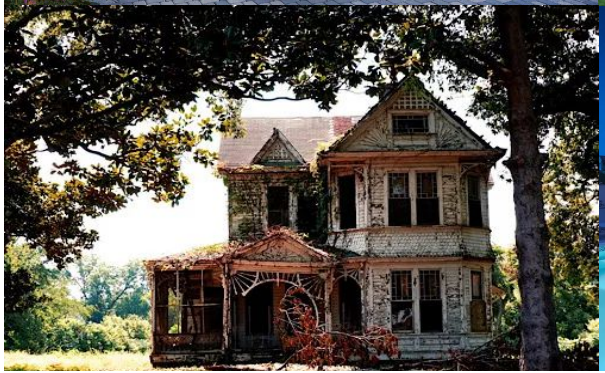
4. **Conclusions & Future Work**

Problem Statement and Motivation

How do you find your dream house?

How can we go through all of REX's listings to **find the perfect home for each user**?





Problem Statement

REX wants to improve users' experience in buying homes.

Our Goal: Develop a **model** that serves **open minded house-hunters** with **personalized matches** for discovering their perfect home

Metric for Success: receiving positive engagement with house listings suggested by our recommender.

How many of the listings we recommend end up being visited by users?

Project Trajectory

Week 2 - Week 5

Data Exploration

- Understand the features in selecting an ideal house
- **Explore** the **datasets** provided
- **Identify** the data that would be helpful to us
- **Assess** quality/missingness

Week 6 - Week 8

Baseline Model

- Create **collaborative filtering** and **content filtering models features** as baselines
- **Request** for **more data** from REX

Week 9 - Week 12

Advanced models

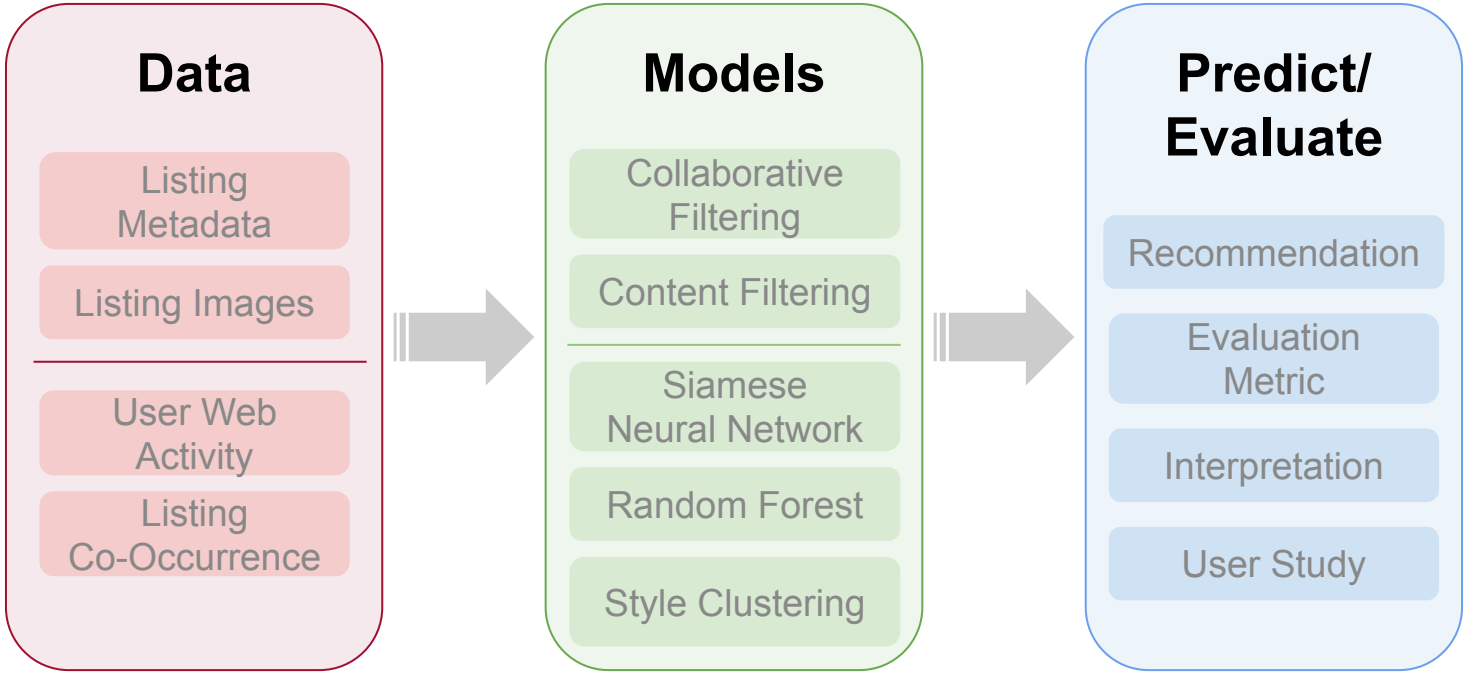
- Incorporate **co-occurrence** information to model **user behavior**
- Develop **Siamese Neural Network Random Forest models**
- **Fine-tune** the models and add image features by **image modeling**

Week 13 - Week 14

Evaluation and user study

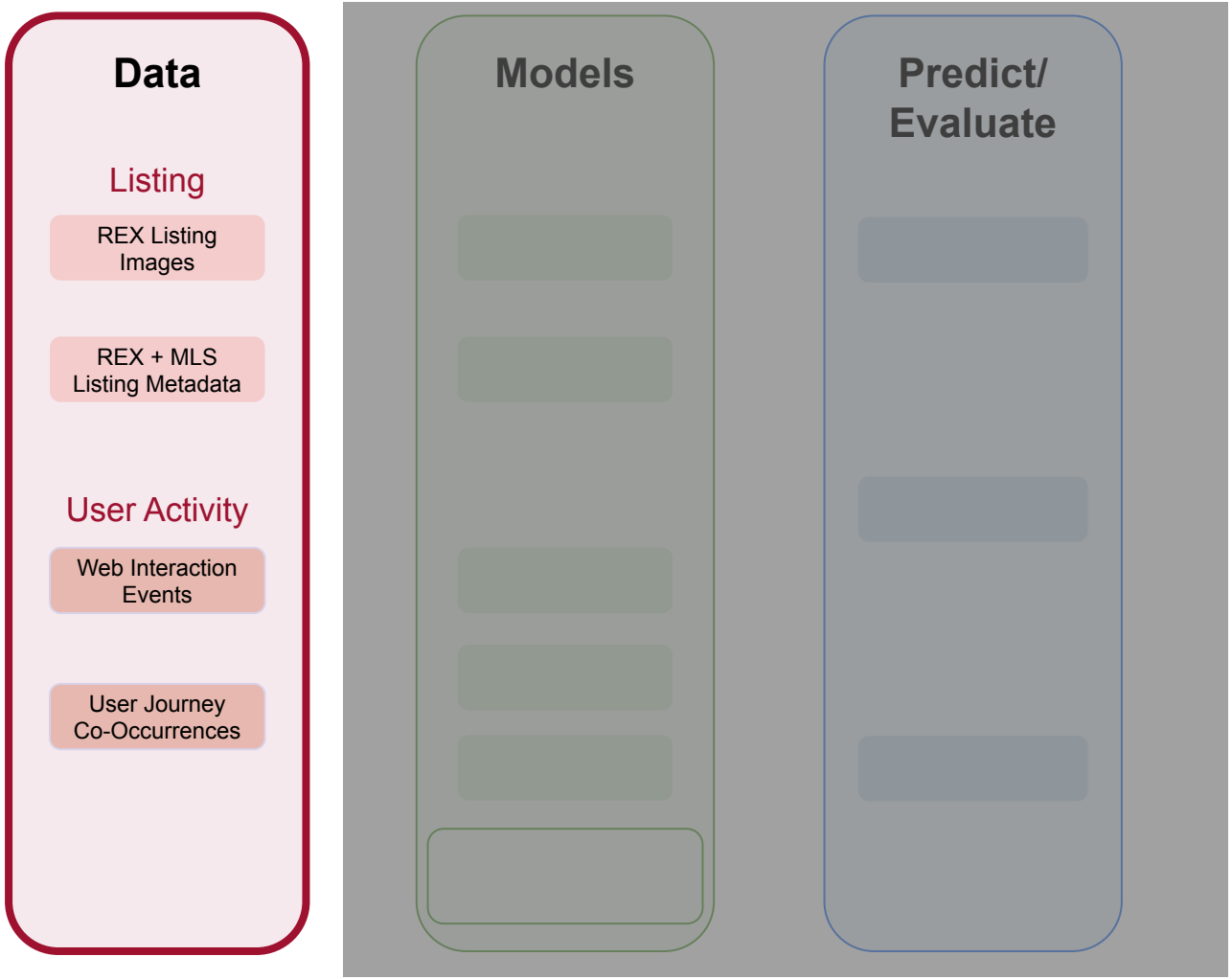
- **Evaluate** and **compare** the results from different models
- Design a **user stud**y to test our assumptions

Project Workflow



Description of Data

Project Workflow



Listing Data: Numeric

Dataset that included **descriptive features** about **LA house listings**

Main Features

6145 Longridge Ave
Van Nuys, CA 91401

Location

Price

$829,000
$466 / sf

3	2	1,778	7,381
bed	bath	sf	lot sf

MAKE AN OFFER

Additional Features

Details

House Type Single-Family	**Stories** 1	**Year Built** 1954	**A/C** Central
Heating Forced Air	**Pool** Yes	**Fireplace** 1	**Parking Spaces** 3 spaces
Parking Type Carport, Driveway	**Basement Sqft** —		

HOA FEE $ —

Sources

- Data provided by listers to REX
- MLS broker database
- Zipcode-level market surveys to encode location information

Listing Data: Images

We extracted features from images of 9.5k listings as supplementary features








Web Interaction Data

Features

Listings viewed by user

Timestamp

Items Clicked

Image carousel, Favorite/Heart, Payment calculator ...

Learn about specific interests of each user

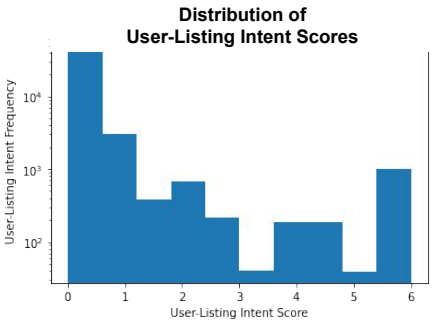
- What kinds of houses were they browsing?

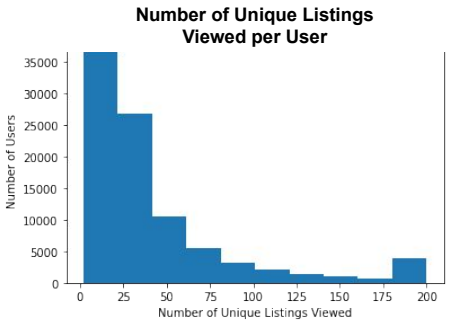
Create user-listing "score" from click data

- **High intent:** Heart, Request Info...
- **Low intent:** scrolls/clicks...

92k users viewed LA listings

533/5.5k listings had interactions (9%)





Co-Occurrences

Goal: Learn from browsing behavior to supervise recommendations

- Assume that co-clicked homes within a window during a session are relatively similar
- Incorporate collaborative information by aggregating across users

We learned: REX browsing is *exploratory*

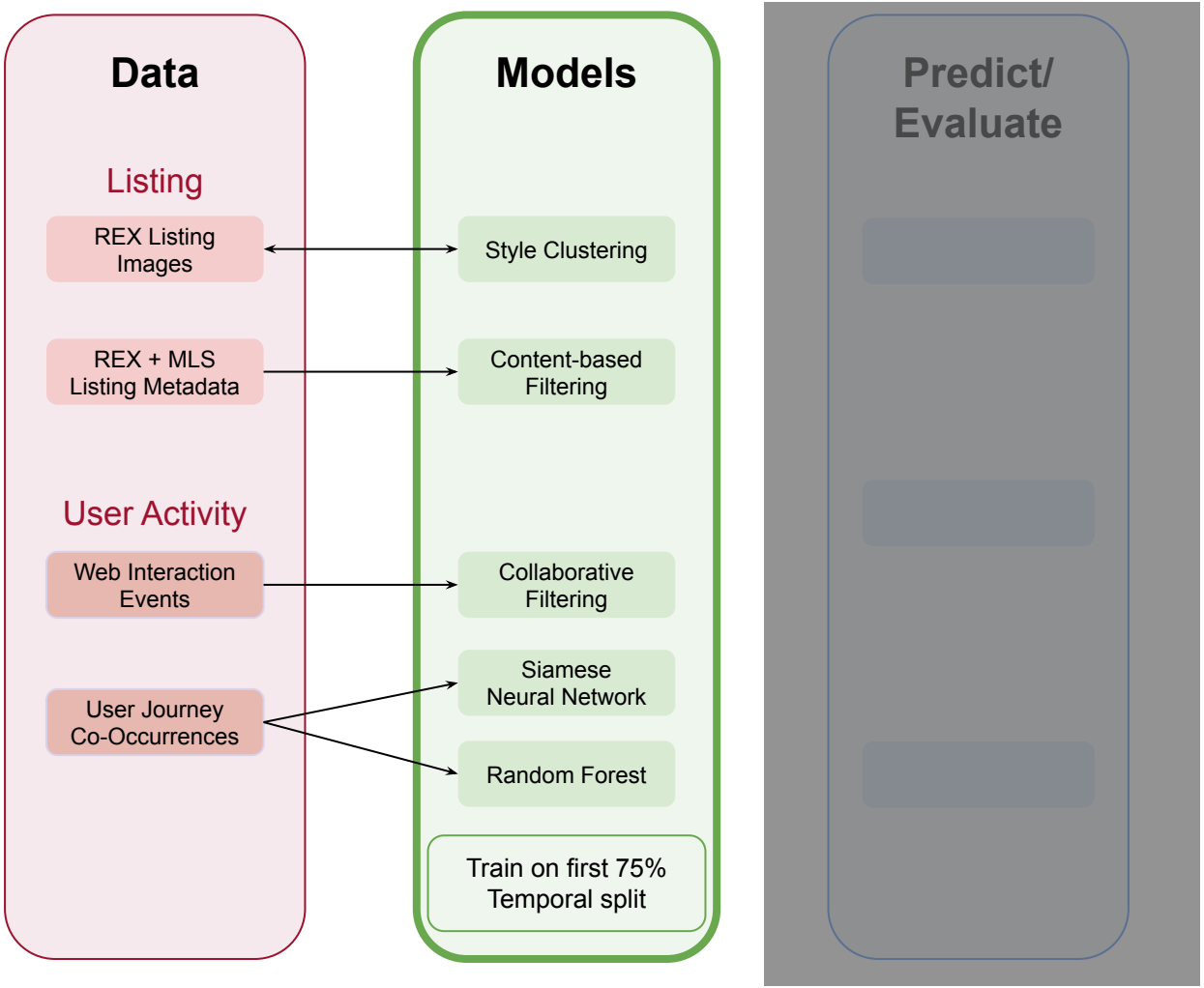
- Users may be interested in a *variety* of houses
- We want to be able to *capture diversity* of viewings in our recos

Total: 504 **co-occurring** listings

ANCHOR	***30x***	***13x***	***8x***	***8x***
2 bd | 1 ba	2 bd | 1 ba	3 bd | 2 ba	2 bd | 3 ba	3 bd | 3 ba
812 sqft	972 sqft	1,126 sqft	1,300 sqft	2,500 sqft
$600k	$450k	$635k	$470k	$3.55m

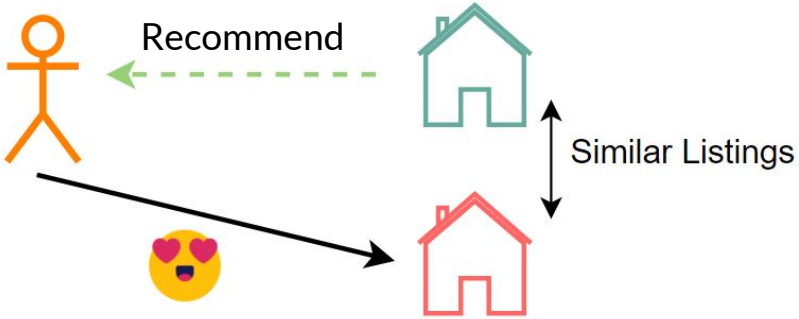
Modeling

Project Workflow



Baseline: Traditional Recommendation Models

Content-based Filtering



Recommend other listings *similar* to what the user likes

Data: Listing features
Method: Cosine-Similarity

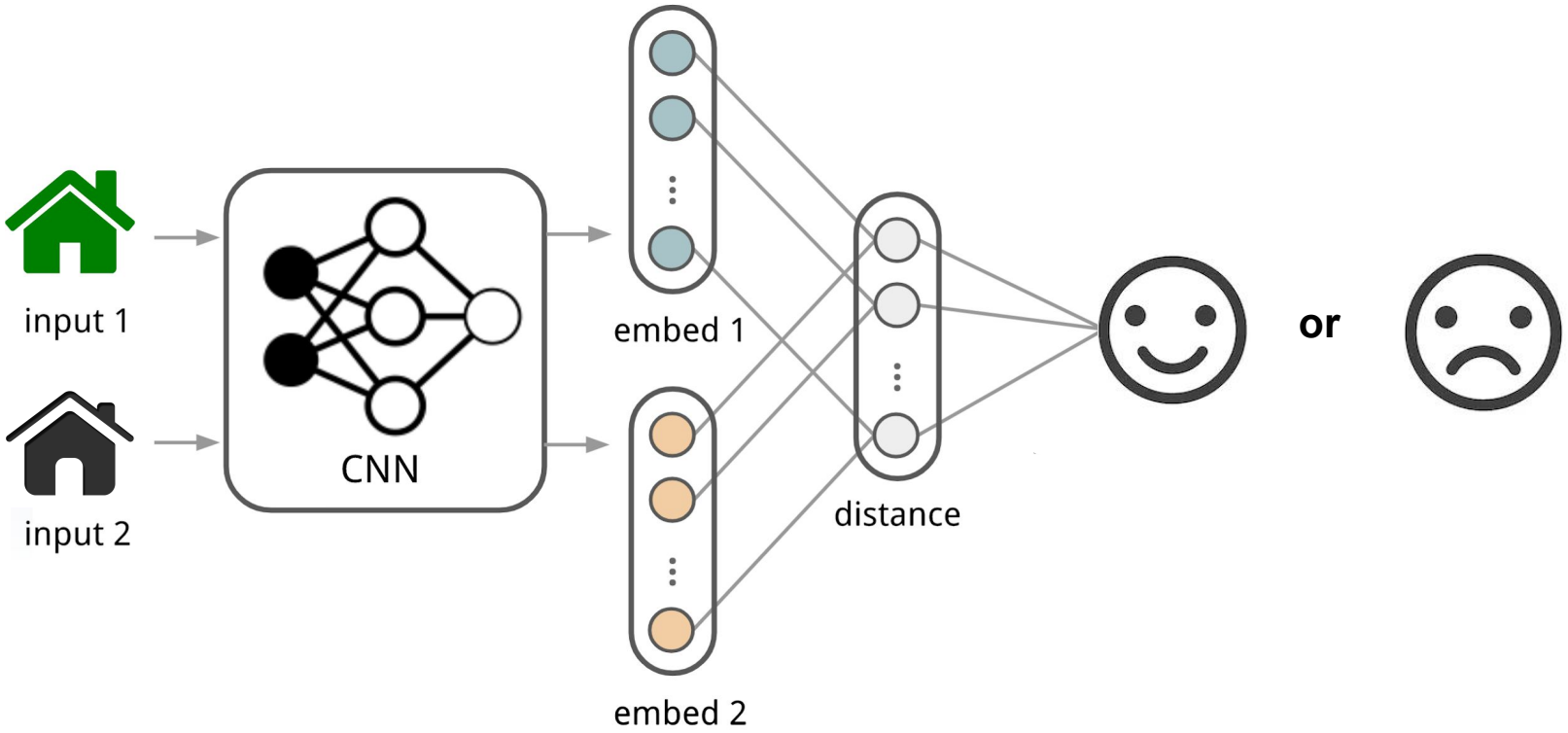
Collaborative Filtering

Use similarities *between users and listings* simultaneously to provide recommendations

Data: User-Listing view "scores"
Method: SVD, NMF

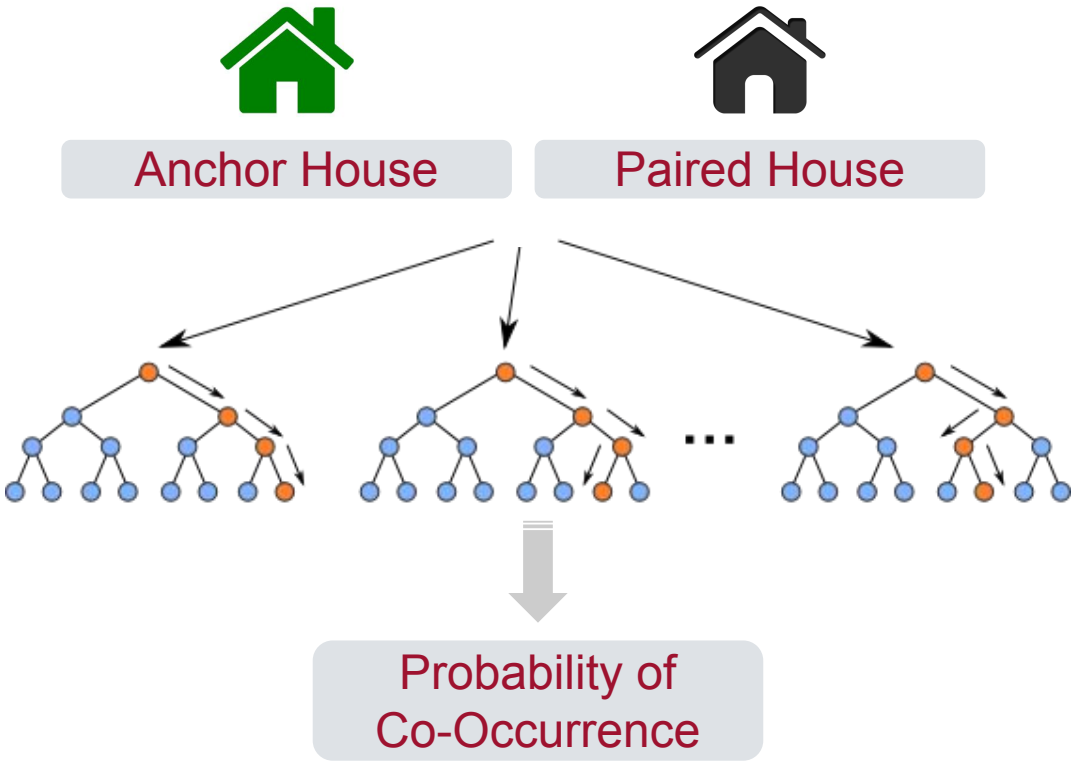
Models: **Siamese Network:**

Content-based approaches informed by **collaborative** information



Models: **Random Forest**

Trained on **co-occurrences**, predict on User Vectors



Extract Features from the Images

Indoor

Outdoor

Living Room

Dining Room

Bedroom

Kitchen

Bathroom

Indoor/Outdoor Classification

Indoor Room Type Classification

Places365

REX API

Style Clusters

For each listing

The output from our model

Outdoor	Style Cluster #	Bedroom	
	0		3
	2		3
	0		

Style Clusters

For each listing

Outdoor

The output from our model

Style Cluster #

0

2

0

Majority Vote

0

Bedroom

3

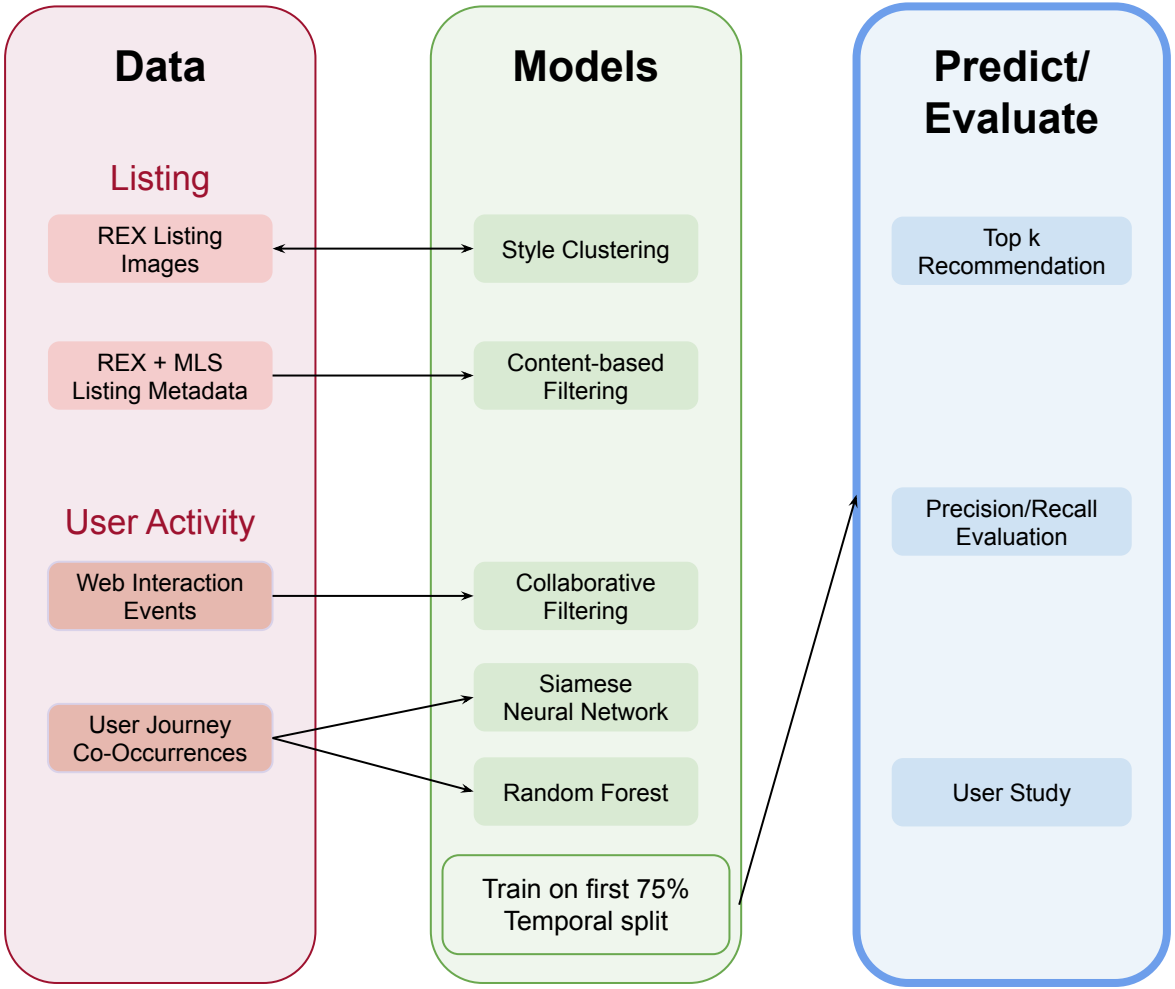
3

3

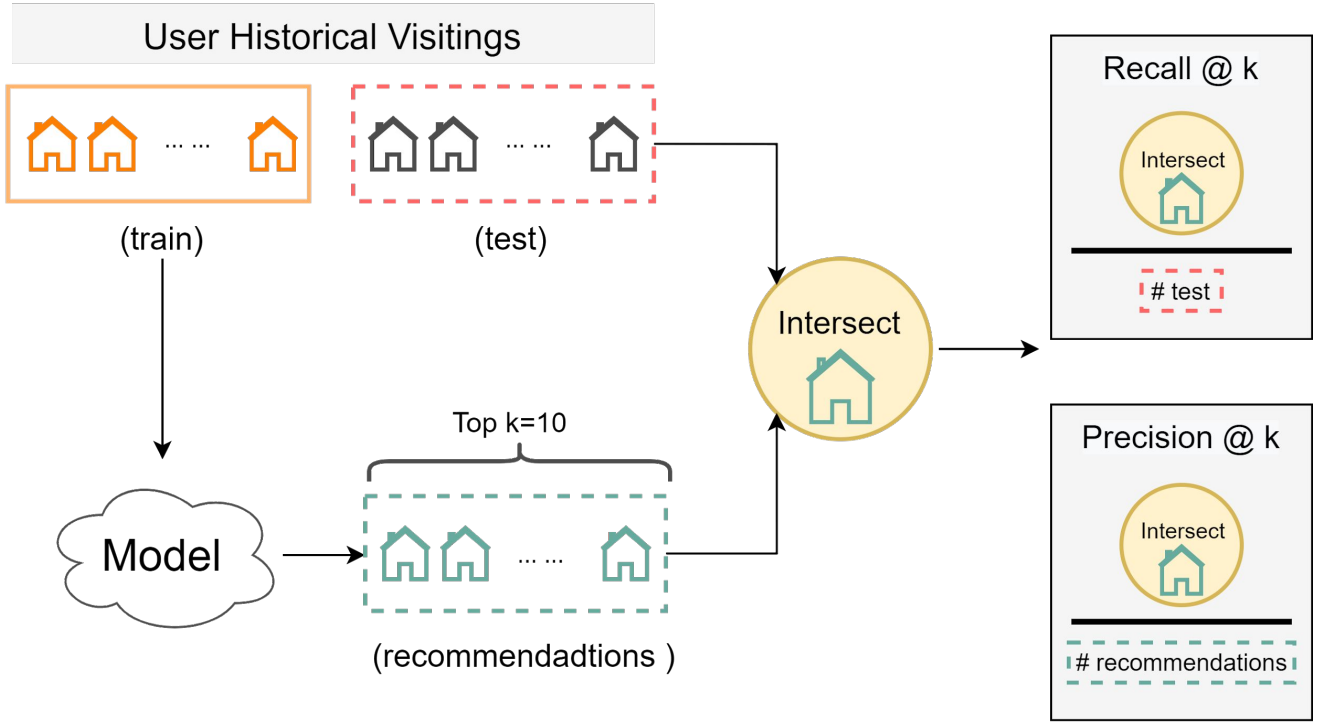
To Listing Metadata as additional features

Evaluation and Interpretation

Project Workflow



Evaluation Metrics: **precision@k**, **recall@k**



Models Summary

Random Forest and Siamese model performed **2x better** than content filter

Model	Precision @ 10 (%)	Recall @ 10 (%)
Content Filtering	0.458	2.77
Collaborative Filtering	0.54	2.78
Siamese NN	**1.04**	**5.94**
Random Forest	**4.2**	**25**

2 X better

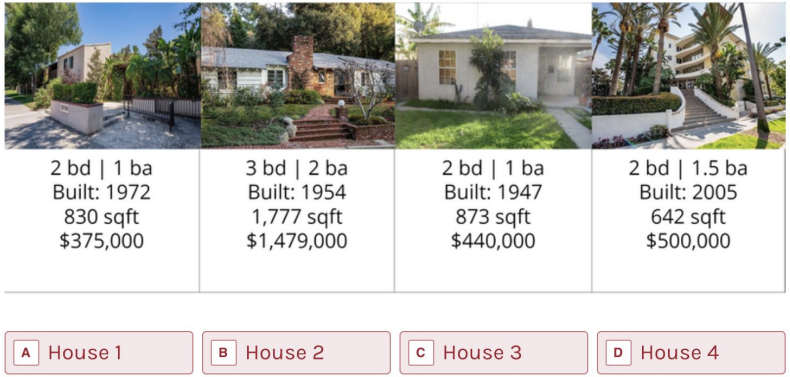
User Study

We **designed a simple user study** to see which model produced recommendations that *real users* would be interested in

1. Pick 3 houses of interest

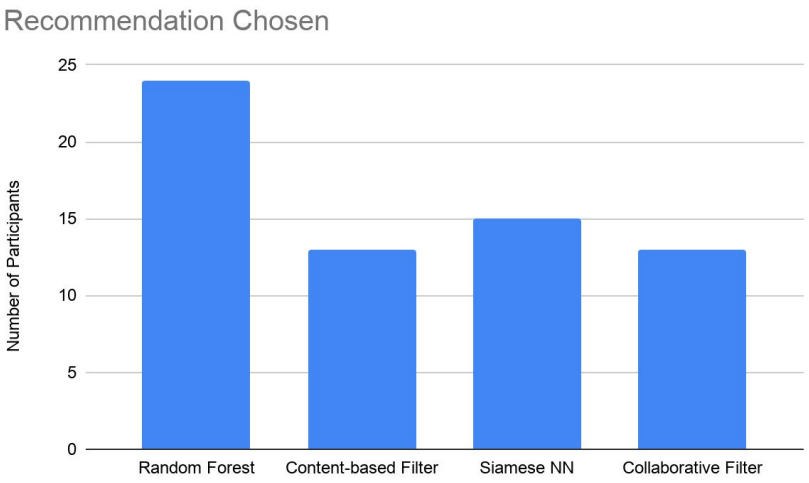
2. Select an additional listing of interest

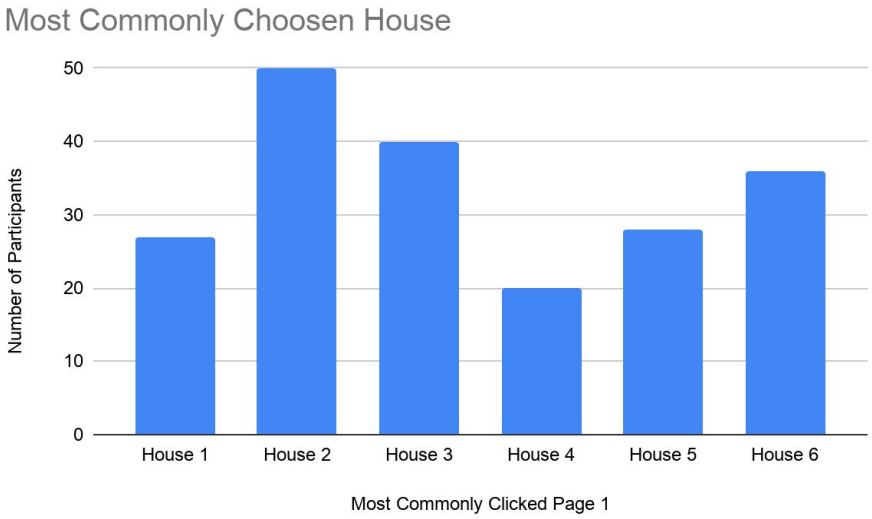
   Which model produced their preferred listing?



User Study Results

Sample Size: 65 (students + REX employees)



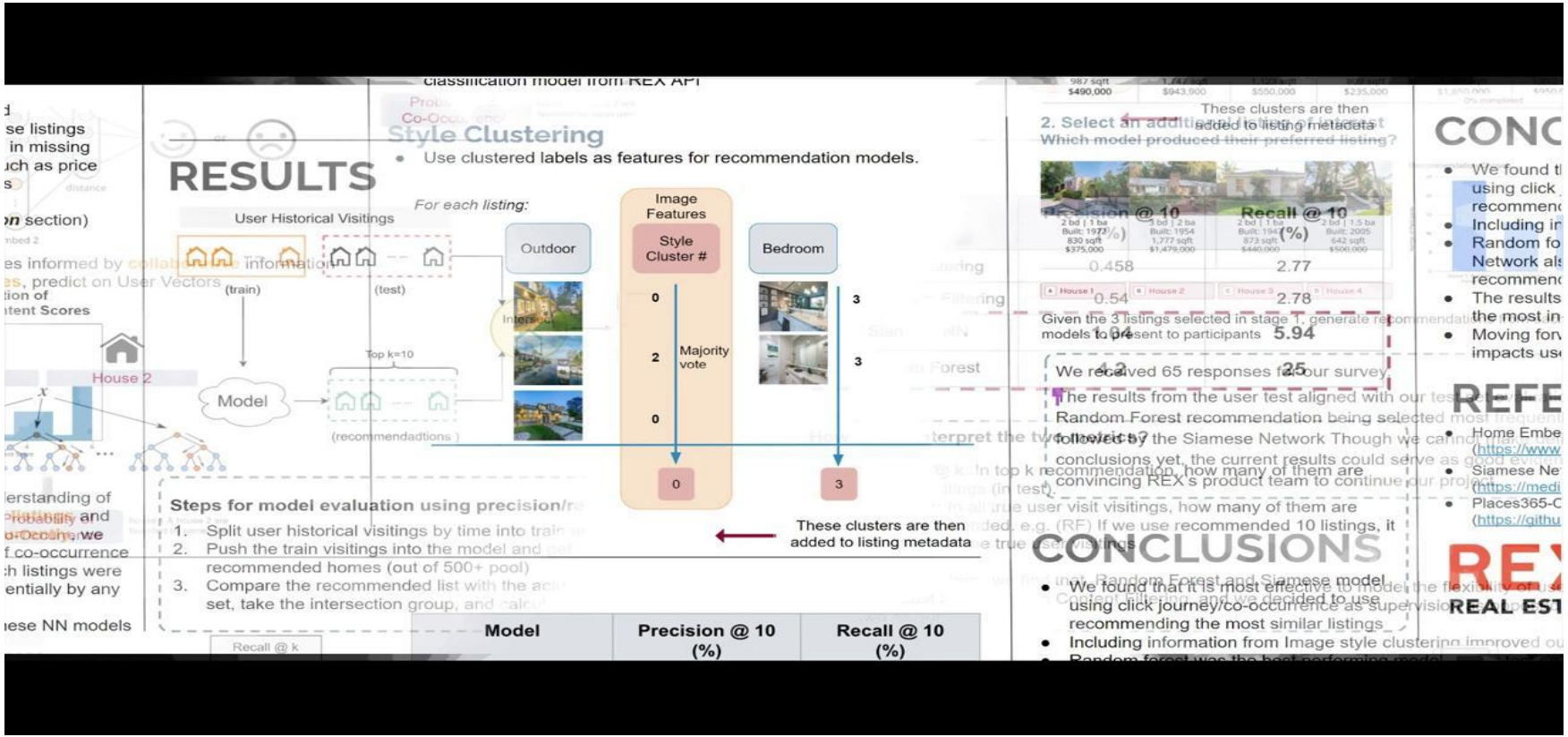


Conclusion & Future Work

- We were able to **generate successful recommendations** that seem to **capture nuances of house hunting behavior**
  - Random forest is the best performing model on test set
  - Random forest also produced the recommendations that participants in our user study were most interested in

- Future Work:
  - We are continuing to collect data for the user study
  - Results can be used to convince REX product managers to continue the project - measure increase in user engagement

Poster Video


RESULTS
Style Clustering
Use clustered labels as features for recommendation models.
For each listing:
User Historical Visitings
(train)
(test)
Model
Top k=10
(recommendations)
Outdoor
Bedroom
Image Features
Style Cluster #
Majority vote
These clusters are then added to listing metadata
Steps for model evaluation using precision/recall
1. Split user historical visitings by time into train
2. Push the train visitings into the model and get recommended homes (out of 500+ pool)
3. Compare the recommended list with the actual set, take the intersection group, and calculate
Model
Precision @ 10 (%)
Recall @ 10 (%)
Precision @ 10
Recall @ 10
0.458
2.77
0.54
2.78
5.94
2.25
We received 65 responses from our survey
The results from the user test aligned with our tests
Random Forest recommendation being selected most frequently followed by the Siamese Network
CONCLUSIONS
Including information from Image style clustering improved

[Link]

Literature & Sources

- Home Embeddings for Similar Home Recommendations
- Siamese Network Keras for Image and Text Similarity
- Places365-CNNs