# Finding Your Dream Home: REX House Recommendations

Weiru Chen, Nikhil Vanderklaauw, Zhenwei (Selina) Wu, Wanxi (Cecilia) Yang

GOAL

The process of finding the perfect is often long and complicated. Rex wants to improve users' experiences in buying homes, and one important step is to provide users an easy way to find their idea home.Therefore, we aim to **develop a model** that serves **open minded house-hunters** with **personalized matches** for discovering their perfect home.

Project Workflow
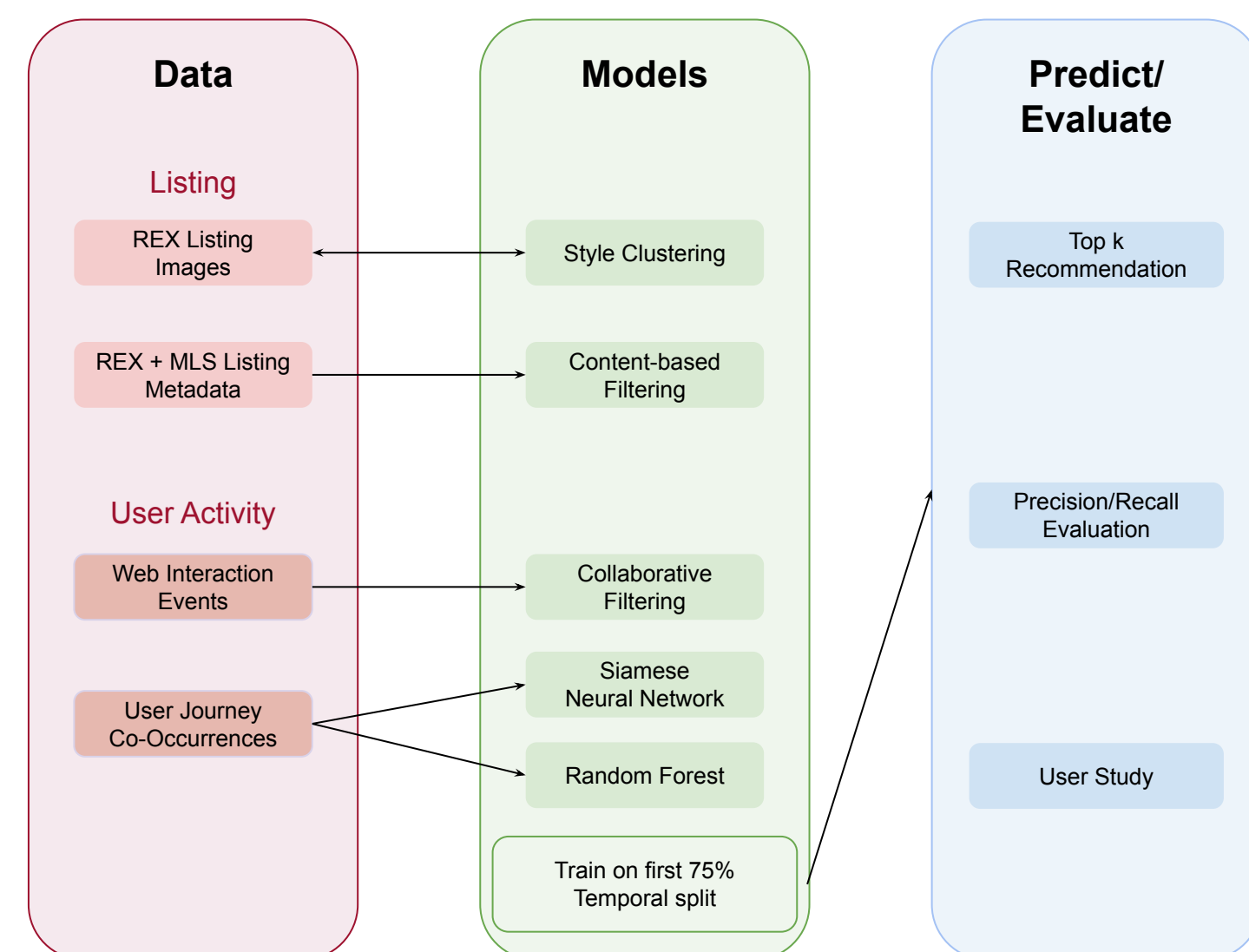
DATA

Listing Data

- REX Listing metadata included descriptive features about house listings
- MLS Metadata was used to fill in missing values in important features such as price
- We used 9.5k listing images as supplementary data (see *Image Feature Extraction* section)

Web Interaction Data

- Used web-log data from REX's website to gain an understanding of **specific users' house interests**
- Ranked interactions to **generate scores of intent** a user has for each listing they engaged with, and implemented the data in our collaborative filtering model

**Distribution of User-Listing Intent Scores**
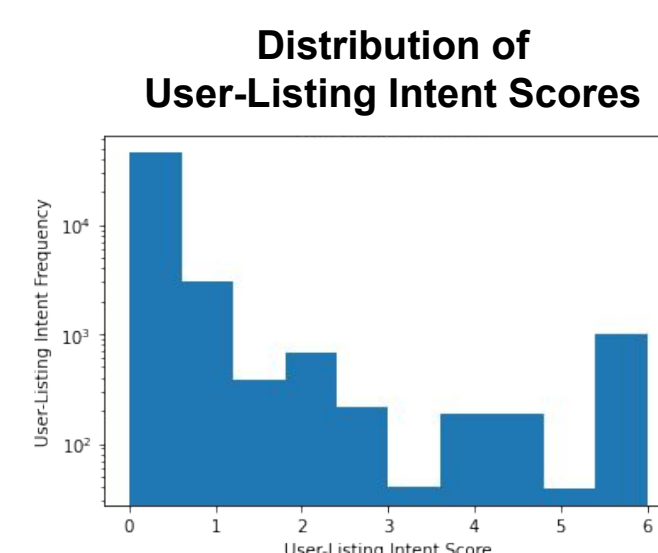
To get a better understanding of how users **browse listings** and **explore the site currently**, we defined windows of co-occurrence to understand which listings were being viewed sequentially by any users
Used in RF + Siamese NN models
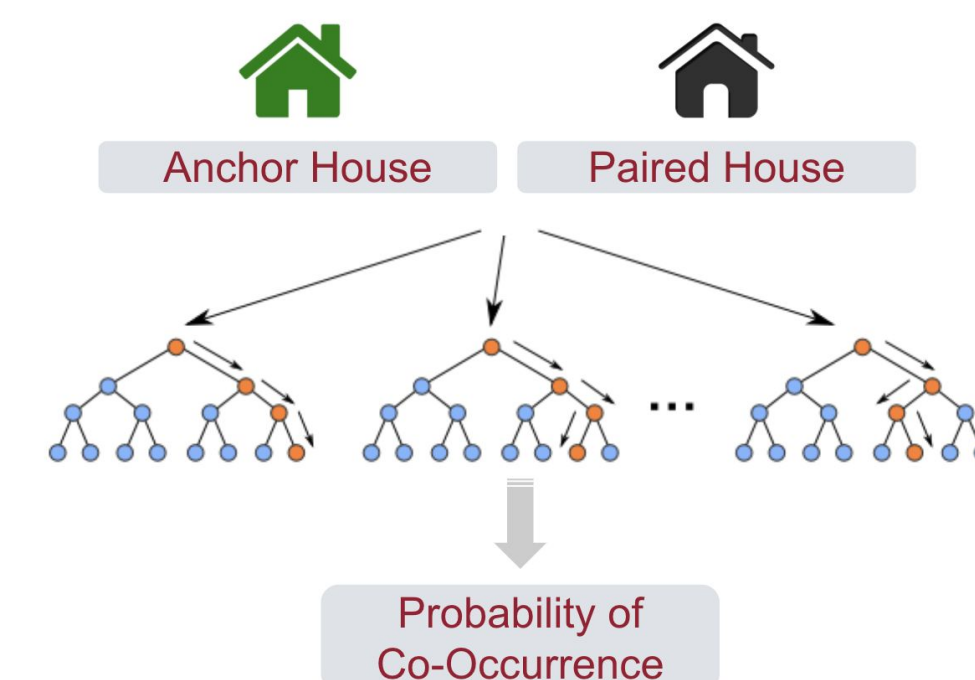
MODELS

Baseline Models

- **Content-based filtering:** recommend other listings similar to what the user likes (cosine-similarity)
- **Collaborative filtering:** recommend using similarities between users and listings simultaneously (matrix methods: SVD, NMF)
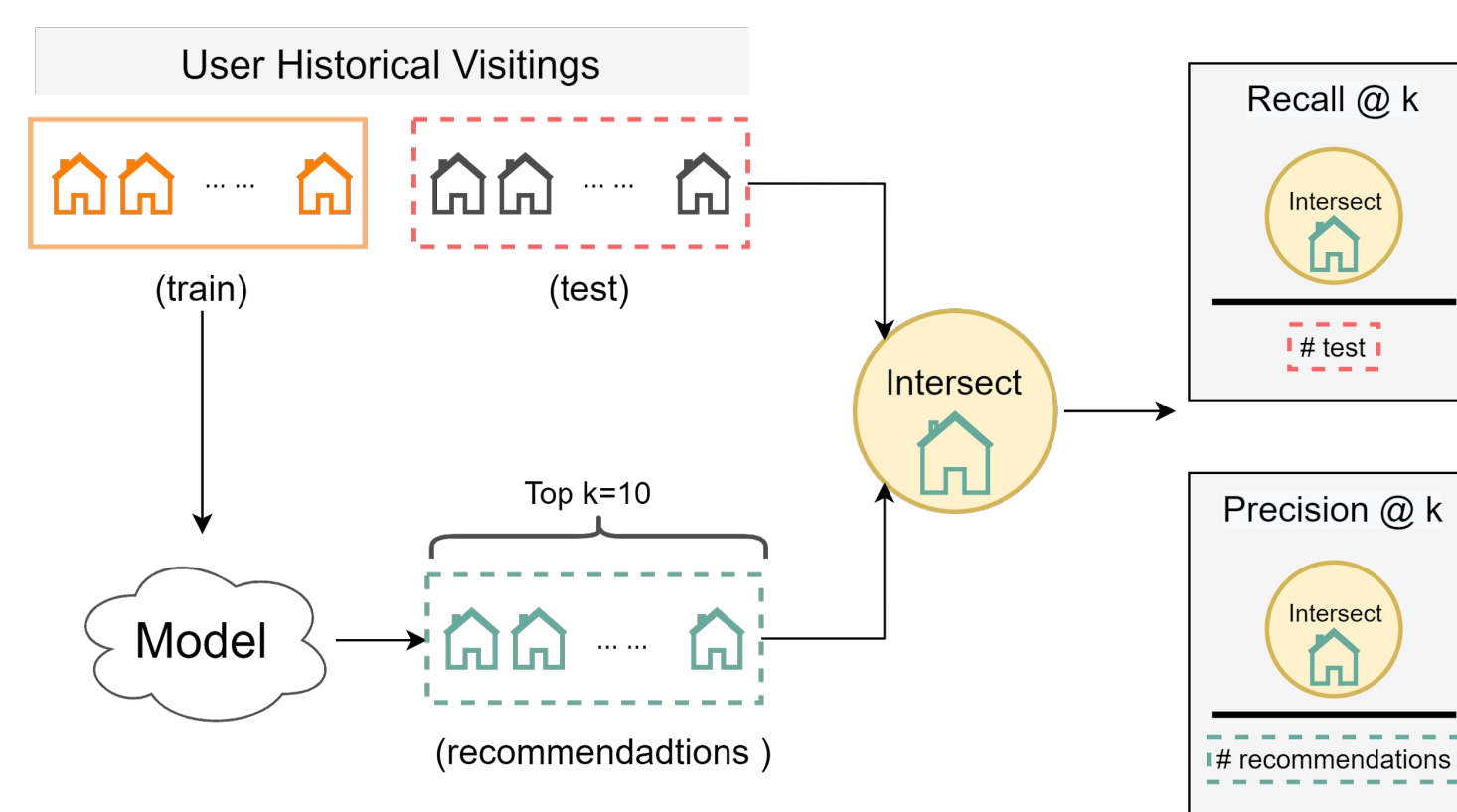
Siamese Network
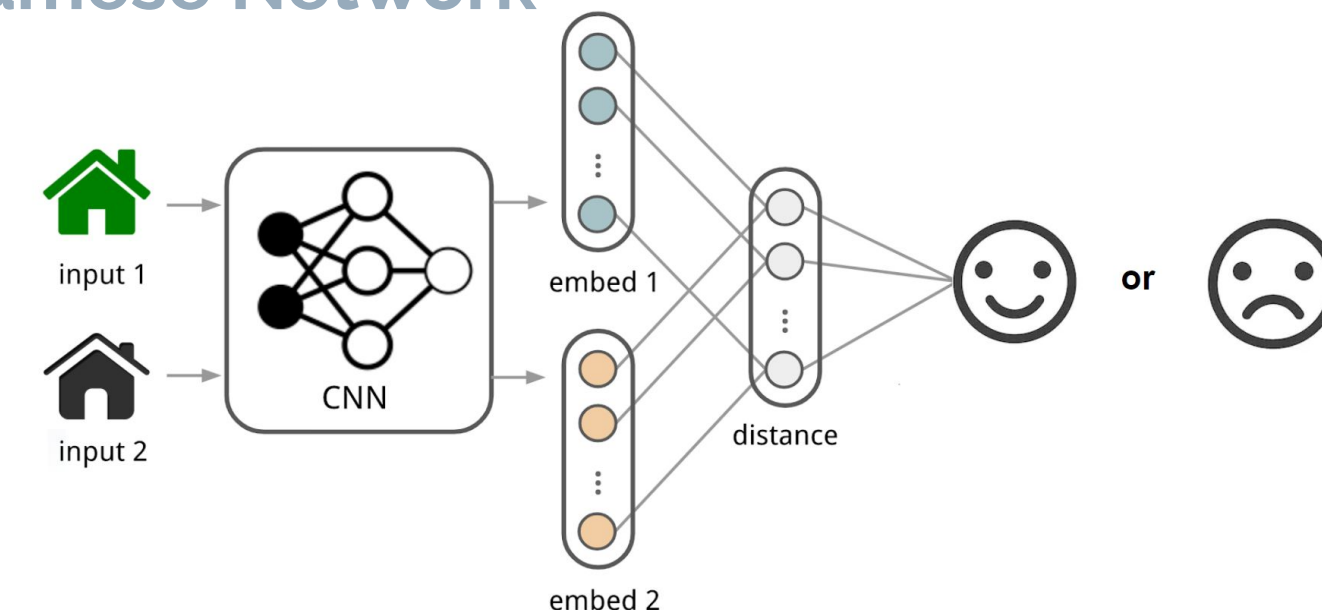
- **Content-based** approaches informed by **collaborative** information
- Trained on **co-occurrences**, predict on User Vectors

Random Forest

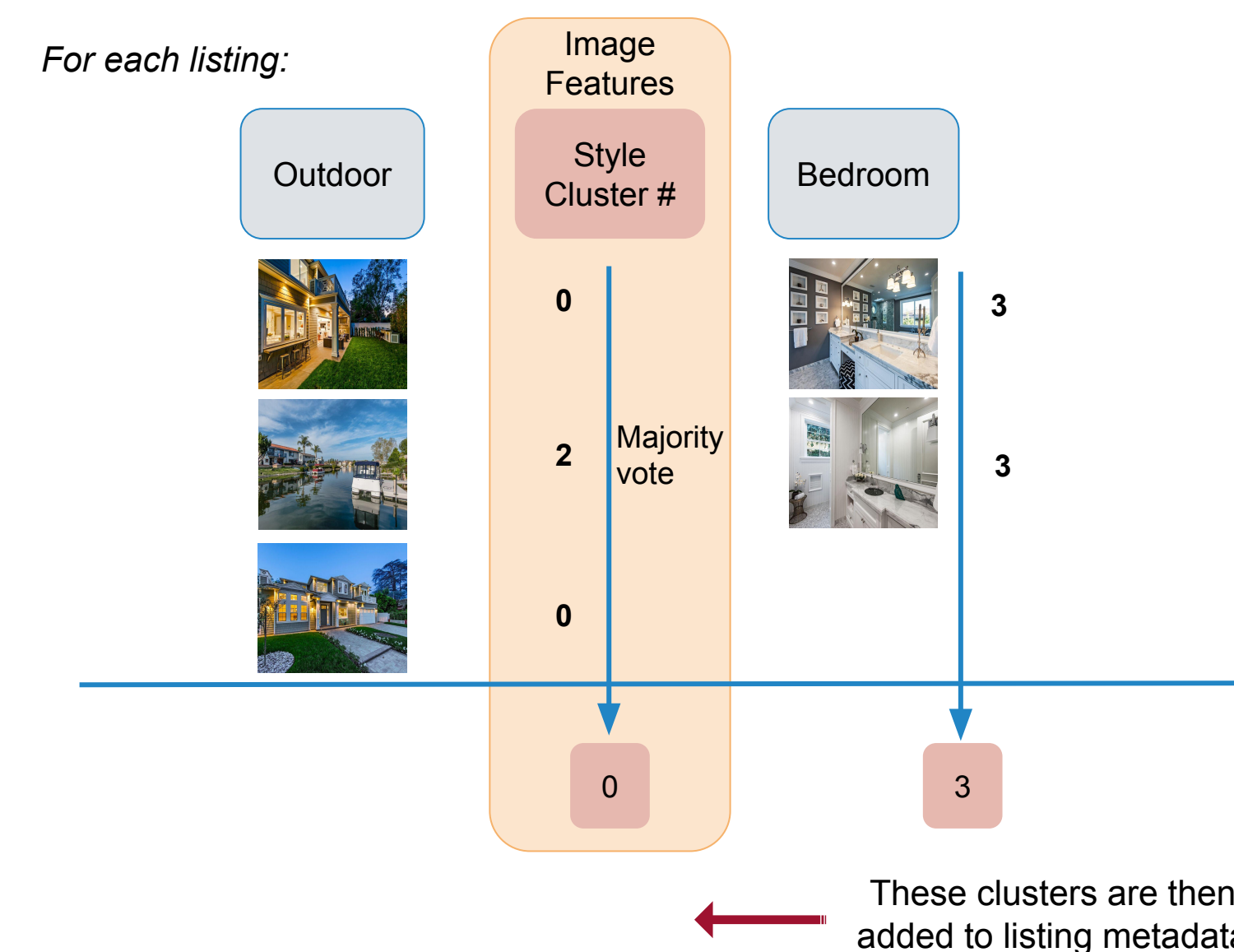### Image Feature Extraction

To extract some additional **image features** from listing images for recommendation models

Location Classification

- Classify listing images as indoor or outdoor using Places365 model
- Classify room types of indoor listing images using room classification model from REX API

Style Clustering

- Use clustered labels as features for recommendation models.

These clusters are then added to listing metadata

RESULTS

Steps for model evaluation using precision/recall @ k:

1. Split user historical visitings by time into train and test sets
2. Push the train visitings into the model and get a set of top 10 recommended homes (out of 500+ pool)
3. Compare the recommended list with the actual user visitings in test set, take the intersection group, and calculate the two metrics

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**

How do we interpret the two metrics?

- **Precision @ k**: In top k recommendation, how many of them are true user visitings (in test).
- **Recall @ k**: In all true user visit visitings, how many of them are being recommended. e.g. (RF) If we use recommended 10 listings, it covers 25% of the true user visitings

Using the two metrics, we find that, Random Forest and Siamese model perform at least 2x better than Content Filtering, and we decided to use these two as our final models
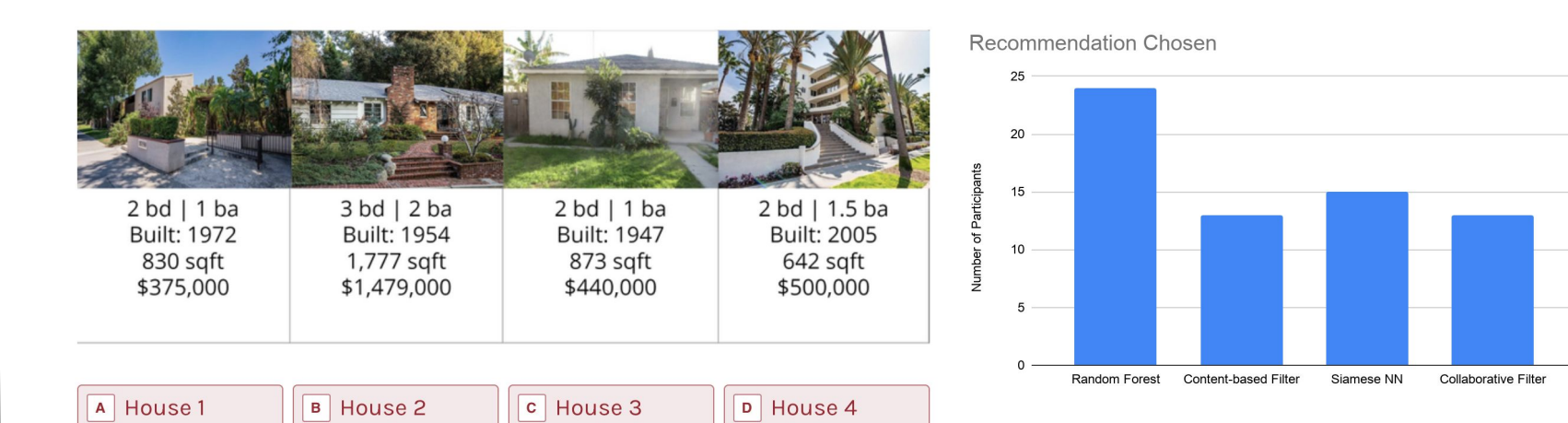
USER STUDY

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?

Given the 3 listings selected in stage 1, generate recommendations from each of our 4 models to present to participants
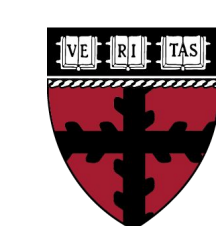
We received 65 responses for our survey.

The results from the user test aligned with our test set evaluation with the Random Forest recommendation being selected most frequently, followed by the Siamese Network Though we cannot make definitive conclusions yet, the current results could serve as good evidence for convincing REX's product team to continue our project.

CONCLUSIONS

- We found that it is most effective to model the flexibility of user behavior using click journey/co-occurrence as supervision as opposed to simply recommending the most similar listings
- Including information from Image style clustering improved our models
- Random forest was the best performing model on our test set; Siamese Network also performs relatively well and outperforms baseline recommendation system models
- The results of our user study also showed that random forest produced the most interesting recommendations
- Moving forward, an A/B test would be useful to test how our model impacts user retention or the efficiency of finding homes on REX's site

REFERENCES

- Home Embeddings for Similar Home Recommendations (https://www.zillow.com/tech/embedding-similar-home-recommendation/)
- Siamese Network Keras for Image and Text Similarity (https://medium.com/@prabhnoor0212/siamese-network-keras-31a3a8f37d04)
- Places365-CNNs (https://github.com/CSAILVision/places365)

REX REAL ESTATE

Harvard John A. Paulson **School of Engineering** and Applied Sciences