Rex House Discovery

Scope of Work (SoW) Document

Prepared by

Zhenwei (Selina) Wu zwu2@g.harvard.edu

Nikhil Vanderklaauw nvanderklaauw@g.harvard.edu

Wanxi (Cecilia) Yang wanxiyang@g.harvard.edu

Weiru Chen wchen@g.harvard.edu

Prepared for

Andy Terrel, aterrel@rexhomes.com

Background

REX is a full-service real estate agency that offers home listing services with its technology platform. It is differentiated from the conventional real estate agencies as REX does not list sellers' homes on multiple listing services (MLSs) that link brokers to other real estate brokers. Instead, REX matches sellers and buyers using data modeling and digital marketing — eliminating the commission for the buyer's agent. By cutting down the middlemen and working with REX's own team of licensed agents, REX saves the home seller the commission fee and provides an overall improved experience in home sales.

The key in helping home buyers find their ideal home lies in the effectiveness of the recommendation algorithms. While there are existing online platforms that allow the user to search for housing with different filters, these platforms often lack the ability to filter for details, do not account for many important features that home buyers seek, and therefore do not provide great searching results. REX wants to use image processing, natural language processing, and other tools to help their users to find the perfect home that satisfies their needs in every aspect. Our goal is to understand the users' needs and improve the current algorithms with various machine learning techniques.

Problem Statement

Our goal is to build an application that serves house-hunters with personalized matches for discovering their perfect home. Our app will quantify customers interests with minimal questioning and go beyond traditional search filtering of simple metadata, providing nuanced personalization that includes property information from the listing images and text descriptions. Our personalized recommendation will also be dynamic and incorporate customer feedback on the quality of recommendations.

After selecting the most salient features to obtain from users, we will approach this project as a recommendation problem and utilize collaborative-filtering, content-filtering and item2vec techniques as our baseline. We will start modeling with the current tagged features provided by REX's existing deep learning models. After completing the baseline, we will consider including additional features (both user-side and housing-side) to REX's NLP and CV models to extract

further features that may be useful for recommendation.

One of the challenges we will face is quantifying the success of the recommendations from a user perspective without deployed user testing. The true success metric will eventually be the purchase rate of recommended homes, but we will proxy this by asking for positive/negative feedback on the results of the model to get an understanding of the quality of recommendations, and the improvement achieved upon updating recommendations from dynamic feedback. We may also try using a hold-out on historical data of purchases to test if users ended up purchasing homes we would have recommended to them.

If time permits, we plan to further build out the deployment of our model in app form, and will run tests to compare the model performance between existing REX feature extraction to our implementations with additional features.

Resources Available

Data available includes:

- Datasets
    - FA Assessor Data
    - House Market Listing Property Data
    - Transaction Data
    - Beeswax Log File
- REX API including computer vision APIs and recommendation endpoint

Other possible resources, depending on the volume of data:

- Amazon Web Services (AWS)
- Apache Spark
- Apache Hadoop

High-Level Project Stages

Our plan is to divide the project into 4 primary stages of development:

1. Understand the features in selecting an ideal house and define the top k primary features that affect customer house selection.

2. Develop a collaborative filtering (CF) model that can recommend housing using features that can be extracted from REX's current models**.** Update the model using feedback from the customer.

3. Extract our own features using image processing and NLP, building on their existing models, in addition to features from other internal datasets, to the CF model from Part (2) in hopes of reaching an improvement in recommendation quality

4. [Stretch Goal] Use this improved model to generate new housing recommendation lists according to user-specified location or other search filters, and present these as potential ideas for implementation at REX.