Real Estate Recommendation System

Project Proposal

Table of Contents

Title: Generating personalized recommendations for customers interested in buy-to-let investment in real-estate sector

Abstract

The idea of generating appropriate recommendations for the user by analyzing available data has grown vastly over the past few years. There are less sufficient researches available on the generation of real-estate recommendations. We will be utilizing customers search history, property view history and past investments history of Mason Verdi clients and will create a self-learning model for generating real-estate recommendations. Based on the generated personalized recommendations, we can show new investment opportunities or send emails to the potential customers with their personalized recommendation. It will not only give more premium feel for customers but we can also have a greater chance of executing the deal.

Keywords: Real-Estate, Machine Learning, Natural Language Processing, Ranker

Introduction

Real estate is the property consisting of land and the buildings on it, along with its natural resources such as crops, minerals or water; the immovable property of this nature and its surrounding environments like hotels, hospitals, schools etc. Real estate management manages the necessaries required including personal property, machinery, instruments, and physical capital assets. The amount of potentially interesting content is increasing day by day. A real estate recommender system can help people cope with the situation. It becomes a nightmare to go through an enormous amount of available potential content to select the best ones. Besides, since the price and demand for renting or buying houses or lands are rising higher now-a-days, the real-estate market is becoming popular day by day all over the world.

Problem Statement and Aim

There is a substantial need of empowering potential customers with personalized recommendations in real-estate sector using Artificial Intelligence. So, we will be generating personalized recommendations for the people visiting Mason Verdi interested in buy-to-let investments to increase the probability of deal execution.

Project Objectives

Using machine learning and natural language processing algorithms to offer personalization for the buy-to-let investments for the customers on the basis of various sets of data including customers search history, property view history and past investments history at Mason Verdi.

Literature Review

Machine Learning and Natural Language Processing

The topics that are spoken, the tone, comments, statements of words are related to the amount of information. However, they are in their own format of unstructured manner. That information doesn't have proper structure and data frames like the numerical data and information (Yse, 2019). The unstructured data can be clearly distinguished from the traditional row and column structure of data storing. In other words, it is significantly messy and need some extra effort to manipulate data to generate more insights (Marr, 2019). So, the natural language like text or speech can be automatically handled and that process is called as Natural Language Processing (NLP).

Clustering

Moreover, the group of documents can be summarized into various clusters based on their similar content. The main idea is that the descriptors which have been built using set of words, explains a common content within the cluster. This unsupervised method will focus on the classify the objects into several groups and that groups can be called as the clusters. In simple terms, dividing the similar text into same group or cluster is the main idea of clustering (Anon., 2020). We can apply clustering on customers search history, property view history and past investments history to make sure that all the matching similar documents are combined into one cluster or group. Because, in the web scraping, most of the time, large amount of data will be scraped and in order to summarize the documents, the summarization application will be used.

Literature published in past

Atisha Sachan et al. granted a survey on recommender system based totally on collaborative filtering where they discuss the method to find out the interest of the user. They designed a book recommender system considering four sorts of filtering approach particularly demographic technique, content-based filtering, collaborative filtering, and hybrid method. They discussed stability and cold-start problem. But they didn't consider sparsity issue in details. Another important work by Prem Melville et al. where they defined different types of recommender system with associated limitations and challenges. Jyoti Gupta and Jayant Gadge analyzed the performance of recommendation systems based on content-based filtering and demographics. They solved the cold-start problem associated with the recommender system.

Proposed Methodology (tentative)

Learning similarity from user behavior

Similarity learning has applications in search ranking, clustering, and recommender systems. Without data, there are many ways we can define a distance or similarity measure, such as the cosine similarity of raw features. While this type of unsupervised approach is simple to implement, the choice of similarity measure may be arbitrary and may not reflect how users actually interact with the items and in our case, listings. In our project, when potential customers will explore homes on Mason Verdi, they will be implicitly providing feedback on listing similarity. A user viewing two listings in quick succession offers some evidence that they are similar to each other. When aggregated over the behavior of lots of users, anonymized listing browsing data offer an informative similarity signal we can use to train our algorithm.

We will be treating each model, either heuristic-based or learned, as a ranker which produces a rank. By using ranks, we avoid dealing with numerical scores produced by each individual component that may vary wildly in range and could be less robust for model ensembling should underlying numbers change. The specific implementation of our ranker ensembling formula involves taking the inverse of the rank output by each ranker, applying a sublinear transformation, and then calculating a weighted average (based on the relative importance of each subranker) as the final model score.
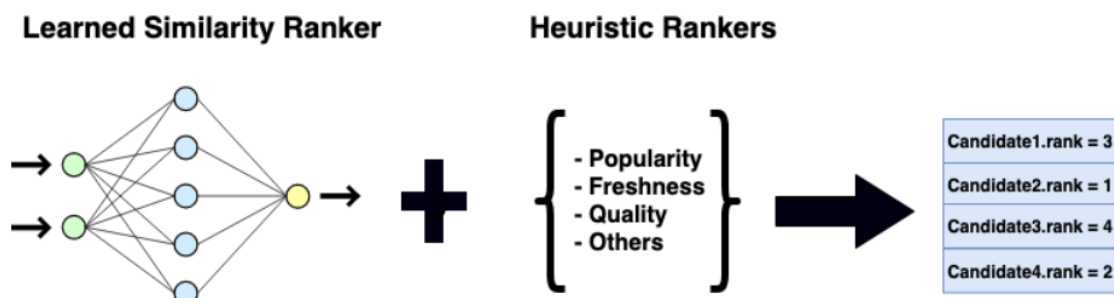

Figure 1

Summary and Conclusion

We will be generating personalized recommendations to potential customers interested in buy-to-let investments at Mason Verdi using similarity as a ranking problem and heuristics and natural language processing. The major motivation for this project is to use machine learning and natural language processing algorithms to offer personalization for the buy-to-let investments for the customers on the basis of various sets of data including customers search history, property view history and past investments history at Mason Verdi to increase the probability of deal execution. We will be learning similarity from the user behavior and then will be using ranker-based model in which we will be calculating a weighted average, based on the relative importance of each subranker as the final model score.

References

Tao, C., Guo, H. & Huang, Z., 2020. Identifying security issues for mobile applications based on user review summarization. Information and Software Technology, Volume 122.

Wang, K., 2017. Text mining technology based on cloud computing. Acta Technica CSAV (Ceskoslovensk Akademie Ved), Volume 62, pp. 72-83.

Anon., 2020. Text mining methods and techniques: The beginner's guide, s.l.: https://roboticsbiz.com/.

Marr, B., 2019. What Is Unstructured Data and Why Is It So Important to Businesses? An Easy Explanation for Anyone, s.l.: https://www.forbes.com/.

Yse, D. L., 2019. Your Guide to Natural Language Processing (NLP), s.l.: https://towardsdatascience.com/.