Realtor Insights' Assessor

1. Introduction

Finding the right home can be a difficult task for individuals lacking experience in the real-estate market. Often, these individuals seek consultation from Realtors, who may not have the client's best interests in mind. There is a clear opportunity to empower these individuals with a real-estate assessment tool that helps them understand the homes that best fit their needs and desires.

2. Proposed Project

2.1. Overview

Realtor Insights' *Assessor* tool helps individuals make an informed purchasing decision by leveraging algorithmic, mathematical, and statistical analysis; presented as a software solution. It collects specific information about an individual which may include (but is not limited to) age, income, ethnicity, and/or occupation; and a particular market location. Based on this information, combined with publicly available real-estate market data and local media coverage, *Assessor* presents the individual with an ideal neighborhood for their next home.

2.2. Technical Details

Finding the ideal neighborhood for an individual involves a highly personal, multivariate analysis. Drawing a parallel to how real consumers choose their next house, the *Assessor* tool looks at three key factors that influence a purchase decision - personal fit, financial outlook for its real estate market, and word-of-mouth details about its quality of life. Each factor is analyzed via distinct machine learning, statistical, and natural language processing models (respectively) to determine the ideal homes for users. These are detailed below.

2.2.1. Analyzing Personal Fit via Machine Learning Classification

The first model seeks to perform predictive classification of neighborhoods that the individual would most likely enjoy based on demographic and social factors. The data that this model leverages is two-fold. First, the model uses the data given by the user (e.g. age, income, occupation etc.) as discussed in Section 2.1. Second, the model uses publicly available data released by cities about the demographic and social details of various neighborhoods (e.g. City of Toronto - Open Data). Using these datasets, a machine learning algorithm can be trained to identify the probability that the user will enjoy living in a given neighborhood. While the exact type of algorithm is still under discussion, the team is investigating clustering methods, Random Forests and Neural Networks.

2.2.2. Analyzing Financial Outlook via Time Series Analysis

The second model produces a financial outlook of the real-estate market in neighborhoods. More specifically, the model employs time series analysis on housing values to identify trend and determine the opportunity cost for a particular property. Alongside typical time series analysis procedures, the team plans to implement recurrent neural networks and LSTMs to utilize past data in similar neighborhoods (i.e. similar location, resident profiles, etc.) to gather comparative predictive analytics - uncovering nuances of the data that is not readily evident. This model may look at broader economic factors beyond housing prices, time permitting.

2.2.3. Analyzing Quality of Life via Natural Language Processing

The third model leverages tools in natural language processing to translate word-of-mouth and media data into a commentary on a neighborhood's quality of life. Data will be harvested from social media (e.g. Twitter), and various media websites through their APIs, and sentiment analysis will be applied to quantitatively determine the general quality of a neighborhood. Moreover, tools in topic modeling (Latent Dirichlet Allocation for long text on news, Biterm Topic Modeling for short text on social media) will be leveraged to summarize key topics of online discussions surrounding the neighborhood.

2.2.4. Consolidating Findings and Presenting Results

The results from each model will be consolidated, and a heat map of neighborhoods will be generated for the individual to inspect (based on the probability that the neighborhood is the individual's ideal location). This map will allow the user to click on particular neighborhoods of interest, and understand what contributed to that location's score - drawing from the analysis conducted by each of the three models presented above. To ensure that the tool provides an end-to-end experience, from research to purchasing decision, we plan to present specific potential properties for the user to view in the neighborhoods the tool recommends. This can be achieved by leveraging a real-estate API that provides information on houses for sale.

3. Closing Remarks

By providing users with a real-estate research tool that is catered to their needs and interests, *Assessor* will help users make a more informed property purchasing decision. The tool is designed to leverage machine learning, statistics, and natural language processing to consolidate huge sources of data into actionable insights for home buyers - allowing users to tap into data at a scale that no traditional realtor can provide.

The Team

Eric Boszin

I am an individual with a love for technology and business and their combined potential to solve problems and create value. Often taking the leader role, I work best in teams and strive to increase my experience with individuals of varying backgrounds and skill sets. Upon graduation, I will be working in the Cybersecurity Consulting industry.

Henry Cho

I am passionate about extracting business insights from data. My work focuses on leveraging tools in software engineering and data science to translate big data into concrete business recommendations. In a team setting, I often engage with project management and execute technical goals. Upon graduation, I will be joining a hedge fund as a data scientist.

Dale Wu

I am driven by the products that software technologies build. My experience is in front end and full stack development of mobile and web applications. In a team, I like to help members engage in the bigger picture and work toward a unified goal. Upon graduation, I will be working as a software engineer for a financial company.

Shinjae Yoo

I am interested in learning how different pieces of a project work together. In a team setting, I often enhance communication and understanding between members working on separate components. My area of expertise is low level software and digital systems. Upon graduation, I will be working in the FPGA industry.