Project schedule

Group 8

2024-10-11

Project Title:

link to dataset: https://www.kaggle.com/datasets/juhibhojani/house-price

Group Member:


```
Jennifer Guo - jeguo@umich.edu
Jintong Hong - jintong@umich.edu
Sihan Chen - csihan@umich.edu
Wenjie Han - wenjieh@umich.edu
```


Project Deadlines:

Project Proposal - Nov.22nd/ Week 13

Presentation - Week 15

Report Deadline - Week 16

Group Rules:

1. Having regular weekly group meeting to check the progress of the project, resolve potential issues, and determine the plan and distribution of assignments for next week.
2. The schedule will update regularly after the meetings; please check the schedule to for any updates of the project.
3. Please update any changes to the main code and report in the group chat (WhatsApp).

Group meeting time:

Regular meeting time: Wednesday at 1pm
Next meeting time: Wednesday 1pm, Week 9

Project Progress:

Week 7:

1. Decided the dataset and topic of project (housing data, to predict the housing price with given variables)

2. Decided the weekly meeting time (Wed. 1 pm, sph I).
3. Assigned work for week 8.

Week 8:

1. Checked variables and we decided to use the following variables:
   Prices (dependent variable)/ location/ Carpet Area/ Status/ Floor/ Transaction/ Furnishing/ facing/ overlooking/ Bathroom/ Balcony/ Car Parking/ Onwership/
2. Decided to use modelling methods: 1. elastic net; 2. Random Forest;

Week 9:

1. Data Cleaning:
   - Keep all missing values and we will put them all together next Tuesday;
   - For elastic net, we might want to delete the extreme values and using numeric version of Bathroom and Balcony; for random forest, we would just keep the extreme values and using categorical Bathroom and Balcony;
2. Data Modeling:
   - Jennifer/ Sihan: elastic net
   - Jintong/ Wenjie: random forest
3. Extra Information Search (for location) - either income or tax
   - we want to use clustering for elastic net when processing the location variable.

Weekly Schedule:

Week 7 - Project Start

1. Find optimal datasets and topic/models for the project – Everyone;
2. Determine the next group meeting time. – Everyone send an optimal time to the group chat;
3. Determine the programming language will be using for the project – Python;
4. Adding group members to github and check if all member can access the repository – Please report any problem in the group chat;
5. Rough plans for rest of the semester – Sheryl;

Week 8 - Data expoloration/ Data cleaning

Everyone:

1. Data exploration.
2. Coding for categorical data.
3. Check if there's any weird things about the dataset itself.
4. Find models we can use.

Week 9 - Data expoloration/ Data cleaning / Start simple data modeling

1. Check extreme values/data patterns/ create plot (scatter for numerical/ barplot for categorical) for each variables/ Do one hot encoding for categorical variables (not for location)/ normalizing the numerical variables if possible (create a separate column for this). Please refer to following link for more information of normalization (more info can also be found in ppt sent during our week 8 meeting):

   https://www.datacamp.com/tutorial/normalization-in-machine-learning

- Sihan: Prices (dependent variable)/ location/ Carpet Area
- Jintong: status/ Floor/ Transaction
- Jennifer: Furnishing/ facing/ overlooking
- Haven: Bathroom/ Balcony/ Car Parking/ Onwership

2. Do data cleaning for the missing values (imputations) - optional
3. Do some research on the models (1. Elastic net; 2. Random Forest) - Everyone

Week 10 - Data modeling preparation

1. Upload all cleaned variables and corresponding graphs before next Tuesdays meeting - **Everyone**
2. Indian location income search - **Wenjie**
3. Write simple sketch for:
    - introduction (background information of housing& purpose of the modeling) - **Jennifer**
    - data cleaning (**Everyone**) + how we are going to separate the datasets to testing and training set - **Wenjie +** Clustering methods used for location (**Jintong/** write for both Isolation Forest or Local Outlier Factor (LOF))
    - modeling methods (random forest/elastic net) - each group write their own version of modeling methods and we will put them together next week during the meeting. – **Everyone**
    - how we are going to test the accuracy of the predictions (cross validation) - **Sihan**
    - references if used any - **Everyone**

Week 11 - Write proposal draft/ Start final report (introduction)/ Data modeling

Week 12 - Revise proposal/ Model testing/ final report (data summary, modeling method)

Week 13 - Finalize and submit proposal (Nov.22nd, Fri.)/ Writing final report (result)

Week 14 - Finalize code/ Prepare for presentation (power point, presenter, etc.)/ Revising final report (conclusion, abstract)

Week 15 - Presentation/ Finalize report (references)

Week 16 - Wrap up and Submit the Final Report and Code