AMSTERDAM RENT PRICE

Report document no. 3

Contents

1. Introduction

In this document I will be addressing how I have commenced Phase 4 of the AI project methodology.

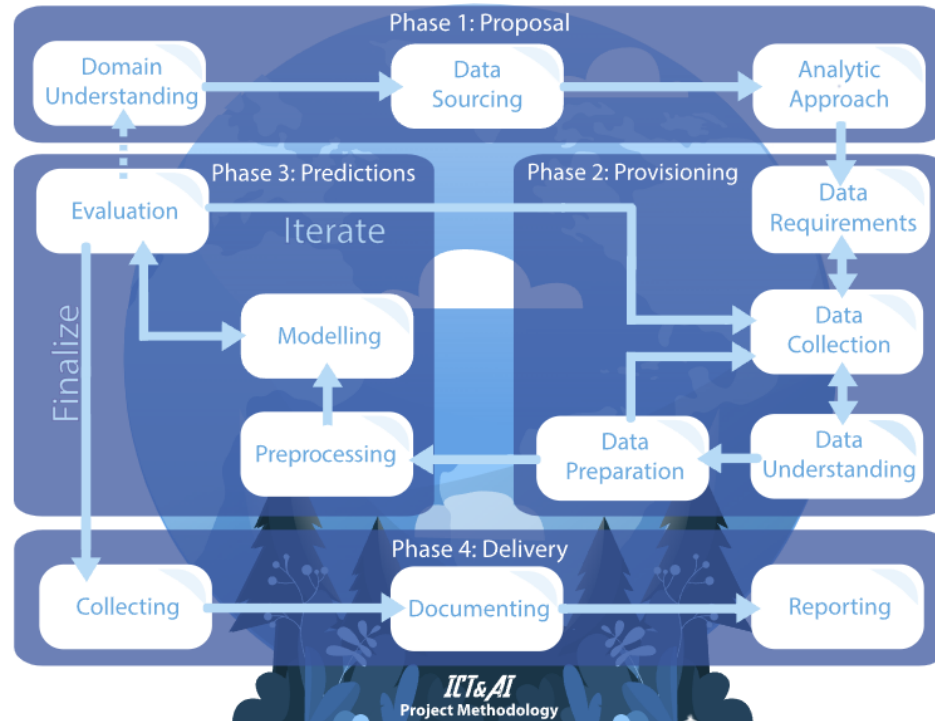This project journey started a bit differently than the AI project methodology.

Figure 1 AI project methodology

As if we look at figure 1, It is established that in the first phase we start with domain understanding, data sourcing and finally analytic approach. This project started with data sourcing, analytic approach and then domain understanding. That is because finding data is hard and nearly impossible, as if the software skills of scrapping data from a website isn't available, then data sourcing would be a problem. Due to that, domain understanding would have been iterated multiple times, due to data not being available. Therefore, after sourcing the data and deciding what the analytic approach should be, came domain understanding. Where the emergence of problem investigation skills happens. Based on the data, research questions were formed, the aim of the project was defined, a lot of research from various reliable sources (i.e. Amsterdam gemeente website) were done, and finally an interview with a domain expert was done.

In the third phase, 4 machine learning models were used, and evaluated. Nevertheless, if this project were to ever be used by the public or the target audience, it would not be very visually pleasing nor very comprehendible, and puzzling, due to the models being in a code form (i.e. Jupyter notebook).

Consequently, It was decided that for the delivery phase, an extra step will be taken. Where an integration between python and power BI will be developed, for a feasible and desired delivery.

2. Power-BI

Power BI is a business intelligence focused application, that was developed by Microsoft. It is used to visualize data interactively, and it can be integrated with python. Before the start of the integration, a version compatibility has to be checked, as in not all python versions work on power BI and especially anaconda.

The dashboard is created for the purpose of feasible visualization of the results and so the aim here is to elaborate more on the dashboard and in the chapter below, the measures used will be addressed.
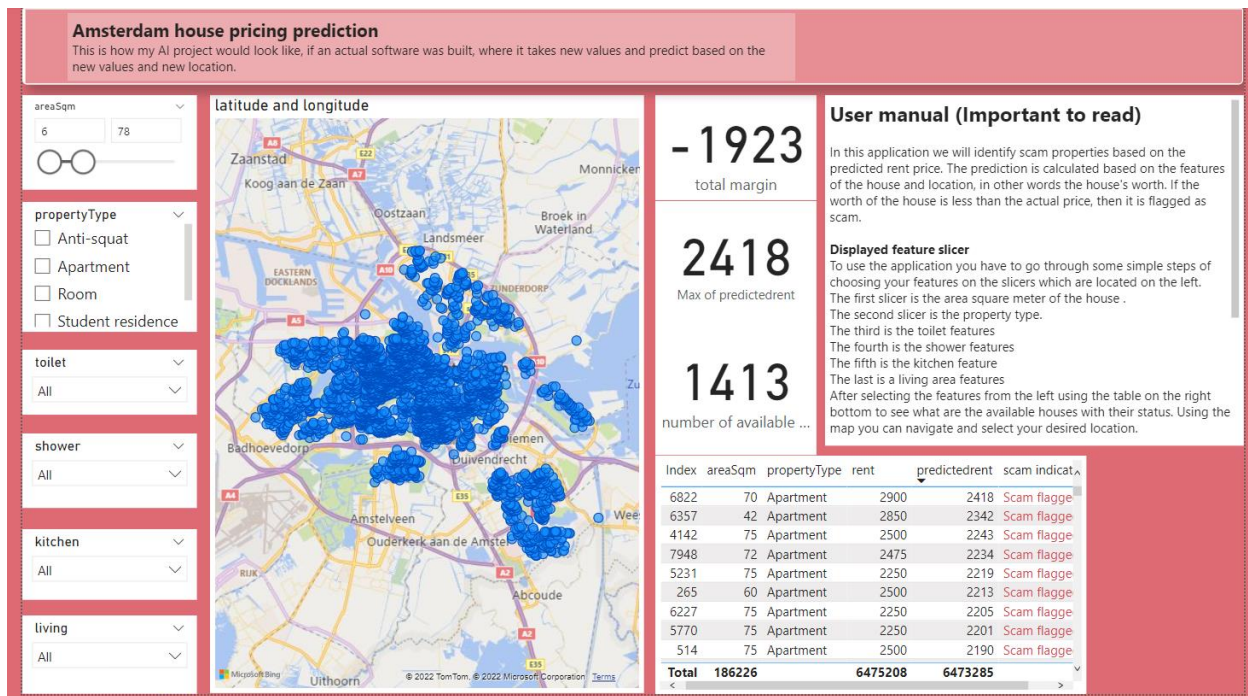

Figure 2 Dashboard page 1

This is the overview of first and only page of the dashboard, Simple, and minimalistic design.

User manual (Important to read)

In this application we will identify scam properties based on the predicted rent price. The prediction is calculated based on the features of the house and location, in other words the house's worth. If the worth of the house is less than the actual price, then it is flagged as scam.

Displayed feature slicer
To use the application you have to go through some simple steps of choosing your features on the slicers which are located on the left.
The first slicer is the area square meter of the house .
The second slicer is the property type.
The third is the toilet features
The fourth is the shower features
The fifth is the kitchen feature
The last is a living area features
After selecting the features from the left using the table on the right bottom to see what are the available houses with their status. Using the map you can navigate and select your desired location.

Figure 3 User manual of dashboard

The 3rd figure shows the user manual of the application, how the application is supposed to work and what it do.

-1923
total margin

2418
Max of predictedrent
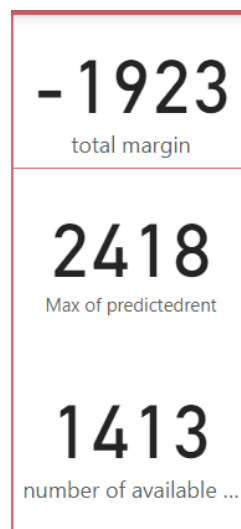
1413
number of available ...

Figure 4 cards

In figure 4, cards were used, due to the simplicity of them in delivering useful information. The measures used for these are discussed in the measures chapter.

Index	areaSqm	propertyType	rent	predictedrent	scam indicat^
6822	70	Apartment	2900	2418	Scam flagge
6357	42	Apartment	2850	2342	Scam flagge
4142	75	Apartment	2500	2243	Scam flagge
7948	72	Apartment	2475	2234	Scam flagge
5231	75	Apartment	2250	2219	Scam flagge
265	60	Apartment	2500	2213	Scam flagge
6227	75	Apartment	2250	2205	Scam flagge
5770	75	Apartment	2250	2201	Scam flagge
514	75	Apartment	2500	2190	Scam flagge
Total	**186226**		**6475208**	**6473285**	

Figure 5 Table of house information

Figure 5 shows the information about the specific house, it is also clickable. The table shows the number of the house, the area of the house, the property type, the rent, the predicted price and the scam indication.
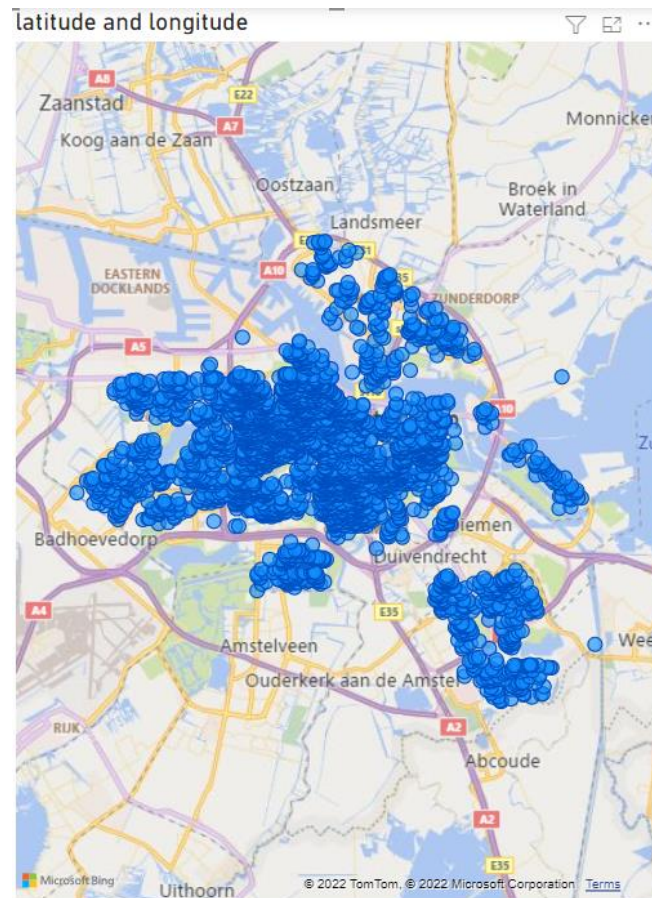

Figure 6 map of amsterdam and the houses

This is a map of Amsterdam that displays the houses and can be filter through the slicers which will be presented below. It is also interactive, the location of choice can be chosen.

Figure 7 house features slicer

These are the filters or the desired features of the house where the features can be selected.

- The first slicer represent the area of the house.
- The second slicer represents the property type
- The third slicer represents the toilet features(i.e. shared, owned)
- The fourth slicer represents the shower features(i.e. shared, owned)
- The fifth slicer represents the kitchen features(i.e. shared, owned)
- The sixth slicer represents the living room features(i.e. shared, owned)

3. Power BI measures

In this chapter the measures used to achieve the desired outcome, will be addressed.

1. Total margin measure: this is the difference between the predicted price and the actual price

   ```
   total margin = [totalpred]-[totprice]
   ```
2. Number of available houses based on the features selected.

   ```
   number of available houses = DISTINCTCOUNT(house[predictedrent])
   ```
3. Scam indication: This is done based on the prediction, as the model predict the rent of the house based on how much it is worth, and if the prediction was lower than the asked rent then it is probably a scam and if it is higher, it is negotiable.

   ```
   scam indication = IF(house[predictedrent]<house[rent],"Scam flagged","within margin" )
   ```

4. Interview summary

To collect feedback and enhance the iteration of my AI challenge I have interviewed Deems Veldkamp. Who is a nurse and also is one of my target audiences. Deems is currently living with his girlfriend in a rented apartment. He has been living there for about a year. It wasn't that hard for him, because before that he moved in with his girlfriend, he used to live in a village, where it was high priced. According to deems, his niece got scammed once in the city of Utrecht, where she used to live in a 7 square meter room for 400 euros, which is insane. (Veldkamp, 2022)

To deems knowledge, it is almost impossible to identify scams, due to landlords perceiving this as an opportunity to scam money from the tenants. The only possible way to identify a scam is by giving the property a visit and conducting your own research, as what deems said "I think nine out of ten times you're getting scammed." (Veldkamp, 2022)

When the power Bi was presented to deems, the question of "what is your first impression ?"was asked, to get feedback on the GUI if it is visually pleasing, and according to deems the colors need some work on. However, everything is clear and comprehendible.

In the opinion of deems, this application would have a huge positive impact on society. That is due the data being collected from various sources and put into 1 system. Hence easier to make your own research in one resourceful place, instead of multiple. Additionally, tenants or target groups can take this into consideration when making a decision over a house and if this was actually deployed and the public commenced using it, it will force the prices to deflate. That is due to landlords being forced to use the program too. Ethic wise, it is ethical as it would flag those people who abuse the system and take advantage of the community.

5. Impact analysis

AI are artifacts that detect something or effect something based on a detected values. As we evolve the capacity of creating such artifacts also evolves. It is a technology that is used daily, and contributes heavily to big data. Throughout the years and through science fiction movies, Humans have developed the idea of Ai being dangerous and that AI would take over our jobs and even us as a life form. (West, 2022)Forgetting the fact that AI is used everywhere, from Siri on your iPhone to the AI that plays chess or cards against you. That might be true in the far future, but what is not realized is that AI now has both positive and negative effects on us. According to Jaroen, all technologies can have a negative and a positive impact on society and it all depends on the intentions of the creator (Koole, 2022). Therefore, even though an AI is an artificial brain, it cannot teach it self the normal ethics, but the creator can. An example of this is Facial recognition AI.

Facial recognition can be used in a good way, as emotion recognition through facial expressions, as how the body world Amsterdam museum is using it. On the other hand, it can be sold to countries like China, that would facially recognize their citizen and since that credit scores are introduced there (lee, 2019), that would affect humans in an unhuman way and would strip them of all their basic human rights, besides the psychological effects on humans (Ng, 2020). The program itself is not evil, but the usage of it is. According to Jeroen and Deems, the prediction of house scams in the Netherlands, would have a huge positive impact on the society (Veldkamp, 2022) (Koole, 2022). As in landlords, would be flagged if they continue scamming, landlords will be forced to use the application to rent their homes, less

individuals and students get scammed and more affordable houses available. In general this project would put landlords under some sort of governance, regulate houses, and find the base amount of price for price negotiation and if anything an AI project like this, won't have a negative impact on society but a positive one, as this would also make houses more affordable.

6. Goal analysis

The main goal of the project was to forecast the housing rent prices for potential local tenants as well as international newcomers with respect to their Budget, financial plan, and preferred features to spread awareness about fraud, scams and to acknowledge the base amount within the housing market by building various regression models. According to the target audience and the stakeholders, the goal of the project has been achieved, with respect to the problem statement. As the product of the project is able to estimate the rent price and the price margin for a tenant in Amsterdam, given a housing with specific coordinates, housing square meter and tenancy agreement (i.e., shared, or individual).

7. Goal process

As to how the goal was achieved, it started by collecting some data requirements (i.e., coordinates, Size and type). After that, it was known that 2 data sets were required and that made the task of data collection a bit complicated, due to the fact that it is hard to source reliable datasets. Therefore, the first data set of "Kamernet" was found on github and the second data set was crawled manually from the website of "Huurwoning". Then, the datasets were cleaned, merged, and some extensive data understanding was commenced, and in the data understanding phase, the study of both datasets is done. Finally, before the prediction of the rent price was commenced, some revison on analytical approach was done, and it was decided to use 4 different regression algorithms. That is due to us predicting the question of "How much?"

These 4 types of regression were used on the dataset from kamernet and the merged dataset (Kamernet and huurwoning), therefore so far it is a total of 8 models that will be compared. In addition to that, since this project is made for studying purposes, it was decided that a study between a tuned hyper parameter and a non-tuned hyper parameters algorithms will be done and compared. So that, we can choose the best possible algorithm for this project and to understand the purpose of each algorithm, so in total 16 models were built, compared and evaluated. Furthermore, to enhance the domain knowledge and to check if this project is on the right path, an interview with a domain expert was conducted. Subsequently, best algorithm was chosen and from that an interactive feasible dashboard was built, to make the end results of the jupyter notebook prediction usable by a non-IT user.

Note: The Dashboard contains a user manual in the first page

8. Report and conclusion

It can be observed that, the power Bi is a better visualization of the results than the jupyter notebook and interactive. In addition to that, as it was mentioned before the power Bi application does not predict the price based on new features, as that is done in the jupyter notebook. The power BI application displays only how the application is supposed to look like if this was developed by an expert in software engineering. Based on the interviews that were made, they all concluded that this project could have an impact on society and it is worth investigating, as it would be revolutionary to the rental market. Additionally, It can also be concluded that, the whole development/progression of this project from the beginning was specific targeting a specific audience, was also measurable by having a lot of references from different sources. The project was also attainable as it was done before the 12th week, also relevant, as it is reasonable and can achieve the goal of the project and finally it was time based, as every AI phase was made in a specific time and had a time limit for motivation.

Furthermore, to use the dashboard, it is important to download the powerBI application and to read the "User manual" that is on the first page of the dashboard. Moreover, for future improvements of the IT side of the project, it is recommended to add more data to enhance the dataset, and to try different tuned models, as what Winston Churchill once said "To improve is to change; to be perfect is to change often."

9. Bibliography

Koole, J. (2022, 04 06). House renting. (R. Fuad, Interviewer)

lee, K.-F. (2019, 07 15). China and silicon valley. (L. Fridman, Interviewer)

Ng, A. (2020, 08 11). *How China uses facial recognition to control human behavior*. Retrieved from Cnet: https://www.cnet.com/news/politics/in-china-facial-recognition-public-shaming-and-control-go-hand-in-hand/

Veldkamp, D. (2022, 04 06). Amsterdam rental price feedback . (R. Fuad, Interviewer)

West, D. (2022, 03 09). *Using AI and machine learning to reduce government fraud*. Retrieved from Brookings: https://www.brookings.edu/research/using-ai-and-machine-learning-to-reduce-government-fraud/

10.   Documents

phase 4 record
Deems.docx

phase 4 record
Deems.aac