**Predictive Modeling of House Prices in Bishkek: A Comprehensive Analysis**

A blue arrow with black background

Description automatically generated

By

Muniza Hashim

In partial fulfillment of the requirement for the degree

Bachelor of Science in Computer Science

Department of Computer Science, School of Arts and Science

University of Central Asia, Naryn Campus Kyrgyz Republic

(2024)

**DECLARATION BY AUTHOR**

I/we certify that this work has not been accepted in substance for any degree and is not concurrently being submitted for any degree other than that of Bachelor of Science in Computer Science being studied at the Department of Computer Science, School of Arts & Science, University of Central Asia, Kyrgyz Republic. I/we also declare that this work is the result of my/our own findings and investigations except where otherwise identified by references and that I/we have not plagiarized another’s

Muniza Hashim

**DECLARATION BY SUPERVISOR**

I, the undersigned hereby certify that I have read this project report and finally approve it with

recommendation that this report may be submitted by the authors above to the final year project.

evaluation committee for final evaluation and presentation, in partial fulfillment of the

requirements for the degree of Bachelor of Science in Computer Science at the Department of

Computer Science, School of Arts & Sciences, University of Central Asia, Kyrgyz Republic.

Dr. Muhammad Fayaz

**Table of Contents**

[ABSTRACT 3](#_Toc134964673)

[ACKNOWLEDGEMENTS 3](#_Toc134964674)

[CHAPTER 1: INTRODUCTION 10](#_Toc134964675)

[1.2 AIMS AND OBJECTIVE 11](#_Toc134964676)

[1.2 AREA 11](#_Toc134964677)

[CHAPTER 2: LITERATURE REVIEW 13](#_Toc134964678)

[CHAPTER 2: PROPOSED METHODOLOGY 14](#_Toc134964679)

[3.1 DATA COLLECTION AND PREPROCESSING 16](#_Toc134964680)

[3.1 REGRESSION 17](#_Toc134964681)

[3.2.1 CATBOOST: 18](#_Toc134964682)

[3.2.2 LIGTGBM 19](#_Toc134964683)

[3.2.3 XGBOOST 19](#_Toc134964684)

[3.2.4: RANDOM FOREST 20](#_Toc134964685)

[3.2.5: EXTRA TREE REGRESSOR 21](#_Toc134964686)

[3.2.6 NEURAL NETWORKS 22](#_Toc134964687)

[3.2.7 KNN ALGORITHM 2](#_Toc134964688)3

[3.2.8 DESIONTREEREGRESSOR 2](#_Toc134964688)4

[CHAPTER 4: IMPLEMENTATION AND RESULTS/DISCUSSION 2](#_Toc134964689)5

[4.1 IMPLEMENTATION STEP 2](#_Toc134964690)6

[4.2 RESULTS AND DISCUSSION FOR THE MODELS 2](#_Toc134964691)7

[4.2.1 RESULTS AND DISCUSSION FOR CATBOOST REGRESSOR 28](#_Toc134964692)

[4.2.2 RESULTS AND DISCUSSION FOR LIGHTGBM 28](#_Toc134964693)

[4.2.3 RESULTS AND DISCUSSION FOR XGBOOST 29](#_Toc134964694)

[4.2.5 RESULTS AND DICUSSION FOR RANDOM FOREST REGRESSOR 30](#_Toc134964695)

[4.2.6 RESULTS AND DICUSSION FOR EXTRA TREE REGRESSOR 31](#_Toc134964696)

[4.2.7 RESULTS AND DICUSSION FOR NEURAL NETWORKS 32](#_Toc134964697)

[4.2.6 RESULTS AND DICUSSION FOR LINEAR-REGRESSION 3](#_Toc134964696)3

[4.2.7 RESULTS AND DICUSSION FOR DESIONTREEREGRESSOR 3](#_Toc134964697)4

[4.3 RESULTS AND DISCUSSION FOR PREDICTING MODEL](#_Toc134964704) 41

[4.3.1:MODEL – I](#_Toc134964705) 42

[4.3.2:MODEL – II](#_Toc134964706) 43

[CHAPTER 5: USER INTERFACE – WEB APPLICATION 47](#_Toc134964709)

[5.1](#_Toc134964710) [PREDICTION BUTTON 49](#_Toc134964711)

[CHAPTER 6: CONCLUSION AND FUTURE WORK 53](#_Toc134964713)

[BIBLIOGRAPHY 55](#_Toc134964714)

[CHAPTER 7: TABLE OF FIGURES 57](#_Toc134964709)

[CHAPTER 8: CONCLUSION AND FUTURE WORK](#_Toc134964713) 60

TABLE OF CONTENTS: TABLES

[**Table 1**: Notations and explanations 12](#_Toc134964740)

[**Table 3.1.1**: Label encoding of categorical values. 17](#_Toc134964741)

[**Table 3.1.2**: Input and output values for regression models 17](#_Toc134964742)

[**Table 3.2.1:** Input values for the classification models 18](#_Toc134964743)

[**Table 4:** : Hardware and Software Specifications used in this project 24](#_Toc134964744)

[**Table 4.1** Details of Regression models 25](#_Toc134964745)

[**Table 4.2:** Details of Classification models 25](#_Toc134964746)

[**Table 4.2:** Evaluation of regression models 26](#_Toc134964747)

[**Table 4.3**: Evaluation of classification models 34](#_Toc134964748)

[**Table 4.3.1:** Evaluation matrices of LightGBM classifier on training dataset 35](#_Toc134964749)

[**Table 4.3.2:** Evaluation matrices for testing dataset 36](#_Toc134964750)

[**Table 4.3.3**: Evaluation matrices of training dataset 37](#_Toc134964751)

[**Table 4.3.4:** Evaluation matrices of LightGBM Classifier 37](#_Toc134964752)

[**Table 4.3.5**: KNN on training dataset 38](#_Toc134964753)

[**Table 4.3.6:** KNN (K-nearest neighbour) algorithm on testing dataset 39](#_Toc134964754)

[**Table 4.3.7:** Accuracy of Neural Networks 39](#_Toc134964755)

[**Table 4.3.8**: Evaluation matrices of XGBoost Classifier 40](#_Toc134964756)

**TABLE OF CONTENTS: FIGURES**

[**Figure 1:** Abstract model of proposed methodology 14](#_Toc134821208)

[**Figure 3.2.1**: Detail structure diagram of different stages 15](#_Toc134821209)

[**Figure 3.3:** Handling NaN values. 16](#_Toc134821210)

[**Figure 3.3.1**: Weak learners in CatBoost Model 18](#_Toc134821211)

[**Figure 3.3.2:** Leaf-wise growth of Decision Trees 19](#_Toc134821212)

[**Figure 3.2.3:** Simplified Structure of XGBoost Model 20](#_Toc134821213)

[**Figure 3.2.4:** Simplified Structure of Random Forest Algorithm 21](#_Toc134821214)

[**Figure 3.2.5:** Extra Trees Algorithm 21](#_Toc134821215)

[**Figure 3.2.6:** Neural Newtork for Multiple Classification 22](#_Toc134821216)

[**Figure 3.2.7:** Working ofKNN algorithm 23](#_Toc134821217)

[**Figure 4.2.1:** R2 Score, MAE and MSE formulae 27](#_Toc134821218)

[**Figure 4.2.1**: Results of CatBoost Regressor on training dataset 27](#_Toc134821219)

[**Figure 4.2.2**: Results of CatBoost Regressor on testing dataset 28](#_Toc134821220)

[**Figure 4.2.3:** Results of LightGBM Regressor on training dataset 28](#_Toc134821221)

[**Figure 4.2.4:** Results of LightGBM Regressor on testing dataset 29](#_Toc134821222)

[**Figure 4.2.5**: Results of XGBoost Regressor on training dataset 29](#_Toc134821223)

[**Figure 4.2.6:** Results of XGBoost Regressor on testing dataset 30](#_Toc134821224)

[**Figure 4.2.7**: Results of Random Forests on training dataset 30](#_Toc134821225)

[**Figure 4.2.8**: Results of Random Forests on testing dataset 31](#_Toc134821226)

[**Figure 4.2.9**: Results of Extra Trees Regressor on training dataset 31](#_Toc134821227)

[**Table 4.2.10:** Results of Extra Trees Regressor on testing dataset 32](#_Toc134821228)

[**Figure 4.2.11:** Results of Neural Network on training dataset 32](#_Toc134821229)

[**Figure 4.2.12**: Results of Neural Network on testing dataset 33](#_Toc134821230)

[**Figure 4.3.1**: Pie chart showing the output categories. 33](#_Toc134821231)

[**Figure 4.3.2:** Evaluation matrices for classification models 34](#_Toc134821232)

[**Figure 4.3.1:** Confusion matrix of training part 35](#_Toc134821233)

[**Figure 4.3.2:** Confusion matrix for the testing dataset 35](#_Toc134821234)

[**Figure 4.3.3:** Confusion matrix on training dataset 36](#_Toc134821235)

[**Figure 4.3.4:** LightGBM classifier on Testing data 37](#_Toc134821236)

[**Figure 4.3.5**: KNN on training dataset 38](#_Toc134821237)

[**Figure 5.1.1**: First graph in Analyze Section 49](#_Toc134821255)

[**Figure 5.1.2:** Second part of Analyze section. 50](#_Toc134821256)

[**Figure 5.1.3:** Last part of Analyze section. 51](#_Toc134821257)

[**Figure 5.2.1:** Predictions of the LightGBM Classifier 52](#_Toc134821258)

[**Figure 5.2.2:** Regression models prediction 53](#_Toc134821259)

[**Figure 5.3.1:** Sidebar section for Forecasting 53](#_Toc134821260)

[**Figure 5.3.2:** Representation of Predictions 54](#_Toc134821261)

# *Abstract*

The project, titled "Predictive Modeling of House Prices in Bishkek: A Comprehensive Analysis," dives deep into the challenges and complication of understanding the real estate market in Bishkek, Kyrgyzstan. It is aimed to analyze the factors that affect the prices of houses in Bishkek and this study leverages a dataset consisting of 10,000 observations, that have gone through some enriching feature engineering to come up with around 30 independent variables. A variety of machine learning models will be used to train the data and will be then evaluated to build most predictive model. The outcome of the application will be a web application helping users to predict the prices of houses in the market from home, based on the kind of house they want. This application will provide all insights into the datasets and methodologies that have been used in the process, with some extra resources that will be available for any user on GitHub for educational purposes.

The importance of this tool is rooted in the dynamic nature of the real estate market in Bishkek, that often leave the owners and the buyers uninformed about the factors that are affecting the property prices. This projects yet targets to provide all the necessary information related to market trends, and facilitate fair transaction among stakeholders, investors, house owners, and importantly renters. The research will cover a thorough literature review of the researchers that have highlighted the historical perspectives and current market conditions influenced certain factors that most of the people are not aware of.

*Keywords:* Machine learning, Bishkek, Real Estate, House Prices, Dataset, User-interface

# **CHAPTER 1: INTRODUCTION**

# *Project Description*

*Introduction:*

Real estate is a very dynamic and constantly evolving field where the prices of any kind of building or houses is affected by several factors (Roseman, 2004). In Bishkek, the capital city of Kyrgyzstan, real estate market is no exception and for many years in the past it has gone through many vivid and big changes that have reshaped the real estate market and have also influenced prices, often leaving the sellers and the buyer with lack of information regarding the main factors behind these variations (Gazette, 2023). In response to this need this project aims to bring up such an approach to predict the house prices in Bishkek.

My final year project “**Predictive Modeling of House Prices in Bishkek: A Comprehensive Analysis**” will be addressing and analyzing the high and low prices of houses in Bishkek based on different factors about which people are not aware. My project will be talking and making clear the asymmetry of the information that the real estate creates in Bishkek. I have a data set of 10000 observations, which probably will have more than 30 independent variables after some feature engineering. There will be a machine learning model on which the data will be trained unless will get a good accuracy. I will be training several models with the data because sometime the accuracy of any model depends on the model as well. Thus, to reach the best accuracy score, the data will be fed to many models and the best model will be chosen for the final project. Then it will have an online application that people can use to find the prices of houses in Bishkek based on their needs. The application has a front page that will give information about the website’s purpose and about the data set I will be using during the research process.

It will also be linked to my GitHub where I will be publishing the data and the code just to make sure if someone wants to learn more about the background work and the way the code is working. It will show all the details and what has been done to achieve the goal. Like people who want to find a house in Bishkek will just search for house by putting in some feature into the app like no of rooms and the area of the house and the number of floors that want the house to be in.

*Aims and Objectives:*

* **Aims:**
  + Analyze, and predict, the house prices in Bishkek using Machine Learning techniques.
* **Objectives:**
  + Develop a predictive model for finding prices of houses in Bishkek.
  + Build a web platform where users can input their requirements and find a house / apartment easily along with a predictive price of the house. They will be able to see the reason behind how high or low the price of the house will be.
  + Any user can use it to search for the specific kind of house they want.
  + Advanced filtering method for house searching.
  + Search Algorithm (Search bar).
  + Requirement input and output (based on the independent variables)

*Understanding the need and the study Area:*

The changes in real estate have left many stakeholders and many buyers confused and seeing the reasons behind these fluctuations not only serves as a potential question to be understood by anyone but also a very socially significant thing. This project seeks to fill the gap between the potential buyers, the owners, renters, and investors. The biggest purpose behind this project is fair transactions, making information about the trends that govern the real estate market. This information can be of a big benefit for many developmental initiatives, infrastructural enhancements, and policies that are related to housing. Eventually I will bring stability into the growth in the real estate business. Besides that, it will become a very powerful tool for many investors, sellers, buyers and renters. It will help them in many ways, like making better decisions and navigating the complexities and future prediction regarding how the real estate will mold. For people who are seeking houses to buy or rent, this project will help them do the pre assessment, according to their needs.

Furthermore, the project has the potential to improve the stability in businesses related to real estate in Bishkek, by fostering right and transparent transactions among people. Most importantly, this project aims to assist people searching for properties that fall in their preferences and requirements.

# *Business benefits of the project:*

The "Predictive Modeling of House Prices in Bishkek" project will offer a number of benefits in the domain of Real Estate. Here are some vital advantages of the project in field of business:

* *Informed Decision Making for Buyers and Sellers*

Transparency in marketing: The project will provide a clear idea of house prices in Bishkek. That will eventually provide an understanding of housing market trends enabling the buyers and sellers to make smart decisions based on accurate price predictions.

Fairness in Prices: the estimation of the prices of the houses will be data driven yet helping buyers to ensure the fair prices and at the same time assisting the sellers in setting competitive prices that ca benefit them.

* *Benefits for Real Estate Agents and Brokers*

Enhancement of the Property Value: the agents and the broker can use the models to enhance and give accurate value of the property, which will enhance the quality of work they do. Besides that understanding of real estate trends will allow the agents to improve their offers to fit the requirements and the budget of the clients.

* *Benefits for the Investors*

Better Investment Decisions: This project will give an insight to the real estate market highlighting the tends in the market for several years, thus it will help them to make better decision and minimize any kind of loss in their investment, this will give them a broader picture if where and when to invest and help them to mitigate the risk of losing money.

* *A new market Insights for Developers*

Demand Analysis: Analyzing the data and understanding the demand of types of properties will guide the developers to focus on what is trending in the market.

* *Urban Planning and Policy Making*

Improvement in infrastructure: The insights from the project will give an idea of what factors are important for a house in terms of business, thus leading to improvement in infrastructure.

* *Rental Market Enhancement*

Assessment of Rental Price: Landlords will use the predictions to ensure the fair prices.

* *Economic Impact*

Local Economy: A stable, fair and transparent real estate market will attract international market leading to more.

# *Literature Review:*

The real estate market in Bishkek has a very big impact on the development of the Kyrgyzstan in the recent years. With the passage of time people are focusing more on understanding and recognizing the factors the amin factors that are responsible for the property prices in Bishkek. The literature survey will dive deep into a number of studies and research that already exists on this topic. It will discuss several sources including articles, research papers and books as a reference for this project. Yet this research-based project will give insights into the recent challenges, and opportunities in the real estate world of Bishkek.

A very comprehensive study has been done by Gray Roseman under the topic, “The Residential Real Estate Market in Bishkek, Kyrgyzstan: Current Conditions and Prospects” in 2004 (Roseman, 2004). This comprehensive study reviews the situation of the prices and has clearly highlighted the main features of any house that affects the prices, and it also talks about. The article highlights the increase in the prices of houses in the residential real estate market of Bishkek. There are many factors that might seem like are not affecting the prices of the houses in Bishkek but that could not be right. So, I found a research study that has been done in Bishkek that talks about what factors affect the real estate in the city and it dives deeper into the dynamic nature of real estate in Bishkek. This research article was published in Burgas Gazette, presents the recent perspectives of people on the Bishkek real estate market, which is titled as “Real Estate Market in BISHKEK, Kyrgyzstan BISHKEK, Kyrgyzstan Real Estate Prices”. This article has covered some major events that happened recently which have affected the prices of houses in Bishkek, and one of those important events is COVID-19. Besides that, it explains the reasons why anyone must or must not invest in the real estate of Bishkek and also explains the benefits and the drawbacks of the purchase people make regarding real estate. This paper also mentions some of the future expectation and insights that are more expected to happen soon (Gazette, 2023).

Much research has been done on the fact that owning a house has a great impact on the policies that apply for housing all around the world. However, we have some people who disagree with this and believe that the property management systems are framed in contemporary privatization policies which belong to the post-socialist cities. Considering the case study in, the article” Globalizing homeownership: Housing privatization schemes and the private rental sector in post-socialist Bishkek, Kyrgyzstan” (Hatcher, 2015) ,examines the limits to which many communities and the governmental itself would prefer to remain focused on homeownership in order to be able to make the developing policies and laws on rental houses and other housing tenures. The research talks about two main points which are very closely related to contemporary housing programs. The programs include the association for homeownership, and the role of government in the settlement on the city’s outskirts. It brought many changes in the real estate markets and one of them was that the tenants remained invisible, or their rights were merely dependent on the owner and were subjected to the owner’s interest. These all events have changed the real estate markets all around the world, which is still a bit complicated for people to understand and figure it out. Moreover, where the rights of the tenets are legally recognized by the government, a big gap still will remain between representations of the properties and their reality which in turn will highlight the post-socialist property right.

Another article” Making Property in Kyrgyzstan: Bishkek’s Urban Periphery as Redistributive Space”, explores the claims that have been made by the people who live in Bishkek in informal settlement, which are locates in the edge of Bishkek city (Hatcher, Making Property in Kyrgyzstan: Bishkek’s Urban, 2017). It talks about how the urban periphery has been growing, along with shifting the property rights that have been introduced by the privatization programs. Bishkek’s Novostroikas are the newly constructed buildings in Bishkek which are built for solving the grassroot inequitable distribution of wealth and private properties in Bishkek. The unrest caused by the Tulip revolution in 2005, and the chaotic events followed by the event in 2010 are considered as decisive moments that have resulted in the unequal distribution of property. Then the article goes a bit deeper into how all the residents of the Novostroikas compose the moral claims over the property that represents the understanding of properties to be clearer and more contextual rather than being predetermined.

**Table 1**: Notations and explanations

|  |  |
| --- | --- |
| **Notations** | **Explanations** |
| Key Words | explain |

# *Similar applications comparison table:*

There are many applications online that are used to find apartments in Bishkek. Some of them give all the information about the houses in Bishkek and some really do not serve any purpose. For example, Lalafo is an online application where people can go and buy any kind of thing possible, they provide information about apartments and hotels and hostels a well, yet their data is very old and the apartment pictures they have provided are from years ago and even if you like the apartment and go find it, it will be totally different and in bad shape. Thus, many sources are not providing the right information about houses in Bishkek. Secondly, there is a big problem that people face when renting an apartment which works through third people, which is real estate agent. Most of the agents are foreigners, because they can communicate with the foreigner and they help the owner find people who can rent the house or apartment, thus the owners find the agents and they do not make sure how the buyer or the rental people really feel about the apartment, yet for the service the agents take big amount of money regardless of the fact either the person buys the apartment or not. Which is very inconvenient for many people in Bishkek. I am foreigner and I have myself faced many such situations and I know how it feels to be in a different country and facing such a stressful situation. I researched each of the similar apps that I have mentioned below and some of the problems they all share are following.

* The data they use is not updated and can be very old and does not give the right and true information to people.
* The prices they show on these platforms are not based on what is going on in the real estate market, but rather only depend on the owner’s perceptions and desires.

# *Table 1.1: Similar applications comparison table*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Applications  Features | Airbnb | Facebook Marketplace | Lalafo | Real Estate Agents | 4321 Property | Booking.com | HousePrices | Total |
|  | Checkbox Checked outline | Checkbox Checked outline | Checkbox Checked outline | Checkbox Checked outline | Checkbox Checked outline | Checkbox Checked outline | Checkbox Checked outline | 5 |
| Recent Picture of Buildings available |  |  |  |  |  |  | Checkbox Checked outline | 1 |
| Authentic resource of data | Checkbox Checked outline | Checkbox Checked outline |  | Checkbox Checked outline |  |  | Checkbox Checked outline | 4 |
| Provide the prices of building according to market |  |  |  |  |  |  | Checkbox Checked outline | 1 |
| data privacy | Checkbox Checked outline |  |  |  |  |  | Checkbox Checked outline | 2 |
| security | Checkbox Checked outline | Checkbox Checked outline |  |  |  |  | Checkbox Checked outline | 3 |
| It will predict the prices of the houses |  |  |  |  |  |  | Checkbox Checked outline | 1 |

# *Technical specification of the project:*

**Note**: Write a clear technical specification. Specify all functional and non-functional requirements of your project. For all requirements provide a description and indicate acceptance criteria.

## *Table 1.2: Functional requirements*

|  |  |
| --- | --- |
| Functional requirement No. | Functional Requirement |
| Data preprocessing | Data must be cleaned and preprocessed using advanced features in machine learning |
| Machine Learning Model | The machine learning model will be built by feeding the data on a model |
| Online Application | Deploying the application online |
| Input and output in application | The user will be asked to choose the input feature and then the model they will prefer to use will predict the prices of the house based on the feature they have choosen |

* *Data preprocessing:*

Data preprocessing is one of the most important parts of any project and the data must be improved before it is fed to the machine learning model, yet the results depend on how clean the data is. The criteria that I will use to handle this part is that I will be using feature engineering to get rid of the missing values, yet I will be creating new relevant features, and I will try to decrease the directionality of my data if possible.

* *Machine Learning Model:*

Predictive modeling will be done in the project where the preprocessed data will be fed to the machine learning model and then a model will be built which will have high accuracy to predict the prices for houses in Bishkek as accurately as possible. To make sure of this part of the project the data will be checked before using it on a model that it is ready to be used and then several machine learning algorithms and model must be used to learn about the data and which model predict accurately. This will produce better results and will make sure that the purpose of the project is achieved.

* *Model deployment (online application):*

Deployment is another important aspect of my project because the reason behind the project is to make the right information accessible to people in Bishkek and the purpose will be served only when I deploy the project online. To make this happen it will be assured that the project is ready to deploy using different libraries in python, yet python is the major programming language being used in the project.

* *Input and output of application:*

The user will be asked to choose some inputs features, like no of rooms in the house, the areas of the house, the number of floors in the building and some ore features. Then the user will also be asked to choose the model that they want to use to predict the price on the basis of the accuracy of the model. The model will then predict the prices on the basis of the features that have been chosen.

## *Table 1.3: Non-functional requirements:*

|  |  |
| --- | --- |
| Non-functional requirement No. | Non-Functional Requirement Description |
| Performance | How quick the application responds to the users’ request |
| Scalability | The application must be quick to respond to many users at the same time |
| Reliability and accessibility | Data will be updated every once in a while, and the performance of the project, eventually the application |
| User tests | tackled the any sort of problem the users might have by doing pre deployment tests |

## *Acceptance criteria for non-functional requirements*

* *Performance:*

The performance of the application depends on May things such as hardware, software, quality of code, database performance, latency of the network, caching, user interface and experience of user, security measures and many other factors. To make sure that the application does not take more the 3 to 5 seconds or exceed this threshold, every step from the starting of the project till the end must be done with deep care and clarity. It also depends on how better the model predicts and how quickly it predicts and for that the best possible model must be used in the backend part of the project.

* *Scalability:*

The scale of any application is crucial because this will define how satisfied the users are with the application and how useful is the application. The application must be quick to respond to many users at the same time, even if the substantial number of users increases the application should not shut down or stop responding to the user.

* *Reliability and accessibility:*

Reliability of the data and the information that the website provides is the main purpose of the project and to make sure of the data will be updated occasionally, and the performance of the project, eventually the application, must be under review constantly. It must be made sure that the downtime of the website is minimal, and it complies with most of the accessibility requirements.

* *User tests:*

Before deploying the application, one must be very sure that they have done the necessary test beforehand and have tackled the any sort of problem the users might have. So, making sure that the application goes thoroughly through the testing process will ensure the quality of information and accessibility and reliability of the application.

* *Data Security:*

This is one of the concerns one must make sure of when working on a project that serves its purpose only when the users interact with the application and use it for their purposes. Thus, at the same time it raises the issue of security of the data. This can be tackled by collecting a wide range of data for the owners of the houses in Bishkek which can expand the project where we can easily bring the buyer and the owners in touch with just a click, the data might contain their consent, emails, and phone numbers. Also, can be tackled by keeping Software and Systems updated.

Here is the detailed budget description that I need for the completion of my project.

# *Table 1.4: Project Budget Estimation*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Category* | *Description* | *Quantity* | *Unit price* | *Total cost* | *% Of Budget* |
| Deployment | I will deploy my application and I want it to be in market before I defend my FYP | 10 months | $ 80 | $ 80.0 | 26.66% |
| Plagiarism report | Two times document check payment with plagiarism checker | 2 times | $ 35 | $ 70 | 23.33% |
| Buying the Domain Name | I will have to buy a domain name for my application | 1 time | $ 30 | $ 30 | 10 % |
| courses | I will take some courses for Streamlit | Not sure | $80 | $100 | 33.33% |
| Miscellaneous | Some unforeseen expenses I might not be aware of | Not sure | $20 | $20 | 6.6% |
| Total Budget Cost Estimate | | | | | $300 |

# *Project plan and schedule:*

**Note**: Provide the project schedule using the Gannt Chart and comment where applicable. Identify milestones. A milestone is a concrete event that one can use to demonstrate progress. Milestones should be clear, concrete, demonstrable achievements (“SMART”).

# *Milestones:*

## *Table 1.4: Milestones of my project*

|  |  |  |  |
| --- | --- | --- | --- |
| S/N | Week | Project | Objectives |
| **Phase 1: Project Initiation (2-4 weeks)** | | |
| 1 | Week 1-2 | Project Planning | Define project scope and objectives. Identify stakeholders and their roles. Create a detailed project plan, including tasks, responsibilities, and timelines. |
| 2 | Week 3-4 | Data Collection and Preprocessing | Conduct exploratory data analysis (EDA) to understand data characteristics. |
| **Phase 2: Data Preparation (4-6 weeks)** | | |
| 3 | Week 5-6 | Data Preparation | Handle missing data, outliers, and duplicates.  Perform feature engineering to extract meaningful variables.  Encode categorical variables and normalize data. |
|  |  | Model Development | Select machine learning algorithms for house price prediction.  Split data into training, validation, and test sets.  Train and evaluate initial models. |
| **Phase 4: Model Optimization and Testing (6-8 weeks)** | | |
| 4 | Week 6-10 | Model Optimization | Fine-tune machine learning models.  Address bias and fairness issues.  Optimize model performance and accuracy |
| 5 | Week 10-15 | Testing and Validation | Conduct extensive testing, including unit, integration, and user acceptance testing. Validate model predictions against real-world data.  Address any identified issues and bugs. |
| **Phase 5: Deployment and Launch (2-4 weeks)** | | |
| 6 | Week 15-20 | Development of the website | Set up deployment infrastructure  Deploy both backend and frontend components. Configure domain and hosting |
| 7 | Week 20-30 | Launching and testing with users | doing the set up and making sure that the website works |
| 8 | Week 30-35 | making changes | getting feedback from users and making changes accordingly. |
| 9 | Week 35-40 | more additions | Making the interface more attractive and user friendly. |

# *Work Breakdown of project*

A diagram of a project planning

Description automatically generated*Figure 1.1: Work breakdown*

# *A timeline with colorful text Description automatically generated with medium confidenceGantt Chart:*

## *Figure 1.2: Gannt Chart*

# *Risk management plan:*

**Note**: Identify your risks and solutions to them. The quality of the plan will impact your midterm exam grade.

## Table 1.5: Technical Risks Assessment Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risks | | Likelihood | Consequence | Overall Risk | Risk Level |
| (1-3) | (1-5) | (1-15) |
| 1 | Dashboard development and deployment on time | 2 | 5 | 10 | Significant |
| 2 | Budget distribution | 2 | 1 | 5 | Minor |
| 3 | Consent from owners of houses | 5 | 5 | 10 | Significant |
| 4 | Operational | 4 | 5 | 10 | Significant |
| 5 | Other resources | 2 | 3 | 5 | Minor |

* *Technical ricks of my project:*
* *Dashboard*: this is something I have never done before, making a dashboard and deploying it takes a lot of time and effort.

*Solution*: So, I am looking forward to just learning Streamlit.Streamlit is an open-source Python library that is used for creating web applications quickly and easily. So, I am looking forward to getting the budget and starting to take a course for it and I will try to make sure that at least one month from the final presentation my website will be ready.

* *Budget distribution:* distribution of budget may cause some problems, like running out of budget before the project is complete.

*Solution*: Identifying and prioritizing the most important aspect of the project help me not this problem.

* *Consent from owners of houses*: This is important for ethical and legal reasons, as I am aware that using personal data without consent form people can lead to serious privacy violations problems.

*Solutions*: Developing clear and transparent communication for the purpose of the project explaining how the data will be used, and the benefits to the house owners and the wider community. Designing a straightforward and legally compliant consent form to make sure the people whose data has been already informed.

* *Operational:* I might have some problems with the users who will interact with the dashboard, like maybe the models will not perform well or the server will be down because of some internal problems or external risks like hackers.

*Solution*: I will have to make sure that I check the website often and update it often to protect it from facing any external risks or any internal malfunctions. Besides that the I will have to do a comprehensive test so that no issues will arise later.

## *Table 1.6: Non-Technical Risks Assessment Table*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risks | | Likelihood | Consequence | Overall Risk | Risk Level |
| (1-3) | (1-5) | (1-15) |
| 1 | Privacy | 2 | 5 | 7 | Major |
| 2 | Security (Hackers) | 1 | 2 | 3 | Minor |

* *Non-Technical ricks of my project:*

One of the ethical issues the project may contain is providing information about someone’s property without their permission.

*Solution*: (This can be tackled by collecting a wide range of data for the owners of the houses in Bishkek which can expand the project where we can easily bring the buyer and the owners in touch with just a click, the data might contain their consent, emails, and phone numbers)

Another issue is when hackers attempt to get user information by attacking my platform using different techniques.

*Solution*: (This can be tackled by keeping Software and Systems updated

## *Data Base diagrams*

**D**atabase diagram is a visual representation of the database of structure of the application. It provide the blueprint for the design for the understanding of the database and is an very important for representation of any project.

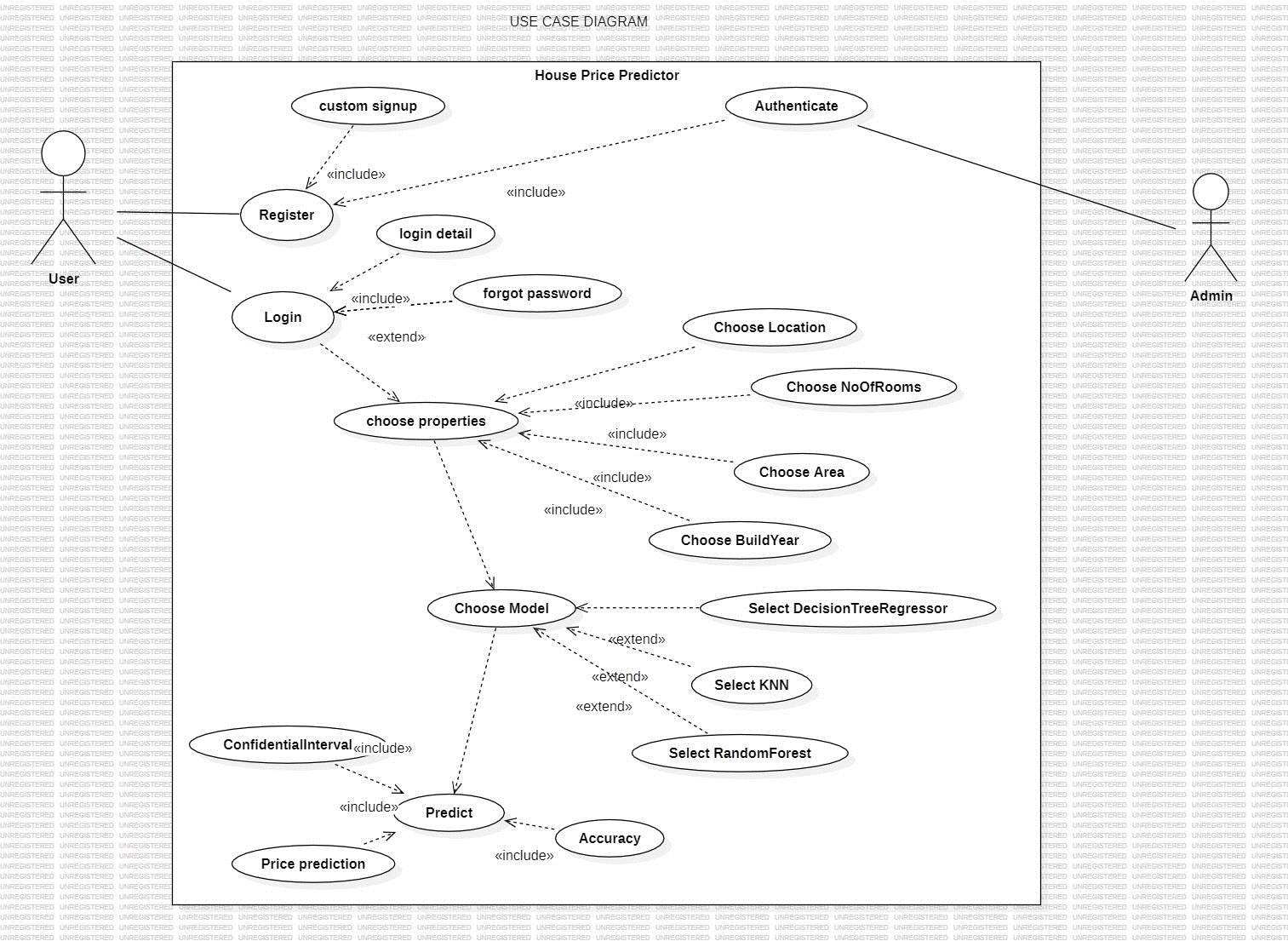
A diagram of a computer

Description automatically generated

## *Figure 1.3. database models (Class Diagram) in UML format*

## *Class diagrams*

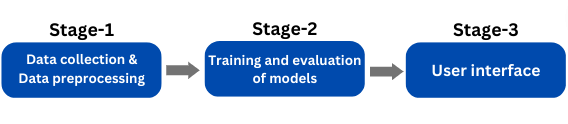
A class diagram is a type of static diagram in software engineering, that can be used to explain the structure of any system by using classes, attributes and operation(methods) and the relation among the classes.



## *Figure 1.3. Class diagram in UML format*

# **CHAPTER 2: PROPOSED METHODOLOGY**

This whole project has been through three main stages that are data collection, the training of the data and finally making interface. The figure below shows the stages of the project.



I got the data “House prices in Bishkek”, in the ML /AL boot camp and that was when I realized it was interesting for me, yet I started working on my Final Year Project. Later I did some research and found the data on Kaggle, under the name “House Price Bishkek” on Kaggle.

# *Bibliography*

Alymbaev, M. (2022). New Settlements in Bishkek: Law, Urban Space, Culture and Socio-Economic Development.

BOSTERIEVA, S. (2014). HOUSING ARCHITECTURE IN SOVIET AND POST-SOVIET CAPITAL OF KYRGYZSTAN: FROM FRUNZE TO BISHKEK .

Cartledge, D. M. (2014). Sustainable Cities in Post-Soviet Central Asia: The Case of Bishkek.

Cestari, A. (2012). Kyrgyz Republic - Bishkek and Osh Urban Infrastructure Project : P104994 - Implementation Status Results Report : Sequence 07.

Craig, H. (n.d.). Globalising homeownership: housing privatisation schemes and the private rental sector in post-socialist Bishkek, Kyrgyzstan. *2015*.

Dzhumataeva, A. (2019). Macroeconomic Determinants of Housing Price. Case of the Eurasian Economic Union (EAEU) Countries (1998-2018).

Erdener Kaynak, A. K. (2022). An empirical investigation of home buying behavior in a high-context culture: a strategic marketing-oriented approach.

Erkin Isaev, B. A.-u. (n.d.). Impact of Climate Change and Air Pollution Forecasting Using Machine Learning Techniques in Bishkek. *2022*.

Erokhin. (2020). Comparative Analysis of Digital Development in Central Asian Countries.

Hatcher, C. (2015). Illegal geographies of the state: the legalisation of a “squatter” settlement in Bishkek, Kyrgyzstan.

Isabaeva, E. (2013). Migration into the "Illegality" and Coping with Difficulties in a Squatter Settlement in Bishkek.

Isabaeva, E. (2021). Transcending Illegality in Kyrgyzstan: The Case of a Squatter Settlement in Bishkek.

Kamchybekov, M. P. (2008). Probabilistic seismic risk of the territory of Bishkek city, Kyrgyzstan.

Kosmarskaya, M. F. (2018). Exploring “North” and “South” in post-Soviet Bishkek: Discourses and perceptions of rural-urban migration.

Kostyukova, I. (2007). The towns of Kyrgyzstan change their faces: Rural‐urban migrants in Bishkek.

Real Estate Market in BISHKEK, Kyrgyzstan BISHKEK, Kyrgyzstan Real Estate Prices. (2023).

ROSEMAN, G. (2004). The Residential Real Estate Market in Bishkek, Kyrgyzstan: Current Conditions and Prospects.

Saidakhmtov, A. (n.d.). Determinants of Apartment Prices in Bishkek. *2017*.

Schröder, P. (2012). Urbanizing' Bishkek.

Schröder, P. (2016). Avoidance and appropriation in Bishkek: dealing with time, space and urbanity in Kyrgyzstan's capital.

Schröder, P. (2017). Neighbourhood Youth and Urban Change in Kyrgyzstan’s Capital.

Tiulegenov, Y. S. (2008). KYRGYZSTAN TODAY.