****

**DATA WAREHOUSING AND DATA MINING**

**REAL ESTATE PRICE PREDICTION**

**REPORT – 2**

By

Arooba Siddiqi               321618

Emaan Bashir          296190

Rimsha Mirza 285718

Faraz Shah 303087

**CONTENTS**

[**ABSTRACT:** 3](#_Toc123332075)

[**INTRODUCTION:** 3](#_Toc123332076)

[**DATA MODEL:** 4](#_Toc123332077)

[**SOFTWARE MODEL:** 4](#_Toc123332078)

[**TIMELINE AND WORK BREAKDOWN:** 5](#_Toc123332079)

[**EXPLANATION:** 6](#_Toc123332080)

# **ABSTRACT**

Real estate price prediction in Pakistan involves forecasting the future prices of properties in the Pakistani market. This can be useful for individuals and businesses looking to buy or sell properties, as well as for investors looking to make informed decisions about their real estate investments. The data model for the real estate price prediction will include real estate properties, transactional data, and geographic data, which will be collected from a commonly used website for real estate transactions in Pakistan through web scraping. Data preprocessing will involve cleaning, formatting, and organizing the data in a way that meets needs of the project. A dashboard will be created for data visualization and a price prediction model will use data mining techniques to identify patterns and relationships in the data. The model will be based on regression techniques and applied to the dataset to generate accurate price predictions about the real estate market in Pakistan.

# **INTRODUCTION**

Real estate price prediction in Pakistan refers to the process of forecasting the future prices of real estate properties in the Pakistani market. This can be useful for individuals and businesses looking to buy or sell properties, as well as for investors looking to make informed decisions about their real estate investments. There are several approaches that can be used to predict real estate prices in Pakistan. These may include analyzing historical price trends, using economic indicators such as inflation and employment rates, and applying data mining techniques to identify patterns and relationships in the data. Overall, real estate price prediction in Pakistan can be a challenging task, but it can also provide valuable insights and help individuals and businesses make informed decisions about the real estate market.

# **DATA MODEL**

The data model for the real estate price prediction will be a combination of real estate properties, transactional data, and geographic data. The real estate properties refer to the property type (such as house, residential plot and building etc.) and size of the property, historical data includes the age of the property, property types, area of the property, sales price, and rental price of real estate and the geographic data refers to the location of the property. The data will be collected from the website most used for real estate in Pakistan. This data model will be used as the foundation for a real estate price prediction solution in Pakistan. It will be used to collect and organize the relevant data needed for the solution, and to identify the relationships between different variables and real estate prices. This data will then be used in conjunction with machine learning algorithms to make accurate price predictions.

# **SOFTWARE MODEL**

The software model for the real estate prediction will be a combination of a dashboard and price prediction model. The dashboard will be used for real-time data visualization to analyze the data, identify patterns and trends that can be used to generate accurate price predictions. The price prediction model will be based on a machine learning algorithm to accurately predict the prices based on the data-driven insights found using data visualization.

# **TIMELINE AND WORK BREAKDOWN:**

|  |  |  |  |
| --- | --- | --- | --- |
| **PHASE** | **TASK NAME** | **TASK DESCRIPTION** | **DURATION** |
| Phase - 1 | Data collection | Data will be collected from a website of real estate in Pakistan using web scrapping | 2 weeks |
|  | Data Preprocessing | Data will be preprocessed. This will be done by handling the null and zero values, convert price to its numerical form and remove outliers. | 1 week |
| Phase – 2 | Data Visualization | Dashboard will be created for the graphical representation of data. This will be done using data visualization tool such as tableau and Power Bi. | 1 week |
| Phase - 3 | Data Mining | Prediction models will be applied for the accurate price prediction. The model will be trained and tested on the preprocessed dataset. | 1 week |

## **EXPLANATION:Top of Form**

**PHASE – 1:**

**DATA COLLECTION:**

The first phase of the project includes data collection and data preprocessing. The data will be collected from a commonly used website of Pakistan for real estate transactions. It will be done through web scraping that is extracting data from a website. The first step of this process will be to identify the specific data needed that is property type, location, price, purpose (property for rent and sale) and area of the property. A web scraper will be designed to access the website, navigate to the relevant pages, and extract the desired data.

**DATA PREPROCESSING:**

The second step in the first phase is data preprocessing. In this step, the data will be cleaned, formatted, and organized in a way that is suitable to meet the needs of project. Some common steps may include identifying and addressing any null or zero values in the data, converting property types to a consistent format and price data into its numerical form. Moreover, outliers will be removed from the data.

**PHASE-2:**

**DATA VISUALIZATION:**

For the graphical representation of data, a dashboard will be created. The dashboard will include multiple graphs to understand and analyze the data. This will help to identify trends, patterns, and underlying data-driven insights in the data. For data visualization, a tool will be used such as Tableau and Power Bi.

**PHASE – 3:**

**DATA MINING:**

A price prediction model will use data mining techniques to identify the impact of factors such as location, property type and size on the prices of real estate properties. This will help to generate more accurate price predictions about the real estate market. For the price prediction model, regression models will be explored and relevant models will be chosen and applied on the dataset.

Top of Form

Bottom of Form