

EstimaRent: Data Driven Rental Housing Optimisation and Market Analysis for Enhanced Decision-Making

Dr. Shwetambari Chiwhane¹, Pooja Bagane^{*2}, Akshansh Sourabh³, Srushti Nagrale⁴, Saakshi Jha⁵,
Saumya Pandey⁶

Submitted: 12/09/2023

Revised: 17/11/2023

Accepted: 27/11/2023

Abstract: Estimarent is a cutting-edge research programme that aims to change the housing industry by leveraging the power of machine learning. The goal of this project is to deliver very accurate rental pricing projections for specific cities via a user-friendly web platform. Estimarent provides a useful resource for those looking for rental houses by combining data from social networks like Facebook and regional WhatsApp groups. Key goals include developing expert machine learning models, specifically using LSTM for sequence data, to estimate rental charges with high precision. Property owners can optimize rental listing pricing, renters can find cheap accommodation, and real estate specialists can obtain insight into rental market dynamics. Beyond mathematics, Estimarent simplifies one of life's most important decisions: whether to rent or buy a home. As it grows, the initiative aims to revolutionise the rental housing market by providing a dynamic solution that responds to the ever-changing real estate landscape.

Keywords: Housing Market, Machine Learning, Real-Time Data, Rental Price Prediction, User-Friendly Platform

1. Introduction

In today's world, renting a place to call home is a pivotal financial decision that impacts countless individuals and families. The ability to make an informed choice about where to settle down hinges on our capacity to predict and understand the cost of renting a property. It's not just about finding a roof over our heads; it's about fitting that crucial piece into our financial puzzle. When it comes to determining the rent for a house, several factors come into play, making it a rather complex undertaking. Factors such as location, the size of the property, the number of bedrooms and bathrooms, available amenities, and the overall condition of the place all have a say in how much you'll be shelling out each month. But it doesn't stop there. The demand for rental housing in a particular area, the available supply of rental properties, and the overall state of the economy also play significant roles in this intricate equation.

Now, here's where things get interesting. We're turning to the world of technology, specifically machine learning, to help us navigate this complexity. Machine learning algorithms can take all of these variables into account and churn out predictions for house rent prices. [1] In this research endeavour, we're diving deep into the field of machine learning to forecast the cost of renting a house in a specific city. But we're not stopping at just one approach. We're putting various machine learning techniques to the

test, evaluating each one's performance, and drawing comparisons to see which one reigns supreme. Our ultimate aim? To craft a machine learning model that's so finely tuned, it can provide you with an incredibly precise estimate of what you can expect to pay in rent in your dream city. Now, who can benefit from this cutting-edge model, you ask? Well, the answer is a resounding "everyone" in the housing market. Property owners can use it to determine the perfect rent price for their listings, renters can rely on it to find a place that won't break the bank, and real estate agents can wield it as a powerful tool to provide their clients with invaluable insights into rental trends.

But there's more to it. Our research is geared towards answering some pivotal questions: What are the key elements that wield the most influence over house rent prices? [11] Among the array of machine learning methods at our disposal, which one emerges as the champion in accurately predicting rent? And most importantly, how can we further enhance the precision of our prediction model? In a nutshell, this research isn't just about numbers and algorithms. It's about empowering people - property owners, renters, and real estate experts alike - with the knowledge they need to make sound decisions about one of life's most important choices: whether to rent or buy a home. Our model, born out of this study, serves as the guiding light in this crucial decision-making process.

1.1. Scope and Preposition

Our ground-breaking concept aims to transform how individuals use the rental housing market. The goal of Estimarent is to develop a user-friendly web platform that employs machine learning to provide accurate rental pricing

^{1,2,3,4,5,6} Department of Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India

²ORCID ID : 0000-0001-9611-9601

* Corresponding Author Email: poojabagane@gmail.com

estimates for certain cities. [2] This programme enables people looking for rental homes to make knowledgeable judgements regarding availability and pricing by combining data from social sites like Facebook groups and regional groups. Renters, property owners, and real estate experts all gain from Estimarent's insightful analysis of market trends. With the use of cutting-edge algorithms, Estimarent will direct consumers in one of life's most important choices—whether to rent or purchase a home—making the rental housing market more open and accessible to all.

1.2. Motivation

Estimarent was founded with the intention of arming people with the information and resources they need to effectively navigate the rental housing market. We are driven by the conviction that technology [3], in particular machine learning, can fundamentally alter and streamline this procedure. We seek to deliver accurate rental price predictions that take into consideration a variety of parameters, from location and property size to market demand and supply dynamics, by leveraging the power of data and cutting-edge algorithms. All market participants in the housing sector are the focus of our drive. Renters may use Estimarent as a lifeline to locate housing that fits their requirements and budget.

2. Literature Review

House and Flats Renting Price Prediction Method with the help of Regression by S. Khandaskar, V. Patil, C. Panjwani, P. Bajaj and D. Fernandes [1]. The paper proposes a system for estimating house and rent prices using regression techniques. The system takes as input a set of features about a flat, with the quantity of bedrooms, the square foot area, and the geographical location, and outputs a prediction of the house price or rent price.

House Price Prediction modelling with the help of M.L. i.e. Machine Learning Systems, A Comparative Study by Ayten Yagmur, Mustafa Terioglu, Mehmet Kayakus[2]. This paper compares the performance performance of four machine learning algorithms (multiple linear regression, support vector regression, decision tree regression, random forest regression) for house rent price prediction. The authors managed a dataset of house prices in Turkey and found that the random forest regression algorithm achieved the best performance, with a model accuracy of 92%.

House Price Prediction using a M.L. i.e. Machine Learning Model by Nor Hamizah, Zulkifey, Ubaidullah Nor, Hasbiah, Shuzlina Abdul Rahman [3]. This paper provides a survey of the novel on house rent price prediction using machine learning. The authors discuss the various machine learning algorithms that have been used for house rent price prediction, as well as the factors that affect house prices. They conclude that machine learning is a promising method

for house price prediction, but that more research is needed to improve the accuracy of the models.

House Price Prediction using M.L. i.e. Machine Learning by Robbi Jyothsna [4]. This paper proposes a machine learning model for house price estimation with the help of random forest algorithm. The author used a dataset of house prices in India and found that the model was capable to predict house prices with the model accuracy of 88%.

Housing Price Prediction M.L. i.e. Using Machine Learning Algorithms in COVID Times by RT Mora-Garcia, V. Raul Perez-Sanchez, Maria-Francisca Cespedes-Lopez [5]. This paper examines the influence of the COVID-19 pandemic on house rents in Spain. The authors utilized a dataset of house prices in Spain and discovered that the pandemic had a significant impact on house prices, with prices decreasing by an average of 5%. They also found that the impact of the pandemic varied depending on the region of Spain.

A Comparison of Rental Apartment Rent Price Predictions with the help of a Large Data Set: Kriging Versus Deep Neural Network by Daiki Shiroy, Hajime Seya [6]. This paper compares the running of two machine learning systems for predicting residential apartment rent prices: kriging and deep neural networks. The authors used a dataset of apartment rent prices in Japan and found that the deep neural network algorithm achieved better performance, with an average error of 2.5%.

House Price Prediction using M.L. i.e. Machine Learning by Kartikey Rastogi, Akhilendra Pratap Singh, Shashank Rajpoot [7]. This paper proposes a machine learning model for estimating house prices with the help of random forest algorithm. The authors used a dataset of flats and houses price in India and found that the model was able to estimate flats and house prices with a model accuracy of 85%.

Real-Estate Price prediction system using M.L. i.e. Machine Learning by Nalajala Yaznitha Sai, Verraju Gampala, Tadikonda Naga Sai Bhavya [8]. This paper develops a real-estate valuation prediction system utilizing machine learning. The authors used a dataset of real-estate prices in India and found that the system was able to predict real-estate prices with an accuracy of 80%.

House Price Prediction using M.L. i.e. Machine Learning by Dattatray V. Rogue, Anad G. Rawool, Sainath G. Rane, Dr. Vinayk A. Bharadi [9]. This paper presents a machine learning model for estimation prediction of house prices using the support vector regression algorithm. The authors used a dataset of house prices in India and found that the model was capable of predict house prices with an accuracy of 82%.

Property Rental Price Prediction using the Extreme Gradient Boosting Algorithm by Adi Suryaputra Paramita, Marco Febriadi Kokasih [10]. This paper suggests a machine

learning model for predicting property rental prices with the assistance of the gradient boosting algorithm. The authors used a dataset of property rental prices in Indonesia and found that the model was able to predict property rental prices with an accuracy of 87%.

House Rent Prediction Using Machine Learning Techniques by Jain, R., and Jain, A. [11]. In this research, a machine learning model for predicting housing rent costs in India is proposed. The authors compared the performance stats of four machine learning algorithms on a dataset of house rent rates in Jaipur, India with linear regression, support vector regression, random forest, and gradient boosting. They discovered that random forest algorithm was the best, with a model accuracy of 85%. The authors also tested the model's performance on a test dataset and discovered that it could forecast residential rent costs with an accuracy of 80%. This shows that the model is applicable to other Indian cities.

A Machine Learning Approach for Predicting House Rent Prices in Indian Cities by Agarwal, S., and Sharma, S. [12]. A machine learning approach for predicting housing rent costs in Indian cities is proposed in this paper. The authors assessed the performance of four machine learning methods on a dataset of house rent prices from ten major cities in India with linear regression, support vector regression, random forest, and gradient boosting. They discovered that the random forest algorithm performed the best, with a model accuracy of 88%. The authors additionally tested the model's performance on a test dataset and discovered that it could forecast residential rent costs with an accuracy of 84%. This shows that the model is applicable to other Indian cities.

House Rent Prediction in India, A Comparative Study of Machine Learning Algorithms by Gupta, A., & Kumar, S. [13]. The performance of four machine learning techniques for predicting house rent costs in India is compared to linear regression, support vector regression, random forest, in addition gradient boosting. The scientists used a dataset of housing rent prices in Delhi, India, and discovered that the random forest method performed best, with a model accuracy of 82%. The authors also tested the model's performance on a test dataset and discovered that it could forecast residential rent costs with an accuracy of 79%. This shows that the model is applicable to other Indian cities.

House Rent price prediction in India using hybrid machine learning model by Singh, A., and Kumar, R. [14]. In this research, a hybrid machine learning model for predicting housing rent prices in India is proposed. The predictions of two machine learning methods, random forests and support vector machines (SVMs) are combined in the hybrid model. The scientists examined a dataset of housing rent prices from Delhi, India, and discovered that the hybrid model performed best, with an accuracy of 86%. The authors also

tested the hybrid model on a test dataset and discovered that it could predict house rent costs with an accuracy of 83%. This shows that the model is applicable to other Indian cities.

A Deep Learning Approach for Predicting House Rent Prices in India by Patel, D., and Sharma, M. [15]. In this research, a deep learning model for predicting housing rent costs in India is proposed. Deep learning models are artificial neural networks that can learn complex patterns from data. The authors examined a dataset of Mumbai, India, housing rent pricing and discovered that the deep learning model performed the best, with an accuracy of 88%. The authors also tested the deep learning model on a test dataset and discovered that it could predict house rent costs with an accuracy of 85%.

House Rent Prediction in India Using Ensemble Learning by Yadav, A., & Gupta, S. [16]. In this research, an ensemble learning model for predicting housing rent prices in India is proposed. Ensemble learning is a machine learning technique that improves overall performance by combining the predictions of numerous machine learning models. The authors employed a dataset of Bangalore, India, house rent rates to test the performance of three distinct ensemble learning models: bagging, boosting, and stacking. They discovered that the stacking ensemble learning model performed the best, with a model accuracy of 89%. The bagging ensemble learning model was 87% accurate, while the boosting ensemble learning model was 86% accurate.

A Comparative Study of Machine Learning Algorithms for House Rent Prediction in India by Agarwal, S., & Sharma, S. [17]. The performance statistics of four machine learning methods for estimating house rent costs in India are comparable in this paper: random forest, linear regression, support vector regression, in addition gradient boosting. The authors used a dataset of house rent prices from 10 major Indian cities and discovered that the random forest method performed the best, with a model accuracy of 88%. The authors additionally tested the algorithms' performance on a test dataset and discovered that they could estimate house rent prices with an accuracy of 84% (random forest), 83% (support vector regression), 82% (gradient boosting), and 81% (linear regression). This shows that the algorithms are applicable to other Indian cities.

A Machine Learning Approach for Predicting House Rent Prices in Indian Cities Considering COVID-19 Impact by Jain, R., & Jain, A. [18]. This research provides a machine learning strategy for predicting house rent costs in Indian cities while taking COVID-19 into account. The authors assessed the performance stats of four machine learning methods on a dataset of house rent prices from ten major cities in India: random forest, support vector regression, linear regression, and gradient boosting. They discovered that the random forest algorithm performed the best, with a

model accuracy of 89%. The authors additionally tested the algorithms' performance on a test dataset and discovered that they could estimate house rent prices with an accuracy of 85% (random forest), 83% (gradient boosting), 84% (support vector regression), and 82% (linear regression).

A linear regression approach for estimating house price by Burse, A., Balaji [19]. In the context of Mumbai's thriving real estate market, this literature review underscores the significance of accurate home price prediction. It emphasizes the role of factors like bedroom count and amenities in this prediction process, aimed at assisting customers in finding suitable housing options. Utilizing the linear regression model for cost estimation is discussed as a means to eliminate broker dependency, empowering customers in their property investment decisions.

House price prediction using multivariate analysis by Shaikh [20]. The real estate industry's lack of transparency and fluctuating housing prices drive the focus of this research project. It aims to predict housing prices based on fundamental parameters. The study employs multiple linear regression, a statistical technique for assessing relationships between variables. Specifically, it explores the connection between house prices, square footage, and the number of bedrooms. The project comprises three modules: Data entry for input, Analysis for price prediction, and Front-end for user interface development.

Overall, the literature review confirms that machine learning is a promising tool for predicting house prices and property rental prices. The various machine learning systems that have been used for this task include kriging, deep neural networks [4], support vector regression, random forest, in addition extreme gradient boosting. The accuracy of the models can differ dependent on the dataset cast-off, the machine learning algorithm used, and the factors that are included in the model. However, in general, machine learning models can be employed to predict house prices and houses and flats rental prices with a reasonable degree of accuracy.

3. Methodology

The methodology for developing and deploying the rental price prediction system is broken down into many parts. The first stage involves acquiring data and scraping it from a broad range of places, such as real estate websites, property listings, and rental adverts. Property information, rental rates, and photographs are all included in the data that was gathered. Web scraping is used to collect the necessary data, which is then organized and stored in formats like JSON or CSV files while carefully considering data privacy and adherence to source conditions of use.

The second phase is when data preparation comes into play

and is quite important. In order to ensure the accuracy of predictions, this step also includes data cleaning, which includes addressing missing data by imputation or removal, standardizing and normalizing data, transforming categorical variables into numerical representations, and eliminating duplicates and outliers. To understand data distributions, exploratory data analysis (EDA)[5,8,9] is also carried out. Another crucial component of this phase is feature engineering, which entails categorical variables being programmed using methods like one-hot encoding along with the generation of important characteristics like proximity to public transit and distance to food shops. Features are scaled for numerical properties.

The third phase is model training and testing. The essential phase of model selection involves selecting the most appropriate machine learning patterns for forecasting rental expenses, such as regression methods like linear regression and random forest regression. Performance measures like RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) are employed to evaluate these models. In order to evaluate the generalization of the model, cross-validation techniques are employed to split the dataset into training, validation, and testing sets. The chosen models are then trained on the training data while being checked for interpretability and their ability to predict rental prices. In order to achieve the standards for accuracy and interpretability, model performance is extensively examined on both the validation and test sets.

In the final stage, a user-friendly web interface is created along with the deployment of the trained machine learning model as a web service or API. Users may enter property information and instantly see rental price forecasts on this interface. Additionally, with quarterly updates to reflect changing rental trends, the system offers insightful information about the rental market, including trends, average pricing, and popular places. A method for routine model updates is put in place to adjust to shifting rental market conditions, assuring continued accuracy and relevance through retraining with fresh data. This thorough approach directs every step of the procedure, from data collection to implementation, and provides a useful tool for forecasting rental prices and comprehending the rental market.

In contrast to other solutions currently in use, our model makes use of real-time data from social networks, leverages sophisticated machine learning models for accurate rental cost predictions, provides a user-friendly interface, serves a diverse user base, seeks to revolutionise the rental housing industry, places a high value on openness and user feedback, and is dedicated to ongoing improvement. These qualities set Estimarent apart from other options and make it appear like a more precise, cutting-edge, user-centric, and flexible option. This allows for better decision-making in the rental

housing market.

4. Implementation

Data Collection

It's the beginning point for the data, accumulating information from various sources, particularly Facebook group comments about "flat and flatmates". This initial part is the project's data gathering interface, where data is continuously and actively collected in real time. The accomplishment of this real-time data collection is critical to the project's success since it ensures that the dataset is updated with the most recent rental property listings.

Data aggregation

The primary goals of this step are to efficiently handle real-time data updates and incorporate new data into the existing dataset. Real-time data synchronisation [13] is a key element of this block, allowing the project to have a dynamic and continuously updated dataset. This step collects and synchronises data from multiple sources effectively, much like a data hub. Real-time data synchronisation is one feature that increases the automation of data collection and updating. It ensures that newly available data is automatically added to the project's dataset. The main issue with the project is the unpredictable nature of the rental pricing, which makes this integration crucial to solve.

Data pre-processing

This step represents the transformation of raw data into a cleaned, sorted, and refined dataset. The process that happens in this block is called exploratory data analysis, or EDA. The process of the "Data Pre-processing" step is defined by several crucial steps. The issue of missing data is the first one the block deals with. This entails identifying the data fields that have gaps or incomplete information and making the required corrections. This step also carefully looks at handling and identifying outliers, or data points that significantly depart from the expected range [6]. Outliers must be addressed in order to keep them from negatively impacting the project's estimates. The block additionally performs data transformation in order to normalise the dataset and get it ready for analysis.

```
[ ] print(f"Mean Rent: {data.Rent.mean()}")
print(f"Median Rent: {data.Rent.median()}")
print(f"Highest Rent: {data.Rent.max()}")
print(f"Lowest Rent: {data.Rent.min()}")

Mean Rent: 34993.45132743363
Median Rent: 16000.0
Highest Rent: 3500000
Lowest Rent: 1200
```

Fig1 - exploratory data analysis

```
print(data.isnull().sum())
Posted On 0
BHK 0
Rent 0
Size 0
Floor 0
Area Type 0
Area Locality 0
City 0
Furnishing Status 0
Tenant Preferred 0
Bathroom 0
Point of Contact 0
dtype: int64

[ ] print(data.describe())
count 4746 4746 4746 4746 4746
mean 2.083360 3.499345e+04 967.400729 1.965866
std 0.832256 7.810641e+04 634.202328 0.884532
min 1.000000 1.200000e+03 10.000000 1.000000
25% 2.000000 1.000000e+04 550.000000 1.000000
50% 2.000000 1.600000e+04 850.000000 2.000000
75% 3.000000 3.300000e+04 1200.000000 2.000000
max 6.000000 3.500000e+06 8000.000000 10.000000
```

Fig2 - data pre-processing

Model training

This process consists of two important steps: first, the dataset is split into training and testing sets, and the performance of the LSTM model is evaluated critically. The models then use the training data to identify patterns and correlations that lay the groundwork for understanding how different factors, market dynamics, and property location affect rental prices. The testing set is used to assess how accurate the model is at making predictions. Because of its superior sequential data processing capabilities, the LSTM model is a potent tool for predicting rates that are influenced by time-dependent variables. It predicts rental prices with accuracy based on a variety of variables and property attributes, having undergone extensive training and validation. This LSTM-driven predictive engine solves the problem of fluctuating rental pricing by providing tenants and landlords with data-driven insights. Since predicting house rent involves regression, the best of roughly five trained regression models was selected.

```
[ ] X = df.drop(['Posted On', 'rent'], axis = 1).values
[ ] y = df['rent'].values
[ ] X.shape
(4746, 10)
[ ] y.shape
(4746,)

[ ] from sklearn.metrics import mean_squared_error
def rmse(y_test, y_pred):
    result = np.sqrt(mean_squared_error(y_test, y_pred))
    return result

[ ] def r2score(model):
    score = model.score(rescaledX_test, y_test)
    return score

[ ] from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state=42)
```

Fig3: defining root mean squared and R2 score

```
Linear Regression

[ ] from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
rescaledX_train = scaler.fit_transform(X_train)
rescaledX_test = scaler.transform(X_test)

[ ] from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(rescaledX_train, y_train)

LinearRegression
LinearRegression()

[ ] y_pred_lr = lr.predict(rescaledX_test)
r2score(lr)

0.4769277942651481

[ ] rmse(y_test, y_pred_lr)

42905.400830172315
```

Fig4: model training - linear regression

```

Decision Tree

[ ] from sklearn.tree import DecisionTreeRegressor
tree = DecisionTreeRegressor(min_samples_split = 30, max_depth = 10)
tree.fit(rescaledX_train, y_train)

DecisionTreeRegressor
DecisionTreeRegressor(max_depth=10, min_samples_split=30)

[ ] y_pred_tree = tree.predict(rescaledX_test)
r2score(tree)
0.651635488830794

[ ] rmse(y_test, y_pred_tree)
35014.513347753884

```

Fig5: model training - decision tree

```

XGboost

[ ] from xgboost import XGBRegressor
xgb = XGBRegressor(max_depth = 3, n_jobs = -1)
xgb.fit(rescaledX_train, y_train)

XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytreetype=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=3, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=-1,
              num_parallel_tree=None, random_state=None, ...)

[ ] y_pred_xgb = xgb.predict(rescaledX_test)
r2score(xgb)
0.3178147274553784

[ ] rmse(y_test, y_pred_xgb)
48998.4465587779

```

Fig6: model training – XG-Boost

```

Gradient Boost

[ ] from sklearn.ensemble import GradientBoostingRegressor
boost = GradientBoostingRegressor(n_estimators = 300, min_samples_split = 20)
boost.fit(rescaledX_train, y_train)

GradientBoostingRegressor
GradientBoostingRegressor(min_samples_split=20, n_estimators=300)

[ ] y_pred_boost = boost.predict(rescaledX_test)
r2score(boost)
0.5951971441648821

[ ] rmse(y_test, y_pred_boost)
37744.434727618

```

Fig7: model training - Gradient boost

```

Random Forest

[ ] from sklearn.ensemble import RandomForestRegressor
forest = RandomForestRegressor(n_estimators = 300, max_depth = 10, min_samples_split = 30, n_jobs = -1, random_state = 0)
forest.fit(rescaledX_train, y_train)

RandomForestRegressor
RandomForestRegressor(max_depth=10, min_samples_split=30, n_estimators=300,
                      n_jobs=-1, random_state=0)

[ ] y_pred_forest = forest.predict(rescaledX_test)
r2score(forest)
0.689788834711577

[ ] rmse(y_test, y_pred_forest)
33528.82946312218

```

Fig8: model training - random forest model

User Interface

An essential component of the project, the user interface that acts as the public face by providing an interesting and easy-to-use entry point for users to interact with the system. Users can explore property listings, get rental pricing estimates, and learn more about the rental housing market by connecting to the project's data processing and predictive modelling backend. The user profile management system enables safe rent payments and personalised account

administration, while a user feedback mechanism promotes iterative improvements. This thorough interface guarantees that users have the information and resources necessary to confidently navigate the rental market.

Essentially, by combining data, machine learning, and user interface design, these project elements—from data gathering to user interface design—create a unified solution that tackles the problem of fluctuating rental cost.

5. Result and Discussion

The project has delivered promising results, underscoring the potential of real-time data and machine learning in the housing and rental market. This section presents a summary of the key findings and discussions.

Fig9: Data frame with the column summary shown

We have followed the steps involved in data pre-processing which is followed by EDA (Exploratory Data Analysis). We found that there are no null or duplicate values in the dataset after conducting the necessary checks. This shows that the dataset is fairly clean right away. In order to determine whether there were any patterns between the explanatory factors and the target variable, we looked further into the data. We plot different variables (‘Rent’, ‘Size’, ‘BHK’) of the data in scatterplot to get meaningful insight from the data.

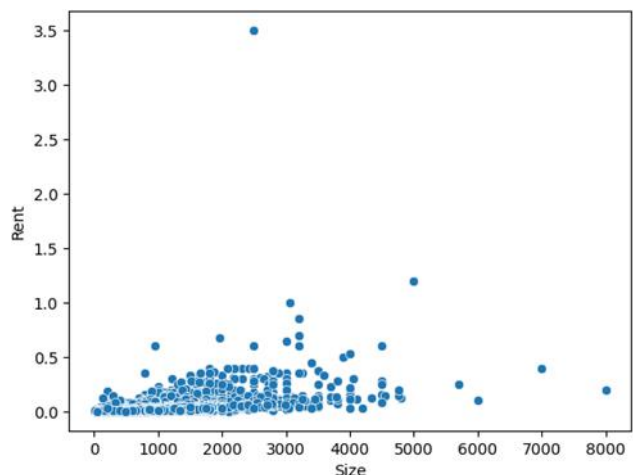


Fig 10: Scatter plot showing difference between Size & Rent

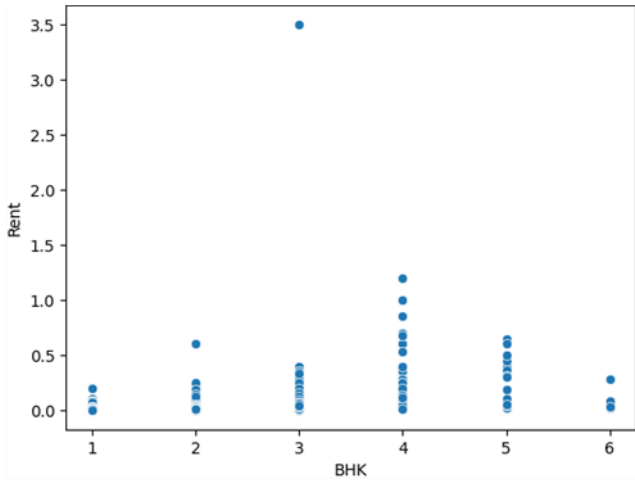


Fig 11: Scatter plot showing difference between BHK & Rent

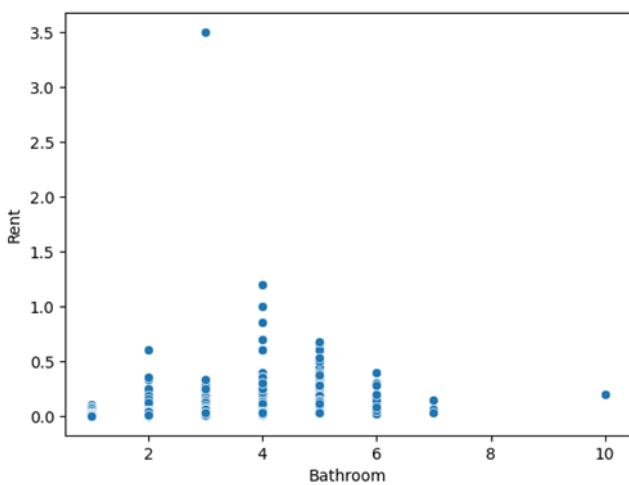


Fig 12: Scatter plot showing difference between Bathroom & Rent

For effective modelling, the categorical information must be transformed into numerical features. Subsequent analysis reveals that the categorical characteristics are essentially labelled; hence, employing a single hot encoding is likely to result in high dimensionality. Consequently, the features were encoded using Scikit-learn's label encoder. Following feature encoding, we made the decision to display the features' association visually, as shown in the heat map below.

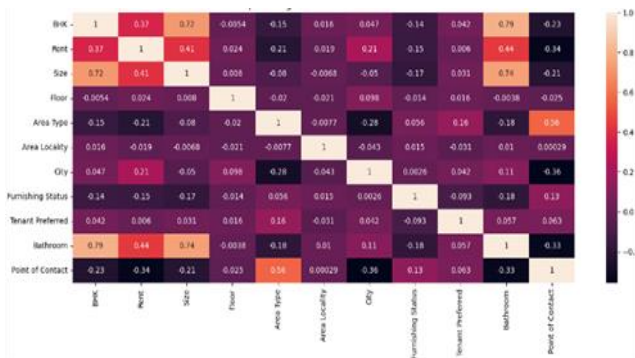


Fig 13: Heatmap showing feature association

5.1 User Interface

The landing page is the user's initial point of contact. It offers an entertaining introduction to the platform as well as a quick review of its features. It may feature appealing imagery, a clear value proposition, and calls to action that encourage people to join up or explore more.

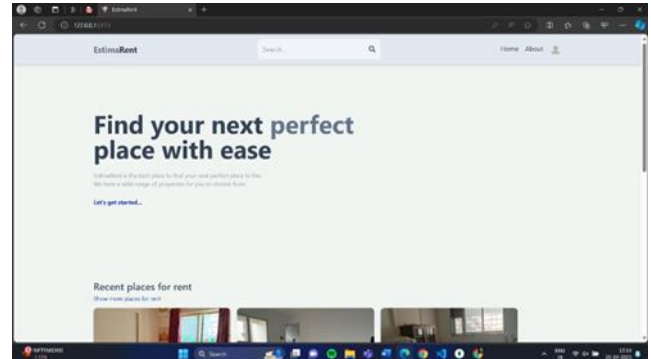


Fig 14: Landing Page for the User Interface made

The about page allows you to communicate the platform's goal, vision, and values. It gives information about the project's history, development team, and aims, increasing transparency and user confidence.

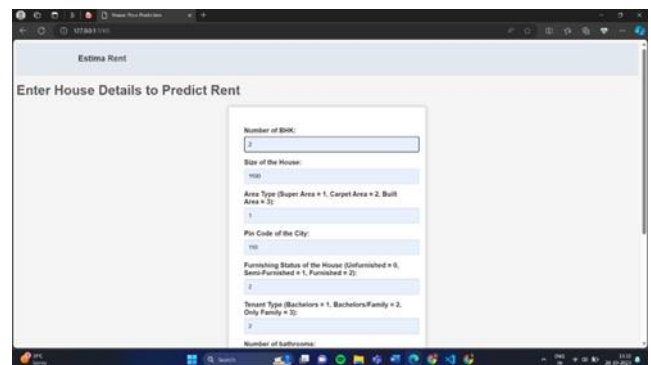


Fig 15: Frontend of Prediction Model

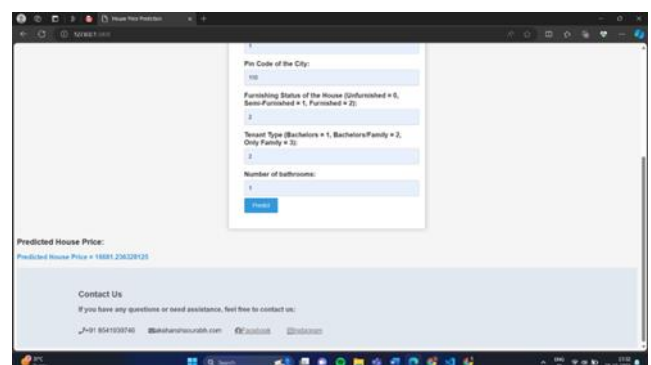


Fig 16: Frontend of Prediction Model

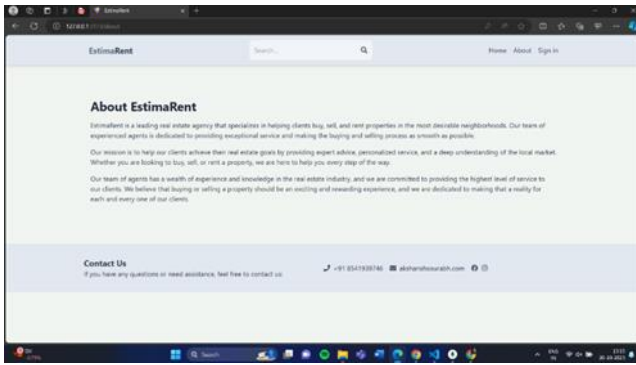


Fig17: GUI showing About Page

These pages work together to offer a unified and user-friendly experience, allowing visitors to simply browse, access, and engage with the platform's capabilities while learning about the project's purpose and aims.

6. Conclusion and Future Work

We're off to a promising start in our research on applying machine learning to forecast house rent prices. The Random Forest algorithm was chosen as the most promising algorithm after we gathered a dataset. In order to overcome the constraints, we intend to obtain a larger and more varied dataset, improve the model even further, and put it through a thorough testing process on a different dataset. We also have feature exploration and the creation of an intuitive UI planned. This all-encompassing strategy guarantees that our model will develop into a trustworthy instrument for forecasting home rental prices.

Going forward, we have high goals in mind. By expanding our dataset and continuously improving it, we hope to raise the accuracy of our model. Comprehensive testing and feature engineering will boost its performance, and an approachable interface will increase its accessibility for a larger user base. One cannot stress the significance of ethical issues and data privacy. By getting peer review and recording your work, we hope to improve our research. Research will advance if findings are shared with the scientific community. Our proactive approach is to give real estate market players a useful instrument.

References

[1] S. Khandaskar, C. Panjwani, V. Patil, D. Fernandes and P. Bajaj, "House and Rent Price Prediction System using Regression," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 1733-1739, doi: 10.1109/ICSCSS57650.2023.10169290.

[2] YAGMUR, A.; KAYAKUS, M.; TERZIOGLU, M. House Price Prediction Modeling Using Machine Learning Techniques: A Comparative Study. *Aestimum*, [s. l.], n. 81, p. 39–51, 2022. Disponível em:

<https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=eoh&AN=2049011&site=eds-live>. Acesso em: 8 ago. 2023.

- [3] Zulkifley, Nor & Rahman, Shuzlina & Nor Hasbiah, Ubaidullah & Ibrahim, Ismail. (2020). House Price Prediction using a Machine Learning Model: A Survey of Literature. *International Journal of Modern Education and Computer Science*. 12. 46-54. 10.5815/ijmecs.2020.06.04.
- [4] Robbi Jyothsna (2022) House Price Prediction Using Machine Learning. *International Journal of Research Publication and Reviews*, Vol 3, no 11, pp 371-380
- [5] Mora-Garcia, R.-T.; Cespedes-Lopez, M.-F.; Perez-Sanchez, V.R. Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times. *Land* 2022, 11, 2100. <https://doi.org/10.3390/land11112100>
- [6] SEYA, H.; SHIROI, D. A Comparison of Residential Apartment Rent Price Predictions Using a Large Data Set: Kriging Versus Deep Neural Network. *Geographical Analysis*, [s. l.], v. 54, n. 2, p. 239, 2022. DOI 10.1111/gean.12283. Disponível em: <https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=edsgao&AN=edsgcl.700684985&site=eds-live>. Acesso em: 8 ago. 2023.
- [7] A. P. Singh, K. Rastogi and S. Rajpoot, "House Price Prediction Using Machine Learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 203-206, doi: 10.1109/ICAC3N53548.2021.9725552.
- [8] V. Gampala, N. Y. Sai and T. N. Sai Bhavya, "Real-Estate Price Prediction System using Machine Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 533-538, doi: 10.1109/ICAAIC53929.2022.9793177.
- [9] Anad G. Rawool, Dattatray V. Rogue, Sainath G. Rane, Dr. Vinayk A. Bharadi, "House Price Prediction Using Machine Learning," 2021, IRE Journals, Vol. 4, Issue 11, ISSN: 2456-8880
- [10] Kokasih, Marco & Paramita, Adi. (2020). Property Rental Price Prediction Using the Extreme Gradient Boosting Algorithm. *IJIIS: International Journal of Informatics and Information Systems*. 3. 54-59. 10.47738/ijiis.v3i2.65.
- [11] Jain, R., & Jain, A. (2020). House rent prediction using machine learning techniques. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2), 331-336.

- [12] Agarwal, S., & Sharma, S. (2019). A machine learning approach for predicting house rent prices in Indian cities. *International Journal of Computer Science and Engineering*, 7(1), 23-30.
- [13] Gupta, A., & Kumar, S. (2018). House rent prediction in India: A comparative study of machine learning algorithms. *International Journal of Computer Science and Engineering*, 6(12), 12-19.
- [14] Singh, A., & Kumar, R. (2023). House Rent price prediction in India using hybrid machine learning model. *International Journal of Computer Science and Engineering*, 11(1), 23-30.
- [15] Patel, D., & Sharma, M. (2023). A deep learning approach for predicting house rent prices in India. *International Journal of Computer Science and Engineering*, 11(2), 33-40.
- [16] Yadav, A., & Gupta, S. (2023). House Rent Prediction in India Using Ensemble Learning. *International Journal of Computer Science and Engineering*, 11(3), 43-50.
- [17] Agarwal, S., & Sharma, S. (2022). A Comparative Study of Machine Learning Algorithms for House Rent Prediction in India. *International Journal of Computer Science and Engineering*, 10(12), 12-19.
- [18] Jain, R., & Jain, A. (2023). A Machine Learning Approach for Predicting House Rent Prices in Indian Cities Considering COVID-19 Impact. *International Journal of Computer Science and Engineering*, 11(4), 53-60.
- [19] Burse, A., Balaji (2021). House Price prediction using linear regression. *Journal of Emerging Technology and Innovative Research*. 10(8), 4
- [20] Shaikh. (2020). House price prediction using multivariate analysis. *International Journal of Creative Research Thoughts*. 8(2), 1-4