Received: 10 December 2025 | Revised: 17 March 2026 | Accepted: 9 May 2026 | Published online: 29 May 2026

RESEARCH ARTICLE

FinTech and Sustainable Innovations (FSI)
2026, Vol. 00(00) 1–10
DOI: 10.47852/bonviewFSI62028758

BON VIEW PUBLISHING

Apartment Purchase Suitability Prediction Using Explainable Machine Learning

Asim Foize Aimon[1] and Mahfuzulhoq Chowdhury[2,*]

[1] *Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh*

[2] *Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh*

Abstract: Apartment purchase suitability checking is important for customers due to apartment price, location, and customer's income issues. Existing research works could not incorporate essential features and ML models for apartment purchase suitability checking. The accuracy of their work is also low. They did not investigate proper feature selection method, data preparation, class imbalance issue solving, and hyper parameter tuning techniques. This paper introduces a ML based framework for predicting the suitability of apartment purchases by considering a number of new factors, advanced preprocessing, feature selection, and hyper parameter tuning methods. This work's preprocessing procedures included class balancing, addressing missing values, normalization, and ensemble feature selection technique. This paper examined 6 ML models for best prediction model selection and Grid Search CV is used for hyper parameter tuning. The random forest appears as most suitable ML model with 89.48% accuracy and 89.50% F1 score. According to the results, the proposed RF model outperformed previous works by offering more than 2 percent gain in accuracy and F1 score. This paper also included feature importance analysis based on SHAP.

Keywords: apartment purchase suitability, machine learning, feature selection, hyperparameter tuning, explainable AI (XAI)

1. Introduction

For individuals and families, buying an apartment is one of the biggest financial decisions they will ever make, especially in fast-growing cities like Dhaka, where rising real estate costs have made affordability a major problem [1]. A number of elements, such as property requirements, market volatility, location desirability, and long-term financial ramifications that may impact household stability for decades, add to the decision's complexity [2]. The majority of real estate prediction research to date has focused on recommendation systems and property valuation models that prioritize buyer preferences such as neighborhood features, size, location, and amenities [3, 4]. Although these methods offer insightful information about market patterns and property matching, they essentially ignore a crucial factor: the buyer's true financial capability and long-term affordability limitations. Furthermore, conventional real estate advising services frequently ignore the complex interactions between individual financial profiles, market dynamics, and property attributes in favor of oversimplified rules-of-thumb like the debt-to-income ratio.

In emerging economies, where income volatility, a lack of credit history, and shifting economic conditions add layers of complexity to the affordability assessment process, this disparity is more noticeable. Decisions about buying a home are typically made using the judgment of real estate brokers or one's own judgment, both of which are subject to bias and estimation errors. Buying an apartment has become especially difficult for middle-class households in rapidly urbanizing nations like Bangladesh due to rising real estate costs and increased urban density. With land prices in Dhaka rising by almost 2.740% and apartment prices rising by 716% during the past 20 years, the gap between ordinary wages and apartment prices in Bangladesh is growing [2].

The economic importance of the real estate industry is demonstrated by the fact that it currently accounts for about 8% of the national GDP according to Wikipedia. The decision-making process for prospective buyers is further complicated by the fact that, despite this development, almost 80% of city dwellers are compelled to rent because they cannot obtain mortgages and have financial limitations. There is currently no objective, data-driven system to assist purchasers in determining if an apartment is financially viable based on their income and property qualities, despite the fact that middle-class buyers are increasingly demanding mid-range residences. Prior research has mostly employed price-to-income ratio approaches [2] or conventional statistical methods [5] to analyze housing affordability, but both lack the predictive power required to assist individual purchasing decisions. This demonstrates the necessity of automated methods that use cutting-edge machine learning (ML) approaches to accurately estimate apartment affordability. Through the analysis of gathered data and attributes, ML-based predictive models can be useful for predicting the suitability of apartment purchases [6–8]. Both apartment attributes and buyer income data were not included in

***Corresponding author:** Mahfuzulhoq Chowdhury, Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh. Email: mahfuz_cse@cuet.ac.bd

the previous studies [1, 9–12]. Adequate feature extraction, data preprocessing, data imbalance handling, and hyperparameter tuning approaches were not included in the previous research [13–16]. The current system's forecast accuracy is insufficient. The integration of buyer income data with apartment attributes for suitability prediction is missing in the previous works. The accuracy of their work is low. They did not select suitable feature selection and ML model selection for apartment prediction with tuning. There is a lack of proper data collection, validation, and usage of proper features.

In this work, apartment attributes and buyer income data are combined to apply ML techniques to evaluate the financial feasibility of apartment acquisitions. This study suggests an ML-based predictive system that categorizes flats as "suitable to buy" or "not suitable to buy," offering buyers in Bangladesh individualized, data-driven decision support. In contrast to current methods that concentrate on credit evaluation or price prediction independently, the suggested methodology combines buyer financial profiles and property attributes to provide thorough recommendations for purchase suitability. This work's primary contributions are elaborated below:

1) This paper gathers data on the suitability of apartment purchases, including property attributes, buyer personal data, apartment location, cost, and buyer income. This dataset was gathered and verified from several Bangladeshi real estate firms. Techniques for feature extraction and adaptive data preparation were applied in this work.
2) To choose the best features, this work employed an ensemble feature selection technique that combined the SHAP, recursive feature elimination (RFE), and chi-square techniques. Issues with dataset imbalance are addressed by the SMOTETomek technique. Grid Search CV was utilized in this work to tune the hyperparameters.
3) The accuracy score and other performance metrics (precision score, F1 score) of six ML models are compared in this paper. The random forest (RF) model is selected for apartment purchase eligibility prediction with the highest accuracy value. The proposed prediction model's performance is compared with previous research. This study presented a SHAP-based explainable AI (XAI) technique to emphasize the significance of features in relation to prediction outcomes.

The explanation of related works is provided in the section that follows. Section 3 provides the ML-based eligibility prediction framework for apartment purchases, together with step-by-step instructions. Section 3 presents the comparison results. In section 5, the proposed work is summarized, and future research is discussed.

2. Literature Review

The literature on predicting eligibility for apartment purchases is illustrated in this second section. By using the XGBoost and BP neural network algorithms on several housing datasets, the work in [3] offers an ensemble learning-based approach for predicting housing prices. Their suggested model is not very accurate. Nevertheless, their work is restricted to estimating home prices solely, not predicting apartment eligibility. For the forecast task, they did not use the buyer's income or financial capacity. The article in [4] used Convolutional Neural Network (CNN) and Natural Language Processing (NLP) algorithms to forecast the increase in real estate property prices. They didn't offer any data on prediction accuracy. Gradient boosting was utilized in the paper in Almaslukh [6] to anticipate home prices using an advanced feature selection method. They only concentrated on increasing the accuracy of property price predictions. Using ML and Deep Learning (DL) approaches, the work in Cheng et al. [7] predicts credit risk related to apartment purchases. When compared to other methods, XGBoost fared well in its work. In order to evaluate credit risk in relation to the purchase of real estate, the article in Biswas et al. [8] employed logistic regression and neural networks. They didn't work on predicting the suitability of apartment purchases. For several apartment price predicting tasks, the study in Mostofi et al. [9] used a Deep Neural Network (DNN) with Principal Component Analysis (PCA), achieving a 90% accuracy rate. Multimodal ML approaches in real estate evaluation have been thoroughly examined in recent surveys [10], with a focus on model interpretability and accuracy. Although these surveys offer insightful information on sophisticated ML methods, their primary focus is on valuation accuracy rather than buyer-centric affordability analysis for apartment buying decisions. Traditional boosting and regression approaches were among the ensemble techniques compared in the work in Pastukh and Khomyshyn [11]. In their work, the gradient boosting approach produced excellent accuracy results for the valuation of real estate properties. With an accuracy of more than 90%, the article in Saini et al. [12] created a loan approval prediction for real estate purchases using the RF method. Through statistical analysis of price-income gaps, the study in Giti et al. [5] offered statistical insights into affordability gaps in the housing market of Dhaka. Tree-based algorithms were utilized in the work in Najib et al. [1] to forecast the price of a rental home. The accuracy of their suggested approach was 86%. Blockchain technology was employed for property valuation in the work in Adilieme et al. [13]. Using social, environmental, and economic factors, the study in Farzana et al. [14] carried out a comparative evaluation of sustainable housing affordability in Khulna. Despite having a broad reach, this study concentrates on policy-level analysis as opposed to buyer-property matching for individual purchase choices. The price-to-income ratio technique was used in the study in Asaduzzaman et al. [2] to evaluate the affordability of housing for middle-class Rajshahi residents. This study lacks ML integration and individualized prediction capabilities for the suitability of individual apartment purchases, although being helpful for macro-level affordability assessments. Extra Trees Regression was utilized in a recent work in Reza et al. [15] to forecast Dhaka rental prices, with an accuracy rate of over 90%. However, this study just looks at rental price prediction; it doesn't take the buyer's financial capacity or purchase affordability into account. Through the forecast of energy poverty in Bangladesh, the study in Karmaker et al. [16] illustrates the potential of ML in household affordability research. This study does not integrate real estate features and does not work inside the housing buying environment, despite demonstrating ML capabilities in socio-economic prediction. In order to value properties, the work in Deng and Zhang [17] employed ensemble learning techniques. It does not, however, address buyer affordability or apartment purchase suitability assessment, as is the case with previous price-focused research. A number of housing affordability-related factors were examined in the paper in Waddel and Besharati [18]. An RF-based residential property price prediction technique with a low Root Mean Square Error (RMSE) value was employed in the study in Yee et al. [19]. The impact of social media marketing on Indian consumers' decisions to buy apartments was examined in the article in Ananthan and Vinayagam [20]. The study [21] looked into how augmented reality-based marketing affected purchasers' decisions

to acquire real estate. The study in Wicaksono and Haris [22] predicted clients' intentions to buy real estate using a fast text-based DL model. A secure property registration system for buyers based on blockchain technology was created in the article in Mehta et al. [23]. The study in Khmelnitsky and Singer [24] looked into the best price strategy for US real estate sales. The XGBoost regressor was utilized in the article in Kumar et al. [25] to estimate Indian home prices. For the Indian population, the authors in [26] created a house rent prediction method based on RFs. In order to anticipate a buyer's eligibility for an apartment purchase, previous research did not look into both buyer income and apartment attributes. Low prediction accuracy, appropriate feature selection, preprocessing, controlling dataset imbalance, and hyperparameter tweaking were among the problems that the previous works failed to address. This work offers an ML-based framework for predicting eligibility for apartment purchases with appropriate feature selection, hyperparameter tuning, data imbalance handling, and preprocessing methods to lessen the drawbacks.

3. Proposed Apartment Purchase Prediction Scheme Using Machine Learning

This section will cover the proposed system model for predicting eligibility for apartment purchases.

3.1. System design

Figure 1 shows the suggested ML-based apartment purchasing appropriateness prediction model. Data gathering, data preprocessing, feature selection, class balance, model performance checking, training, hyperparameter tweaking, testing, and apartment purchase suitability prediction using the chosen model are some of the crucial steps in this study's approach. Encoding categorical variables and changing inconsistent column names (such as salary to income) were the first steps in cleaning and transforming the raw dataset. Missing values were dealt with proper checking by the validators. To find the most pertinent characteristics impacting price appropriateness, feature selection was carried out utilizing methods such as chi-square scoring, RFE, and the SHAP approach. SMOTETomek, a hybrid oversampling and undersampling technique, was used to balance the dataset due to the class imbalance in the labeled target (appropriate vs. not suitable), which improved the model's generalization across both classes. Several ML and DL models were used, and the best model was chosen to forecast eligibility for apartment purchases with the highest accuracy. To improve performance, Grid Search CV was used for hyperparameter tweaking. A random selection of attributes (e.g., flat price, income, bedroom, bathroom, floor size, and distances to amenities, loc1, and loc2) is paired with each of the varied data subsets produced by bootstrap sampling. At the leaf level, distinct suitable and not suitable classifications result from the generic nodes that are used to specify each decision tree (DT). The best precision in this work is achieved by using a majority vote to select the final product.

3.2. Dataset preparation

A dataset gathered from several well-known real estate firms and internet resources in Bangladesh was used in this investigation. The dataset used in this investigation includes 4141 apartment purchase records in total. Property attributes like location, cost, and floor area, and apartment room numbers were systematically compiled as part of the data collection process. With this method, the dataset is guaranteed to represent a diverse and accurate sample of the Bangladeshi real estate market, making it appropriate for ML-based research of real estate purchasing patterns. The dataset was gathered from a number of Bangladeshi real estate firms, including Holy Homes, Sheltech Group Limited, and Navana Real Estate Company. The dataset contains labeled samples that distinguish between apartments that are acceptable and those that are not, according to a number of characteristics like location, cost, and income, as seen in Table 1. The

Figure 1
Methodology diagram for proposed apartment purchase eligibility prediction

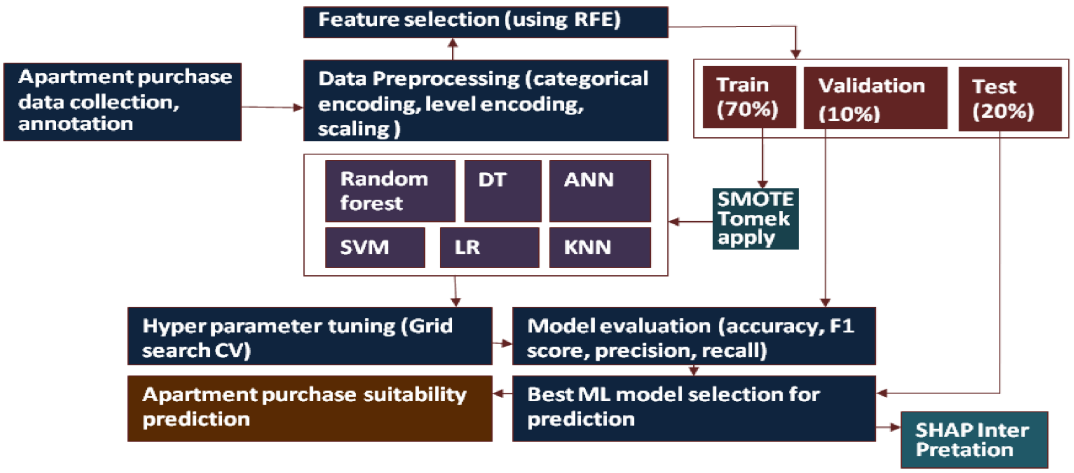


Table 1
Proportion of dataset entries

ID	Loc1, Loc2	Bedroom, bathroom	Floor area (sq ft)	Flat price (BDT)	Nearest mosque (meter)
2993	DOHS, Mirpur	3, 3	1200	10000K	310
4124	Road 7, Banani	3, 3	1240	13500K	990
2044	Ashulia, Savar	8, 4	2146	8500K	237
1108	Sector 13, Uttara	2, 1	1000	2050K	107
Nearest hospital (m)	Nearest shop (m)	Nearest school (m)	Income (month)	Label	Validation
1559	1310	1130	142K	0 (unsuitable)	Yes
1970	1256	1385	266K	0 (unsuitable)	Yes
758	946	1066	149K	0 (unsuitable)	Yes
200	135	528	272K	1 (suitable)	Yes

Figure 2
Data collection sources distribution

Data collection sources distribution (total=4141)

Company name	Data collection percentage
Navana	11.4%
BD property	22.5%
Holy homes	17.2%
Sheltech	48.9%

Figure 3
Features list with their explanation

Features name	Indication	Features name	Indication
Loc1	Primary area or district	Flat price	Property price in BDT
Loc2	Sub area or neighborhood	Nearest shopping	Distance to nearby shopping center in meters
bedroom	Number of bedrooms	Nearest hospital	Distance to nearby hospital in meters
bathroom	Number of bathrooms	salary	Monthly buyer income in BDT
floor	Floor level	Nearest school	Distance to nearby school in meters
area	Floor area in square feet	Nearest mosque	Distance to nearby mosque in meters
lavel	Suitability indicator (0=not suitable, 1=suitable)		

distribution of data gathering sources is displayed in Figure 2. Thirteen features, including flat price, buyer income, appropriateness classification, location, floor area, bedroom, bathroom number, and proximity metrics, are included in the dataset (see Figure 3). In order to categorize whether a buyer can actually afford a certain apartment, the target variable (label) was created, taking into account the buyer's financial status in addition to the property's qualities. When a buyer's monthly income is not enough to cover the entire cost of the purchase, they are assigned the level 0 (unsuitability) of apartment purchase in the dataset. If the yearly income is greater than the flat cost, the purchase suitability is true. Moreover, flat price, location, flat room number, shop distance, and school location distance are also checked for suitability. Apartments, when the buyer has sufficient financial means, are given the level 1 appropriateness rating for apartment purchases. Five real estate industry experts, two clients, and a university professor validate the dataset. Based on majority voting, the final label is determined (see Table 2).

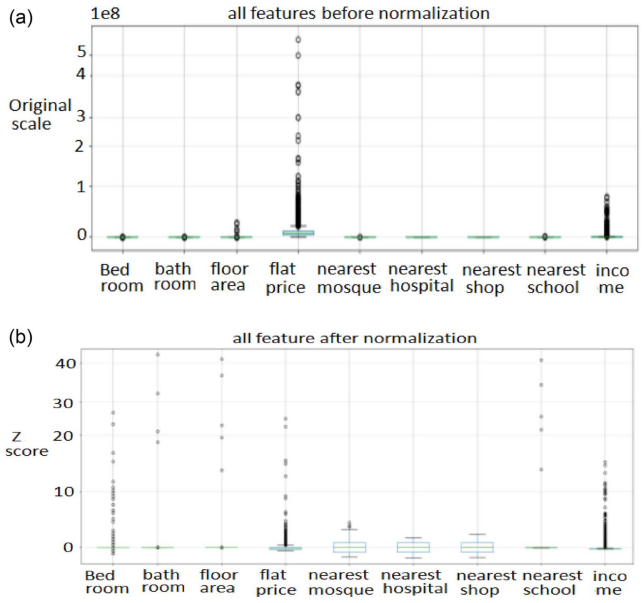
3.3. Data preprocessing

The dataset's missing and null values are examined to guarantee the dataset's dependability and relevance. Encoding categorical variables and changing inconsistent column names (such as salary to income) were the first steps in cleaning and transforming the raw dataset. Missing values were dealt with properly. IQR technique was incorporated to remove outliers for numerical properties like price, income, and distance features,

Table 2
Class annotation using majority voting

Flat location	Bedroom, bathroom requirement	Floor area, nearest mosque, hospital, shop, school distance requirement	Flat price greater than yearly income	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Majority voting
Match with requirements	Match with requirements	Match with requirements	Yes	Suitable	Suitable	Suitable	Suitable	Suitable	Suitable
Match with requirements	Not match with requirements	Not match with requirements	No	Unsuitable	Unsuitable	Unsuitable	Unsuitable	Unsuitable	Unsuitable

Figure 4
Features before and after normalization



while label encoding was used to consistently encode categorical variables like location. One-hot encoding is used to transform location-based features (loc1, loc2) into a numeric representation so that ML models can use them. By generating binary indicator variables for every categorical value, this transformation enables algorithms to efficiently handle categorical data. Z-score normalization is used to standardize feature scales after a methodical data cleaning step in the preprocessing stage. *Z*-score normalization is used to standardize numerical features including income, distances, and flat prices. Figure 4(b) displays the standardized version following normalization, whereas Figure 4(a) displays the distribution of all attributes before normalization. The SMOTE technique with Tomek connections is used for balanced dataset construction in order to handle possible class imbalance issues. Tomek linkages are used to find noisy or borderline samples, which are cases from different classes that are closest to one another. SMOTE interpolates between real occurrences to create synthetic examples of the minority class. The training dataset class count before and after the SMOTETomek was applied is displayed in Table 3. To avoid data leakage, SMOTETomek is applied only to the training data. Cohen's kappa score is also shown in Table 3 for inter-annotator agreement measurement. Majority voting is used for the final labeling of the class. In this work, "suitable apartment purchase for buyer" refers to the buyer's salary being adequate to cover the cost of the flat. The room number, shop, school location, and apartment location that match the buyer's requirements are additional criteria for suitability. The flat is not suitable for buyers if their income is insufficient to cover the cost. Additionally, mismatches in the locations of the apartment, store, school, and room number are other elements that contribute to unsuitability.

3.4. Feature selection

In order to systematically find the most discriminative predictors for multiclass classification, this work used a wide range of statistical, model-based, and XAI feature selection strategies, all of which have their roots in rigorous mathematical concepts. A number of competitive feature selection techniques were used to select the most suitable features for prediction. The feature selection methods used in this work are RFE, chi-square, and SHAP for training. Borda count aggregation, a rank-based fusion technique that allocates scores based on each feature's position across many selection criteria, was used to combine feature ranks from all approaches. The important feature rankings and scores for various feature selection techniques are displayed in Figure 5(a) and Figure 5(b), respectively.

This work carried out several iterations of the same methodology to ascertain the ideal number of traits to eliminate. This study evaluated model performance at various feature subset sizes using RF as the base classifier because features cannot be removed at random. We determined the saturation point, the point at which additional feature removal no longer significantly improves (or starts to deteriorate) model performance by tracking the accuracy, F1 score, precision, recall, and Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) throughout these rounds. Train, validation, and test sets are divided into 70%, 10%, and 20%, respectively.

Figure 5
(a, b) Important feature selection and ranking

(a) Features importance scores

Features	Chi2 score	RFE score	SHAP score
Flat price	1247.8	.342	.89
income	892.56	.29	.75
area	634.2	.18	.62
bedroom	421.6	.06	.23
Loc1	387.4	.08	.18
bathroom	298.3	.05	.16
Loc2	267.8	.12	.14
floor	156.7	.03	.09
Nearest school	134.2	.07	.08
Nearest hospital	89.67	.04	.07
Nearest mosque	67.45	.021	.05
Nearest shop	45.89	.015	.04

(b) Features importance ranking

Features	Chi2 rank	RFE rank	SHAP rank	Ensemble rank
Flat price	1	1	1	1
income	2	2	2	2
area	3	3	3	3
bedroom	4	6	4	4
Loc1	5	5	5	5
bathroom	6	7	6	6
Loc2	7	4	7	7
floor	8	9	8	8
Nearest school	9	8	9	9
Nearest hospital	10	10	10	10
Nearest mosque	11	11	11	11
Nearest shop	12	12	12	12

Table 3
Class balancing using SMOTETomek and Cohen's kappa score

Class name	Training data before SMOTETomek	Training data after SMOTETomek	P_0	P_e	Cohen's kappa score
Unsuitable	999	2000	0.85	0.1	0.83
Suitable	2000	2000	0.91	0.1	0.90

Figure 6
(a–c) Hyperparameter tuning. (d–f) Hyperparameter tuning

(a)

Random forest parameters	Values
n estimators	[100,200,300, 400, 500]
Max depth	[10, 15, 20, 25, none]
Min samples split	[2,5, 10]
Min samples leaf	[1,2,4]
Max features	[Sqrt, log2]

(b)

Decision tree parameters	Values
Max depth	[10, 15, 20, 25, none]
Min samples split	[2,5, 10,15]
Min samples leaf	[1,2,4,8]
criterion	[gini, entropy]

(c)

SVM parameters	Values
C	[.1,1,10,100, 1000]
kernel	[Rbf,linear, poly]
gamma	[Scale, auto, .001,.01,.1]
Decision function shape	[ovo,ovr]

(e)

ANN parameters	Values
Hidden layer sizes	[50, 100, (50, 50), (100,50)]
activation	[relu, tanh, logistic]
solver	[adam, lbfgs, sgd]
alpha	[.0001, .0001, .001]
Learning rate	[Constant, adaptive]

(d)

Logistic regression parameters	Values
C	[.01,.1,1,10,100]
Max iter	[100,200,300,400,500]
penalty	[l1,l2]
Solver	[Lbfgs,liblinear,saga]
Class weight	[None,balanced]

(f)

KNN parameters	Values
N neighbors	[3,5,7,9,11,15]
weights	[uniform,distance]
metric	[Euclidean,manhattan,minkowski]

Table 4
Performance comparison machine learning model (chi-square feature selection technique)

Model (using RFE)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Random forest	89.48	89.49	89.51	89.50
Decision tree	82	86	85	85.5
ANN	85	84	85	84.50
Logistic regression	78	81	78	79.50
SVM	66	74	66	70
KNN	61	72	61	66.5

Table 5
Performance comparison machine learning models (RFE) and statistical significance test

Model (using RFE)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	McNemar tests T-statistic (%)	McNemar tests *p*-value
Random forest	89.48	89.49	89.51	89.50	05	0.035
Decision tree	82	86	85	85.5	10	0.430
ANN	85	84	85	84.50	04	0.070
Logistic regression	78	81	78	79.50	10	0.830
SVM	66	74	66	70	13	0.930
KNN	61	72	61	66.5	14	1.000

Table 6
Performance comparison of machine learning models (SHAP)

Model (using SHAP)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Random forest	88.40	87.80	88.20	88
Decision tree	79	83	81	82
ANN	85	84	85	84.50
Logistic regression	81	80	81	80.50
SVM	79	82	79	80.50
KNN	57	68	57	62.50

3.5. Model selection, tuning, statistical test

Grid Search CV was utilized to maximize the proposed ML model performance with k-fold cross-validation (k = 10). The highest F1 score was used to determine the optimal hyperparameters. The hyperparameter grid of various ML models (logistic regression [LR], DT, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), RF) is displayed in Figure 6. The accuracy value and F1 score value of six comparable ML models are then used to determine the best model. Tables 4, 5, and 6 highlight the comparative performance of conventional ML models (e.g., RF, DT, SVM, KNN, ANN, LR) utilizing different feature selection strategies (chi-square, RFE, and SHAP), respectively. Since the RF model with RFE feature selection has the highest accuracy value out of all the ML models that were compared, it is chosen as the best model. Table 5 shows the statistical significance test using McNemar's test. The best RF classifiers (with RFE feature selection) empirical accuracy results (89.48%) and p-value (p < 0.035) were validated through a statistical hypothesis test. A statistically significant difference between the model's predictions and the ground truth is indicated by a p-value less than 0.05 when comparing a forecast to the ground truth using McNemar's test. Table 5 displays the McNemar's test p-value and T-statistic values to assess whether the RF classifier's performance differs statistically from those of other classifiers.

Figure 7
(a) Confusion matrix for random forest. (b) ROC-AUC curve for random forest

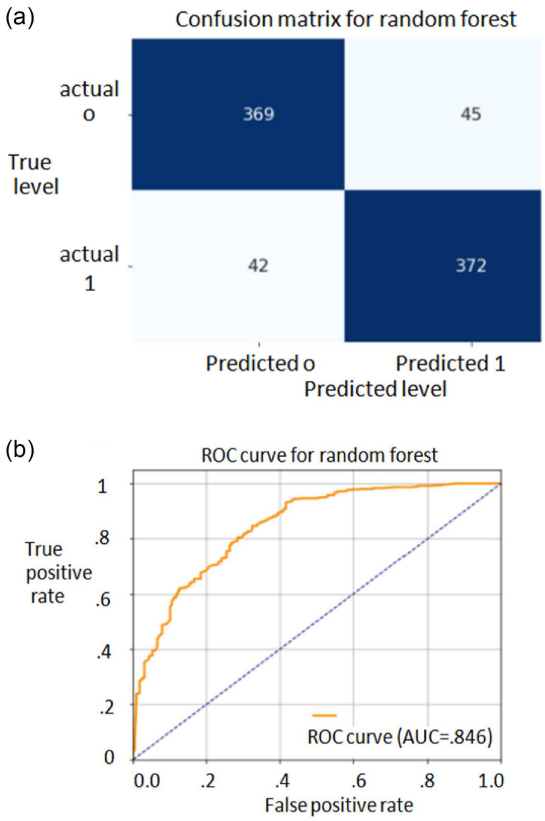


Figure 7(a) shows the confusion matrix result, and Figure 7(b) displays the ROC-AUC curve of the selected RF scheme for prediction. With an outstanding AUC score of 0.846, RF is the best classifier for predicting the suitability of apartment purchases. For every prediction class, the confusion matrix curve demonstrates that the proposed RF approach has the highest true purchase suitability prediction rates. The false positive prediction rates are also small in the proposed method. The SHAP-based feature significance analysis is displayed in Figure 8. With a SHAP score of 0.891, the figure demonstrates that the flat price is the most significant predictor of apartment purchasing appropriateness. The second most important element is income (SHAP score: 0.756). Area comes in third (SHAP score: 0.623), suggesting that apartment size has a significant impact on purchasing appropriateness. Bedroom count has a considerable impact on apartment purchasing appropriateness, albeit being less significant (SHAP score: 0.234). With SHAP scores of 0.189 and 0.145, location factors like loc1 and loc2 have a moderate contribution. The remaining characteristics, such as the floor number, the closest hospital, school, mosque, and superstore, contribute less, suggesting that while they are important, they are not the primary factors in evaluating whether an apartment is a good buy.

4. Comparison with Existing Works

Table 7 outlines the results of the comparison between the proposed ML-based apartment purchase suitability work and the existing schemes. The table demonstrates that, in comparison to other literary approaches (e.g., [1, 7, 11]), the suggested RF-based apartment purchase eligibility scheme delivers the greatest accuracy (89.48%) and F1 score value (89.50%). Compared to current efforts, the suggested strategy provides at least 2% higher accuracy value and 4% more F1 score value. Data collection with multiple apartment and buyer income information

Figure 8
SHAP-based feature importance analysis

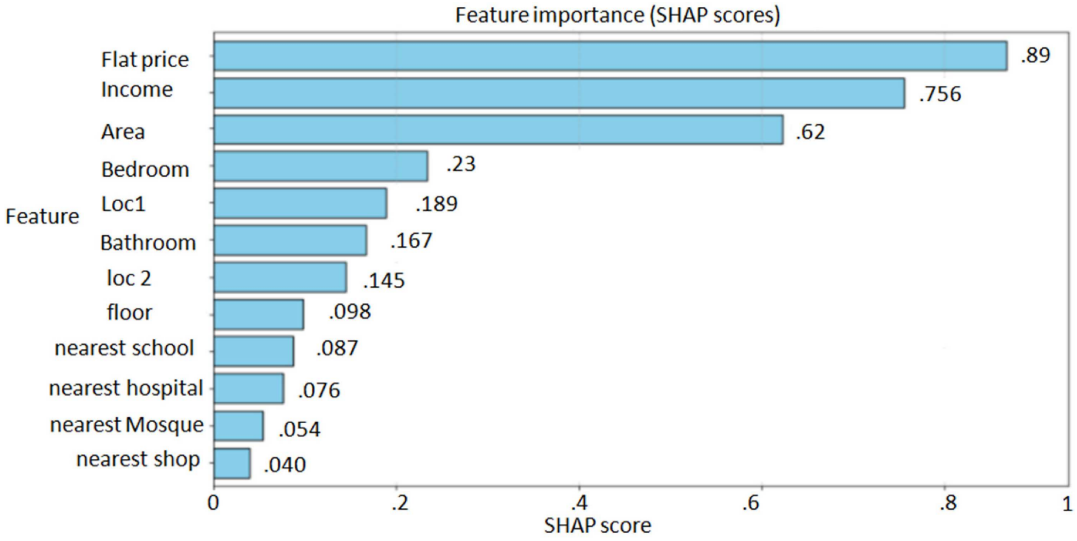


Table 7
Comparison results

Work name	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Used method
[11]	72.40	73.20	72.50	73.50	Gradient boosting
[7]	87.10	85.22	88.35	83.34	XGBoost
[1]	86.00	85.20	84.60	85.50	Ensemble ML algorithm
Proposed work	89.48	89.49	89.51	89.50	Random forest classifier with RFE

features, advanced data preprocessing, RFE-based feature selection, SMOTETomek-based class imbalance problem solving, and Grid Search CV-based hyperparameter tuning technique are the primary factors contributing to the suggested schemes' suitability.

5. Conclusion

This paper presented a technique for predicting if an apartment purchase would be suitable for prospective buyers based on ML. This paper gathered data on apartment purchase suitability prediction from several real estate firms, including location, buyer income, and apartment details. Data pretreatment was carried out in this work, including resolving missing and null values, encoding, normalization, and class imbalance issues utilizing the SMOTETomek method. Using the SHAP, chi-square, and RFE techniques, an ensemble feature selection method is used to choose appropriate characteristics for predicting eligibility for apartment purchases. In this work, hyperparameter tweaking is done using Grid Search CV. Six ML models were tested in order to choose the best one for predicting eligibility for apartment purchases. For predicting apartment buying eligibility, RF with the RFE feature selection approach is chosen since it provides greater accuracy than the compared ML models. The evaluation results confirmed that the proposed RF model outperforms the existing comparative works by more than 2% gain in accuracy and F1 score. The proposed ML-based apartment purchase suitability prediction is essential for buyers, developers, and policy-makers to identify ideal properties based on personal data, promoting sustainable urban development, project planning, addressing the housing shortage problem, equitable housing facility access decision-making process, risk mitigation, and optimizing resource allocation in the real estate market. This paper explains feature importance analysis based on SHAP. The limitations of this work are the lack of addition of different socio-economic factors and geographical location data for a sustainable apartment purchase suitability prediction system. Future research challenges include, but are not limited to, Internet of Things (IoT) and DL-based real-time eligibility evaluation for apartment purchases, apartment categorization based on a person's income and location, Reinforcement Learning (RL) and Large Language Model (LLM)-based buyer feedback analysis, transfer learning-based appropriate real estate company selection, LLM-based appropriate apartment selection for purchase, and IoT and blockchain-based owner verification and secure apartment sales process.

Recommendation

This work suggested that the RF algorithm is most efficient for apartment purchase eligibility prediction with proper feature selection and hyperparameter tuning techniques.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Asim Foize Aimon: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Visualization. **Mahfuzulhoq Chowdhury:** Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Najib, T., Muntasir, F., & Wasi, W. W. (2023). Transparency in house rent of Dhaka: Explainable AI based predictive framework. In *Proceedings of the 6th Industrial Engineering and Operations Management Bangladesh Conference*, 26–28. https://doi.org/10.46254/BA06.20230146

[2] Asaduzzaman, M., & Sheikh, H. (2021). Measuring affordability of the middle income group for residential house price in real estate sector of Rajshahi, Bangladesh. *American Journal Scientific Research Journal for Engineering, technology, and Sciences*, *82*, 1–10.

[3] Yang, Z., Zhu, X., Zhang, Y., Nie, P., & Liu, X. (2023). A housing price prediction method based on stacking ensemble learning optimization method. In *10th International Conference on Cyber Security and Cloud Computing*, 96–101. 10.1109/CSCloud-EdgeCom58631.2023.00025

[4] Wandhe, A., Sehgal, L., Sumra, H., Choudhary, A., & Dhone, M. (2023). Real estate prediction system using ML. In *2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing*, 1–4.

[5] Giti, A. (2018). *Measuring ownership housing affordability of middle income people in Dhaka city*. Retrieved from: https://www.bip.org.bd/admin/uploads/bip-publication/publication-19/paper/20181204075017.pdf

[6] Almaslukh, B. (2020). A gradient boosting method for effective prediction of housing prices in complex real estate systems. In *2020 International Conference on Technologies and Applications of Artificial Intelligence*, 217–222. https://doi.org/10.1109/TAAI51410.2020.00047

[7] Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and deep learning: A study on credit card customers. *Risks*, *12*(11), 174. https://doi.org/10.3390/risks12110174

[8] Biswas, N., Mondal, A. S., Kusumastuti, A., Saha, S., & Mondal, K. C. (2022). Automated credit assessment framework using ETL process and machine learning. *Innovations in Systems and Software Engineering*, *21*(1), 257–270. https://doi.org/10.1007/s11334-022-00522-x

[9] Mostofi, F., Toğan, V., & Başağa, H. B. (2022). Real-estate price prediction with deep neural network and principal component analysis. *Organization, Technology and Management in Construction: an International Journal*, *14*(1), 2741–2759. https://doi.org/10.2478/otmcj-2022-0016

[10] Huang, C., Liang, B., Li, Z., & Chen, F. (2025). Multimodal machine learning for real estate appraisal: A comprehensive survey. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 345–361. https://doi.org/10.1007/978-981-96-8183-9_26

[11] Pastukh, O., & Khomyshyn, V. (2024). Using ensemble methods of machine learning to predict real estate prices. *arXiv Preprint:2504.04303*

[12] Saini, P. S., Bhatnagar, A., & Rani, L. (2023). Loan approval prediction using machine learning: A comparative analysis of classification algorithms. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering*, 1821–1826. https://doi.org/10.1109/ICACITE57410.2023.10182799

[13] Adilieme, C. M., Abidoye, R. B., & Lee, C. L. (2024). Reducing clients' influence in property valuation: An exploration of a blockchain-based solution. *Habitat International*, *154*, 103217. https://doi.org/10.1016/j.habitatint.2024.103217

[14] Farzana, F., Al Mahmood, T., Biswas, M., & Islam, L. (2023). Assessment of sustainable housing affordability: A comparative study between two planned residential areas of Khulna city. *Khulna University Studies*, *20*(2), 160–168. https://doi.org/10.53808/KUS.2023.20.02.310-se

[15] Reza, H., Tareq, N. I., Dehan, T. A., & Joy, U. M. (2024). Predicting apartment rental prices in Bangladesh: A machine learning approach. In *2024 International Conference on Recent Progresses in Science, Engineering and Technology*, 1–4. https://doi.org/10.1109/ICRPSET64863.2024.10955925

[16] Karmaker, S. C., Rjbongshi, A., Pal, B., Sen, K. K., & Chapman, A. J. (2025). Machine learning-based prediction of energy poverty in Bangladesh: Unveiling key socioeconomic drivers for targeted policy actions. *Socio-Economic Planning Sciences*, *99*, 102213. https://doi.org/10.1016/j.seps.2025.102213

[17] Deng, L., & Zhang, X. (2025). Boosting the accuracy of property valuation with ensemble learning and explainable artificial intelligence: The case of Hong Kong. *The Annals of Regional Science*, *74*(1), 1–15. https://doi.org/10.1007/s00168-025-01365-7

[18] Waddel, P., & Besharati, A. (2023). *Data-driven multiscale planning for housing affordability*. Joint Center for Housing Studies of Harvard University. https://www.jchs.harvard.edu/sites/default/files/research/files/harvard_jchs_digitalization_panel5_waddell_2023.pdf

[19] Yee, L. W., Bakar, N. A. A., Hassan, N. H., Zainuddin, N. M., Yusoff, R. C. M., & Ab Rahim, N. Z. (2021). Using machine learning to forecast residential property prices in overcoming the property overhang issue. In *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology*, 1–6. https://doi.org/10.1109/IICAIET51634.2021.9573830

[20] Ananthan, J., & Vinayagam, K. (2025). A study on impact of social media marketing on flats purchase decision of consumers with respect to real estate in Puducherry UT. In *2025 International Conference on Automation and Computation*, 182–186. https://doi.org/10.1109/AUTOCOM64127.2025.10956836

[21] Rosadi, I., Cindarbumi, D., & Ananda, A. S. (2024). Enhancing housing property: The mediating role of marketing activities in augmented reality and its impact on purchase

intentions. In *2024 International Conference on Informatics, Multimedia, Cyber and Information System*, 650–655. https://doi.org/10.1109/ICIMCIS63449.2024.10957390

[22] Wicaksono, B. D., & Haris, M. (2023). Predicting customer intentions in purchasing property units using deep learning. In *2023 International Conference on Information Technology Research and Innovation*, 103–108. https://doi.org/10.1109/ICITRI59340.2023.10249704

[23] Mehta, N., Rani, R., & Kalra, N. (2023). Blockchain based system for secure property registration. In *2023 International Conference on Sustainable Computing and Smart Systems*, 1426–1433. https://doi.org/10.1109/ICSCSS57650.2023.10169812

[24] Khmelnitsky, E., & Singer, G. (2023). Optimal real estate pricing and offer acceptance strategy. *IEEE Access*, *11*, 58644–58653. https://doi.org/10.1109/ACCESS.2023.3284549

[25] Kumar, D., Rawat, A. S., Jha, S., & Yadav, D. (2023). Machine learning-based prediction of home prices. In *2023 5th International Conference on Advances in Computing, Communication Control and Networking*, 423–428. https://doi.org/10.1109/ICAC3N60023.2023.10541614

[26] Singh, H. V., Srivastava, S. K., Pandey, N. K., & Kumar, N. (2025). Integrating random forest and XGBoost for accurate house rent prediction. In *2025 International Conference on Pervasive Computational Technologies*, 956–961. https://doi.org/10.1109/ICPCT64145.2025.10939118

How to Cite: Aimon, A. F., & Chowdhury, M. (2026). Apartment Purchase Suitability Prediction Using Explainable Machine Learning. *FinTech and Sustainable Innovations (FSI)*. https://doi.org/10.47852/bonviewFSI62028758