APARTMENT PRICE PREDICTION OF BANGLADESH USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

BY

MD. MUSTAK AHMED
ID: 202-15-14418

AND

MAHJABEEN HOSSAIN
ID: 202-15-14435

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "**Apartment Price Prediction of Bangladesh using Machine Learning and Deep Learning Algorithms**", submitted by **Md. Mustak Ahmed** and **Mahjabeen Hossain** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **01-07-2024**.

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah (MIJ) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Assaduzzaman (MA) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Ms. Sharun Akter Khushbu (SAK) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza (DWR) — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Fahad Faisal, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Fahad Faisal
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. S. M. Aminul Haque
Professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:

Md. Mustak Ahmed
ID: 202-15-14418
Department of CSE
Daffodil International University

Mahjabeen Hossain
ID: 202-15-14435
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Fahad Faisal**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning and Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The overall aim of this study is to analyze the effectiveness of different machine learning and deep learning models for apartment price prediction in terms of their accuracy and error measures, so as to determine how well these models can predict apartment prices. The experimental setup was done using exhaustive preprocessing and utilized the Scikit-Learn, TensorFlow frameworks for training the model. Also considered were Random Forest, Decision Tree, K-Nearest Neighbor, Gradient Boosting, Extreme Gradient Boosting and Convolutional Neural Network (CNN), Artificial Neural Network (ANN) among others. They also used evaluation indicators like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Accuracy (R^2). It was found that the Extreme Gradient Boosting Regressor had the highest accuracy of 97.53%, indicating its reliability in forecasting apartment prices. Following this closely was Decision Tree Regressor with an accuracy rate of 96.68%, proving that it is still efficient even though it is a simpler model. On top of that, Random Forest Regressor has also shown good performance with an accuracy score of 96.15% owing to its ensemble nature. Conversely; reduced prediction accuracies were observed for both K-Nearest Neighbor Regressor and Artificial Neural Network (ANN) having R^2 values at 87.55% and 89.73% respectively. Despite their lesser accuracy, these models still offered useful insights and could be further improved. The Convolutional Neural Network (CNN) demonstrated competitive performance, suggesting its potentials on intricate patterns in the data. This study points to the best fit of Decision Tree Regressor and Extreme Gradient Boosting Regressor among other models for forecasting apartment prices. These results indicate that selection of a model and preprocessing are important in predicting real estate prices accurately.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
DL	Deep Learning
RF	Random Forest
DT	Decision Tree
KNN	K-Nearest Neighbor
GBOOST	Gradient Boosting
XGBOOST	Extreme Gradient Boosting
CNN	Convolutional Neural Network
ANN	Artificial Neural Network
R^2	R-squared
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error

CHAPTER 1

Introduction

1.1 Overview

The demand for residential apartments in Bangladesh is clearly on the rise with major cities like Dhaka, one of the world's largest mega cities, and other cities including Chattogram, Khulna, Sylhet, Barisal, Narayanganj, Gazipur, Rajshahi, Cumilla, Rangpur and Mymensingh. Properly estimating apartment prices during this boom is crucial in-order to meet the needs of various stakeholders trying to make sense of this dynamic market. Sophisticated methods are required for an accurate assessment and forecast of these values, given that pricing is multi-dimensional and depends on a complex interplay of factors.

Machine learning and deep learning techniques have become more important in real estate business as far as the use and study of machine learning for apartment price forecast are concerned. Accurately projecting apartment prices has become possible and more significant for a variety of stakeholders including buyers, sellers, investors and real estate brokers thanks to the growth of data and computational techniques [1].
Apartment price prediction projects have made extensive use support vector machines (SVM), decision trees (DT), random forests (RF) as well as linear regression (LR). These algorithms utilize historical data to detect patterns or correlations that enable them predict future or unseen flats listings [2]. Moreover, it provides detailed price analysis and data on trends to help understand the features of the crowd, as well as property evaluations based on historical sales and pertinent characteristics [3].
Since deep learning is a branch of machine learning that can analyze large and complex data sets, it is also used in the prediction of prices of apartments. Deep learning can also learn complex patterns from raw input data [4]. In-order to upgrade the predictive accomplishment of models, deep neural networks—in particular, convolutional neural networks (CNNs) and artificial neural networks (ANNs)—have shown encouraging results in the extraction of spatial and temporal features from time-series data, textual descriptions, and apartment images [5].
This research will be an upgrade on the conventional valuation methodologies through the use of modern data analytics mechanism, machine learning, and deep learning techniques to extract patterns that affect flat pricing in major cities of Bangladesh. We grasp the range of parameters representing different areas, the square footage, the number of bedrooms and bathrooms, their proximity to educational and commercial hubs, and their historical market behavior. Moreover, the base of our study is a large dataset.

Analyzing this enormous dataset will help us achieve two objectives: it will be demonstrated in the first place, how one variable affects the price of the apartment and what is its relative importance; secondly, it will empower us to produce prediction models that make use of these variables together and, thus, get as much information as possible from them. The models of value prediction can do not only precise price forecasting, but by analyzing the influencing factors, they can offer useful inferences into property valuations in the dynamic real estate market of Bangladesh.

Finally, the design of this study exceeds just predicting the future trends of prices straightforwardly. The stakeholders receive the detailed description about market dynamics which helps them to get awareness about the changes in the market, spot emergence of new trends and to make right decisions about choosing property projects, investments and policymaking.

1.2 Background and Present State

Predicting apartment prices is an important portion of the real estate industry that affects investors, buyers, sellers, and legislators. Conventional approaches frequently depended on statistical models that were too simple and domain-specific to represent the complex, nonlinear relationships seen in real estate data. Machine learning and deep learning have become effective tools for making predictions that are more precise and efficient with the introduction of large data and sophisticated computational approaches. Hedonic pricing models, which took into account attributes like size, location, and amenities, were employed in earlier methods. However, these systems were constrained by linear assumptions and could not manage huge, complicated datasets. Though they were an improvement over earlier techniques, statistical models like multiple linear regression and time series analysis still had issues with nonlinearity and high-dimensional data.

The accuracy and dependability of apartment price prediction have increased dramatically with the integration of deep learning and machine learning. Decision trees, Random Forests, Gradient Boosting Machines, Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, K-Nearest Neighbors, and Long Short-Term Memory are some of the methods used. Cloud computing, AI frameworks, and big data are examples of technological developments. These models are used to generate real-time price predictions, pinpoint undervalued properties, and inform policies on housing affordability and urban growth in real estate platforms, investment strategies, and policy making. Governments and urban planners utilize these models to understand the dynamics of the housing market, while companies like Redfin and Zillow use them to improve user experience and market transparency. Subsequent investigations will concentrate on improving these models, adding fresh data sources, and guaranteeing moral and environmentally friendly methods in their use.

1.3 Problem Statement

Apartment price prediction is a challenging task because of the complex pricing dynamics impacted by factors like location, property size, amenities, market trends, and socioeconomic situations. These dynamics are evident in Bangladesh's various cities, including Dhaka, Chattogram, Khulna, Sylhet, Barisal, Narayanganj, Gazipur, Rajshahi, Comilla, Rangpur, and Mymensingh. By using advanced data-directed valuation approaches, specifically machine learning, to address the information gap and forecasting challenges in the complex pricing dynamics influenced by location, property size, amenities, market trends, and socioeconomic situations, the project seeks to bridge the gap between traditional valuation techniques and the complexities of the real estate market.

1.4 Objectives

• Develop dependable prediction models with machine learning to understand the intricacies of the Bangladeshi real estate market.

• Identify the main factors that affect pricing, such as size, location, facilities, and market trends.

• Enhance decision-making by providing stakeholders with useful data derived from prediction models.

• Increase market transparency by decreasing information asymmetry and providing a clear knowledge of price dynamics.

• Reduce investment risks by allowing investors and developers to analyze and reduce property investment hazards.

• Help shape policy by offering data-driven insights to guide policies on urban growth, housing initiatives, and regulatory frameworks.

• Demonstrate technological potential in real estate by proving the usefulness of machine learning and data analytics.

1.5 Scope and Limitations

The study's purpose regards to create a paradigm and employ a machine-learning method prior to estimate apartment selling prices in Bangladesh's different cities, like Dhaka, Chattogram, Khulna, Sylhet, Barisal, Narayanganj, Gazipur, Rajshahi, Comilla, Rangpur, and Mymensingh. Geography has a big impact on apartment costs; therefore, it's hard to get accurate pricing estimates. We can estimate the most likely price by paying greater attention to data noise. As a result, we will concentrate on many factors, such as several data source integrations along with suitable machine learning approaches, to provide higher accurate and dependable results. We would concentrate on location, bed, bath, and size to estimate apartment prices and would remove superfluous characteristics that would produce an uneven scenario.

Finally, the following are the criteria by which we will evaluate this proposal:
1. Create a dependable price prediction model.
2. Examine the prototype's predictions for accuracy.
3. Acknowledge the critical house rate characteristics so that they help the model forecast.
4. Use advanced regression algorithms to try to combine suitable patterns.

1.6 Report Organization

A thorough analysis of pertinent literature and earlier research is provided in the background chapter. An extensive grasp of the methods, theoretical underpinnings, and technical developments pertaining to deep learning, machine learning and apartment price prediction is provided in this section. It lays forth the background for the present investigation and identifies knowledge gaps.

Chapter 1: Introduction

The research is introduced in the first chapter, which also gives background information on the importance of predicting apartment prices, the use of deep learning and machine learning, and the necessity of comparing different prediction methods. It provides an overview of the study’s goals, and general arrangement.

Chapter 2: Background

A thorough analysis of pertinent literature and earlier research is provided in the background chapter. An extensive grasp of the methods, theoretical underpinnings, and technical developments pertaining to deep learning, machine learning and apartment price prediction is provided in this section. It lays forth the background for the present investigation and identifies knowledge gaps.

Chapter 3: Research Methodology

The strategy used to carry out the study is described in the research methodology chapter. It sheds light on the model architecture, data collection techniques, study design, and the reasoning behind particular methodology selections. The ethical issues and potential constraints on the investigation are also covered in this chapter.

Chapter 4: Experimental Result

The experimental findings from using deep learning and machine learning to the prediction of apartment prices are presented in this chapter. Comprehensive evaluations of the effectiveness of various prediction techniques are offered, accompanied by graphical depictions and statistical metrics to verify the results.

Chapter 5: Impact on Society

The chapter on social effect derives into the wider ramifications of the research findings concerning the forecasting of apartments prices. It talks about how better techniques for predicting prices might benefit buyers, sellers, broker and owners as well as the real estate industry as-a-whole. Future technological developments and their effects on society are taken-into account.

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research

The whole report is summarized in this last chapter. It provides recommendations for real world applications and more research in addition to summarizing the main results and restating the research's conclusions. The ramifications for further investigation are explored, offering a guide for scientists who want to build on the work that is being done now.

The reader is guided through the study process from the wider implications and suggestions by these organized arrangements. Which guarantees a logical flow of information. It makes it possible to have a thorough grasp of the study methodology and how it could affect the practical and scholarly applications of machine learning and deep learning for apartment price forecast.

1.7 Summary

Significant progress has been made and would make more progress with further study in the field of apartment price prediction using machine learning and deep learning, with the goal of providing reliable, transparent, and data-driven insights into the real estate market. Deep learning techniques, particularly neural networks, are being investigated and would influence researchers more in-order to capture subtle patterns in apartment pricing data and assess external factors such as economic indicators and demographic trends. Gaps in feature engineering, model interpretability, spatial-temporal integration, deep learning applications, and external factor analysis should be addressed in further researches.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

Predicting apartment prices requires both deep learning (DL) and machine learning (ML). In-order to generate predictions or judgements without the need for human interaction, machine learning algorithms construct mathematical models using training data or sample data. CNN, ANN and other deep learning algorithms can identify intricate patterns in vast volumes of data. In machine learning and deep learning, regression analysis is a statistical technique that is used to model the connection between one or more liberated factors and a dependent variable, in this example, the price of an apartment. Deep learning uses neural networks, which mirrored the human brain, to identify patterns and generate predictions. The act of choosing, altering, and producing variables from unprocessed data in-order to increase model accuracy is known as feature engineering. An input-output pair dataset is called training data when it is used to train ML or DL models. In-order to adjust model parameters and avoid overfitting, test and validation data are utilized. When a model learns the training data too well, or when it is too easy to capture underlying patterns in the data, overfitting and underfitting are the problems that result. Cross validation is a method for assessing machine learning models, where hyperparameters specify the model architecture and learning process. A dimensionality reduction method called Principal Component Analysis (PCA) aids in lowering computing complexity and enhancing model performance. Several decision trees are used in the Random Forest ensemble learning technique to increase prediction accuracy and reduce overfitting.

2.2 Related Works

There are quite a few studies regarding apartment price prediction in Bangladesh, especially focusing on Dhaka Megacity, as it is the ninth largest populated city in the world. Most people live in Dhaka because of its better facilities in terms of jobs, education, business, and living. Mentioning the related previously done studies are:

F. M. Rafi [6] study is about the prediction and analysis of house prices in Dhaka city using advanced regression techniques like lasso, bayesian, ridge, gradient boost, and so on in 2023. Of all these methods, Gradient Boost provided the highest accuracy.

M. A. Hossen and Md. M. Rahman [7] studied the prediction of house rent in Dhaka city utilizing different machine learning algorithms in 2023. They used linear regression, ridge regression, bayesian regression, and so on. From the advanced techniques mentioned, ridge regression showed its performance to be the highest.

A. Begum et al. [8] researched residence and predicted the price using a machine learning algorithm in 2022. They researched using decision trees, random forests, and linear regression. Random Forest performed best among the mentioned methods.

K. Xu and H. Nguyen [9] studied the prediction of housing prices and analyzed the real estate market in Chicago suburbs using machine learning algorithms in 2022. They employed linear regression, support vector regression, random forest regression, and so on. XGBoost's performance appeared to be the best among them.

J. J. Jui et al. [10] used linear regression and random forest regression in their research on flat price prediction in 2020. Random Forest gave the highest accuracy between the two machine learning algorithms.

In 2020, M. Modi et al. [11] used a variety of machine learning approaches to estimate housing prices, including logistic regression, SVM, naive bayes, stochastic gradient descent, additional trees, KNN, and voting classifiers. During implementation, the voting classifier had the greatest accuracy rate.

I. N. Y. Saputra et al. [12] predicted the apartment price record in Indonesia by analyzing random forest, multiple regression, and backpropagation in 2021. After implementation, backpropagation gave the best result.

In 2020, S. B. Jha et al. [13] studied predicting house market problems using different machine learning algorithms, like CatBoost, random forest, lasso, XGBoost, voting regressor, linear regressor, and support vector regressor. Among the mentioned methods, XGBoost gave the best result.

S. Özöğür-Akyüz et al. [14] studied a novel house price prediction model implementing linear regression, ridge regression, K-means, KNN, NuSVR, and hybrid NuSVR in 2020. After implementation, the hybrid NuSVR gave the best result.

D. [15] studied house price prediction employing machine learning algorithms using decision tree regression and KNN in 2023. The decision tree regressor proved to be the best predictor.

M. Shahhosseini et al. [16] approach the optimization for ensemble generation proposed in this paper maximizes the mean squared error of base learners. It was evaluated using seven algorithms (Lasso, Random Forests, Neural Network, XGBoost, SVM (poly), SVM (RBF), SVM (sigmoid)) on housing datasets from Ames and Boston. Appliable to several datasets, the optimized ensemble beat the benchmark ensemble and all individual base learners.

J. Kim et al. [17] in order to predict 52,900 land values in Seoul, South Korea, from January 2017 to December 2020, this study used the Random Forest and XGBoost algorithms. The models were trained over a range of time periods and land usage. In residential regions, Random Forest demonstrated more accuracy in predictions than XGBoost. The report recommends more research that takes into account submarkets based on location and price volatility. Although the study is unable to conclude which of the two ensemble models is better, the limited findings point to the need for more research that takes into account other aspects of real estate localities.

M. Yue et al. [18] using an internet platform, 33224 pieces of Chengdu housing rental data are analyzed in this article. Tenant demand for small-area apartments and joint rentals is greater, according to this data. Next, models like Random Forest Regressor, XGBoost, and LightGBM are used to examine the data. With its 0.85 forecast accuracy, the XGBoost model offers the best value as a benchmark for rental home costs. Ninety-six percent of rental areas are under $30m^2$ for shared rental, suggesting a stronger demand for properties under $90m^2$. With its superior prediction effects, the XGBoost model advances Chengdu's study on housing rent prediction.

A. Oust et al. [19] in order to value residential properties, the study investigates the application of hedonic regression (regression kriging, mixed regressive-spatial autoregressive, and geographically weighted regression) and repeat sales in real estate analysis. It makes use of a dataset including 16,417 Oslo, Norway, real estate transactions. The findings demonstrate that adding repeat sales increases the median absolute percentage inaccuracy of hedonic models by 6.8–9.5%, suggesting that repeat sales forecasts might give automated valuation algorithms useful geographical data.

A. Deaconu et al. [20] In order to estimate the prices of 900 apartments in Cluj-Napoca, Romania, the study compares the Generalized Linear Model (GLM) with Artificial Neural Networking (ANN). The findings indicate that ANN performs better than GLM in stabilizing and predicting selling prices. In response to concerns regarding transparency, ANN additionally employs relevant statistical indicators to highlight the importance of real estate qualities. For the purpose of real estate appraisal, these results may be helpful.

F. Mostafi et al. [21] study uses 1381 real estate price data from Trabzon, Turkey, to investigate the impact of skewness on a deep neural network model's ability to forecast prices accurately. The square root transformation, cube root transformation, and logarithmic transformation were the three skewness reduction methods applied. The findings indicated that whereas SRT and LT led to high error levels and overfitting problems, CRT enhanced prediction accuracy and computing time.

M. Štubňová et al. [22] precise real estate property pricing assessment for the banking and insurance industries is covered in the paper. Using data from Nitra, Slovakia, it contrasts conventional and cutting-edge approaches and shows how artificial neural networks trained with the Levenberg-Marquart, Bayesian Regularization, and Scaled Conjugate Gradient learning algorithms greatly increase prediction accuracy in estimating residential property market prices. According to the study's findings, these techniques can greatly improve real estate market price estimation.

2.3 Comparison between existing works

TABLE 2.3.1: Comparative Analysis of Existing Work

SL No.	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	F. M. Rafi [6]	Lasso, Bayesian Ridge, Ridge, XGBoost, KNN, ANN, Random Forest, Support Vector, LightGBM, Catboost, Gradient Boost	Gradient Boost: 92.3%
2.	M. A. Hossen and Md. M. Rahman [7]	Linear Regression, Ridge Regression, Bayesian Regression, and Lasso Regression	Ridge Regression: 91.54%
3.	A. Begum et al. [8]	Decision Tree, Random Forest, and Linear Regression	Random Forest: 94.4%
4.	K. Xu and H. Nguyen [9]	Linear Regression, Support Vector Regression, Random Forest Regression, Decision Trees, XGBoost Regression	XGBoost Regression: 92.5%
5.	J. J. Jui et al. [10]	Linear Regression, Random Forest Regression	Random Forest Regression: 93%
6.	S. B. Jha et al. [13]	CatBoost, Random Forest, Lasso, XGBoost, Voting Regressor, Linear Regressor, and Support Vector Regressor	XGBoost: 97%
7.	S. Özöğür-Akyüz et al. [14]	Linear Regression, Ridge regression, K-means, KNN, NuSVR, Hybrid NuSVR	Hybrid NuSVR: 77.9%
8	D. [15]	Decision Tree Regressor, KNN	Decision Tree Regressor: 89%
9	M. Shahhosseini et al. [16]	Lasso, Random Forests, Neural Network, XGBoost, SVM (poly), SVM (RBF), SVM (sigmoid)	Random Forest: 92%
10	J. Kim et al. [17]	Random Forest and XGBoost	XGBoost: 87%
11	M. Yue et al. [18]	Random Forest Regressor, XGBoost, and LightGBM	XGBoost: 85%
12	A. Oust et al. [19]	Enhanced Hedonic Regression – Ordinary regression, Regression Kriging, Mixed Regressive-Spatial Autoregressive, and Geographically Weighted Regression	Ordinary Regression: 75%

13	A. Deaconu et al. [20]	Generalized Linear Model (GLM) Artificial Neural Networking (ANN)	ANN: 80%
14	F. Mostafi et al. [21]	DNN, PCA-DNN	PCA-DNN: 85%
15	M. Štubňová et al. [22]	ANN	ANN: 97%

2.4 Open Issues

• Missing data, conflicting data sources, and feature engineering are examples of open problems.

• Models do not perform comparably, as studies differ in how they tune hyperparameters and use evaluation criteria.

• Interpretability and transparency of the model are important, and methods such as SHAP and LIME are still in the early phases of development.

• The persistence of overfitting and geographical transferability necessitates the use of techniques like dropout layers, cross-validation, and regularization.

• There is a lack of developed integration with other data sources.

• Bias, privacy, and environmental concerns are among the ethical and social ramifications.

• Deep learning model training may be expensive and computationally demanding, requiring technology and techniques that use less energy.

• Real-world application and scalability are crucial, with difficulties in system integration, real-time prediction, and model maintenance.

• The success of the model depends on user acceptability.

2.5 Summary

One of the most significant aspects of the immovable property industry is estimating apartment prices. It is frequently difficult to anticipate a true price because pricing ranges vary from location to the area. Many fraudulent real estate agents take advantage of this chance to defraud consumers. As a result, this research was performed in-order to reckon an estimated price for the purchaser. Data concerning study was gathered from a variety of sources. Previous research had many gaps, such as a lack of data, which we will attempt to fill up in our study. We will use many strategies for cleaning and combining the data. Many more sophisticated regression techniques will be used by us to predict the anticipated pricing. By doing this investigation and collecting data, we discovered that, despite having identical bed and bath amenities, there is a significant difference across areas.

CHAPTER 3
METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

In-order to predict apartment prices, the study topic “Apartment Price Prediction of Bangladesh using Machine Learning and Deep Learning Algorithms” focuses on applying sophisticated artificial intelligence techniques to different features regarding apartment like location, size and other amenities. Apartments are human habitats that are necessary for contemporary life, give security, promote communal living, and provide a host of other advantages including meeting practical, social, environmental, and economic demands. The study explores the drawbacks of conventional apartment prediction techniques and looks for ways to use deep learning and machine learning techniques to get over these restrictions.

Artificial intelligence’s newest idea, machine learning, is used to apply information from unrelated activities to the difficult field of price prediction. Deep learning techniques, which are well-known for their effectiveness in prediction with numerous characteristics, are used to extract complex patterns, which help to make predictions that are more precise and nuanced.

The research’s comparative analysis section examines the variations and complementarities among different prediction techniques. The process of prediction includes determining variances in terms of location, size and surrounding environment, as well as the existence of rooms, building condition, and facilities like security, an elevator, and a lending facility. Though a close examination of various approaches, the study hopes to provide a thorough grasp of their advantages and disadvantages, assisting in the creation of more reliable and efficient predictive tools.

The study topic basically focuses on the confluence of artificial intelligence, cost, location, size and other amenities that are given, with an emphasis on improving pricing prediction via the creative use of deep learning and machine learning. The results are intended to add to the scholarly conversation as well as the real estate industry’s practical environment, with the goal of perhaps increasing forecast accuracy and decreasing fraudulent and overpriced transactions.

3.2 Proposed Methodology/ System Design

The first stage of the proposed framework is the collection of data for better study, followed by the preprocessing of the data, the selection of features, the development of the model, model evaluation and lastly the result.

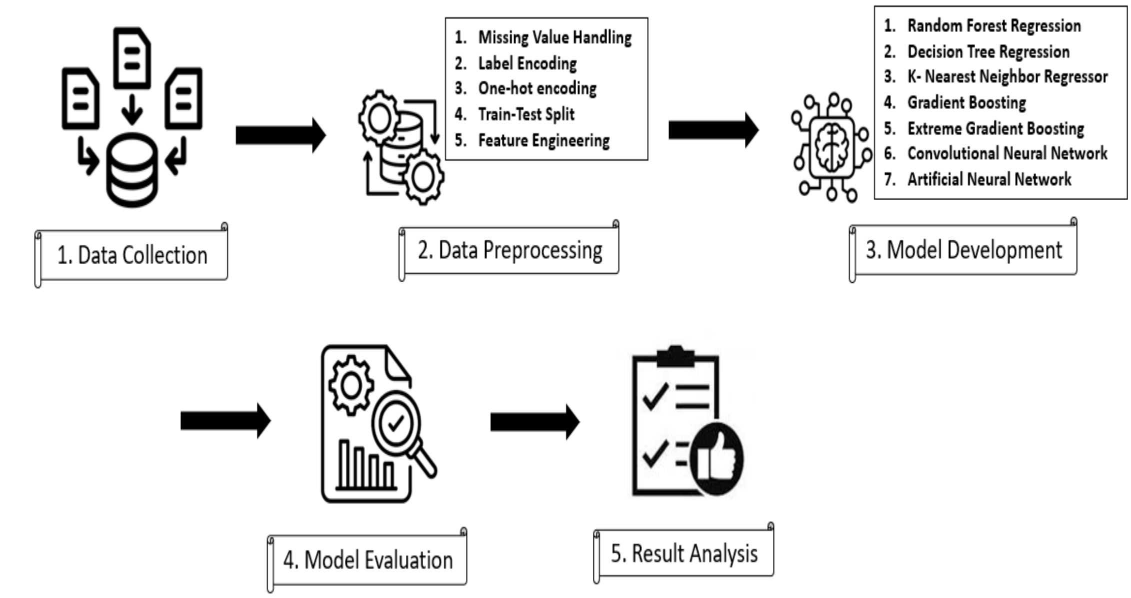

Figure 3.2.1: Flowchart of Proposed Framework

1. Data Collection Procedure

This study makes use of both original data collected manually from numerous real estate firms and data scraped from well-known real estate websites (e.g., bproperty.com, bikroy.com). The collection includes the apartment description for Bangladesh's megacity, Dhaka, as well as neighboring cities such as Chattogram, Khulna, Sylhet, Barisal, Narayanganj, Gazipur, Rajshahi, Cumilla, Rangpur, Mymensingh, and so on. We collected 8255 data in total.

Location	Region	Total (Sqft	Bedroom	Bathroom	Balcony	Drawing ar	Kitchen	Servant ro	Condition	Parking spa	Elevator (L	Electricity	Loan facili	Installmen	Price (BDT)
Firingee B	Chattogra	965	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5200000
Mohakhal	Dhaka	775	2	1	1	Yes	Yes	No	Complete	No	No	Yes	No	No	5000000
Mohakhal	Dhaka	850	2	1	1	Yes	Yes	No	Complete	No	No	Yes	No	No	5200000
North Pah	Chattogra	815	2	2	1	Yes	Yes	No	Complete	No	No	Yes	No	No	3700000
North Pah	Chattogra	850	2	2	1	Yes	Yes	No	Complete	No	No	Yes	No	No	4400000
North Pah	Chattogra	900	2	2	1	Yes	Yes	No	Complete	No	No	Yes	No	No	4500000
North Pah	Chattogra	1025	2	2	1	Yes	Yes	No	Complete	Yes	Yes	Yes	No	No	4500000
North Pah	Chattogra	1050	2	2	1	Yes	Yes	No	Complete	Yes	Yes	Yes	No	No	4600000
North Pah	Chattogra	1052	2	2	1	Yes	Yes	No	Complete	No	No	Yes	No	No	4700000
North Pah	Chattogra	1110	2	2	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	4700000
North Pah	Chattogra	1150	2	2	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5000000
North Pah	Chattogra	1175	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5200000
North Pah	Chattogra	1192	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5200000
North Pah	Chattogra	1200	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5600000
North Pah	Chattogra	1200	3	3	2	Yes	Yes	No	Complete	No	No	Yes	No	No	5600000
North Pah	Chattogra	1200	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5600000
North Pah	Chattogra	1200	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5600000
North Pah	Chattogra	1200	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5600000
North Pah	Chattogra	1210	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5875000
North Pah	Chattogra	1210	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	5875000
North Pah	Chattogra	1240	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6000000
North Pah	Chattogra	1250	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6200000
North Pah	Chattogra	1250	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6200000
North Pah	Chattogra	1250	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6200000
North Pah	Chattogra	1250	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6200000
North Pah	Chattogra	1250	3	3	2	Yes	Yes	No	Complete	Yes	Yes	Yes	Yes	Yes	6200000

Figure 3.2.2: Dataset Sample

Our primary dataset has 16 features in total and they are location, the measurement of an apartment in sq ft, the quantity of bedrooms, bathrooms, balconies, the presence of drawing and dining, a kitchen, the condition of the building, the availability of different amenities like a servant room, an elevator, parking space, a generator for electricity backup, a loan, and an installment facility to increase the buyer's opportunity of purchasing an apartment with the price.

2. Preprocessing

Since the dataset we gathered is statistical in nature, it must go through a few preprocessing steps in-order to be prepared for model maintenance. There may be some mistakes in the dataset because it is being collected manually. We could apply various preprocessing methods to fix those mistakes. They are as follows:

- **Missing Value Handling:** This step checks for missing values in the dataset and deals with missing values [23].
- **Label Encoder:** An effective preprocessing method for ordinal categorical data is label encoding. It provides a simple method for transforming category data into numerical data representations that machine learning algorithms may use [24].
- **One-Hot Encoding for multi-class categorical variables:** One-hot encoding turns multi-class categorical data into a binary matrix where each category is represented by a unique binary vector. Create a matrix with categories in columns and data points in rows after creating a list of each unique category. Set all other columns to 0 and the column corresponding to the category of each data item to 1. This results in a matrix where each category is represented by a unique set of binary values, enabling the effective handling of categorical data by algorithms [25].

- **Train-Test Split:** To evaluate the performance of the model, separate the dataset into training and test sets. Typically, 20% is used for testing and 80% is for training.

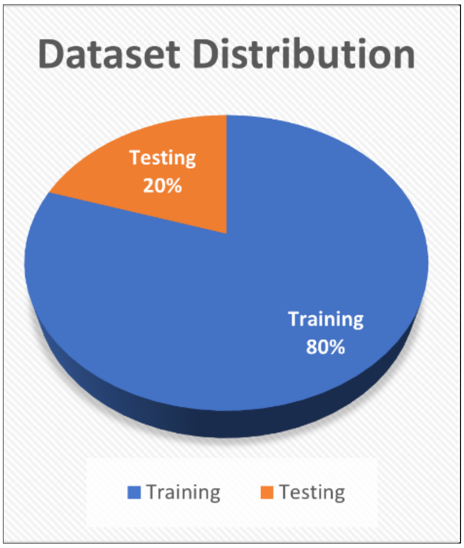

Figure 3.2.3: Dataset Distribution

- **Feature Engineering:** A lot of machine learning techniques depends on feature engineering, and standardization in particular. We may normalize features by abolishing the mean and scaling to unit variance using StandardScaler from sklearn.preprocessing.

3. Model Development

A critical step in developing a machine learning and deep learning system is model construction, which entails choosing, training, and assessing models to guarantee performance on unknown data. These models will be put into practice, the dataset will be divided for training and testing, and the selected methods will be selected.

- **Random Forest**

Random Forest is a popular ensemble learning method in machine learning that may be used for both classification and regression issues. It creates a lot of decision trees in the training phase, and at the conclusion, it produces a class that is the average prediction (regression) or the class mode (classification) of every single tree [26].

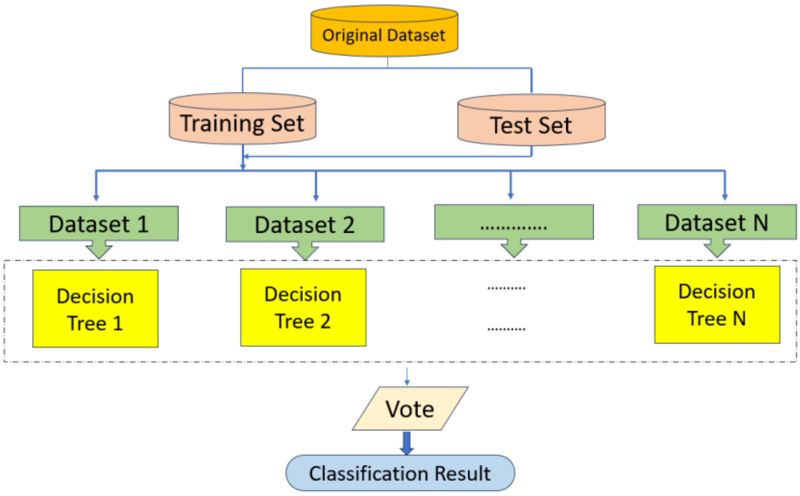

Figure 3.2.4: Architecture of Random Forest Regressor

❖ Decision Tree Regressor

One well-liked machine learning approach for both regression and classification applications is the decision tree. In-order to operate, the dataset is divided into subsets according to the input feature values. This creates a tree structure, with each leaf node representing the result, each branch representing a decision rule, and each node representing a feature [27].

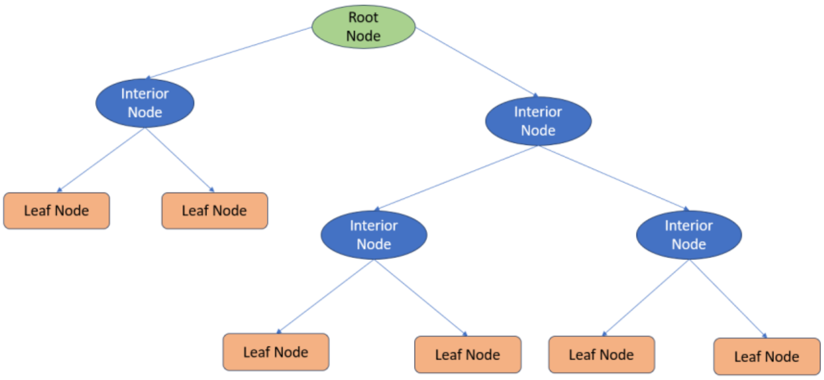

Figure 3.2.5: Architecture of Decision Tree Regressor

❖ K-Nearest Neighbor Regressor

For both classification and regression applications, the K-Nearest Neighbors (KNN) method is a effortless yet effective machine learning technique. By averaging the values of the k-nearest neighbors in the feature space, KNN predicts the value of a target variable in regression. As a non-parametric, lazy learning method, it only processes data during the prediction stage and makes no assumptions about the distribution of the data [28].

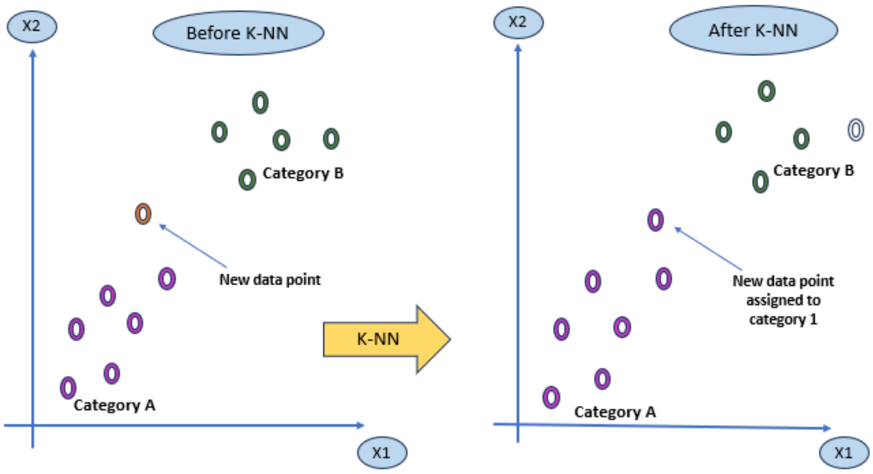

Figure 3.2.6: Architecture of K-NN Regressor

❖ Gradient Boosting:

A potent ensemble machine learning method for classification and regression applications is gradient boosting. It develops models one after the other, trying with every iteration to fix the mistakes of the earlier models. To create a strong learner, the technique entails integrating the outputs of several weak learners, usually decision trees [29].

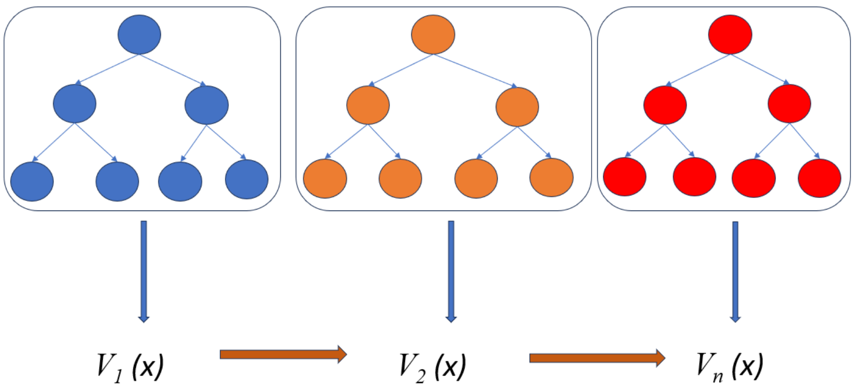

Figure 3.2.7: Architecture of Gradient Boosting

❖ Extreme Gradient Boosting:

XGBoosting, short for Extreme Gradient Boosting, is a powerful machine learning algorithm that belongs to the family of ensemble learning techniques. Its foundation is gradient boosting, which builds an ensemble of weak learners (typically decision trees) gradually by having each new learner correct the errors of the prior one [30].

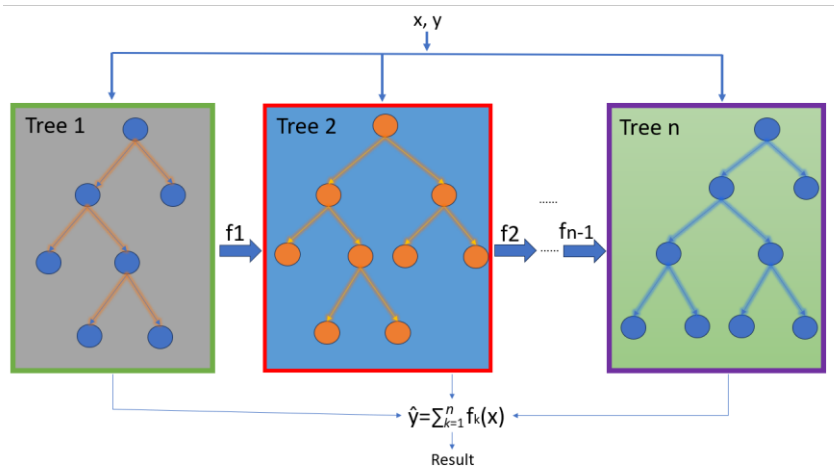

Figure 3.2.8: Architecture of XGBoosting

❖ CNN

A family of deep learning methods called convolutional neural networks, or CNNs, are generally used to picture data. They may, however, be modified for tabular data, such as apartment price prediction, with the appropriate preprocessing and feature engineering. Because CNNs can automatically identify and learn patterns and characteristics from incoming data, they are very powerful [31].

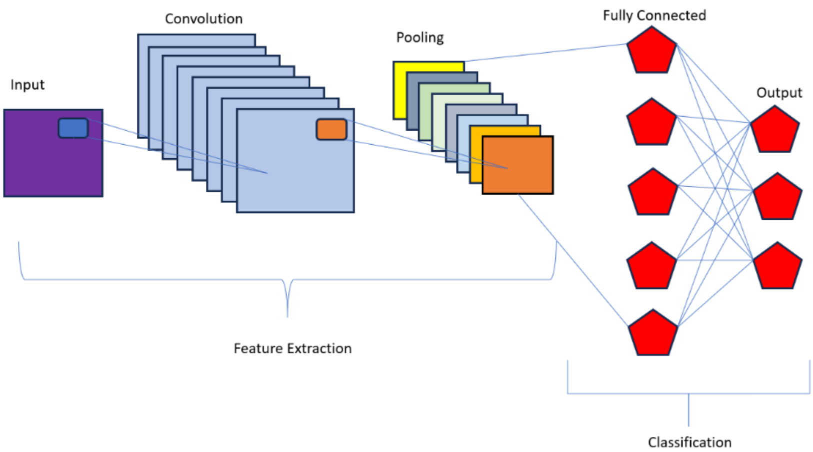

Figure 3.2.9: Architecture of CNN

❖ ANN

The foundation of deep learning is Artificial Neural Networks (ANNs), which use numerous layers of linked neurons to capture complex patterns in data. ANNs can be quite useful for predicting apartment prices since they can learn from both category and numerical input [32].

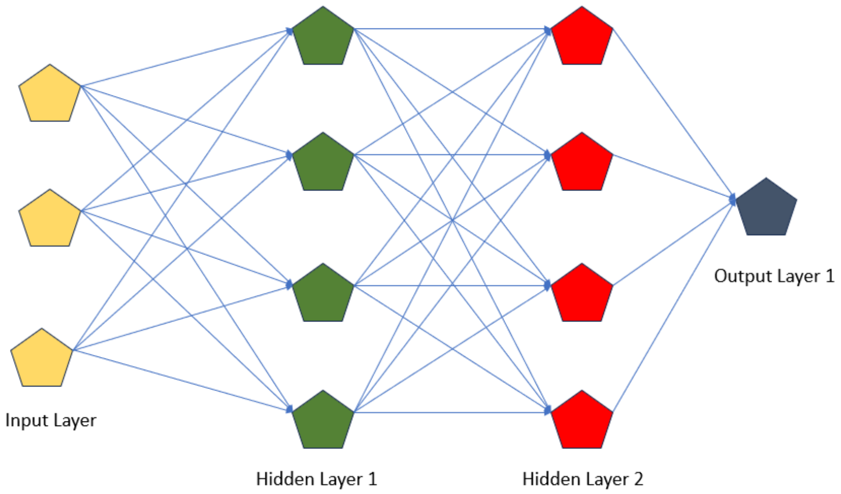

Figure 3.2.10: Architecture of ANN

3.3 Hardware/ Software Requirement

There are particular prerequisites for the implementation of the research on "Apartment Price Prediction of Bangladesh using Machine Learning and Deep Learning Algorithms" in-order to guarantee the study's effective completion. The criteria for implementation cover aspects pertaining to data gathering and model assessment, in addition to hardware and software components.

3.3.1 Background Study

- Review previous or related works to identify gaps and conduct a thorough analysis.

3.3.2 System Requirements

1. Data Collection
2. Feature Selection
3. Data Processing
4. Model Development
5. Feature Engineering
6. Model Evaluation
7. Real-World Applicability
8. Ethical Considerations

3.3.3 Tools for Data Collection

1. **Google Sheets or Excel**
    - These spreadsheet tools provide advanced sorting, filtering, and preliminary analysis capabilities.

3.3.4 Resources

1. **Programming Languages**
    - Python, due to its versatility and widespread use in machine learning and data science.
    - Key libraries include NumPy, Pandas, Scikit-Learn, TensorFlow, Matplotlib, and Seaborn.
    - Development environments such as Jupyter Notebook or Google Colaboratory.
2. **Hardware Requirements**
    - **Processor (CPU):** A multi-core processor to handle computational tasks efficiently.
    - **Memory (RAM):** Sufficient RAM to manage large datasets and facilitate smooth processing.
    - **Storage:** Ample storage capacity to store datasets, intermediate files, and model outputs.
    - **Graphics Processing Unit (GPU):** A graphics processing unit to accelerate deep learning and computationally intensive tasks.
3. **Software Requirements**
    - **Jupyter Notebook or Google Colaboratory:** Allows users to write and execute Python code**.**
    - **Draw.io:** Online diagramming tool for creating flowcharts, diagrams, and more.

3.4 Project Management and Financial Analysis

3.4.1 Project Management

3.4.1.1 Project Objectives:

- Develop dependable prediction models with machine learning to understand the intricacies of the Bangladeshi real estate market.
- Identify the main factors that affect pricing, such as size, location, facilities, and market trends.
- Enhance decision-making by providing stakeholders with useful data derived from prediction models.

- Increase market transparency by decreasing information asymmetry and providing a clear knowledge of price dynamics.

3.4.1.2 Project Timeline:

We have collected a dataset of 8255 successfully and completed the whole implementation as well. The following four-month schedule phases have been passed and involve:

A. Month 1-2: Model Development and Preliminary Evaluation

- **Week 2-4:** Based on the data that is currently accessible, we chose machine learning (RF Regressor, DT Regressor, KNN Regressor, GBoost Regressor and XGBoost Regressor) models and deep learning (CNN and ANN) models. Feature engineering, baseline model implementation, and performance evaluation via cross-validation took place during this period.

B. Month 3: Feature Engineering and Model Refinement

- **Week 1-2:** We explored feature engineering in greater detail and tried out various transformations and combinations in-order to enhance our models. The models were fine-tuned, hyperparameters optimized, and ensemble approaches investigated.

C. Month 4: Evaluations, Reporting and Presentation Preparation

- **Week 3-4:** The updated models had been thoroughly assessed using test datasets and, outside validation. After that, we compiled the results of our investigation, wrote a comprehensive report, and started working on a presentation.

3.4.1.3 Risk Management:

SWOT analysis typically involves identifying the Strengths, Weaknesses, Opportunities, and Threats associated with a particular situation or decision. Here is a SWOT analysis of apartment price prediction of Bangladesh using machine learning (ML) and deep learning (DL) algorithms:

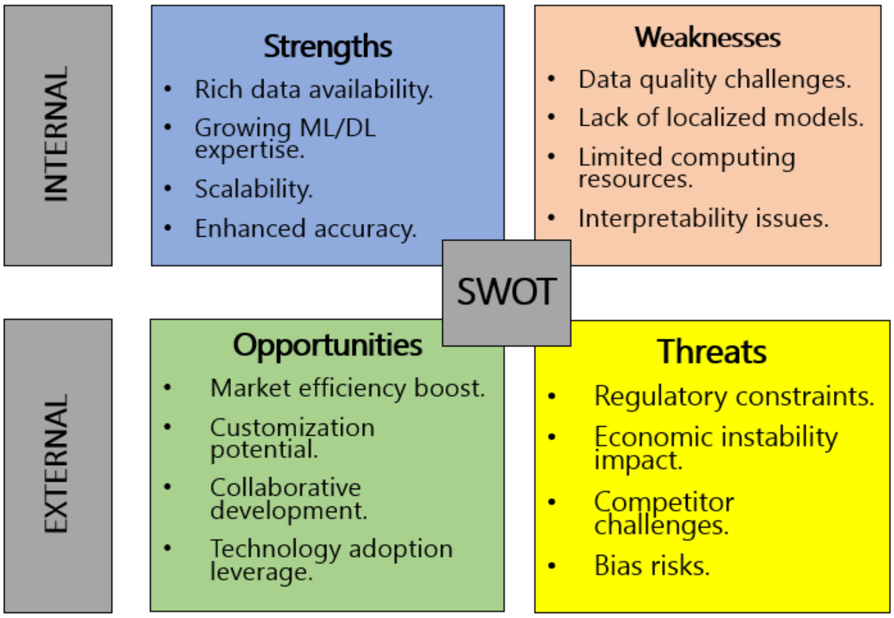

Figure 3.4.1.3.1: SWOT Analysis

3.4.2 Financial Analysis

Real-world issues are the foundation of our project. gathering information from several real estate companies, meeting with clients and subject matter experts, and then carrying out and launching the completed project. The cost of all these measures is money. As a result, the following is the financial analysis for this project:

TABLE 3.4.2.1: Estimated Cost for Machine Learning and Deep Learning Based Project

SL. No.	Components	Estimated Cost (BDT)
01	Hardware and Tools	80,000 - 90,000 BDT
02	Visiting Stakeholders	3000 - 4000 BDT
03	Software and Tools	4000 - 5000 BDT
04	Data Collection and Processing	500 - 1000 BDT
05	Documentation and Report Writing	1500 - 2000 BDT
06	Miscellaneous	1000 - 1500 BDT
07	Contingency (10% of total)	9000-10350 BDT
	Total Estimated Cost	99000 - 1,13,850 BDT

3.5 Summary

To produce accurate and practical insights in the changing real estate market, the apartment price forecast approach blends rigorous data processing, innovative modeling tools, and excellent project management. Data collection is done properly to obtain the highest accuracy after following all the steps of system design.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

Model evaluation is critical in determining how well different regression models predict apartment prices. Important metrics such as MAE, MSE, RMSE, and R^2 are used to gauge the performance of various models including Random Forest Regressor, Decision Tree Regressor, K-Nearest Neighbors Regressor, Gradient Boosting Regressor, XGBoost Regressor, CNN, and ANN.

4.2 Train Model/ Prototype Design

Data Loading and Preprocessing:

- ❖ Load the data from a CSV file.
- ❖ Handle missing values and data types.
- ❖ Encode binary categorical columns using LabelEncoder.
- ❖ Create dummy variables for categorical columns using One-hot Encoder.
- ❖ Separate the target variable (Price (BDT)) and features.
- ❖ Split the data into training and testing sets.
- ❖ Scale the feature values using StandardScaler.

Model Training and Evaluation:

- ❖ Train and evaluate the Random Forest Regressor:
  - ▪ Fit the model on scaled training data.
  - ▪ Predict and calculate performance metrics (MAE, MSE, RMSE, R^2).
- ❖ Train and evaluate the Decision Tree Regressor with similar steps.
- ❖ Train and evaluate the K-Neighbors Regressor.
- ❖ Train and evaluate the Gradient Boosting Regression model.
- ❖ Train and evaluate the Extreme Gradient Boosting Regression model.
- ❖ Train and evaluate a simple Artificial Neural Network (ANN) using TensorFlow:
  - ▪ Build the ANN model.

- Compile and fit the model.
- Predict and calculate performance metrics.
- Plot the training and validation loss.

❖ Train and evaluate a Convolutional Neural Network (CNN) model using TensorFlow:

- Reshape the data.
- Build the CNN model.
- Compile and fit the model.
- Predict and calculate performance metrics.
- Plot the training and validation loss.

4.3 System Testing/ Model Evaluation

An essential first step in machine learning and deep learning initiatives is model assessment. It guarantees that your model generalizes properly to new cases and helps you assess how well it works on data that hasn't been seen before. In this instance, our evaluation will center on how well RF Regressor, DT Regressor, KNN Regressor, GBoosting Regressor, XGBoosting Regressor, CNN and ANN models predict apartment prices. Mean Absolute Error (MAE), Mean Squared Error (MSE), and Accuracy (R^2) are important metrics for regression issues.

➢ **Accuracy (R^2):**

The percentage of the variation for the dependent variable that can be accounted for by the independent variables in a regression model is known as R-Squared (R^2), a statistical metric. It shows how effectively the independent factors account for the dependent variable's variability.

$$R^2 = \frac{\text{RegSS}}{\text{TSS}} = \frac{\sum_i \left(\hat{y}_i - \bar{y}\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \quad \text{---(1)}$$

➢ **Mean Absolute Error (MAE):**

An extensively used statistic for assessing regression models is the Mean Absolute Error (MAE). Without taking into an account the direction of the mistakes, it calculates the average size of the errors in a series of forecasts. The average absolute difference between the actual and anticipated values is computed. It's a simple metric that gives an intelligible assessment of model performance.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \quad \text{---(2)}$$

- **Mean Squared Error (MSE):**

Regression model performance is commonly assessed using the Mean Squared Error (MSE) measure. The average squared difference between the actual and anticipated values is what's measured. MSE is susceptible to outliers since it assigns a higher weight to bigger mistakes.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{---(3)}$$

- **Root Mean Squared Error (RMSE):**

Regression models are frequently assessed using the Root Mean Squared Error (RMSE) metric, particularly when Gaussian errors are anticipated. The square root of the average of the squared discrepancies between the expected and actual values is measured by RMSE. Much like MSE, but in the original units of the target variable, it assigns a higher weight to bigger mistakes resulting from the square root operation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \quad \text{---(4)}$$

4.4 Summary

This section emphasizes the importance of rigorous model evaluation to ensure that models generalize well to new data. Various regression models are trained and evaluated using key metrics like MAE, MSE, RMSE, and R^2, providing insights into their predictive performance. Visualization of training and validation losses for ANN and CNN models aids in understanding their learning processes and potential overfitting issues.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

Use an organized method while setting up tests to estimate apartment prices using machine learning. Establish the goals and study topic, compile a large dataset, clean it, and choose predictive characteristics. Create a baseline model, use cross-validation strategies, divide the dataset into training and testing sets, and design the experimental setup. Pick suitable machine learning algorithms and performance measures such as R-squared, MAE, MSE, and RMSE. Carry out the experiment through training, assessment, and parameter tweaking. Use statistical tests to analyze the data and interpret the findings in-light of the study's goals. Record the specifics of the experimental setup and write an extensive report that includes a summary of the results, conclusions, and suggestions. Talk about potential paths and restrictions, including restrictions on the dataset, model presumptions, and outside variables influencing the dynamics of apartment prices. In real estate applications, this systematic arrangement guarantees accurate findings and well-informed decision-making.

5.2 Experimental/ Simulation Result

In this study, we have employed a total of five machine learning algorithms; Random Forest, Decision Tree, KNN, Gradient Boosting and Extreme Gradient Boosting. We calculated the evaluation metrics using k fold cross validation techniques and calculated the R2 Score, MAE, MSE and RMSE for each of the algorithms. we also have employed two deep learning algorithms; Artificial Neural Network(ANN) and Convolutional Neural Network(CNN).

TABLE 5.2.1: Experimental Result of all Models

Models	Accuracy (R^2)	MAE	MSE	RMSE
Random Forest Regressor	96.15%	444135.66	1709094993842.77	1307323.59
Decision Tree Regressor	96.68%	385739.13	1474762190850.35	1214397.87
K-Nearest Neighbor Regressor	89.73%	726504.65	4548904387007.87	2132816.06
Gradient Boosting Regressor	94.21%	940349.16	2564647872111.46	1601451.80
Extreme Gradient Boosting Regressor	97.53%	506740.81	1091503837228.54	1044750.61
Convolutional Neural Network (CNN)	92.34%	1218796.96	5522033750524.54	2349900.79
Artificial Neural Network (ANN)	87.55%	1115045.06	3394257327499.21	1842351.03

Based on their performance metrics—Accuracy (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE)—the table presents the results of several regression models.

Achieving an accuracy of 96.15%, the **Random Forest Regressor** has an MAE of 444135.66, an MSE of 1709094993842.77, and an RMSE of 130732.59. At 96.68% accuracy, the **Decision Tree Regressor** performs somewhat better than the other models. Its error metrics are lower, with an MAE of 385739.13, an MSE of 1474762190850.35, and an RMSE of 121439.87.

In comparison to the preceding models, the **K-Nearest Neighbor Regressor** performs worse, as evidenced by its lower accuracy of 89.73% and higher error values of 726504.65, 4548904387007.87, and 213281.06. Although the **Gradient Boosting Regressor** has a 94.21% accuracy rate, it has higher error metrics than the **Random Forest and Decision Tree Regressors** (MAE: 940349.16, MSE: 2564647872111.46, RMSE: 160145.80).

The top performing model overall is the **Extreme Gradient Boosting Regressor**, which stands out with the highest accuracy of 97.53% and lower error values: MAE of 506740.81, MSE of 1091503837228.54, and RMSE of 104475.61. With a 92.34% accuracy rate, the **Convolutional Neural Network (CNN)** performs moderately when

compared to the best-performing models, as seen by its higher error measures (MAE: 1218796.96, MSE: 5522033750524.54, RMSE: 234990.79).
Last but not least, among the models given, the **Artificial Neural Network (ANN)** performs the least well, with the lowest accuracy of 87.55% and quite high error values: MAE of 1115045.06, MSE of 3394257327499.21, and RMSE of 184235.03.

In conclusion, the Artificial Neural Network (ANN) performs the worst overall, while the Extreme Gradient Boosting Regressor has the greatest overall performance with the highest accuracy and lowest error metrics.

Training and Validation Loss for Deep Learning Models:

Represents the training and validation loss of the deep learning models, with loss percentages on the x-axis and the number of epochs on the y-axis. In the figure, training and validation data are adequately divided, and there is no overfitting.

Convolutional Neural Network (CNN)

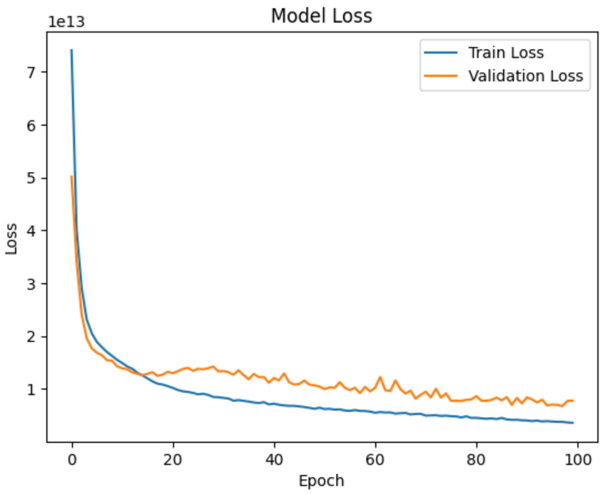

Figure 5.2.2: Training and validation loss of CNN

The training and validation loss of a model during 100 epochs is displayed in Fig. 5.3.2. Both losses are large at beginning, but within the first 20 epochs, the training loss drops off quickly, suggesting swift learning. Initially, the validation loss likewise declines, but it begins to vary at Epoch 20, which may indicate overfitting. The validation loss varies but usually follows the training loss pattern with somewhat larger values from Epoch 20 to Epoch 60, while the training loss continues to diminish more slowly.

The training loss stabilizes in the latter phase (Epoch 60 to 100), indicating better performance on training data. Nonetheless, a little rising trend in the validation loss stabilizes, indicating some overfitting. All things considered; the model learns well.

Artificial Neural Network (ANN)

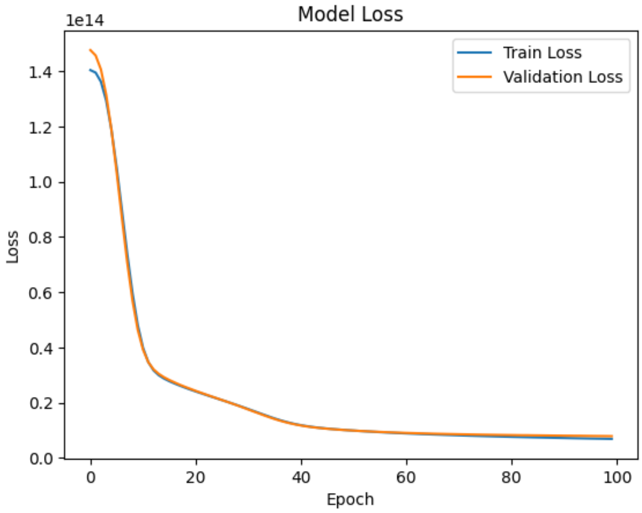

Figure 5.2.3: Training and validation loss of ANN

The training and validation loss of an artificial neural network (ANN) across 100 epochs is shown in Fig. 5.3.3. At the beginning, there is a large amount of mistake shown by the very high training and validation losses. But these losses drop very quickly in the first 20 epochs, showing that the model is picking up new skills and becoming more efficient. The losses decline, albeit more slowly, from Epoch 20 until Epoch 60, when they start to stabilize. The training and validation losses level out in the last phase, which spans Epochs 60 to 100.

This suggests that the model's performance has stabilized and shows no decline. The training and validation loss curves near alignment throughout the course of the epochs indicates that the model has low overfitting and good generalization.

5.3 Performance/ Comparative Analysis

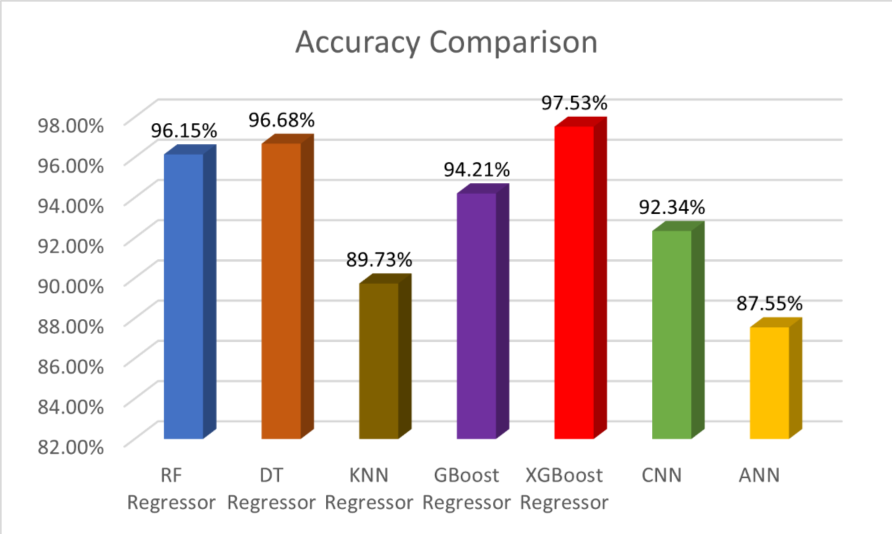

Figure 5.3.1: ALL Model Accuracy Comparison

The "Accuracy Comparison" bar chart in Fig. 5.3.1 evaluates the performance of various machine learning and deep learning models based on accuracy. The XGBoost Regressor excels with the highest accuracy of 97.53%, showcasing its strong predictive power. The Random Forest (RF) Regressor and Decision Tree (DT) Regressor also perform well, with accuracies of 96.15% and 96.68%, respectively. This indicates that ensemble techniques like XGBoost, DT, and RF are highly effective for the dataset.

The Gradient Boosting (GBoost) Regressor also shows good performance with an accuracy of 94.21%. The Convolutional Neural Network (CNN) model achieves 92.34% accuracy, indicating it is reasonably suitable but not as strong as the top three models. The K-Nearest Neighbors (KNN) Regressor, with 89.73% accuracy, and the Artificial Neural Network (ANN) model, with 87.55% accuracy, are less effective, suggesting they may not adequately capture the data's complexity.

In summary, the XGBoost Regressor is the most accurate model, followed closely by the DT and RF Regressors. Ensemble methods are particularly effective, while ANN and KNN are less suitable. The CNN model shows decent performance, indicating it can be useful in some contexts.

❖ Error Matrix of All Model

The model's performance is also evaluated using error metrics like MAE, MSE and RMSE which gives strong evidence to the accuracy of model performance. The less the error value, the better the accuracy.

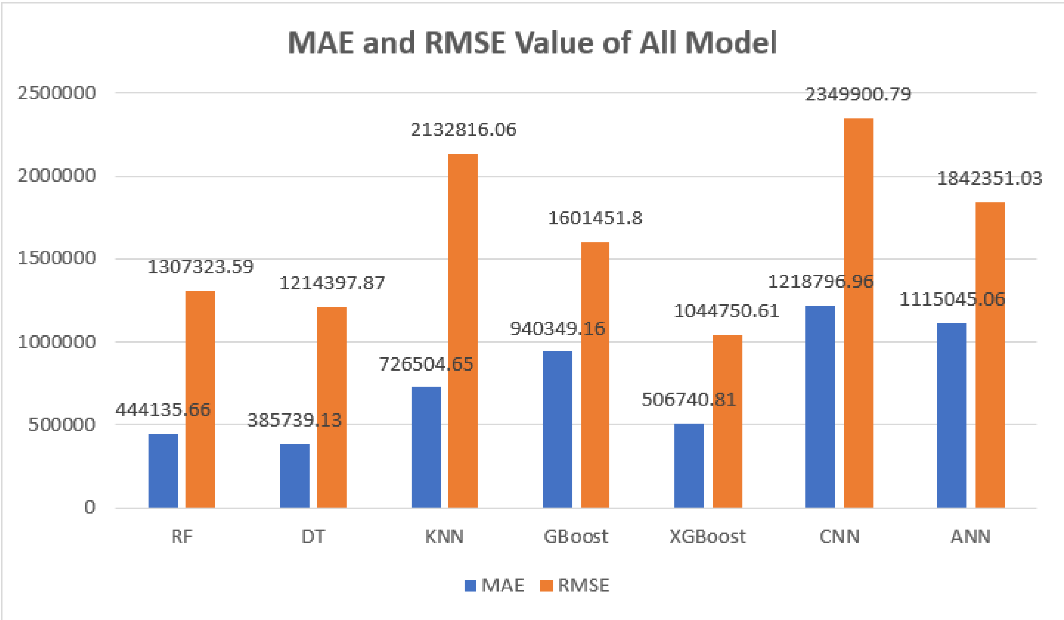

Figure 5.3.2: MAE and RMSE Value of All Model

The figure 5.3.2 shows that the Convolutional Neural Network (CNN) model has the highest error rates, with MAE of 1,218,796.96 and RMSE of 2,349,900.79, among all the models. Whereas, the Decision Tree (DT) model demonstrates the lowest error rates, with an MAE of 3,85,739.13 and the XGBoost model has the lowest RMSE value of 1,044,750.61.

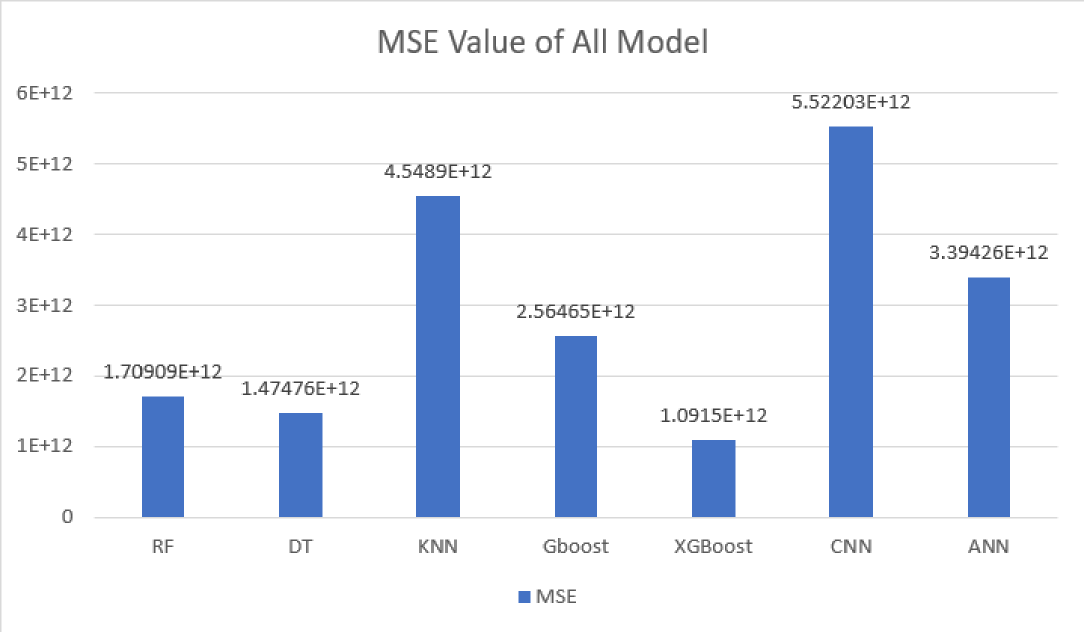

Figure 5.3.3: MSE Value of All Model

The figure 5.3.3 shows that the Convolutional Neural Network (CNN) model has the highest error rates, with an MSE of 5522033750524.54 among all the models. Whereas, the XGBoost model demonstrates the lowest error rates, with MSE of 1091503837228.54.

TABLE 5.3.1: Comparison between Previously Existing Work and Our Work

SL No.	Author Name	Dataset	Algorithms	Accuracy
1	F. M. Rafi [6]	1374 data (collected manually from different renowned websites)	Lasso, Bayesian Ridge, Ridge, XGBoost, KNN, ANN, Random Forest, Support Vector, LightGBM, Catboost, Gradient Boost	Gradient Boost: 92.3%
2	M. A. Hossen and Md. M. Rahman [7]	28000 data (Dhaka City rental house)	Linear Regression, Ridge Regression, Bayesian Regression, and Lasso Regression	Ridge Regression: 91.54%
3	A. Begum et al. [8]	506 data (January 2015 to November 2019, StatLib library Carnegie Mellon University)	Decision Tree, Random Forest, and Linear Regression	Random Forest: 94.4%

4	K. Xu and H. Nguyen [9]	The real estate market in Bolingbrook and Naperville	Linear Regression, Support Vector Regression, Random Forest Regression, Decision Trees, XGBoost Regression	XGBoost Regression: 92.5%
5	J. J. Jui et al. [10]	-	Linear Regression, Random Forest Regression	Random Forest Regression: 93%
6	S. B. Jha et al. [13]	62,723 data (January 2015 to November 2019, Florida's Volusia County Property Appraiser website)	CatBoost, Random Forest, Lasso, XGBoost, Voting Regressor, Linear Regressor, and Support Vector Regressor	XGBoost: 97%
7	S. Özöğür-Akyüz et al. [14]	Manual web scraping of Kadıköy district in Istanbul and kaggle	Linear Regression, Ridge regression, K-means, KNN, NuSVR, Hybrid NuSVR	Hybrid NuSVR: 77.9%
8	D. [15]	-	Decision Tree Regressor, KNN	Decision Tree Regressor: 89%
9	M. Shahhosseini et al. [16]	506 data (Boston metropolitan area, U.S. Census Service) 1460 data (Ames, Iowa from 2006 to 2010)	Lasso, Random Forests, Neural Network, XGBoost, SVM (poly), SVM (RBF), SVM (sigmoid)	Random Forest: 92%
10	J. Kim et al. [17]	52,900 data (Land values in Seoul, South Korea)	Random Forest and XGBoost	XGBoost: 87%
11	M. Yue et al. [18]	33224 data (Chengdu housing rental)	Random Forest Regressor, XGBoost, and LightGBM	XGBoost: 85%
12	A. Oust et al. [19]	16,417 data	Enhanced Hedonic Regression – Ordinary Regression, Regression Kriging, Mixed	Ordinary Regression: 75%

		(Oslo, Norway real estate transactions)	Regressive-Spatial Autoregressive, and Geographically Weighted Regression	
13	A. Deaconu et al. [20]	900 data (apartments in Cluj-Napoca, Romania)	Generalized Linear Model (GLM) Artificial Neural Networking (ANN)	ANN: 80%
14	F. Mostafi et al. [21]	1381 data (real estate price from Trabzon, Turkey)	DNN, PCA-DNN	PCA-DNN: 85%
15	M. Štubňová et al. [22]	256 data (Nitra in the Slovak Republic)	ANN	ANN: 97%
16	Our Work	8255 data (manually collected from bikroy.com and bproperty.com)	RF Regressor, Decision Tree Regressor, KNN, GBoost, XGBoost, CNN and ANN	XGBoost: 97.53%

5.4 Summary

The study's main goal is to forecast apartment prices by utilizing deep learning and machine learning methods. Training and testing sets of the dataset are separated, and an experimental setup is created. The experiment entails training, evaluating, and fine-tuning parameters using a variety of machine learning methods and performance indicators. With the lowest error metrics, the Extreme Gradient Boosting Regressor attains the best accuracy of 97.53%. High accuracy and low mistake rates are also displayed by other models, such as Random Forest and Decision Tree Regressors. CNNs are reasonably accurate, however ANNs are not as accurate. The study comes to the conclusion that, ML and DL models can accurately forecast apartment prices. Subsequent investigations may tackle constraints, investigate supplementary attributes, refine data standards, and augment the comprehensibility of the model.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The use of deep learning and machine learning to the prediction of apartment prices has important environmental and socioeconomic implications. It increases the affordability of housing, strengthens investment choices, promotes equitable pricing policies, and helps with urban planning and development. Additionally, by enabling buyers and sellers to bargain more effectively based on factual information, it increases market efficiency. Additionally helpful in risk management, predictive models let investors and real estate firms take preemptive steps to reduce losses. They also help with sustainable development by pointing out patterns that back environmentally friendly home developments and making resource allocation more effective. Nonetheless, safeguarding data privacy, guaranteeing equity, and ensuring data privacy are ethical issues. Maintaining impartial and accurate models requires regular checks and modifications. In-order to ensure that models can interact with stakeholders and adjust to shifting market conditions, a sustainability plan is necessary. To develop the concept, cooperation with real estate companies, legislators, and urban planners is essential. Involving the community also enhances the relevance and accuracy of the model. In general, the economy, the environment, and society will be greatly impacted by the forecast of apartment prices through the use of deep learning and machine learning.

6.2 Impact on Society

The use of deep learning (DL) and machine learning (ML) to the prediction of apartment prices has profound effects on many facets of society in addition to the technological domain.

- **Improved Market Efficiency**

Apartment price prediction using deep learning and machine learning has the potential to greatly increase social fairness, accessibility, market efficiency, and economic stability. Precise pricing results in more equitable property assessments, diminishing disparities between listed costs and true market values. Price estimating methods can be automated to expedite transactions and enable buyers and sellers to respond more quickly on the basis of accurate information.

- **Enhanced Accessibility and Affordability**

Accurate pricing forecasts provide greater affordability and accessibility while highlighting patterns and trends in the supply of affordable housing. Policy makers then utilize this knowledge to build and promote these kinds of housing alternatives. Precise forecasts also help purchasers make well informed decisions by assisting them in locating houses that fit their budgets and preventing them from overspending.

- **Economic Benefits**

Investments possibilities, market stability, social fairness, equitable property values, resource allocation, well-informed urban planning and development, and sustainable development are among the economic advantages. Predictive modeling technology may lead to the stimulation of interdisciplinary research, bringing together ideas from the social sciences, economics, data science, and urban planning.

- **Social Equity**

Data-driven models improve transparency and trust, fostering confidence among buyers, sellers, and real estate agents. Additionally, consumer empowerment is promoted, resulting in an educated and engaged market. The environment will also be significantly impacted; improved forecasts will result in more economical resource utilization and will encourage the development of green housing solutions.

- **Challenges and Complications**

However, there are issues and things to think about, such prejudice and justice, the digital gap and data privacy. Ensuring data security and responsible information usage is essential, as is providing everyone access to sophisticated prediction tools. To find and reduce biases in training data, ongoing efforts are required.

6.3 Impact on Environment

Predicting apartment prices through the use of deep learning (DL) and machine learning (ML) can have a variety of environmental effects, both good and bad.

- **Promoting Sustainable Environment**

Predicting apartment prices with machine learning and deep learning can benefit the environment in ways like encouraging sustainable development, maximizing resource use, cutting down on building waste, assisting with the integration of renewable energy sources. These models can recognize energy efficient and green structures, which promotes the creation and acquisition of ecologically responsible real estate.

- **Green Energy Usage**

Unfortunately, these models also have detrimental effects on the environment. For example, their high energy consumption stems from the computing power they demand for inference and training, as well as the fact that their training requires sophisticated hardware. A greater carbon footprint can also result from the usage of non-renewable energy sources like hydroelectric, solar, or wind power. Using model architecture, and making cloud investments are some mitigation techniques for environmental effects.

- **Electronic Waste**

Electronic waste may be decreased by employing sustainable hardware techniques like recycling and reusing electronic devices. To lower the frequency of replacements and the ensuring electronic waste, businesses should invest in robust and long-lasting technology and develop take back programs for outdated gear.

- **Smarter Urban Planning**

Better urban planning, data driven choices supporting sustainable development strategies, transportation efficiency, educated consumer choices, and a lower carbon footprint are some of the long-term environmental advantages. Predictive models may be used by urban planners to make data driven decisions that support environmentally friendly infrastructure, lessen urban sprawl, and promote sustainable development practices.

- **Efficient Model Design**

Predictive models can help consumes make more informed decisions by raising their knowledge of how their housing decisions will affect the environment. Precise forecasting may result in more energy efficient appliances, improved insulation, and other environmentally friendly housing upgrades, all of which can reduce carbon emissions.

In conclusion, there are environmental concerns associated with the integration of deep learning and machine learning into apartment pricing prediction; nevertheless, there are also significant potential to promote sustainability. The good effects on resource usage, urban planning, and consumer behavior can result in long-term environmental benefits, while the negative effects can be reduced by putting mitigation strategies like model optimization, renewable energy use, and sustainable hardware practices into practice.

6.4 Ethical Aspects

There are several ethical issues with the application of deep learning and machine learning in apartment price prediction. In-order to guarantee that technology is utilized properly and benefits all stakeholders equally, it is imperative that these challenges be addressed.

- **Data Privacy and Security**

Ethical questions are raised by the use of deep learning and machine learning in apartment price prediction. Because they depend on sensitive data, data privacy and security are essential. To combat misuse, data collection and usage regulations must be transparent. Unfair price projections might result from algorithmic bias perpetuating preexisting prejudices. It is important to ensure equitable representation of all demographic groups and to implement strategies for detecting and mitigating prejudice. Gaining the trust of users and stakeholders requires openness and explainability. Accountability for decisions is essential to comprehending the variables affecting pricing forecasts.

- **Impact on Housing Market Dynamics**

With the possibility of fake inflation and market manipulation, the effect on the dynamics of the housing market is likewise noteworthy. Maintaining accessibility is essential to halting the growing disparity between various income groups. Technology usages ethics ought to be in line with moral principles, legal requirements and economic disparities. Regulatory frameworks, stakeholder participation, transparency, data governance, bias reduction, and education are examples of mitigation techniques. Potential problems can be found and fixed with regular ethical audits. Predictive models have the potential to generate a housing market that is more equitable, inclusive, and efficient by cultivating trust, and efficient by cultivating trust, advancing fairness, and guaranteeing adherence to moral principles.

6.5 Sustainability Plan

Formulating a sustainable plan for apartment price prediction by machine learning and deep learning entails formulating approaches that guarantee the technology's long-term sustainability, low environmental effect, and favorable social consequences.

- **Energy Efficiency and Green Computing**

Energy efficiency, data management, environment impact reduction, social responsibility, transparency, regulatory compliance, economic, economic sustainability, and a long-term vision should be the main priorities strategy for machine learning and deep learning in apartment price prediction. This includes lowering environment effect through lifecycle management, creating hardware recycling programs, implementing carbon offset programs, working with renewable energy sources, optimizing computing resources, and implementing data management and privacy methods.

- **Social Responsibility and Equity**

Along with bias detection and mitigation, inclusive datasets, community participation, explainable AI, ethical audits, regulatory compliance, cost management, and long-term strategy, social responsibility and equity are also essential. This entails putting explainable AI strategies into practice, carrying out frequent audits, and making sure that rules and regulations are followed.

- **Economic Sustainability**

In-order to achieve economic sustainability, grants and financing should be sought for sustainable technical development, as well as cost-effective methods for training models and handling data. Sustainable expansion, ongoing development, and cooperation with educational institutions, research centers, and business partners are all components of a long-term vision. By implementing these tactics, business can make sure their predictive modeling initiatives are economically, socially, and environmentally sound, improving both society and environment.

6.6 Summary

Apartment prices are predicted using machine learning and deep learning, which can have substantial socioeconomic and environmental effects. These models enhance pricing strategies, urban planning, investment choices, and housing affordability. They support sustainable development and risk management as well. But ethical issues like data privacy and data privacy need to be taken into account. Model improvement requires coordination with real estate corporations, regulators, and urban planners, as well as regular audits and sustainability strategies. Involving the community also improves model accuracy. All things considered, these forecasts have a big impact on the environment, the economy, and society.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

The real estate sector has seen a transformation thanks to the accurate and efficient integration of deep learning and machine learning into apartment price prediction of Bangladesh, which outperforms conventional techniques. All parties gain from these sophisticated model's ability to analyze enormous volumes of data, find hidden patterns, and adjust to changing market conditions. More accurate and reliable price forecasts are produced by deep learning models because they are particularly good at managing high dimensional data and capturing intricate, non-linear relationships. In-order to improve model accuracy, the study highlights how crucial it is to use a variety of data sources, including historical pricing, economic indicators, and property attributes. It's also critical to take ethical factors like data privacy, algorithmic bias, and equal access into account. The research highlights the necessity of maximizing computing capacity and employing sustainable energy sources to reduce the carbon impact associated with model training and deployment. Deep learning and machine learning have a significant positive social impact by improving consumer information, reducing market inefficiencies, and valuing real estate more fairly.

7.2 Further Suggested Works

Many interesting consequences, including technical breakthroughs, social advantages, and policy development, might result from more study into the application of machine learning and deep learning in apartment price prediction of Bangladesh.

It is anticipated that the use of deep learning (DL) and machine learning (ML) in apartment pricing prediction would result in substantial technological improvements as well as moral behavior, positive social effects, and environmental sustainability. This includes interaction with IoT, real time forecasts, hybrid models, and improved model accuracy. The geographic accuracy of forecasts may be improved by expanding data sources, such as social media and geospatial data. By identifying additional features, sophisticated feature engineering approaches can increase the accuracy and insights of the model. Techniques for reducing prejudice, equitable AI development, and openness and explainability are all ethical issues. This research has implications for the social and economic sphere, such as improved market efficiency, informed decision-making, decreased volatility, accessibility and affordability, and governmental assistance. Predictive models may also support energy-efficient computing, resource optimization, and sustainable urban design. The consequences are not limited to real estate industry, it provides insights and approaches that may be extended to other areas, so contributing to a future that is more sustainable, egalitarian, and informed.

7.3 Limitation/Conflicts of Interest

Apartment price prediction uses machine learning and deep learning; nevertheless, these approaches have drawbacks, including partial data, data bias, privacy concerns, complexity, generalizability, variances in local markets, shifts in the economy and market, technological difficulties, and conflicts of interest. Model interpretability is important, as are data availability and quality. Model generalization may be hampered by overfitting. Computational resources and model upkeep are two technological obstacles that smaller firms may find prohibitive. Commercial bias, data manipulation, ethical issues, regulatory compliance, market manipulation, sustainability vs. profit, and environmental effect are examples of conflicts of interest. Developing trustworthy prediction models requires resolving these issues and tensions. Updating these models often, including stakeholders, and adhering to regulations are required to reduce possible conflicts and improve these model's trust in the real estate industry.

REFERENCE

[1] 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.

[2] H. Kim, Y. Kwon, and Y. Choi, “Assessing the Impact of Public Rental Housing on the Housing Prices in Proximity: Based on the Regional and Local Level of Price Prediction Models Using Long Short-Term Memory (LSTM),” Sustainability, vol. 12, no. 18, p. 7520, Sep. 2020, doi: https://doi.org/10.3390/su12187520.

[3] R.-T. Mora-Garcia, M.-F. Cespedes-Lopez, and V. R. Perez-Sanchez, “Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times,” Land, vol. 11, no. 11, p. 2100, Nov. 2022, doi: https://doi.org/10.3390/land11112100.

[4] L. Wen, K. Zhou, and S. Yang, “Load demand forecasting of residential buildings using a deep learning model,” Electric Power Systems Research, vol. 179, p. 106073, Feb. 2020, doi: https://doi.org/10.1016/j.epsr.2019.106073.

[5] M. Yazdani, “Machine Learning, Deep Learning, and Hedonic Methods for Real Estate Price Prediction,” 2021. Accessed: Jun. 25, 2024. https://arxiv.org/pdf/2110.07151

[6] F. M. Rafi, “Prediction and Analysis of Flat Price in Dhaka Using Advanced Regression Techniques,”Mar. 04, 2023. http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/9792

[7] M. A. Hossen and Md. M. Rahman, “House Rent Prediction in Dhaka City Using Machine LearningApproach,”Apr.01,2023.http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/10015

[8] A. Begum, N., N. J. Kheya, Md. Z. Rahman, N. Author_id, and N. Author_id, “Housing Price Prediction with Machine Learning,” International journal of innovative technology and exploring engineering, Jan. 30, 2022. https://www.researchgate.net/publication/363371217-Housing-Price-Prediction-with-Machine-Learning

[9] K. Xu and H. Nguyen, “Predicting housing prices and analyzing real estate market in the Chicago suburbs using Machine Learning,” arXiv.org, Oct. 12, 2022. https://arxiv.org/abs/2210.06261

[10] J. J. Jui, M. M. I. Molla, B. S. Bari, M. Rashid, and J. Hasan, “Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques,” Lecture notes in electrical engineering, Jan. 01, 2020. https://www.researchgate.net/publication/342793117-Flat-Price-Prediction-Using-Linear-and-Random-Forest-Regression-Based-on-Machine-Learning-Techniques

[11] M. Modi, A. Sharma, and P. Madhavan, “Applied Research On House Price Prediction Using Diverse Machine Learning Techniques.” https://www.ijstr.org/final-print/apr2020/Applied-Research-On-House-Price-Prediction-Using-Diverse-Machine Learning-Techniques.pdf

[12] I. N. Y. Saputra, S. Saadah, and P. E. Yunanto, “Analysis of Random Forest, Multiple Regression, and Backpropagation Methods in Predicting Apartment Price Index in Indonesia,” Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, vol. 7, no. 2, p. 238, Jul. 2021, doi: 10.26555/jiteki.v7i2.20997.

[13] S. B. Jha, R. F. Babiceanu, V. Pandey, and R. K. Jha, “Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study,” arXiv.org, Jun. 17, 2020. https://arxiv.org/abs/2006.10092

[14] S. Özöğür-Akyüz, B. E. Erdoğan, Ö. Yıldız, and P. K. Ataş, “A Novel Hybrid House Price Prediction Model,” Computational economics, Sep. 16, 2022. https://link.springer.com/article/10.1007/s10614-022-10298-8

[15] D., “House Price Prediction Using Machine Learning,” International Journal for Research in Applied Science and Engineering Technology, vol. 11, no. 7, pp. 693–698, Jul. 2023, doi: 10.22214/ijraset.2023.54739.

[16] M. Shahhosseini, G. Hu, and H. Pham, “Optimizing Ensemble Weights for Machine Learning Models: A Case Study for Housing Price Prediction,” *Smart Service Systems, Operations Management, and Analytics*, pp. 87–97, Nov. 2019, doi: https://doi.org/10.1007/978-3-030-30967-1_9.

[17] J. Kim, J. Won, H. Kim, and J. Heo, “Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea,” *Sustainability*, vol. 13, no. 23, p. 13088, Nov. 2021, doi: https://doi.org/10.3390/su132313088.

[18] M. Yue, J. Zhang, J. Qi, T. Liao, M. Wang, and L. Zhang, “Prediction and Analysis of Chengdu Housing Rent Based on XGBoost Algorithm,” Sep. 2020, doi: https://doi.org/10.1145/3422713.3422720.

[19] A. Oust, S. N. Hansen, and T. R. Pettrem, “Combining Property Price Predictions from Repeat Sales and Spatially Enhanced Hedonic Regressions,” *The Journal of Real Estate Finance and Economics*, Aug. 2019, doi: https://doi.org/10.1007/s11146-019-09723-x.

[20] A. Deaconu, A. Buiga, and H. Tothăzan, “REAL ESTATE VALUATION MODELS PERFORMANCE IN PRICE PREDICTION,” *International Journal of Strategic Property Management*, vol. 26, no. 2, pp. 86–105, Feb. 2022, doi: https://doi.org/10.3846/ijspm.2022.15962.

[21] F. Mostofi, Vedat Toğan, and Hasan Basri Başağa, “House price prediction: A data-centric aspect approach on performance of combined principal component analysis with deep neural network model,” vol. 04, no. 2, pp. 106-116, June. 2021, doi: https://doi.org/10.31462/jcemi.2021.02106116

[22] M. Štubňová, M. Urbaníková, J. Hudáková, and V. Papcunová, “Estimation of Residential Property Market Price: Comparison of Artificial Neural Networks and Hedonic Pricing Model,” Emerging Science Journal, vol. 4, no. 6, pp. 530–538, Dec. 2020, doi: 10.28991/esj-2020-01250.

[23] M. Geerts, Seppe vanden Broucke, and Jochen De Weerdt, “A Survey of Methods and Input Data Types for House Price Prediction,” vol. 12, no. 5, pp. 200–200, May 2023, doi: https://doi.org/10.3390/ijgi12050200.

[24] J. Manasa, R. Gupta and N. S. Narahari, "Machine Learning based Predicting House Prices using Regression Techniques," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 624-630, doi: 10.1109/ICIMIA48430.2020.9074952.

[25] S. Gangarapu and V. S. A. Mernedi, ‘Predicting Airbnb Prices in European Cities Using Machine Learning’, Dissertation, 2023.

[26] J. Hong, H. Choi, and W. Kim, “A HOUSE PRICE VALUATION BASED ON THE RANDOM FOREST APPROACH: THE MASS APPRAISAL OF RESIDENTIAL PROPERTY IN SOUTH KOREA,” International Journal of Strategic Property Management, vol. 24, no. 3, pp. 140–152, Feb. 2020, doi: https://doi.org/10.3846/ijspm.2020.11544.

[27] T.-C. Peng and C.-C. Wang, “The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area,” Social science Japan journal, vol. 25, no. 1, pp. 3–28, Aug. 2021, doi: https://doi.org/10.1093/ssjj/jyab029.

[28] T. Mohd, Nur Izzah Jamil, Noraini Johari, L. Abdullah, and Suraya Masrom, “An Overview of Real Estate Modelling Techniques for House Price Prediction,” pp. 321–338, Jan. 2020, doi: https://doi.org/10.1007/978-981-15-3859-9_28.

[29] L. Mrsic, H. Jerkovic, and M. Balkovic, “Real Estate Market Price Prediction Framework Based on Public Data Sources with Case Study from Croatia,” Intelligent Information and Database Systems, p. 13, 2020, doi: https://doi.org/10.1007/978-981-15-3380-8_2.

[30] S. Yang, "Learning-based Airbnb Price Prediction Model," 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT), Hangzhou, China, 2021, pp. 283-288, doi: 10.1109/ECIT52743.2021.00068.

[31] L. Hyunjae; Son, “Prediction and factors of Seoul apartment price using convolutional neural networks,” The Korean Journal of Applied Statistics, vol. 33, no. 5, pp. 603–614, 2020, doi: https://doi.org/10.5351/KJAS.2020.33.5.603.

[32] J. Lee and J. P. Ryu, “Prediction of Housing Price Index Using Artificial Neural Network,” Journal of the Korea Academia-Industrial cooperation Society, vol. 22, no. 4, pp. 228–234, 2021, doi: https://doi.org/10.5762/KAIS.2021.22.4.228.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
Apartment Price Prediction of Bangladesh using Machine Learning and Deep Learning Algorithms

Student ID: 202-15-14418, 202-15-14435

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **apartment price prediction of Bangladesh** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design	**PO10**

	documentation, as well as provide and receive clear instructions throughout this FYDP.	
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project exhibits **fundamental engineering (K3)** principles by utilize machine learning and deep learning's diverse models for preprocessing and prediction tasks. The project demonstrates **specialist knowledge (K4)** by conducting machine learning with RF, DT, KNN, GBoost, and XGBoost and deep learning with CNN and ANN architectures, enhancing apartment price prediction accuracy in tabular data.	**Page no:** [01] **Section:** [1.1] **Page no:** [11] **Section:** [3.2]
				The project applies **engineering practice & design (K5)** to the method of experiments. The project addresses **engineering practice & technology**	**Page no:** [11] **Section:** [3.2] **Page no:**

				(K6) by employing RF, DT, KNN, GBoost, and XGBoost and deep learning with CNN and ANN model.	[18] **Section:** [3.3]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to advance apartment price prediction using machine learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [06] **Section:** [2.2]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in apartment price prediction, including the limitations of traditional methods and the complexities of integrating machine learning (RF, DT, KNN, GBoost, XGBoost) and deep learning models (CNN and ANN). Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining prediction methodologies.	**Page no:** [02,03, 13, 30] **Section:** [1.2, 1.3, 3.2 (2), 5.3]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting Machine Learning and Deep Learning as the chosen significant solution for enhancing price prediction amidst multiple potential approaches.	**Page no:** [26, 30] **Section:** [5.2, 5.3]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting the real estate industry like apartment price prediction and contributing to advancements in public benefits, which indicates **EP-4**.	**Page no:** [35, 35, 36, 37] **Section:** [6.1, 6.2, 6.3, 6.4]
5.	**EP5: Extends of application codes**	Yes	CO5	To implement the algorithms, we have done python programming code with its libraries, which covers **K4** and **K6.**	**Page no:** [23] **Section:** [4.2]
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	Yes	CO8	We have collected data from well-known real estate websites for our research work, where diverse groups of users or customers and sellers are involved with widely varying needs, where we will	**Page No:** [12, 18] **Section:** [3.2 (1), 3.3]

				determine a certain and reliable data for our research work in.	
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in apartment price prediction which ensures **EP-7**.	**Page no:** [11, 18] **Section:** [3.2, 3.3]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, machine learning, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in apartment price predictions through machine learning and deep learning.	**Page no:** [18, 23,37] **Section:** [3.3, 4.2, 6.4]
2.	**EA2: Level of interaction**	No		N/A	N/A

3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving knowledge through advanced apartment price prediction methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [35, 35, 36,37] **Section:** [6.1, 6.2, 6.3, 6.4]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining apartment price prediction through machine learning and deep learning demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [06, 09] **Section:** [2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [19] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing client privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge	**Page no:** [36] **Section:** [6.3]

			dissemination and societal well-being through the ethical application of advanced real estate technologies that comply with **CO9**.	
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [12, 26, 30, 35, 36] **Section:** [3.2 (1), 5.2, 5.3, 6.2, 6.3]

APARTMENT PRICE PREDICTION OF BANGLADESH USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

ORIGINALITY REPORT

11%	7%	5%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	fastercapital.com Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
4	Submitted to Daffodil International University Student Paper	<1%
5	export.arxiv.org Internet Source	<1%
6	ebin.pub Internet Source	<1%
7	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
8	Submitted to Liverpool John Moores University Student Paper	<1%