Internship Final Report

Student Name: Siri Vennela Amaragani
University: Methodist College of Engineering and Technology
Major: Computer Science Engineering
Internship Duration: 2nd January 2026 – 31st January 2026
Company: ShadowFox
Domain: AIML
Coordinator: Mr.Hari

Objectives

My primary objectives for this internship project were to:

1. Develop a strong understanding of data analytics and machine learning concepts through a real-world regression problem.

2. Gain hands-on experience in preprocessing, analyzing, and modeling structured datasets using Python.

3. Design and implement a regression-based machine learning model to accurately predict Boston house prices.

4. Learn how to evaluate and fine-tune machine learning models using appropriate performance metrics such as Mean Squared Error (MSE) and R^2 score.

5. Apply theoretical knowledge of statistics and machine learning to practical decision-making scenarios in the real estate domain.

Tasks and Responsibilities

During the internship, I was involved in the following key tasks:

● **Data Collection and Understanding:** Worked with the Boston Housing dataset containing socio-economic and geographic attributes such as crime rate, number of rooms, and population indicators. Analyzed the structure and significance of each feature.

● **Data Preprocessing:** Cleaned and prepared the dataset by handling missing values, checking data types, and identifying outliers to ensure data quality and reliability for modeling.

● **Exploratory Data Analysis (EDA):** Performed statistical analysis and data visualization using libraries such as Pandas, Matplotlib, and Seaborn to understand feature distributions and relationships with house prices.

● **Model Selection:** Implemented multiple regression models including Linear Regression, Decision Tree Regressor, and Gradient Boosting Regressor to compare predictive performance.

● **Model Training and Testing:** Split the dataset into training and testing sets and trained the selected models to learn patterns from historical housing data.

● **Model Evaluation:** Evaluated model performance using Mean Squared Error (MSE) and R^2 score to assess prediction accuracy and generalization capability.

● **Hyperparameter Tuning:** Applied GridSearchCV to fine-tune the Gradient Boosting Regressor parameters to improve model performance.

● **Result Analysis:** Compared actual and predicted house prices using visualizations to validate the effectiveness of the final model.

Learning Outcomes

● **Technical Proficiency:** Gained practical experience in Python-based data analytics and machine learning using libraries such as NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn.

● **Understanding of Machine Learning Workflow:** Developed a clear understanding of the end-to-end machine learning lifecycle, from data preprocessing and feature selection to model training, evaluation, and optimization.

● **Analytical Thinking:** Improved ability to analyze datasets, interpret patterns, and make data-driven decisions based on model outputs and evaluation metrics.

● **Problem-Solving Skills:** Enhanced skills in debugging data issues, selecting suitable models, and improving model accuracy through fine-tuning.

● **Professional Development:** Learned how to structure a data science project, document results, and present findings in a clear and professional manner.

Challenges and Solutions

● **Data Quality Issues:** Handling missing values and detecting outliers was a challenge during preprocessing. This was addressed by applying appropriate data imputation techniques and visual inspections using box plots.
● **Model Selection:** Choosing the most suitable regression model required experimentation and comparison. Multiple models were evaluated, and Gradient Boosting Regressor was selected based on superior performance metrics.
● **Overfitting Risk:** Tree-based models initially showed signs of overfitting. This issue was mitigated by tuning hyperparameters and using cross-validation techniques.
● **Performance Optimization:** Achieving higher predictive accuracy required careful tuning of learning rate, depth, and number of estimators, which was resolved using GridSearchCV

Conclusion

This internship project provided valuable hands-on experience in data analytics and machine learning through the development of a Boston House Price Prediction model. By applying regression techniques and performance evaluation methods, I successfully built a robust model capable of accurately predicting housing prices. The project strengthened my understanding of data-driven problem-solving and reinforced my interest in pursuing a career in data analytics and artificial intelligence.

Acknowledgments

I would like to express my sincere gratitude to my internship mentors and coordinators for their guidance and continuous support throughout this project. I am also thankful to my institution for providing the opportunity to work on this practical and industry-relevant internship project. This experience has played a significant role in enhancing my technical skills, confidence, and professional growth.