Industrial Internship Report on

House Price Prediction using Machine Learning

Prepared by

Vishal Kumar Thakur

Executive Summary

This report provides details of the Industrial Internship provided by Upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship focused on solving a real-world machine learning problem related to predicting house prices based on housing features. The project involved data preprocessing, model training, evaluation, and comparison of multiple machine learning algorithms.

My project was **House Price Prediction using Machine Learning**, where I built and compared Linear Regression, Decision Tree, and Random Forest models to predict housing prices based on features such as area, bedrooms, bathrooms, and other housing attributes.

This internship provided valuable exposure to industrial problem-solving and real-world implementation of machine learning solutions. It helped me gain practical knowledge of data science workflows and improved my analytical and programming skills.

---

1. Preface

This internship was conducted over a period of six weeks focusing on solving a real-world machine learning problem. The main objective was to gain practical exposure to industry-level projects and understand the implementation of machine learning algorithms.

The project involved building a house price prediction system using different machine learning models. The internship helped in understanding data preprocessing, model training, evaluation, and performance comparison.

This internship provided hands-on experience with Python, data analysis, and machine learning libraries such as Pandas, Scikit-learn, and Matplotlib.

I would like to thank Upskill Campus, The IoT Academy, and UniConverge Technologies Pvt Ltd for providing this opportunity.

---

2. Introduction

2.1 About UniConverge Technologies Pvt Ltd

UniConverge Technologies Pvt Ltd (UCT) is a company established in 2013 that provides digital transformation solutions focusing on sustainability and return on investment. The company works in technologies such as IoT, cybersecurity, machine learning, cloud computing, and industrial monitoring.

2.2 About Upskill Campus

Upskill Campus is a career development platform that provides industry-based training and internship programs to improve students' technical and professional skills.

2.3 Objective of Internship

The objectives of this internship were:

- To gain practical industry experience
- To solve real-world problems using machine learning
- To understand data science workflows
- To improve problem-solving and programming skills
- To enhance career readiness

2.4 Reference

- Scikit-learn documentation
- Pandas documentation
- Machine Learning online resources

2.5 Glossary

Term	Meaning
ML	Machine Learning
R2 Score	Model accuracy measure
MSE	Mean Squared Error

---

3. Problem Statement

The problem was to build a machine learning system capable of predicting house prices based on housing features such as area, number of bedrooms, number of bathrooms, and other property characteristics.

The challenge was to build an accurate prediction model and compare multiple algorithms to determine the best performing model.

---

4. Existing and Proposed Solution

Existing Solution

Traditional house price estimation methods rely on manual calculations and limited statistical techniques, which may not provide accurate predictions.

Proposed Solution

A machine learning-based system was developed to predict house prices automatically using historical housing data. Multiple models were trained and evaluated to select the best performing model.

Value Addition

- Automated prediction system
- Improved accuracy
- Model comparison
- Feature importance analysis

4.1 Code Submission (GitHub Link)

https://github.com/VishalKumarThakur7/house_price_detection.git

---

4.2 Report Submission (GitHub Link)

---

5. Proposed Design / Model

The system follows these steps:

1. Data loading and preprocessing
2. Feature selection
3. Train-test data split
4. Model training
5. Model evaluation
6. Model comparison
7. Prediction output

5.1 High Level Design

Dataset → Data Preprocessing → Model Training → Prediction → Evaluation