INTERNSHIP FINAL REPORT

STUDENT NAME **:** Vignesh Madivala

UNIVERSITY **:** PES university

MAJOR : CSE (Artificial Intelligence and Machine Learning)

INTERNSHIP DURATION : August 1st , 2025 – August 31st , 2025

COMPANY **:** ShadowFox

DOMAIN : AI/ML

MENTOR **:** Hariharan

COORDINATOR **:** Aakash S

OBJECTIVES

The main purpose of my internship was to get real, hands-on experience in machine learning and data science by working on practical projects. I didn't just want to stay at the theory level — my aim was to actually apply different tools, models, and techniques in real situations.

More specifically, I wanted to:

1. Apply machine learning concepts to solve real-world problems.
2. Build and test predictive models for datasets like housing prices and loan approvals.
3. Learn and experiment with advanced NLP through BERT.
4. Improve my technical skills in Python, data preprocessing, model evaluation, and visualization.
5. Develop a structured way of solving problems — starting from exploring the data to analyzing results and finally presenting them clearly.

TASKS AND RESPONSIBILITIES

During the internship, I got to work on tasks that gave me a full picture of the data science process:

1. Data Cleaning and Preprocessing: Dealt with missing values, encoded categorical data, and prepared datasets for modeling.
2. Exploratory Data Analysis (EDA): Explored datasets with charts and stats to find patterns — for example, I noticed that the average number of rooms ("RM") was strongly linked to Boston house prices.
3. Predictive Modeling:
    3.1. Built regression models for predicting housing prices using Linear Regression, Decision Trees, and Random Forest.
    3.2. Created classification models for loan approval prediction with Logistic Regression, Decision Trees, and XGBoost.
4. Model Evaluation: Compared models with metrics like RMSE, R^2, accuracy, precision, recall, and F1-score to pick the best ones.

5. NLP with BERT: Implemented and analyzed BERT to understand how transformer models handle language tasks.
6. Documentation and Reporting: Organized work into Jupyter Notebooks and scripts, and summarized results in clear reports.

LEARNING OUTCOMES

This internship helped me grow in several ways:

1. Technical Skills: Got more comfortable with Python and libraries like scikit-learn, pandas, numpy, matplotlib, and Hugging Face Transformers.
2. ML Knowledge: Gained practical experience with regression, classification, and NLP.
3. Analytical Thinking: Learned how to identify key features in data and choose suitable models.
4. Problem-Solving: Learned to break down challenges step by step — from preparing the data to interpreting results.
5. Professional Skills: Improved at documenting work, presenting results, and managing time effectively.

CHALLENGES AND SOLUTIONS

Like any real project, there were challenges, but each one taught me something:

1. Messy Data: Some datasets weren't clean. I fixed this through imputation, encoding, and scaling.
2. Accuracy Issues: Reaching high accuracy for loan approval prediction was tough. I solved it by testing multiple algorithms and tuning them with GridSearchCV.
3. Working with BERT: At first, BERT was hard to understand and needed a lot of resources. I tackled this by reading Hugging Face docs, reducing batch sizes, and following online tutorials.
4. Resource Management: Some models took long to run. I handled this by using Google Colab and writing more efficient code.

CONCLUSION

Overall, this internship was a great learning experience. From predicting house prices to loan approvals and exploring BERT, I got to apply theory in practice and build full workflows. These projects not only improved my technical skills but also boosted my confidence in handling real-world datasets. It was a solid step forward in my journey toward becoming a data professional**.**

ACKNOWLEDGMENTS

I'm really thankful for the chance to work on these projects. They helped me grow technically and also taught me how to connect classroom learning with real-world practice.

I also want to acknowledge the huge role of online resources. Google, open-source documentation, research articles, and community forums were a constant help whenever I got stuck. They guided me through tough concepts, debugging issues, and learning new techniques.