PROJECTS REPORT

AI /ML INTERNSHIP

Presented To
UNIFIED MENTOR

Presented By
SAMIR SHARMA
UMID02082553309

__Introduction__

In today's world, Artificial Intelligence (AI) and Machine Learning (ML) have become the foundation of innovation and automation. My internship at Unified Mentor was a significant milestone in my journey toward becoming a skilled data and AI professional.

This internship offered me an opportunity to apply theoretical knowledge to practical, real-world challenges. Over the course of the internship, I developed and implemented multiple AI/ML projects involving classification, regression, deep learning, and anomaly detection tasks.

Each project not only helped me understand the complete lifecycle of a machine learning pipeline but also improved my ability to clean data, engineer features, tune hyperparameters, and evaluate model performance effectively.

This report summarizes the projects I completed, my learnings, methodologies used, and the overall professional growth achieved during my tenure as an AI/ML Intern at Unified Mentor.

Objectives of the Internship

The main goals of my internship were:
- To gain hands-on experience in building machine learning models from scratch.
- To strengthen my understanding of supervised learning, ensemble methods, and neural networks.
- To explore real-world datasets and derive insights through Exploratory Data Analysis (EDA).
- To enhance practical knowledge in libraries like scikit-learn, XGBoost, LightGBM, TensorFlow, and Keras.
- To develop industry-standard documentation, GitHub project presentation, and report-writing skills.
- To experience how machine learning is applied to different domains such as healthcare, finance, and environmental studies.

Tools and Technologies Used

Category	Tools / Technologies
Programming Language	Python 3.11
Libraries	NumPy, Pandas, Matplotlib,
ML Frameworks	Scikit-learn, XGBoost,
IDEs	Jupyter Notebook, Google
Deployment	Streamlit (for demo apps),
Version Control	Git and GitHub
Data Handling	CSV, NumPy arrays, Pickle,
Visualization	Matplotlib, Seaborn, Plotly

Internship Overview

Throughout this internship, I worked on six major machine learning projects, each targeting a specific use case and domain.

   These projects helped me understand not just how algorithms work, but also when and why to use them.
From predicting product prices to detecting fraud and classifying images, each project expanded my skill set across diverse data problems.

Project Summaries

1) Forest Cover Type Prediction
GitHub Repository: <u>Forest-Cover-Prediction</u>
Objective: Predict the type of forest cover (1–7) for a 30×30m plot of land based on terrain and environmental features.
Approach:
  • Conducted EDA and visualized correlations between elevation, soil types, and forest cover.
  • Implemented Logistic Regression, Random Forest, XGBoost, and LightGBM models.
  • Adjusted label encoding for multi-class compatibility (0–6 index for boosting models).
  • Evaluated models on accuracy, precision, and confusion matrix visualization.
Results:
 LightGBM achieved the best performance with an accuracy of 88.5%, followed closely by XGBoost (88.3%).
Key Learnings:
  • Handling multi-class datasets effectively.
  • Importance of proper label encoding for boosted models.
  • Balancing model performance and computational efficiency.

2) Vehicle Price Prediction

GitHub Repository: <u>Vehicle-Price-Prediction</u>

Objective: Predict the price of used vehicles based on make, model, mileage, year, and other specifications.

Approach:

- Cleaned data and engineered new features (extracted horsepower, engine displacement, text lengths).
- Used regression algorithms: Linear Regression, Ridge, Random Forest, XGBoost, and LightGBM.
- Evaluated models using MAE, RMSE, and R^2 metrics.
- Visualized model residuals and feature importance.

Results:

Ridge Regression performed best with R^2 = 0.847 and RMSE ≈ 6830, indicating good generalization.

Key Learnings:

- Regularization (Ridge) can outperform complex ensemble models when data is clean and structured.
- Importance of feature scaling and preprocessing pipelines.
- How textual data (car names, engine specs) can enhance numeric prediction models.

---

3) Mobile Price Prediction

GitHub Repository: <u>Mobile-Price-Prediction</u>

Objective: Classify mobile phones into one of four price categories (0–3) based on their specifications.

Approach:

- Explored the dataset through EDA and correlation plots.
- Applied Logistic Regression, Random Forest, XGBoost, and LightGBM models.
- Balanced data using stratified splits and standardized feature scaling.
- Compared models using classification metrics (accuracy, precision, recall).

Results:

Logistic Regression achieved the highest performance, confirming the linear separability of the data.

Key Learnings:

- Feature importance visualization and interpretability.
- The value of simple linear models in well-structured datasets.
- Model deployment using joblib serialization.

4) ASL Image Classification

GitHub Repository: <u>ASL-Image-Classification</u>

Objective: Recognize hand signs (A–Z) in American Sign Language using deep learning.

Approach:

- Preprocessed images (resizing, normalization).
- Built a Convolutional Neural Network (CNN) using TensorFlow & Keras.
- Used data augmentation to prevent overfitting.
- Evaluated model performance on training and validation sets.

Results:

Achieved 94% validation accuracy on ASL dataset.

Key Learnings:

- Implementing CNNs from scratch.
- Understanding convolution, pooling, and dropout layers.
- Managing overfitting using data augmentation.

---

5) Heart Disease Detection

GitHub Repository: <u>Detect-Heart-Disease</u>

Objective: Predict the presence of heart disease based on medical parameters.

Approach:

- Cleaned dataset and handled missing values.
- Trained Logistic Regression, Random Forest, and XGBoost models.
- Evaluated using accuracy, recall, precision, and ROC-AUC scores.

Results:

Random Forest achieved 86% accuracy with strong recall performance.

Key Learnings:

- Handling medical datasets with care due to sensitivity and imbalance.
- Interpreting confusion matrices and ROC curves.
- Importance of recall in health-critical classification tasks.

6) Fraud Transaction Detection
GitHub Repository: <u>Fraud-Transaction-Detection</u>
Objective: Detect fraudulent financial transactions using transaction data.
Approach:
- Addressed class imbalance using undersampling and precision-recall analysis.
- Trained Logistic Regression, Random Forest, and XGBoost models.
- Optimized models for recall and F1-score to detect rare fraud cases.

Results:
 XGBoost achieved 98% accuracy and a very high recall for fraud class.

Key Learnings:
- Managing highly imbalanced datasets.
- Using precision-recall trade-off for fraud detection.
- Understanding anomaly detection in real-world financial data.

<u>Key Learnings</u>

- Understood how to build complete ML pipelines from raw data to model deployment.
- Gained confidence in using advanced ML algorithms like XGBoost and LightGBM.
- Learned how to evaluate models using diverse performance metrics suited to each problem type.
- Improved problem-solving mindset and data-driven decision-making.
- Enhanced understanding of real-world data challenges, including missing values, imbalance, and overfitting.
- Developed skills in presenting ML results clearly through reports, charts, and documentation.

Challenges Faced

- Managing inconsistent and unclean datasets, especially during preprocessing.
- Understanding how hyperparameter tuning affects performance in ensemble methods.
- Maintaining balanced datasets while avoiding model bias.
- Training deep learning models under hardware constraints (GPU memory limits).
- Creating clear visualizations and interpretable results for non-technical audiences.

Outcomes and Impact

- Through this internship, I transitioned from learning ML concepts theoretically to implementing them practically.
- By completing six end-to-end projects, I built a diverse portfolio of applied AI/ML work.
- This experience strengthened my resume, enhanced my technical foundation, and provided me with real-world insight into how data science adds value across industries such as automotive, healthcare, finance, and environmental analysis.

Conclusion

My journey as an AI/ML Intern at Unified Mentor has been both educational and transformative.

I have developed a comprehensive understanding of how machine learning systems are built, tested, and optimized. This internship also honed my technical communication, collaboration, and analytical skills — essential for success in the AI/ML industry.

By applying AI techniques to solve real-world problems, I've become more confident in pursuing a professional career as a Data Scientist or Machine Learning Engineer.

Declaration

I, Samir Sharma, hereby declare that the work presented in this report is my original work carried out during my internship as an AI/ML Intern at Unified Mentor, under Internship ID UMID02082553309.
This report accurately represents my efforts and learnings throughout the internship period.

References

- Unified Mentor Internship Portal
- Scikit-learn Documentation (https://scikit-learn.org)
- XGBoost & LightGBM Official Documentation
- TensorFlow/Keras Documentation
- Kaggle Datasets and Articles