Project 2
Part 3: Model Evaluation and Final Report

The goal of this project was to predict whether a house in California is priced above the median value using various classification algorithms. After performing exploratory data analysis and preprocessing the data, we trained and evaluated four supervised machine learning models:

- K-Nearest Neighbors (KNN)
- Decision Tree Classifier
- Random Forest Classifier
- AdaBoost Classifier

In this section, we evaluate these models based on key performance metrics and provide a thorough analysis of their behavior, advantages, and shortcomings. The final recommendation will be based on both model performance and relevance to the real-world application of house price classification.

The following table summarizes the performance of each model based on the testing set:

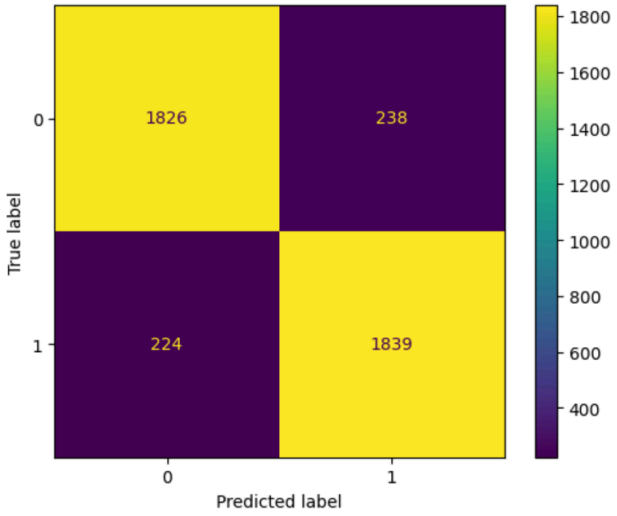
Model	***Accuracy***	***Precision***	***Recall***	***F1 Score***
KNN	0.8320814150714805	0.8252611585944919	0.842462433349491	0.8337730870712402
Decision Tree	0.8475890477344318	0.8298068077276909	0.8744546776539021	0.8515459051215483
Random Forest	0.8880542767143204	0.8854116514203177	0.8914202617547261	0.8884057971014493
AdaBoost	0.8706081899685001	0.8621506395073425	0.8822103732428502	0.872065165309056

For this particular binary classification problem – predicting whether a house is priced above the median value – we must choose the most relevant metric for evaluation. We will not only rely on accuracy, as while accuracy gives us an overall sense of correctness, it can be misleading, especially if the dataset is imbalanced. In such cases, a model could achieve high accuracy simply by favoring the majority class.

Therefore, F1 score matters more in this case. The F1 score is the harmonic mean of precision and recall, which makes it a better choice when there is a tradeoff between false positives and false negatives. In this context, precision is important because we don't want to mistakenly label an inexpensive house as being above the median to prevent misleading buyers or investors.

Recall is equally important because we want to capture as many truly expensive homes as possible. Missing out on them means underestimating valuable properties. Thus, the F1 score is the most appropriate metric to evaluate performance in this scenario, as it balances both precision and recall effectively.

To interpret the confusion matrix of the two top-performing models (Random Forest and AdaBoost), in the Random Forest model shown below, both False Positives (FP) and False Negative (FN) values were relatively low, showing that the model was well-balanced. It had high True Positives with correct prediction of above-median houses and True Negatives, which the model correctly classified as below-median houses, explaining its strong F1 score and accuracy.



AdaBoost showed similar trends but had slightly higher FP and FN counts than Random Forest. However, it still maintained a strong F1 score and generalizability. The model might benefit from more estimators, which could push its performance further.

This clearly compares the strengths and weaknesses of each model. KNN benefits from its simple, interpretable, and decent performance on standardized data. However, it is sensitive to irrelevant features and scaling and slower on large datasets. Therefore, it would be more useful on smaller, well-scaled datasets where interpretability matters. Decision Tree Classifier is easy to understand and visualize, and it has the advantage of being able to handle non-linear relationships, but it is prone to overfitting unless pruned or depth-limited. Random Forest classifier, the best performer out of the four, is able to handle both variance and bias, robust to outliers, and performance automatic feature selection. However, it is a miss that it is less interpreable than a single decision tree. It is very useful in most situations involving structured tabular data, and also works well out of the box. Lastly, AdaBoost classifier is good that it focuses on hard-to-classify instances, and performs well with fewer estimators. Therefore, it is useful when dealing with clean data that we want to reduce bias.

Based on the performance metrics, confusion matrix analysis, and model strengths, Random Fores Calssifier is the most suitable model for this Problem. It had the highest F1 score, indicating balanced precision and recall, and it is robust to noise and outliers in the data. Not only that, but it also requires minimal preprocessing and still performs well. A runner-up choice would be AdaBoost, especially in environments where boosting may improve generalization or when ensemble size needs to be smaller for efficiency.

In this project, we successfully explored the dataset and understood its distribution, applied several machine learning classification models, compared them using relevant evaluation metrics, and identified the best-performing model. This approach provides a strong foundation for classifying house prices and can be extended to real estate valuation, investment analysis, or housing market predictions.