Sri Sivasubramaniya Nadar College of Engineering, Chennai

(An autonomous Institution affiliated to Anna University)

Degree & Branch	M.Tech (Integrated) Computer Science & Engineering	**Semester**	V
Subject Code & Name	ICS1512 – Machine Learning Algorithms Laboratory		
Academic Year	2025–2026 (Odd)	**Batch**	2023–2028

Experiment 3: Ensemble Prediction and Decision Tree Model Evaluation

Objective

To build classifiers such as Decision Tree, AdaBoost, Gradient Boosting, XGBoost, Random Forest, and Stacked Models (using SVM, Naïve Bayes, Decision Tree) and evaluate their performance through 5-Fold Cross-Validation and hyperparameter tuning.

Dataset

Download: Wisconsin Diagnostic Dataset
569 samples and 30 numerical features representing cell nuclei characteristics from digitized images. Target labels are binary.

Steps for Implementation

1. Load and preprocess dataset: encode labels, handle missing values, standardize features.
2. Perform EDA (class balance, feature correlation).
3. Split dataset into training and test sets (e.g., 80-20).
4. Train the following models:
   - Decision Tree
   - AdaBoost
   - Gradient Boosting
   - XGBoost
   - Random Forest
   - Stacking Classifier (SVM + Naïve Bayes + Decision Tree)
5. Use GridSearchCV or RandomizedSearchCV to tune hyperparameters.
6. Record best hyperparameters and evaluate with 5-Fold Cross-Validation.
7. Plot ROC curves and compute performance metrics.

Important Hyperparameters (To Explore)

- **Decision Tree:** `criterion`, `max_depth`, `min_samples_split`, `min_samples_leaf`
- **AdaBoost:** `n_estimators`, `learning_rate`, `base_estimator`
- **Gradient Boosting:** `n_estimators`, `learning_rate`, `max_depth`, `subsample`
- **XGBoost:** `n_estimators`, `learning_rate`, `max_depth`, `gamma`, `subsample`, `colsample_bytree`
- **Random Forest:** `n_estimators`, `max_depth`, `criterion`, `max_features`, `min_samples_split`
- **Stacked Ensemble:** Base models + `final_estimator` (e.g., Logistic Regression)

Decision Tree Model

Hyperparameter Trials

Table 1: Decision Tree - Hyperparameter Tuning

criterion	max_depth	Accuracy	F1 Score

AdaBoost Model

Hyperparameter Trials

Table 2: AdaBoost - Hyperparameter Tuning

n_estimators	learning_rate	Accuracy	F1 Score

Gradient Boosting Model

Hyperparameter Trials

Table 3: Gradient Boosting - Hyperparameter Tuning

n_estimators	learning_rate	max_depth	Accuracy	F1 Score

XGBoost Model

Hyperparameter Trials

Table 4: XGBoost - Hyperparameter Tuning

n_estimators	learning_rate	max_depth	gamma	Accuracy	F1 Score

Random Forest Model

Hyperparameter Trials

Table 5: Random Forest - Hyperparameter Tuning

n_estimators	max_depth	criterion	Accuracy	F1 Score

Stacked Ensemble Model

Hyperparameter Trials

Table 6: Stacked Ensemble - Hyperparameter Tuning

Base Models	Final Estimator	Accuracy / F1 Score
SVM, Naïve Bayes, Decision Tree	Logistic Regression	
SVM, Naïve Bayes, Decision Tree	Random Forest	
SVM, Decision Tree, KNN	Logistic Regression	

5-Fold Cross-Validation Results

Observation Questions

- Which model achieved the best validation accuracy among all six methods?
- How does Decision Tree performance compare to ensemble methods?
- Did the Random Forest benefit from tuning `max_depth` or `n_estimators`?
- Which model showed the best generalization? Any overfitting?
- Did stacking improve performance over base models?

Table 7: 5-Fold Cross Validation Results for All Models

Model	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Average Accuracy**
Decision Tree						
AdaBoost						
Gradient Boosting						
XGBoost						
Random Forest						
Stacked Model						

Report Checklist

- Aim and Objective
- Libraries Used
- Code for All Variants and Models
- Confusion Matrix and ROC for Each
- Hyperparameter Tuning Tables
- Cross-Validation Results Table
- Feature Importance Visuals
- All Comparison Tables
- Observations and Conclusions

References

- scikit-learn: Decision Tree
- scikit-learn: Ensemble Methods
- XGBoost Documentation
- UCI Dataset