Assignment-03

This project explores the implementation and evaluation of various ensemble learning techniques on a classification dataset. The methods applied include Bagging (Random Forest), Boosting (AdaBoost and Gradient Boosting), Stacking, and Voting (Hard and Soft). Performance was assessed using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC, supported by cross-validation for reliability. Explainable AI tools like SHAP and LIME were employed to interpret model predictions and identify significant features. The study highlights the trade-offs and strengths of different ensemble methods, offering insights into their suitability for imbalanced data scenarios.

Ensemble Learning Techniques

Bagging

- **Implemented Method**: Random Forest
- **Description**: A bagging ensemble method that uses multiple decision trees and aggregates their predictions.
- **Performance**:
    - Accuracy: 85.65%
    - Precision: 44.78%
    - Recall: 17.48%
    - F1 Score: 25.14%
    - ROC-AUC: 57.01%
- **Analysis**: Random Forest performed moderately, showing strong accuracy but relatively low recall, which indicates it struggles to identify positive cases.

Boosting

- **Implemented Methods**: AdaBoost and Gradient Boosting
- **Description**:
    - AdaBoost combines weak classifiers iteratively to minimize errors from previous iterations.
    - Gradient Boosting optimizes predictions by minimizing loss using gradient descent.
- **Performance**:
    - AdaBoost: Balanced precision (52.10%) and recall (19.85%), F1 score: 28.75%, ROC-AUC: 58.47%
    - Gradient Boosting: High precision (60.97%) but low recall (6.83%), F1 score: 12.28%, ROC-AUC: 53.06%
- **Analysis**: Boosting techniques showed strong precision but required trade-offs in recall.

Stacking

- **Implemented Method**: Two base models (Random Forest, Gradient Boosting) and a Logistic Regression meta-model.
- **Model Selection**: Base models were chosen for their strengths (Random Forest for stability and Gradient Boosting for precision). Logistic Regression was selected as the meta-model due to its simplicity and effectiveness in combining outputs.
- **Performance**:

  - Accuracy: 86.65%
  - Precision: 56.48%
  - Recall: 13.83%
  - F1 Score: 22.22%
  - ROC-AUC: 56.06%
- **Analysis**: Stacking achieved better balance than individual base models, though recall remained a challenge.

Voting

- **Implemented Methods**: Hard and Soft Voting
- **Description**: Combined three models (Random Forest, Gradient Boosting, AdaBoost) using majority voting for hard voting and weighted averaging for soft voting.
- **Performance**:
  - Hard Voting: Accuracy: 86.60%, Precision: 58.95%, Recall: 9.36%, F1 Score: 16.15%, ROC-AUC: 54.15%
  - Soft Voting: Accuracy: 84.49%, Precision: 40.28%, Recall: 25.83%, F1 Score: 31.47%, ROC-AUC: 59.85%
- **Analysis**: Soft Voting performed better in recall and F1-score, while Hard Voting excelled in accuracy.

Cross-Validation

- **Methodology**: Applied 5-fold cross-validation for each ensemble method to assess stability and reliability.
- **Findings**:
  - Random Forest: Stable accuracy, slightly inconsistent recall across folds.
  - Boosting (AdaBoost, Gradient Boosting): Showed variability in recall across folds but maintained consistent precision.
  - Stacking: Provided steady accuracy and F1 scores.
  - Voting: Soft Voting showed more consistency than Hard Voting.

Explainable AI (XAI)

SHAP

- **Applied Model**: Soft Voting (Best performer for recall and F1 score).
- **Visualization**: SHAP summary plots highlighted the most significant features influencing predictions. Feature contributions for true positive predictions were particularly analyzed.
- **Insights**: Key features were identified, with clear interpretation of how they increased or decreased the likelihood of positive classifications.

LIME

- **Test Instances**: A few challenging cases (false positives and false negatives) were analyzed.

- **Explanation**: LIME provided localized explanations for individual predictions, helping understand why specific misclassifications occurred.
- **Insights**: Indicated possible biases in the dataset affecting predictions.

Performance Analysis

Comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC-AUC**
Random Forest	85.65%	44.78%	17.48%	25.14%	57.01%
Gradient Boosting	86.55%	60.97%	6.83%	12.28%	53.06%
AdaBoost	86.43%	52.10%	19.85%	28.75%	58.47%
Stacking Classifier	86.65%	56.48%	13.83%	22.22%	56.06%
Hard Voting	86.60%	58.95%	9.36%	16.15%	54.15%
Soft Voting	84.49%	40.28%	25.83%	31.47%	59.85%

Trade-offs and Insights

- Precision vs. Recall: Boosting methods and Stacking provided high precision but struggled with recall.
- Soft Voting offered the best balance, achieving the highest recall and F1 score, making it more suitable for imbalanced datasets.
- Ensemble methods demonstrated varying strengths; model selection depends on the specific application and metric priorities.