**Sri Lanka Institute of Information Technology**

A logo with a lion and a blue and orange design

AI-generated content may be incorrect.

**IT2011**

**Artificial Intelligence & Machine Learning**

**2025**

**Year 2 Semester 1**

**Group ID: 2025-Y2-S1-MLB-B11G1-05**

**Project Topic: Stroke Prediction Using Clinical Features**

**1.Models Developed**

1. Decision Tree (DecisionTreeClassifier)
2. K-Nearest Neighbors (KNN) (KNeighborsClassifier)
3. XGBoost (XGBClassifier)
4. Random Forest (RandomForestClassifier)
5. Logistic Regression (LogisticRegression)
6. LightGBM (LGBMClassifier)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Cross-Validation Accuracy | Test Accuracy | Precision | Recall | F1-Score | ROC-AUC |
| Logistic Regression (Baseline) | None | 0.837573 | 0.166667 | 0.419355 | 0.238532 | 0.641969 |
| Random Forest (Baseline) | None | 0.94910 | 0.00000 | 0.00000 | 0.00000 | 0.50000 |
| KNN (Baseline) | None | 0.936399 | 0.000000 | 0.000000 | 0.000000 | 0.498437 |
| XGBoost (Baseline) | None | 0.937378 | 0.428571 | 0.096774 | 0.157895 | 0.544220 |
| LightGBM (Baseline) | None | 0.937378 | 0.400000 | 0.064516 | 0.111111 | 0.529133 |
| Logistic Regression (Tuned) | 0.871961 | 0.746575 | 0.168350 | 0.806451 | 0.278551 | 0.774579 |
| Random Forest (Tuned) | 0.959375 | 0.940313 | 0.666667 | 0.032258 | 0.061538 | 0.515608 |
| Decision Tree (Tuned) | 0.928740 | 0.89726 | 0.247059 | 0.33871 | 0.285714 | 0.636022 |
| KNN (Tuned) | 0.933864 | 0.885519 | 0.227723 | 0.370968 | 0.282209 | 0.644859 |
| XGBoost (Tuned) | 0.959375 | 0.915851 | 0.184211 | 0.112903 | 0.140000 | 0.540306 |
| LightGBM (Tuned) | 0.871961 | 0.934442 | 0.40000 | 0.16129 | 0.229885 | 0.572833 |

**2.Final Selected Model**

* The Logistic Regression (Tuned with Class Balance) model was selected due to its optimal balance of high Recall (essential for patient safety) and High Interpretability meeting both performance and ethical criteria.

| **Model** | **Test Accuracy** | **Recall** | **F1-Score** | **ROC-AUC** | **Interpretability** |
| --- | --- | --- | --- | --- | --- |
| **Logistic Regression (Tuned with Class Balance)** | **0.746575** | **0.806451** | **0.278551** | **0.774579** | **High** |