**Bail Reckoner**

**Module 2: Model Training**

* In the second module, we have used three algorithms to train the model, here are the following algorithms:

**1. Gradient Boosting Algorithm: 82% Acuuracy**

* Derive new features from existing ones or perform feature interactions.
* Use techniques like one-hot encoding, label encoding, or polynomial features.
* Perform feature selection to eliminate irrelevant or redundant features.

**2. XGBoost Algorithm: 100% Acuuracy**

* Faster and more efficient than traditional Gradient Boosting.
* Handles missing values internally.
* Highly effective for structured/tabular data.

1. **CatBoost Algorithm: 86% Accuracy**

* Handles categorical features automatically.
* Reduces overfitting through robust regularization.
* Performs well on both numerical and categorical data.
* **Comparision between the 3 Algorithms**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Strengths** | **Weaknesses** |
| **Gradient Boosting** | 82% | - Simple to implement - Handles overfitting better than many models | - Slower training time - Can be outperformed by newer boosting methods |
| **XGBoost** | 100% | - Extremely high performance - Regularization to reduce overfitting - Fast and scalable | - Risk of overfitting if not properly tuned - Complex to interpret |
| **CatBoost** | 86% | - Handles categorical variables natively - Less data preprocessing needed | - Slightly slower than XGBoost - Less community support than XGBoost |

* **Our aim is to select Catboost Algorithm; Following are the Reasons to Select CatBoost Algorithm**

1. **Handles Categorical Data Efficiently**
   * CatBoost natively supports categorical features, which are common in legal datasets (e.g., offense type, court category).
2. **High Accuracy (86%)**
   * Provides a strong balance between performance and generalization without overfitting like XGBoost might.
3. **Less Data Preprocessing Needed**
   * Reduces development time and avoids errors by automatically managing missing values and encoding.
4. **Robust to Overfitting**
   * CatBoost uses techniques like ordered boosting, making it more stable and less likely to overfit.
5. **Better Interpretability**
   * Offers tools like feature importance that help legal professionals understand model decisions more transparently.