**Model Development Phase**

**Project Name :** Covid - 19 Infant growth Analysis and Prediction

## **Model Selection Report**

In this project, multiple machine learning models were evaluated for predicting infant developmental outcomes. Factors such as accuracy, confusion matrix, and classification report were considered to select the most suitable model.

| Model | Description |
| --- | --- |
| Model - 1 : TabPFNClassifier | • A transformer-based probabilistic neural network designed for tabular data.  • Outperformed Logistic Regression in accuracy and robustness.  • Selected as the final model. |
| Model- 2 : XGBClassifier | • A gradient boosting–based ensemble model that builds multiple weak learners (decision trees) sequentially.  • Handles both numerical and categorical features efficiently, with built-in regularization to prevent overfitting.  • Achieved strong predictive performance and interpretability through feature importance analysis.  • Considered as an alternative advanced model alongside TabPFNClassifier. |

**Model Development Steps :**

| **Model** | **Description** |
| --- | --- |
| Model - 1: Dataset Loading and Preprocessing | Loaded Infant\_development\_dataset.csv, handled missing values (mean for numeric, mode for categorical), and applied Label encoder for categorical features and target labels. |
| Model - 2 : Splitting dataset | Split the dataset into training (75%) and testing (25%) |
| Model - 3: Feature Encoding | Encoded categorical variables (period) and the target class labels using Label encoder. |
| Model - 4 : Training the models | Trained XGBClassifier (baseline) and TabPFNClassifier (final model). |
| Model - 5 : Evaluation | Evaluated models using accuracy, precision, recall, and F1-score. TabPFNClassifier achieved higher performance. |
| Model - 6 : Prediction | Used TabPFNClassifier to predict infant development outcomes for unseen input records. |