**Data Collection and Preprocessing**

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

Prepare the infant development dataset for machine learning classification using TabPFN.

**PREPROCESSING STEPS :**

| **SECTION** | **DESCRIPTION** |
| --- | --- |
| Data Overview | Loaded the dataset, shuffled dataset, inspected columns, and checked missing values. |
| Handling missing data | Filled missing numeric values (age\_months, height\_cm, weight\_kg, speech\_score, milestone\_score) with their column means. For categorical column period replace missing values with the mode. |
| Feature & Target Split | Split dataset into features x and y |
| Encoding | Applied LabelEncoder to transform categorical target labels into numeric form. |
| Splitting dataset | Divided data into training (75%) and testing (25%) sets |
| Model preparation | Installed and initialized TabPFNClassifier for training. |
| Visualization | Generated scatter plots to compare training and testing predictions versus actual labels. |

**Data Preprocessing Code Screenshots :**

| **Section** | **Code** |
| --- | --- |
| Data Overview |  |
| Handling missing data |  |
| Feature & Target Split |  |
| Encoding |  |
| Splitting dataset |  |
| Model preparation |  |
| Visualization |  |