# CatBoost Algorithm

Shah Md. Arshad Rahman Ziban

## Introduction

CatBoost (Categorical Boosting) is a gradient boosting algorithm developed by Yandex, specifically designed to handle categorical data efficiently. It is a powerful machine learning algorithm, often used for classification and regression problems, and is known for its high performance with minimal hyperparameter tuning.

## Key Features of CatBoost

### Handles Categorical Data Efficiently

Unlike other boosting algorithms (e.g., XGBoost, LightGBM), CatBoost does not require one-hot encoding or target encoding. It uses an advanced method called *ordered boosting* to process categorical features dynamically.

### Faster Training & Inference

- Supports GPU acceleration.
- Efficient memory usage compared to XGBoost and LightGBM.

### Robust to Overfitting

CatBoost uses ordered boosting and oblivious decision trees to prevent target leakage and reduce overfitting.

### Great for Small Datasets

Works well with small datasets where deep learning may not perform well.

### Built-in Missing Value Handling

CatBoost can handle missing values automatically without requiring imputation.

### Supports Multiclass Classification & Regression

Can be used for binary classification, multiclass classification, and regression problems.

## How CatBoost Works

### Oblivious Decision Trees

CatBoost uses symmetric (balanced) trees, where the same splitting rule is applied to both left and right nodes. This improves training stability and makes parallelization easier.

### Ordered Boosting

Instead of using traditional boosting, CatBoost processes categorical data using a special method to avoid target leakage.

### Efficient Handling of Categorical Features

Uses combination encoding to encode categorical features in an optimal way.