

CatBoost

- CatBoost is a recently open-sourced machine learning algorithm from Yandex.
- It can work with diverse data types to help solve a wide range of problems that businesses face today.
- **“CatBoost” name comes from two words “Category” and “Boosting”.**
- It is a machine learning algorithm which allows users to quickly handle categorical features for a large data set and this differentiates it from XGBoost & LightGBM.
- It yields state-of-the-art results without extensive data training typically required by other machine learning methods, and Provides powerful out-of-the-box support for the more descriptive data formats that accompany many business problems.
- **“Boost”** comes from gradient boosting machine learning algorithm as this library is based on gradient boosting.
- It can also return very good result with relatively less data, unlike Deep Learning models that need to learn from a massive amount of data.

Catboost advantages

- **Performance:** CatBoost provides state of the art results and it is competitive with any leading machine learning algorithm on the performance front.
- **Handling Categorical features automatically:** We can use CatBoost without any explicit pre-processing to convert categories into numbers. CatBoost converts categorical values into numbers using various statistics on combinations of categorical features and combinations of categorical and numerical features.
- **Robust:** It reduces the need for extensive hyper-parameter tuning and lower the chances of overfitting also which leads to more generalized models. Although, CatBoost has multiple parameters to tune and it contains parameters like the number of trees, learning rate, regularization, tree depth, fold size, bagging temperature and others.
- **Easy-to-use:** You can use CatBoost from the command line, using an user-friendly API for both Python and R.