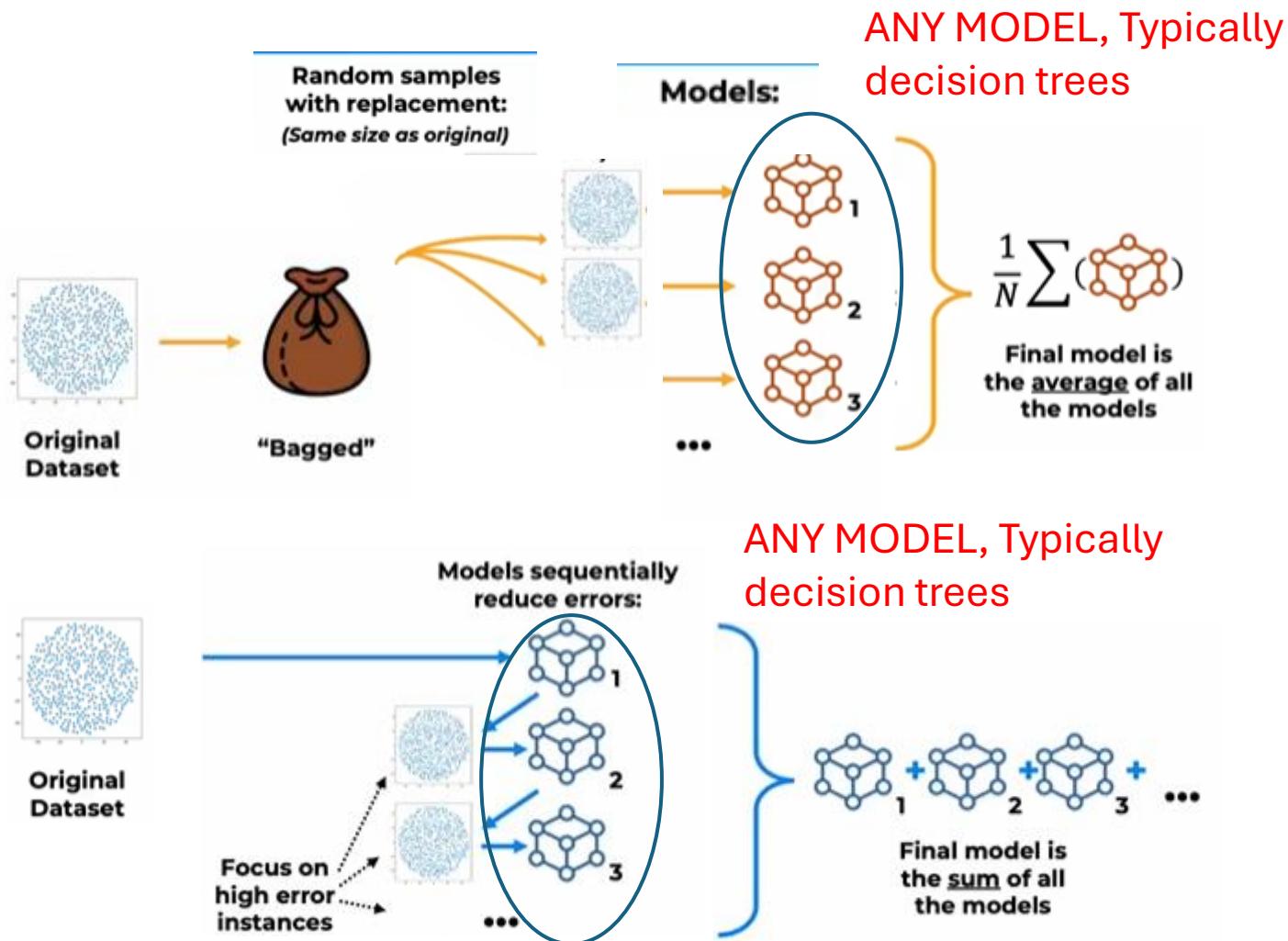
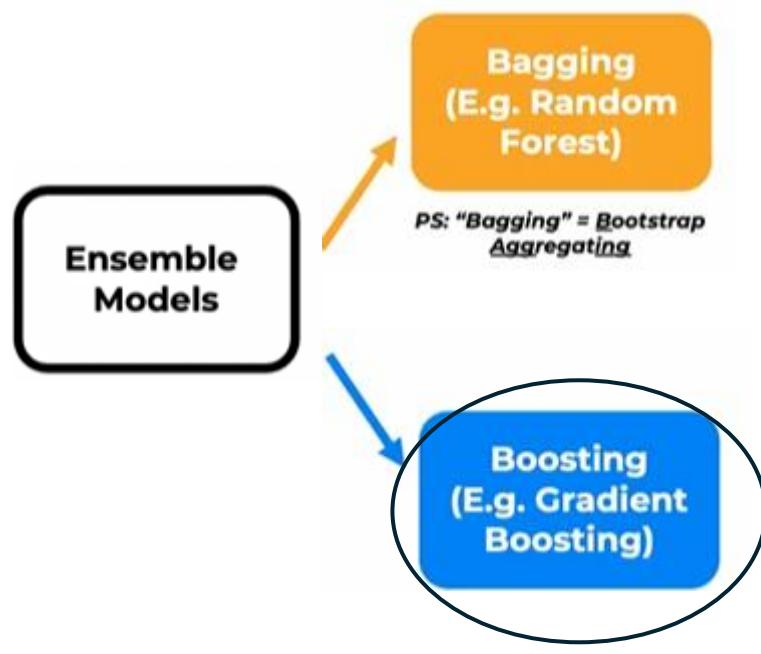


Machine Learning

Ensemble Models



LightGBM (Microsoft, 2016)

Welcome to LightGBM's documentation! ☺

LightGBM is a gradient boosting framework that uses tree based learning algorithms. It is designed to be distributed and efficient with the following advantages:

- Faster training speed and higher efficiency.
- Lower memory usage.
- Better accuracy.
- Support of parallel, distributed, and GPU learning.
- Capable of handling large-scale data.

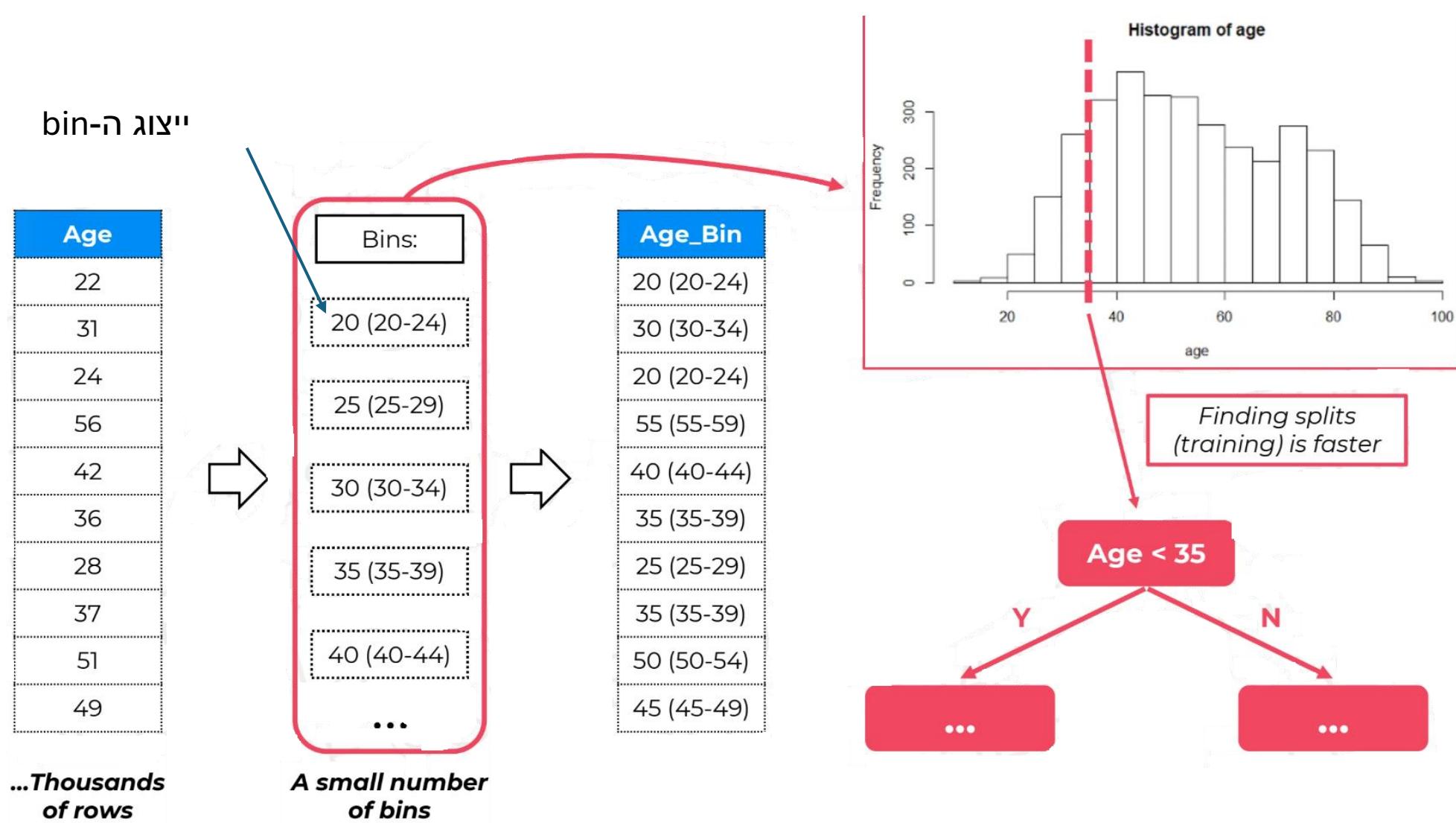
<https://lightgbm.readthedocs.io/en/stable/>



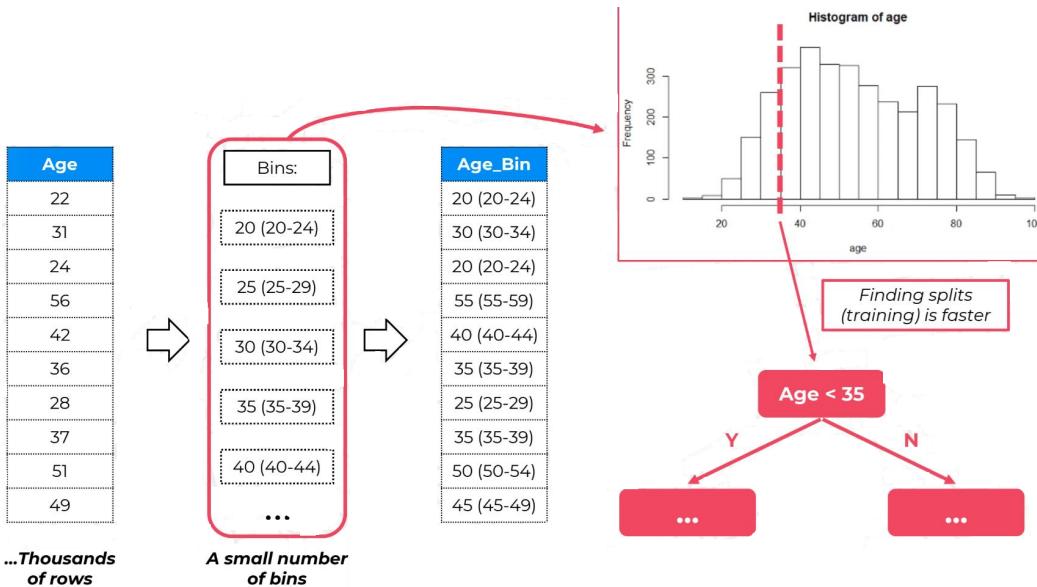
LightGBM

LightGBM compared to XGboost

LightGBM – Histogram Based Training



LightGBM – Histogram Based Training



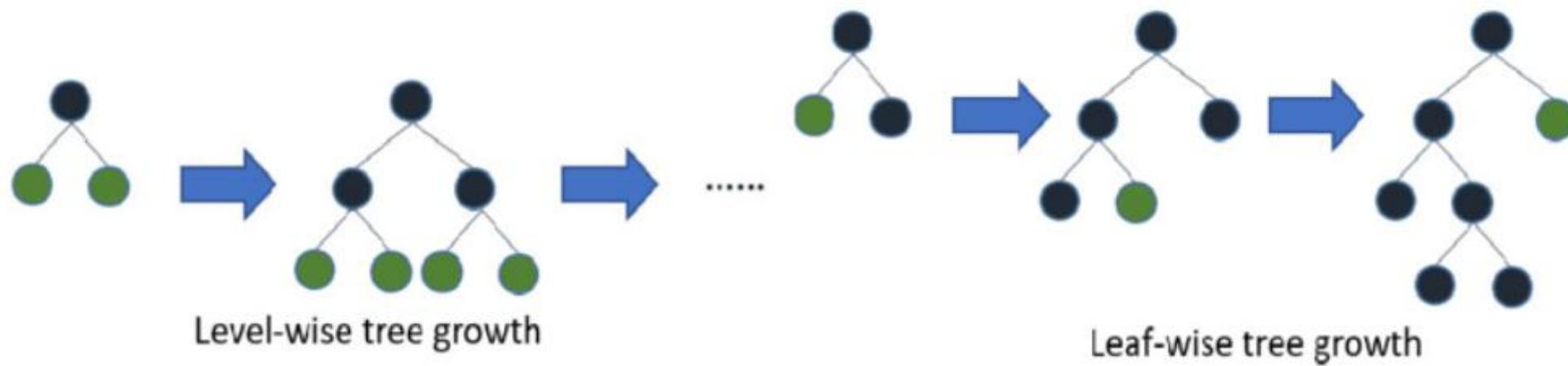
(LightGBM) ב-^מ Histogram-Based Training

- במוקם לבדוק כל ערך רציף אפשרי כדי למצוא את הפיצול הכי טוב — LightGBM מקובץ את הערכים ל-`bins` (טוווחים).
- הוא בונה היסטוגרמה של הערכים, ואז מחפש את נקודת הפיצול הכי טובה בתחום הבינים.

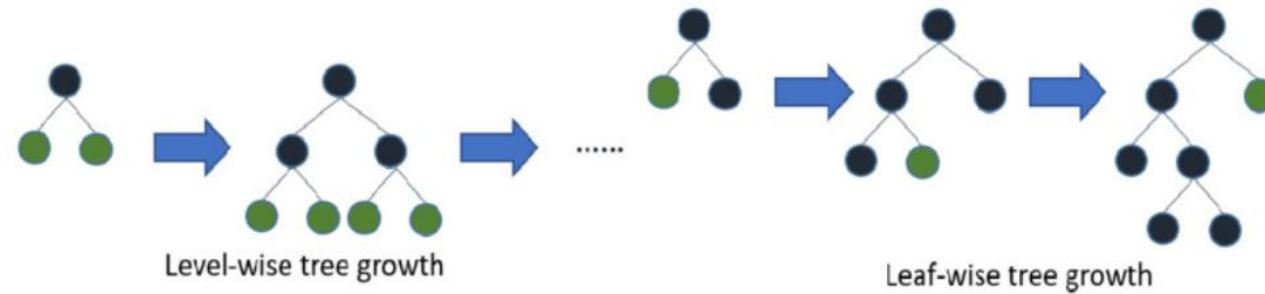
יתרונות:

- ריבבה יותר מהיר
- צורך פחות זיכרון
- סקלביי לדאטה עצום

LightGBM – Leaf Wise Tree Growth



LightGBM – Leaf Wise Tree Growth



1. Level-wise (כמו ב-XGBoost ברירת מחדל)

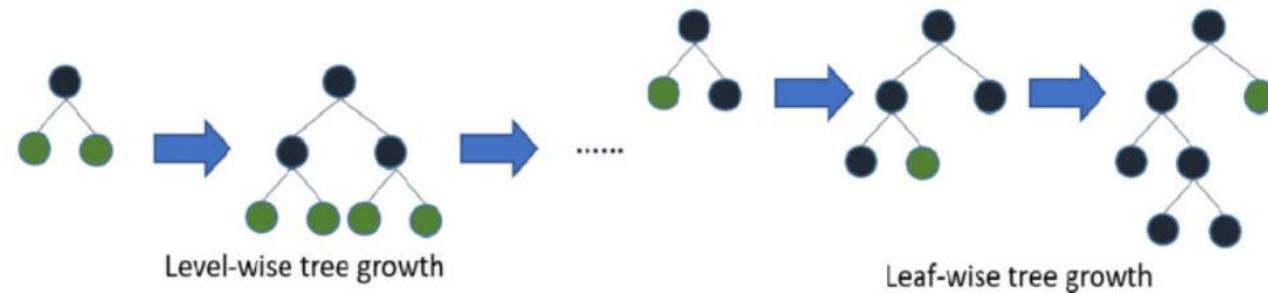
- בונים את העץ שכבה אחר שכבה – כלומר כל הרמות נבנו יחד:
- קודם מפצלים את השורש (root)
- אחר כך את כל הצמתים ברמה השנייה
- ואז את כל הרמה השלישית, וכן הלאה...

נבנה בצורה סימטרית יחסית

יציב

יכול להיות פחות מדויק, כי לא מתחמק היקן שיש כי הרבה טעות.

LightGBM – Leaf Wise Tree Growth



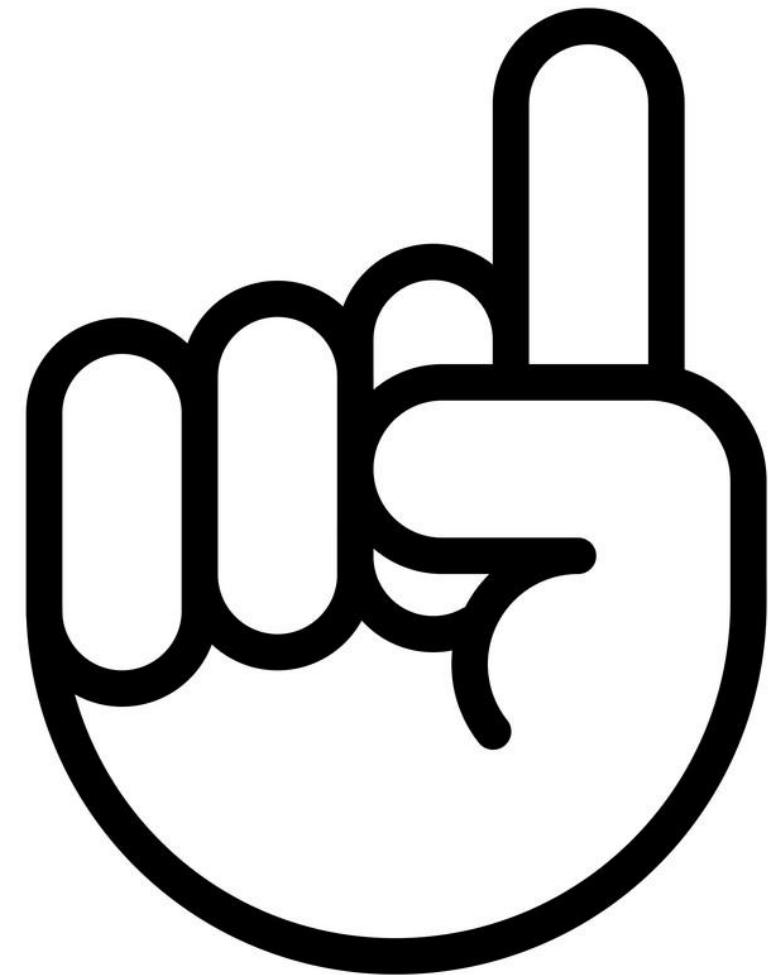
(LightGBM) שיטת Leaf-wise .2

- בכל שלב בוחרים את העלה (leaf) שיביא את שיפור הטעות הכי גדול — ומפצלים רק אותו.
- התוצאה:
- יש ענפים מסוימים בעץ שיגדלו עמוק מאוד
- אחרים ישארו שטוחים

✓ מדויק מאוד – מתמקד באזוריים שקשה לחזות

⚠ נוטה לאוברפייטינג אם לא מגבלים (כגון עצים עמוקים)

lightgbm_classifier



CatBoost (Yandex, 2017)

חכם לפיצ'רים קטגוריאליים CatBoost – Gradient Boosting ✨

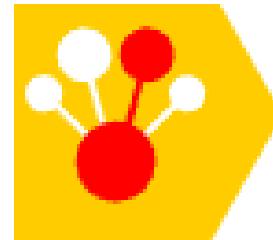
מה זה CatBoost

- ספריית Boosting שפותחה על ידי Yandex.
- מיועדת בעיקר לDATA TABLEAI (structured/tabular).
- תומכת מוצן בפיצ'רים קטגוריאליים – בלי צורך בHot-One או Label Encoding.

```
from catboost import CatBoostRegressor  
model = CatBoostRegressor(cat_features=[0, 2])  
model.fit(X_train, y_train)
```

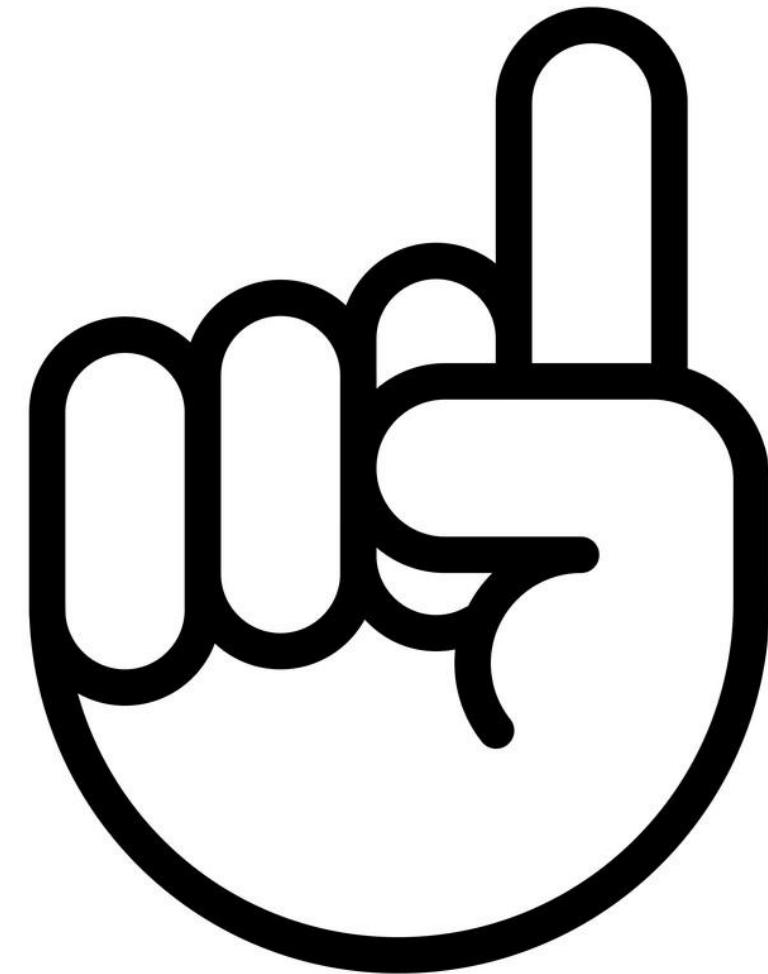
.

<https://catboost.ai/>



CatBoost

catboost_classifier



CatBoost (Yandex, 2017)

```
from catboost import CatBoostRegressor  
model = CatBoostRegressor(cat_features=[0, 2])  
model.fit(X_train, y_train)
```

"העמודות מספר 0 ו-2 הן קטגוריאליות, ולכן CatBoost צריך לטפל בהן בהתאם".

המספרים כאן הם אינדקסים של עמודות, כלומר לפי הסדר ב- `X_train`.

לדוגמה:

Type	Column Name	Index
(<code>male</code> , <code>female</code>) קטגוריאלי	<code>gender</code>	0
מספרי	<code>age</code>	1
(<code>north</code> , <code>south</code>) קטגוריאלי	<code>region</code>	2

-

از במכירה זהה `[2, 0]` [cat_features=[2, 0]]:

• `region` ו- `gender` הן קטגוריות

• `age` הוא רציף (ולא נכנס לרשימה)

Xgboost vs LightGBM vs CatBoost

CatBoost	LightGBM	XGBoost	קריטריון
מהיר מאוד (ב-CPU)	מהיר מוד (histogram)	מהיר	מהירות 
Symmetric trees	Leaf-wise	Level-wise	בנייה עצים 
<input checked="" type="checkbox"/> מלא	<input checked="" type="checkbox"/> CN (מוכנה + מתקדם)	<input checked="" type="checkbox"/> לא (דורש encoding)	תמיכה בפיצ'רים קטגורים 
(Ordered Boosting)	גבואה (עצים عمוקים)	בינונית	טיפול בערכים אבודים 