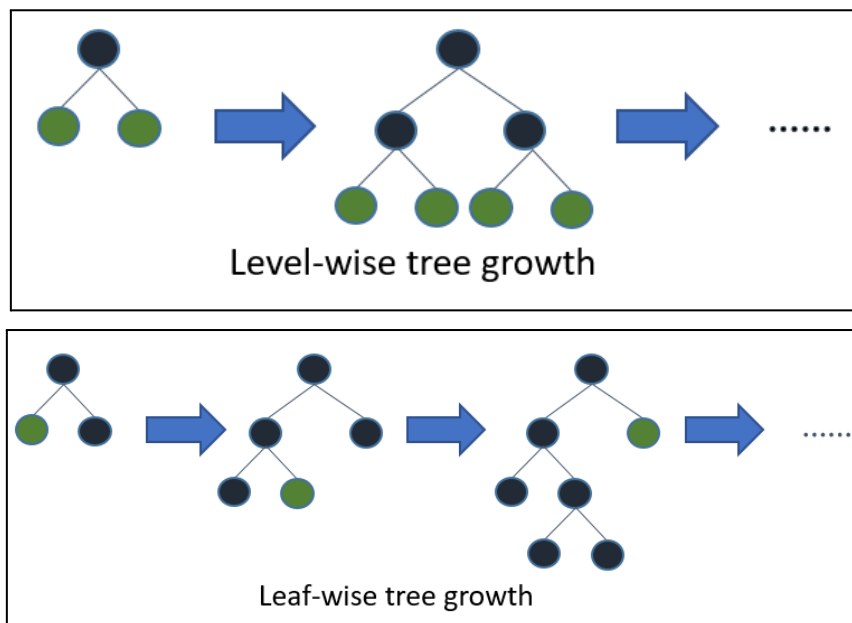


LightGBM (Light Gradient Boosting Machine)

- Gradient boosting machine that combines many weak learners to create a strong predictive model
- Works for classification and regression
- Why is it “Light”?
 - Efficient memory management
 - Speed
 - Fast training time
 - Works on large datasets

How does it work?

- Leaf-wise tree splitting strategy
 - For reference, XGBoost used a level-wise tree splitting strategy
 - **LightGBM is more time efficient**
 - Level-wise splitting makes the tree more stable, and less prone to overfitting → **LightGBM is likely to overfit!**
- The splitting happens at the node with the highest error (loss value)
 - Can result in a deeper and asymmetrical tree
 - More susceptible to overfitting for small datasets (not really applicable for us thankfully)
 - Minimizes cost for the entire model at each iteration



Gradient-based one sided sampling (GOSS)

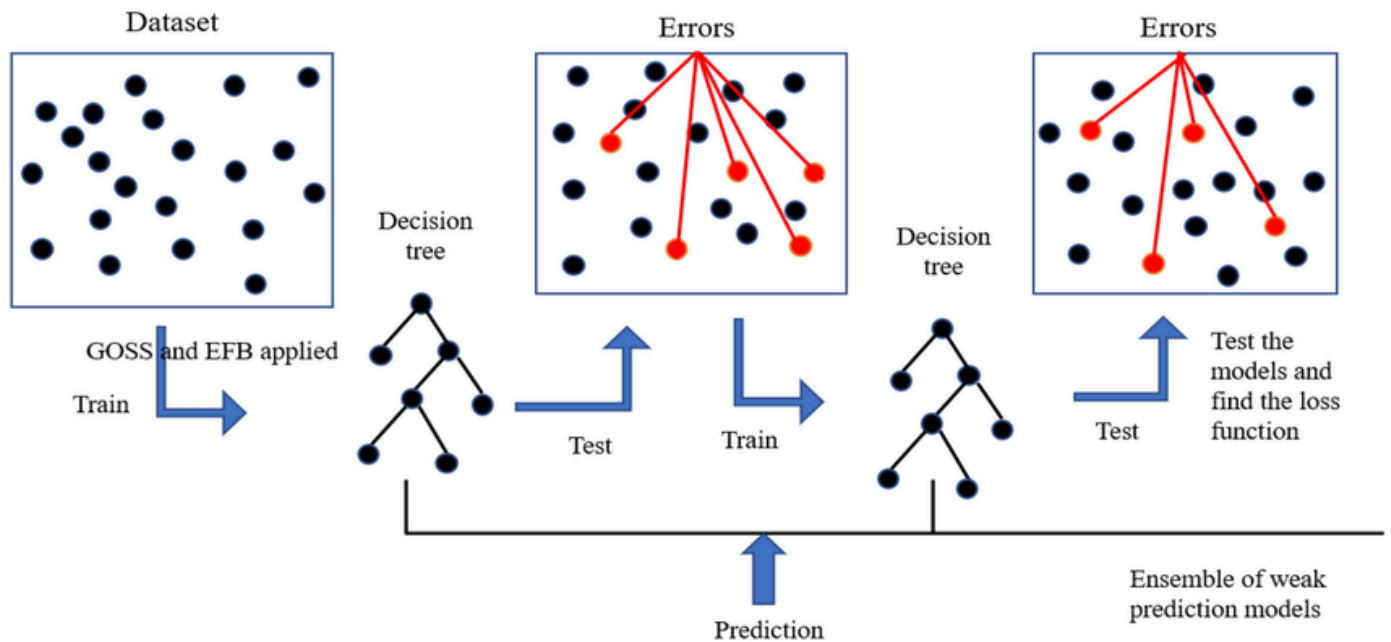
- This is essentially the reason why LightGBM is very time effective
- GOSS keeps in mind data points with high gradients (high cost) and samples ones with small gradients (low cost)

- In this way, GOSS focuses on the data points that are harder to predict, making the training data set smaller → speeding training

Exclusive feature bundling (EFB)

- A way to workaround the increase in dimensionality as there is an increase in categorical data features
 - This is useful for us especially for the themes column
- This technique groups features → mutually exclusive features
 - Features that cannot occur together in a single data point
 - Could work for genres, since a lot have a very low correlation as per Memet's correlation matrix

**There are many built in functions in the model that handle the way that categorical data is used. One-hot encodings can be extremely time consuming and suboptimal for trees specifically, so LightGBM splits categorical features at each node.



Installing: <https://lightgbm.readthedocs.io/en/stable/Python-Intro.html>

Citations

Microsoft. (2025). *Parameters – LightGBM 4.6.0 documentation*. Readthedocs.io.
https://lightgbm.readthedocs.io/en/stable/Parameters.html#categorical_feature

Handling Categorical Features using LightGBM. (2023, November 28).

GeeksforGeeks.

<https://www.geeksforgeeks.org/handling-categorical-features-using-lightgbm/>

AI for Non-Experts - Basics to Applications. (2024, June 27). *LightGBM (Light Gradient Boosting Machine)*. YouTube.

<https://www.youtube.com/watch?v=iUxi1O5MA1g>