



WEBINAR

Live Code-Along: Machine Learning with XGBoost in Python

🕒 Monday, November 15th, 11 AM ET



Lis Sulmont

Workspace Architect at DataCamp

Our Mission

Our mission is to democratize data science education by building the best platform to learn and teach data skills and make data fluency accessible to millions of people and businesses around the world.



Speaker



Lis Sulmont

Workspace Architect

- Master's degree in Computer Science from McGill University focused on computer science education research and applied machine learning
- Worked on the DataCamp curriculum for over 2 years, including instructing Tableau, Power BI, and For Everyone courses
- Now focusing on building out the template and dataset library for DataCamp Workspace!



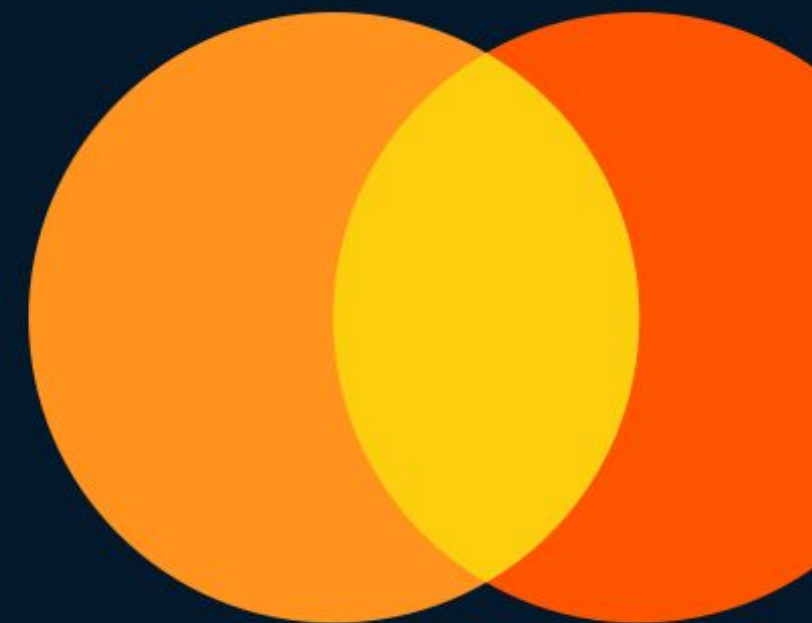
Timeline



- 1 Intro by Lis (approx. 5 min.)
- 2 Getting to know our data (approx. 10 min.)
- 3 Your First XGBoost Classifier (approx. 15 min.)
- 4 Q&A 1 (approx. 10 min.)
- 5 Digging into Parameters (approx. 25 min.)
- 6 Hyperparameter tuning with Random Search (approx. 10 min.)
- 7 Q&A 2 (approx. 10 min.)
- 8 Final Remarks



Gradient Boosting Recap

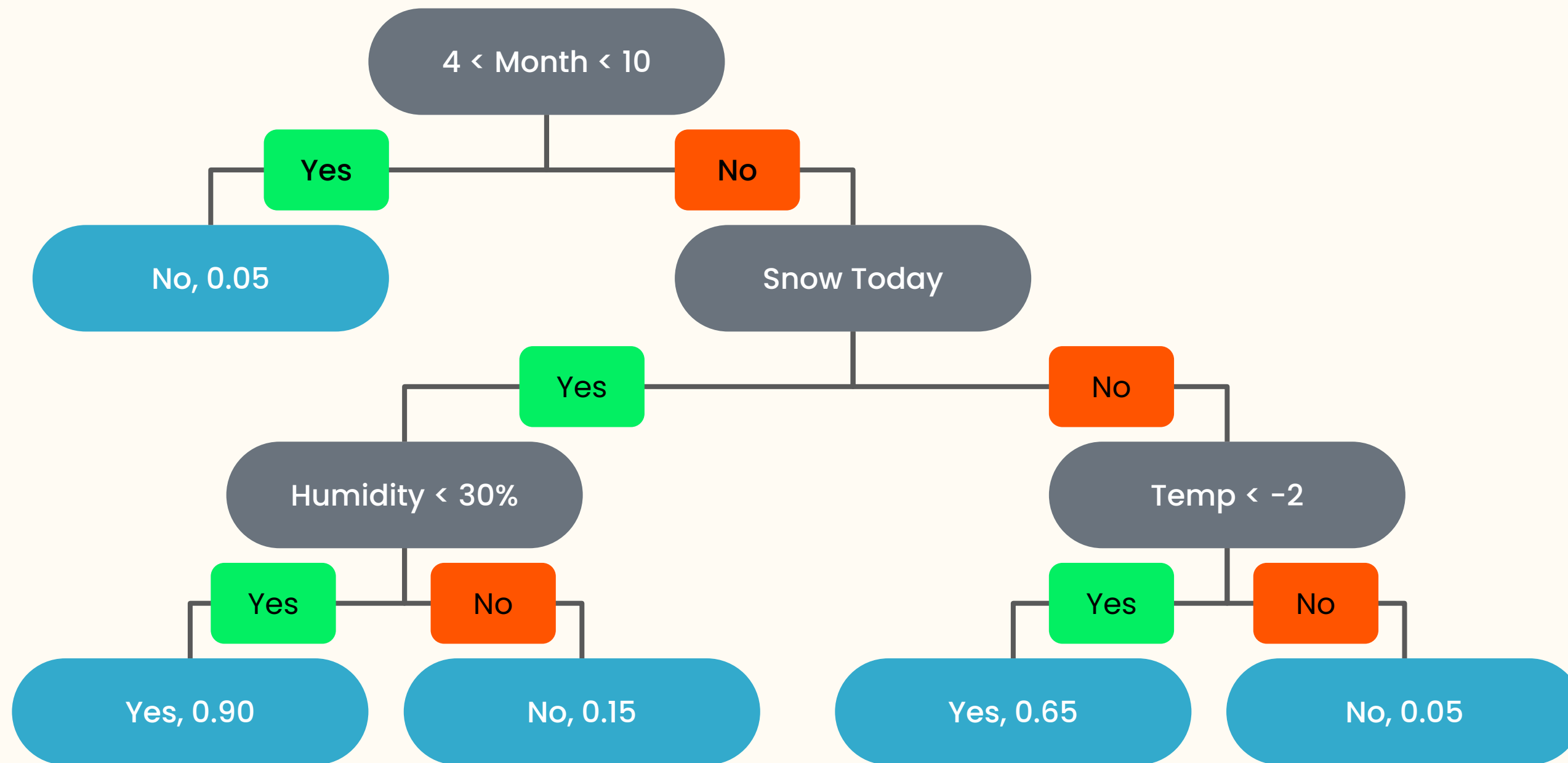


Decision Trees



- Machine learning technique that uses tree structures
- At each decision node, the data is split into two based on a feature
- Split by finding the best information gain possible
- Constructed iteratively until stopping criteria is met
- Works for regression and classification problems
- Classification and Regression Trees (CART)
 - Each leaf node contains a prediction score, not only the decision

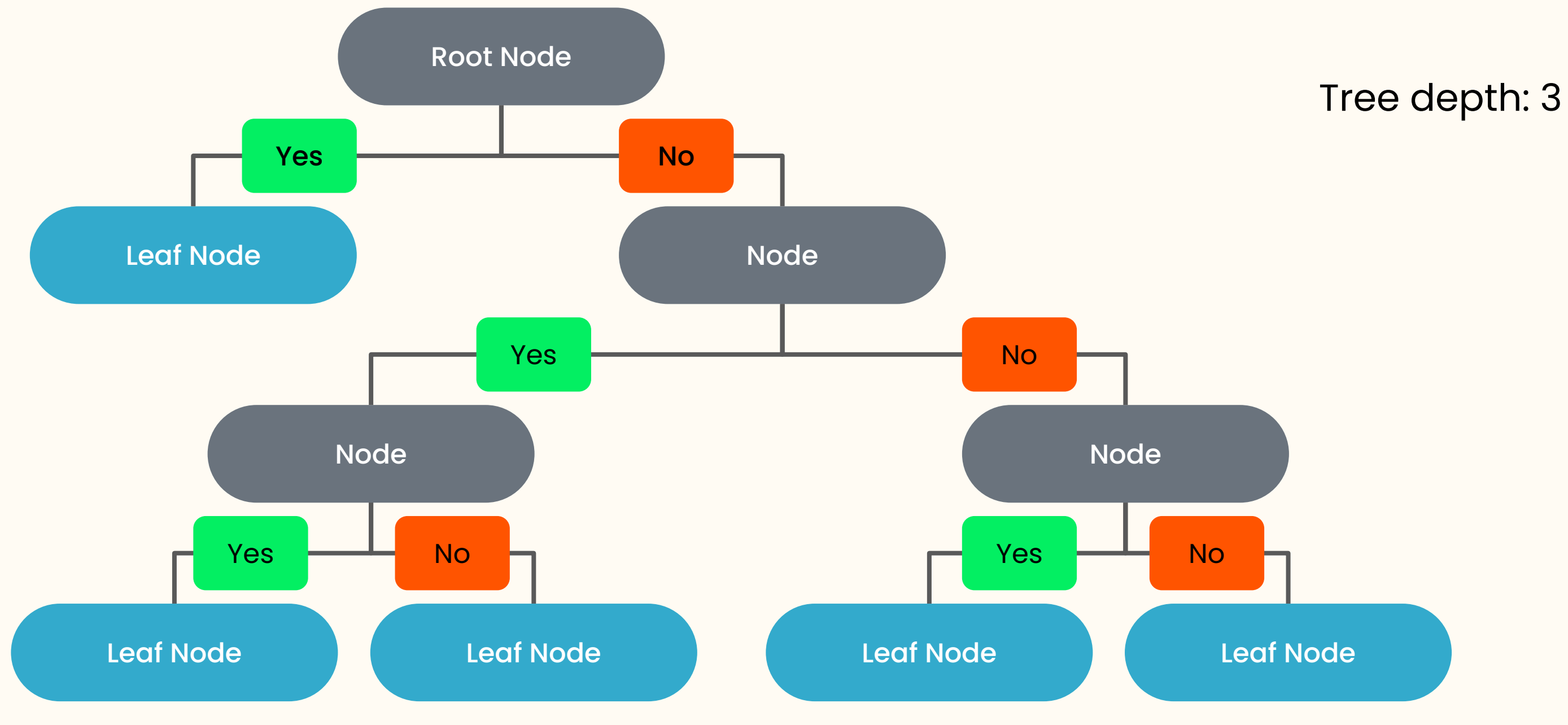
Example: Will it snow tomorrow?



Threshold: 0.5



Tree Terminology

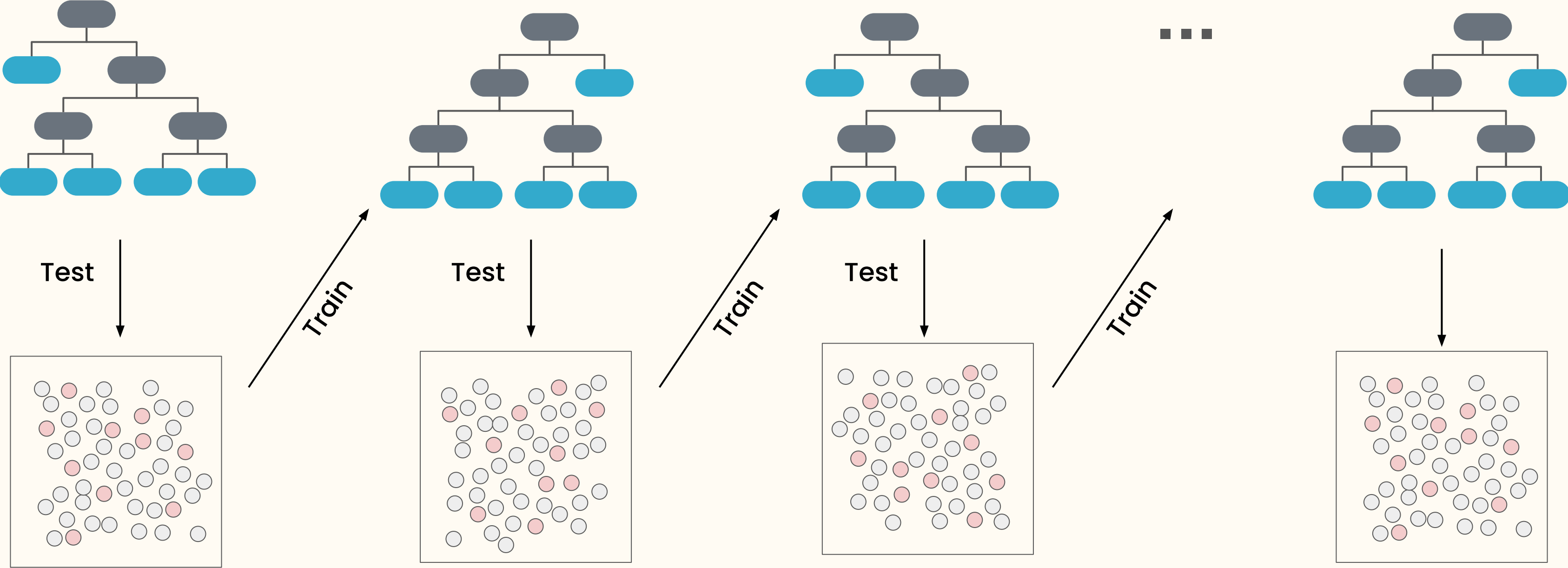


Boosting



- Ensemble method: convert many weak learners into a strong learner
 - Weak learners = slightly better than chance
 - Decision trees are great weak learners
- Boosting is accomplished by
 - Sequentially train weak learners to correct its predecessor

Gradient Boosting



Fit next predictor to the predecessor's residual errors



XGBoost Weak Learners



- Weak learners, aka base learners, boosters
- Decision tree (most common)
 - `booster=gbtree`
- Generalized linear regression
 - `booster=gblinear`



WEBINAR

Live Code-Along: Machine Learning with XGBoost in Python

🕒 Monday, November 15th, 11 AM ET



Lis Sulmont

Workspace Architect at DataCamp