📊 ***Train-Test Split: Why It's Crucial in Machine Learning*** 🚀
As a budding data scientist or machine learning enthusiast, you've probably heard the phrase: "Never test on your training data!" But why is this so important? Let's break it down!

Summary
Train-Test Split is a foundational concept in machine learning that involves splitting your dataset into two parts:
1️⃣ ***Training set:*** Used to train the model.
2️⃣ ***Test set:*** Used to evaluate how well the model performs on unseen data.
This separation ensures that your model learns patterns from one portion of the data and is validated on another, mimicking real-world scenarios where predictions are made on new data.

Highlights
🎯 ***Purpose:*** To prevent overfitting and evaluate the model's generalizability to unseen data.
📐 ***Typical Split Ratios:*** 80% training, 20% testing
70% training, 30% testing (or other custom splits depending on dataset size).
⚖️ ***Balancing Act:*** The training set should be large enough for the model to learn meaningful patterns.
The test set must be sufficiently large to provide reliable performance metrics.

Key Insights
📉 ***Avoid Overfitting:*** A model trained and tested on the same data may memorize the data instead of learning patterns, leading to poor performance on new data.
📊 ***Model Evaluation:*** Performance metrics like accuracy, precision, recall, and RMSE are calculated on the test set to measure the model's real-world readiness.
🔄 ***Variants:*** Validation Set: Sometimes, a third set is created (training, validation, and test) for hyperparameter tuning.
Cross-Validation: Splits the data into multiple training and test sets to improve reliability.

GitHub code : https://github.com/NafisAnsari786/Machine-Learning-Algorithms/blob/main/6%20Train%20Test%20Split/Train%20Test%20Split.ipynb