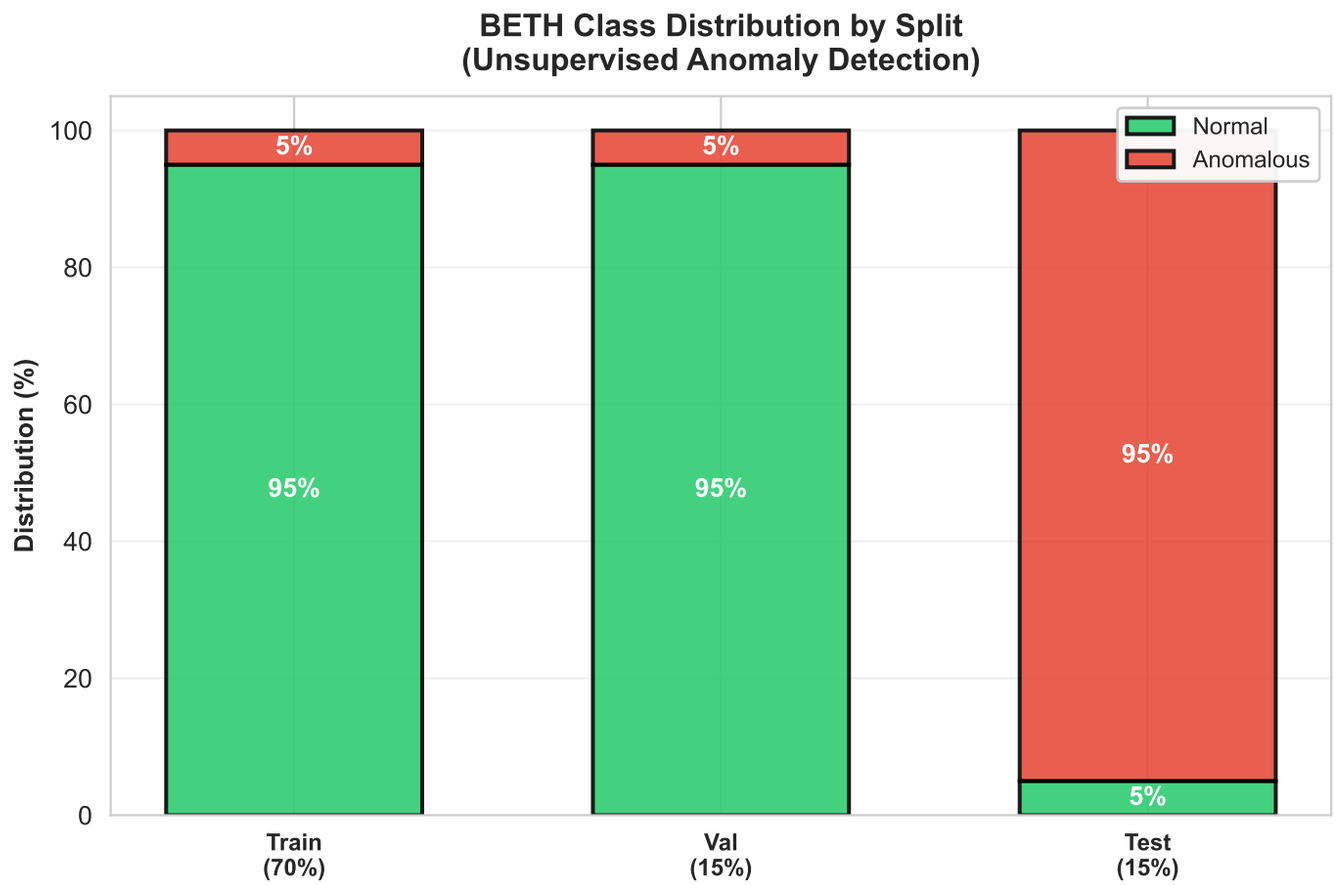
Understanding Our Two Datasets - BETH vs UNSW-NB15

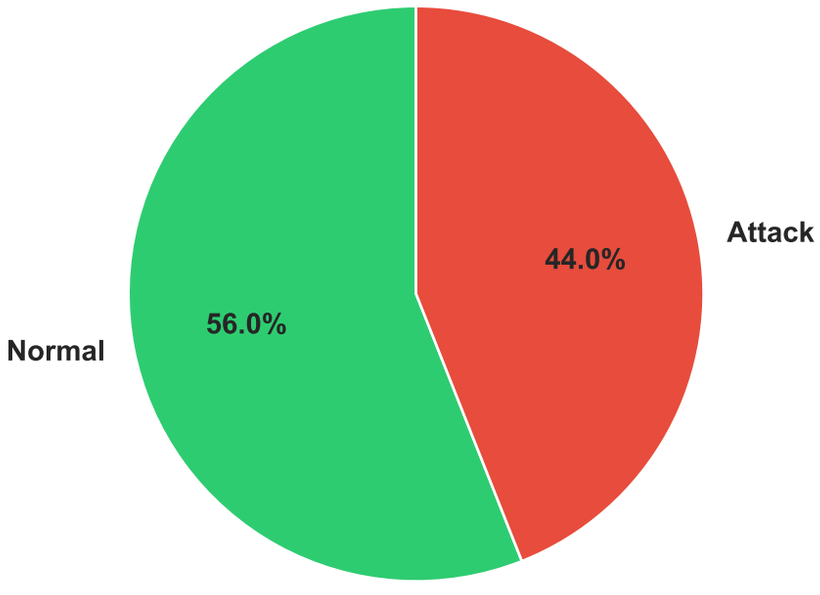
Dataset Characteristics

Dataset	BETH	UNSW-NB15
Domain	System Calls	Network Traffic
Learning Type	Unsupervised	Supervised
Task	Anomaly Detection	Classification
Training Samples	~763K	~175K
Test Samples	~164K	~82K
Features	~40	~45



Train/Val: Overwhelmingly Normal | Test: Overwhelmingly Anomalous

UNSW-NB15 Class Distribution



Models & Techniques

Dataset	Models Used
BETH	K-Means Clustering
	DBSCAN
	Gaussian Mixture Model
UNSW-NB15	Logistic Regression
	Random Forest (RF)
	XGBoost