Accuracy vs ROC-AUC on the Cricketers War Dataset

Accuracy is Misleading on Imbalanced Data

- Compared classifiers using ROC-AUC instead of accuracy

- Investigated the effect of class imbalance

- Verified that high accuracy does not mean good classification

Results

Model	Accuracy	ROC-AUC	Observation
Linear SVM	~96%	low (~0.6)	Classifies almost everything as non-KIA
Decision Tree	~97%	higher (~0.75)	Detects some KIA regions
KNN (k=15)	~97%	highest (~0.80+)	Best at identifying war deaths.

Hyperparameter Tuning Improves KNN Performance

Effect of Number of Neighbors (k) in KNN

- Tried different values of k:
- k = 3
- k = 15 (default in worksheet)
- k = 50

Results

k value	Behavior	Performance	
k = 3	Very flexible, noisy	Overfits	⚠️
k = 15	Balanced	Best performance	✅
k = 50	Too smooth	Misses small KIA regions	⬇️