

PET-TURTLE: A DEEP UNSUPERVISED SVM FRAMEWORK FOR ROBUST LEARNING ON IMBALANCED DATA CLUSTERS

Author's: T. Siddharthan¹, R. Surendiran², K. Sujith³, N. Vanjulavalli⁴

Abstract

The rapid growth of large-scale datasets in modern applications has created significant challenges in identifying meaningful patterns when the data distribution is highly imbalanced. Traditional machine learning models often struggle to detect minority patterns or anomalous structures hidden within dominant clusters. This project proposes PET-TURTLE, a deep unsupervised learning framework that integrates deep representation learning with Support Vector Machine-based anomaly detection to effectively analyze imbalanced data clusters. The proposed system first generates synthetic datasets containing majority clusters, minority micro-clusters, and global outliers to simulate real-world imbalance scenarios. A deep autoencoder network is then trained to learn compact latent representations of the data in an unsupervised manner. These latent embeddings capture the intrinsic structure of the dataset and reduce noise and dimensionality. Subsequently, K-Means clustering is applied in the latent space to discover underlying cluster groups without requiring labeled data. To model cluster boundaries and detect abnormal instances, a One-Class Support Vector Machine (OC-SVM) is trained separately for each discovered cluster. This cluster-wise support vector modeling enables the system to learn tight decision boundaries around normal data distributions while identifying points that fall outside these boundaries as potential anomalies. The anomaly scores from cluster-specific SVMs are aggregated to produce a global anomaly detection score for each data instance.