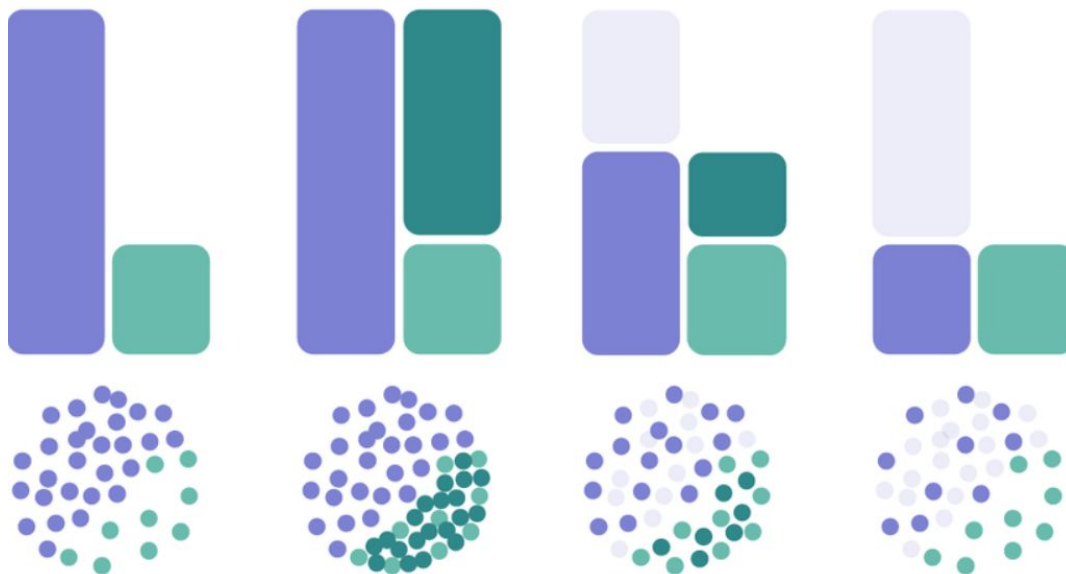


Data Sampling for Imbalanced Datasets

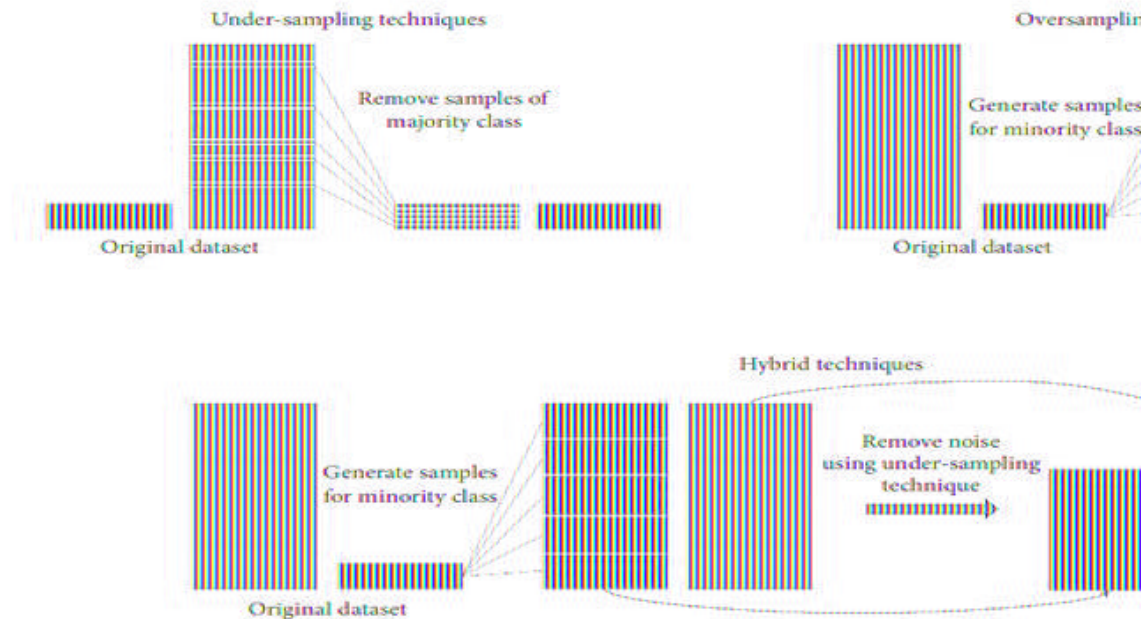


Nội dung

1. Failure of k-Fold Cross-Validation
2. Data Sampling

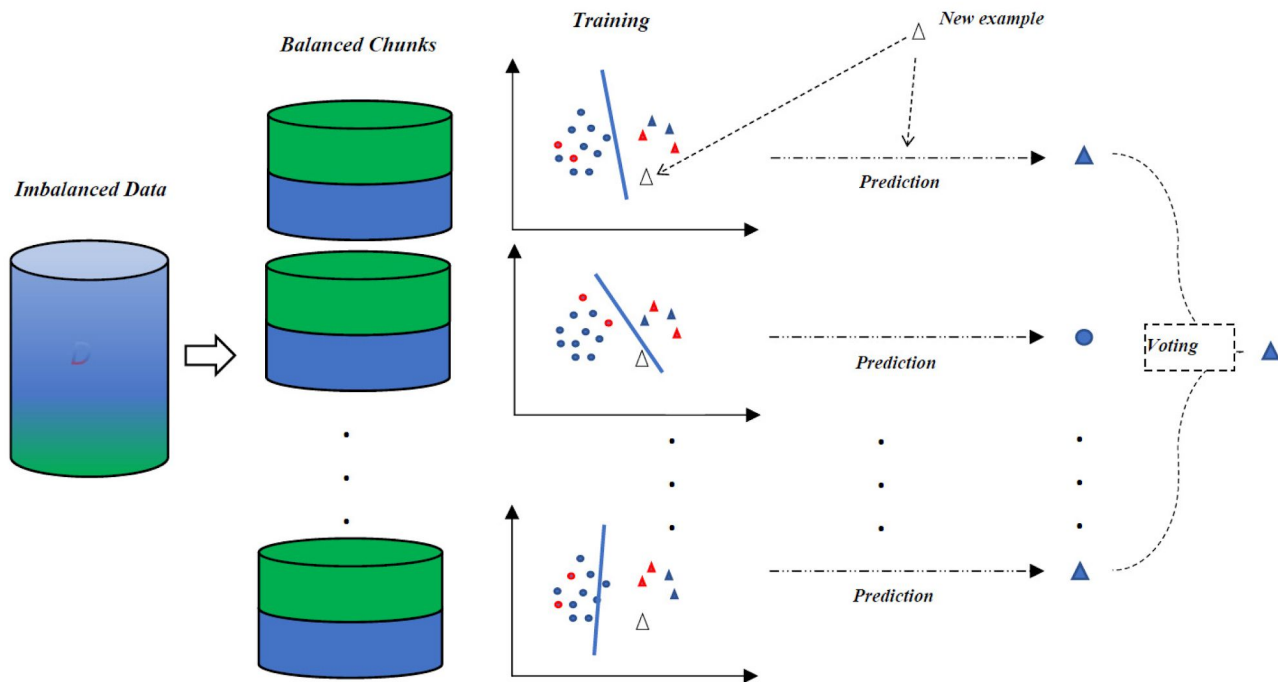
Data Sampling

Imbalance Problem - Một số phương pháp giải quyết imbalance



Data Sampling

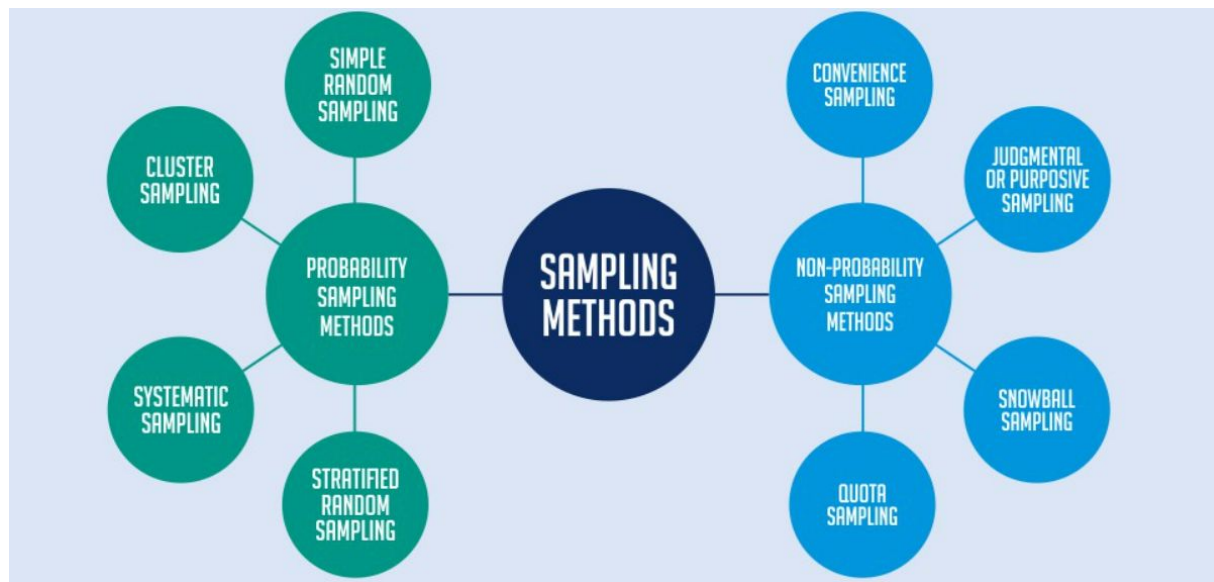
Imbalance Problem - Một số phương pháp giải quyết imbalance



Data Sampling

Có 3 kỹ thuật Data Sampling phổ biến:

- Oversampling Techniques
- Undersampling Techniques
- Combinations of Techniques



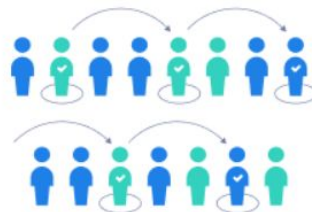
Tour of Model Evaluation Metrics

Random Data Sampling - Random Sampling

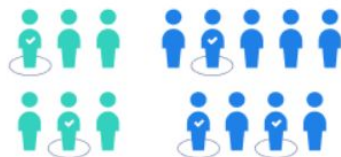
Simple random sample



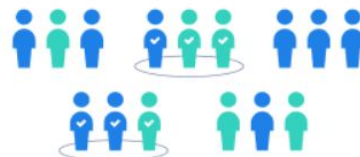
Systematic sample



Stratified sample

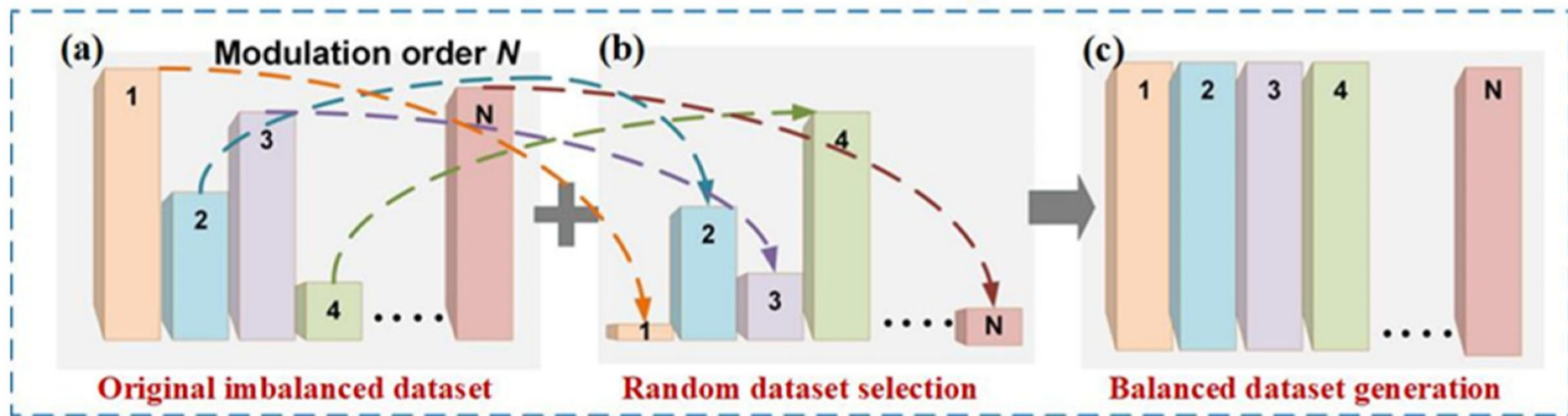


Cluster sample



Tour of Model Evaluation Metrics

Random Data Sampling - Random Oversampling



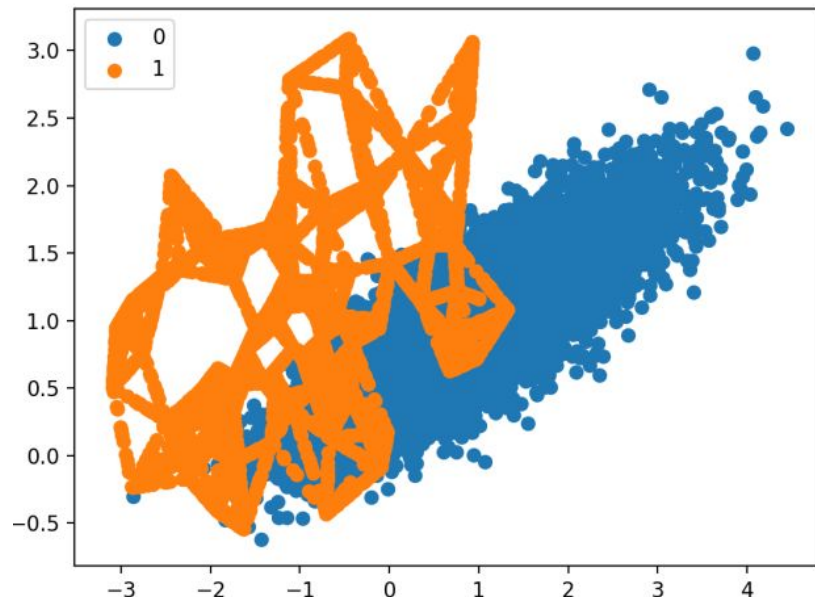
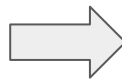
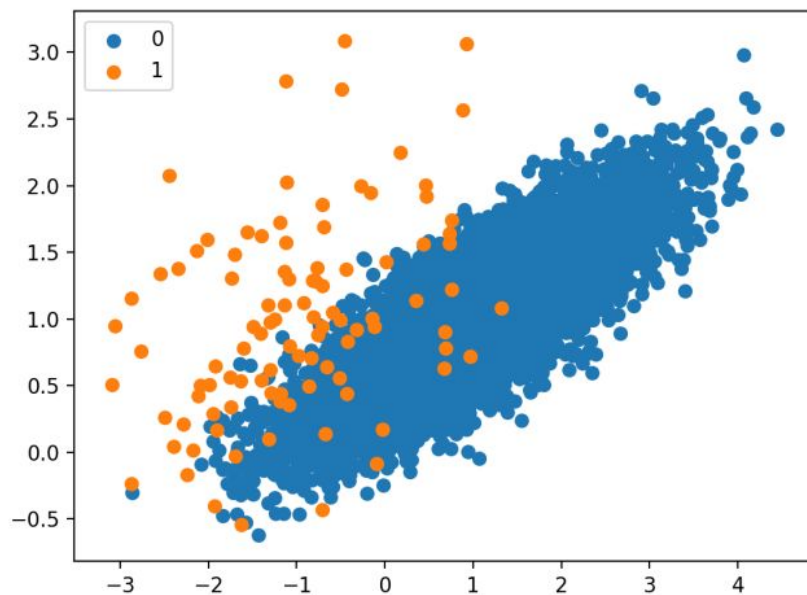
Tour of Model Evaluation Metrics

Random Data Sampling - Random Undersampling



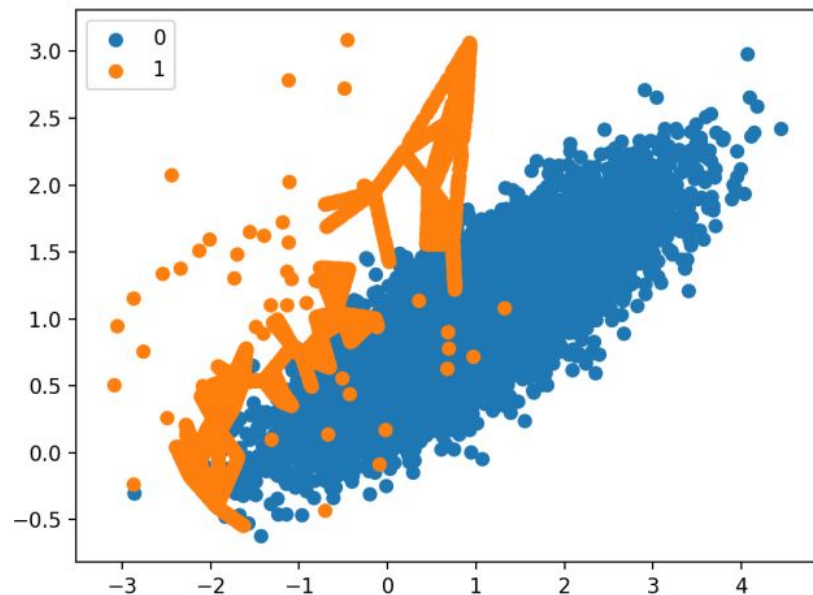
Tour of Model Evaluation Metrics

Oversampling Methods - SMOTE

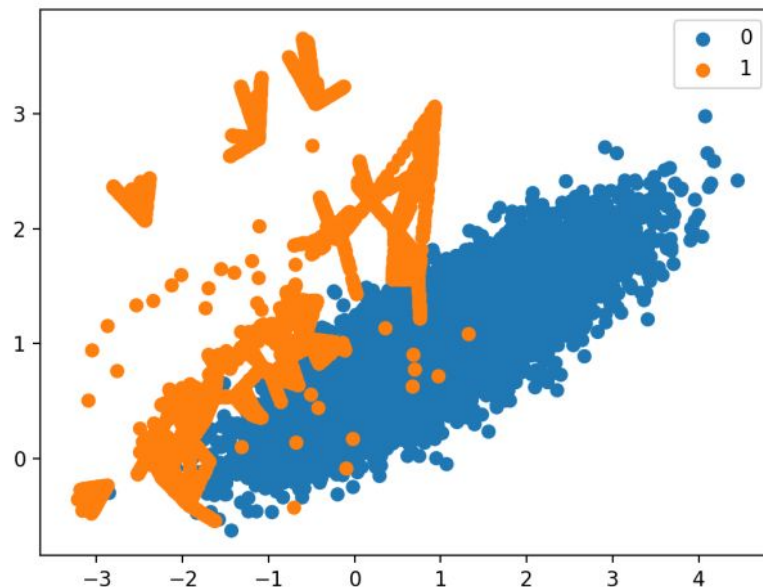


Tour of Model Evaluation Metrics

Oversampling Methods - SMOTE With Selective Sample Generation



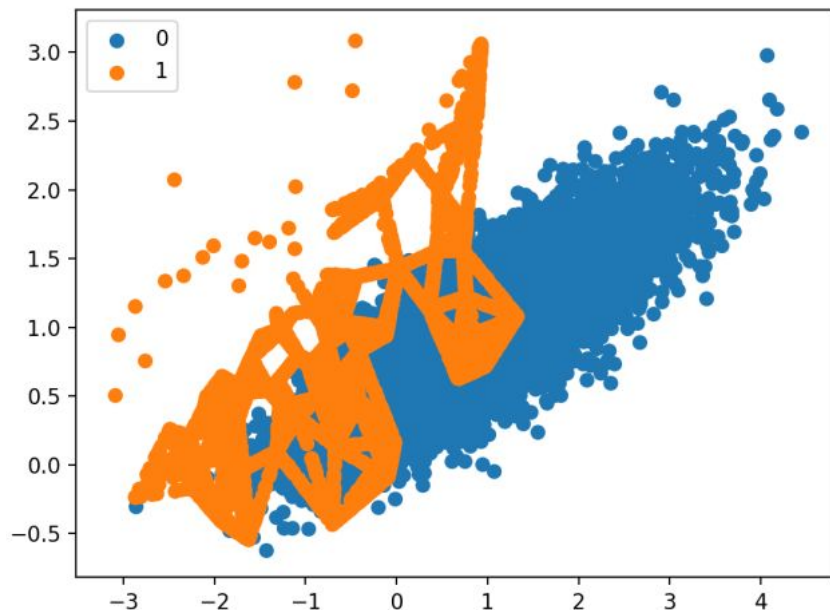
Borderline-SMOTE



Borderline-SMOTE SVM

Tour of Model Evaluation Metrics

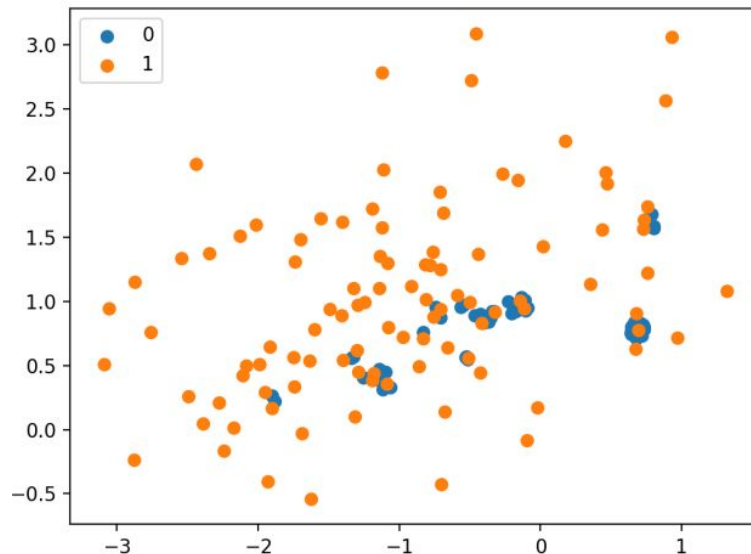
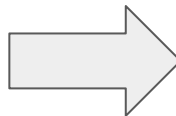
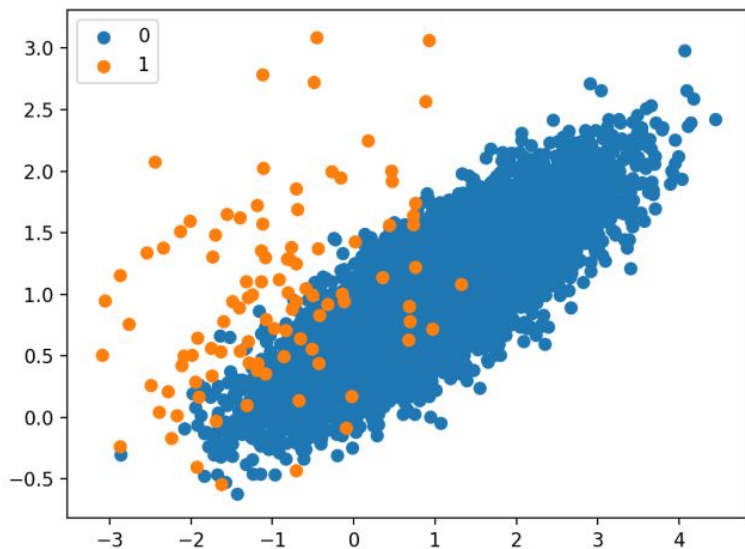
Oversampling Methods - SMOTE With Selective Sample Generation



Adaptive Synthetic Sampling

Tour of Model Evaluation Metrics

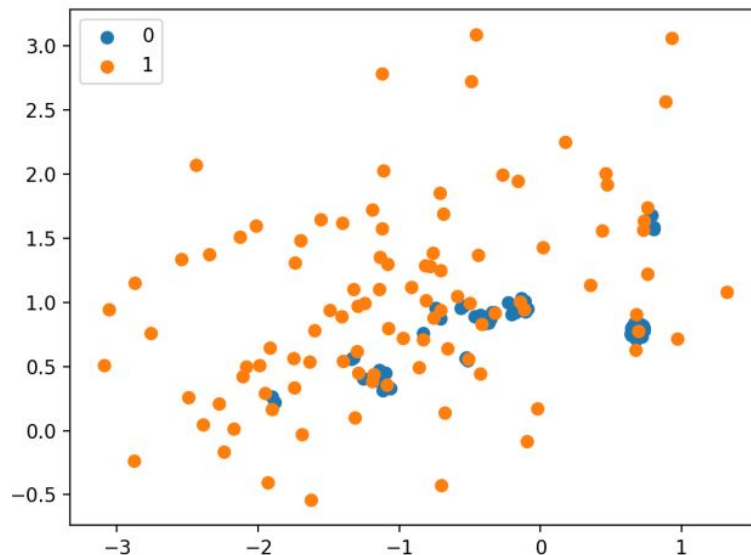
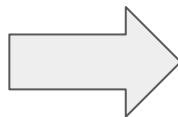
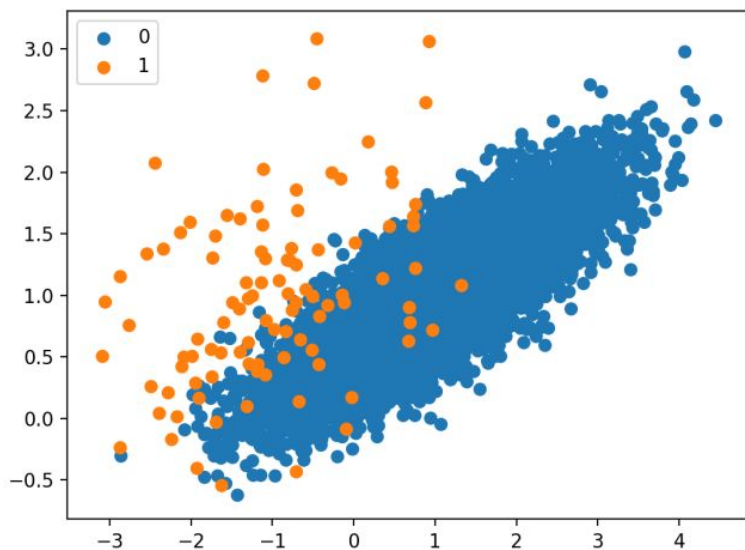
Undersampling Methods - Near Miss Undersampling



Adaptive Synthetic Sampling

Tour of Model Evaluation Metrics

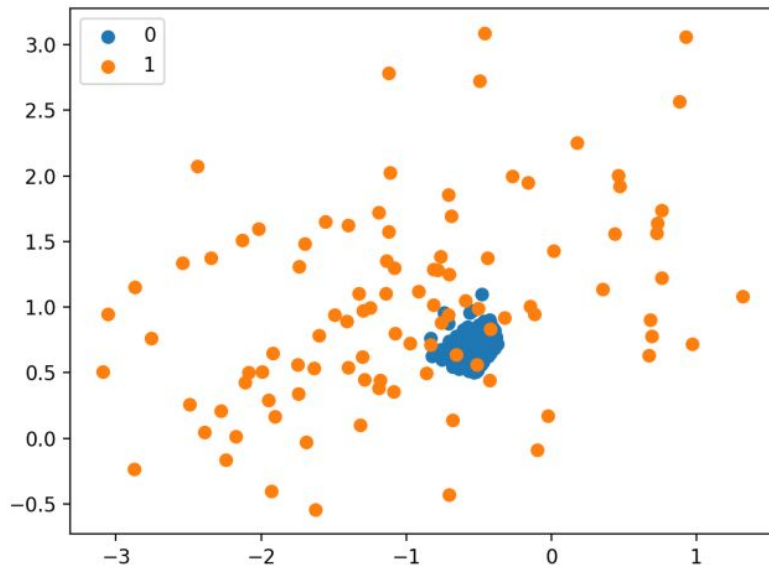
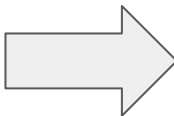
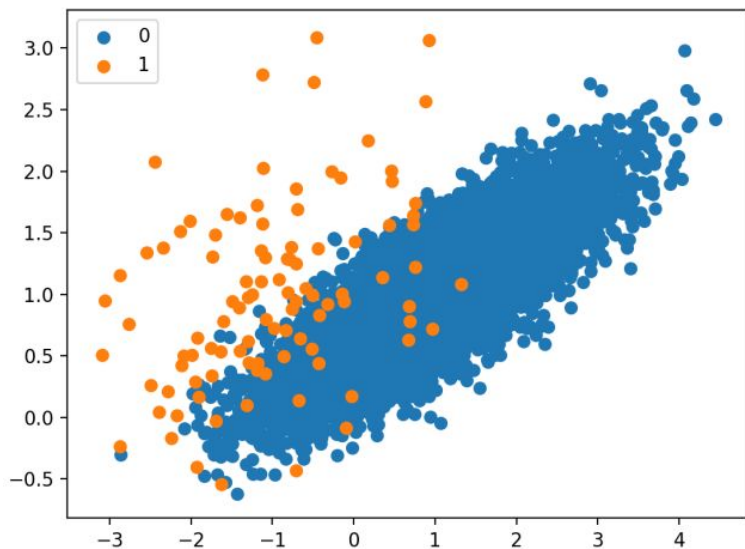
Undersampling Methods - Near Miss Undersampling



NearMiss-1

Tour of Model Evaluation Metrics

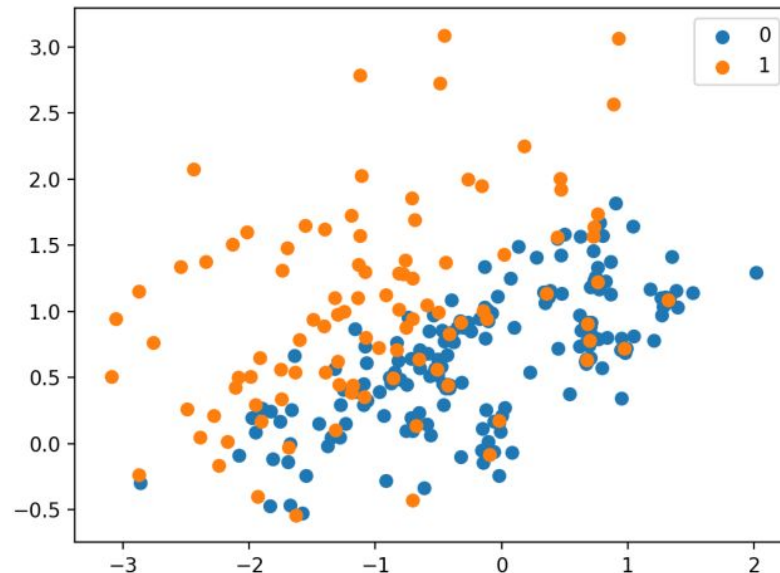
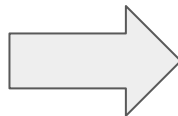
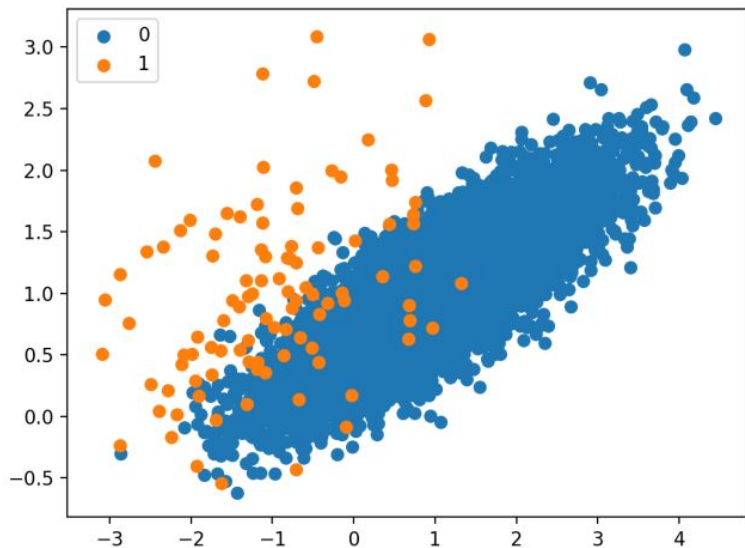
Undersampling Methods - Near Miss Undersampling



NearMiss-2

Tour of Model Evaluation Metrics

Undersampling Methods - Condensed Nearest Neighbor Rule



Tour of Model Evaluation Metrics

Undersampling Methods - One-Sided Selection

