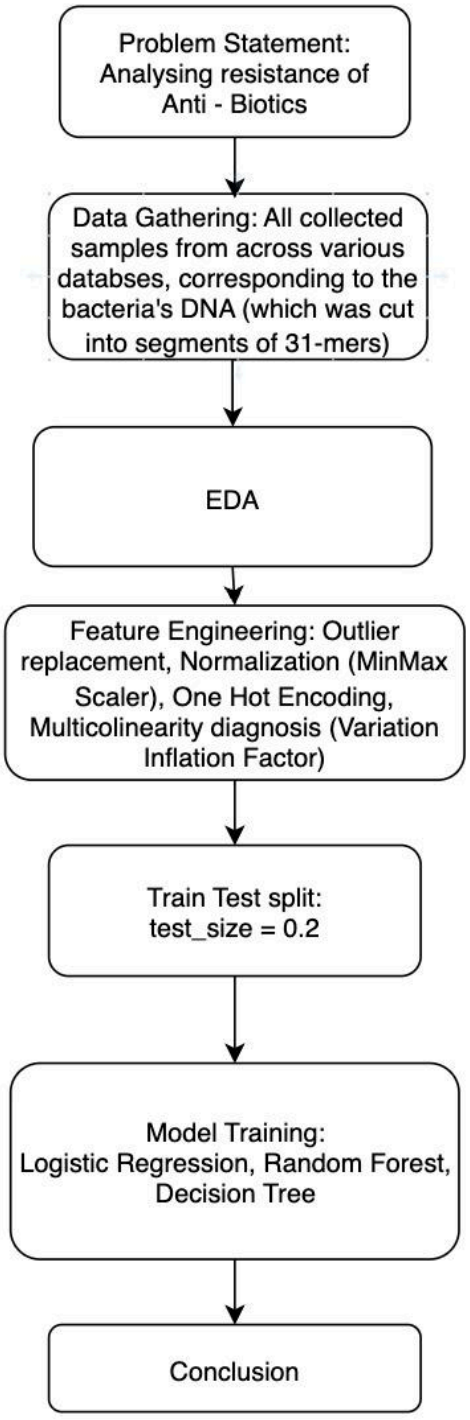
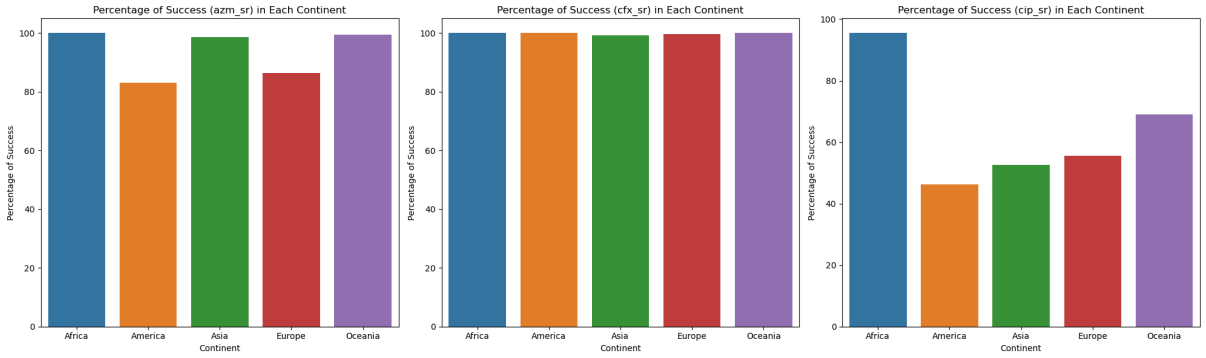
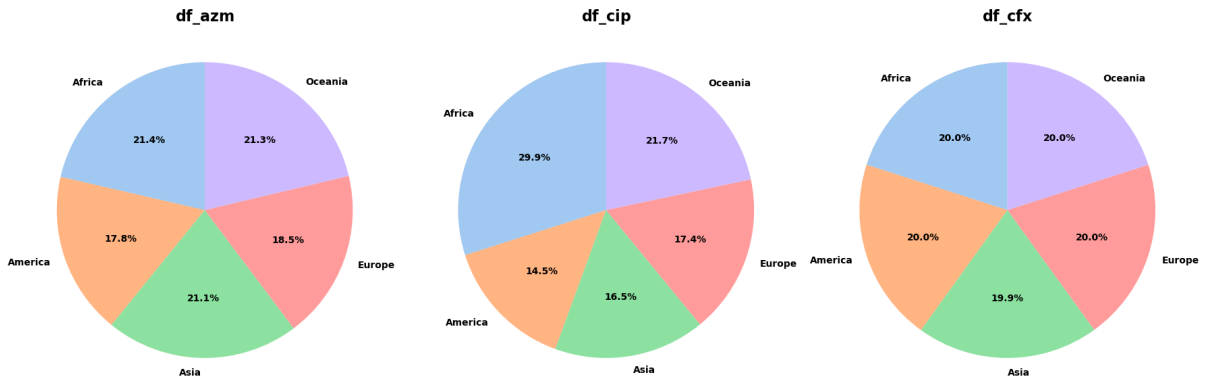


Exploratory Data Analysis:

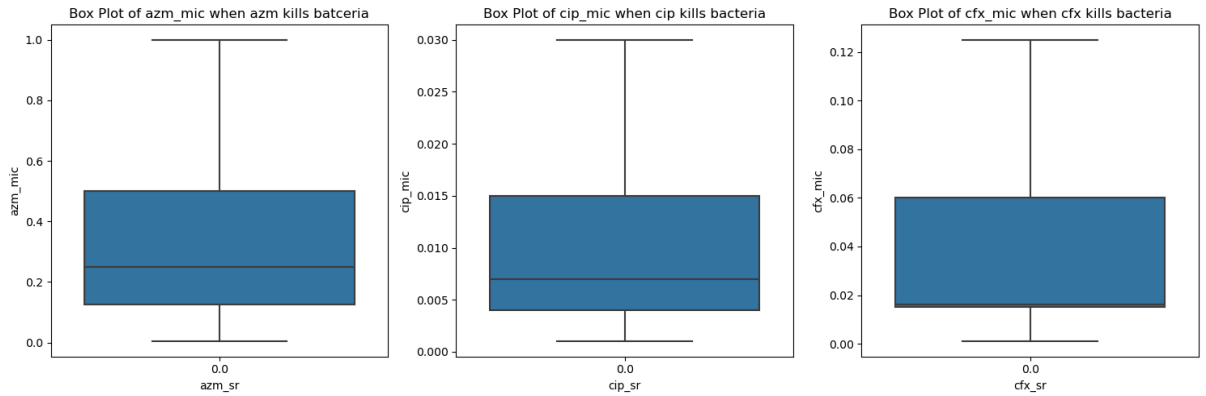
A) Continent Vs anti - biotic





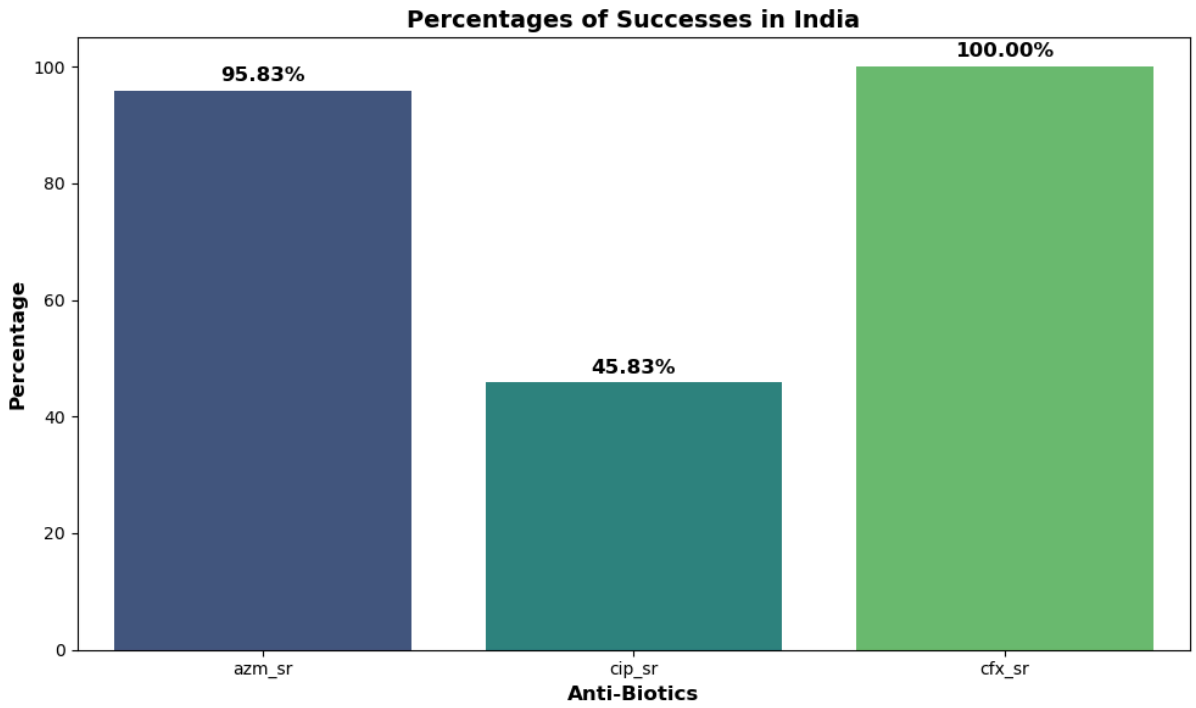
1) **Cfx is useful throughout.**
2) **Azm is most useful in africa, asia and oceania**
3) **CIP is most useful in Africa**

B) MIC values vs Concentration:



1) **Cfx : Successful in moderate concentrations (0.02 to 0.06)**
2) **Azm : Successful in high concentrations (0.175 to 0.5)**
3) **Cip : Successful in small concentrations (0.005 to 0.015)**

C) Performance in India



1) In India, Most successful is Cfx, followed by Azm and lastly Cip

Feature Engineering

The following techniques are used in case of Categorical and Continuous variables to verify relation with respect to the target variable:

target \ features	Categorical	Numerical
Categorical	Chi squared Mutual information	t-test Mutual information
Numerical	t-test Mutual information	Pearson correlation Spearman's rank correlation Mutual information

Fig 1: Tests and techniques to check relation between different types of variable

Source: "Feature Selection Techniques" [5]

Variable	Type	Test	Conclusion
Continent	Categorical	Chi2	Yes, there is a relation
NG Mast	Continuous	Point-Biserial	No, there is no relation

		Correlation	
Group	Continuous	Point-Biserial Correlation	Yes, there is no relation.

Regression Analysis

Training set of Azm has 2759 examples

Test set of Azm has 690 examples

Training set of Cip has 2452 examples

Test set of Cip has 614 examples

Training set of Cfx has 2700 examples

Test set of Cfx has 675 examples

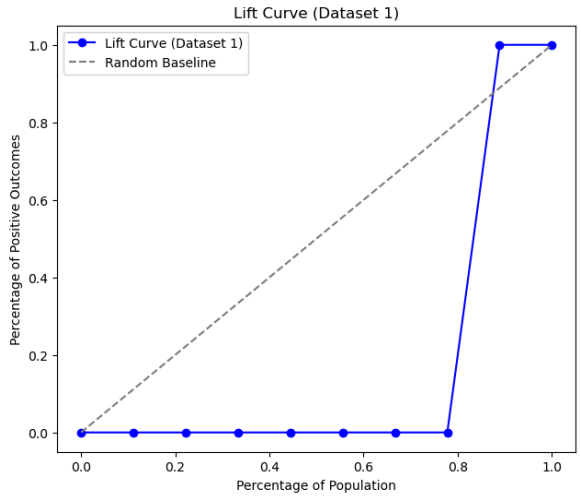
1) Logistic Regression:

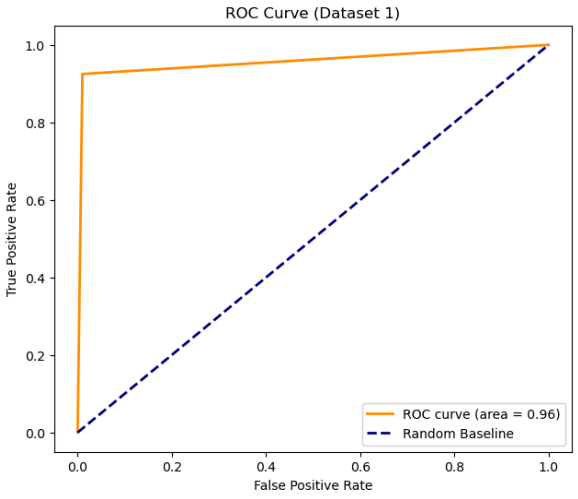
Best Parameters (AZM): {'C': 100, 'penalty': 'l2'}

Best Cross-Validated Accuracy (AZM): 1.0

Test Accuracy (AZM): 1.0

ROC AUC (AZM): 1.0



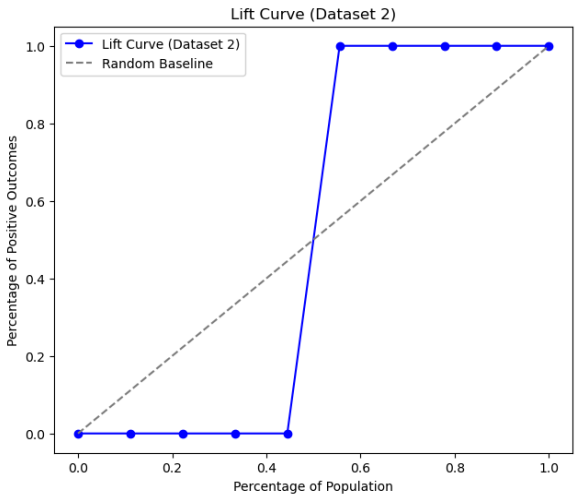


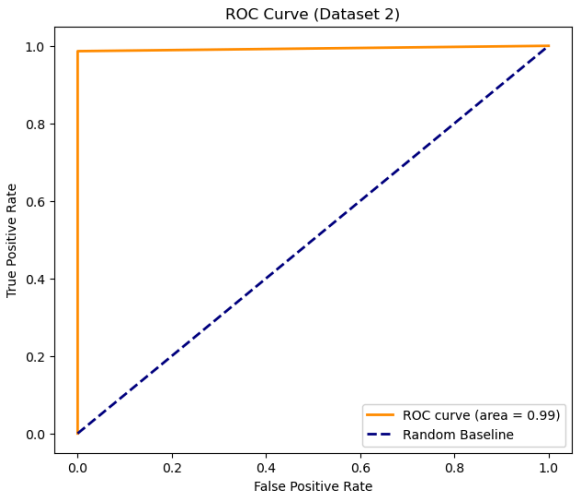
Best Parameters (CIP): {'C': 100, 'penalty': 'l2'}

Best Cross-Validated Accuracy (CFX): 0.9979641612742658

Test Accuracy (CIP): 0.996742671009772

ROC AUC (CIP): 0.996551724137931



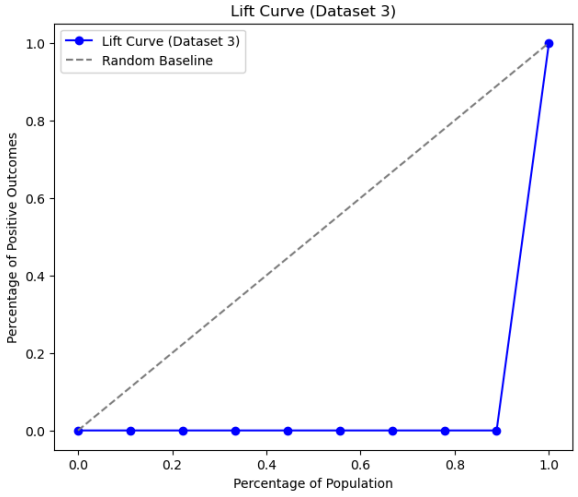


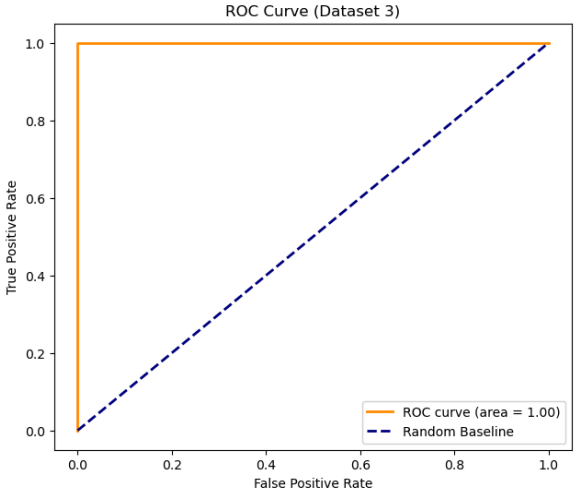
Best Parameters (CFX): {'C': 100, 'penalty': 'l2'}

Best Cross-Validated Accuracy (CFX): 0.9651851851851851

Test Accuracy (CFX): 0.9555555555555556

ROC AUC (CFX): 0.7284546805349182





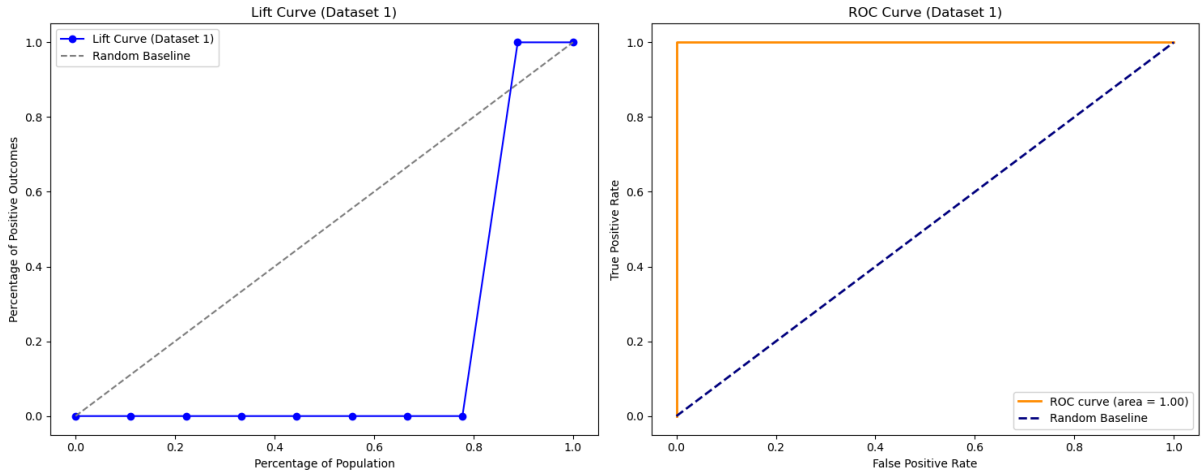
2) Random Forest Classifier

Best Parameters (AZM): {'bootstrap': True, 'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100, 'verbose': 0}

Best Accuracy (AZM): 1.0

Test Accuracy (AZM): 1.0

ROC AUC (AZM): 1.0

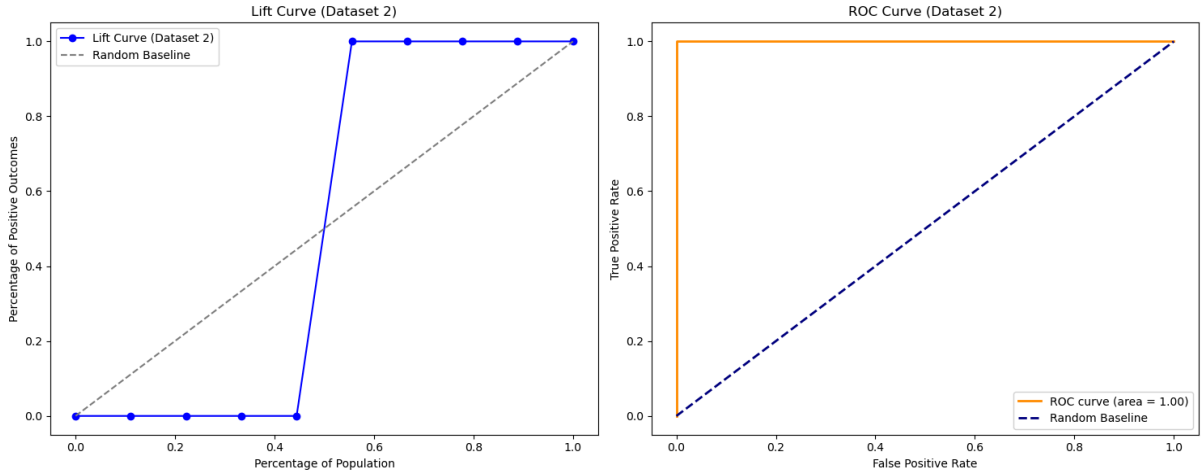


Best Parameters (CIP): {'bootstrap': True, 'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100, 'verbose': 0}

Best Accuracy (CIP): 1.0

Test Accuracy (CIP): 1.0

ROC AUC (CIP): 1.0

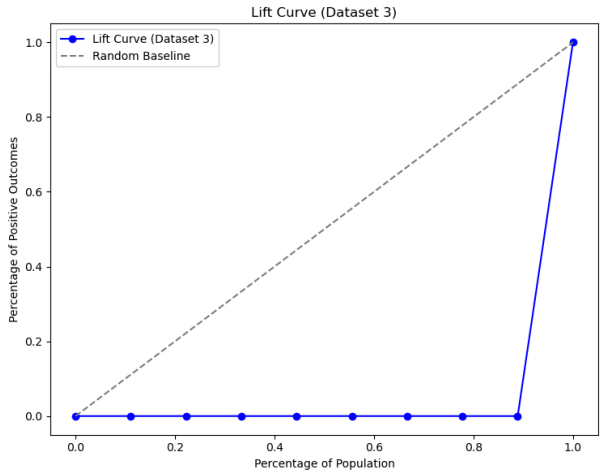


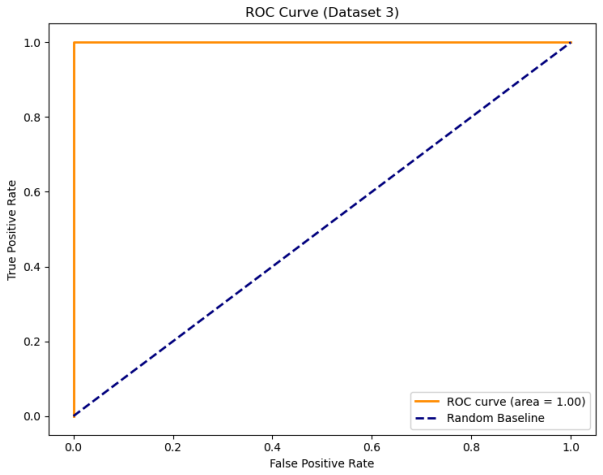
Best Parameters (CFX): {'bootstrap': True, 'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100, 'verbose': 0}

Best Accuracy (CFX): 0.9818518518518518

Test Accuracy (CFX): 0.9822222222222222

ROC AUC (CFX): 0.7418276374442793





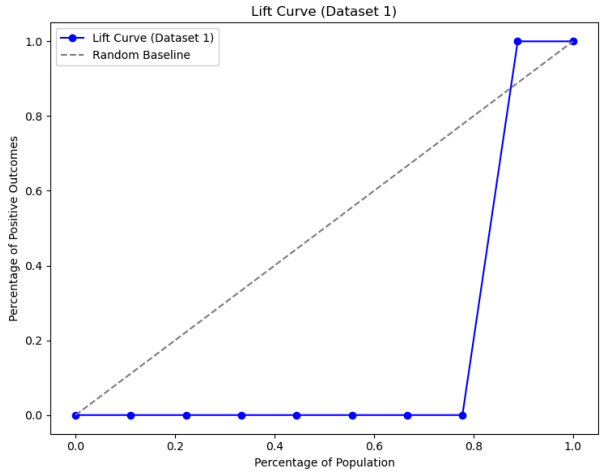
3) Decision Tree Algorithm

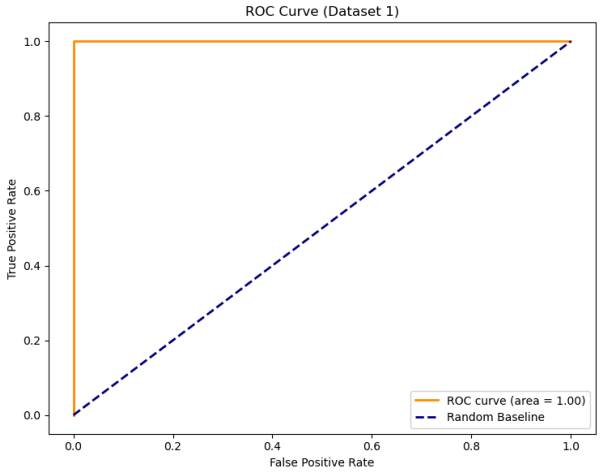
Best Parameters (AZM): {'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 2}

Best Accuracy (AZM): 1.0

Test Accuracy (AZM): 1.0

ROC AUC (AZM): 1.0



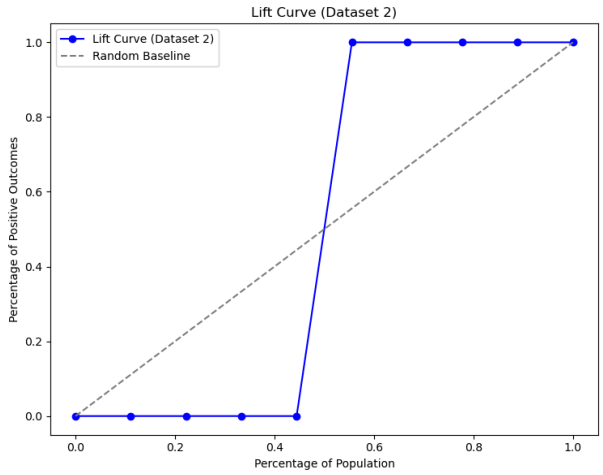


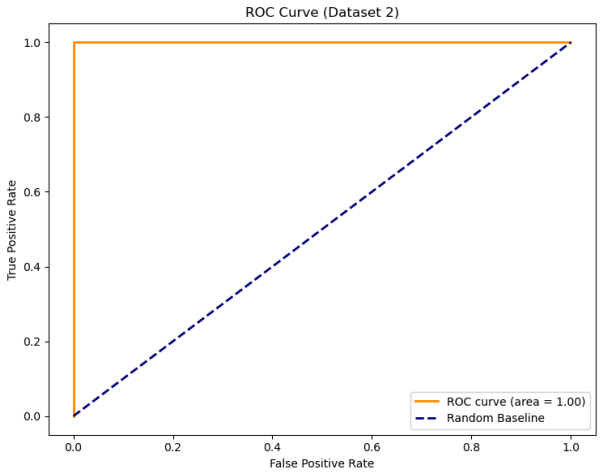
Best Parameters (CIP): {'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 2}

Best Accuracy (CIP): 1.0

Test Accuracy (CIP): 1.0

ROC AUC (CIP): 1.0





Best Parameters (CFX): {'class_weight': None, 'criterion': 'gini', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 2}

Best Accuracy (CFX): 0.9988888888888889

Test Accuracy (CFX): 0.997037037037037

ROC AUC (CFX): 0.5