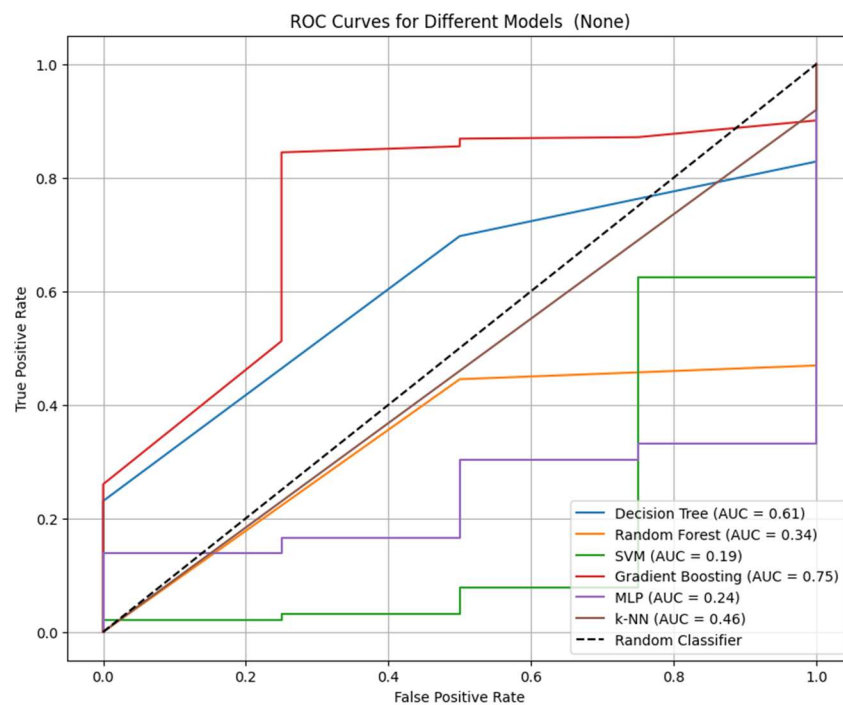# CHOCOLATE AND MAGIC MUSHROOMS DATASETS
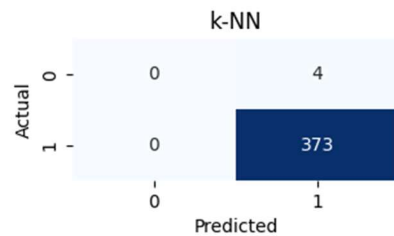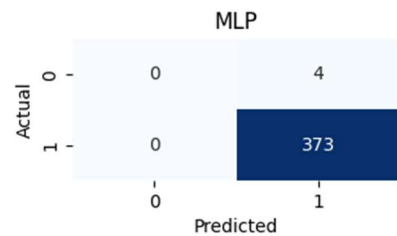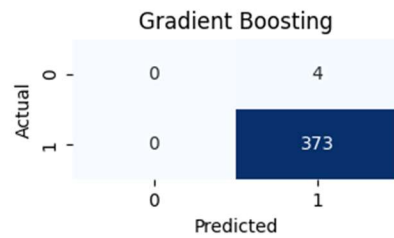
Vrishab Prasanth Davey <300438343>

Best parameters	
Confusion matrix	
Precision,recall,F1 score	

1. Chocolate(Imbalanced Dataset)

i. Original Dataset

ROC Curves for Different Models (None)

	Decision Tree	
Actual 0	0	4
Actual 1	0	373
	Predicted 0	Predicted 1

	Random Forest	
Actual 0	0	43
Actual 1	0	373
	Predicted 0	Predicted 1

	SVM	
Actual 0	0	4
Actual 1	0	373
	Predicted 0	Predicted 1

	Gradient Boosting	
Actual 0	0	4
Actual 1	0	373
	Predicted 0	Predicted 1

	MLP	
Actual 0	0	4
Actual 1	0	373
	Predicted 0	Predicted 1

	k-NN	
Actual 0	0	4
Actual 1	0	373
	Predicted 0	Predicted 1
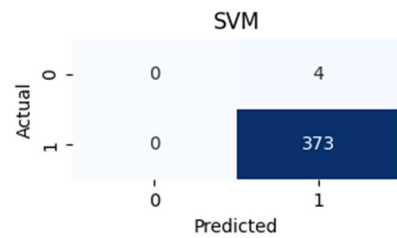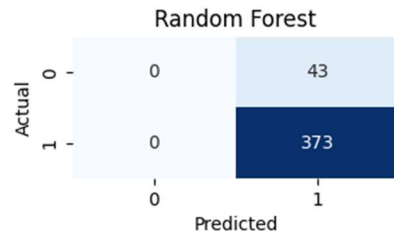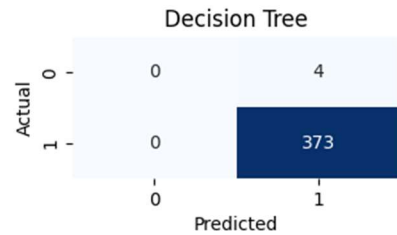
Best parameters for Decision Tree: {'min_samples_split': 5, 'min_samples_leaf': 2, 'max_depth': 3, 'criterion': 'entropy'}

Confusion Matrix for Decision Tree:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

Best parameters for Random Forest: {'n_estimators': 50, 'min_samples_split': 5, 'min_samples_leaf': 2, 'max_depth': None, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

Best parameters for SVM: {'kernel': 'linear', 'gamma': 'scale', 'degree': 3, 'C': 100}

Confusion Matrix for SVM:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

Best parameters for Gradient Boosting: {'subsample': 0.8, 'n_estimators': 50, 'min_samples_split': 5, 'max_depth': 3, 'learning_rate': 0.01}

Confusion Matrix for Gradient Boosting:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

Best parameters for MLP: {'solver': 'sgd', 'learning_rate': 'constant', 'hidden_layer_sizes': (100, 100), 'alpha': 0.001, 'activation': 'relu'}

Confusion Matrix for MLP:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

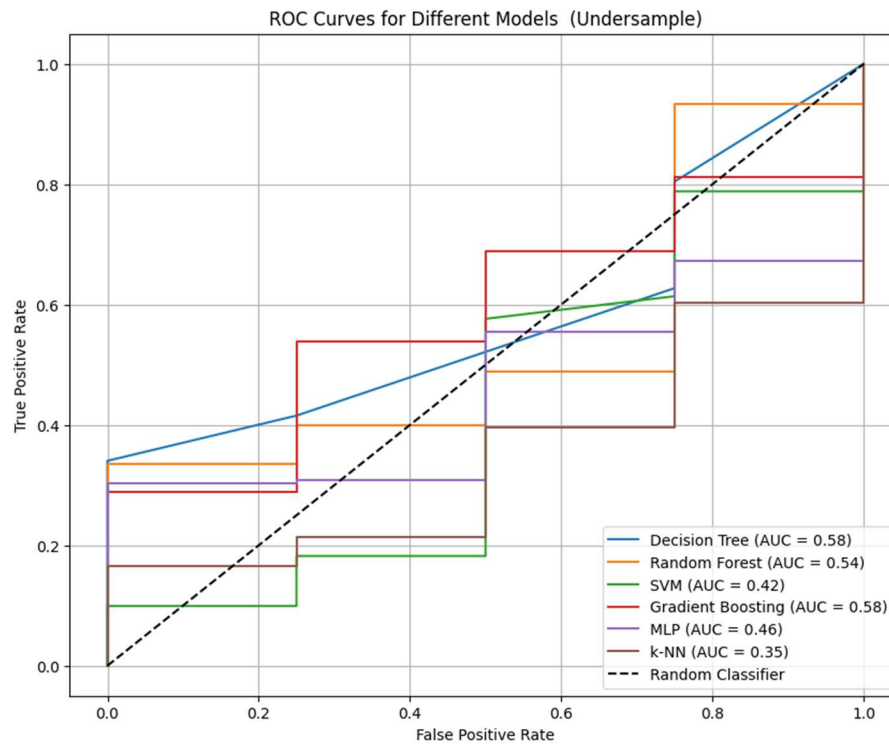Best parameters for k-NN: {'weights': 'uniform', 'p': 1, 'n_neighbors': 5}

Confusion Matrix for k-NN:

[[0   4]

 [0 373]]

Precision: 0.99, Recall: 1.00, F1-score: 0.99

ii.   Undersampled

ROC Curves for Different Models (Undersample)

Decision Tree (AUC = 0.58)
Random Forest (AUC = 0.54)
SVM (AUC = 0.42)
Gradient Boosting (AUC = 0.58)
MLP (AUC = 0.46)
k-NN (AUC = 0.35)
Random Classifier

Decision Tree		
3	1	
218	155	

Random Forest		
1	3	
176	197	

SVM		
2	2	
166	207	

Gradient Boosting		
2	2	
143	230	

MLP		
2	2	
185	188	

k-NN		
1	3	
160	213	

Best parameters for Decision Tree: {'min_samples_split': 5, 'min_samples_leaf': 1, 'max_depth': 5, 'criterion': 'entropy'}

Confusion Matrix for Decision Tree:

[[3   1]

 [218 155]]

Precision: 0.99, Recall: 0.42, F1-score: 0.59

Best parameters for Random Forest: {'n_estimators': 200, 'min_samples_split': 2, 'min_samples_leaf': 4, 'max_depth': 10, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[1   3]

 [176 197]]

Precision: 0.98, Recall: 0.53, F1-score: 0.69

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'auto', 'degree': 2, 'C': 100}

Confusion Matrix for SVM:

[[2   2]

 [166 207]]

Precision: 0.99, Recall: 0.55, F1-score: 0.71

Best parameters for Gradient Boosting: {'subsample': 1.0, 'n_estimators': 200, 'min_samples_split': 2, 'max_depth': 5, 'learning_rate': 0.1}

Confusion Matrix for Gradient Boosting:

[[2   2]

 [143 230]]

Precision: 0.99, Recall: 0.62, F1-score: 0.76

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'constant', 'hidden_layer_sizes': (50,), 'alpha': 0.0001, 'activation': 'tanh'}

Confusion Matrix for MLP:

[[2   2]

 [185 188]]

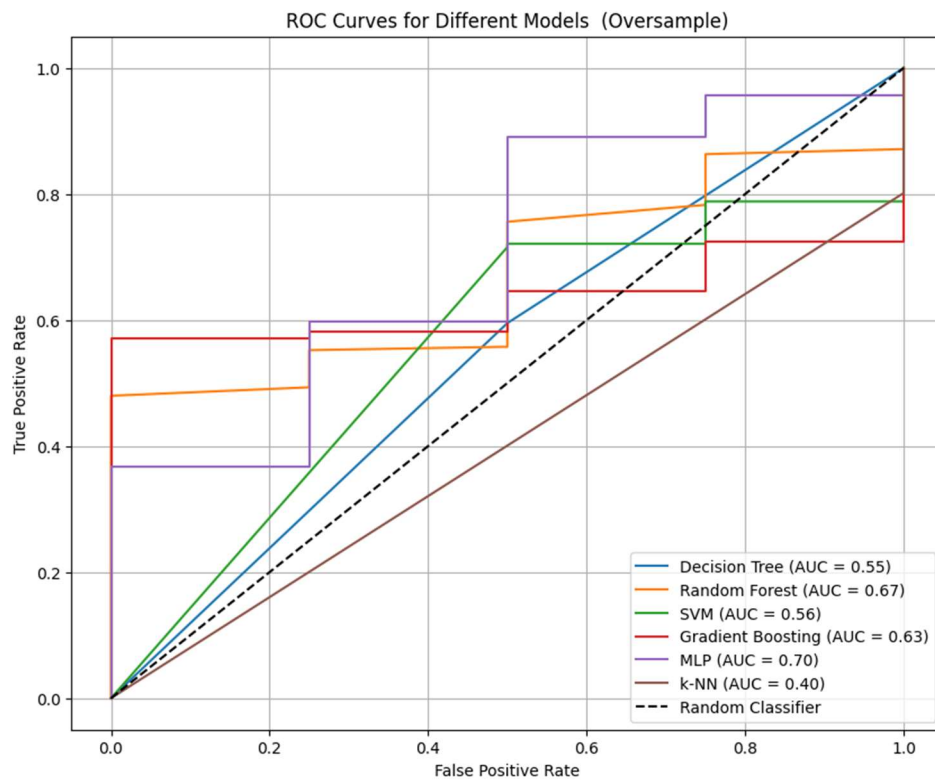Precision: 0.99, Recall: 0.50, F1-score: 0.67

Best parameters for k-NN: {'weights': 'distance', 'p': 2, 'n_neighbors': 3}
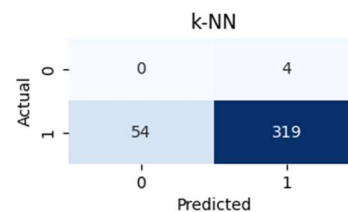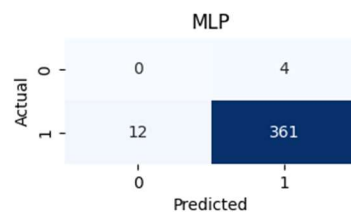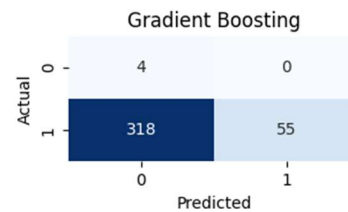
Confusion Matrix for k-NN:

[[1   3]

 [160 213]]

Precision: 0.99, Recall: 0.57, F1-score: 0.72

iii.   Oversampled
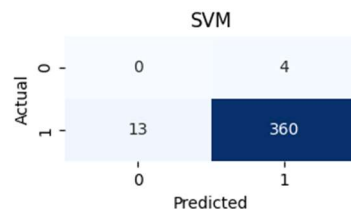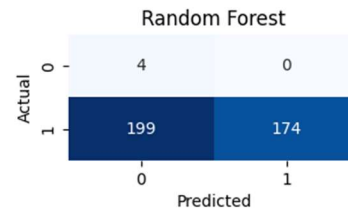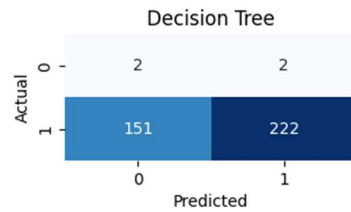
ROC Curves for Different Models (Oversample)

Decision Tree

Actual 0	2	2
Actual 1	151	222
	0	1
	Predicted	

Random Forest

Actual 0	4	0
Actual 1	199	174
	0	1
	Predicted	

SVM

Actual 0	0	4
Actual 1	13	360
	0	1
	Predicted	

Gradient Boosting

Actual 0	4	0
Actual 1	318	55
	0	1
	Predicted	

MLP

Actual 0	0	4
Actual 1	12	361
	0	1
	Predicted	

k-NN

Actual 0	0	4
Actual 1	54	319
	0	1
	Predicted	

Best parameters for Decision Tree: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'criterion': 'entropy'}

Confusion Matrix for Decision Tree:

[[2   2]

 [151 222]]

Precision: 0.99, Recall: 0.60, F1-score: 0.74

Best parameters for Random Forest: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[4   0]

 [199 174]]

Precision: 1.00, Recall: 0.47, F1-score: 0.64

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'scale', 'degree': 2, 'C': 100}

Confusion Matrix for SVM:

[[0   4]

 [13 360]]

Precision: 0.99, Recall: 0.97, F1-score: 0.98

Best parameters for Gradient Boosting: {'subsample': 0.8, 'n_estimators': 100, 'min_samples_split': 2, 'max_depth': 10, 'learning_rate': 0.2}

Confusion Matrix for Gradient Boosting:

[[4   0]

 [318 55]]

Precision: 1.00, Recall: 0.15, F1-score: 0.26

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'constant', 'hidden_layer_sizes': (50,), 'alpha': 0.001, 'activation': 'relu'}

Confusion Matrix for MLP:

[[0   4]

 [12 361]]

Precision: 0.99, Recall: 0.97, F1-score: 0.98

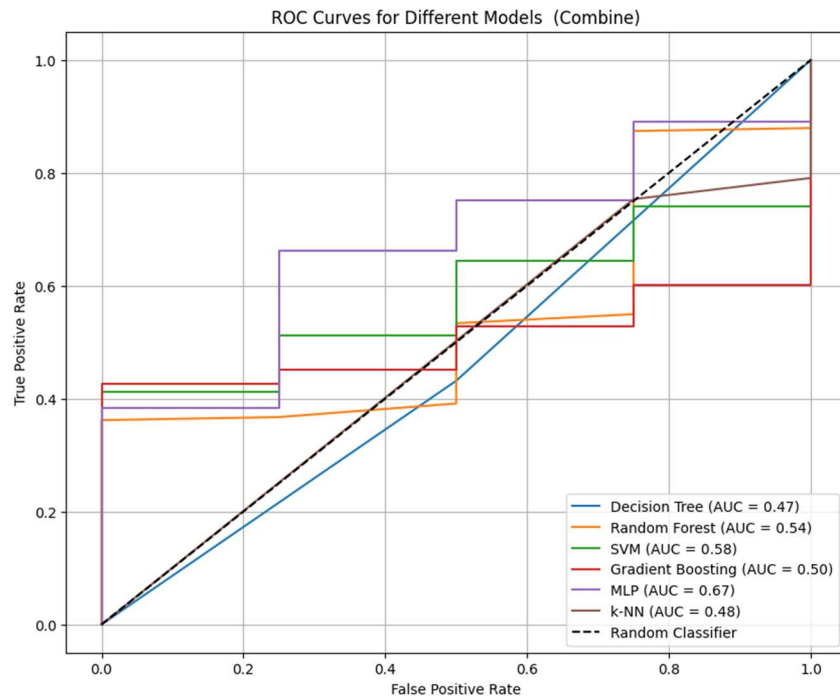Best parameters for k-NN: {'weights': 'uniform', 'p': 1, 'n_neighbors': 3}

Confusion Matrix for k-NN:

[[0   4]

 [54 319]]

Precision: 0.99, Recall: 0.86, F1-score: 0.92

iv.   Combined

ROC Curves for Different Models (Combine)

- Decision Tree (AUC = 0.47)
- Random Forest (AUC = 0.54)
- SVM (AUC = 0.58)
- Gradient Boosting (AUC = 0.50)
- MLP (AUC = 0.67)
- k-NN (AUC = 0.48)
- Random Classifier

Decision Tree

	Predicted 0	Predicted 1
Actual 0	2	2
Actual 1	212	161

Random Forest

	Predicted 0	Predicted 1
Actual 0	2	2
Actual 1	194	179

SVM

	Predicted 0	Predicted 1
Actual 0	0	4
Actual 1	32	341

Gradient Boosting

	Predicted 0	Predicted 1
Actual 0	4	0
Actual 1	340	43

MLP

	Predicted 0	Predicted 1
Actual 0	0	4
Actual 1	31	342

k-NN

	Predicted 0	Predicted 1
Actual 0	0	4
Actual 1	78	295

Best parameters for Decision Tree: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'criterion': 'entropy'}

Confusion Matrix for Decision Tree:

[[2   2]

 [212 161]]

Precision: 0.99, Recall: 0.43, F1-score: 0.60

Best parameters for Random Forest: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[2   2]

 [194 179]]

Precision: 0.99, Recall: 0.48, F1-score: 0.65

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'auto', 'degree': 4, 'C': 10}

Confusion Matrix for SVM:

[[0   4]

 [32 341]]

Precision: 0.99, Recall: 0.91, F1-score: 0.95

Best parameters for Gradient Boosting: {'subsample': 1.0, 'n_estimators': 200, 'min_samples_split': 10, 'max_depth': 5, 'learning_rate': 0.5}

Confusion Matrix for Gradient Boosting:

[[4   0]

 [330 43]]

Precision: 1.00, Recall: 0.12, F1-score: 0.21

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'adaptive', 'hidden_layer_sizes': (100, 100), 'alpha': 0.01, 'activation': 'relu'}

Confusion Matrix for MLP:

[[0   4]

 [31 342]]

Precision: 0.99, Recall: 0.92, F1-score: 0.95

Best parameters for k-NN: {'weights': 'uniform', 'p': 2, 'n_neighbors': 3}
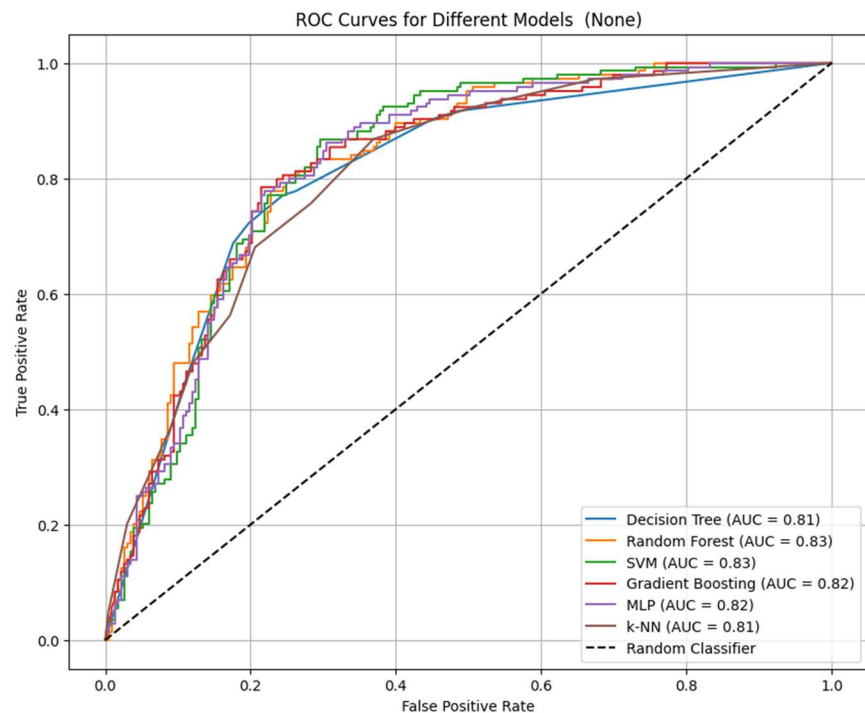
Confusion Matrix for k-NN:

[[0   4]

 [78 295]]

Precision: 0.99, Recall: 0.79, F1-score: 0.88

2. MAGIC MUSHROOMS DATASET(Balanced Dataset)
i.   Original Dataset

ROC Curves for Different Models (None)

Decision Tree — Random Forest — SVM — Gradient Boosting — MLP — k-NN confusion matrices

Best parameters for Decision Tree: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 3, 'criterion': 'gini'}

Confusion Matrix for Decision Tree:

[[187 46]

 [40 104]]

Precision: 0.69, Recall: 0.72, F1-score: 0.71

Best parameters for Random Forest: {'n_estimators': 50, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_depth': 5, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[192 41]

 [53 91]]

Precision: 0.69, Recall: 0.63, F1-score: 0.66

Best parameters for SVM: {'kernel': 'linear', 'gamma': 'scale', 'degree': 2, 'C': 0.1}

Confusion Matrix for SVM:

[[191 42]

 [46 98]]

Precision: 0.70, Recall: 0.68, F1-score: 0.69

Best parameters for Gradient Boosting: {'subsample': 0.9, 'n_estimators': 200, 'min_samples_split': 10, 'max_depth': 3, 'learning_rate': 0.01}

Confusion Matrix for Gradient Boosting:

[[194 39]

 [52 92]]

Precision: 0.70, Recall: 0.64, F1-score: 0.67

Best parameters for MLP: {'solver': 'sgd', 'learning_rate': 'constant', 'hidden_layer_sizes': (100,), 'alpha': 0.001, 'activation': 'relu'}

Confusion Matrix for MLP:

[[194 39]

 [51 93]]

Precision: 0.70, Recall: 0.65, F1-score: 0.67

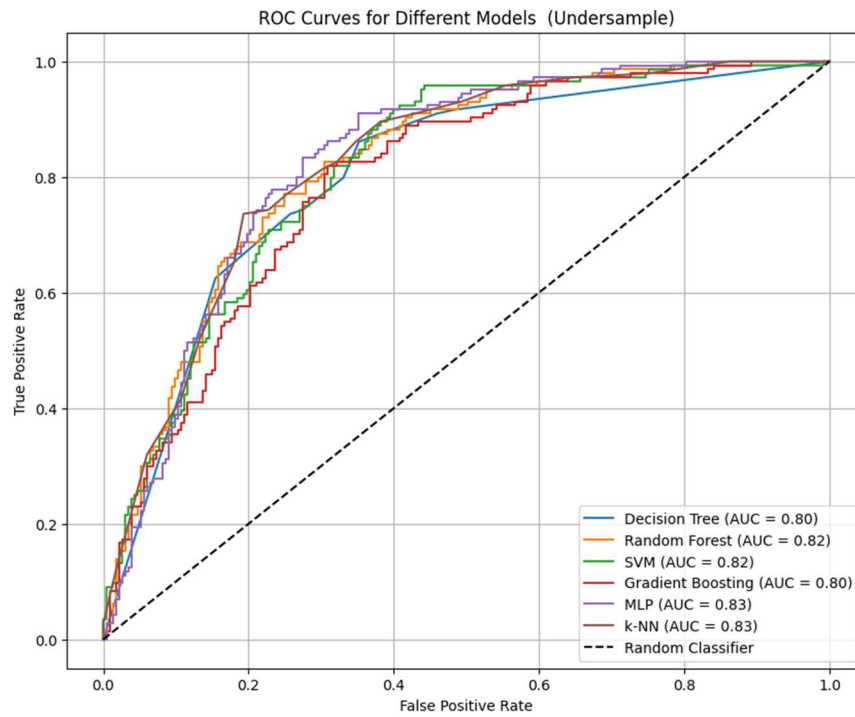Best parameters for k-NN: {'weights': 'uniform', 'p': 1, 'n_neighbors': 10}

Confusion Matrix for k-NN:

[[193 40]

 [63 81]]

Precision: 0.67, Recall: 0.56, F1-score: 0.61

ii.   Undersampled

ROC Curves for Different Models (Undersample)

Legend:
- Decision Tree (AUC = 0.80)
- Random Forest (AUC = 0.82)
- SVM (AUC = 0.82)
- Gradient Boosting (AUC = 0.80)
- MLP (AUC = 0.83)
- k-NN (AUC = 0.83)
- Random Classifier

Decision Tree

	Predicted 0	Predicted 1
Actual 0	156	77
Actual 1	29	115

Random Forest

	Predicted 0	Predicted 1
Actual 0	166	67
Actual 1	30	114

SVM

	Predicted 0	Predicted 1
Actual 0	160	73
Actual 1	31	113

Gradient Boosting

	Predicted 0	Predicted 1
Actual 0	166	67
Actual 1	34	110

MLP

	Predicted 0	Predicted 1
Actual 0	164	69
Actual 1	22	122

k-NN

	Predicted 0	Predicted 1
Actual 0	163	70
Actual 1	27	117

Best parameters for Decision Tree: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 3, 'criterion': 'gini'}

Confusion Matrix for Decision Tree:

[[156 77]

 [29 115]]

Precision: 0.60, Recall: 0.80, F1-score: 0.68

Best parameters for Random Forest: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 4, 'max_depth': 10, 'bootstrap': True}

Confusion Matrix for Random Forest:

[[166 67]

 [30 114]]

Precision: 0.63, Recall: 0.79, F1-score: 0.70

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'scale', 'degree': 3, 'C': 1}

Confusion Matrix for SVM:

[[160 73]

 [31 113]]

Precision: 0.61, Recall: 0.78, F1-score: 0.68

Best parameters for Gradient Boosting: {'subsample': 0.8, 'n_estimators': 50, 'min_samples_split': 2, 'max_depth': 5, 'learning_rate': 0.1}

Confusion Matrix for Gradient Boosting:

[[166 67]

 [34 110]]

Precision: 0.62, Recall: 0.76, F1-score: 0.69

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'adaptive', 'hidden_layer_sizes': (50, 50), 'alpha': 0.0001, 'activation': 'tanh'}

Precision: 0.64, Recall: 0.85, F1-score: 0.73

Precision: 0.63, Recall: 0.81, F1-score: 0.71

iii. Oversampled

ROC Curves for Different Models (Oversample)

Decision Tree / Random Forest / SVM / Gradient Boosting / MLP / k-NN confusion matrices

Confusion Matrix for Decision Tree:

[[80 153]

 [8 136]]

Precision: 0.47, Recall: 0.94, F1-score: 0.63

Best parameters for Random Forest: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[132 101]

 [25 119]]

Precision: 0.54, Recall: 0.83, F1-score: 0.65

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'auto', 'degree': 2, 'C': 100}

Confusion Matrix for SVM:

[[170 63]

 [54 90]]

Precision: 0.59, Recall: 0.62, F1-score: 0.61

Best parameters for Gradient Boosting: {'subsample': 0.8, 'n_estimators': 100, 'min_samples_split': 2, 'max_depth': 10, 'learning_rate': 0.2

Confusion Matrix for Gradient Boosting:

[[112 121]

 [18 126]]

Precision: 0.51, Recall: 0.88, F1-score: 0.64

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'adaptive', 'hidden_layer_sizes': (50, 50), 'alpha': 0.0001, 'activation': 'tanh'}

Confusion Matrix for MLP:

[[166 67]

 [43 101]]

Precision: 0.60, Recall: 0.70, F1-score: 0.65

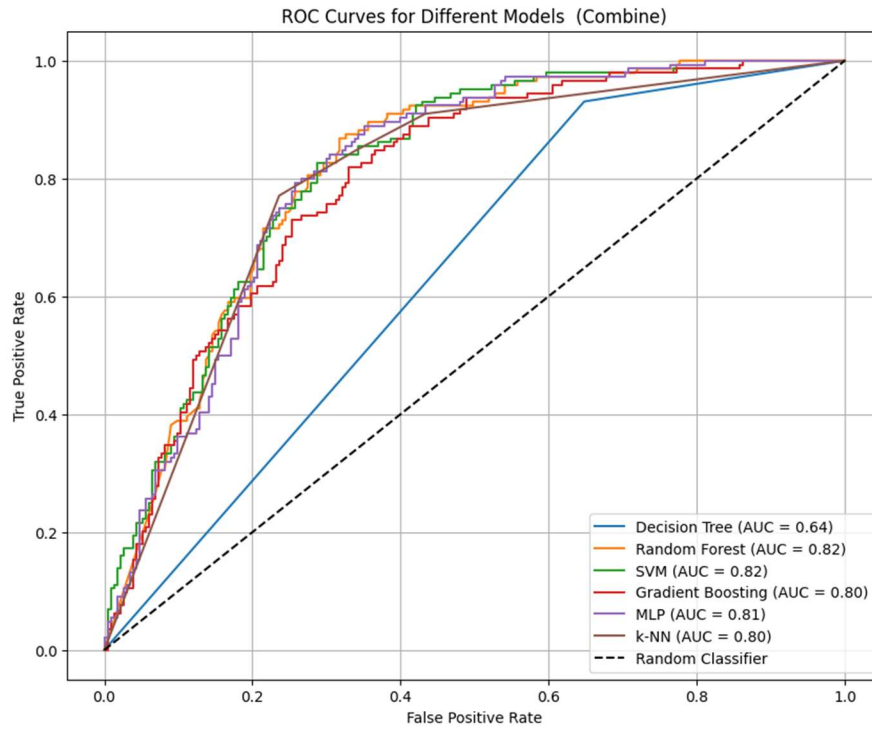Best parameters for k-NN: {'weights': 'distance', 'p': 1, 'n_neighbors': 5}

Confusion Matrix for k-NN:

[[151 82]

 [33 111]]

Precision: 0.58, Recall: 0.77, F1-score: 0.66

iv. Combined

ROC Curves for Different Models (Combine)

- Decision Tree (AUC = 0.64)
- Random Forest (AUC = 0.82)
- SVM (AUC = 0.82)
- Gradient Boosting (AUC = 0.80)
- MLP (AUC = 0.81)
- k-NN (AUC = 0.80)
- Random Classifier

Best parameters for Decision Tree: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'criterion': 'entropy'}

Confusion Matrix for Decision Tree:

[[82 151]

 [10 134]]

Precision: 0.47, Recall: 0.93, F1-score: 0.62

Best parameters for Random Forest: {'n_estimators': 50, 'min_samples_split': 5, 'min_samples_leaf': 2, 'max_depth': None, 'bootstrap': False}

Confusion Matrix for Random Forest:

[[135 98]

 [11 133]]

Precision: 0.58, Recall: 0.92, F1-score: 0.71

Best parameters for SVM: {'kernel': 'rbf', 'gamma': 'scale', 'degree': 3, 'C': 1}

Confusion Matrix for SVM:

[[153 80]

 [21 123]]

Precision: 0.61, Recall: 0.85, F1-score: 0.71

Best parameters for Gradient Boosting: {'subsample': 0.9, 'n_estimators': 200, 'min_samples_split': 2, 'max_depth': 5, 'learning_rate': 0.2}

Confusion Matrix for Gradient Boosting:

[[118 115]

 [9 135]]

Precision: 0.54, Recall: 0.94, F1-score: 0.69

Best parameters for MLP: {'solver': 'adam', 'learning_rate': 'constant', 'hidden_layer_sizes': (50,), 'alpha': 0.01, 'activation': 'relu'}

Confusion Matrix for MLP:

[[153 80]

 [18 126]]

Precision: 0.61, Recall: 0.88, F1-score: 0.72

Best parameters for k-NN: {'weights': 'uniform', 'p': 1, 'n_neighbors': 3}

Confusion Matrix for k-NN:

[[154 79]

 [22 122]]

Precision: 0.61, Recall: 0.85, F1-score: 0.71