

Table 1: Whole dataset

Table 1A: HoldOut Cross-validation or Train-Test Split

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	39	0.57	0.49	0.52	0.49	0.40
F1-score	39	0.57	0.49	0.52	0.47	0.38
Precision	39	0.57	0.49	0.51	0.47	0.44
Recall	39	0.58	0.49	0.52	0.49	0.40

Table 1B: Repeated kFold Cross Validation-- n_splits=10, n_repeats=5

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	39	0.58	0.50	0.52	0.42	0.40
F1-score	39	0.57	0.50	0.52	0.40	0.38
Precision	39	0.57	0.50	0.52	0.42	0.44
Recall	39	0.58	0.50	0.52	0.42	0.40

Table 2: Applying Filter Method--Removing 3 attributes with too much missingness

Table 2A: HoldOut Cross-validation or Train-Test Split

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	36	0.57	0.49	0.52	0.48	0.41
F1-score	36	0.56	0.49	0.52	0.47	0.39
Precision	36	0.56	0.49	0.51	0.47	0.39
Recall	36	0.57	0.49	0.52	0.49	0.41

Table 2B: Repeated kFold Cross Validation-- n_splits=10, n_repeats=5

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	36	0.57	0.69	0.52	0.45	0.42
F1-score	36	0.57	0.69	0.52	0.44	0.39
Precision	36	0.57	0.69	0.52	0.44	0.40
Recall	36	0.57	0.69	0.52	0.45	0.40

Table 3: Applying Filter Method - Information Gain [Selecting k=28 best, where MI \geq 0.01]

Table 3A: HoldOut Cross-validation or Train-Test Split

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	28	0.57	0.49	0.52	0.48	0.40
F1-score	28	0.57	0.49	0.51	0.47	0.40
Precision	28	0.57	0.49	0.51	0.47	0.44
Recall	28	0.57	0.49	0.52	0.48	0.40

Table 3B: Repeated kFold Cross Validation-- n_splits=10, n_repeats=5

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	28	0.57	0.49	0.52	0.42	0.39
F1-score	28	0.57	0.49	0.52	0.40	0.38
Precision	28	0.57	0.49	0.52	0.42	0.44
Recall	28	0.57	0.49	0.52	0.42	0.39

Table 4: Applying Wrapper Method – Forward Selection/Backward Elimination- Optimal 31 features

Table 4A: HoldOut Cross-validation or Train-Test Split

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	31	0.56	0.48	0.50	0.48	0.39
F1-score	31	0.56	0.48	0.50	0.47	0.37
Precision	31	0.55	0.48	0.50	0.47	0.43
Recall	31	0.56	0.48	0.51	0.49	0.39

Table 4B: Repeated kFold Cross Validation-- n_splits=10, n_repeats=5

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	31	0.56	0.48	0.51	0.42	0.39
F1-score	31	0.56	0.48	0.50	0.40	0.37
Precision	31	0.56	0.48	0.50	0.42	0.44
Recall	31	0.56	0.48	0.51	0.42	0.39

Table 5: Applying Embedded Method – Sequential Forward Feature Selection- Optimal 31 features

Table 5A: HoldOut Cross-validation or Train-Test Split

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	29	0.57	0.49	0.52	0.49	0.40
F1-score	29	0.57	0.49	0.51	0.47	0.38
Precision	29	0.57	0.49	0.51	0.47	0.44
Recall	29	0.57	0.49	0.52	0.49	0.40

Table 5B: Repeated kFold Cross Validation-- n_splits=10, n_repeats=5

Parameter	Number of Attributes	Random Forest	Decision tree	KNN	Multinomial Logistic Regression	Naïve Bayes
Accuracy	29	0.58	0.50	0.52	0.42	0.40
F1-score	29	0.57	0.50	0.52	0.40	0.38
Precision	29	0.57	0.50	0.52	0.42	0.44
Recall	29	0.58	0.50	0.52	0.42	0.40