Unsupervised Feature Selection for Efficient Exploration of High Dimensional Data

Evaluations Results

Datasets

No	Dataset Description	No. of Features	No. of Classes	Missing Values	Source
1	Automobile Dataset	25	2	Yes	UCI ML
2	Breast Cancer Dataset	30	2 (Skewed Classes)	No	UCI ML
3	AutoUniv Dataset	39	8	No	UCI ML
4	QSAR-Biodegradation DataSet	41	2	Yes	openml
5	Sonar Dataset	60	2	No	Kaggle
6	Emotions Dataset	78	2	No	openml
7	Robot-Failures Dataset	91	5	No	openml
8	Spectrometer	125	2 (Skewed Classes)	Yes	openml
9	Musk	168	2	No	Kaggle

Table 1.1: Selected Datasets for Evaluation

Automobile Dataset
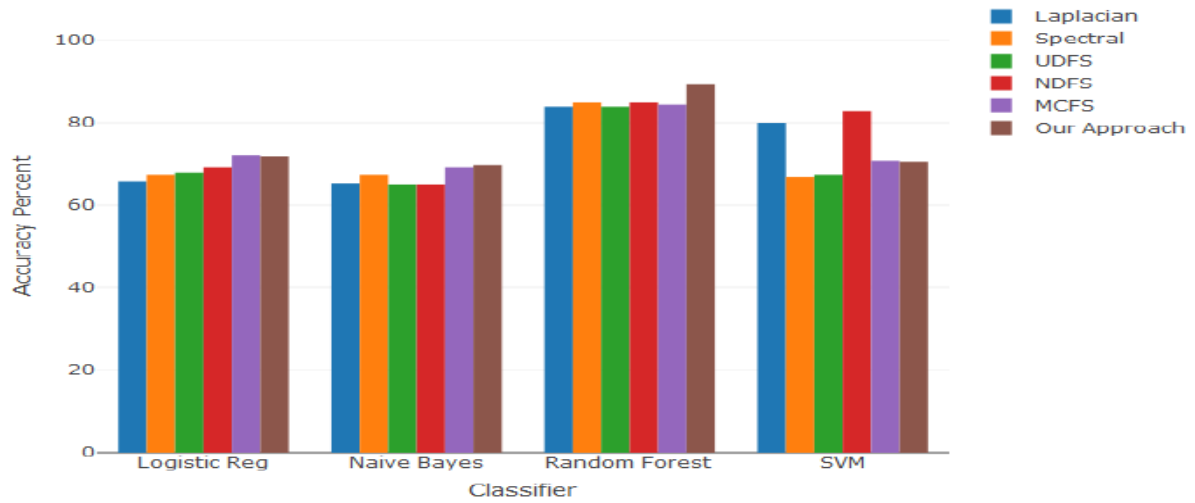**Source :** https://archive.ics.uci.edu/ml/datasets/Automobile
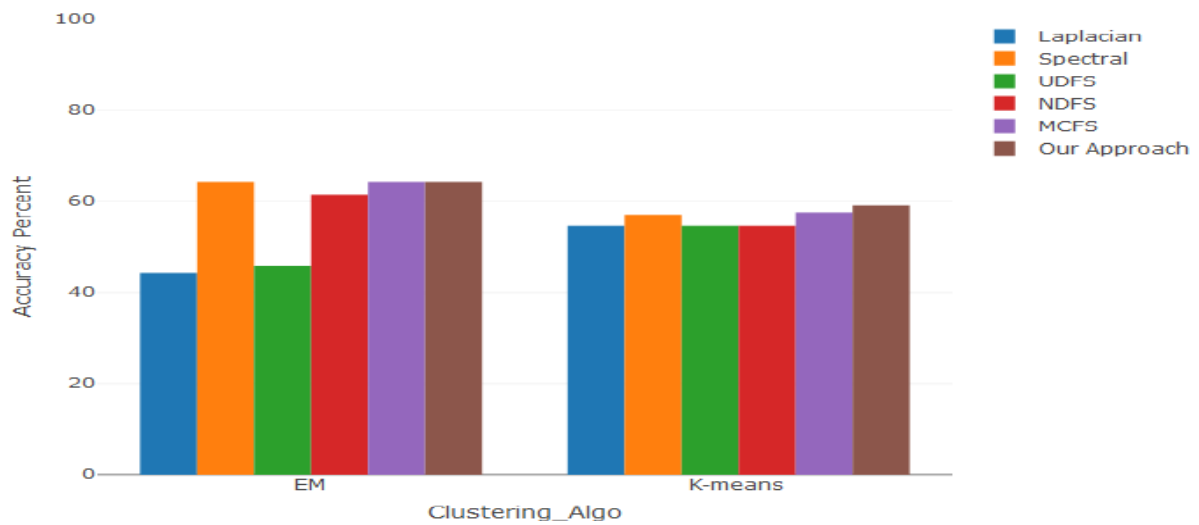

Figure 1.1: Classification Accuracy (Automobile Dataset)
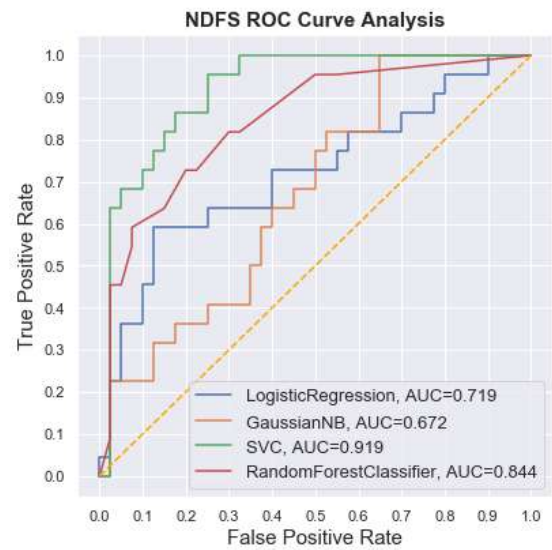

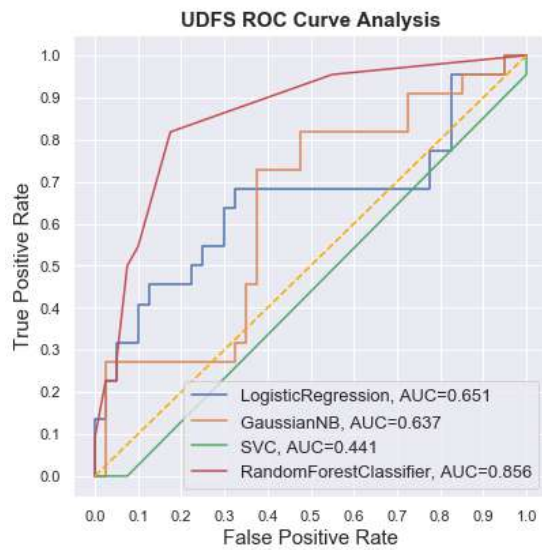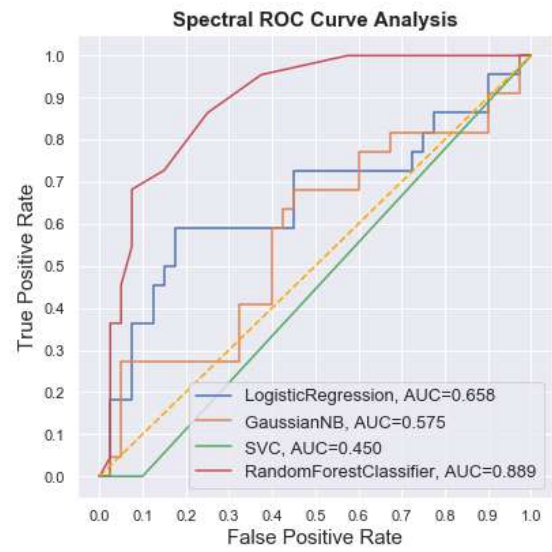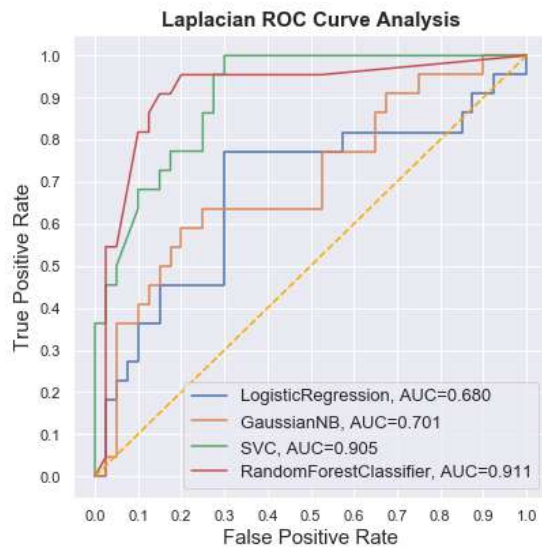Figure 1.2: Clustering Accuracy (Automobile Dataset)

Figure 1.3: ROC Curves (Automobile Dataset) 1
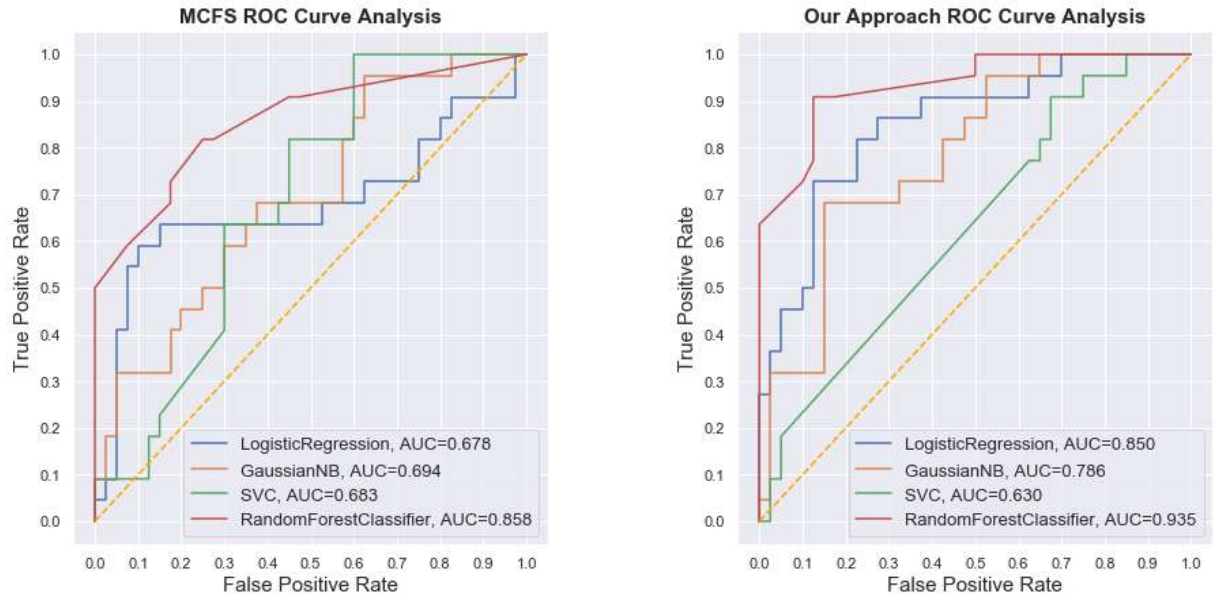
Figure 1.4: ROC Curves (Automobile Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.3462	0.00186	0.0150	0.216	0.00181	0.0162

Table 1.2: Representation Entropy (Automobile Dataset)

Breast Cancer Dataset
Source: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
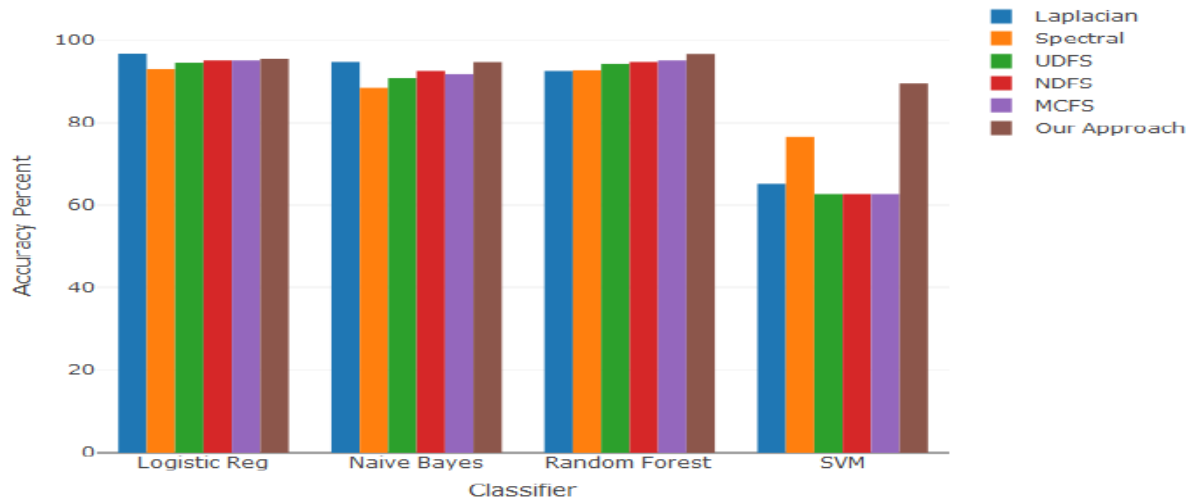
Figure 1.5: Classification Accuracy (Breast Cancer Dataset)
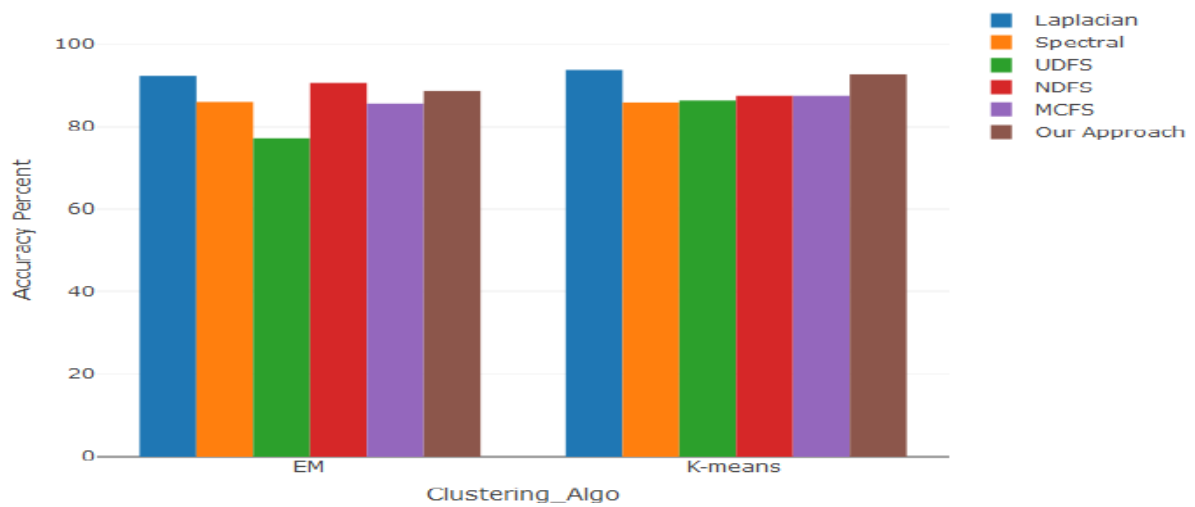
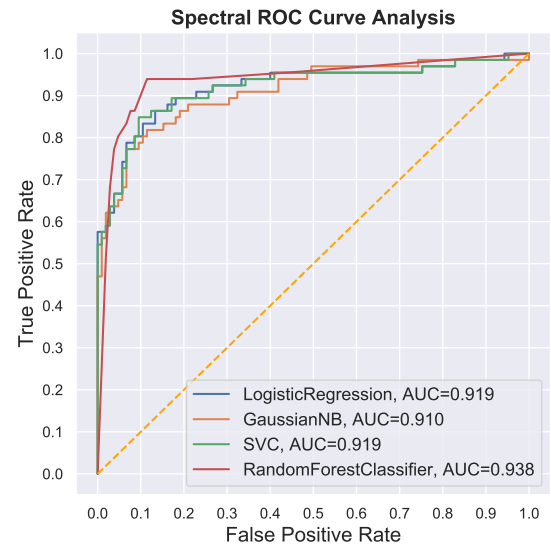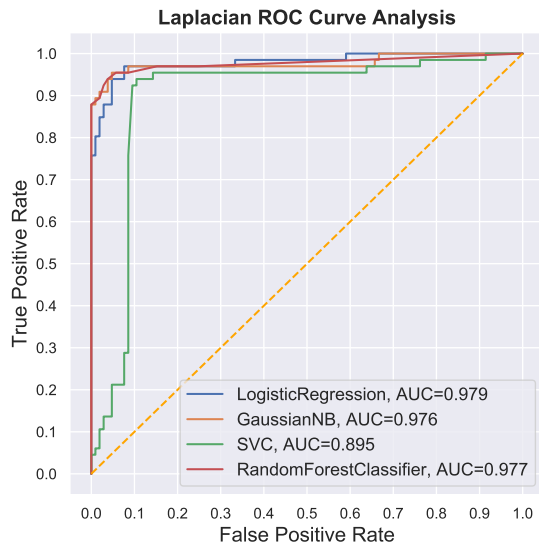Figure 1.6: Clustering Accuracy (Breast Cancer Dataset)
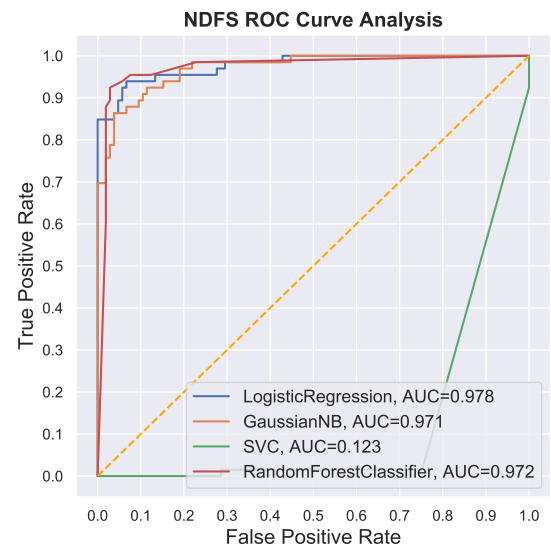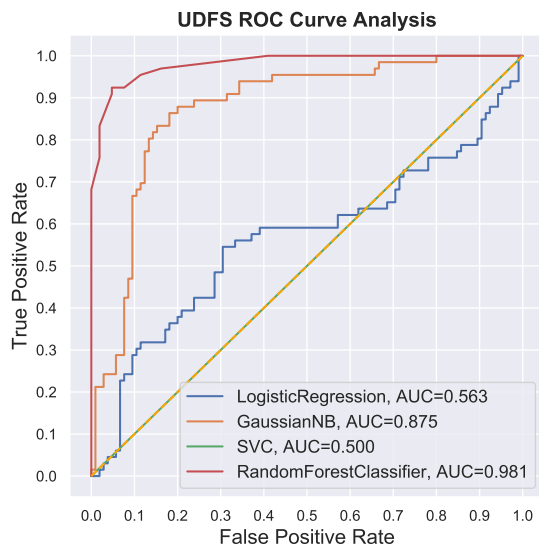
Figure 1.7: ROC Curves (Breast Cancer Dataset) 1
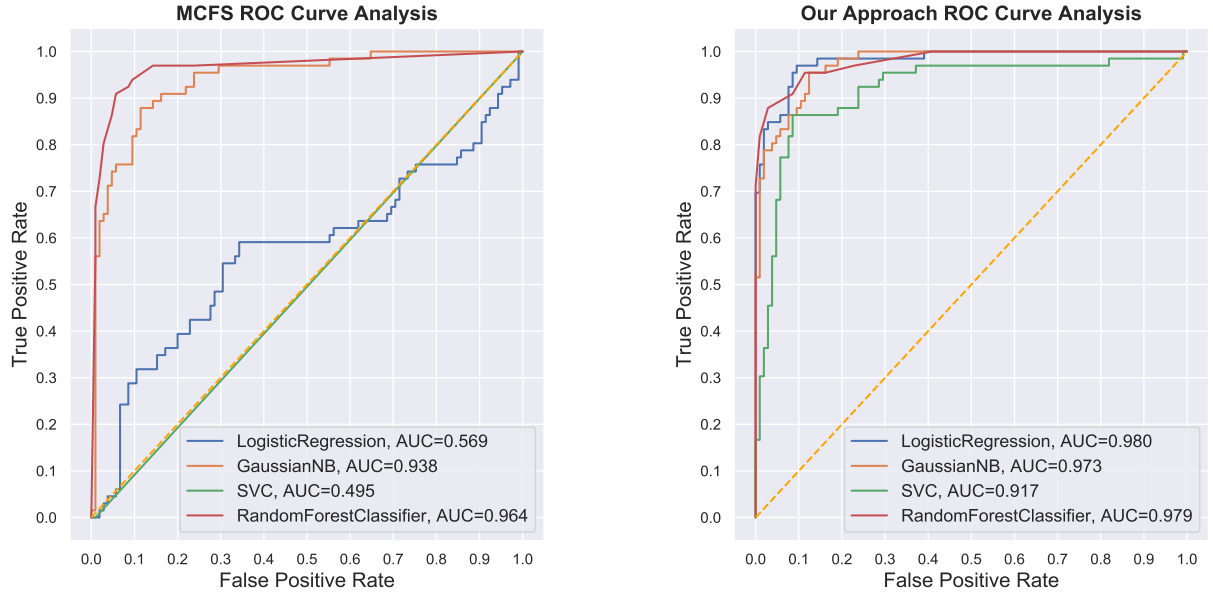
Figure 1.8: ROC Curves (Breast Cancer Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.0012	0.359	0.00091	0.04194	-	0.00920

Table 1.3: Representation Entropy (Breast Cancer Dataset)
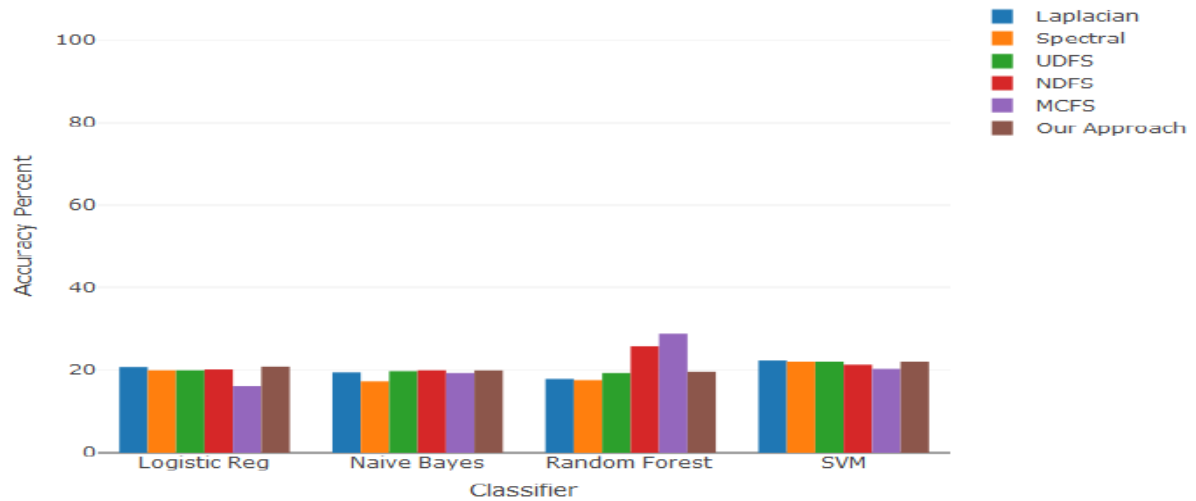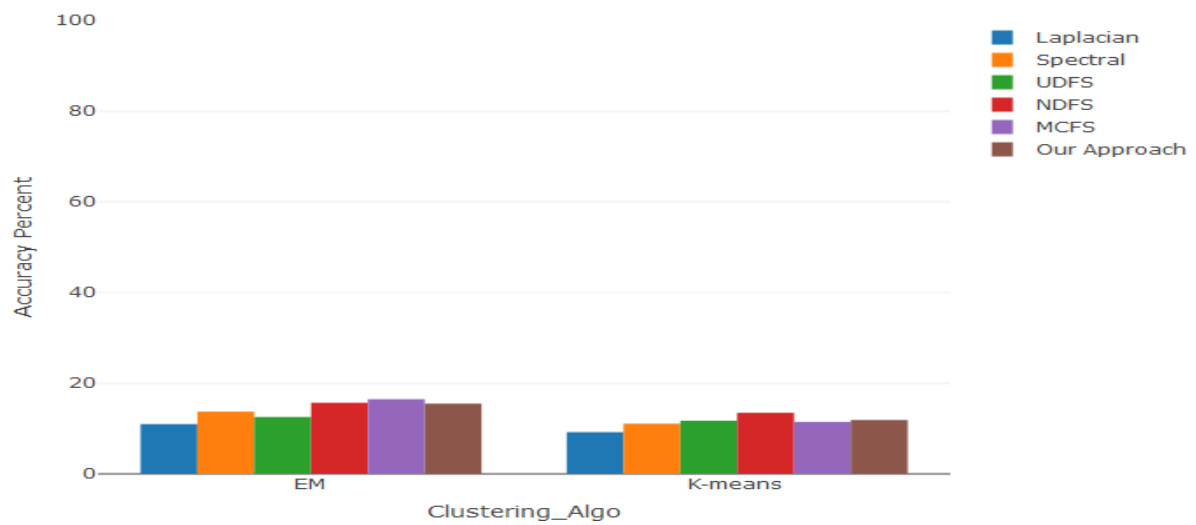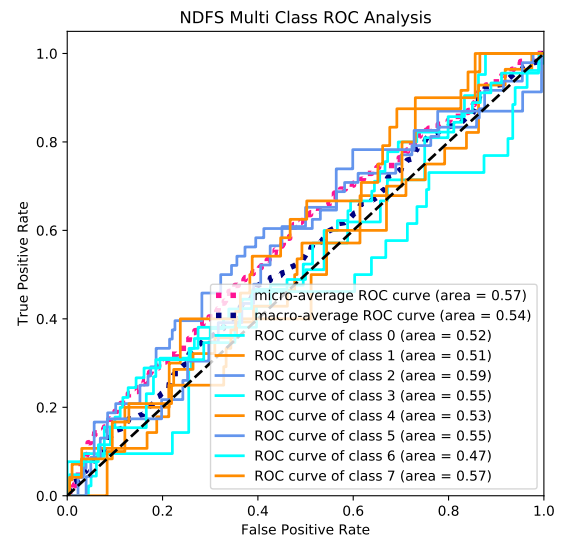
Auto Univ Dataset
Source: https://www.openml.org/d/1555

Figure 1.9: Classification Accuracy (Auto Univ Dataset)

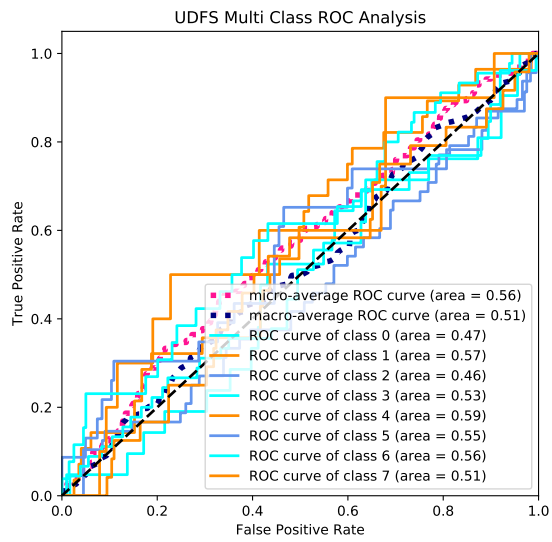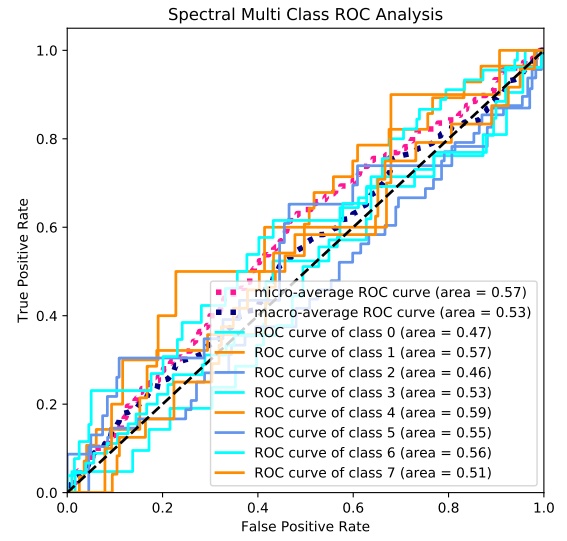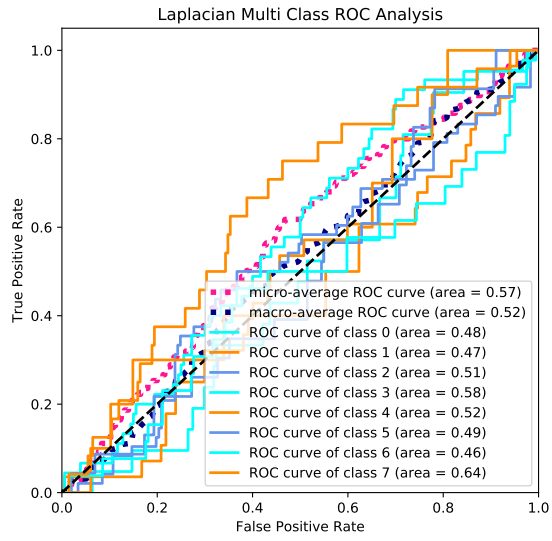Figure 1.10: Clustering Accuracy (Auto Univ Dataset)

Figure 1.11: ROC Curves (Auto Univ Dataset) 1
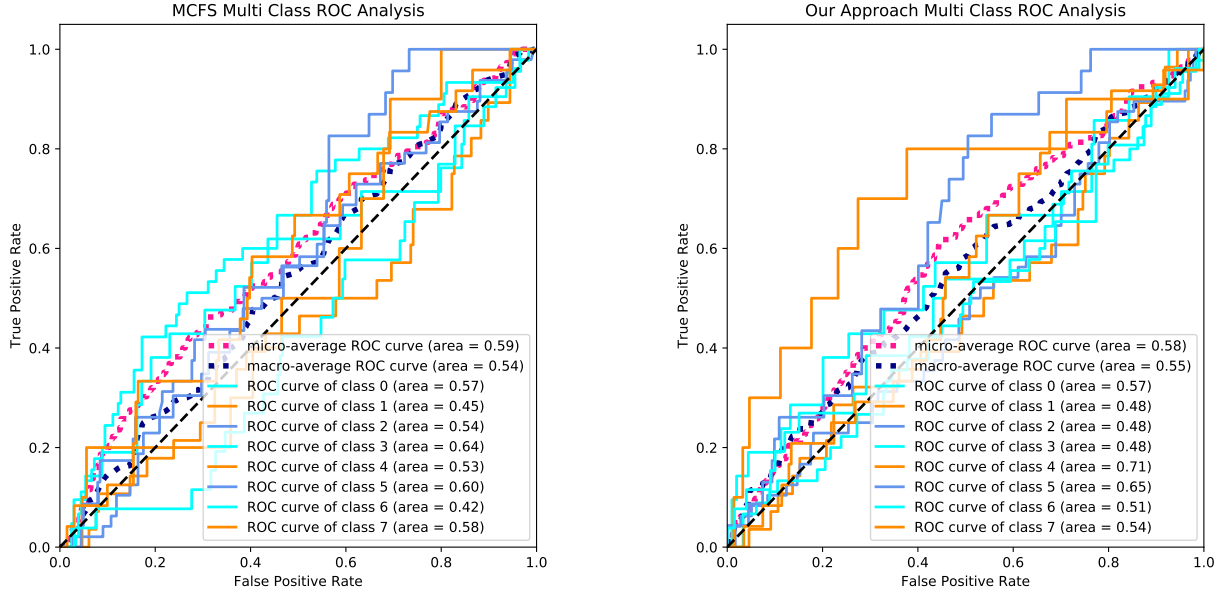
Figure 1.12: ROC Curves (Auto Univ Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.05386	0.003875	0.05718	0.90212	0.77475	0.00391

Table 1.4: Representation Entropy (Auto Univ Dataset)
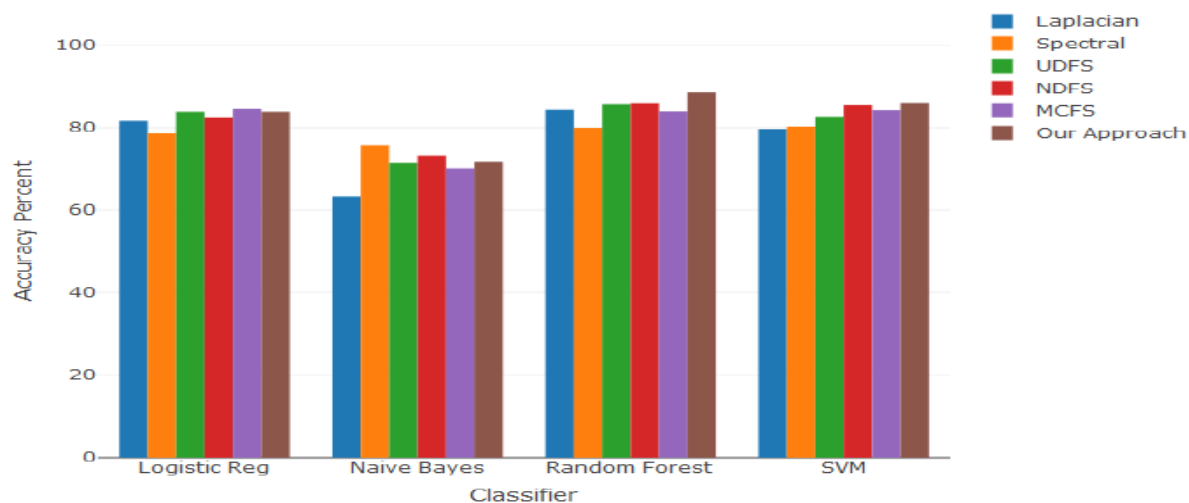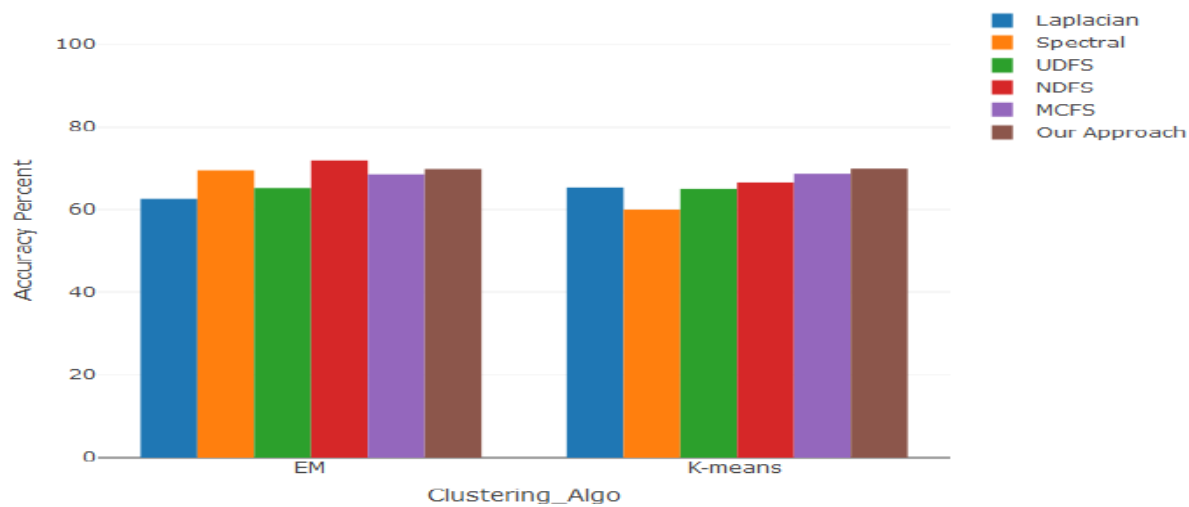
QSAR Biodegradation Dataset (No Missing Values)

Source: https://archive.ics.uci.edu/ml/datasets/QSAR+biodegradation

Figure 1.13: Classification Accuracy (QSAR Bio Dataset)
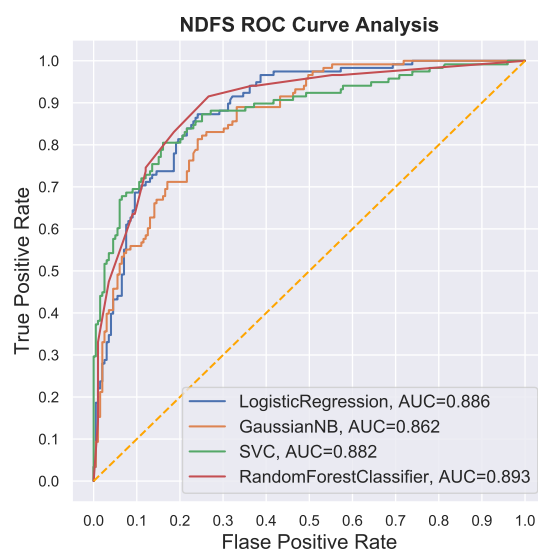

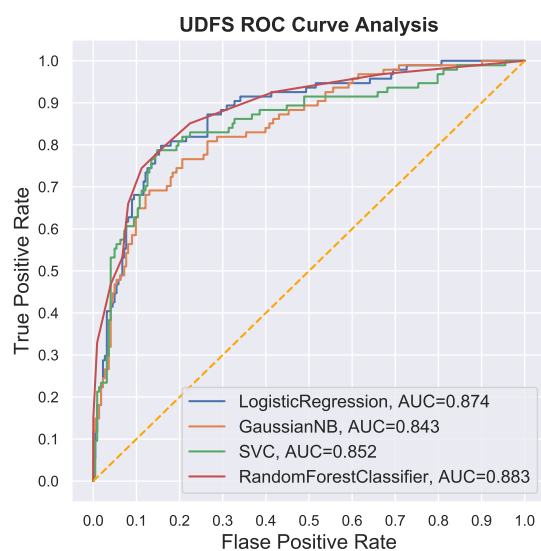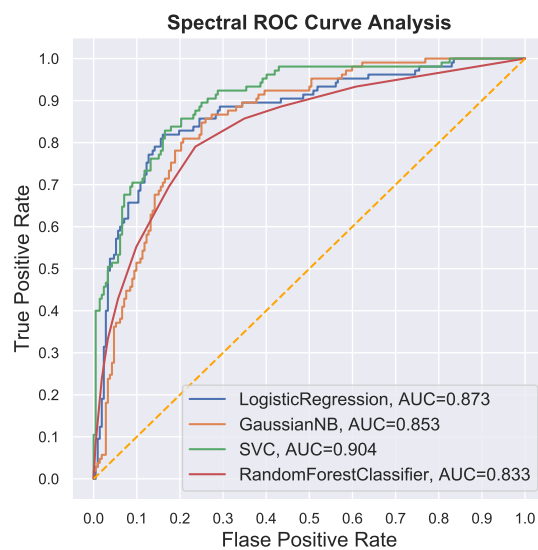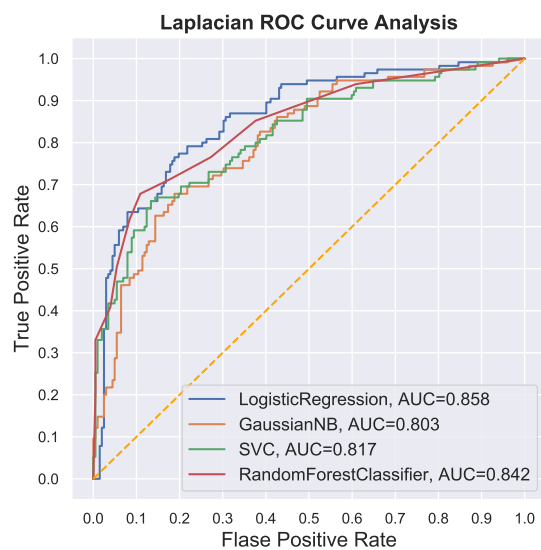Figure 1.14: Clustering Accuracy (QSAR Bio Dataset)
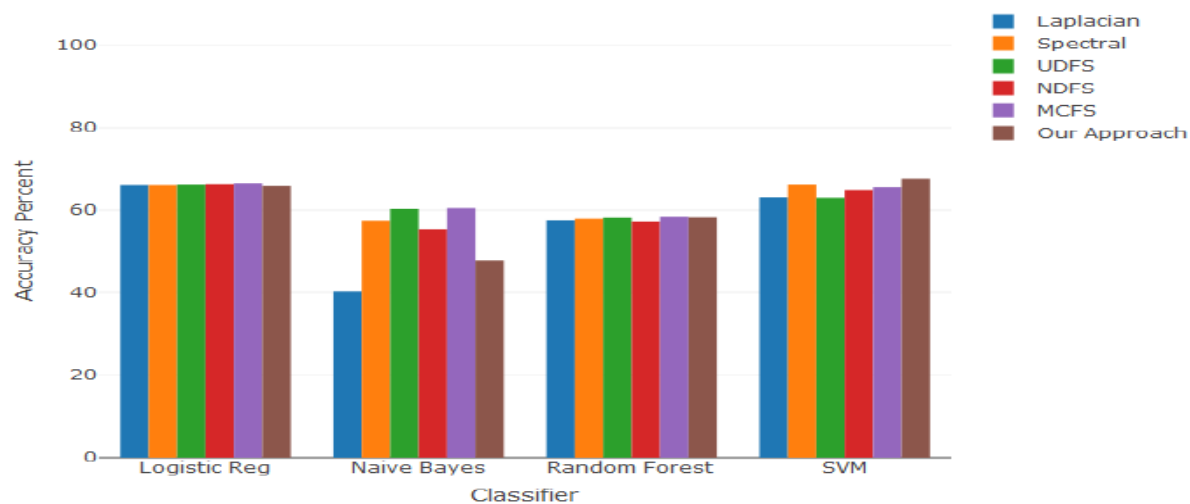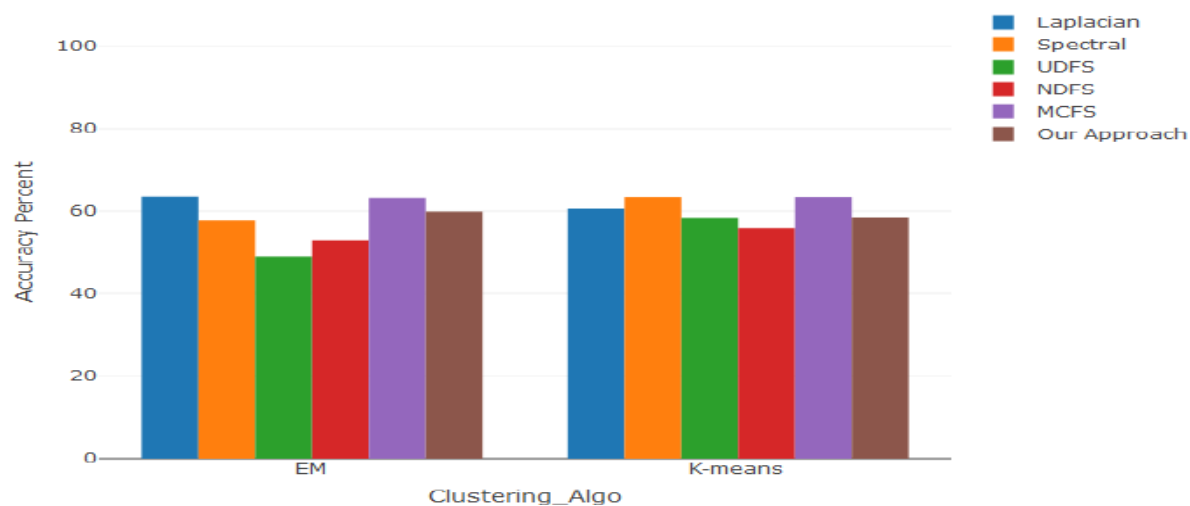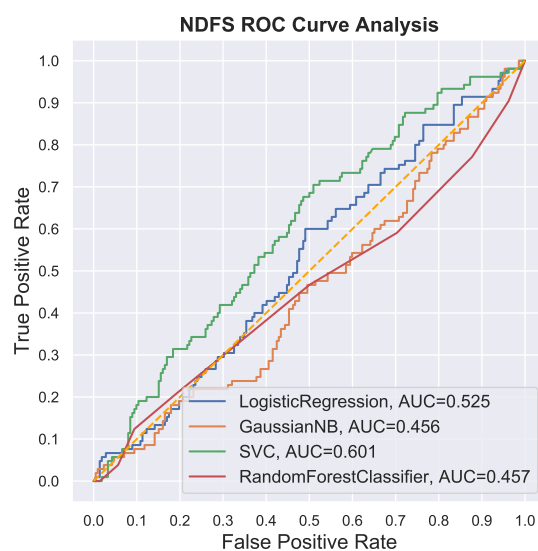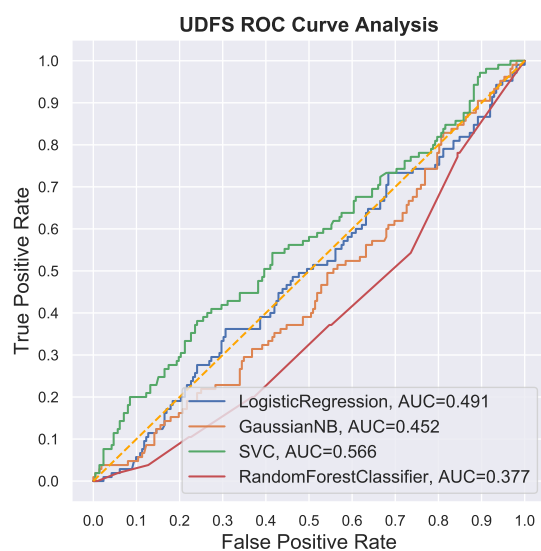
Figure 1.15: ROC Curves (QSAR Bio Dataset) 1
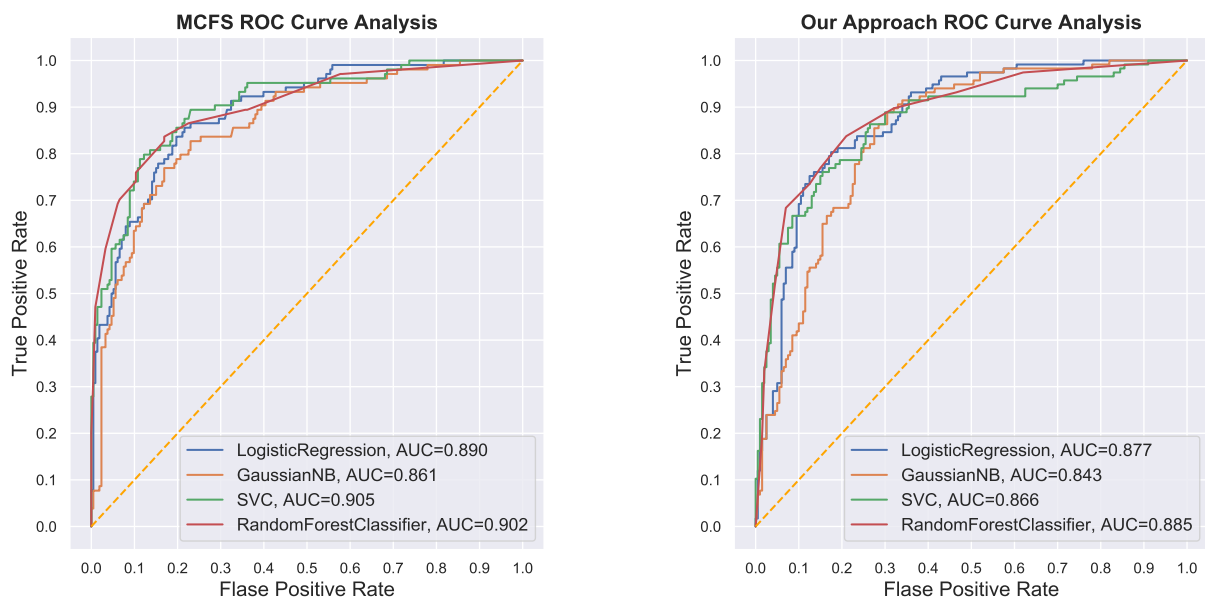
Figure 1.16: ROC Curves (QSAR Bio Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.47426	0.70142	0.5449	0.33878	0.6993	0.65399

Table 1.5: Representation Entropy (QSAR Bio Dataset)

QSAR Biodegradation Dataset (With Missing Values)
Source: https://archive.ics.uci.edu/ml/datasets/QSAR+biodegradation

Figure 1.17: Classification Accuracy (QSAR Bio Dataset)

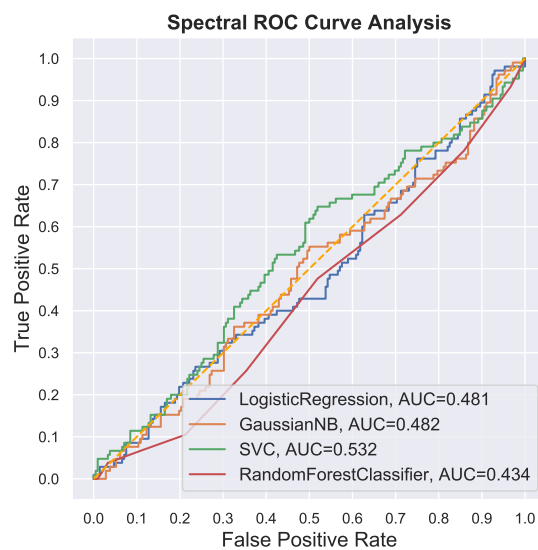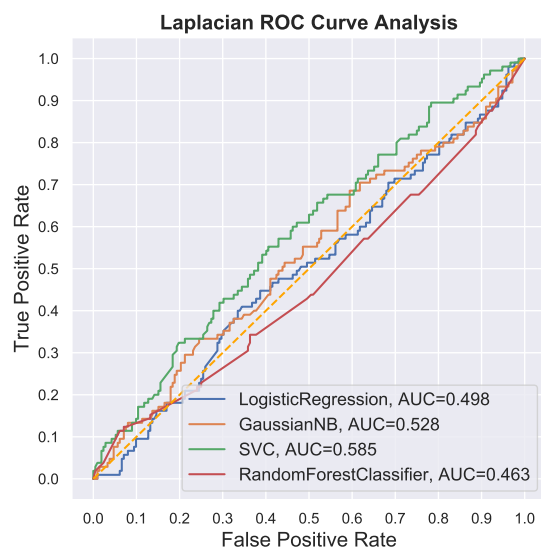Figure 1.18: Clustering Accuracy (QSAR Bio Dataset)
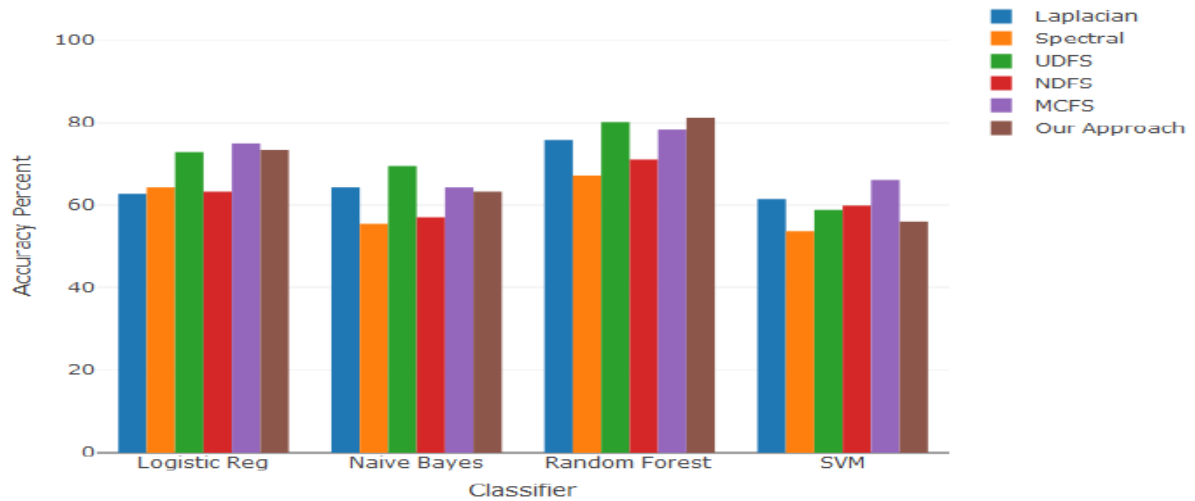
Figure 1.19: ROC Curves (QSAR Bio Dataset) 1
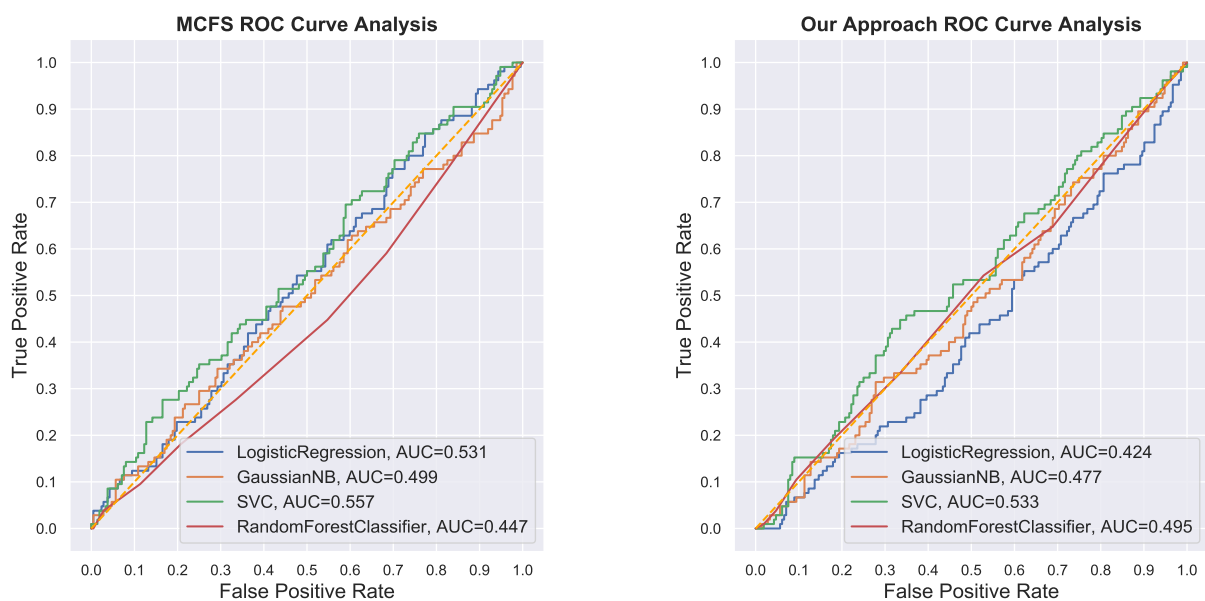
Figure 1.20: ROC Curves (QSAR Bio Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.57004	0.68327	0.54075	0.42013	0.69525	0.62873

Table 1.6: Representation Entropy (QSAR Bio Dataset)

16

Sonar Dataset

Source: http://archive.ics.uci.edu/ml/datasets/connectionist+bench+(sonar,+mines+vs.+rocks)
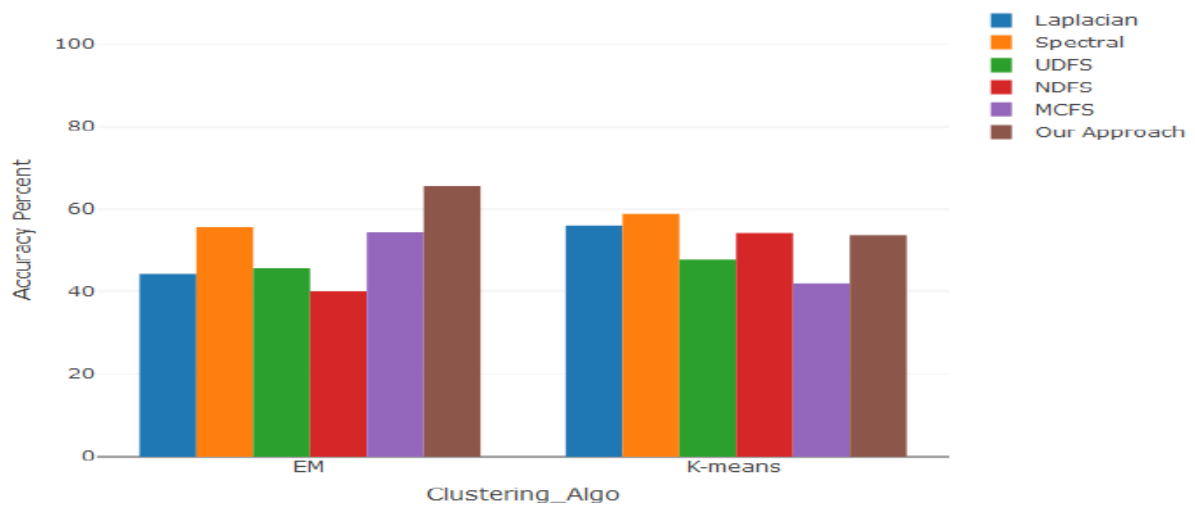

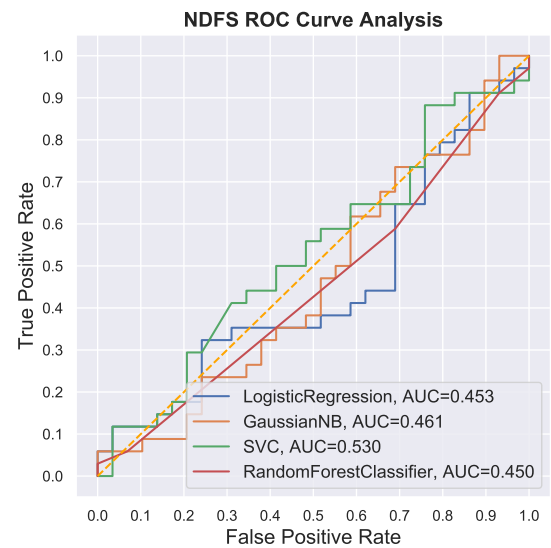Figure 1.21: Classification Accuracy (Sonar Dataset)

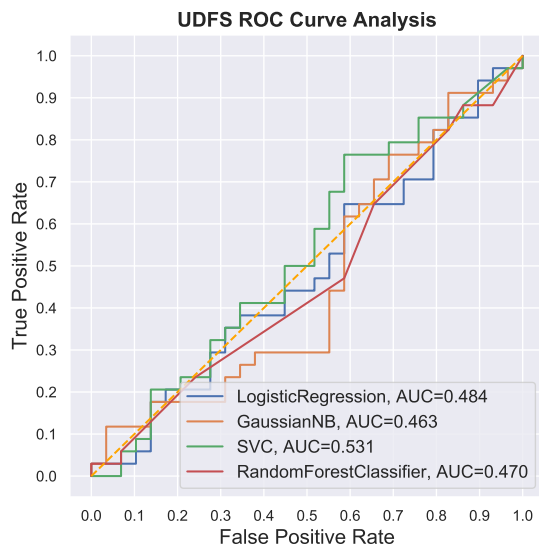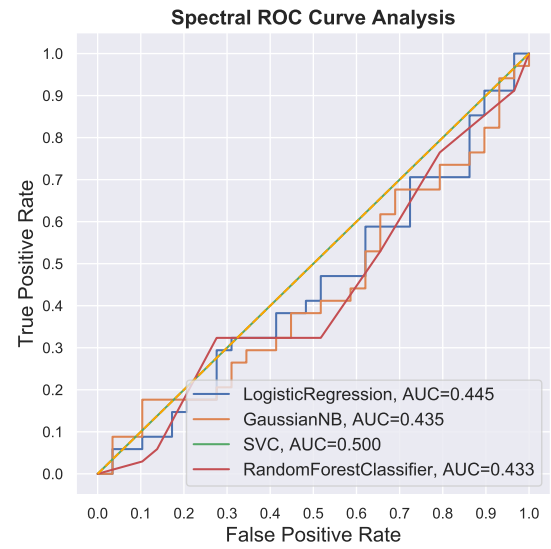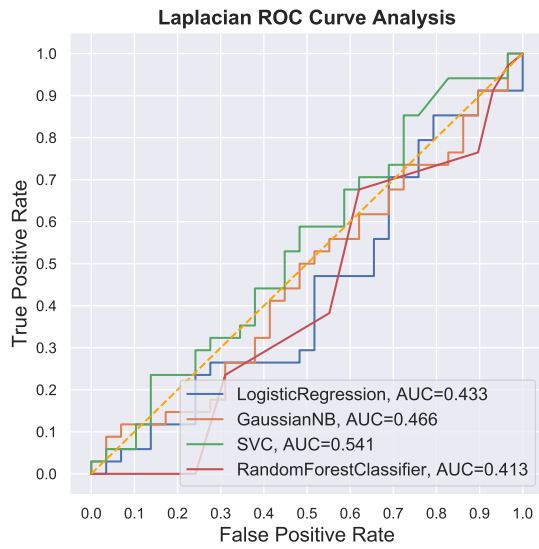Figure 1.22: Clustering Accuracy (Sonar Dataset)

Figure 1.23: ROC Curves (Sonar Dataset) 1
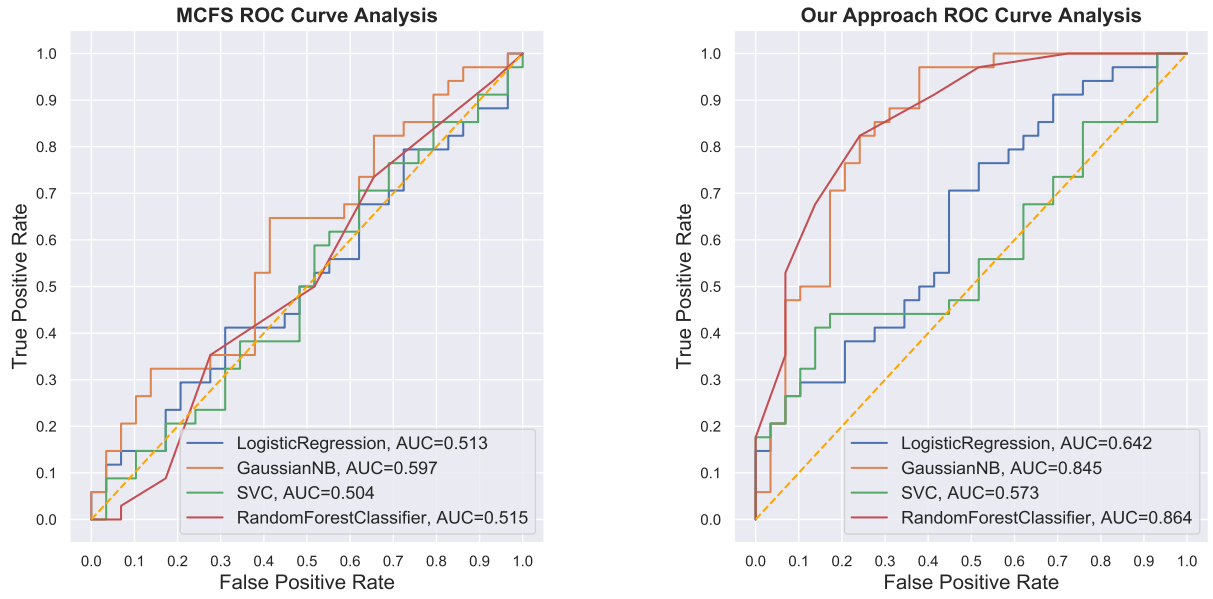
Figure 1.24: ROC Curves (Sonar Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.72008	0.527482	0.796367	0.76677	0.79827	0.76724

Table 1.7: Representation Entropy (Sonar Dataset)

Emotions Dataset
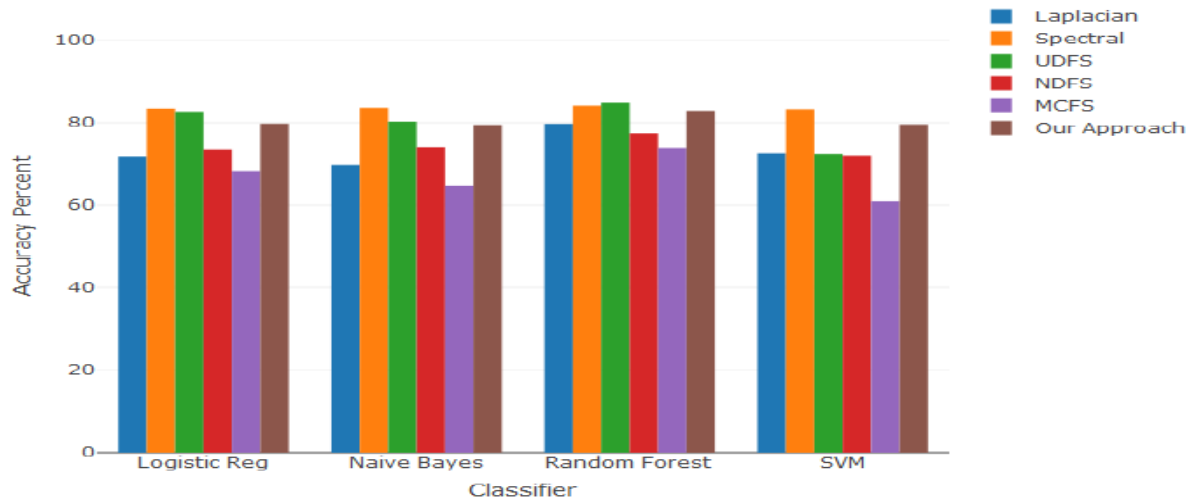
Source: https://www.openml.org/d/40589
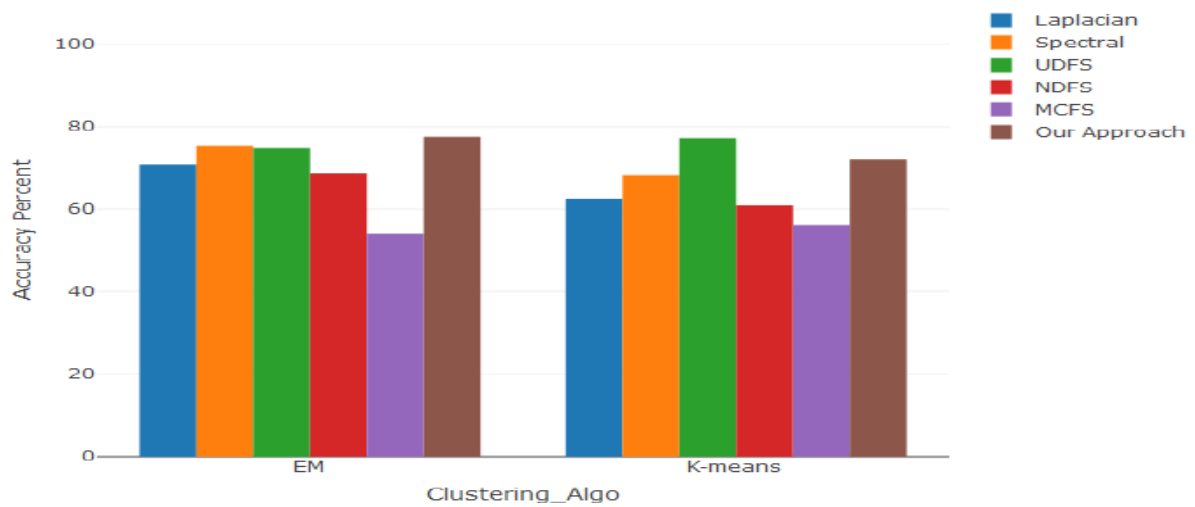

Figure 1.25: Classification Accuracy (Emotions Dataset)
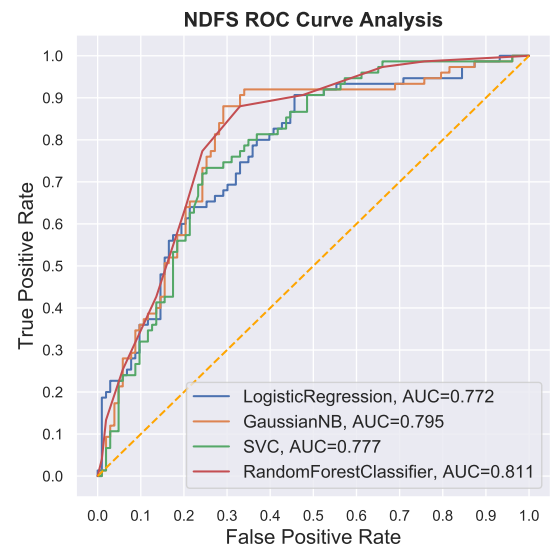

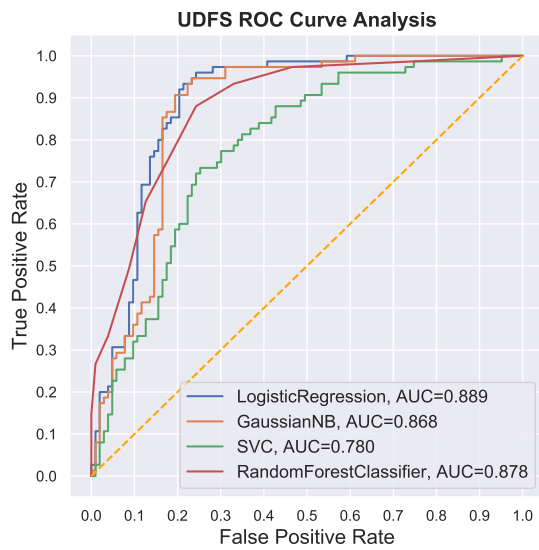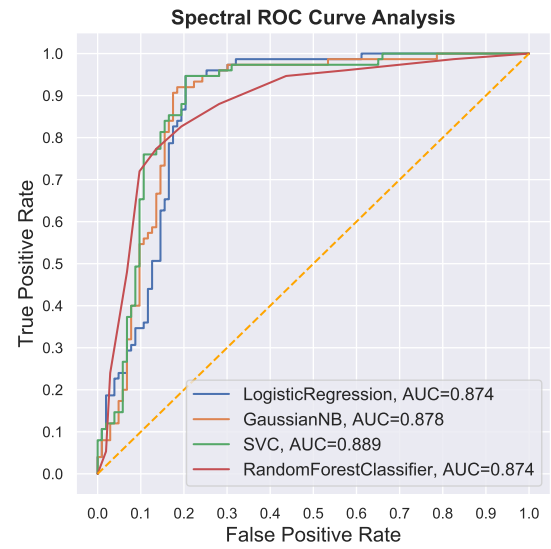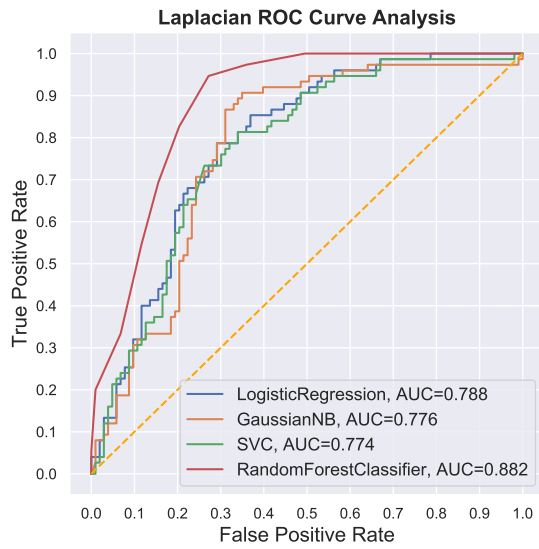Figure 1.26: Clustering Accuracy (Emotions Dataset)

Figure 1.27: ROC Curves (Emotions Dataset) 1
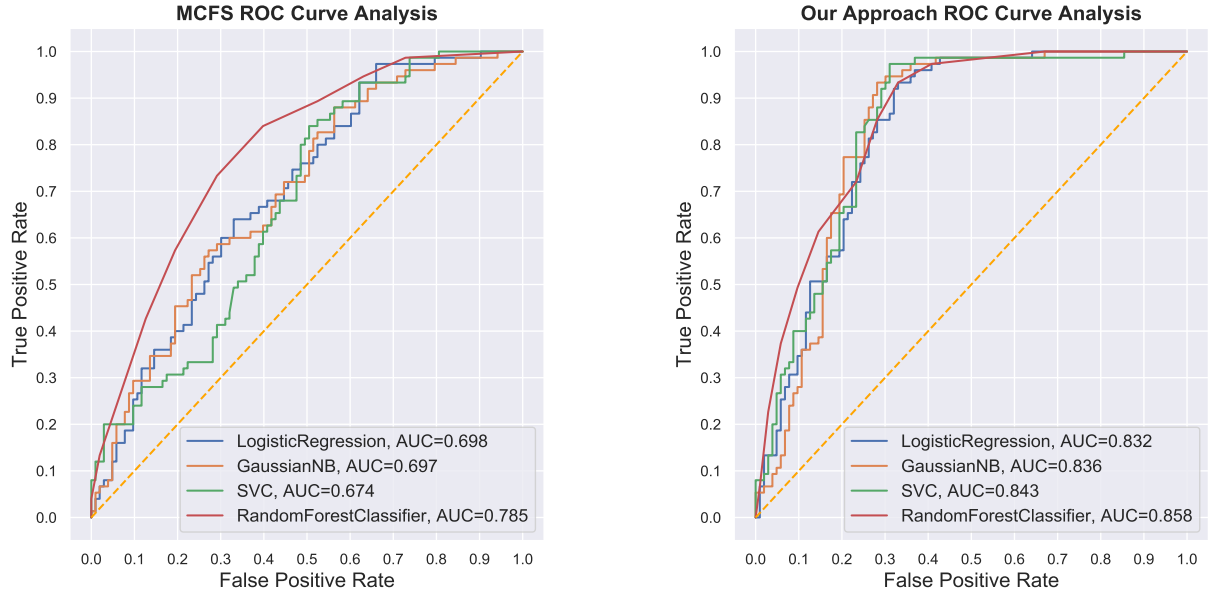
Figure 1.28: ROC Curves (Emotions Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.20329	0.91409	0.20628	0.20617	0.01400	0.46283

Table 1.8: Representation Entropy (Emotions Dataset)

Robot-Fail Dataset
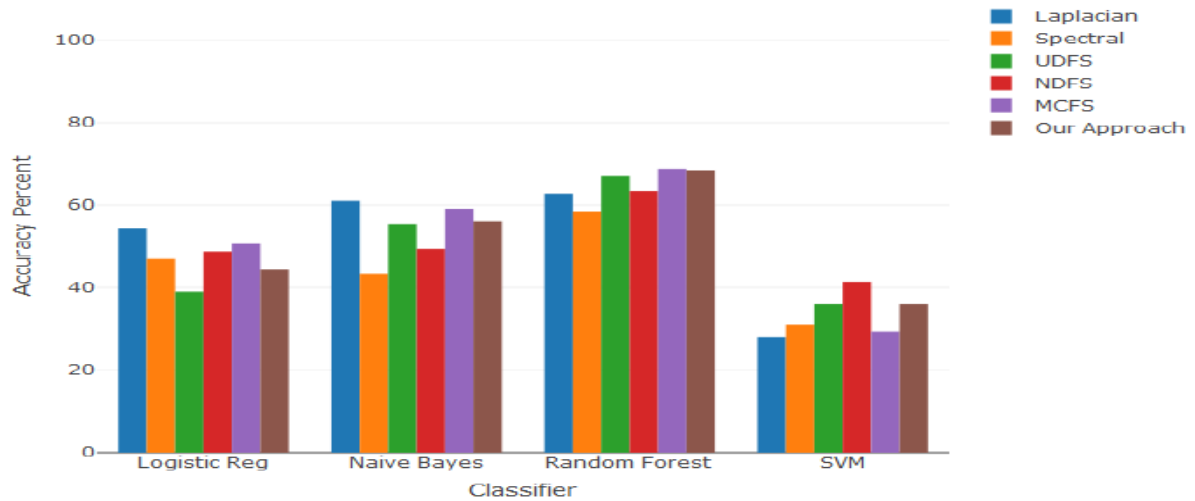**Source:** https://www.openml.org/d/1520
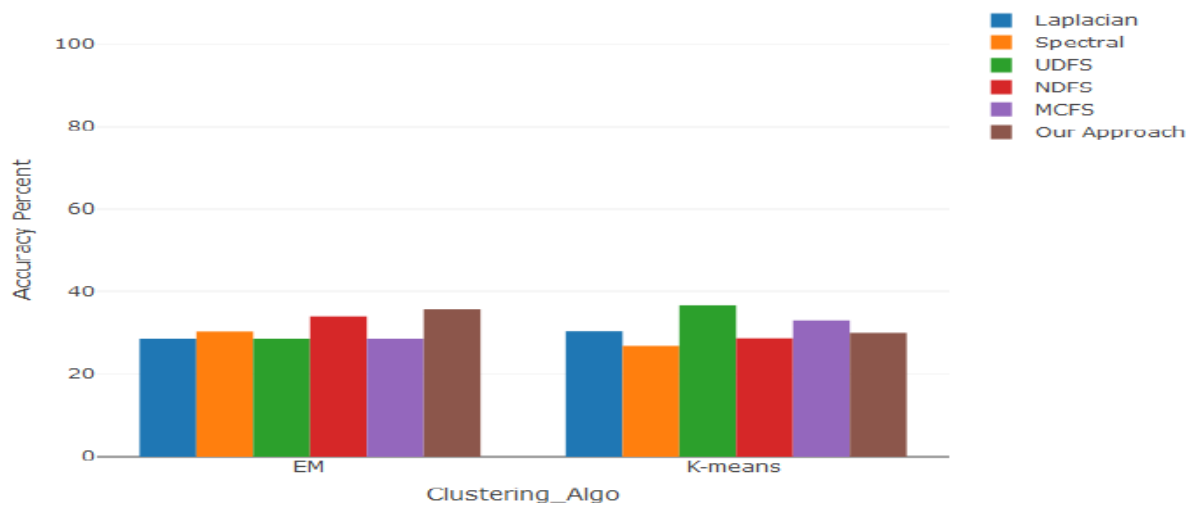

Figure 1.29: Classification Accuracy (Robot-Fail Dataset)
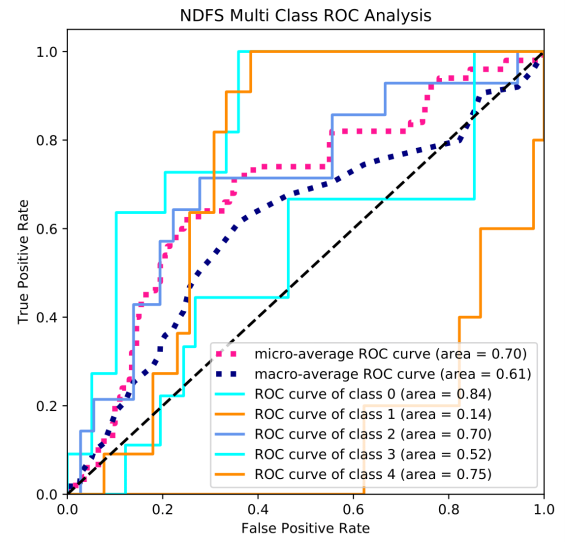

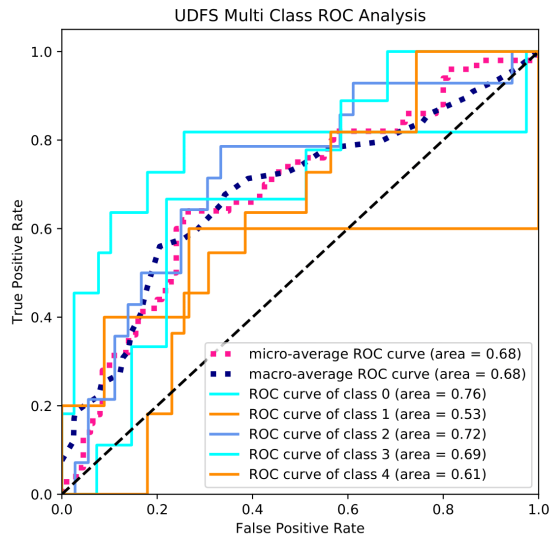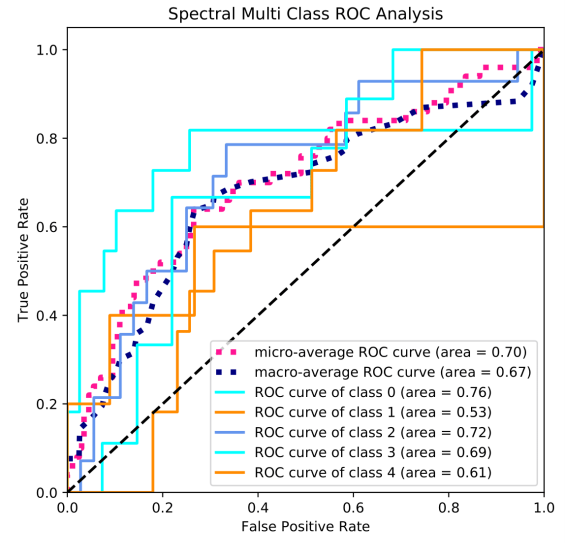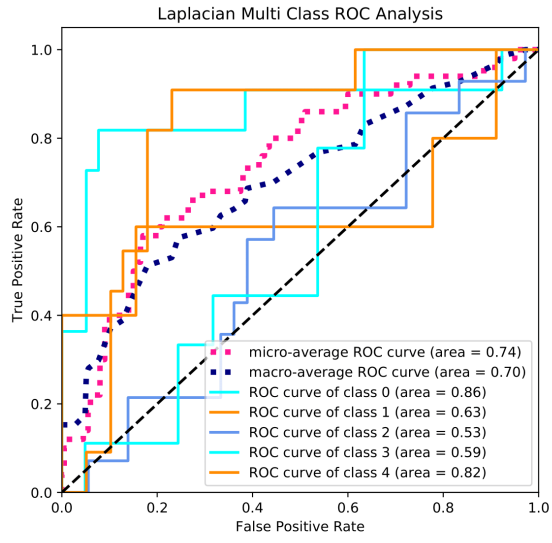Figure 1.30: Clustering Accuracy (Robot-Fail Dataset)

Figure 1.31: ROC Curves (Robot-Fail Dataset)(1) Dataset) 1

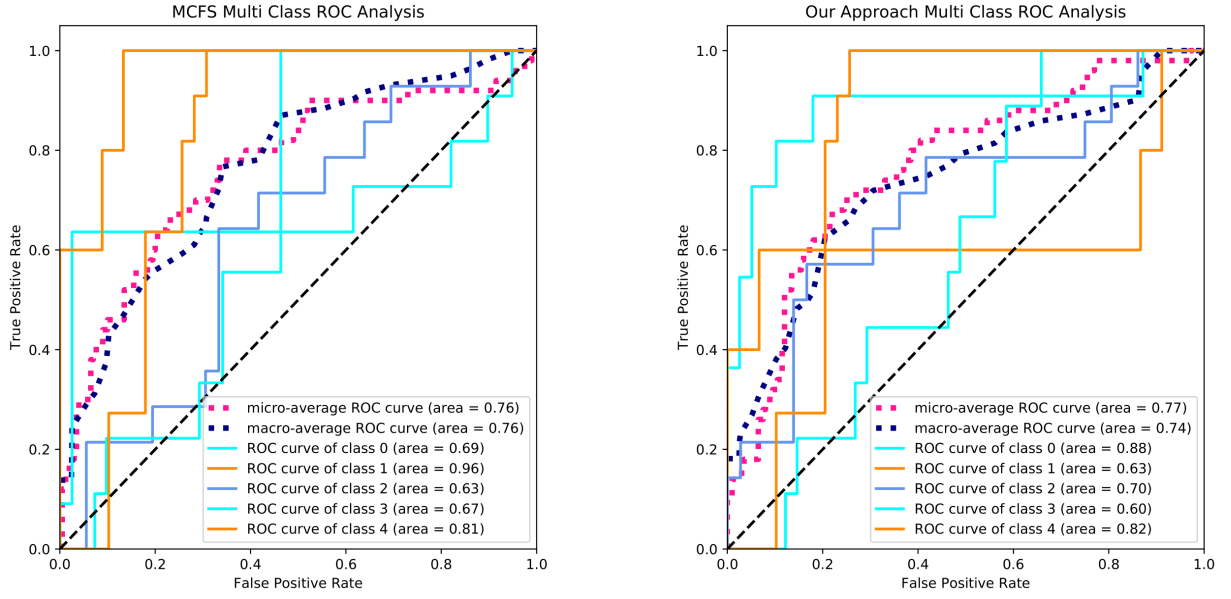Figure 1.32: ROC Curves (Robot-Fail Dataset)(1) Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.38221	0.87985	0.54304	0.92084	0.29013	0.44478

Table 1.9: Representation Entropy (Robot-Fail Dataset)

Spectrometer Dataset
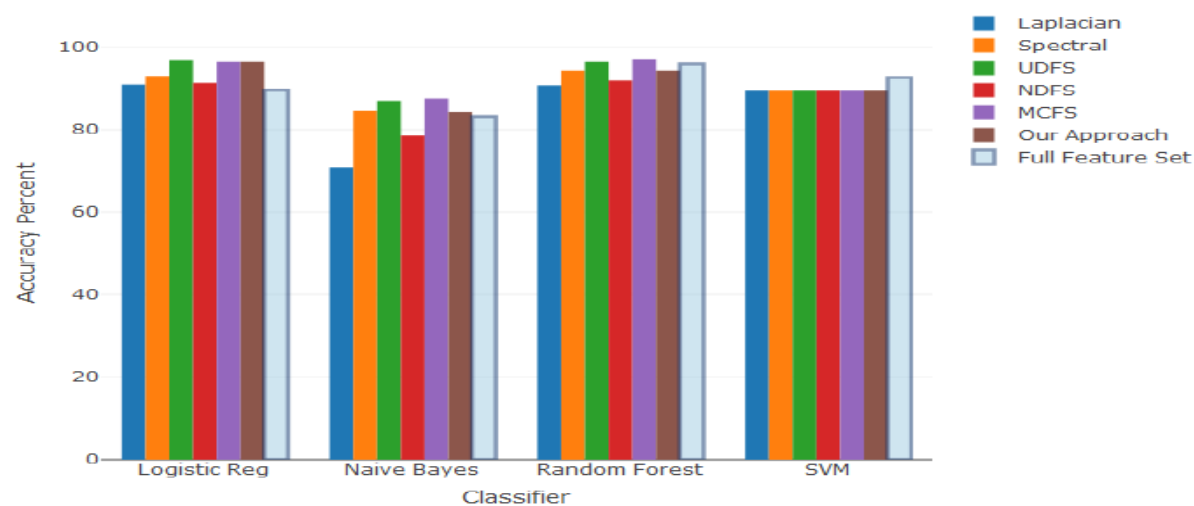**Source:** https://www.openml.org/d/954
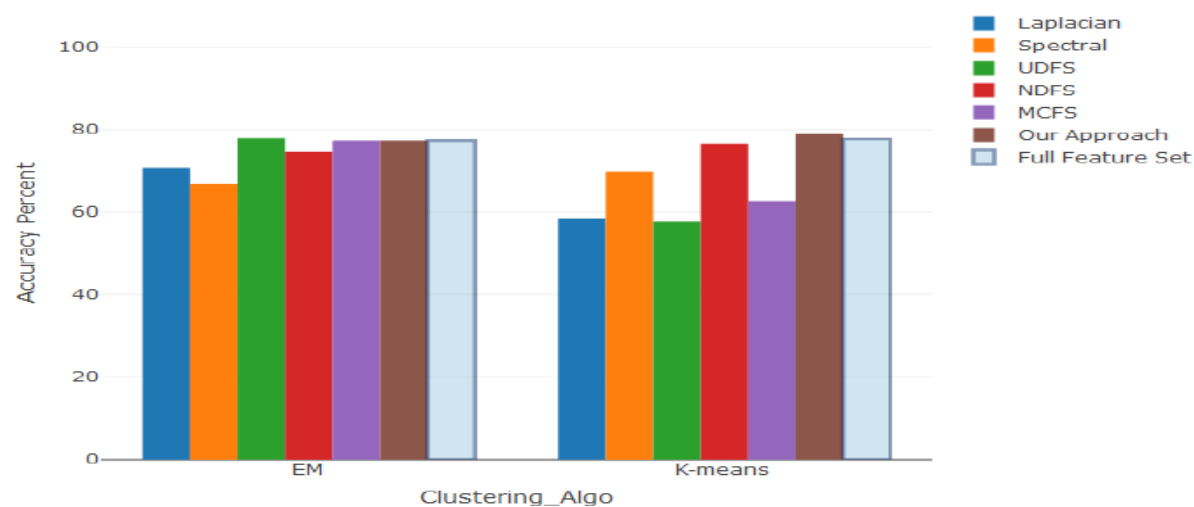

Figure 1.33: Classification Accuracy (Spectrometer Dataset)
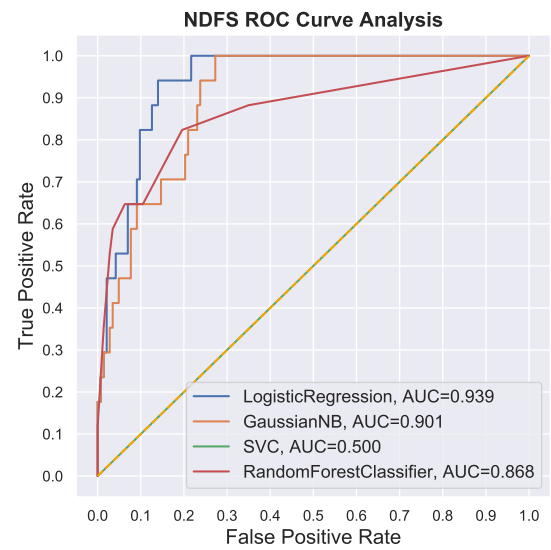

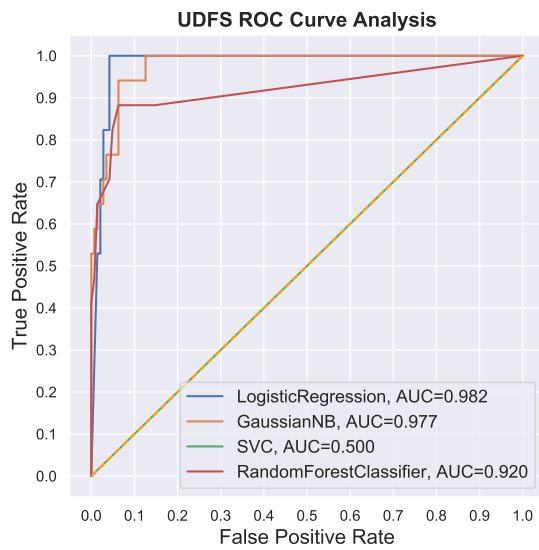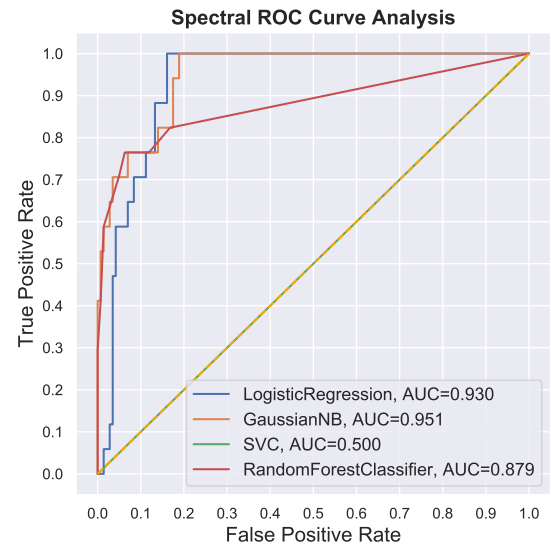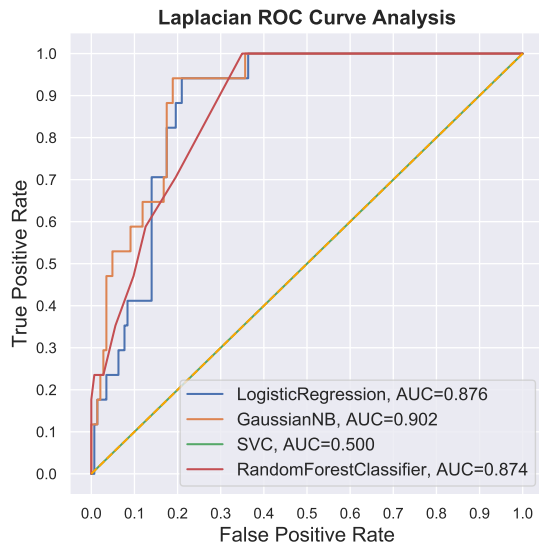Figure 1.34: Clustering Accuracy (Spectrometer Dataset)

Figure 1.35: ROC Curves (Spectrometer Dataset) 1
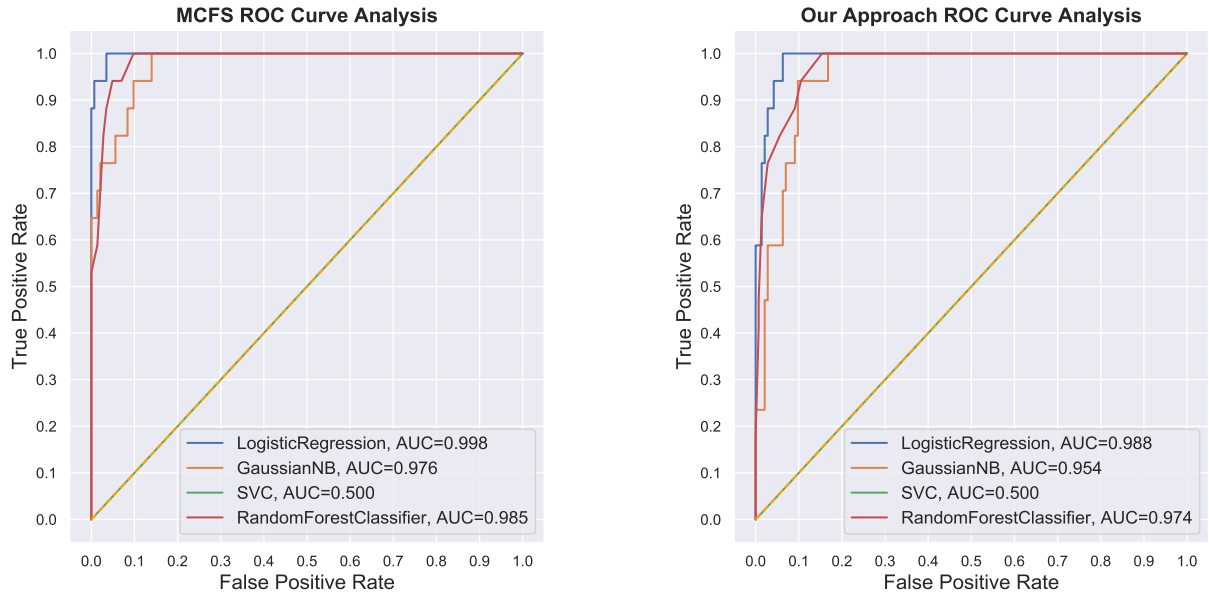
Figure 1.36: ROC Curves (Spectrometer Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.28753	-	0.37981	0.44524	0.43421	0.4492

Table 1.10: Representation Entropy (Spectrometer Dataset)
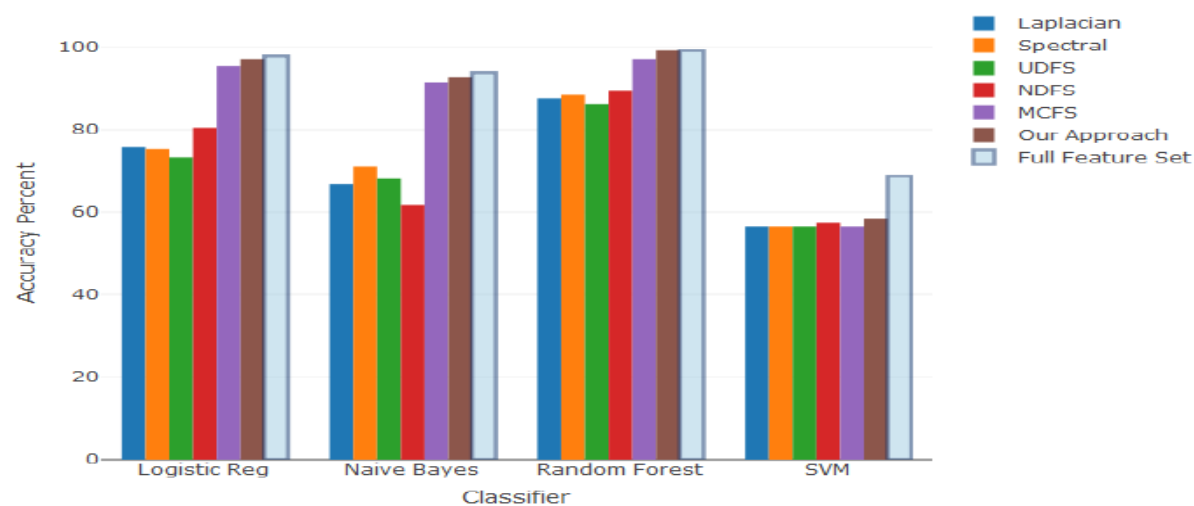
Musk Dataset
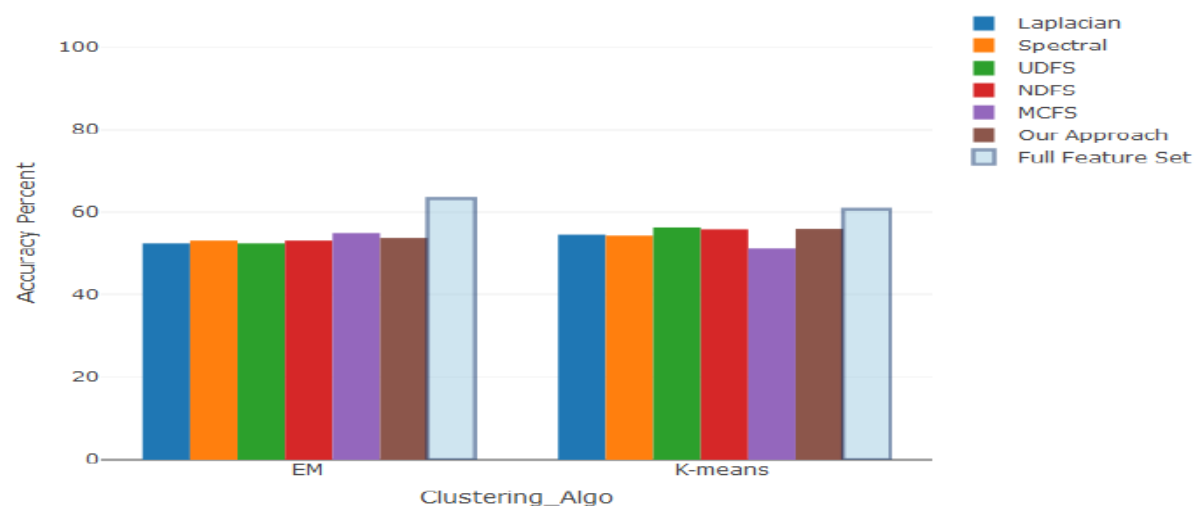
Source:

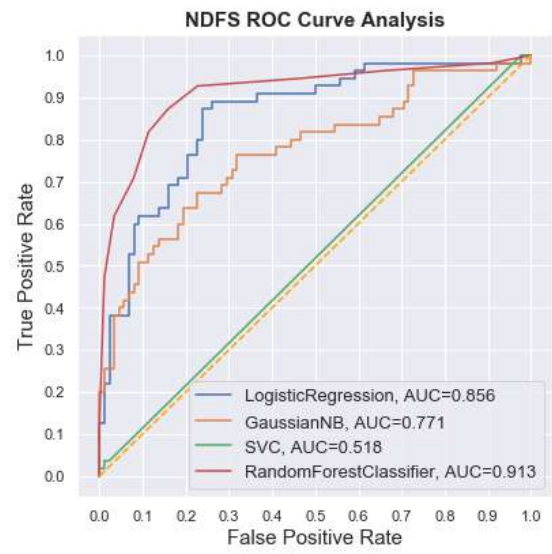Figure 1.37: Classification Accuracy (Musk Dataset)
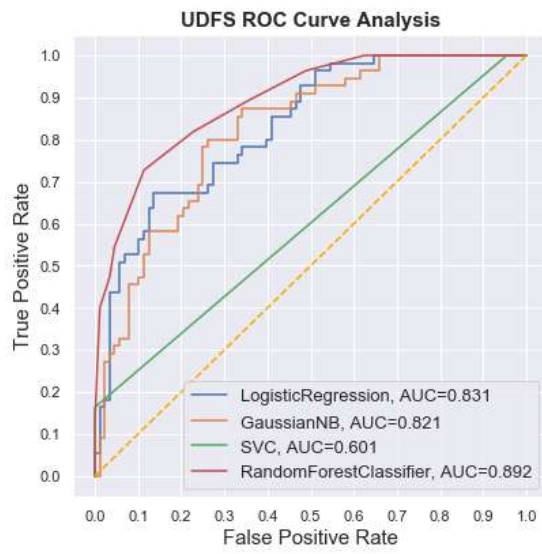

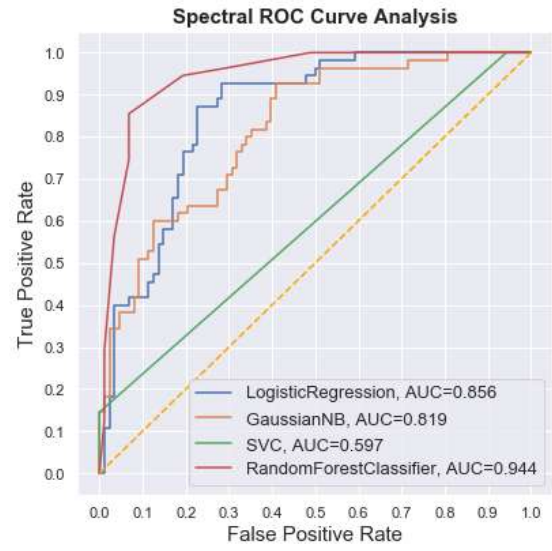Figure 1.38: Clustering Accuracy (Musk Dataset)
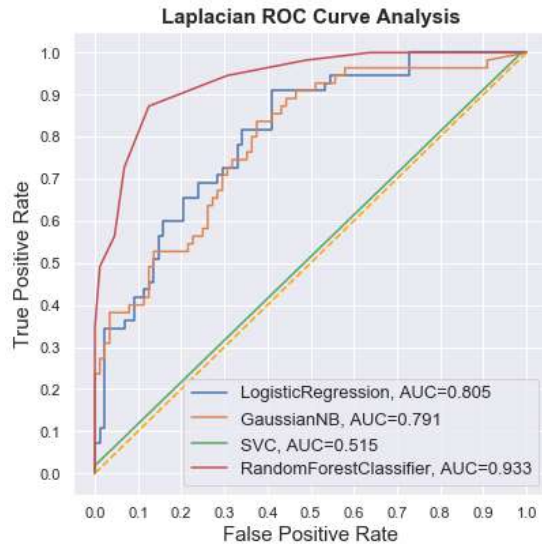
Figure 1.39: ROC Curves (Musk Dataset) 1
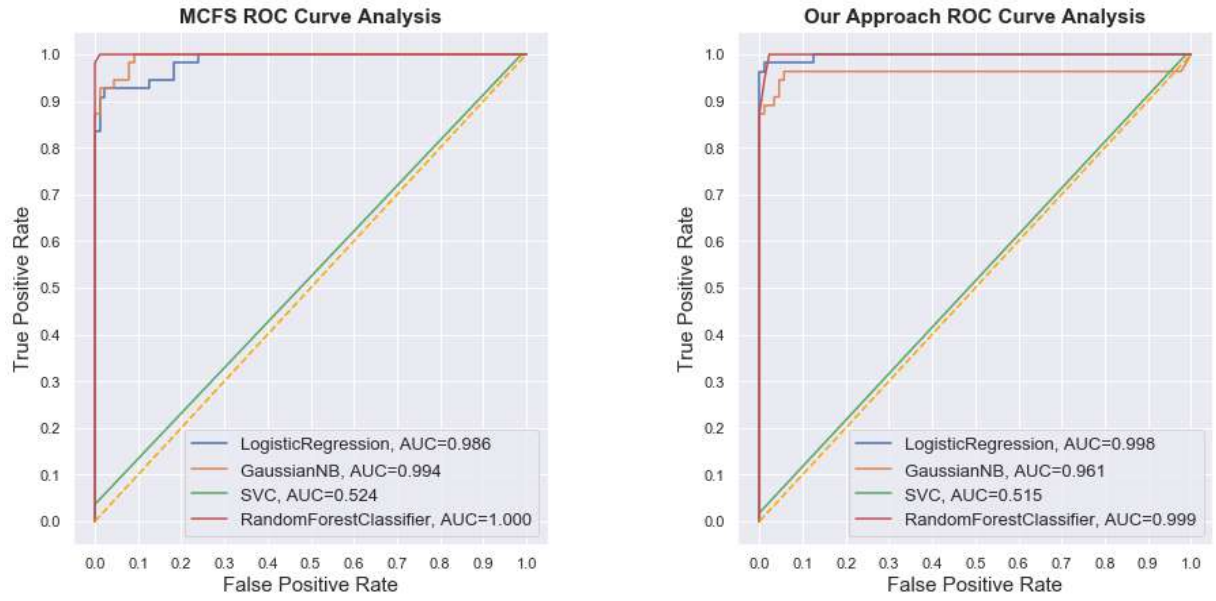
Figure 1.40: ROC Curves (Musk Dataset) 2

	Laplacian	Spectral	UDFS	NDFS	MCFS	Our Approach
Rep Entropy	0.91315	0.90626	0.84794	0.7526	0.94620	1.0111

Table 1.11: Representation Entropy (Musk Dataset)