

Overall Results Report of DWDM Mini Project

A. Libraries

Items	Sotatus
time	used
numpy	used
pandas	used
matplotlib.pyplot	used
seaborn	used
matplotlib.patches.TextContainer	used
sklearn.preprocessing.OneHotEncoder	used
sklearn.preprocessing.LabelBinarizer	used
sklearn.metrics.Similarity	used
sklearn.preprocessing.StandardScaler	used
sklearn.preprocessing.MinMaxScaler	used
sklearn.cluster.MPST_Markov	used
sklearn.cluster.MKMeans	used
sklearn.metrics.silhouette_score	used
sklearn.metrics.adjusted_rand_score	used
sklearn.metrics.adjusted_mutual_info_score	used
mutual feature selection.SequentialFeatureSelector	used
sklearn.decomposition.PCA	used
sklearn.model_selection.train_test_split	used
sklearn.metrics.grid_search	used
sklearn.metrics.classification_report	used

B. Data Preview

Items	Sotatus
basic operations	done
description	done
information	done
memory utilization	done

C. Data Cleaning

Items	Items subtype	Status
remove irrelevant observations	-	done
remove duplicate	-	done
filter unwanted outliers	-	done
missing data	-	done
	dropping - column	done
	dropping - rows	done
nan	filling - rows	done
	filling - columns	not used
	simplification	done
	equi-width	tried
binning	equi-depth	done
	mean	done
	id	done
	name	done
	values	done
replacements	elimination	done
	classes - dependent	done
	classes - independent	tried
	classes - priority	done

D. Data Encoding

Items	Status
One Hot Encoder	used
Label Binarizer	used
Label Encoder	not used
TransactionEncoder	used

E. Data Normalization

Items	Satatus
Standard Scalar	used
Min Max Scalar	used
Z Score Scalar - std	tried
Z Score Scalar - var	not used
Decimol Scalar	tried

F. Unsupervised on Data

[illegible]

G. Features on Data

Items	Items subtypes	Items subtypes	Checked(best case)	Status	Score	Hyper Parameter Features left with	
Feature Selection	Filter Methods	Observation correlation	corr	used	normal	70	
		Chi-squared Score	chi	not used	-	59	
		Variance Threshold	V1	not used	-	-	
	Wrapper Methods	Forward Selection	SFS	tried	-	-	-
		Backward Elimination	SFS	used	0.91	40	-
		Subselection Elimination	SFS	not used	-	-	-
		Recursive/Exhaustive Feature Elimination	RFE	tried	-	-	-
			EFS	tried	-	-	-
			D1	tried	-	-	-
	Embedded Methods	-	lasso regression(1)	not used	-	-	-
			ridge regression(2)	not used	-	-	-
			elastic net(3)	tried	-	-	-
Feature Extraction Methods		-	PCA	used	0.95	27	
			Vectorizer	tried	-	-	
Winner is Observation, corr, SFS, PCA giving final features as 27							

H. Supervised on Data

Items	Items subtypes	Checked(test case)	Status	Score		Hyper Parameter
				Training	Testing	N/Xmax depth/gamma
Classification	KN	mean	used	0.88	normal	3
	DT	entropy	used	0.85		11
	gini	gini	used	0.87		11
	RF	entropy	used	0.98		10
	gini	gini	used	0.99	0.955	11
	SVM	linear	used	0.98	0.97	0.17/0/500
		rbf	used	0.4	1	
		poly	used	1	0.91	0.17/0/500
Regression	LR	-	tried	-	-	-
	GD	-	-	not used	-	-
	PCA	-	used	1	0.88	30
	PolyR	-	tried	-	-	-
Winner is RF: for max depth 10						