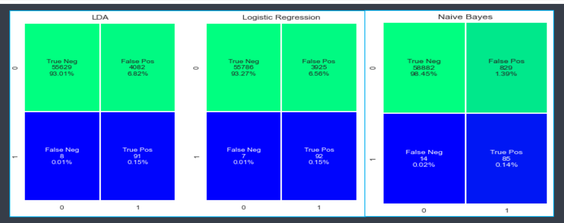
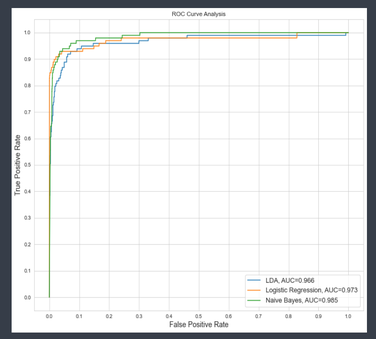
FRAUD TRANSACTION IDENTIFICATION

ML MODELS TO DETECT FRAUD TRANSACTIONS FROM HEAVILY IMBALANCED DATASET

Final Model Selection

ModelName	Features	Approach	Accuracy	Precision	Recall	confmatrix_TN	confmatrix_FP	confmatrix_FN	confmatrix_TP	auc
Logistic Regression	all	Direct	0.93426	0.02290	0.92929	55786	3925	7	92	0.972997
LDA	all	Direct	0.93162	0.02181	0.91919	55629	4082	8	91	0.966454
Naive Bayes	all	Direct	0.98591	0.09300	0.85859	58882	829	14	85	0.985198





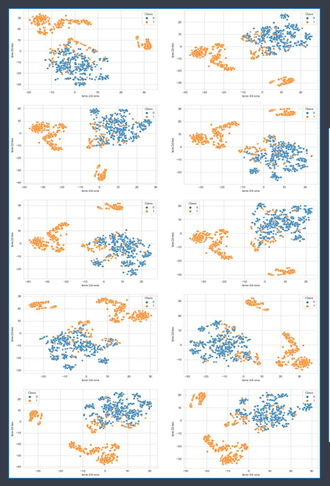
- Hyper Tuning parameters
- Best performing models – LDA, Logistic Regression, Gaussian Naïve Bayes

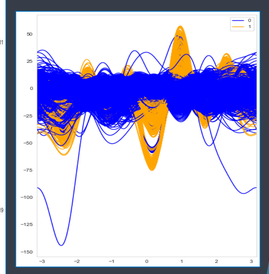
Model Build and Selection Process

- Primarily 2 approaches used against Selective Features and All Features files:
  - SMOTE applied on each of K-Folds Training Set
  - ADASYN applied on Training Set
- Model selection based on Recall, False Positives

ModelName	Features	Approach	Accuracy	Precision	Recall	confmatrix_TN	confmatrix_FP	confmatrix_FN	confmatrix_TP	auc
Logistic Regression	all	Direct	0.93426	0.02290	0.92929	55786	3925	7	92	0.972997
Logistic Regression	selective	Direct	0.91404	0.01760	0.92929	54577	5134	7	92	0.975582
Naive Bayes	all	Direct	0.96407	0.04077	0.91919	57570	2141	8	91	0.981288
LDA	all	Direct	0.93162	0.02181	0.91919	55629	4082	8	91	0.966454
Logistic Regression	all	kFold	0.97228	0.05175	0.90909	58062	1649	9	90	0.967842
Logistic Regression	selective	kFold	0.97184	0.05048	0.89899	58037	1674	10	89	0.974815
Naive Bayes	selective	Direct	0.97089	0.04890	0.89899	57980	1731	10	89	0.977431
Naive Bayes	all	kFold	0.97604	0.05828	0.88889	58289	1422	11	88	0.981534
Naive Bayes	selective	kFold	0.98699	0.10106	0.86869	58946	765	13	86	0.975682
LDA	selective	Direct	0.92498	0.01886	0.86869	55237	4474	13	86	0.960935

High Dimension Visualizations

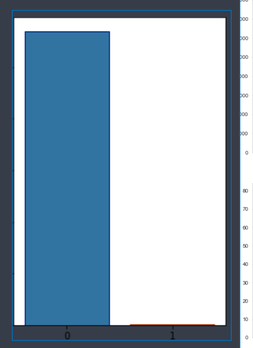


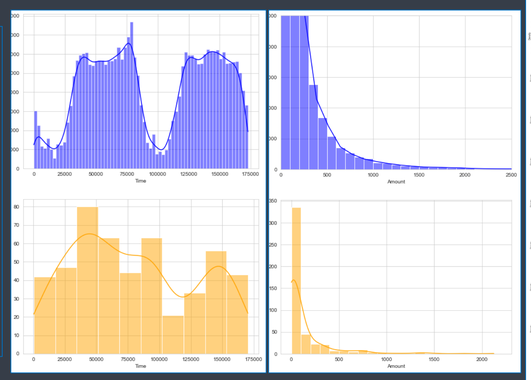


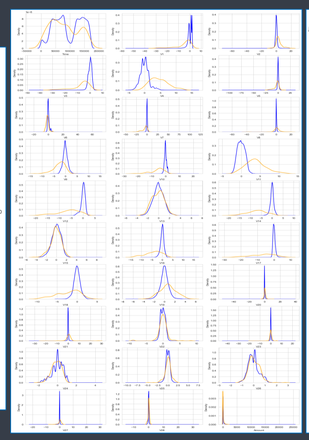
- Some features demonstrate more distinguishing characteristics.

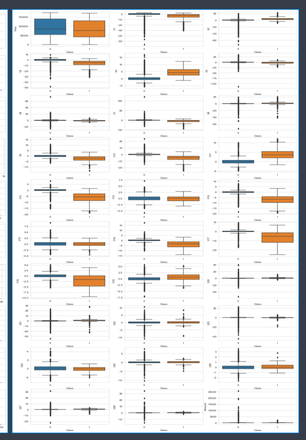
EDA

- Highly Imbalanced Data *(99.8% Authentic, 0.2% Fraud)*
- Interpretable Columns (Time, Amount) not providing major distinction.
- Individual Features show shifted distributions for different classes.









DEBJIT RAY (12010066)