Intrusion Detection using Machine Learning

By Muhammad Ali Haider

Intrusion Detection

- An intrusion detection for malicious activity
- Malicious traffic is any suspicious link, file or connection over the network.
- It is any software intentionally designed to cause harm or make changes to a computer, system, or data—without appropriate consent

	duration	protocol_type	service	flag	src_bytes	dst_bytes
0	0.00000	tcp	ftp_data	SF	491.00000	0.00000
1	0.00000	udp	other	SF	146.00000	0.00000
2	0.00000	tcp	private	S0	0.00000	0.00000
3	0.00000	tcp	http	SF	232.00000	8153.0000
4	0.00000	tcp	http	SF	199.00000	420.00000
5	0.00000	tcp	private	REJ	0.00000	0.00000

...

dst_host_srv_serror_rate	dst_host_rerror_rate	dst_host_srv_rerror_rate	class
0.00000	0.05000	0.00000	normal
0.00000	0.00000	0.00000	normal
1.00000	0.00000	0.00000	anomaly
0.01000	0.00000	0.01000	normal
0.00000	0.00000	0.00000	normal
0.00000	1.00000	1.00000	anomaly

Dataset

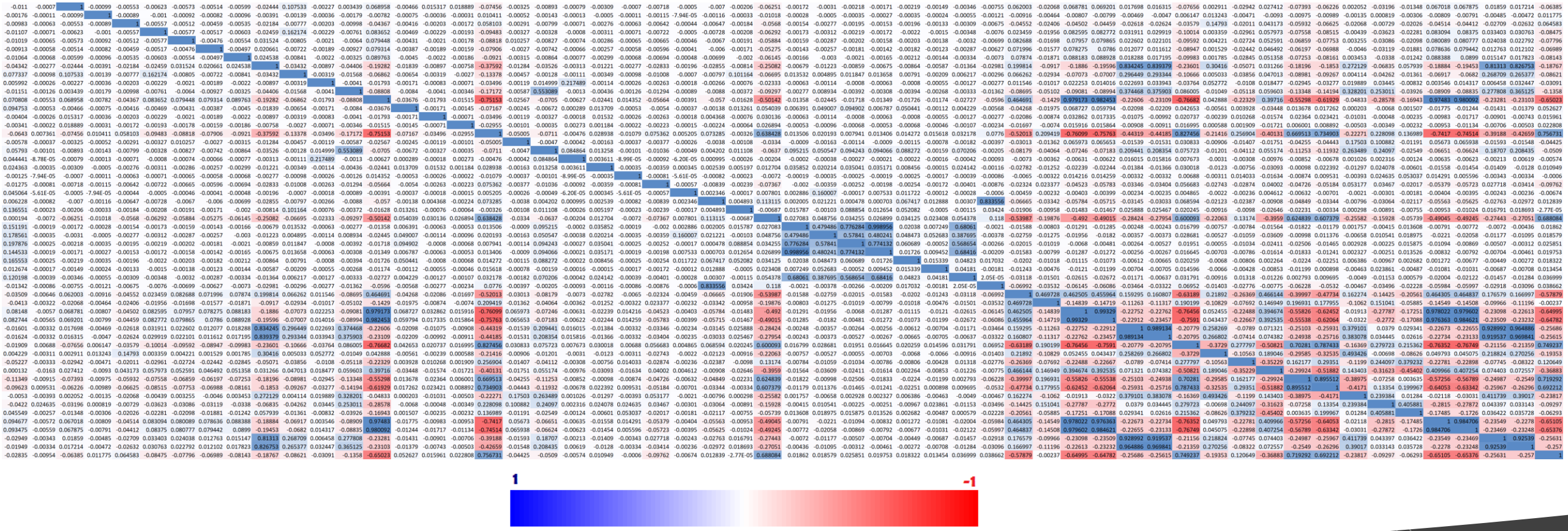
- 41 independent variables (e.g. service, flag,src_bytes,…...)
- 1 response variable of with categories (normal, anomaly)
- 25,192 rows
- Data source: https://www.kaggle.com/sampadab17/network-intrusion-detection

Steps

- Data cleaning
- Data visualization
- Data Manipulation (Feature scaling, PCA, etc)
- Build the model and run
- Results

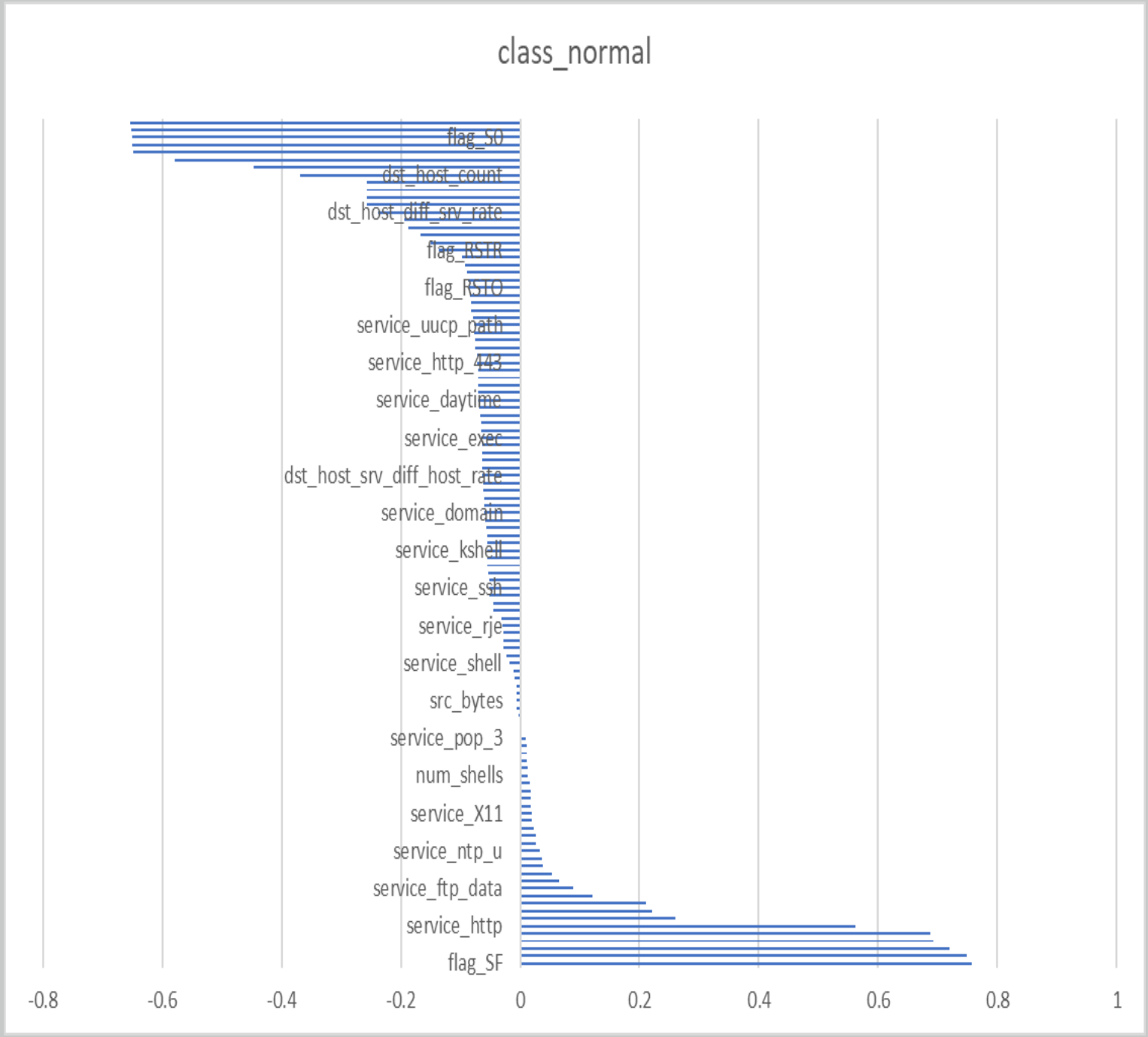
Data Cleaning

- No missing values found
- No outliers found (Z Scores and IQR)
- No inconsistent data found
- No data inconsistencies found
- No imbalance data



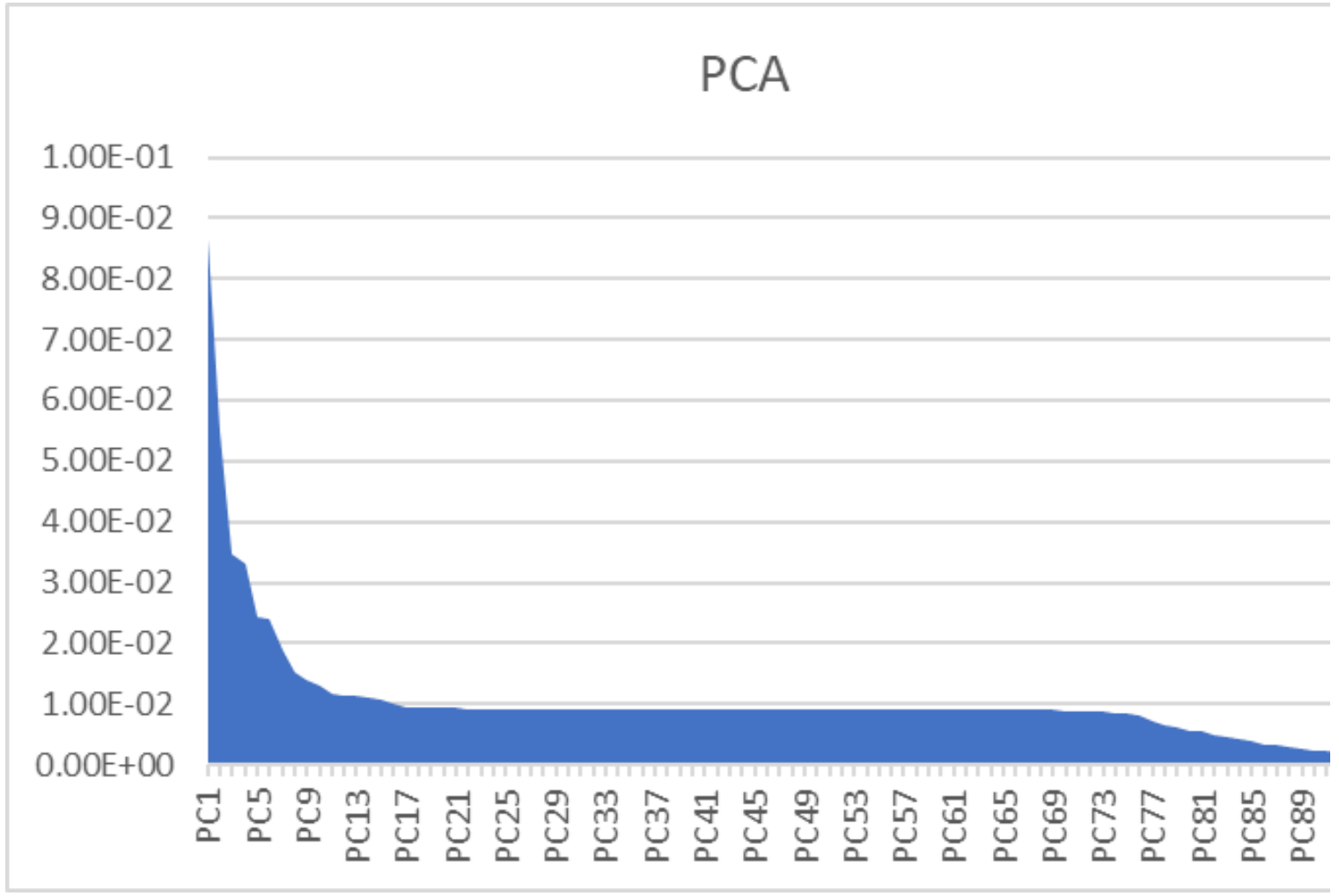
Correlation matrix

- Majority of the chart shows white which means close to zero correlation
- Drop 3 features with high correlations

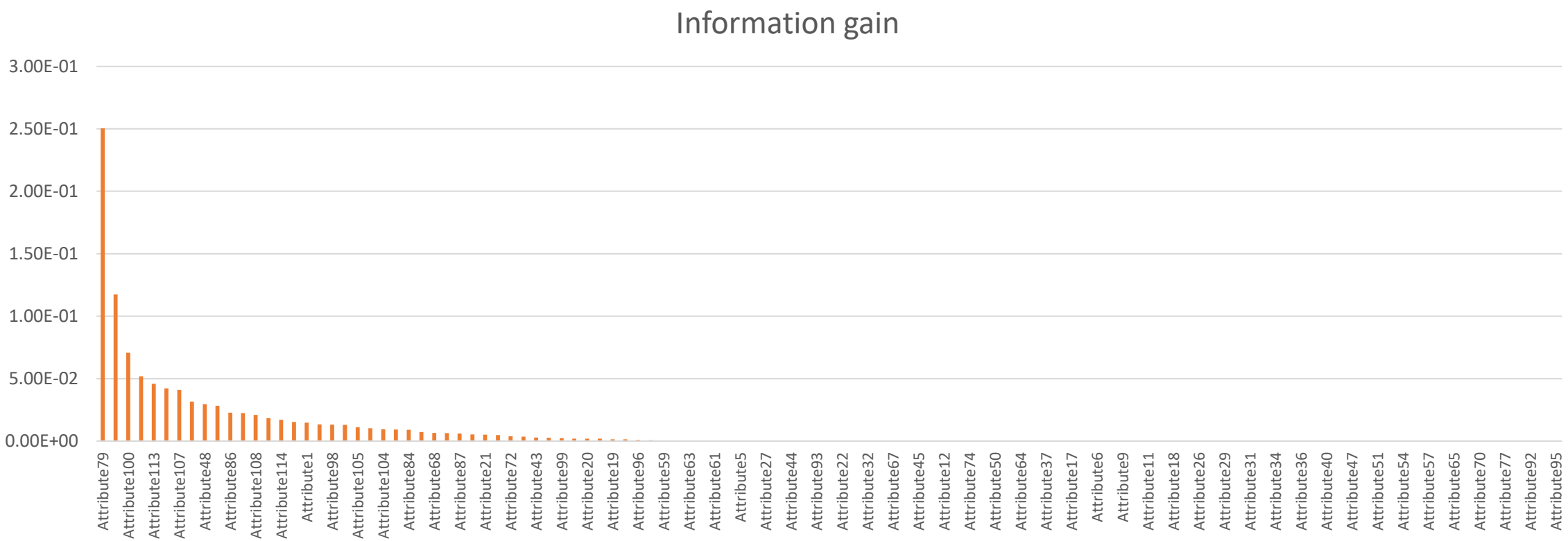


Correlation with response variable

- Tail is not long relatively
- Drop another 9 variables with poor correlation

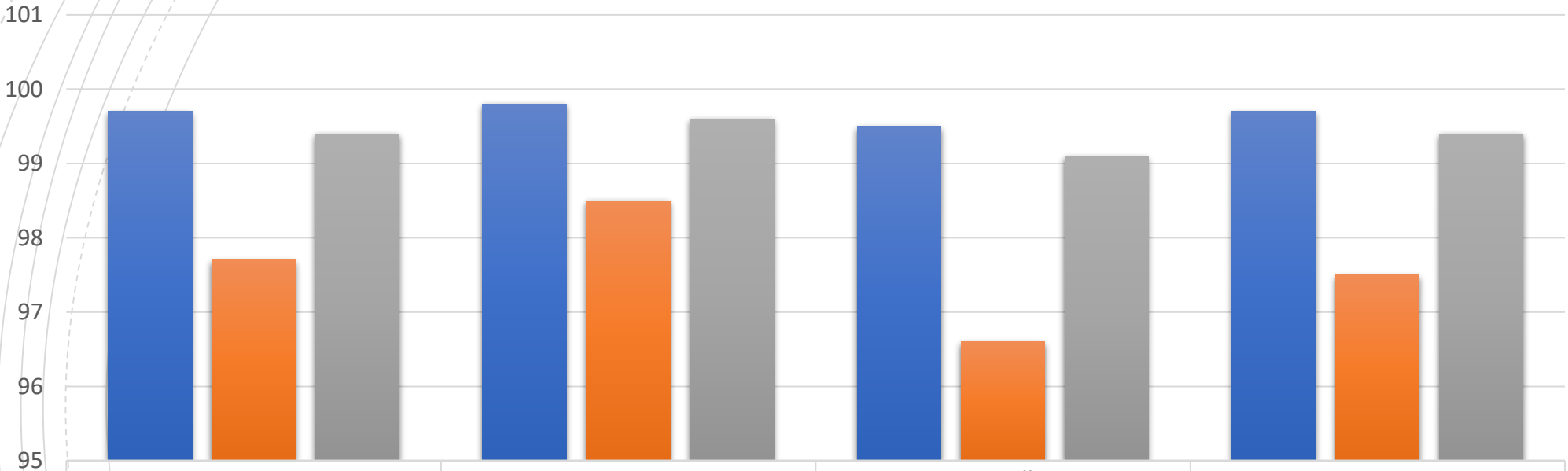


Principal component analysis



Random forest- Importance of attributes

Results comparison different Algorithums



	Accuracy	Precision	Recall	F1
Random Forest	99.7	99.8	99.5	99.7
SVM	97.7	98.5	96.6	97.5
Deep Learning	99.4	99.6	99.1	99.4

Random Forest SVM Deep Learning