CS631 Cyber Security of Critical Infrastructure
Homework 5 – Report

Sanjeev Lal (19111077)

Avesh Kumar Agrawal (19111020)

Manish Patel (19111051)

Pratik Kumar (19111065)

Virendra Nishad (19111099)

Decision Tree Model With All Features

1) Converted categorical columns into numerical value by LabelEncoding.
2) Generalize various attack labels into 5 categories (Normal, DoS, Probe, R2L, U2R).
3) Split full labelled data as 75% train and 25% test.
4) Normalized the train and test data.
5) We have trained our model on 75% train data using Cross Validation with 3 fold and parameters as criterion = 'entropy', max_depth = 7, min_samples_leaf = 4.

Accuracy
99.96%

Confusion matrix

		Predicted Class				
		Normal (0)	DoS (1)	Probe (2)	R2L (3)	U2R (4)
Actual Class	Normal (0)	243605	7	60	13	0
	DoS (1)	23	970325	8	0	0
	Probe (2)	148	38	10079	0	0
	R2L (3)	147	0	0	129	2
	U2R (4)	11	0	0	7	3

Decision Tree Model With Feature Selection

1) Converted categorical columns into numerical value by LabelEncoding.
2) Generalize various attack labels into 5 categories (Normal, DoS, Probe, R2L, U2R).
3) Split full labelled data as 75% train and 25% test.
4) Normalized the train and test data.
5) Selected 14 features using Recursive Feature Elimination (RFE) Technique namely ['protocol_type', 'service', 'flag', 'src_bytes', 'dst_bytes', 'wrong_fragment', 'num_compromised', 'count', 'diff_srv_rate', 'dst_host_diff_srv_rate', 'dst_host_same_src_port_rate', 'dst_host_srv_diff_host_rate', 'dst_host_serror_rate', 'dst_host_rerror_rate']
6) We have trained our model on 75% train data using Cross Validation with 3 fold and parameters as criterion = 'entropy', max_depth = 7, min_samples_leaf = 4.

Accuracy
99.96%

Confusion matrix

		Predicted Class				
		Normal (0)	DoS (1)	Probe (2)	R2L (3)	U2R (4)
Actual Class	Normal (0)	243607	5	60	13	0
	DoS (1)	23	970325	7	1	0
	Probe (2)	148	38	10079	0	0
	R2L (3)	148	0	4	129	0
	U2R (4)	12	0	0	7	2