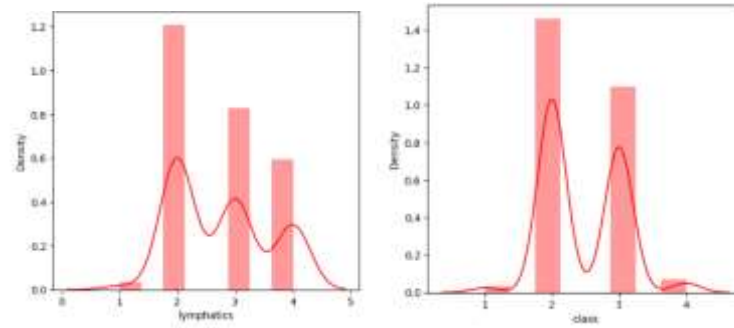
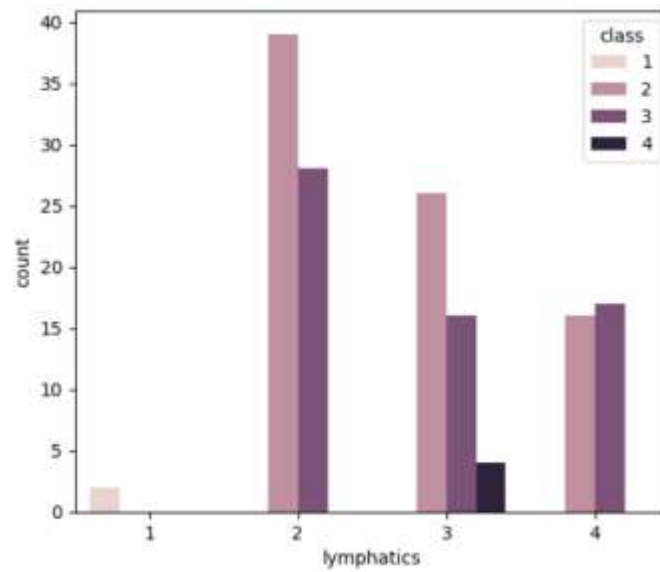


[illegible]

Univariate Analysis

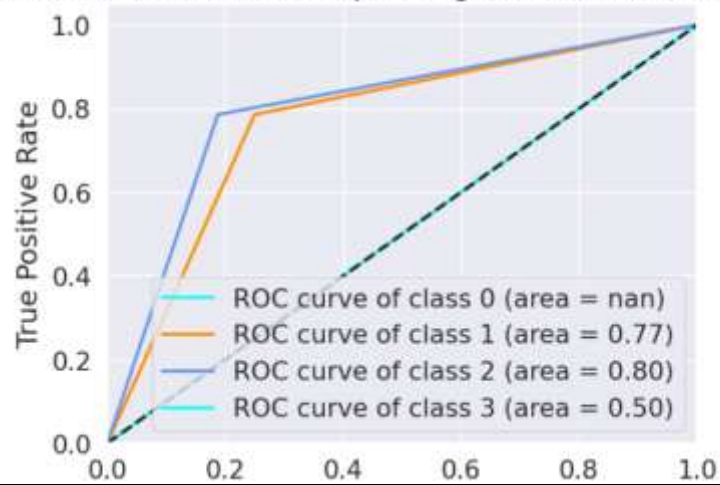


Bivariate Analysis



Multivariate
Analysis

Some extension of Receiver operating characteristic to multi-class



Outliers
and
Anomalies

-

Data Preprocessing Code Screenshots

Loading
Data

```
1 df=pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/lymphography/lymphography.data",names=col_names)
2 print("size of dataset:",df.shape)
3 df.head()
```

size of dataset: (148, 19)

	class	lymphatic	block of affere	bl. of lymph	bl. of lymph	of pass	ectromassates	regeneration of	early uptake in	lym.nodes disin	lym.nodes enlar	changes in lym.	defect in node	charges in node	changes in stru	special force	dislocation of	exclusion of no	no. of nodes in
0	3	4	2	1	1	1	1	1	2	1	2	2	2	4	8	1	1	2	2
1	2	3	2	1	1	2	2	1	2	1	3	3	2	5	4	2	2	2	2
2	3	3	2	2	2	2	2	2	2	1	4	3	3	4	8	3	2	2	1
3	3	3	1	1	1	1	2	1	2	1	3	3	4	4	4	3	1	2	8
4	2	3	1	1	1	1	1	1	1	1	2	2	4	5	8	1	2	2	1

Handling Outliers	<pre> 1 for col in df.columns: 2 q1 = np.quantile(df[col],0.25) 3 q3 = np.quantile(df[col],0.75) 4 iqr = q3-q1 5 lower_bound = q1 - (1.5*iqr) 6 upper_bound = q3 + (1.5*iqr) 7 df[col] = np.where(df[col]> upper_bound,upper_bound,df[col]) 8 df[col] = np.where(df[col]< lower_bound,lower_bound,df[col]) 9 sns.boxplot(df[col]) 10 print("") 11 plt.show() </pre>
Training and Testing	<pre> [23] 1 # Assuming 'class' is your target variable and the rest are features 2 y = df['class'] # Create y to hold your target variable 3 x = df.drop('class', axis=1) # Create x to hold your features 4 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42) 1 ##check shape to make sure it is all in order 2 print("size of x_train: {} \t size of x_test: {} \t size of y_train: {} \t size of y_test: {}".format(x_train.shape,x_test.shape,y_train.shape,y_test.shape)) </pre>
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-