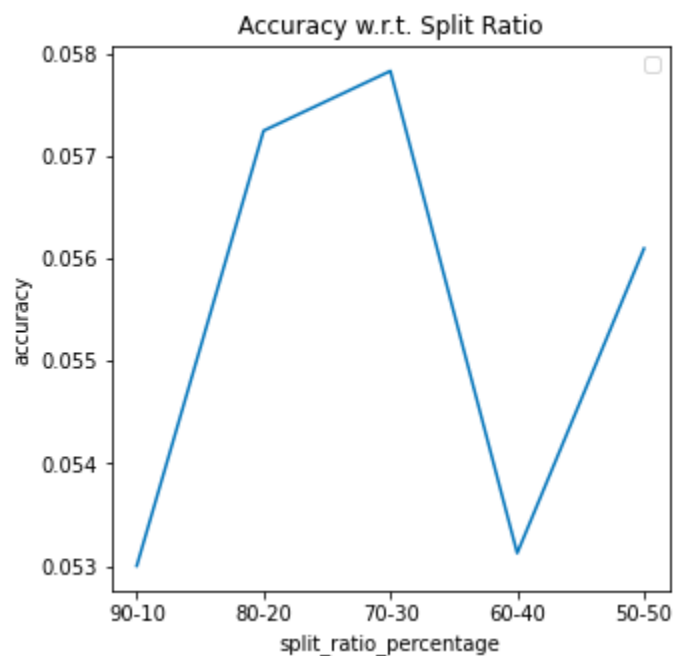


## CSE643: Artificial Intelligence

### Assignment-4

- The dataset is loaded and then all the attributes that are of object type are converted to integer type so that model can be applied to the data.
- The data is first split into x and y where y denotes the end goal which is “Suggested Job Role”. After which the data is further splitted into different ratios to check accuracy for each split.
- It can be seen that there are 33 different types of classification in y for which accuracy is maximum for 70-30 split.
- The table below shows the Accuracy for the corresponding Split Ratio Percentage.

	Split Ratio Percentage	Accuracy
0	90-10	0.053000
1	80-20	0.057250
2	70-30	0.057833
3	60-40	0.053125
4	50-50	0.056100



- Now we can see that dataframe for “Suggested Job Role” contains values as follows:

```
array(['Database Developer', 'Portal Administrator',
```

```
'Systems Security Administrator', 'Business Systems Analyst',
'Software Systems Engineer', 'Business Intelligence Analyst',
'CRM Technical Developer', 'Mobile Applications Developer',
'UX Designer', 'Quality Assurance Associate', 'Web Developer',
'Information Security Analyst', 'CRM Business Analyst',
'Technical Support', 'Project Manager',
'Information Technology Manager', 'Programmer Analyst',
'Design & UX', 'Solutions Architect', 'Systems Analyst',
'Network Security Administrator', 'Data Architect',
'Software Developer', 'E-Commerce Analyst',
'Technical Services/Help Desk/Tech Support',
'Information Technology Auditor', 'Database Manager',
'Applications Developer', 'Database Administrator',
'Network Engineer', 'Software Engineer', 'Technical Engineer',
'Network Security Engineer',
'Software Quality Assurance (QA) / Testing']], dtype=object)
```

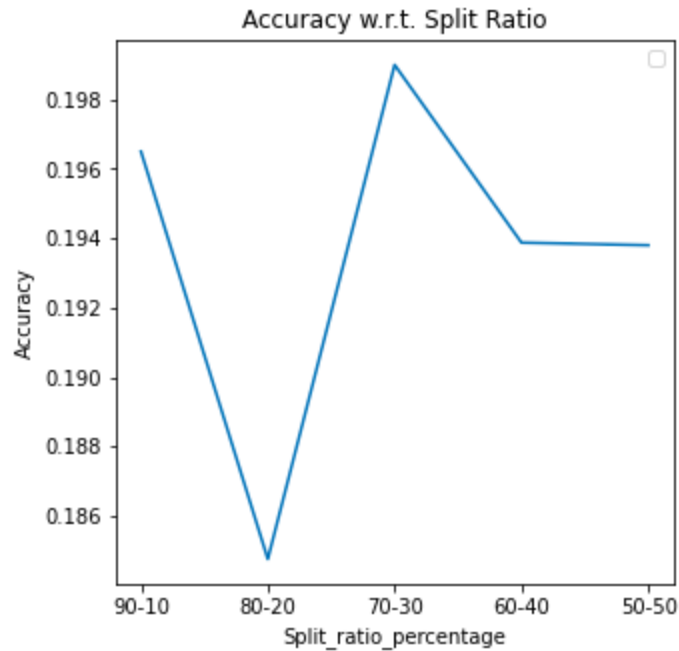
- But now we club all the similar type of jobs together i.e. club some job roles like “Business Intelligence Analyst” and “Business Systems Analyst” into “Analyst”, etc. and reduce the classes for “Suggested Job Role” which is denoted by y, hence y now consists of unique values as follows:

```
array(['Developer', 'Administrator', 'Analyst', 'Engineer', 'UX Designer',
'Quality Assurance Associate', 'Support', 'Manager', 'Design & UX',
'Architect', 'Information Technology Auditor',
'Software Quality Assurance (QA) / Testing'], dtype=object)
```

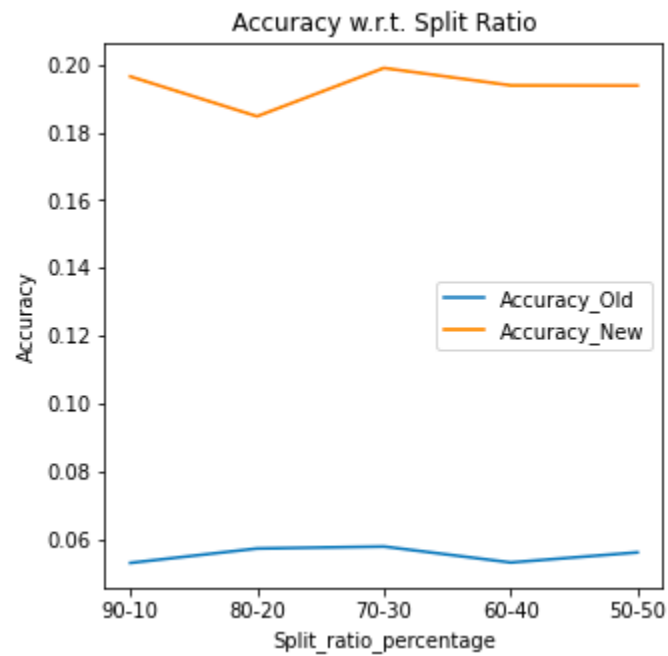
- Now again we apply the same technique as earlier and check accuracy, confusion matrix, etc. for different splits like we performed earlier.
- The Accuracy\_Old and Accuracy\_New are as follows:

	Split Ratio Percentage	Accuracy_old	Accuracy_new
0	90-10	0.053000	0.196500
1	80-20	0.057250	0.184750
2	70-30	0.057833	0.199000
3	60-40	0.053125	0.193875
4	50-50	0.056100	0.193800

- Accuracy vs Split Ratio Percentage graph for new accuracy is as follows:



- Here also we can see that we get maximum accuracy for 70-30 split.
- To compare plotting both accuracy together i.e. Accuracy\_new and Accuracy\_Old.



##All the outputs can be seen in the python file hence not included in the report.