

Artificial Intelligence | Assignment 4

Report

Data Preparation/ Modification

- To study the data properly, firstly the first few rows were sampled.

```
1 df.head()
```

	Acedamic percentage in Operating Systems	percentage in Algorithms	Percentage in Programming Concepts	Percentage in Software Engineering	Percentage in Computer Networks	Percentage in Electronics Subjects	Percentage in Computer Architecture	Percentage in Mathematics	Percentage in Communication skills	Hours working per day	Interested Type of Books	Salary Range Expected
0	69	63	78	87	94	94	87	84	61	9	Prayer books	salary
1	78	62	73	60	71	70	73	84	91	12	Childrens	salary
2	71	86	91	87	61	81	72	72	94	11	Travel	Work
3	76	87	60	84	89	73	62	88	69	7	Romance	Work
4	92	62	90	67	71	89	73	71	73	4	Cookbooks	salary

5 rows x 39 columns

- Describe was called on the dataframe to describe the data

```
1 df.describe()
```

	Acedamic percentage in Operating Systems	percentage in Algorithms	Percentage in Programming Concepts	Percentage in Software Engineering	Percentage in Computer Networks	Percentage in Electronics Subjects	Percentage in Computer Architecture	Percentage in Mathematics	Percentage in Communication skills	Hours working per day	Logical quotient rating	hackathor
count	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000
mean	77.002300	76.948200	77.017550	77.094500	76.958200	77.015550	77.069850	76.913100	76.921450	7.985950	4.995050	2.992800
std	10.085697	10.101733	10.134815	10.087837	10.020088	10.168888	10.069059	10.138555	10.103494	2.593798	2.578383	2.005700
min	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	4.000000	1.000000	0.000000
25%	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	6.000000	3.000000	1.000000
50%	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	8.000000	5.000000	3.000000
75%	86.000000	86.000000	86.000000	86.000000	85.000000	86.000000	86.000000	86.000000	86.000000	10.000000	7.000000	5.000000
max	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	12.000000	9.000000	6.000000

- info was called on the dataset

#	Column	Non-Null Count		Dtype
---	-----	-----	-----	-----
0	Acedamic percentage in Operating Systems	20000	non-null	int64
1	percentage in Algorithms	20000	non-null	int64
2	Percentage in Programming Concepts	20000	non-null	int64
3	Percentage in Software Engineering	20000	non-null	int64
4	Percentage in Computer Networks	20000	non-null	int64
5	Percentage in Electronics Subjects	20000	non-null	int64
6	Percentage in Computer Architecture	20000	non-null	int64
7	Percentage in Mathematics	20000	non-null	int64
8	Percentage in Communication skills	20000	non-null	int64
9	Hours working per day	20000	non-null	int64
10	Logical quotient rating	20000	non-null	int64
11	hackathons	20000	non-null	int64
12	coding skills rating	20000	non-null	int64
13	public speaking points	20000	non-null	int64
14	can work long time before system?	20000	non-null	object
15	self-learning capability?	20000	non-null	object
16	Extra-courses did	20000	non-null	object
17	certifications	20000	non-null	object
18	workshops	20000	non-null	object
19	talenttests taken?	20000	non-null	object
20	olympiads	20000	non-null	object
21	reading and writing skills	20000	non-null	object
22	memory capability score	20000	non-null	object
23	Interested subjects	20000	non-null	object
24	interested career area	20000	non-null	object
25	Job/Higher Studies?	20000	non-null	object
26	Type of company want to settle in?	20000	non-null	object
27	Taken inputs from seniors or elders	20000	non-null	object
28	interested in games	20000	non-null	object
29	Interested Type of Books	20000	non-null	object
30	Salary Range Expected	20000	non-null	object
31	In a Realtionship?	20000	non-null	object
32	Gentle or Tuff behaviour?	20000	non-null	object
33	Management or Technical	20000	non-null	object
34	Salary/work	20000	non-null	object

- Unique job roles were seen in the data

```
1 df['Suggested Job Role'].unique()
```

```
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)
```

- The bucketization of the job roles was done as follower:

```
finalDf = df.replace(to_replace = ['Database Manager', 'Project  
Manager', 'Information Technology Manager'], value = "Manager")  
finalDf = finalDf.replace(to_replace = ['Database  
Administrator', 'Network Security Administrator', 'Portal  
Administrator', 'Systems Security Administrator'], value =  
"Administrator")  
finalDf = finalDf.replace(to_replace = ['E-Commerce Analyst',  
'Systems Analyst', 'Programmer Analyst', 'Business Systems  
Analyst', 'Business Intelligence Analyst', 'Information Security  
Analyst', 'CRM Business Analyst'], value = "Analyst")  
finalDf = finalDf.replace(to_replace = ['Network Engineer',  
'Software Engineer', 'Technical Engineer', 'Network Security  
Engineer', 'Software Developer', 'Software Systems Engineer'],  
value = "Engineer")  
finalDf = finalDf.replace(to_replace = ['Applications  
Developer', 'CRM Technical Developer', 'Mobile Applications  
Developer', 'Web Developer', 'Database Developer'], value =  
"Developer")  
finalDf = finalDf.replace(to_replace = ["UX Designer", "Design &  
UX"], value = "Designer")  
finalDf = finalDf.replace(to_replace = ['Software Quality  
Assurance (QA) / Testing', 'Information Technology Auditor',  
'Technical Services/Help Desk/Tech Support', 'Data Architect',
```

```
'Solutions Architect', "Quality Assurance Associate",
"Technical Support"], value =
"Support/QA/Testing/Auditor/Architect")
```

- Finally the data was split into X and y

```
X = finalDf.drop(['Suggested Job Role'], axis=1)
y = finalDf['Suggested Job Role']
```

- Categorical columns and non-categorical columns were found
- the categorical columns were divided into one hot encoded vectors
- for the non-categorical columns, there were divided into high, medium and low values depending upon the range of the values that particular column took.
- The final rows were sampled and studied.

```
1 x.head()
```

	Academic percentage in Operating Systems	percentage in Algorithms	Percentage in Programming Concepts	Percentage in Software Engineering	Percentage in Computer Networks	Percentage in Electronics Subjects	Percentage in Computer Architecture	Percentage in Mathematics	Percentage in Communication skills	Hours working per day	...	Management or Technical_Management
0	1	1	2	3	3	3	3	3	1	2	...	1
1	2	1	2	1	1	1	2	3	3	3	...	0
2	1	3	3	3	1	2	2	2	3	3	...	1
3	2	3	1	3	3	2	1	3	1	2	...	1
4	3	1	3	1	1	3	2	1	2	1	...	1

5 rows x 126 columns

Experiments Performed

- About 20 experiments were performed with varying train test split sizes with test size ranging from 10% of the data to 90% of the data.
- MLP Classifier with 2 hidden layer of sizes 100 and 15 were used.
- Adam optimizer was used with Adaptive learning rate.
- Early stopping was implemented with n_iter_no_change set to 10.
- Max iters were set at 300.
- First this was done without shuffling the data, then with shuffling the data.
- The data was also stratified over the y column.
- Classification Report, Confusion matrices and Classwise accuracies were printed.

Results

Without Randomizing the data

Train Test Split of: 90%-10%

	precision	recall	f1-score	support
Administrator	0.14	0.09	0.11	286
Analyst	0.19	0.23	0.21	387
Designer	0.00	0.00	0.00	118
Developer	0.06	0.00	0.01	281
Engineer	0.19	0.13	0.15	356
Manager	0.00	0.00	0.00	176
Support/QA/Testing/Auditor/Architect	0.20	0.54	0.29	396
accuracy		0.19	2000	
macro avg	0.11	0.14	0.11	2000
weighted avg	0.14	0.19	0.14	2000

Accuracy of class Administrator is : 9.44055944055944%

Accuracy of class Analyst is : 22.739018087855296%

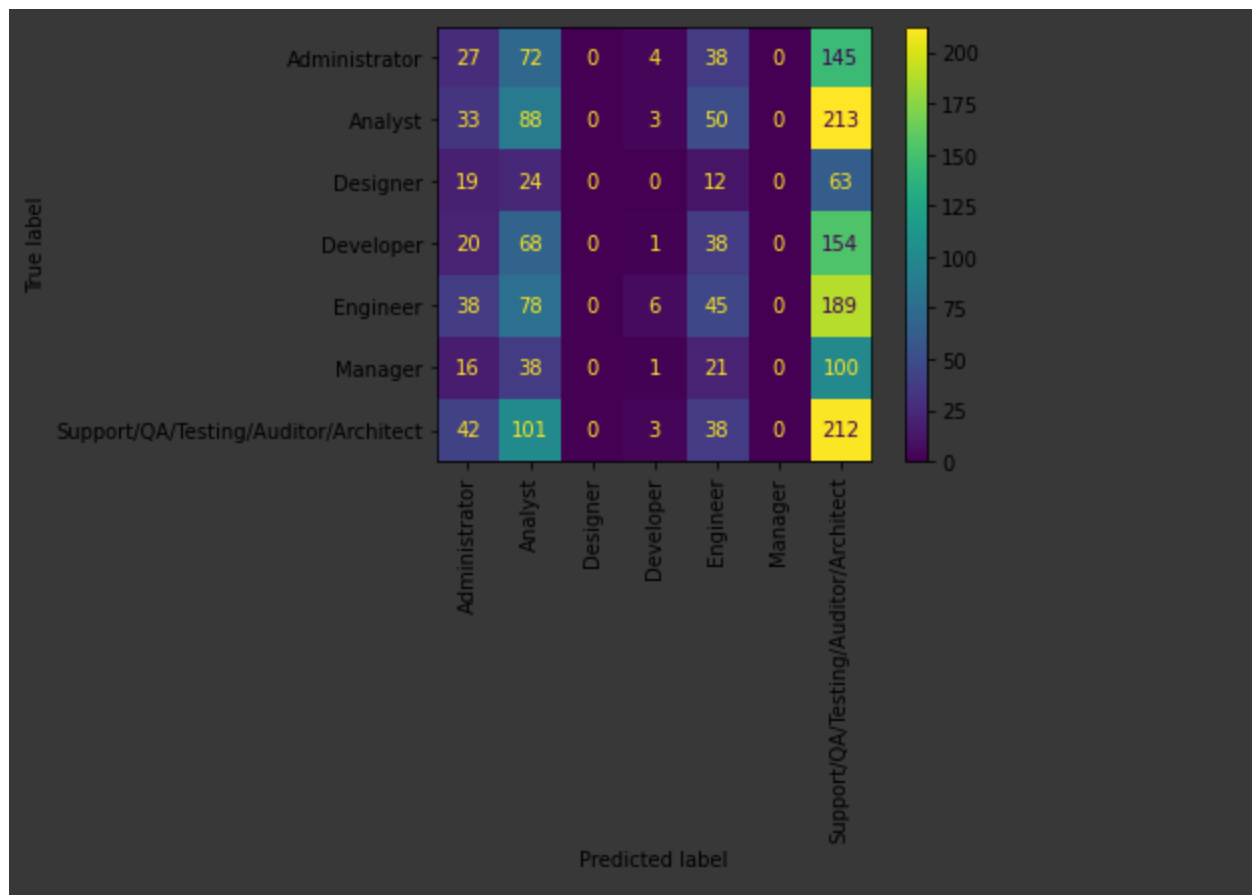
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.3558718861209964%

Accuracy of class Engineer is : 12.640449438202248%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 53.535353535353536%



Train Test Split of: 80%-20%

precision recall f1-score support

Administrator 0.15 0.02 0.04 572

Analyst 0.19 0.35 0.25 775

Designer 0.00 0.00 0.00 235

Developer 0.06 0.00 0.01 561

Engineer 0.16 0.22 0.18 712

Manager 0.00 0.00 0.00 353

Support/QA/Testing/Auditor/Architect 0.19 0.37 0.25 792

accuracy 0.18 4000

macro avg 0.11 0.14 0.11 4000

weighted avg 0.13 0.18 0.14 4000

Accuracy of class Administrator is : 2.4475524475524475%

Accuracy of class Analyst is : 34.58064516129033%

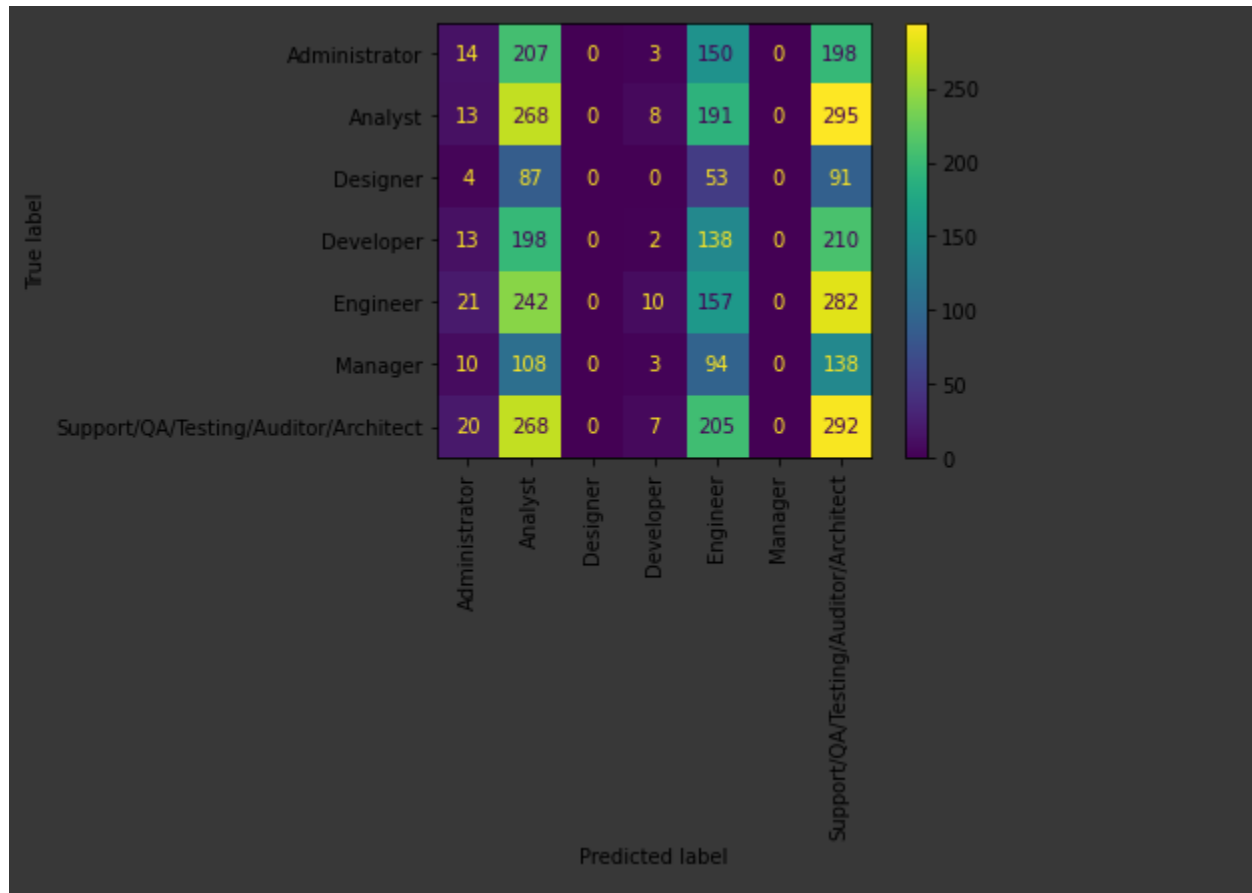
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.35650623885918004%

Accuracy of class Engineer is : 22.05056179775281%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 36.868686868686865%



Train Test Split of: 70%-30%

precision recall f1-score support

Administrator 0.14 0.18 0.16 858

Analyst 0.20 0.18 0.19 1162

Designer 0.25 0.00 0.01 353

Developer 0.16 0.10 0.12 842

Engineer 0.17 0.24 0.20 1068

Manager 0.11 0.00 0.01 529

Support/QA/Testing/Auditor/Architect 0.21 0.33 0.25 1188

accuracy 0.18 6000

macro avg 0.18 0.15 0.13 6000

weighted avg 0.18 0.18 0.16 6000

Accuracy of class Administrator is : 17.715617715617714%

Accuracy of class Analyst is : 17.814113597246127%

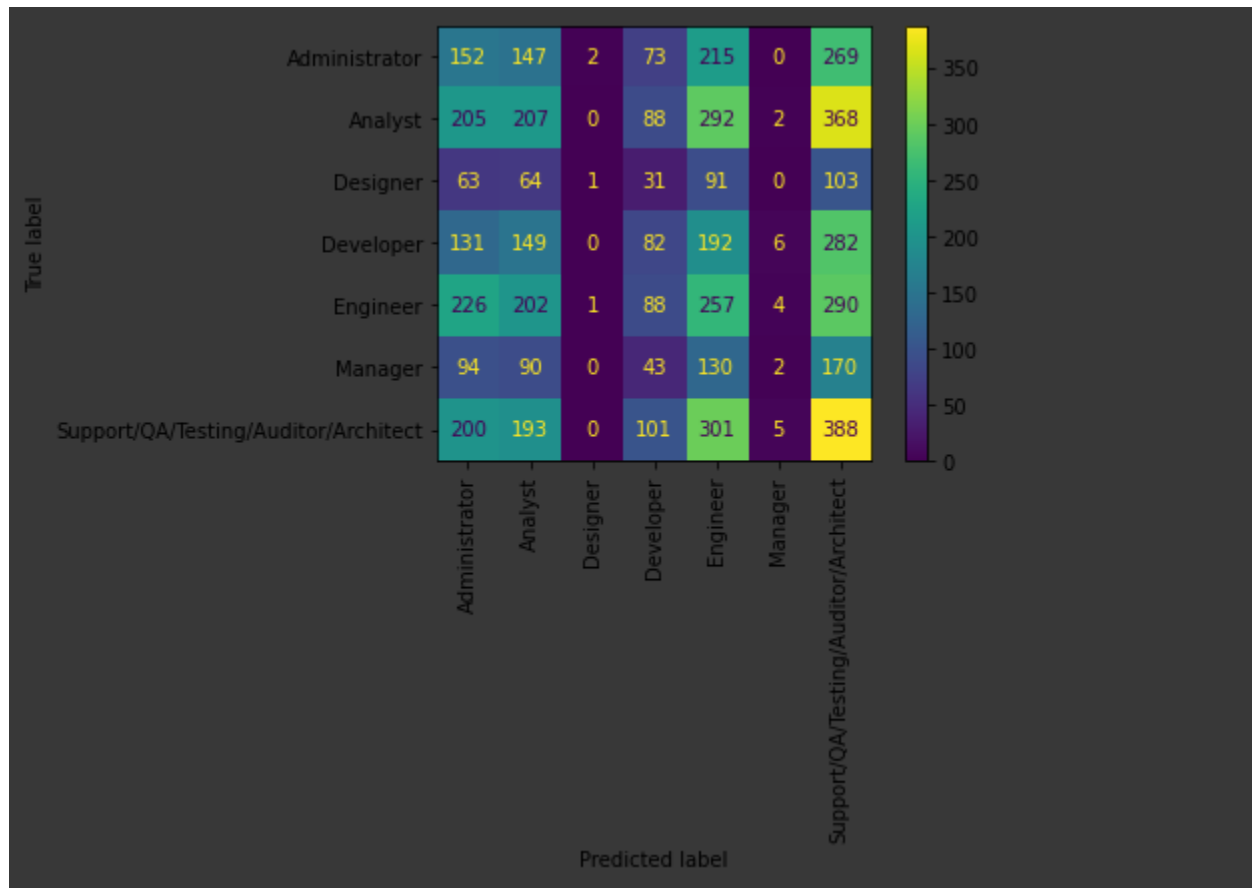
Accuracy of class Designer is : 0.28328611898017%

Accuracy of class Developer is : 9.73871733966746%

Accuracy of class Engineer is : 24.06367041198502%

Accuracy of class Manager is : 0.3780718336483932%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 32.659932659932664%



Train Test Split of: 60%-40%

precision recall f1-score support

Administrator	0.18	0.02	0.03	1144
Analyst	0.20	0.47	0.28	1549
Designer	0.00	0.00	0.00	471
Developer	0.12	0.00	0.01	1123
Engineer	0.18	0.16	0.17	1424
Manager	0.00	0.00	0.00	705

Support/QA/Testing/Auditor/Architect	0.20	0.36	0.25	1584
--------------------------------------	------	------	------	------

accuracy	0.19	8000		
macro avg	0.12	0.14	0.11	8000

weighted avg 0.15 0.19 0.14 8000

Accuracy of class Administrator is : 1.8356643356643356%

Accuracy of class Analyst is : 46.804389928986446%

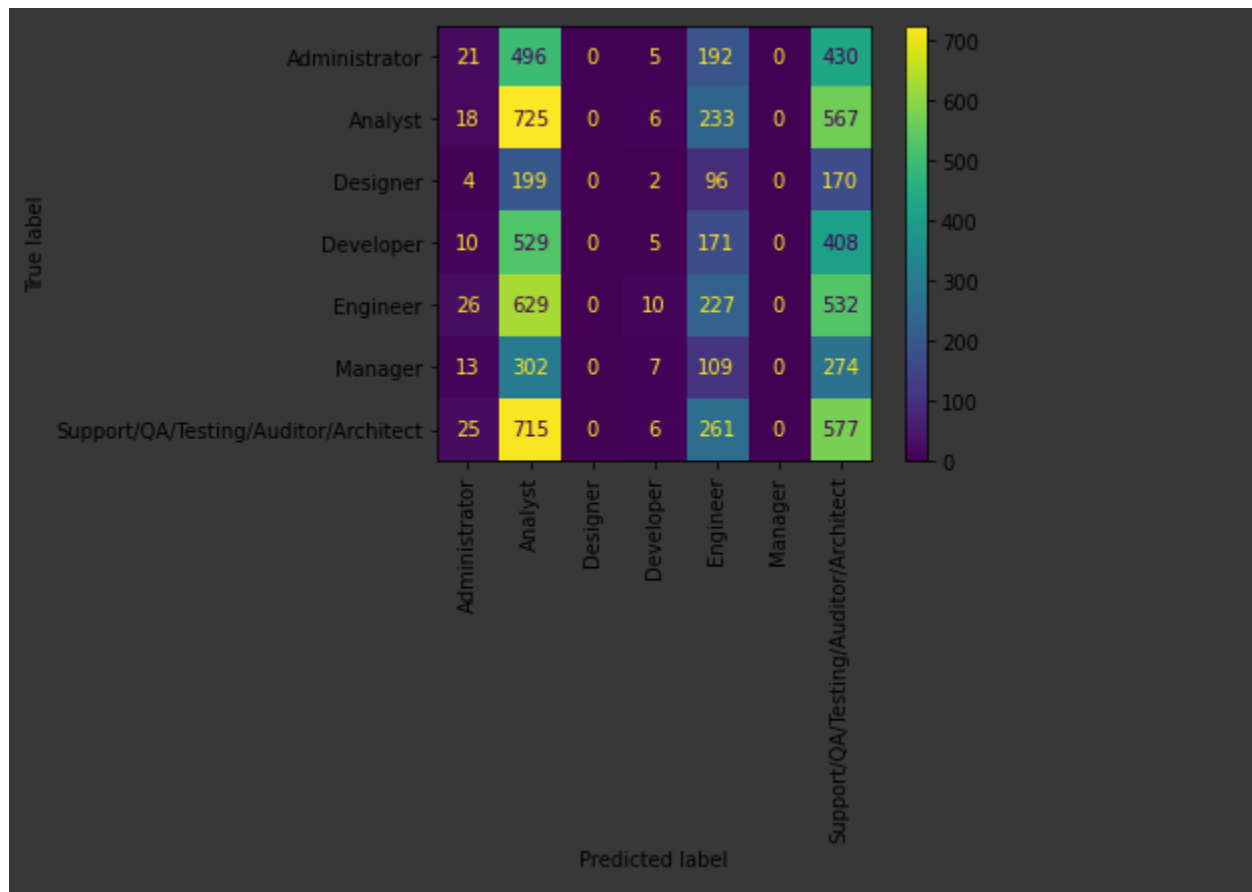
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.44523597506678536%

Accuracy of class Engineer is : 15.941011235955054%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 36.426767676767675%



Train Test Split of: 50%-50%

precision recall f1-score support

Administrator 0.12 0.01 0.02 1430

Analyst 0.19 0.27 0.22 1937

Designer 0.00 0.00 0.00 589

Developer 0.14 0.01 0.03 1403

Engineer 0.17 0.29 0.22 1780

Manager 0.00 0.00 0.00 881

Support/QA/Testing/Auditor/Architect 0.20 0.40 0.27 1980

accuracy		0.19	10000
macro avg	0.12	0.14	0.11 10000
weighted avg	0.14	0.19	0.14 10000

Accuracy of class Administrator is : 1.3286713286713288%

Accuracy of class Analyst is : 27.20702116675271%

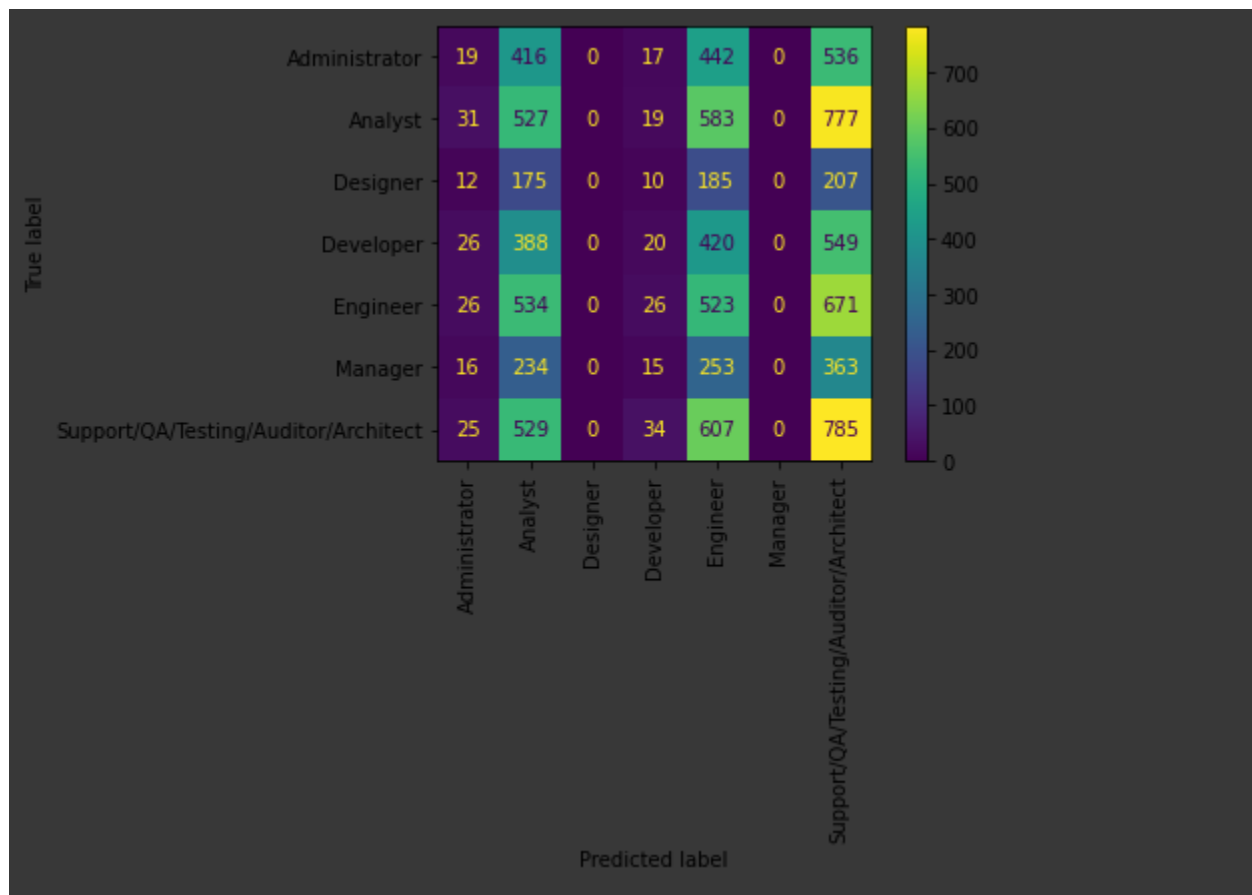
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 1.4255167498218104%

Accuracy of class Engineer is : 29.382022471910112%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 39.64646464646464%



Train Test Split of: 40%-60%

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

Administrator	0.13	0.01	0.02	1716
Analyst	0.19	0.38	0.26	2324
Designer	0.00	0.00	0.00	706
Developer	0.12	0.00	0.00	1684

Engineer	0.18	0.40	0.25	2136
Manager	1.00	0.00	0.00	1058
Support/QA/Testing/Auditor/Architect	0.19	0.20	0.20	2376

accuracy	0.19	12000		
macro avg	0.26	0.14	0.10	12000
weighted avg	0.23	0.19	0.14	12000

Accuracy of class Administrator is : 0.9324009324009324%

Accuracy of class Analyst is : 38.38209982788296%

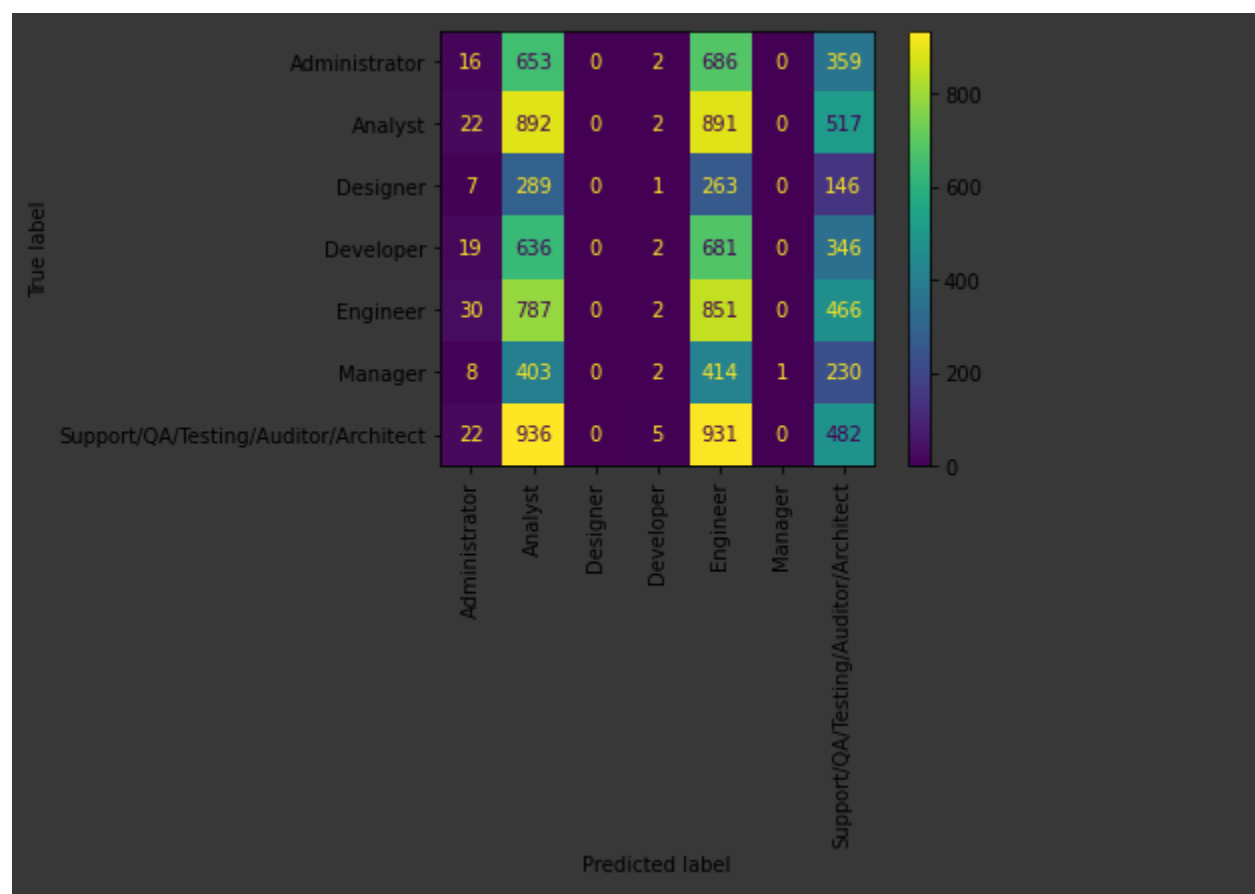
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.11876484560570072%

Accuracy of class Engineer is : 39.840823970037455%

Accuracy of class Manager is : 0.0945179584120983%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 20.286195286195287%



Train Test Split of: 30%-70%

precision recall f1-score support

Administrator	0.15	0.04	0.06	2002
---------------	------	------	------	------

Analyst	0.19	0.40	0.26	2712
Designer	0.00	0.00	0.00	824
Developer	0.10	0.01	0.02	1965
Engineer	0.20	0.14	0.17	2492
Manager	0.00	0.00	0.00	1234
Support/QA/Testing/Auditor/Architect	0.20	0.41	0.27	2771

accuracy		0.19	14000
macro avg	0.12	0.14	0.11
weighted avg	0.15	0.19	0.14

Accuracy of class Administrator is : 3.5464535464535465%

Accuracy of class Analyst is : 40.15486725663717%

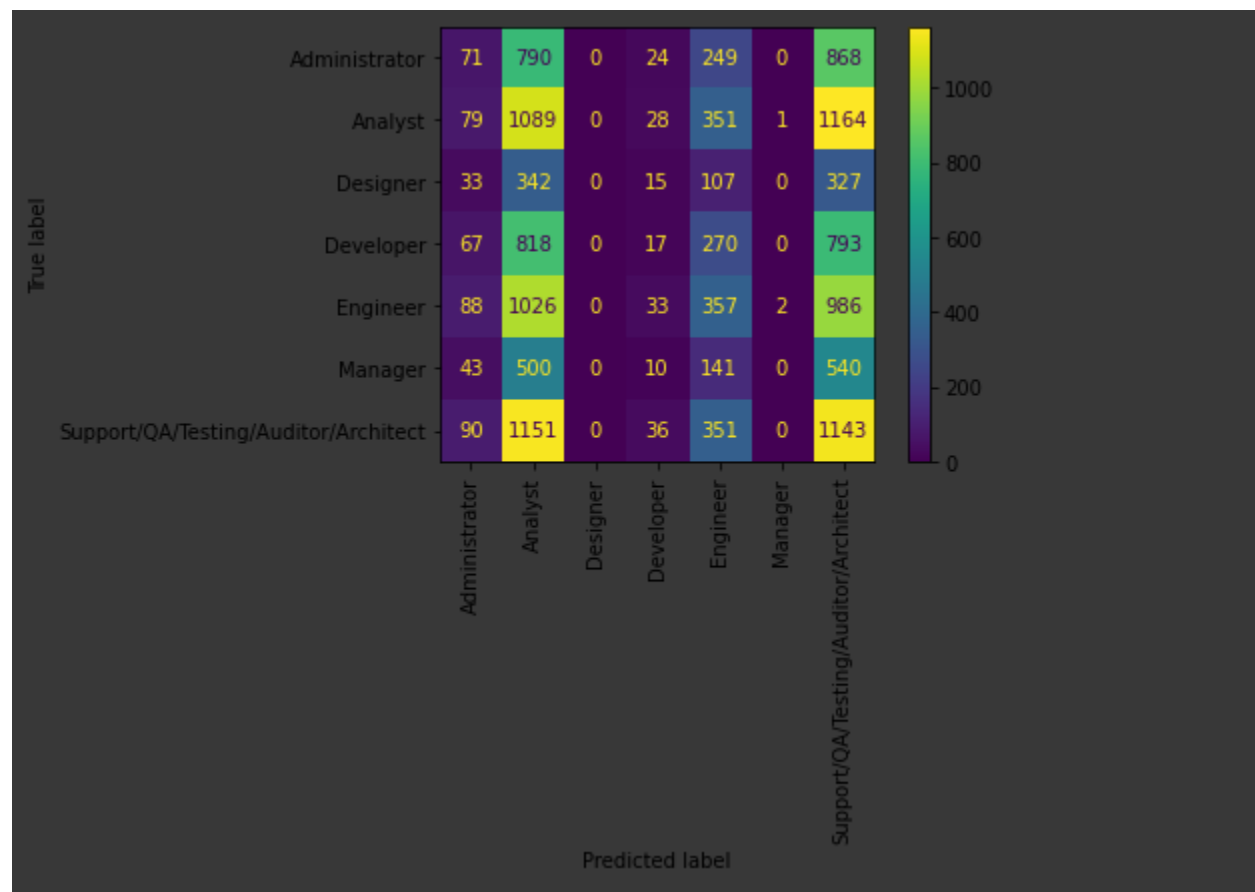
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.8651399491094147%

Accuracy of class Engineer is : 14.325842696629213%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 41.24864669794298%



Train Test Split of: 20%-80%

	precision	recall	f1-score	support
Administrator	0.15	0.09	0.11	2288
Analyst	0.18	0.13	0.15	3099
Designer	0.00	0.00	0.00	942
Developer	0.14	0.01	0.01	2246
Engineer	0.19	0.13	0.15	2848
Manager	0.00	0.00	0.00	1410
Support/QA/Testing/Auditor/Architect	0.20	0.65	0.30	3167
accuracy		0.19		16000
macro avg	0.12	0.14	0.10	16000
weighted avg	0.15	0.19	0.13	16000

Accuracy of class Administrator is : 9.090909090909092%

Accuracy of class Analyst is : 12.584704743465633%

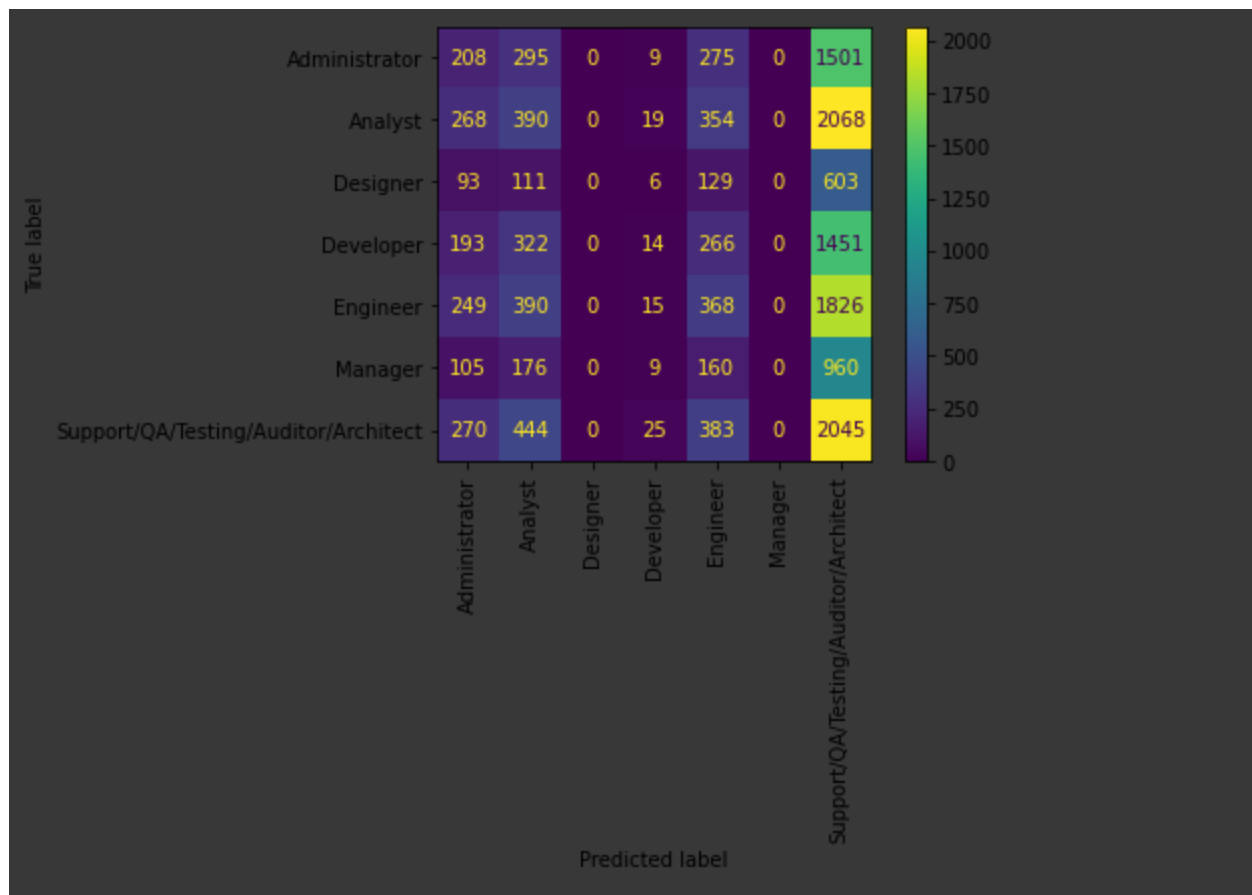
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.6233303650934996%

Accuracy of class Engineer is : 12.921348314606742%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 64.57215029996843%



Train Test Split of: 10%-90%

precision recall f1-score support

Administrator 0.15 0.11 0.12 2574

Analyst 0.19 0.31 0.24 3487

Designer 0.00 0.00 0.00 1059

Developer 0.15 0.02 0.04 2526

Engineer 0.18 0.29 0.22 3204

Manager 0.00 0.00 0.00 1587

Support/QA/Testing/Auditor/Architect 0.20 0.28 0.23 3563

accuracy 0.19 18000

macro avg 0.12 0.14 0.12 18000

weighted avg 0.15 0.19 0.15 18000

Accuracy of class Administrator is : 10.606060606060606%

Accuracy of class Analyst is : 30.97218239174075%

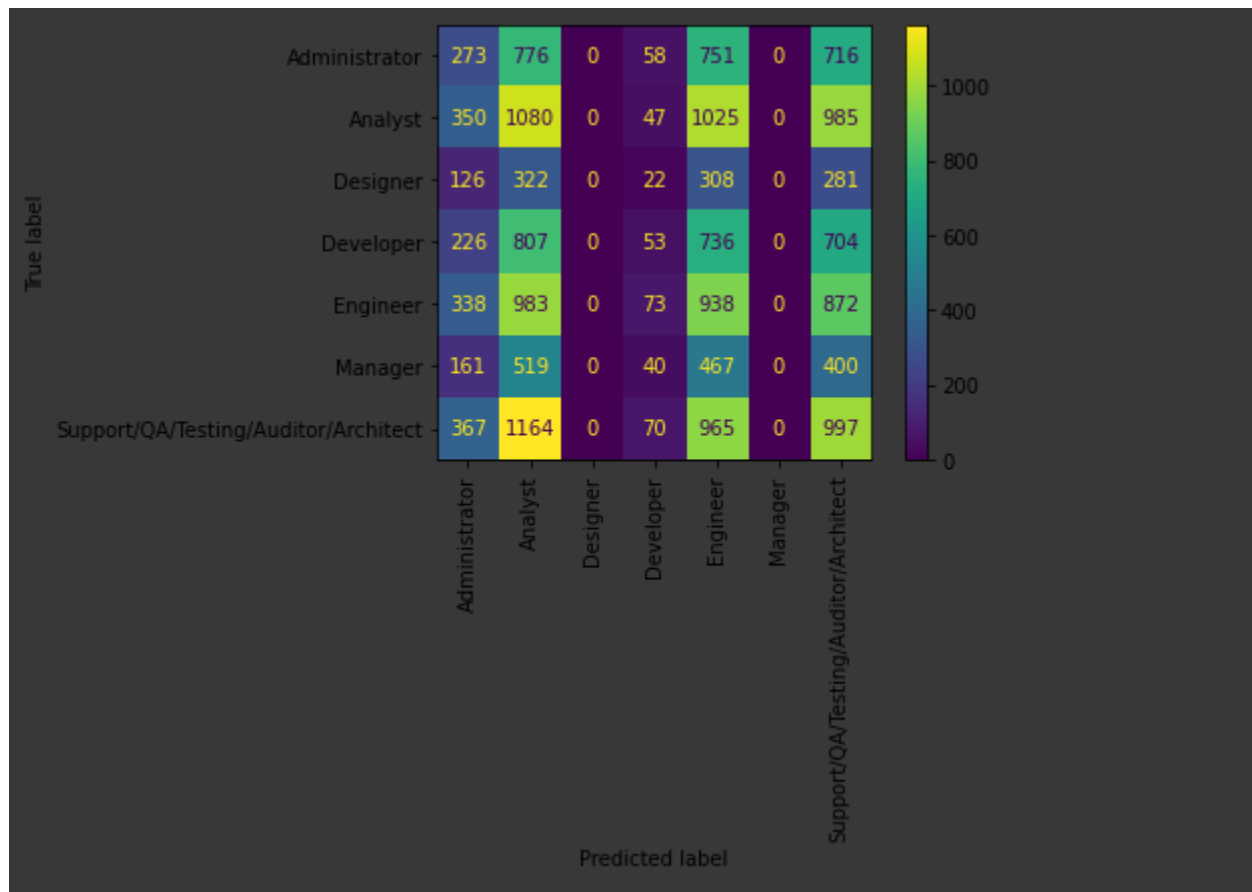
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 2.098178939034046%

Accuracy of class Engineer is : 29.27590511860175%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 27.982037608756666%



Randomizing the data

Train Test Split of: 90%-10%

	precision	recall	f1-score	support
Administrator	0.23	0.04	0.07	286
Analyst	0.21	0.20	0.21	387
Designer	0.00	0.00	0.00	118
Developer	0.29	0.01	0.01	281
Engineer	0.18	0.39	0.24	356
Manager	0.00	0.00	0.00	176
Support/QA/Testing/Auditor/Architect	0.21	0.42	0.28	396
accuracy		0.20		2000
macro avg	0.16	0.15	0.12	2000
weighted avg	0.19	0.20	0.15	2000

Accuracy of class Administrator is : 3.8461538461538463%

Accuracy of class Analyst is : 20.41343669250646%

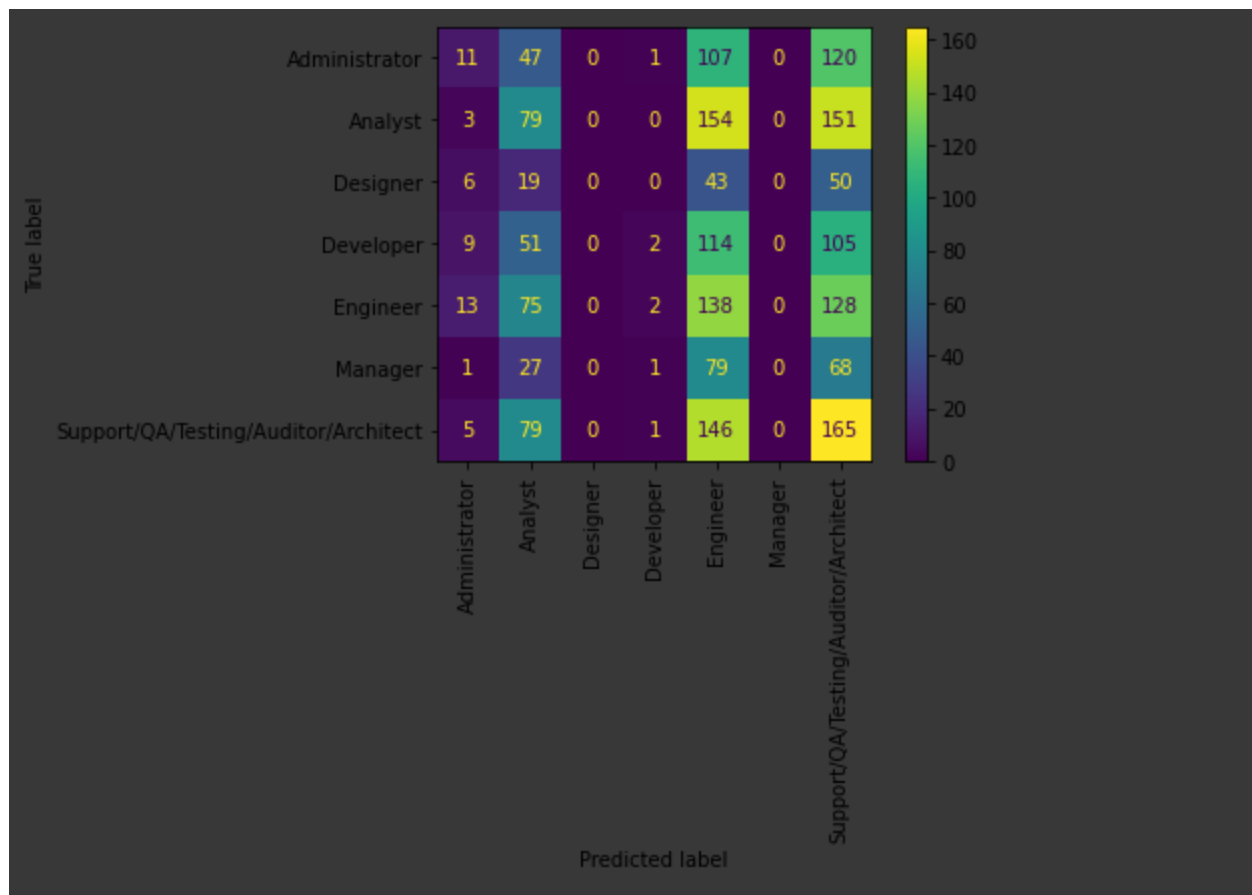
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.7117437722419928%

Accuracy of class Engineer is : 38.764044943820224%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 41.66666666666667%



Train Test Split of: 80%-20%

precision recall f1-score support

Administrator 0.11 0.02 0.04 572

Analyst 0.19 0.21 0.20 775

Designer 0.00 0.00 0.00 235

Developer 0.15 0.05 0.07 561

Engineer 0.17 0.29 0.21 712

Manager 0.00 0.00 0.00 353

Support/QA/Testing/Auditor/Architect 0.21 0.45 0.29 792

accuracy 0.19 4000

macro avg 0.12 0.14 0.12 4000

weighted avg 0.15 0.19 0.15 4000

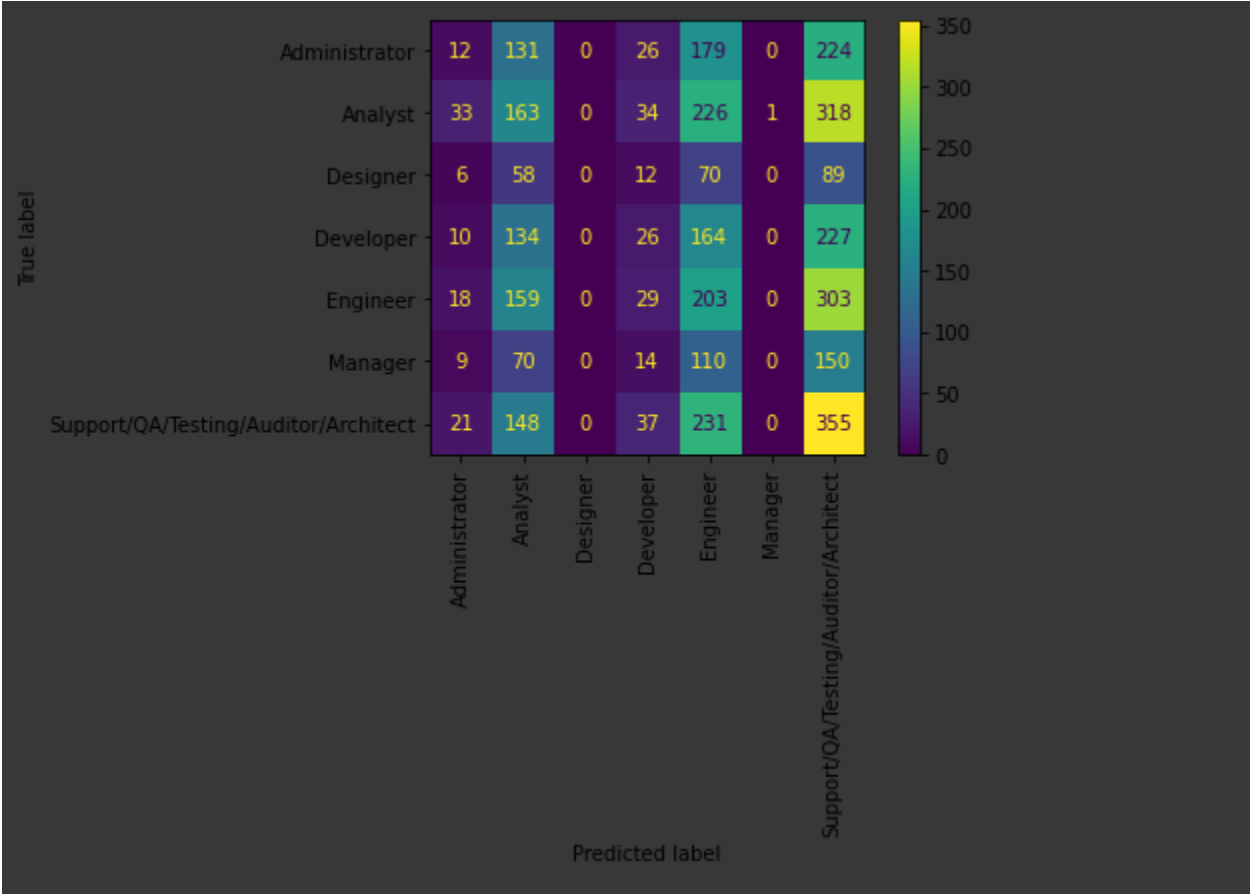
Accuracy of class Administrator is : 2.097902097902098%

Accuracy of class Analyst is : 21.032258064516128%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 4.634581105169341%

Accuracy of class Engineer is : 28.51123595505618%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is : 44.8232323232325%



Train Test Split of: 70%-30%

	precision	recall	f1-score	support		
Administrator	0.12	0.03	0.04	858		
Analyst	0.20	0.26	0.23	1162		
Designer	0.00	0.00	0.00	353		
Developer	0.13	0.01	0.03	842		
Engineer	0.18	0.22	0.20	1068		
Manager	0.00	0.00	0.00	529		
Support/QA/Testing/Auditor/Architect			0.20	0.48	0.28	1188
accuracy			0.19	6000		
macro avg	0.12	0.14	0.11	6000		
weighted avg	0.14	0.19	0.14	6000		

Accuracy of class Administrator is : 2.564102564102564%

Accuracy of class Analyst is : 26.41996557659208%

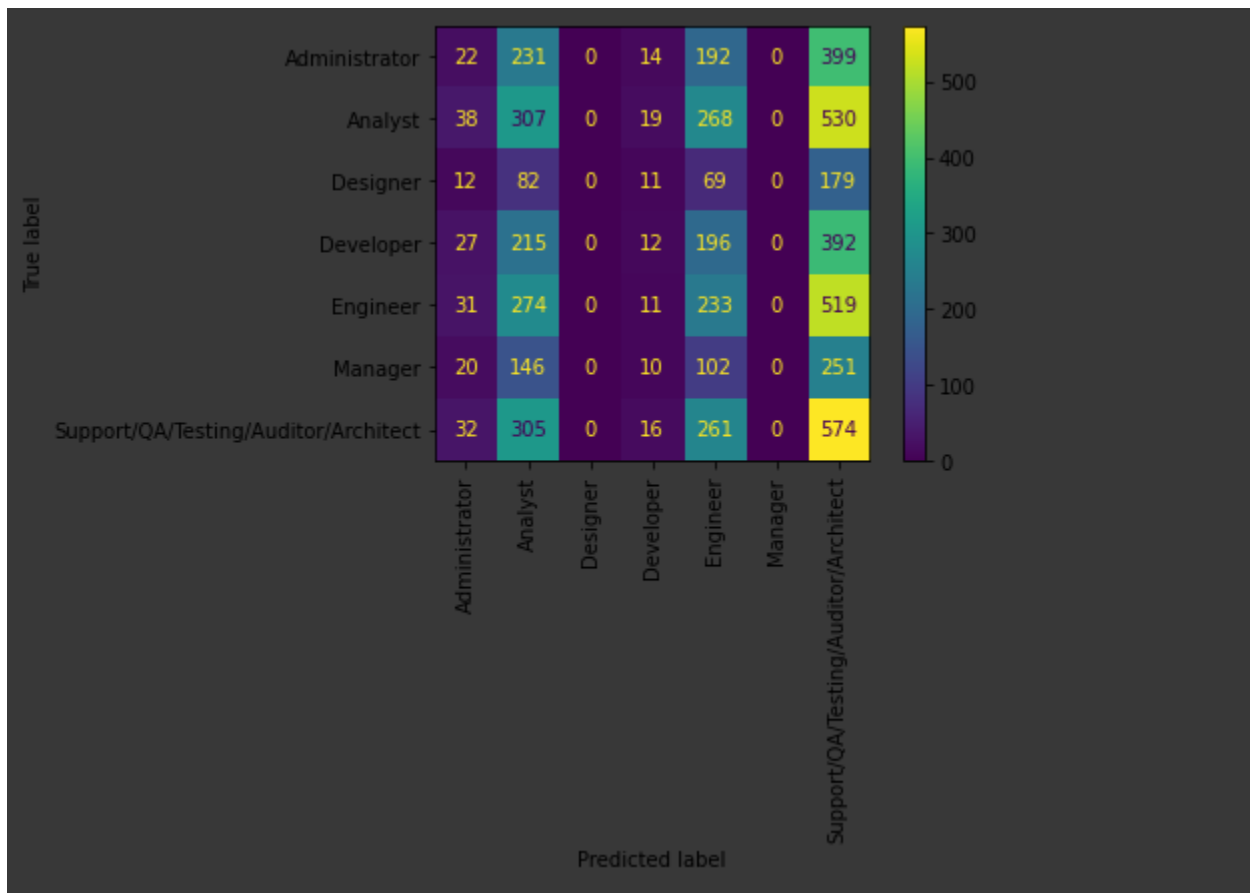
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 1.4251781472684086%

Accuracy of class Engineer is : 21.816479400749063%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 48.31649831649832%



Train Test Split of: 60%-40%

precision recall f1-score support

Administrator	0.13	0.02	0.04	1144
Analyst	0.19	0.28	0.23	1549
Designer	0.00	0.00	0.00	471
Developer	0.07	0.00	0.01	1123
Engineer	0.17	0.21	0.19	1424
Manager	0.00	0.00	0.00	705

Support/QA/Testing/Auditor/Architect	0.20	0.47	0.28	1584
--------------------------------------	------	------	------	------

accuracy	0.19	8000
----------	------	------

macro avg	0.11	0.14	0.11	8000
weighted avg	0.14	0.19	0.14	8000

Accuracy of class Administrator is : 2.36013986013986%

Accuracy of class Analyst is : 28.276307295029053%

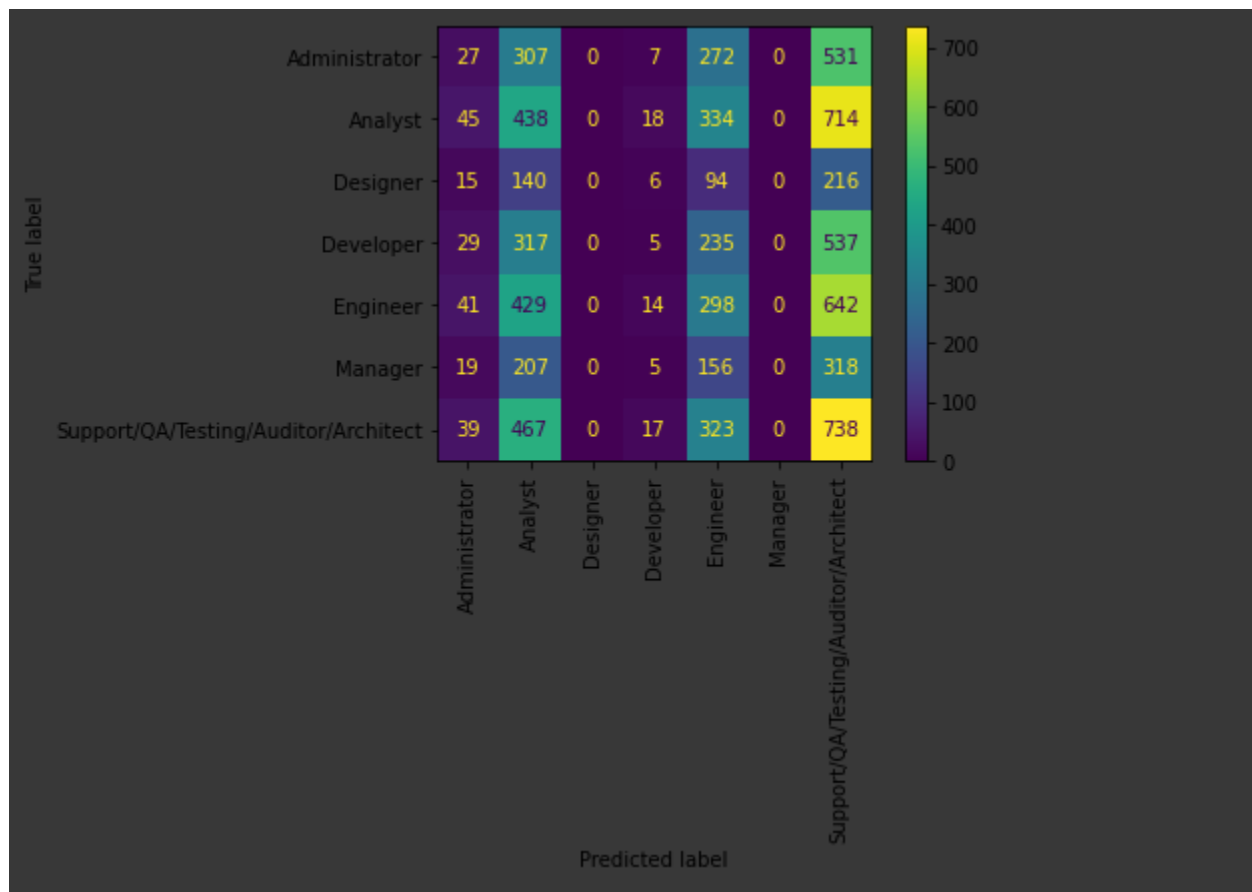
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.44523597506678536%

Accuracy of class Engineer is : 20.926966292134832%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 46.590909090909086%



Train Test Split of: 50%-50%

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

Administrator	0.12	0.02	0.03	1430
Analyst	0.20	0.08	0.11	1937
Designer	0.00	0.00	0.00	588
Developer	0.12	0.02	0.04	1404
Engineer	0.17	0.22	0.19	1780
Manager	0.00	0.00	0.00	881

Support/QA/Testing/Auditor/Architect	0.20	0.65	0.30	1980
--------------------------------------	------	------	------	------

accuracy		0.19	10000
macro avg	0.12	0.14	0.10
weighted avg	0.14	0.19	0.13

Accuracy of class Administrator is : 1.5384615384615385%

Accuracy of class Analyst is : 8.053691275167784%

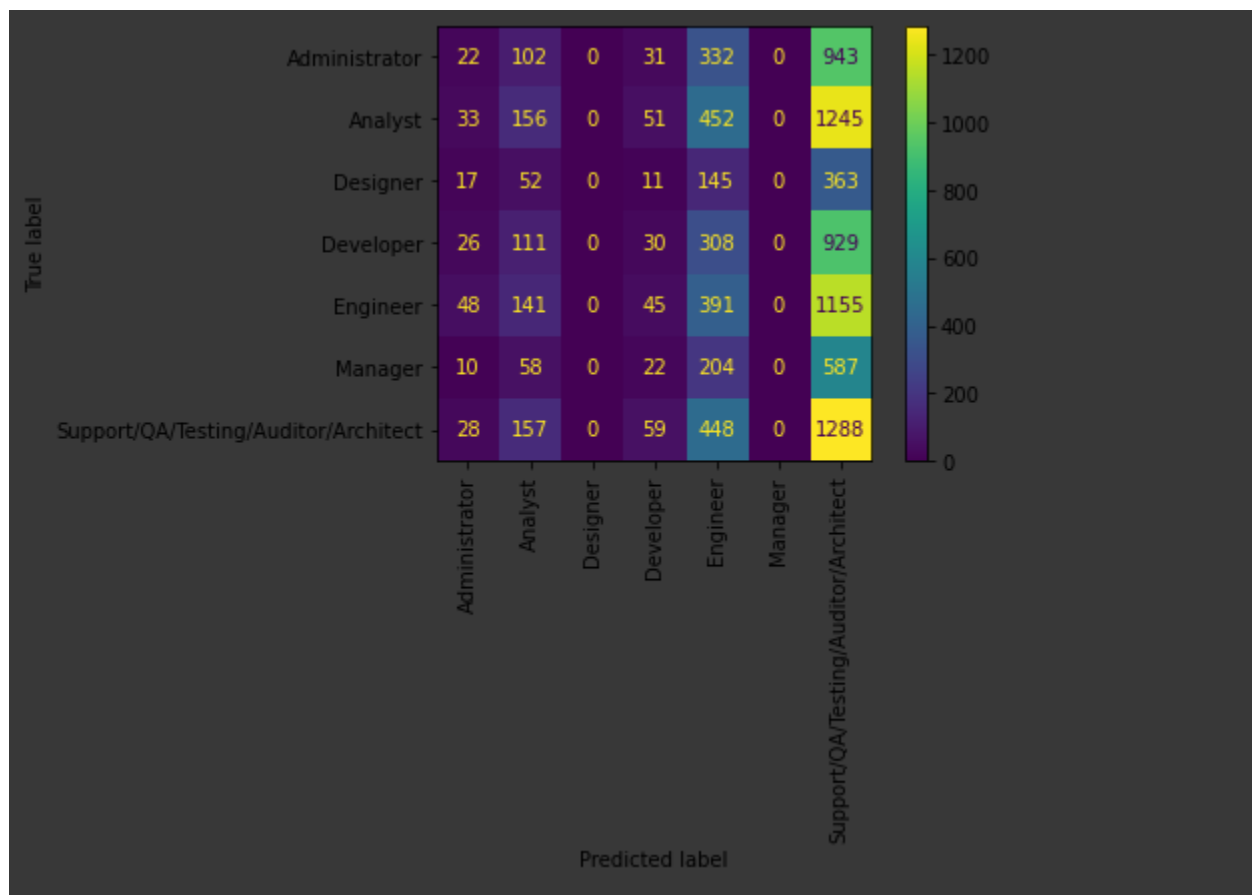
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 2.1367521367521367%

Accuracy of class Engineer is : 21.966292134831463%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 65.05050505050505%



Train Test Split of: 40%-60%

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

Administrator	0.13	0.02	0.04	1716
Analyst	0.19	0.24	0.21	2324
Designer	0.00	0.00	0.00	706

Developer	0.13	0.01	0.03	1684
Engineer	0.19	0.16	0.17	2136
Manager	0.00	0.00	0.00	1058
Support/QA/Testing/Auditor/Architect	0.20	0.56	0.29	2376
accuracy		0.19		12000
macro avg	0.12	0.14	0.11	12000
weighted avg	0.15	0.19	0.14	12000

Accuracy of class Administrator is : 2.3892773892773893%

Accuracy of class Analyst is : 23.580034423407916%

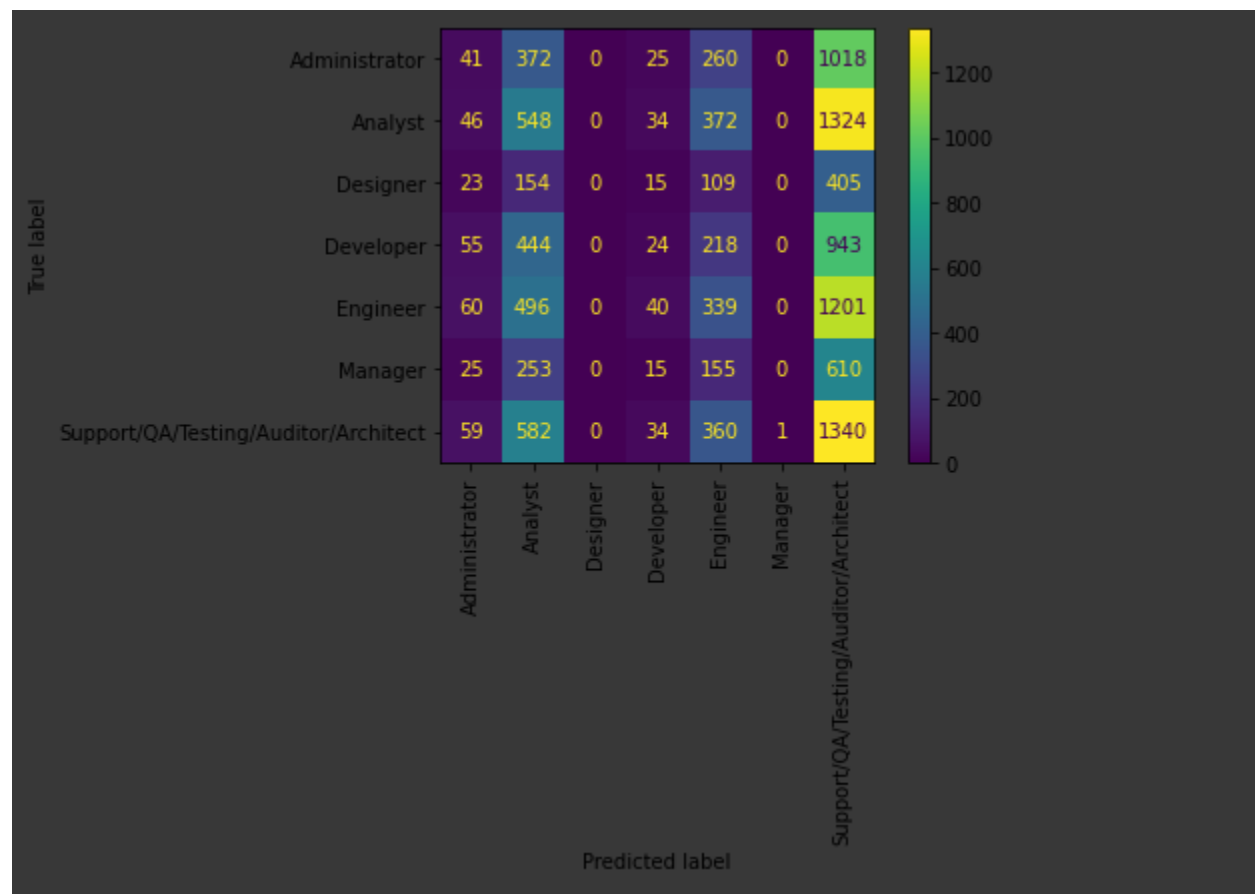
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 1.4251781472684086%

Accuracy of class Engineer is : 15.870786516853933%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 56.3973063973064%



Train Test Split of: 30%-70%

precision recall f1-score support

Administrator	0.15	0.03	0.05	2002
Analyst	0.19	0.29	0.23	2712
Designer	0.00	0.00	0.00	824
Developer	0.13	0.03	0.05	1965
Engineer	0.19	0.09	0.12	2492
Manager	0.00	0.00	0.00	1234
Support/QA/Testing/Auditor/Architect	0.20	0.57	0.30	2771
accuracy		0.19		14000
macro avg	0.12	0.14	0.11	14000
weighted avg	0.15	0.19	0.14	14000

Accuracy of class Administrator is : 2.647352647352647%

Accuracy of class Analyst is : 29.166666666666668%

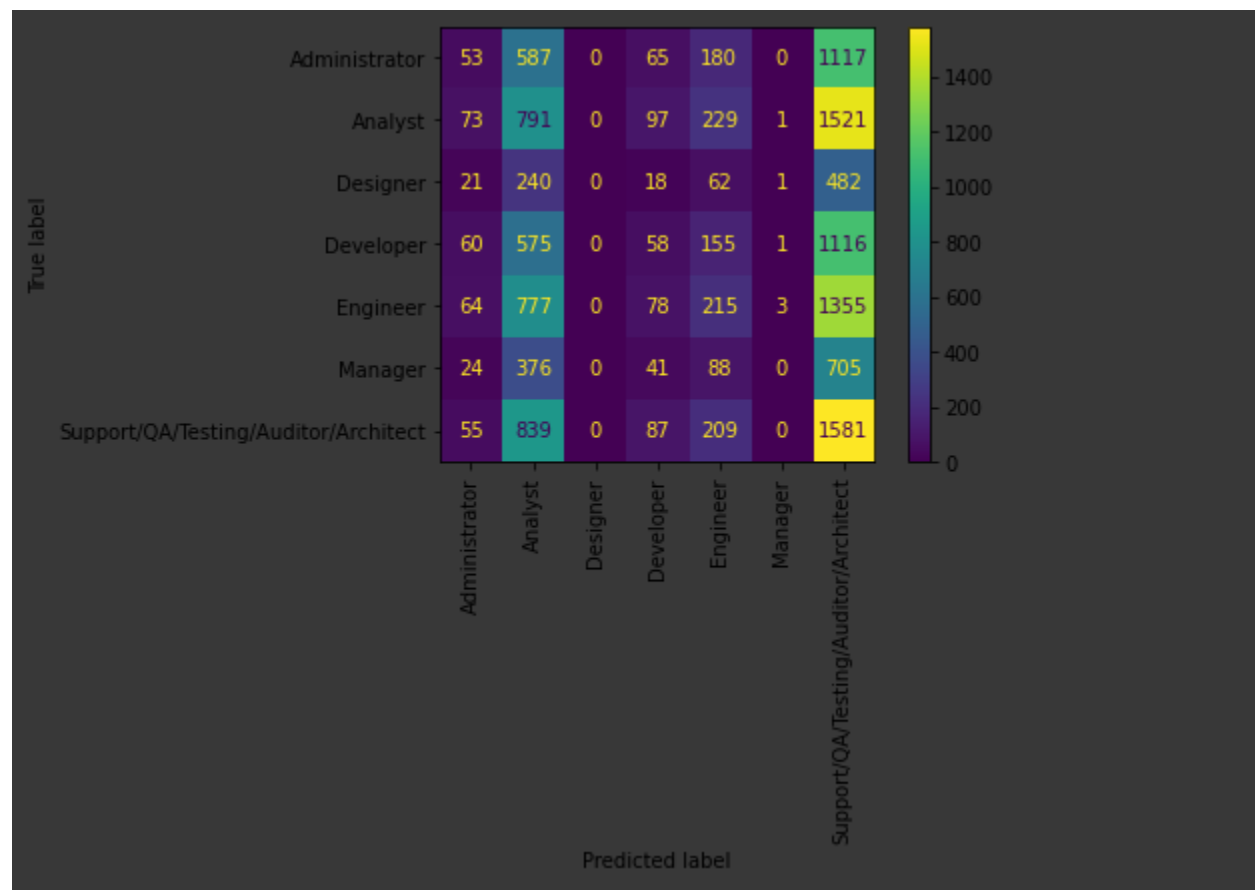
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 2.9516539440203564%

Accuracy of class Engineer is : 8.62760834670947%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 57.05521472392638%



Train Test Split of: 20%-80%

	precision	recall	f1-score	support
Administrator	0.14	0.02	0.04	2288
Analyst	0.19	0.24	0.21	3099
Designer	0.06	0.00	0.01	942
Developer	0.14	0.47	0.21	2246
Engineer	0.50	0.00	0.00	2848
Manager	0.25	0.00	0.00	1410
Support/QA/Testing/Auditor/Architect		0.20	0.26	0.23 3167
accuracy		0.17	16000	
macro avg	0.21	0.14	0.10	16000
weighted avg	0.23	0.17	0.12	16000

Accuracy of class Administrator is : 2.1853146853146854%

Accuracy of class Analyst is : 23.78186511777993%

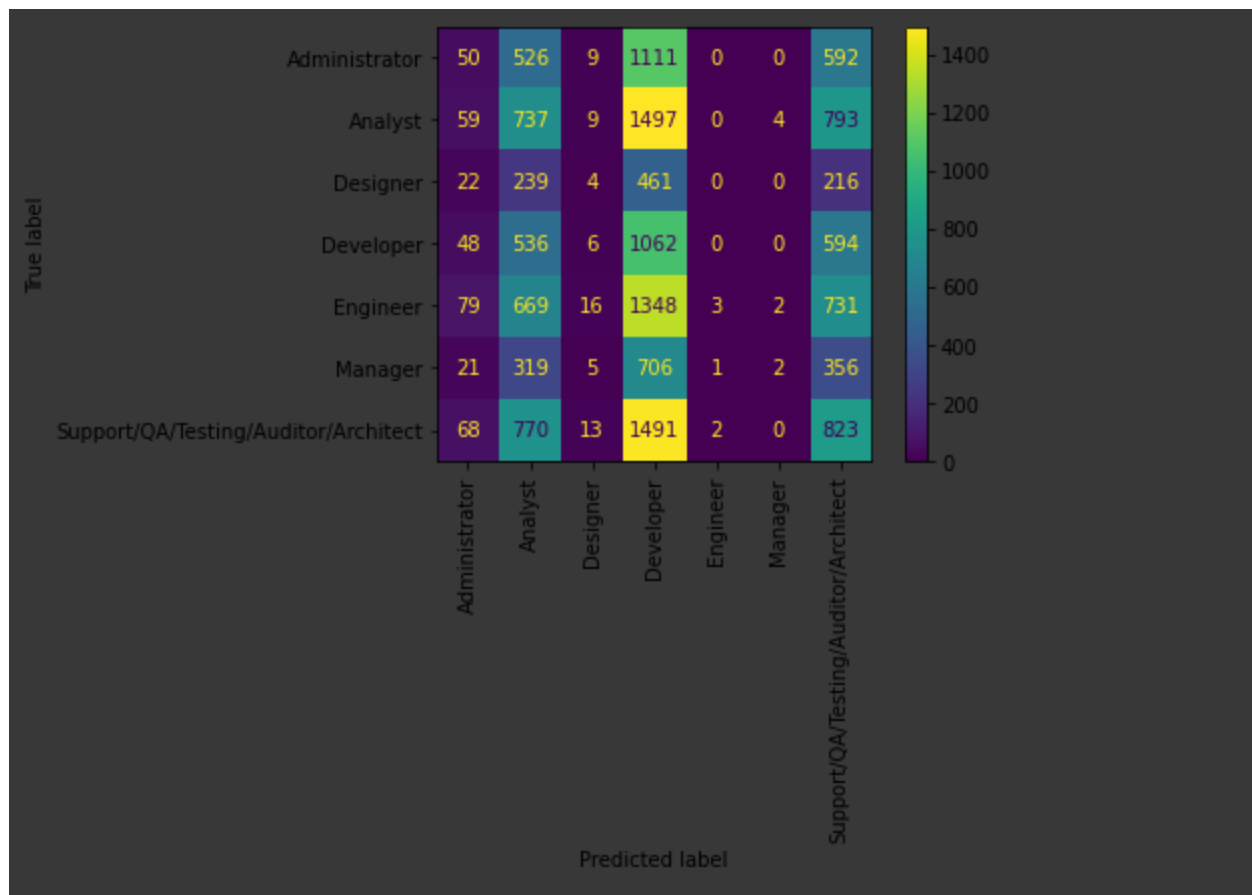
Accuracy of class Designer is : 0.42462845010615713%

Accuracy of class Developer is : 47.28406055209261%

Accuracy of class Engineer is : 0.10533707865168539%

Accuracy of class Manager is : 0.14184397163120568%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 25.986738238080203%



Train Test Split of: 10%-90%

precision recall f1-score support

Administrator 0.15 0.07 0.09 2574

Analyst 0.19 0.33 0.24 3487

Designer 0.00 0.00 0.00 1059

Developer 0.12 0.01 0.02 2526

Engineer 0.18 0.21 0.20 3204

Manager 0.08 0.00 0.00 1587

Support/QA/Testing/Auditor/Architect 0.20 0.39 0.26 3563

accuracy 0.19 18000

macro avg 0.13 0.14 0.12 18000

weighted avg 0.15 0.19 0.15 18000

Accuracy of class Administrator is : 6.915306915306915%

Accuracy of class Analyst is : 32.66418124462289%

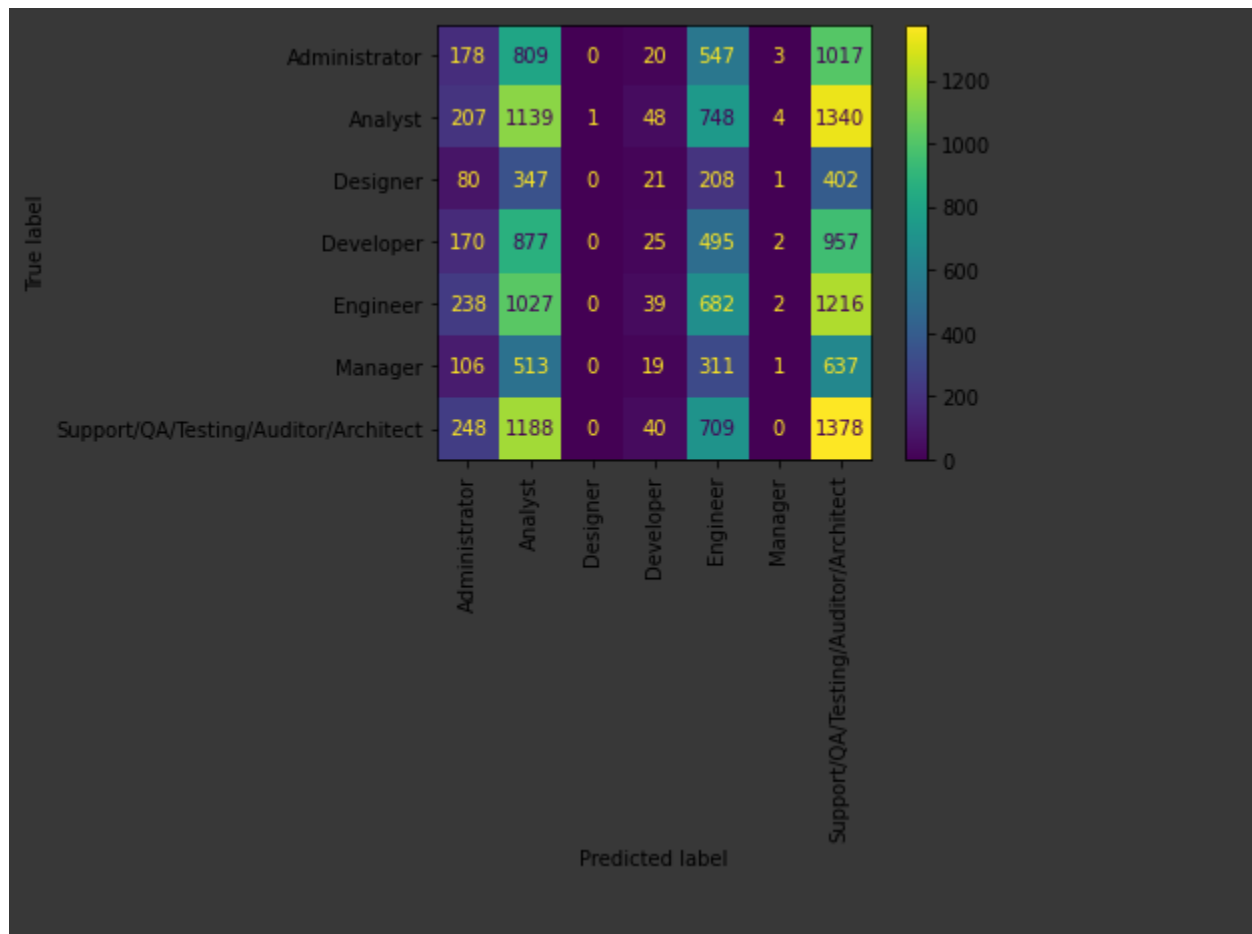
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.9897070467141725%

Accuracy of class Engineer is : 21.28589263420724%

Accuracy of class Manager is : 0.0630119722747322%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 38.6752736458041%



Analysis

- As it can be seen that the obtained results depend a lot on the initial classification of the different job classifications.
- As the classification is done by humans, it might not be always possible to classify it as it is intended to be.
- We can still see that this classification is also providing some accuracy which is better than the random classification.
- If stronger models are used, we can definitely improve this too.

Connecting it with Assignment 1

- One way to do it would be to connect the columns:
 - Academic percentage in Operating Systems
 - percentage in Algorithms
 - Percentage in Programming Concepts
 - Percentage in Software Engineering
 - Percentage in Computer Networks
 - Percentage in Electronics Subjects
 - Percentage in Computer Architecture
 - Percentage in Mathematics
 - Percentage in Communication skills
- to the columns:
 - Grade in Operating Systems
 - Grade in ADA
 - Grade in IP
 - Grade in DSA
 - Grade in Computer Networks
 - Grade in Basic Electronics
 - Grade in Computer Organization
 - Grade in Linear Algebra
 - Grade in Communication skills
- Respectively
- Another way can be to link the columns:
 - Academic percentage in Operating Systems
 - percentage in Algorithms
 - Percentage in Programming Concepts
 - Percentage in Software Engineering
 - Percentage in Computer Networks
 - Percentage in Electronics Subjects
 - Percentage in Computer Architecture
 - Percentage in Mathematics
 - Percentage in Communication skills
- to the columns:
 - Course done or not: Operating Systems
 - Course done or not: ADA
 - Course done or not: IP
 - Course done or not: DSA
 - Course done or not: Computer Networks
 - Course done or not: Basic Electronics
 - Course done or not: Computer Organization
 - Course done or not: Linear Algebra

- Course done or not: Communication skills
- This can be done by taking some threshold, that if the student has achieved $\geq x\%$ of grade then the student has taken the course, else not.
- These will suffice for electives and their grades.
- For interests, we can select the following columns:
 - Interested subjects
 - interested career area
 - interested in games
 - Interested Type of Books
- We can keep the data same, this will lead us to get the same results, or we can edit the data, by dropping the columns which are not required.
- This is not modification of data per se as we are just dropping the unnecessary columns for this purpose.
- Then the same steps are followed again, which were followed to generate the results.
- The results obtained are as follows:
 - Without Shuffling Data:

	precision	recall	f1-score	support
Administrator	0.09	0.00	0.01	286
Analyst	0.20	0.24	0.22	387
Designer	0.00	0.00	0.00	118
Developer	0.00	0.00	0.00	281
Engineer	0.20	0.19	0.20	356
Manager	0.00	0.00	0.00	176
Support/QA/Testing/Auditor/Architect	0.20	0.59	0.30	396
accuracy		0.20	2000	
macro avg	0.10	0.15	0.10	2000
weighted avg	0.13	0.20	0.14	2000

Accuracy of class Administrator is : 0.34965034965034963%

Accuracy of class Analyst is : 24.289405684754524%

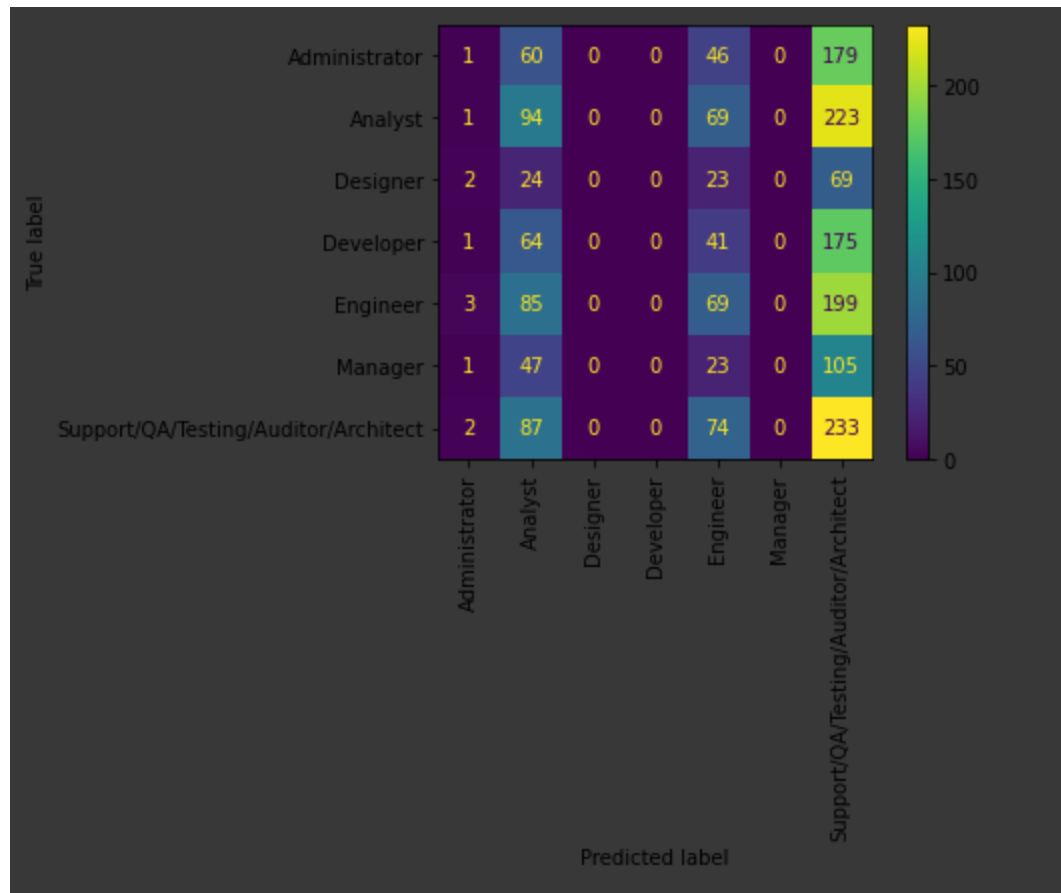
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 19.382022471910112%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 58.838383838383834%



Train Test Split of: 80%-20%

precision recall f1-score support

Administrator	0.00	0.00	0.00	572
Analyst	0.19	0.44	0.27	775
Designer	0.00	0.00	0.00	235
Developer	0.20	0.00	0.00	561
Engineer	0.20	0.12	0.15	712
Manager	0.00	0.00	0.00	353

Support/QA/Testing/Auditor/Architect	0.21	0.46	0.29	792
--------------------------------------	------	------	------	-----

accuracy		0.20	4000	
macro avg	0.11	0.15	0.10	4000
weighted avg	0.14	0.20	0.14	4000

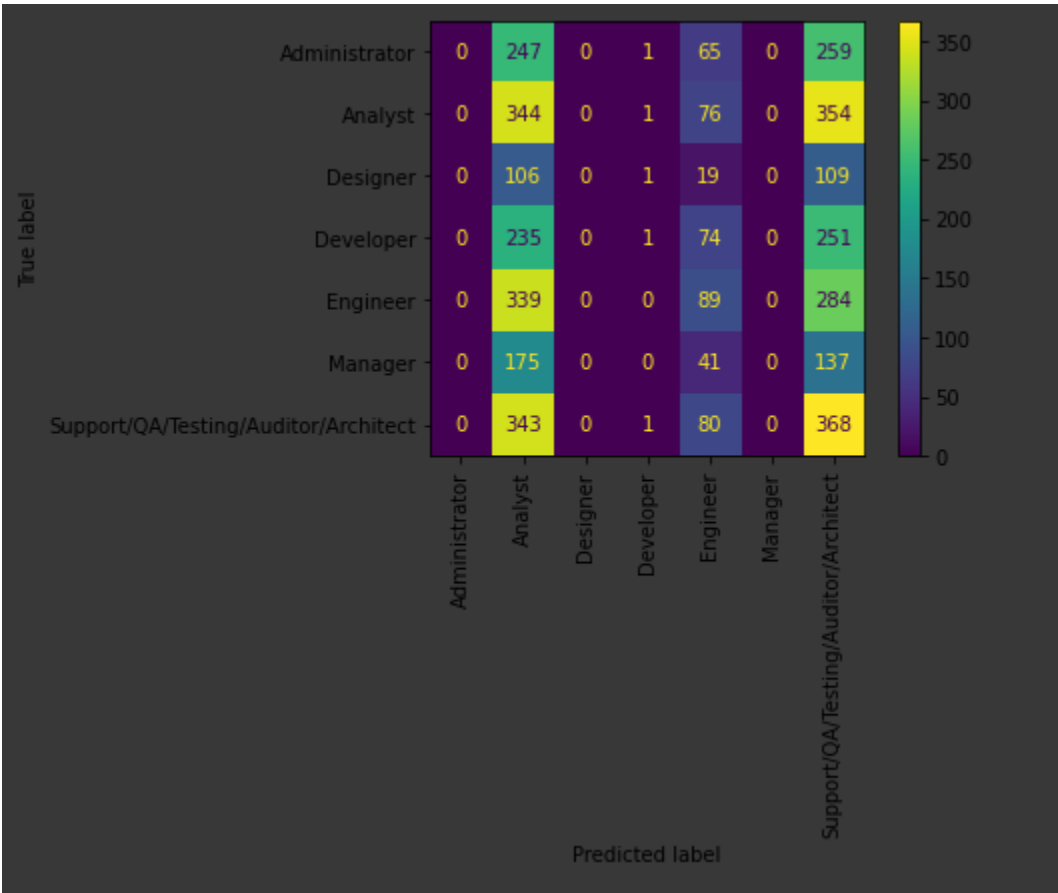
Accuracy of class Administrator is : 0.0%

Accuracy of class Analyst is : 44.387096774193544%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.17825311942959002%

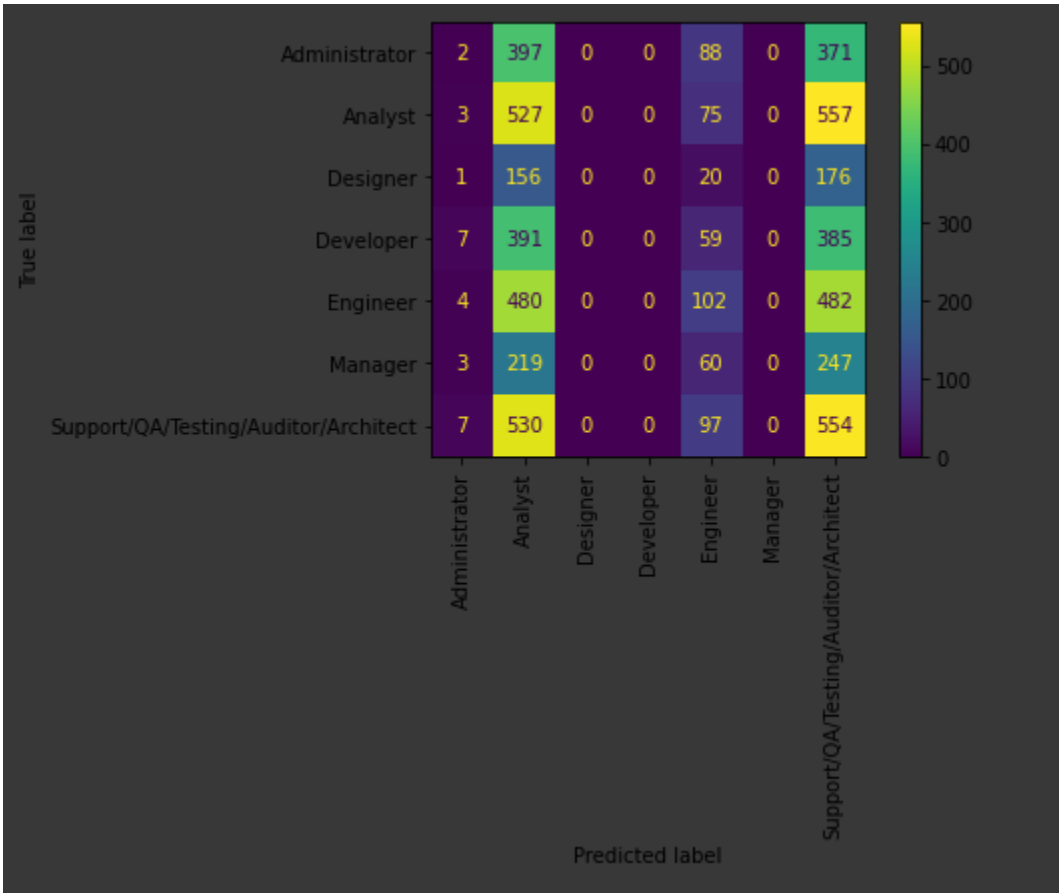
Accuracy of class Engineer is : 12.5%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is :
46.4646464646464%



Train Test Split of: 70%-30%

	precision	recall	f1-score	support	
Administrator	0.07	0.00	0.00	858	
Analyst	0.20	0.45	0.27	1162	
Designer	0.00	0.00	0.00	353	
Developer	0.00	0.00	0.00	842	
Engineer	0.20	0.10	0.13	1068	
Manager	0.00	0.00	0.00	529	
Support/QA/Testing/Auditor/Architect		0.20	0.47	0.28	1188
accuracy		0.20	6000		
macro avg	0.10	0.15	0.10	6000	
weighted avg	0.12	0.20	0.13	6000	

Accuracy of class Administrator is : 0.2331002331002331%
Accuracy of class Analyst is : 45.352839931153184%
Accuracy of class Designer is : 0.0%
Accuracy of class Developer is : 0.0%
Accuracy of class Engineer is : 9.550561797752808%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is :
46.63299663299663%



Train Test Split of: 60%-40%

	precision	recall	f1-score	support	
Administrator	0.22	0.00	0.00	1144	
Analyst	0.20	0.45	0.27	1549	
Designer	0.00	0.00	0.00	471	
Developer	0.00	0.00	0.00	1123	
Engineer	0.16	0.16	0.16	1424	
Manager	0.00	0.00	0.00	705	
Support/QA/Testing/Auditor/Architect		0.20	0.38	0.26	1584

accuracy		0.19	8000
macro avg	0.11	0.14	0.10
weighted avg	0.14	0.19	0.13

Accuracy of class Administrator is : 0.17482517482517482%

Accuracy of class Analyst is : 45.12588766946417%

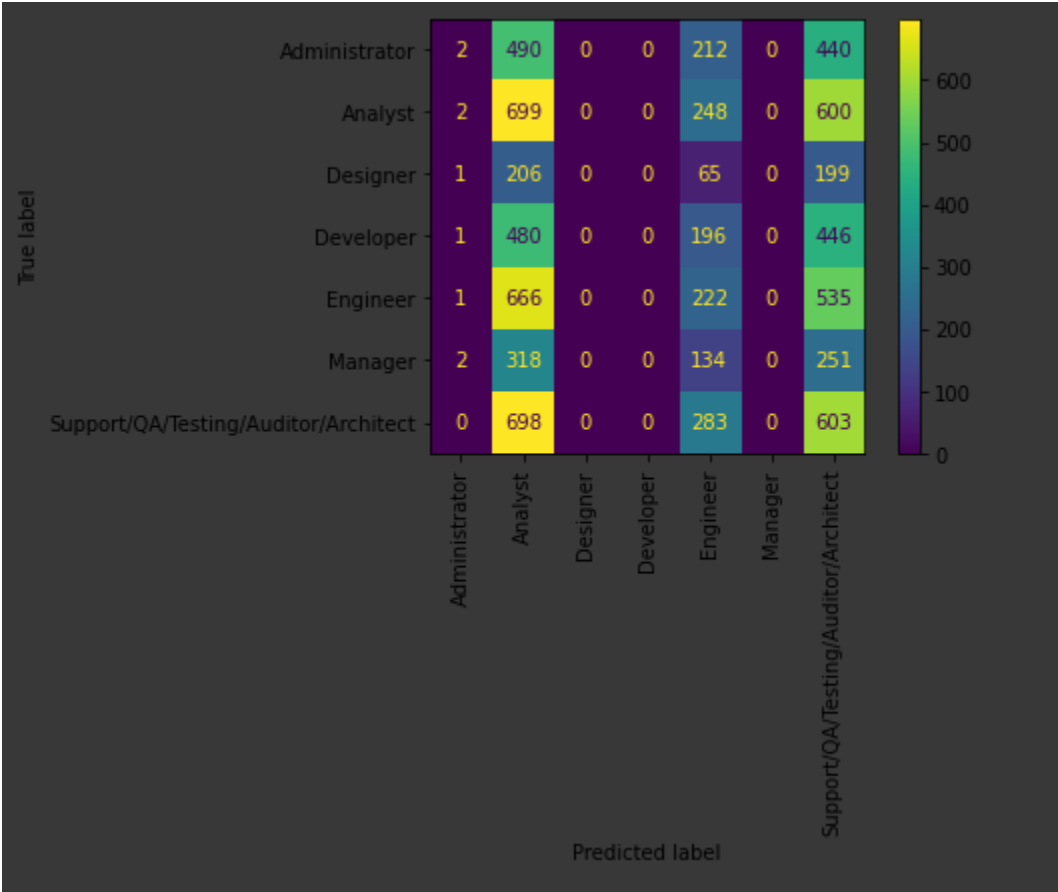
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 15.589887640449437%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 38.06818181818182%



Train Test Split of: 50%-50%

	precision	recall	f1-score	support
Administrator	0.33	0.00	0.00	1430
Analyst	0.19	0.38	0.25	1937
Designer	0.00	0.00	0.00	589
Developer	0.00	0.00	0.00	1403

Engineer	0.16	0.15	0.16	1780
Manager	0.00	0.00	0.00	881
Support/QA/Testing/Auditor/Architect	0.20	0.44	0.27	1980

accuracy		0.19	10000
macro avg	0.13	0.14	0.10
weighted avg	0.15	0.19	0.13

Accuracy of class Administrator is : 0.06993006993006994%

Accuracy of class Analyst is : 38.0485286525555%

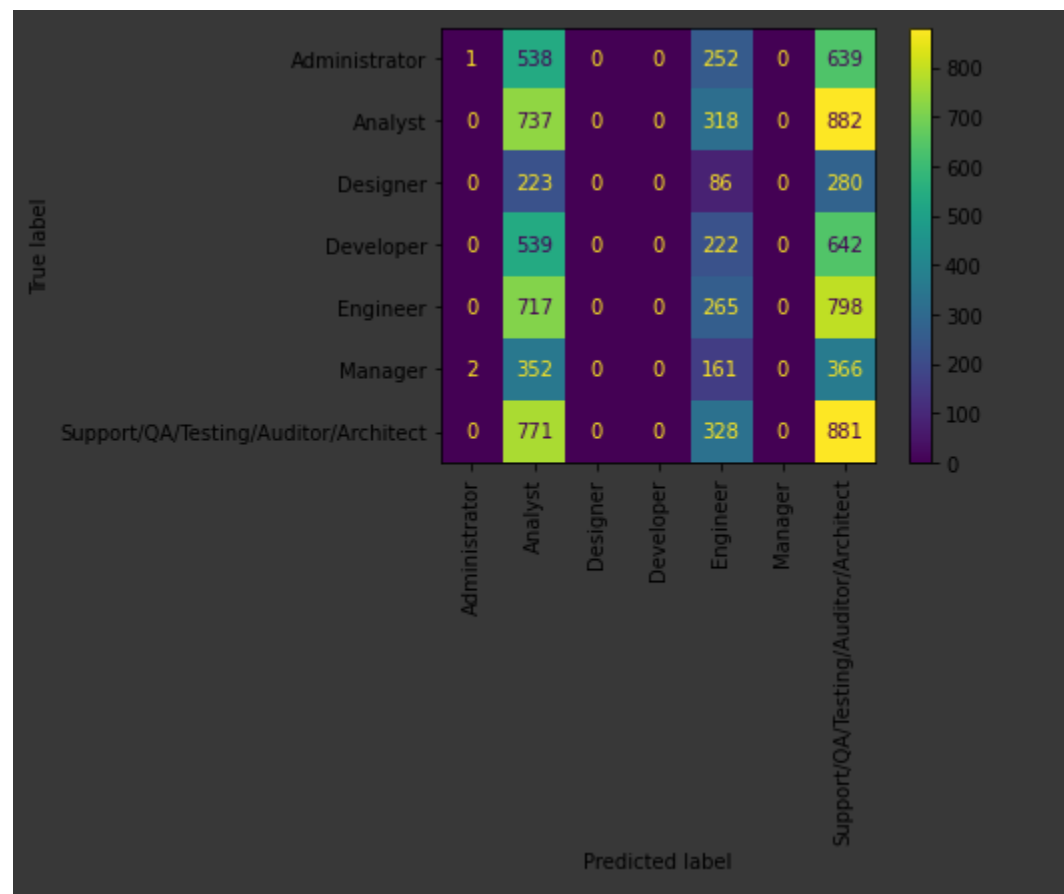
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 14.887640449438203%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 44.494949494949495%



Train Test Split of: 40%-60%

precision recall f1-score support

Administrator	0.16	0.01	0.01	1716
Analyst	0.20	0.67	0.31	2324
Designer	0.00	0.00	0.00	706
Developer	0.18	0.01	0.02	1684
Engineer	0.17	0.08	0.11	2136
Manager	0.50	0.00	0.00	1058
Support/QA/Testing/Auditor/Architect		0.21	0.27	0.24 2376
accuracy		0.20	12000	
macro avg	0.20	0.15	0.10	12000
weighted avg	0.20	0.20	0.13	12000

Accuracy of class Administrator is : 0.5827505827505828%

Accuracy of class Analyst is : 67.38382099827884%

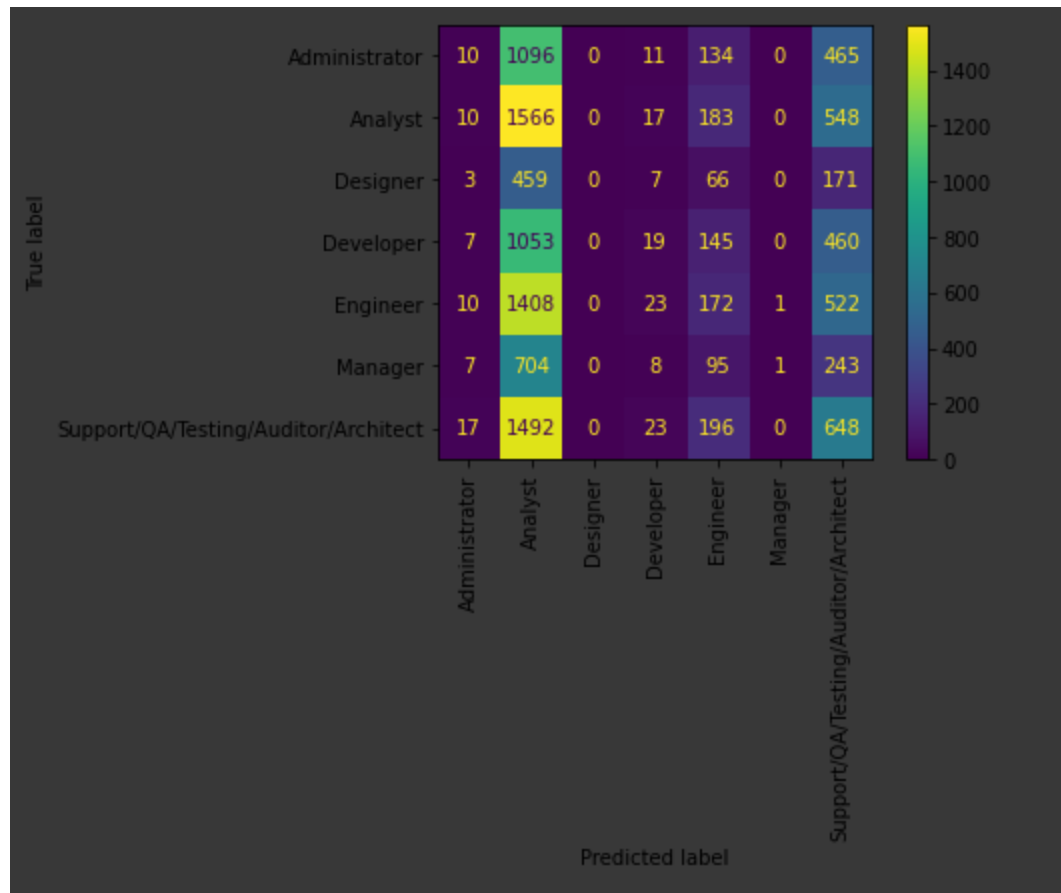
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 1.1282660332541568%

Accuracy of class Engineer is : 8.05243445692884%

Accuracy of class Manager is : 0.0945179584120983%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 27.27272727272727%



Train Test Split of: 30%-70%

	precision	recall	f1-score	support
Administrator	0.05	0.00	0.00	2002
Analyst	0.19	0.43	0.27	2712
Designer	0.00	0.00	0.00	824
Developer	0.00	0.00	0.00	1965
Engineer	0.18	0.11	0.14	2492
Manager	0.00	0.00	0.00	1234
Support/QA/Testing/Auditor/Architect		0.20	0.47	0.28 2771
accuracy		0.20	14000	
macro avg	0.09	0.14	0.10	14000
weighted avg	0.12	0.20	0.13	14000

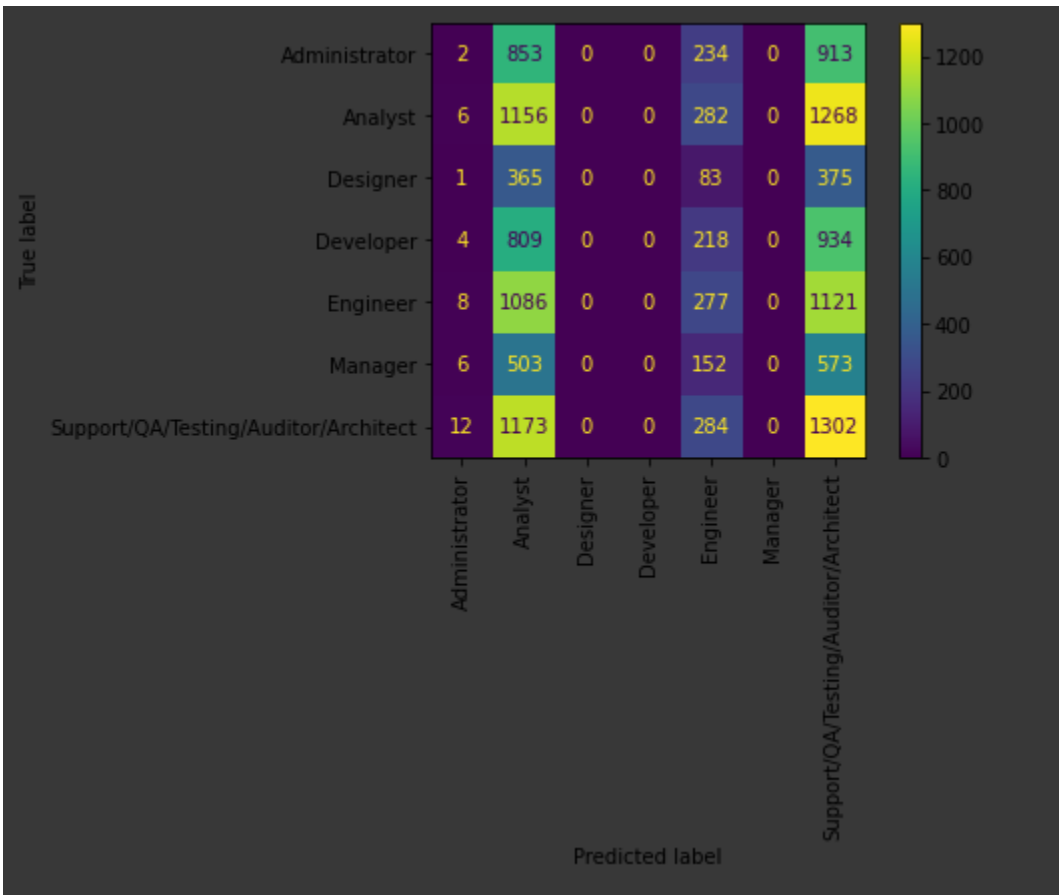
Accuracy of class Administrator is : 0.0999000999000999%

Accuracy of class Analyst is : 42.62536873156343%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

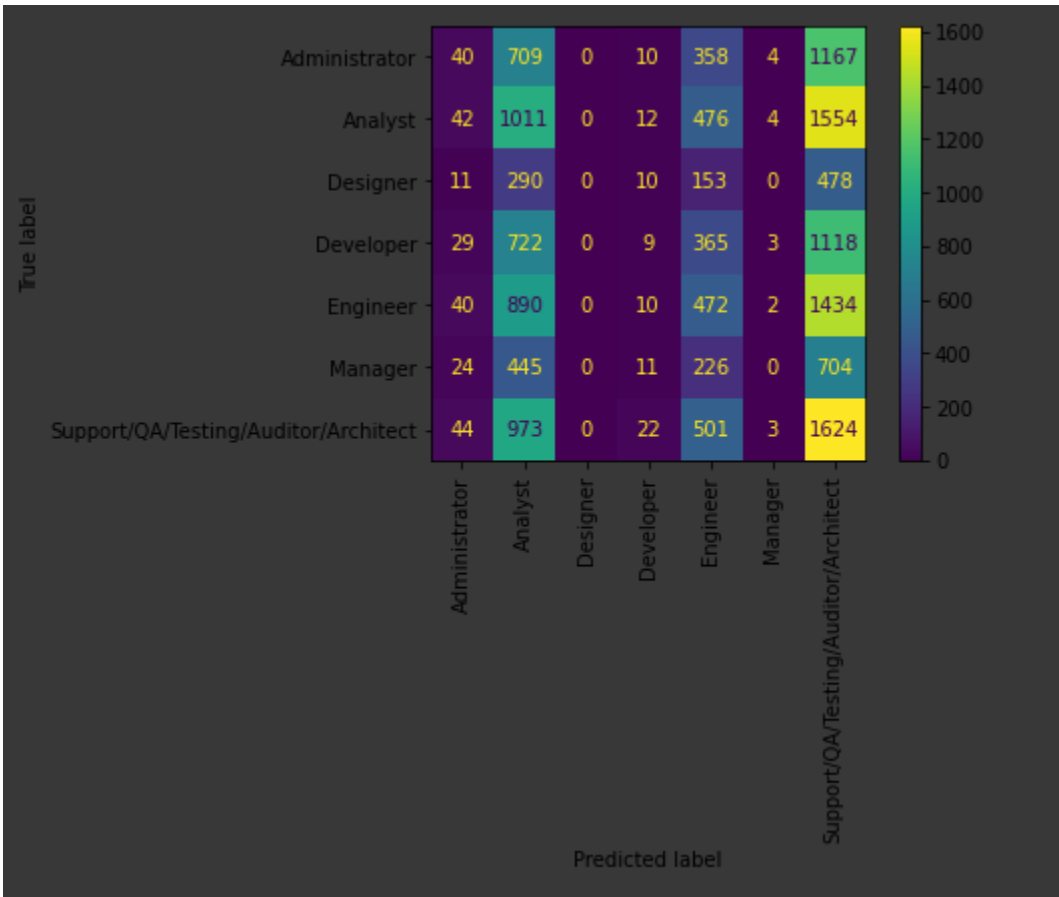
Accuracy of class Engineer is : 11.115569823434992%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is :
46.98664741970408%



Train Test Split of: 20%-80%

	precision	recall	f1-score	support
Administrator	0.17	0.02	0.03	2288
Analyst	0.20	0.33	0.25	3099
Designer	0.00	0.00	0.00	942
Developer	0.11	0.00	0.01	2246
Engineer	0.19	0.17	0.17	2848
Manager	0.00	0.00	0.00	1410
Support/QA/Testing/Auditor/Architect		0.20	0.51	0.29 3167
accuracy		0.20	16000	
macro avg	0.12	0.15	0.11	16000
weighted avg	0.15	0.20	0.14	16000

Accuracy of class Administrator is : 1.7482517482517483%
Accuracy of class Analyst is : 32.62342691190707%
Accuracy of class Designer is : 0.0%
Accuracy of class Developer is : 0.4007123775601068%
Accuracy of class Engineer is : 16.573033707865168%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is : 51.278812756551936%



Train Test Split of: 10%-90%

	precision	recall	f1-score	support		
Administrator	0.14	0.12	0.13	2574		
Analyst	0.21	0.12	0.16	3487		
Designer	0.00	0.00	0.00	1059		
Developer	0.14	0.00	0.00	2526		
Engineer	0.18	0.77	0.29	3204		
Manager	0.00	0.00	0.00	1587		
Support/QA/Testing/Auditor/Architect			0.13	0.00	0.00	3563

accuracy		0.18	18000
macro avg	0.12	0.14	0.08
weighted avg	0.14	0.18	0.10

Accuracy of class Administrator is : 11.771561771561771%

Accuracy of class Analyst is : 12.41755090335532%

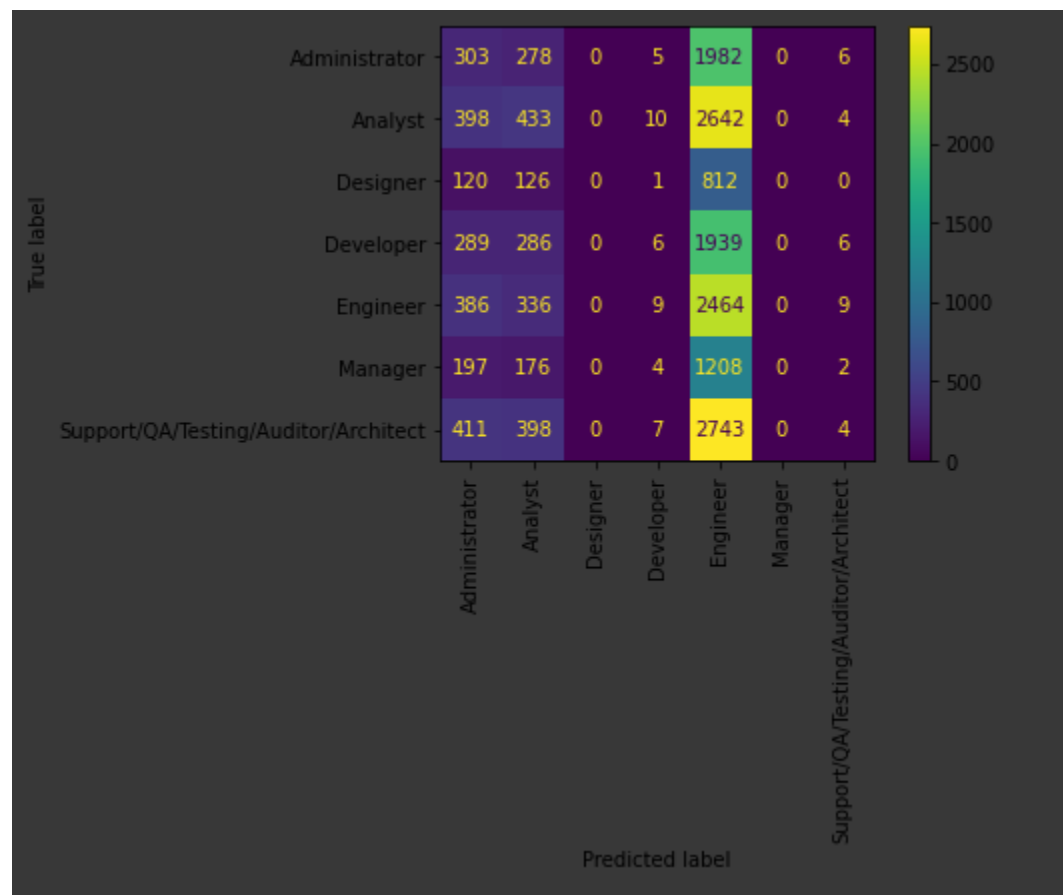
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.23752969121140144%

Accuracy of class Engineer is : 76.90387016229712%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is :
0.11226494527083919%



- With Shuffling Data:

Train Test Split of: 90%-10%

	precision	recall	f1-score	support
Administrator	0.00	0.00	0.00	286

Analyst	0.19	0.40	0.26	387
Designer	0.00	0.00	0.00	118
Developer	0.00	0.00	0.00	281
Engineer	0.00	0.00	0.00	356
Manager	0.00	0.00	0.00	176
Support/QA/Testing/Auditor/Architect	0.20	0.61	0.30	396
accuracy		0.20		2000
macro avg	0.06	0.14	0.08	2000
weighted avg	0.08	0.20	0.11	2000

Accuracy of class Administrator is : 0.0%

Accuracy of class Analyst is : 39.53488372093023%

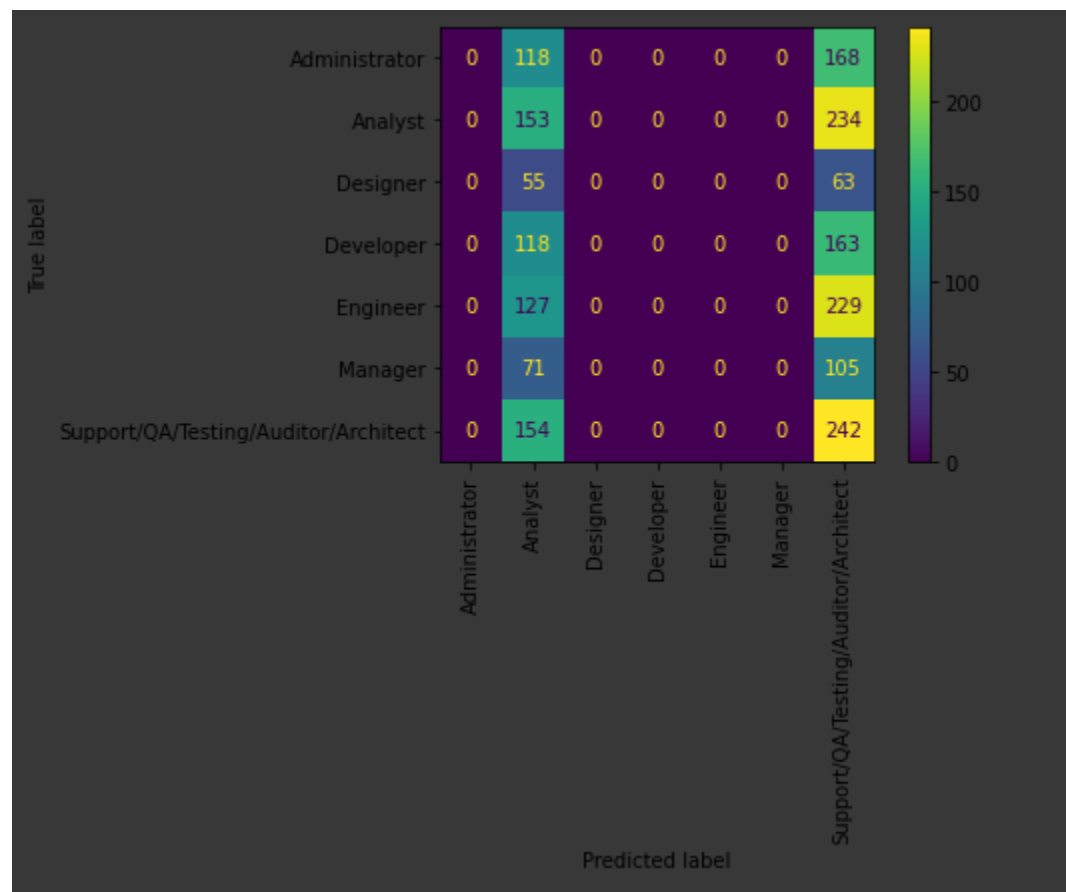
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 0.0%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is :
61.111111111111114%



Train Test Split of: 80%-20%

	precision	recall	f1-score	support
Administrator	0.00	0.00	0.00	572
Analyst	0.18	0.35	0.24	775
Designer	0.00	0.00	0.00	235
Developer	0.00	0.00	0.00	561
Engineer	0.17	0.03	0.06	712
Manager	0.00	0.00	0.00	353
Support/QA/Testing/Auditor/Architect		0.20	0.59	0.29 792
accuracy		0.19	4000	
macro avg	0.08	0.14	0.08	4000
weighted avg	0.10	0.19	0.11	4000

Accuracy of class Administrator is : 0.0%

Accuracy of class Analyst is : 34.83870967741935%

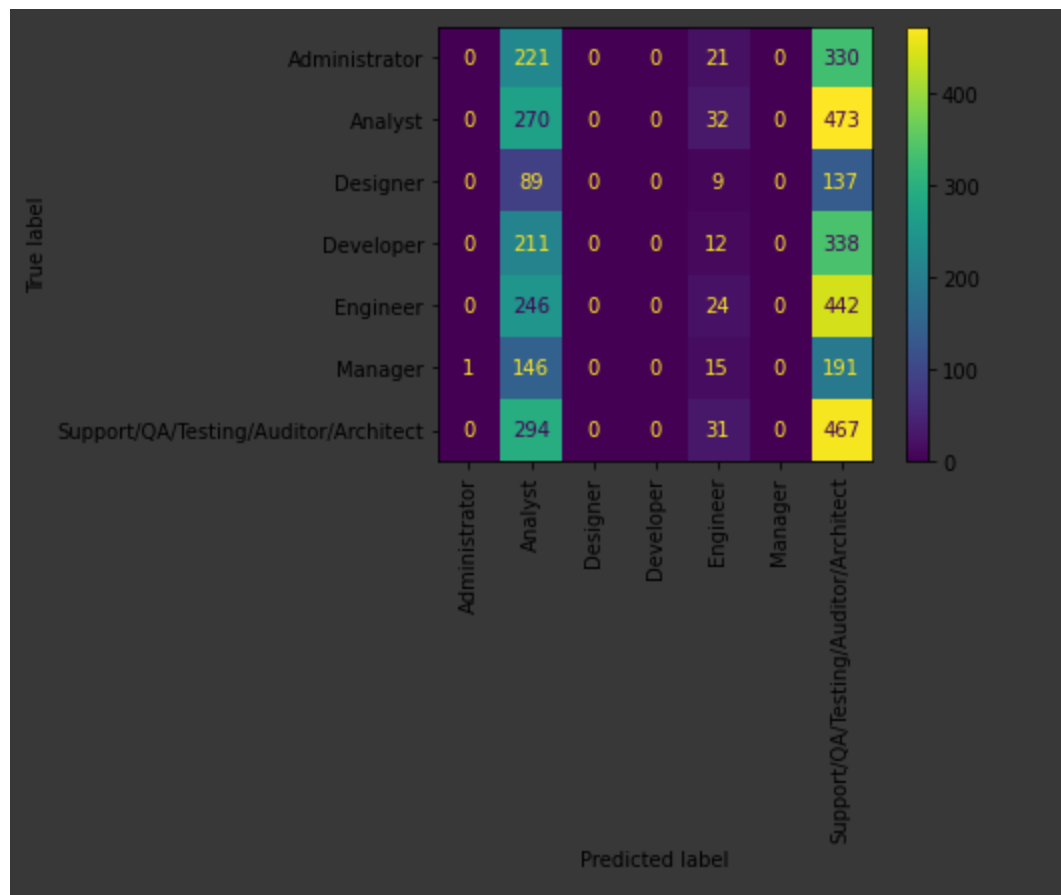
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 3.3707865168539324%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is :
58.964646464646464%



Train Test Split of: 70%-30%

precision recall f1-score support

Administrator	0.00	0.00	0.00	858
Analyst	0.21	0.18	0.19	1162
Designer	0.00	0.00	0.00	353
Developer	0.00	0.00	0.00	842
Engineer	0.16	0.05	0.07	1068
Manager	0.00	0.00	0.00	529

Support/QA/Testing/Auditor/Architect	0.20	0.79	0.32	1188
--------------------------------------	------	------	------	------

accuracy		0.20	6000	
macro avg	0.08	0.14	0.08	6000
weighted avg	0.11	0.20	0.11	6000

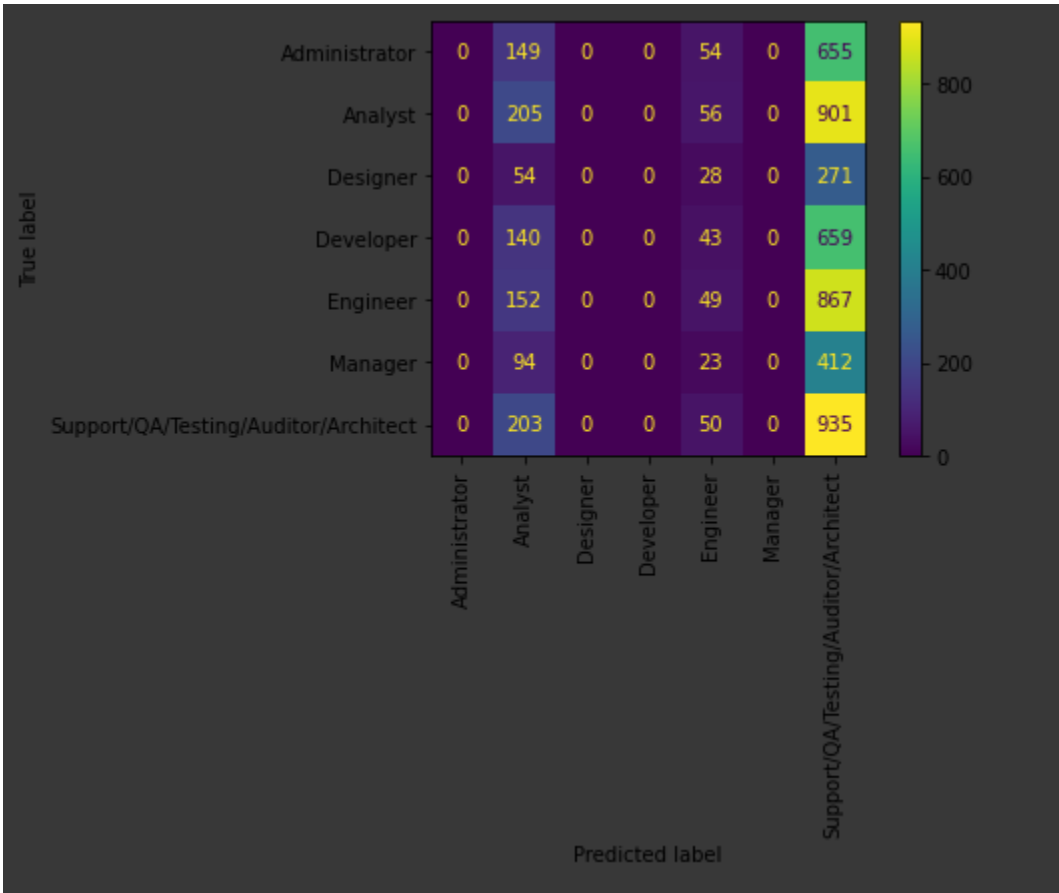
Accuracy of class Administrator is : 0.0%

Accuracy of class Analyst is : 17.64199655765921%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

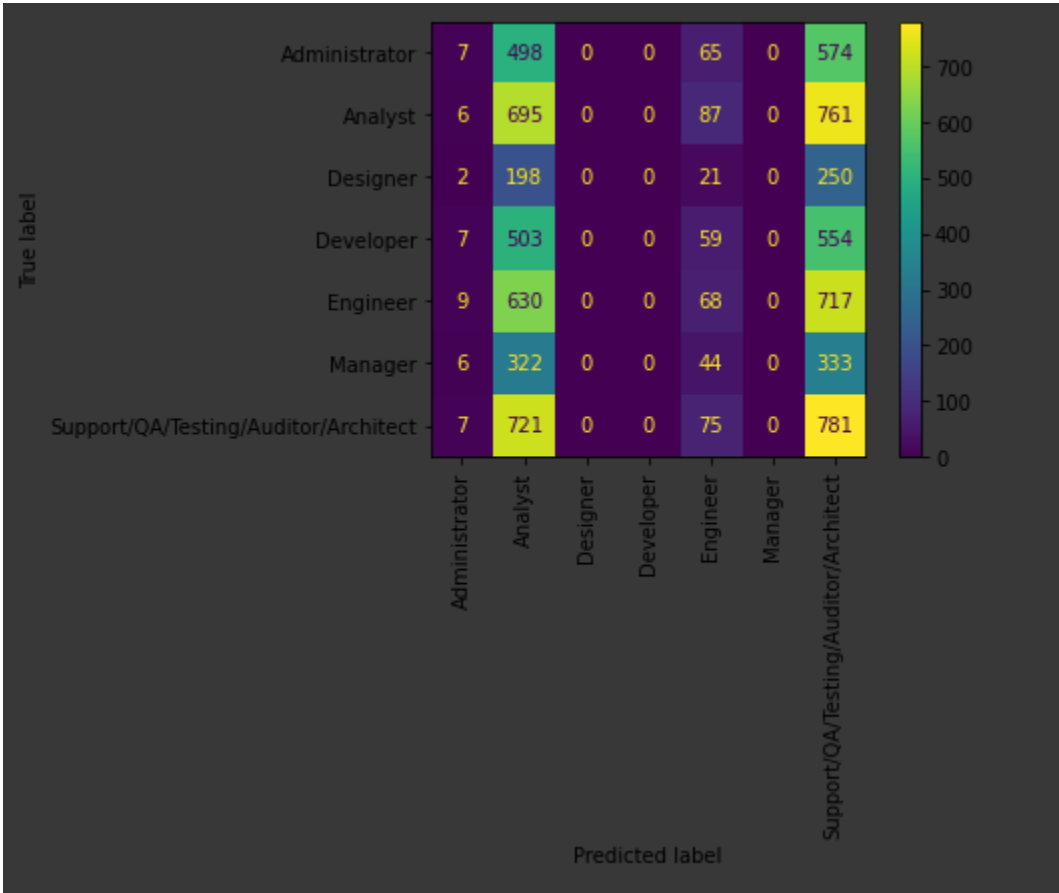
Accuracy of class Engineer is : 4.588014981273409%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is : 78.70370370370371%



Train Test Split of: 60%-40%

	precision	recall	f1-score	support
Administrator	0.16	0.01	0.01	1144
Analyst	0.19	0.45	0.27	1549
Designer	0.00	0.00	0.00	471
Developer	0.00	0.00	0.00	1123
Engineer	0.16	0.05	0.07	1424
Manager	0.00	0.00	0.00	705
Support/QA/Testing/Auditor/Architect		0.20	0.49	0.281584
accuracy		0.19		8000
macro avg	0.10	0.14	0.09	8000
weighted avg	0.13	0.19	0.12	8000

Accuracy of class Administrator is : 0.6118881118881119%
Accuracy of class Analyst is : 44.86765655261459%
Accuracy of class Designer is : 0.0%
Accuracy of class Developer is : 0.0%
Accuracy of class Engineer is : 4.775280898876404%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is : 49.30555555555556%



Train Test Split of: 50%-50%

	precision	recall	f1-score	support
Administrator	0.11	0.01	0.01	1430
Analyst	0.20	0.44	0.27	1937
Designer	0.00	0.00	0.00	588
Developer	0.00	0.00	0.00	1404
Engineer	0.19	0.11	0.14	1780
Manager	0.00	0.00	0.00	881
Support/QA/Testing/Auditor/Architect		0.20	0.46	0.28 1980

accuracy		0.20	10000
macro avg	0.10	0.15	10000
weighted avg	0.13	0.20	10000

Accuracy of class Administrator is : 0.6293706293706294%

Accuracy of class Analyst is : 44.29530201342282%

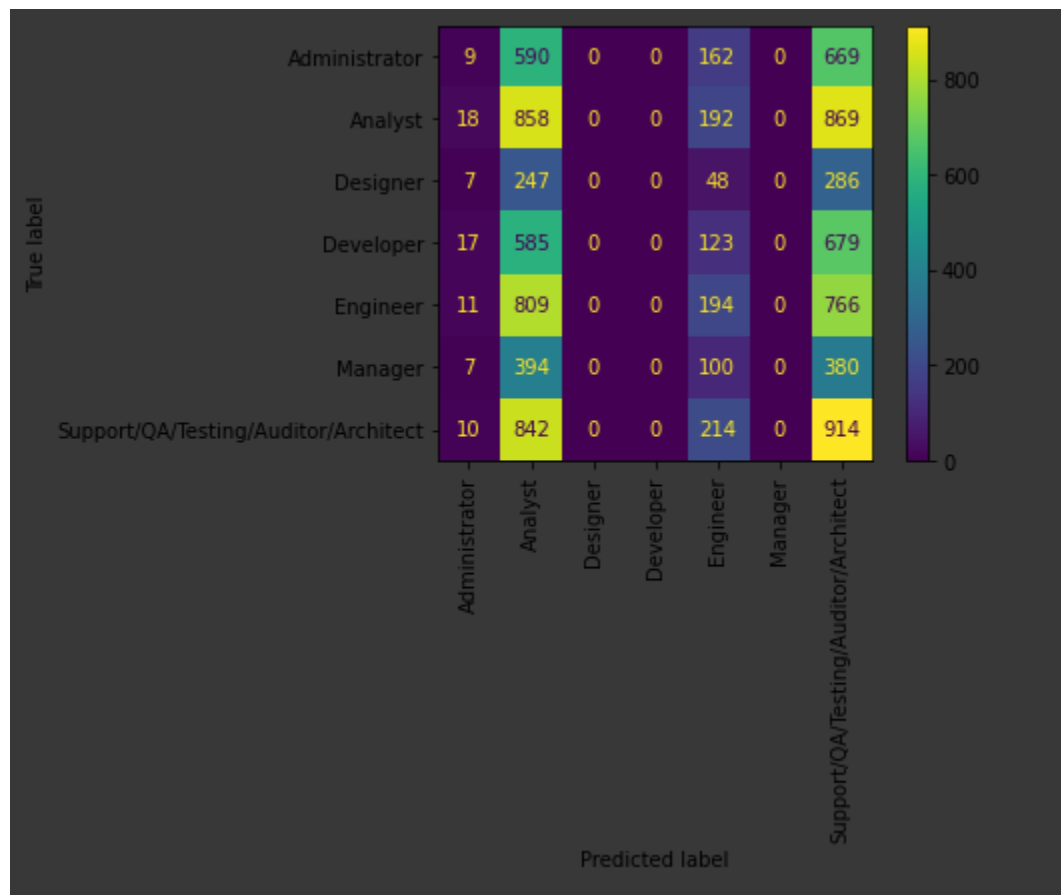
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 10.89887640449438%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 46.16161616161616%



Train Test Split of: 40%-60%

	precision	recall	f1-score	support
Administrator	0.10	0.00	0.00	1716
Analyst	0.19	0.60	0.29	2324
Designer	0.00	0.00	0.00	706
Developer	0.00	0.00	0.00	1684

Engineer	0.18	0.16	0.17	2136	
Manager	0.00	0.00	0.00	1058	
Support/QA/Testing/Auditor/Architect		0.20	0.24	0.22	2376

accuracy		0.19	12000	
macro avg	0.10	0.14	0.10	12000
weighted avg	0.12	0.19	0.13	12000

Accuracy of class Administrator is : 0.17482517482517482%

Accuracy of class Analyst is : 59.93975903614458%

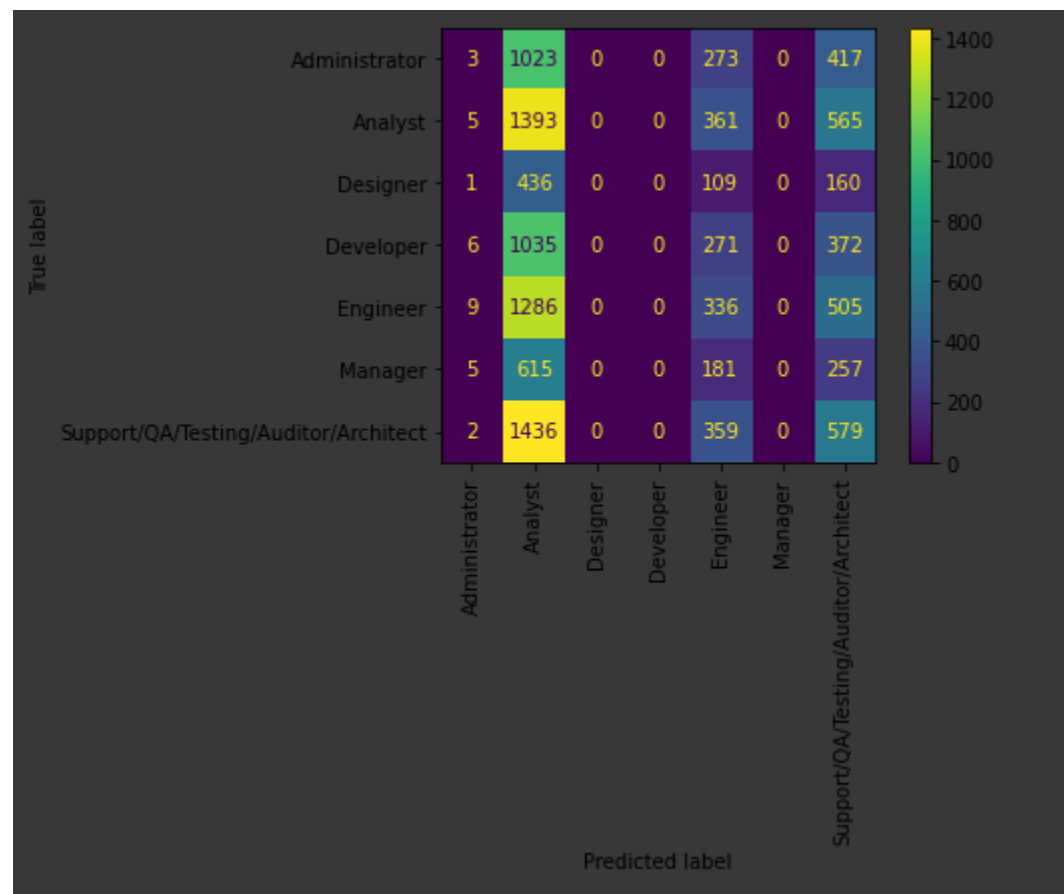
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.0%

Accuracy of class Engineer is : 15.730337078651685%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is :
24.36868686868687%



Train Test Split of: 30%-70%

precision recall f1-score support

Administrator	0.12	0.01	0.01	2002
Analyst	0.20	0.27	0.23	2712
Designer	0.00	0.00	0.00	824
Developer	0.33	0.00	0.00	1965
Engineer	0.17	0.17	0.17	2492
Manager	0.00	0.00	0.00	1234
Support/QA/Testing/Auditor/Architect		0.20	0.55	0.29 2771

accuracy		0.19	14000
macro avg	0.15	0.14	0.10 14000
weighted avg	0.17	0.19	0.13 14000

Accuracy of class Administrator is : 0.5494505494505495%

Accuracy of class Analyst is : 27.47050147492625%

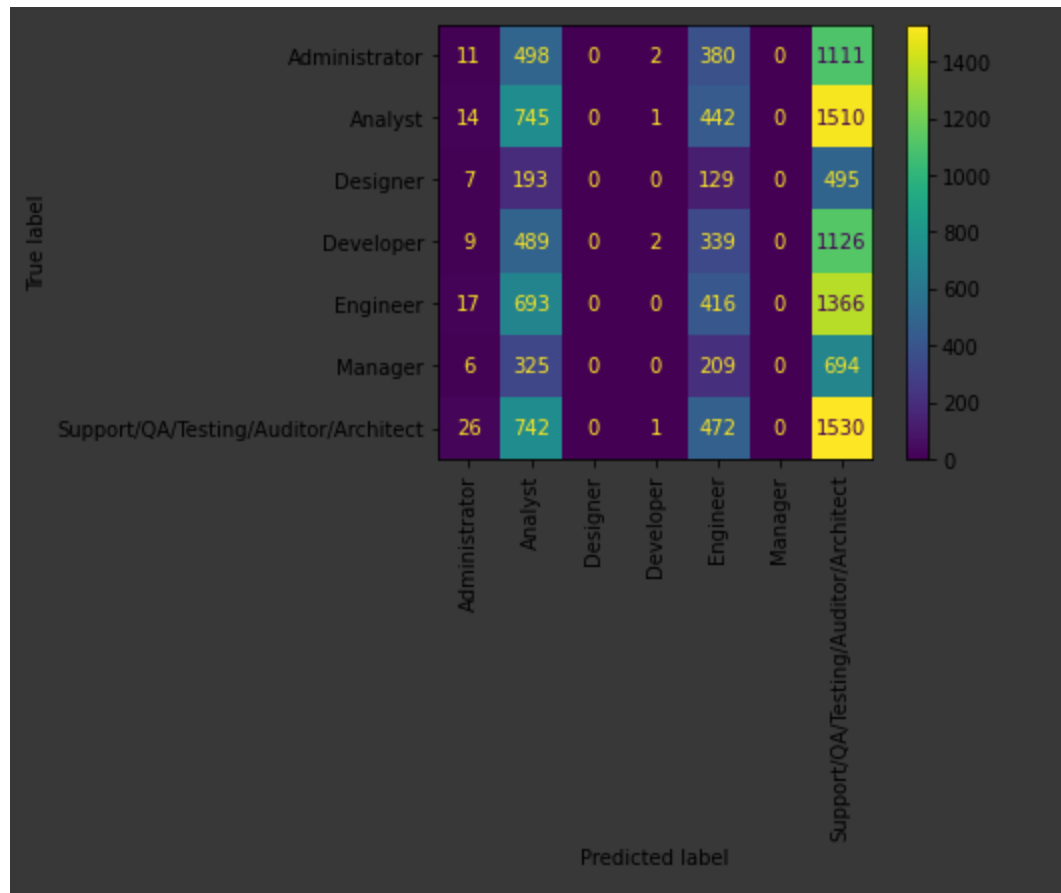
Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.10178117048346055%

Accuracy of class Engineer is : 16.693418940609952%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is : 55.21472392638037%



Train Test Split of: 20%-80%

precision recall f1-score support

Administrator	0.11	0.01	0.01	2288
Analyst	0.20	0.37	0.26	3099
Designer	0.00	0.00	0.00	942
Developer	0.20	0.00	0.00	2246
Engineer	0.16	0.10	0.13	2848
Manager	0.00	0.00	0.00	1410

Support/QA/Testing/Auditor/Architect	0.20	0.51	0.28	3167
--------------------------------------	------	------	------	------

accuracy		0.19	16000
macro avg	0.12	0.14	0.10
weighted avg	0.15	0.19	0.13

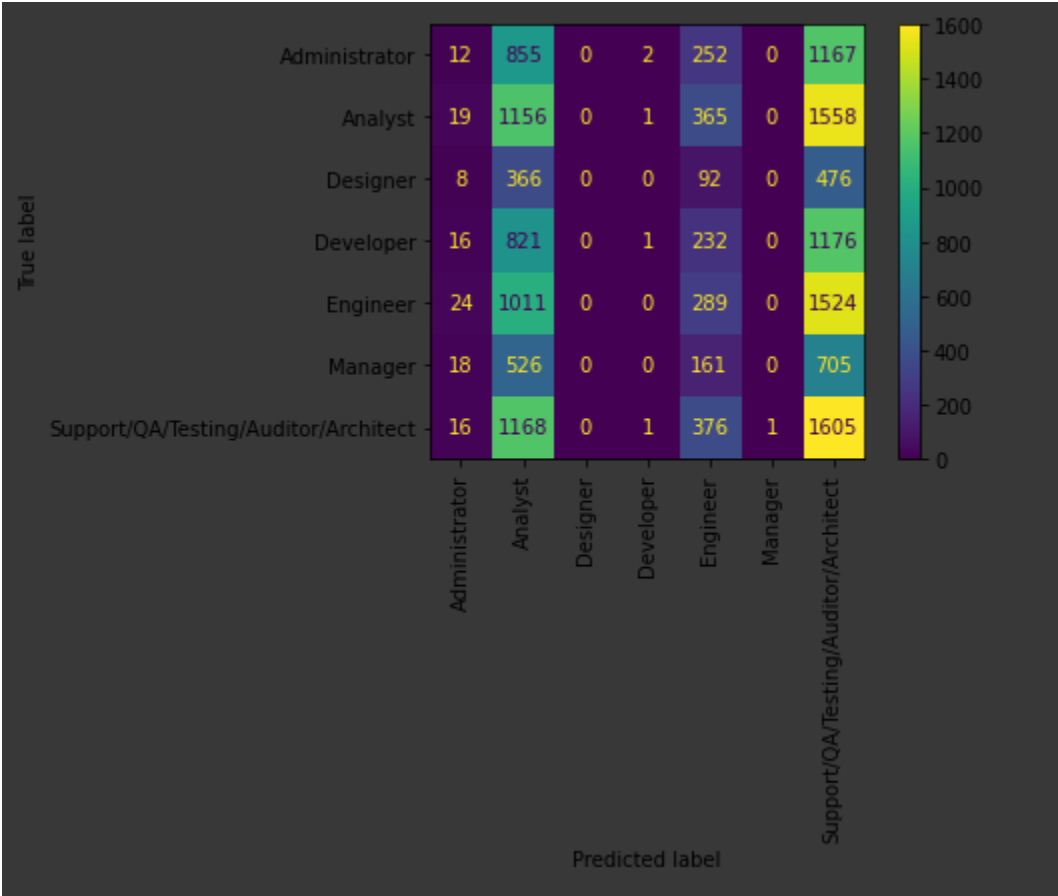
Accuracy of class Administrator is : 0.5244755244755245%

Accuracy of class Analyst is : 37.30235559858019%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.04452359750667854%

Accuracy of class Engineer is : 10.14747191011236%
Accuracy of class Manager is : 0.0%
Accuracy of class Support/QA/Testing/Auditor/Architect is : 50.67887590779918%



Train Test Split of: 10%-90%

	precision	recall	f1-score	support		
Administrator	0.12	0.02	0.04	2574		
Analyst	0.20	0.41	0.27	3487		
Designer	0.00	0.00	0.00	1059		
Developer	0.10	0.00	0.00	2526		
Engineer	0.17	0.17	0.17	3204		
Manager	0.00	0.00	0.00	1587		
Support/QA/Testing/Auditor/Architect			0.20	0.39	0.26	3563
accuracy			0.19	18000		
macro avg	0.11	0.14	0.11	18000		
weighted avg	0.14	0.19	0.14	18000		

Accuracy of class Administrator is : 2.253302253302253%

Accuracy of class Analyst is : 41.03814166905649%

Accuracy of class Designer is : 0.0%

Accuracy of class Developer is : 0.11876484560570072%

Accuracy of class Engineer is : 17.228464419475657%

Accuracy of class Manager is : 0.0%

Accuracy of class Support/QA/Testing/Auditor/Architect is :
39.068200954252035%

