In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import datasets, preprocessing
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn import svm
from sklearn import preprocessing, model_selection, neighbors, discriminant_analysis
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn import datasets, linear_model
from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
```

In [2]:
```
file = ("glass.data")
df = pd.read_csv(file,delim_whitespace=False, header=None)
columns = ["Id", "Ri", "Na", "Mg", "Al", "Si", "K", "Ca", "Ba", "Fe", "Type"]
df.columns = columns
df.head()
```

Out[2]:

	Id	Ri	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
0	1	1.52101	13.64	4.49	1.10	71.78	0.06	8.75	0.0	0.0	1
1	2	1.51761	13.89	3.60	1.36	72.73	0.48	7.83	0.0	0.0	1
2	3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0.0	0.0	1
3	4	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0.0	0.0	1
4	5	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0.0	0.0	1

In [3]:
```
min_max_scaler = preprocessing.MinMaxScaler()
normalized = min_max_scaler.fit_transform(df)
print(normalized)
```

```
[[0.         0.43283582 0.43759398 ... 0.         0.         0.        ]
 [0.00469484 0.28358209 0.47518797 ... 0.         0.         0.        ]
 [0.00938967 0.22080773 0.42105263 ... 0.         0.         0.        ]
 ...
 [0.99061033 0.41703248 0.54586466 ... 0.52063492 0.         1.        ]
 [0.99530516 0.23529412 0.54887218 ... 0.4984127  0.         1.        ]
 [1.         0.26163301 0.52631579 ... 0.53015873 0.         1.        ]]
```

In [4]:
```
normalized_df = pd.DataFrame(normalized, columns=df.columns)
print(normalized_df)
```

```
           Id        Ri        Na        Mg        Al        Si         K  \
0    0.000000  0.432836  0.437594  1.000000  0.252336  0.351786  0.009662
1    0.004695  0.283582  0.475188  0.801782  0.333333  0.521429  0.077295
2    0.009390  0.220808  0.421053  0.790646  0.389408  0.567857  0.062802
3    0.014085  0.285777  0.372932  0.821826  0.311526  0.500000  0.091787
4    0.018779  0.275241  0.381955  0.806236  0.295950  0.583929  0.088567
..        ...       ...       ...       ...       ...       ...       ...
209  0.981221  0.223003  0.512782  0.000000  0.806854  0.500000  0.012882
210  0.985915  0.250219  0.630075  0.000000  0.529595  0.580357  0.000000
211  0.990610  0.417032  0.545865  0.000000  0.538941  0.644643  0.000000
212  0.995305  0.235294  0.548872  0.000000  0.514019  0.678571  0.000000
213  1.000000  0.261633  0.526316  0.000000  0.557632  0.633929  0.000000

           Ca        Ba   Fe  Type
0    0.308550  0.000000  0.0   0.0
1    0.223048  0.000000  0.0   0.0
2    0.218401  0.000000  0.0   0.0
3    0.259294  0.000000  0.0   0.0
4    0.245353  0.000000  0.0   0.0
..        ...       ...  ...   ...
209  0.348513  0.336508  0.0   1.0
210  0.276022  0.504762  0.0   1.0
211  0.279740  0.520635  0.0   1.0
212  0.283457  0.498413  0.0   1.0
213  0.296468  0.530159  0.0   1.0

[214 rows x 11 columns]
```

In [5]:
```
df = normalized_df.iloc[:, 0:10]
Target = normalized_df.Type
df,Target
```

Out[5]:
```
(           Id        Ri        Na        Mg        Al        Si         K  \
 0    0.000000  0.432836  0.437594  1.000000  0.252336  0.351786  0.009662
 1    0.004695  0.283582  0.475188  0.801782  0.333333  0.521429  0.077295
 2    0.009390  0.220808  0.421053  0.790646  0.389408  0.567857  0.062802
 3    0.014085  0.285777  0.372932  0.821826  0.311526  0.500000  0.091787
 4    0.018779  0.275241  0.381955  0.806236  0.295950  0.583929  0.088567
 ..        ...       ...       ...       ...       ...       ...       ...
 209  0.981221  0.223003  0.512782  0.000000  0.806854  0.500000  0.012882
 210  0.985915  0.250219  0.630075  0.000000  0.529595  0.580357  0.000000
 211  0.990610  0.417032  0.545865  0.000000  0.538941  0.644643  0.000000
 212  0.995305  0.235294  0.548872  0.000000  0.514019  0.678571  0.000000
 213  1.000000  0.261633  0.526316  0.000000  0.557632  0.633929  0.000000

           Ca        Ba   Fe
 0    0.308550  0.000000  0.0
 1    0.223048  0.000000  0.0
 2    0.218401  0.000000  0.0
 3    0.259294  0.000000  0.0
 4    0.245353  0.000000  0.0
 ..        ...       ...  ...
 209  0.348513  0.336508  0.0
 210  0.276022  0.504762  0.0
 211  0.279740  0.520635  0.0
 212  0.283457  0.498413  0.0
 213  0.296468  0.530159  0.0

 [214 rows x 10 columns],
 0      0.0
 1      0.0
 2      0.0
 3      0.0
 4      0.0
       ...
 209    1.0
 210    1.0
 211    1.0
 212    1.0
 213    1.0
 Name: Type, Length: 214, dtype: float64)
```

In [6]:
```
x_train, x_test, y_train, y_test = model_selection.train_test_split(df, Target, test_size=0.3,random_state=2)
```

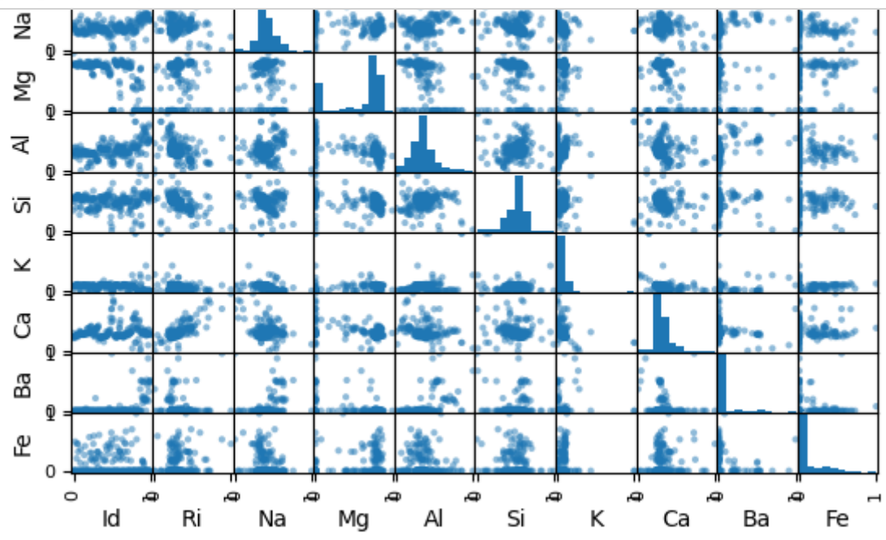
In [7]:
```
x_train,y_train
```

Out[7]:
```
(           Id        Ri        Na        Mg        Al        Si         K  \
109  0.511737  0.308604  0.449624  0.000000  0.084112  0.828571  0.000000
146  0.685446  0.287094  0.439098  0.815145  0.255452  0.528571  0.017713
178  0.835681  0.313433  0.560902  0.498886  0.414330  0.458929  0.000000
184  0.863850  0.000000  1.000000  0.000000  0.015576  1.000000  0.000000
48   0.225352  0.486392  0.372932  0.839644  0.155763  0.389286  0.020934
..        ...       ...       ...       ...       ...       ...       ...
43   0.201878  0.480685  0.451128  0.855234  0.133956  0.348214  0.027375
22   0.103286  0.272608  0.308271  0.806236  0.311526  0.532143  0.095008
72   0.338028  0.209833  0.354887  0.799555  0.383178  0.587500  0.107890
15   0.070423  0.283582  0.312782  0.788419  0.292835  0.612500  0.093398
168  0.788732  0.241879  0.320301  0.000000  0.479751  0.726786  0.156200

           Ca   Ba   Fe
109  0.516729  0.0  0.0
146  0.294610  0.0  0.0
178  0.355948  0.0  0.0
184  0.113383  0.0  0.0
48   0.426580  0.0  0.0
..        ...  ...  ...
43   0.400558  0.0  0.0
22   0.303903  0.0  0.0
72   0.223048  0.0  0.0
15   0.275093  0.0  0.0
168  0.440520  0.0  0.0

[149 rows x 10 columns],
109    0.166667
146    0.333333
178    0.833333
184    0.833333
48     0.000000
         ...
43     0.000000
22     0.000000
72     0.166667
15     0.000000
168    0.666667
Name: Type, Length: 149, dtype: float64)
```

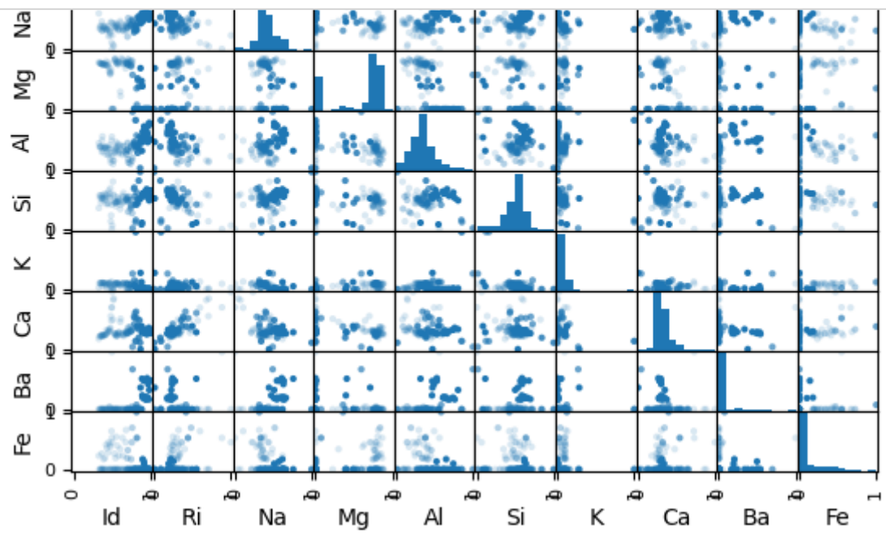
In [8]:
```
print(x_train.shape,y_train.shape)
print(x_test.shape, y_test.shape)
```

```
(149, 10) (149,)
(65, 10) (65,)
```

In [9]:
```
pd.plotting.scatter_matrix(df)
```



In [10]:
```
pd.plotting.scatter_matrix(x_train,y_train)
```



In [11]:
```
pd.plotting.scatter_matrix(x_test,y_test)
```