Mining Quality Prediction

August 3, 2023

1 TASK #1: UNDERSTAND THE PROBLEM STATEMENT AND BUSINESS CASE

2 Applications in which mining companies leverage the power of Artificial Intelligence and Machine Learning.

- Mineral Explorations
- Autonomous Drillers
- Minerals sorting

3 TASK #2: IMPORT LIBRARIES/DATASETS AND PERFORM DATA EXPLORATION

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import zipfile
# setting the style of the notebook to be monokai theme
# this line of code is important to ensure that we are able to see the x and y␣
↪axes clearly
# If you don't run this code line, you will notice that the xlabel and ylabel␣
↪on any plot is black on black and it will be hard to see them.
```

[2]:
```
mining_df = pd.read_csv('mining_data.csv')
mining_df
```

[2]:
```
        % Iron Feed  % Silica Feed  Starch Flow  Amina Flow  Ore Pulp Flow  \
0             55.20          16.98  3196.680000  542.694333     396.284000
1             55.20          16.98  3213.673333  540.649333     397.949333
2             55.20          16.98  3180.080000  535.929333     397.305000
3             55.20          16.98  3196.713333  535.102000     397.010667
4             55.20          16.98  3111.723333  532.735000     395.263667
…             …              …       …            …           …
245695        49.75          23.20  2961.490000  452.658667     381.784667
245696        49.75          23.20  2851.376667  470.843333     384.250000
```

245697	49.75	23.20	2729.876667	463.040667	383.073667
245698	49.75	23.20	2698.756667	460.303667	384.490667
245699	49.75	23.20	1673.480000	486.685000	384.424333

	Ore Pulp pH	Ore Pulp Density	Flotation Column 01 Air Flow	\
0	10.158367	1.668070	249.796333	
1	10.156600	1.664973	249.536000	
2	10.154800	1.661877	249.576000	
3	10.153067	1.658780	249.380333	
4	10.151300	1.655680	249.426667	
…	…	…	…	
245695	9.615990	1.654507	300.351000	
245696	9.617760	1.654413	299.978667	
245697	9.619523	1.654070	302.416667	
245698	9.620873	1.653653	302.603000	
245699	9.616857	1.653240	302.189000	

	Flotation Column 02 Air Flow	Flotation Column 03 Air Flow	…	\
0	250.275667	248.668000	…	
1	250.752000	250.968333	…	
2	250.279667	251.001333	…	
3	248.799333	250.241333	…	
4	252.209667	249.243333	…	
…	…	…	…	
245695	295.703667	298.312667	…	
245696	304.478333	301.176667	…	
245697	299.060000	299.929000	…	
245698	299.953333	299.483667	…	
245699	300.740333	298.719000	…	

	Flotation Column 07 Air Flow	Flotation Column 01 Level	\
0	250.547000	464.978667	
1	249.807000	445.001000	
2	249.686667	443.574667	
3	249.926333	440.731333	
4	249.975667	445.851667	
…	…	…	
245695	321.464444	404.055667	
245696	335.351222	417.104667	
245697	291.529000	386.853000	
245698	285.945000	395.502667	
245699	280.703000	409.253333	

	Flotation Column 02 Level	Flotation Column 03 Level	\
0	490.450333	443.465000	
1	362.894667	442.748333	
2	478.916333	432.779333	

3	488.994000	452.461333
4	418.860000	462.936667
…	…	…
245695	523.933667	880.286833
245696	548.184333	883.347000
245697	546.765000	870.141667
245698	443.776333	872.247333
245699	460.097000	875.170667

	Flotation Column 04 Level	Flotation Column 05 Level \
0	442.856333	438.782333
1	471.045333	445.239667
2	437.401667	441.761000
3	439.572667	434.027333
4	454.948333	453.571667
…	…	…
245695	331.448333	500.245000
245696	388.821000	482.577333
245697	435.600333	488.983667
245698	416.832667	495.943333
245699	401.933000	501.938333

	Flotation Column 06 Level	Flotation Column 07 Level \
0	452.248333	466.300667
1	443.630667	426.921667
2	490.824667	478.046667
3	457.083667	458.815667
4	446.831667	426.600000
…	…	…
245695	424.037667	400.798000
245696	338.373000	371.803333
245697	420.578667	366.426333
245698	437.163000	417.842667
245699	350.411000	436.092667

	% Iron Concentrate	% Silica Concentrate
0	67.06	1.11
1	67.06	1.11
2	67.06	1.11
3	67.06	1.11
4	67.06	1.11
…	…	…
245695	64.27	1.71
245696	64.27	1.71
245697	64.27	1.71
245698	64.27	1.71
245699	64.27	1.71

```
[245700 rows x 23 columns]
```

[3]:
```python
mining_df.dtypes
```

[3]:
```
% Iron Feed                     float64
% Silica Feed                   float64
Starch Flow                     float64
Amina Flow                      float64
Ore Pulp Flow                   float64
Ore Pulp pH                     float64
Ore Pulp Density                float64
Flotation Column 01 Air Flow    float64
Flotation Column 02 Air Flow    float64
Flotation Column 03 Air Flow    float64
Flotation Column 04 Air Flow    float64
Flotation Column 05 Air Flow    float64
Flotation Column 06 Air Flow    float64
Flotation Column 07 Air Flow    float64
Flotation Column 01 Level       float64
Flotation Column 02 Level       float64
Flotation Column 03 Level       float64
Flotation Column 04 Level       float64
Flotation Column 05 Level       float64
Flotation Column 06 Level       float64
Flotation Column 07 Level       float64
% Iron Concentrate              float64
% Silica Concentrate            float64
dtype: object
```

[4]:
```python
# check the number of null elements in the dataframe
mining_df.isnull().sum()
```

[4]:
```
% Iron Feed                     0
% Silica Feed                   0
Starch Flow                     0
Amina Flow                      0
Ore Pulp Flow                   0
Ore Pulp pH                     0
Ore Pulp Density                0
Flotation Column 01 Air Flow    0
Flotation Column 02 Air Flow    0
Flotation Column 03 Air Flow    0
Flotation Column 04 Air Flow    0
Flotation Column 05 Air Flow    0
Flotation Column 06 Air Flow    0
Flotation Column 07 Air Flow    0
```

```
Flotation Column 01 Level       0
Flotation Column 02 Level       0
Flotation Column 03 Level       0
Flotation Column 04 Level       0
Flotation Column 05 Level       0
Flotation Column 06 Level       0
Flotation Column 07 Level       0
% Iron Concentrate              0
% Silica Concentrate            0
dtype: int64
```

[5]:
```
mining_df.describe() # Performs different mathematical operation done on the dataset
```

[5]:
```
          % Iron Feed  % Silica Feed    Starch Flow    Amina Flow  \
count  245700.000000  245700.000000  245700.000000  245700.000000
mean       56.294974      14.651438    2869.241181     488.144186
std         5.158958       6.808961    1187.990184      90.736360
min        42.740000       1.310000       0.074147     241.699632
25%        52.670000       8.940000    2073.322500     432.204667
50%        56.080000      13.850000    2994.311667     504.510667
75%        59.720000      19.600000    3712.951667     553.479083
max        65.780000      33.400000    6295.130657     739.422405

        Ore Pulp Flow    Ore Pulp pH  Ore Pulp Density  \
count  245700.000000  245700.000000     245700.000000
mean      397.577332       9.767534          1.680348
std         9.468496       0.387036          0.069213
min       376.272600       8.753370          1.519829
25%       395.212583       9.527158          1.647197
50%       399.354833       9.797963          1.697560
75%       402.458750      10.037833          1.728257
max       418.625439      10.808046          1.853229

       Flotation Column 01 Air Flow  Flotation Column 02 Air Flow  \
count                 245700.000000                 245700.000000
mean                     280.166032                    277.172893
std                       29.616570                     29.936823
min                      175.666333                    175.923177
25%                      250.268667                    250.367333
50%                      299.418000                    297.433000
75%                      300.127333                    300.435000
max                      372.387588                    369.550000

       Flotation Column 03 Air Flow  …  Flotation Column 07 Air Flow  \
count                 245700.000000  …                 245700.000000
mean                     281.097236  …                    290.774336
```

std	28.537193	…	28.158596
min	176.471917	…	186.074077
25%	250.693667	…	263.524333
50%	299.048333	…	299.350833
75%	300.308667	…	301.239667
max	359.948635	…	370.190800

	Flotation Column 01 Level	Flotation Column 02 Level	\
count	245700.000000	245700.000000	
mean	520.242050	522.648563	
std	130.389539	127.450562	
min	149.451600	211.266111	
25%	413.516320	442.291000	
50%	492.971167	496.380667	
75%	594.960083	595.989167	
max	862.197932	828.593000	

	Flotation Column 03 Level	Flotation Column 04 Level	\
count	245700.000000	245700.000000	
mean	531.355055	420.306805	
std	150.614529	90.566437	
min	126.352031	162.293185	
25%	410.134583	356.440167	
50%	494.859500	410.511667	
75%	601.060000	486.533417	
max	886.820204	680.019967	

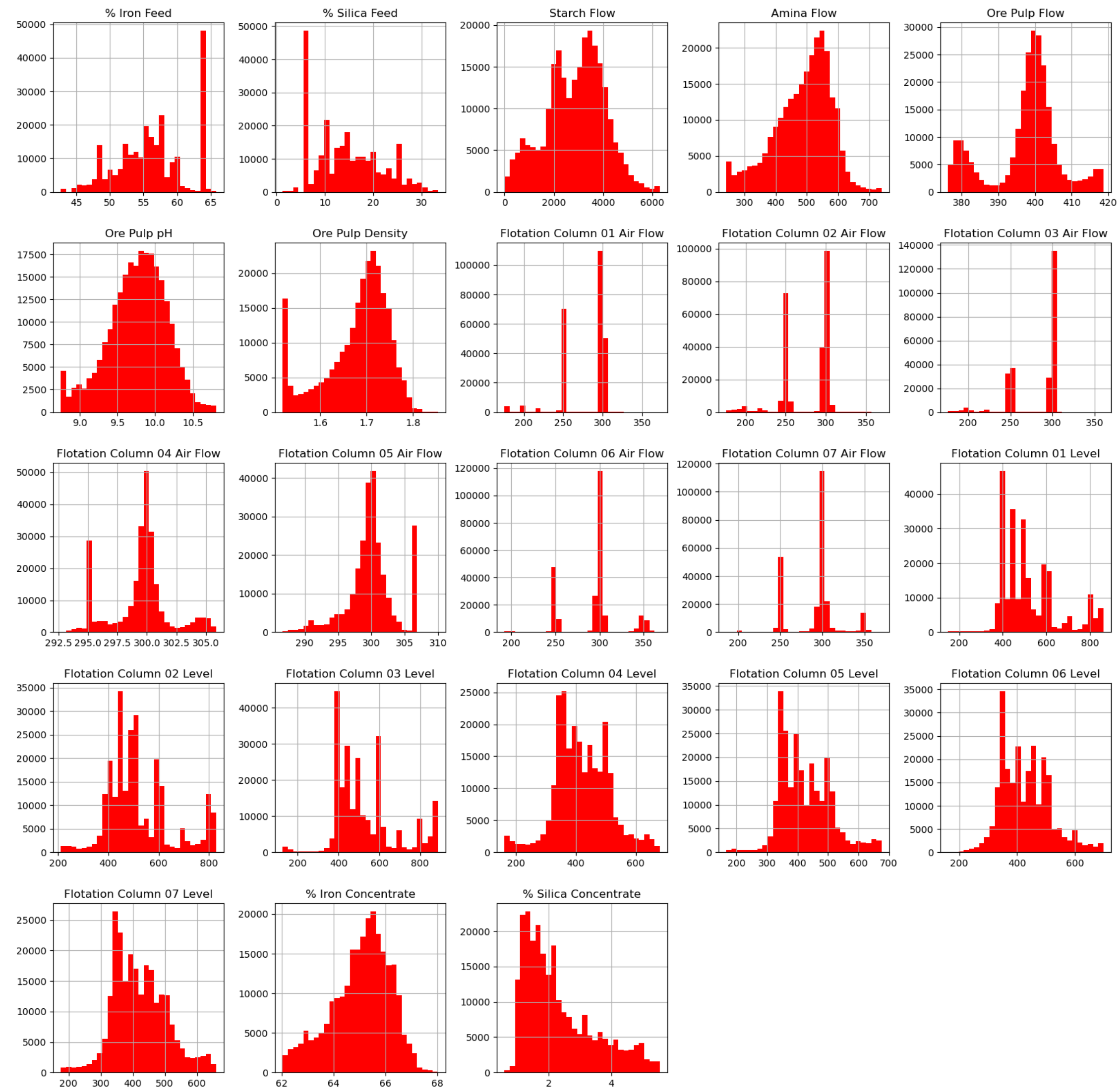
	Flotation Column 05 Level	Flotation Column 06 Level	\
count	245700.000000	245700.000000	
mean	425.237994	429.927646	
std	83.601851	85.320602	
min	167.139620	161.485667	
25%	357.074583	358.078583	
50%	408.022833	419.931167	
75%	485.580833	490.725500	
max	675.571459	698.621871	

	Flotation Column 07 Level	% Iron Concentrate	% Silica Concentrate
count	245700.000000	245700.000000	245700.000000
mean	421.006767	65.049435	2.327228
std	83.736727	1.118479	1.125623
min	175.908240	62.050000	0.600000
25%	356.567833	64.370000	1.440000
50%	410.043333	65.210000	2.000000
75%	475.922283	65.860000	3.010000
max	659.618696	68.010000	5.530000

```
[8 rows x 23 columns]
```

4 TASK #3: PERFORM DATA VISUALIZATION

[6]:
```python
mining_df.hist(bins = 30, figsize = (20, 20), color = 'r')
plt.show()
```



[7]:
```python
# Obtain the correlation matrix
mining_df.corr()
```

[7]:
```
                                   % Iron Feed  % Silica Feed  Starch Flow  \
% Iron Feed                           1.000000      -0.971837     0.028644
```

```
% Silica Feed                 -0.971837       1.000000    -0.016489
Starch Flow                    0.028644      -0.016489     1.000000
Amina Flow                    -0.088595       0.079065     0.261934
Ore Pulp Flow                  0.164676      -0.154508     0.044926
Ore Pulp pH                    0.007496       0.007953     0.269135
Ore Pulp Density              -0.125583       0.104567     0.231530
Flotation Column 01 Air Flow  -0.062448       0.091833     0.195538
Flotation Column 02 Air Flow  -0.170716       0.183953     0.192244
Flotation Column 03 Air Flow  -0.062318       0.084325     0.187602
Flotation Column 04 Air Flow  -0.141811       0.166009     0.098634
Flotation Column 05 Air Flow   0.124844      -0.134294    -0.093094
Flotation Column 06 Air Flow  -0.043084       0.068399     0.223137
Flotation Column 07 Air Flow  -0.016722       0.046193     0.202817
Flotation Column 01 Level      0.029619      -0.038417    -0.302567
Flotation Column 02 Level      0.020928      -0.037113    -0.286336
Flotation Column 03 Level      0.015189      -0.022653    -0.309177
Flotation Column 04 Level      0.016346      -0.034534    -0.013037
Flotation Column 05 Level      0.023571      -0.044270    -0.057417
Flotation Column 06 Level      0.003910      -0.020568    -0.077105
Flotation Column 07 Level      0.026100      -0.047079    -0.039584
% Iron Concentrate             0.055526      -0.044292     0.071303
% Silica Concentrate          -0.077108       0.072778    -0.068008

                              Amina Flow  Ore Pulp Flow  Ore Pulp pH  \
% Iron Feed                    -0.088595       0.164676     0.007496
% Silica Feed                   0.079065      -0.154508     0.007953
Starch Flow                     0.261934       0.044926     0.269135
Amina Flow                      1.000000       0.219162     0.124254
Ore Pulp Flow                   0.219162       1.000000     0.121256
Ore Pulp pH                     0.124254       0.121256     1.000000
Ore Pulp Density                0.659461       0.012571     0.115073
Flotation Column 01 Air Flow    0.111994      -0.141215     0.163433
Flotation Column 02 Air Flow    0.124490      -0.122322     0.146264
Flotation Column 03 Air Flow    0.102632      -0.144232     0.193887
Flotation Column 04 Air Flow    0.016685      -0.164685     0.155388
Flotation Column 05 Air Flow   -0.045642       0.176281    -0.144273
Flotation Column 06 Air Flow    0.122063      -0.164935     0.115448
Flotation Column 07 Air Flow    0.081561      -0.146904     0.178322
Flotation Column 01 Level      -0.270753      -0.008877    -0.166525
Flotation Column 02 Level      -0.236452      -0.008335    -0.154391
Flotation Column 03 Level      -0.288484      -0.055890    -0.140251
Flotation Column 04 Level      -0.183739      -0.010212     0.046933
Flotation Column 05 Level      -0.215334       0.002158     0.013095
Flotation Column 06 Level      -0.209496      -0.018051    -0.079157
Flotation Column 07 Level      -0.171384       0.017185     0.019193
% Iron Concentrate             -0.111045       0.085098     0.192705
% Silica Concentrate            0.157644       0.008519    -0.147550
```

	Ore Pulp Density	Flotation Column 01 Air Flow \
% Iron Feed	-0.125583	-0.062448
% Silica Feed	0.104567	0.091833
Starch Flow	0.231530	0.195538
Amina Flow	0.659461	0.111994
Ore Pulp Flow	0.012571	-0.141215
Ore Pulp pH	0.115073	0.163433
Ore Pulp Density	1.000000	-0.004695
Flotation Column 01 Air Flow	-0.004695	1.000000
Flotation Column 02 Air Flow	0.026413	0.854298
Flotation Column 03 Air Flow	-0.031237	0.955256
Flotation Column 04 Air Flow	-0.099155	0.488605
Flotation Column 05 Air Flow	0.045261	-0.209335
Flotation Column 06 Air Flow	-0.030314	0.670130
Flotation Column 07 Air Flow	-0.052655	0.658199
Flotation Column 01 Level	-0.112201	-0.443870
Flotation Column 02 Level	-0.093542	-0.428294
Flotation Column 03 Level	-0.149823	-0.372533
Flotation Column 04 Level	-0.089066	-0.062665
Flotation Column 05 Level	-0.142675	-0.092254
Flotation Column 06 Level	-0.105346	-0.125196
Flotation Column 07 Level	-0.087454	-0.112074
% Iron Concentrate	0.015116	0.099239
% Silica Concentrate	0.047413	-0.219698

	Flotation Column 02 Air Flow \
% Iron Feed	-0.170716
% Silica Feed	0.183953
Starch Flow	0.192244
Amina Flow	0.124490
Ore Pulp Flow	-0.122322
Ore Pulp pH	0.146264
Ore Pulp Density	0.026413
Flotation Column 01 Air Flow	0.854298
Flotation Column 02 Air Flow	1.000000
Flotation Column 03 Air Flow	0.866927
Flotation Column 04 Air Flow	0.449124
Flotation Column 05 Air Flow	-0.188583
Flotation Column 06 Air Flow	0.598969
Flotation Column 07 Air Flow	0.592888
Flotation Column 01 Level	-0.437313
Flotation Column 02 Level	-0.428303
Flotation Column 03 Level	-0.368913
Flotation Column 04 Level	-0.075833
Flotation Column 05 Level	-0.112465
Flotation Column 06 Level	-0.124478

```
Flotation Column 07 Level                         -0.114693
% Iron Concentrate                                 0.059709
% Silica Concentrate                              -0.169029

                              Flotation Column 03 Air Flow  …  \
% Iron Feed                                       -0.062318  …
% Silica Feed                                      0.084325  …
Starch Flow                                        0.187602  …
Amina Flow                                         0.102632  …
Ore Pulp Flow                                     -0.144232  …
Ore Pulp pH                                        0.193887  …
Ore Pulp Density                                  -0.031237  …
Flotation Column 01 Air Flow                       0.955256  …
Flotation Column 02 Air Flow                       0.866927  …
Flotation Column 03 Air Flow                       1.000000  …
Flotation Column 04 Air Flow                       0.505138  …
Flotation Column 05 Air Flow                      -0.231556  …
Flotation Column 06 Air Flow                       0.667008  …
Flotation Column 07 Air Flow                       0.661261  …
Flotation Column 01 Level                         -0.443454  …
Flotation Column 02 Level                         -0.423836  …
Flotation Column 03 Level                         -0.365800  …
Flotation Column 04 Level                         -0.047216  …
Flotation Column 05 Level                         -0.064355  …
Flotation Column 06 Level                         -0.105673  …
Flotation Column 07 Level                         -0.086448  …
% Iron Concentrate                                 0.100888  …
% Silica Concentrate                              -0.219593  …

                              Flotation Column 07 Air Flow  \
% Iron Feed                                       -0.016722
% Silica Feed                                      0.046193
Starch Flow                                        0.202817
Amina Flow                                         0.081561
Ore Pulp Flow                                     -0.146904
Ore Pulp pH                                        0.178322
Ore Pulp Density                                  -0.052655
Flotation Column 01 Air Flow                       0.658199
Flotation Column 02 Air Flow                       0.592888
Flotation Column 03 Air Flow                       0.661261
Flotation Column 04 Air Flow                       0.579148
Flotation Column 05 Air Flow                      -0.267964
Flotation Column 06 Air Flow                       0.878873
Flotation Column 07 Air Flow                       1.000000
Flotation Column 01 Level                         -0.380922
Flotation Column 02 Level                         -0.359771
Flotation Column 03 Level                         -0.349285
```

```
Flotation Column 04 Level                        -0.178906
Flotation Column 05 Level                        -0.215236
Flotation Column 06 Level                        -0.262289
Flotation Column 07 Level                        -0.244693
% Iron Concentrate                               -0.001676
% Silica Concentrate                             -0.073707

                              Flotation Column 01 Level  \
% Iron Feed                                     0.029619
% Silica Feed                                  -0.038417
Starch Flow                                    -0.302567
Amina Flow                                     -0.270753
Ore Pulp Flow                                  -0.008877
Ore Pulp pH                                    -0.166525
Ore Pulp Density                               -0.112201
Flotation Column 01 Air Flow                   -0.443870
Flotation Column 02 Air Flow                   -0.437313
Flotation Column 03 Air Flow                   -0.443454
Flotation Column 04 Air Flow                   -0.299247
Flotation Column 05 Air Flow                    0.221151
Flotation Column 06 Air Flow                   -0.390170
Flotation Column 07 Air Flow                   -0.380922
Flotation Column 01 Level                       1.000000
Flotation Column 02 Level                       0.724455
Flotation Column 03 Level                       0.729648
Flotation Column 04 Level                       0.265465
Flotation Column 05 Level                       0.332957
Flotation Column 06 Level                       0.316801
Flotation Column 07 Level                       0.309139
% Iron Concentrate                             -0.014195
% Silica Concentrate                            0.017315

                              Flotation Column 02 Level  \
% Iron Feed                                     0.020928
% Silica Feed                                  -0.037113
Starch Flow                                    -0.286336
Amina Flow                                     -0.236452
Ore Pulp Flow                                  -0.008335
Ore Pulp pH                                    -0.154391
Ore Pulp Density                               -0.093542
Flotation Column 01 Air Flow                   -0.428294
Flotation Column 02 Air Flow                   -0.428303
Flotation Column 03 Air Flow                   -0.423836
Flotation Column 04 Air Flow                   -0.294471
Flotation Column 05 Air Flow                    0.214655
Flotation Column 06 Air Flow                   -0.349831
Flotation Column 07 Air Flow                   -0.359771
```

```
Flotation Column 01 Level                       0.724455
Flotation Column 02 Level                       1.000000
Flotation Column 03 Level                       0.658129
Flotation Column 04 Level                       0.195648
Flotation Column 05 Level                       0.272166
Flotation Column 06 Level                       0.302109
Flotation Column 07 Level                       0.239788
% Iron Concentrate                             -0.026083
% Silica Concentrate                            0.031137

                              Flotation Column 03 Level  \
% Iron Feed                                     0.015189
% Silica Feed                                  -0.022653
Starch Flow                                    -0.309177
Amina Flow                                     -0.288484
Ore Pulp Flow                                  -0.055890
Ore Pulp pH                                    -0.140251
Ore Pulp Density                               -0.149823
Flotation Column 01 Air Flow                   -0.372533
Flotation Column 02 Air Flow                   -0.368913
Flotation Column 03 Air Flow                   -0.365800
Flotation Column 04 Air Flow                   -0.231198
Flotation Column 05 Air Flow                    0.183885
Flotation Column 06 Air Flow                   -0.377330
Flotation Column 07 Air Flow                   -0.349285
Flotation Column 01 Level                       0.729648
Flotation Column 02 Level                       0.658129
Flotation Column 03 Level                       1.000000
Flotation Column 04 Level                       0.247080
Flotation Column 05 Level                       0.327009
Flotation Column 06 Level                       0.317418
Flotation Column 07 Level                       0.268676
% Iron Concentrate                             -0.022000
% Silica Concentrate                            0.014250

                              Flotation Column 04 Level  \
% Iron Feed                                     0.016346
% Silica Feed                                  -0.034534
Starch Flow                                    -0.013037
Amina Flow                                     -0.183739
Ore Pulp Flow                                  -0.010212
Ore Pulp pH                                     0.046933
Ore Pulp Density                               -0.089066
Flotation Column 01 Air Flow                   -0.062665
Flotation Column 02 Air Flow                   -0.075833
Flotation Column 03 Air Flow                   -0.047216
Flotation Column 04 Air Flow                   -0.146936
```

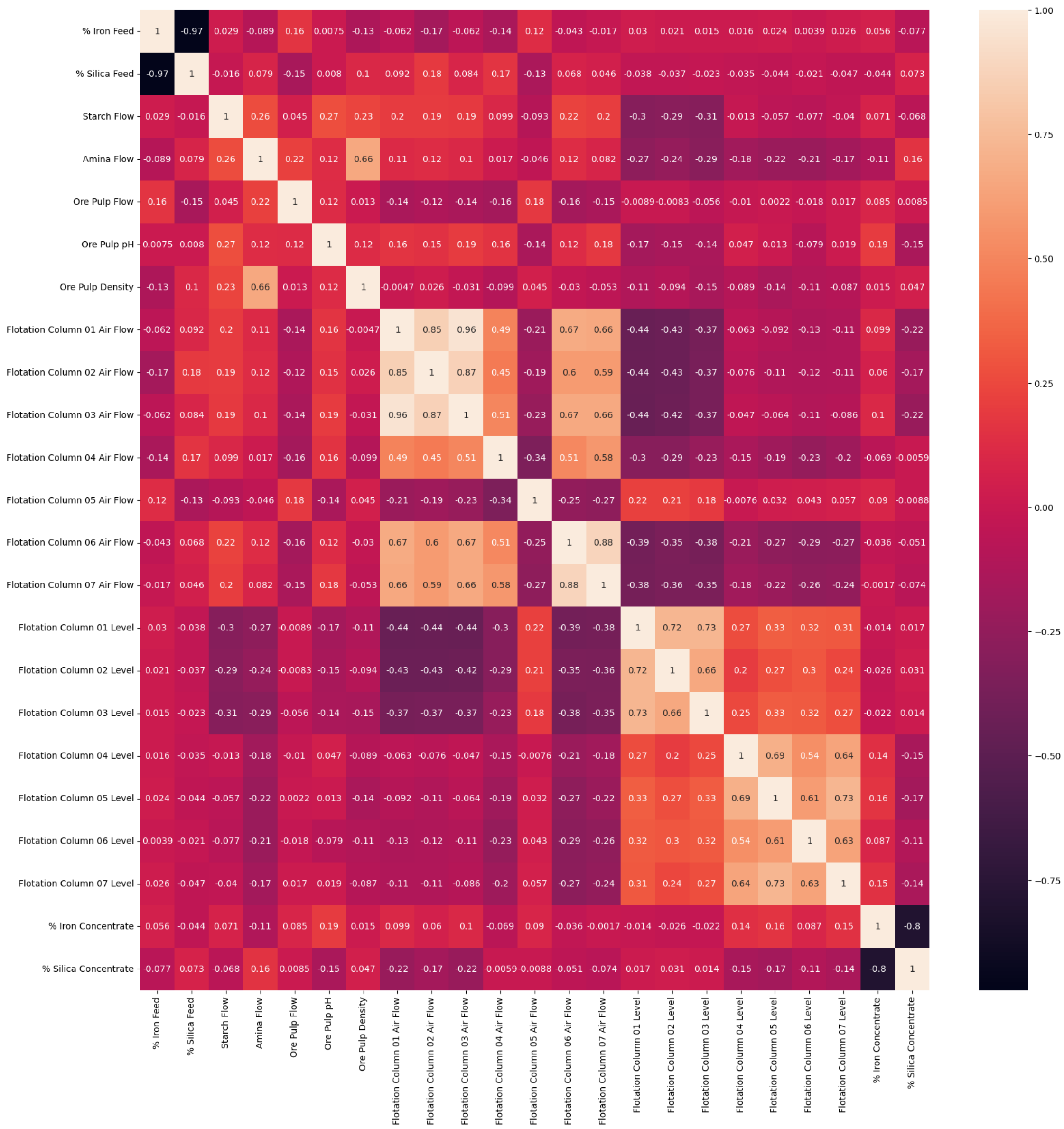
```
Flotation Column 05 Air Flow                     -0.007646
Flotation Column 06 Air Flow                     -0.212949
Flotation Column 07 Air Flow                     -0.178906
Flotation Column 01 Level                         0.265465
Flotation Column 02 Level                         0.195648
Flotation Column 03 Level                         0.247080
Flotation Column 04 Level                         1.000000
Flotation Column 05 Level                         0.692128
Flotation Column 06 Level                         0.544777
Flotation Column 07 Level                         0.641676
% Iron Concentrate                                0.138780
% Silica Concentrate                             -0.151411

                              Flotation Column 05 Level  \
% Iron Feed                                       0.023571
% Silica Feed                                    -0.044270
Starch Flow                                      -0.057417
Amina Flow                                       -0.215334
Ore Pulp Flow                                     0.002158
Ore Pulp pH                                       0.013095
Ore Pulp Density                                 -0.142675
Flotation Column 01 Air Flow                     -0.092254
Flotation Column 02 Air Flow                     -0.112465
Flotation Column 03 Air Flow                     -0.064355
Flotation Column 04 Air Flow                     -0.189317
Flotation Column 05 Air Flow                      0.032117
Flotation Column 06 Air Flow                     -0.269520
Flotation Column 07 Air Flow                     -0.215236
Flotation Column 01 Level                         0.332957
Flotation Column 02 Level                         0.272166
Flotation Column 03 Level                         0.327009
Flotation Column 04 Level                         0.692128
Flotation Column 05 Level                         1.000000
Flotation Column 06 Level                         0.605338
Flotation Column 07 Level                         0.729287
% Iron Concentrate                                0.162523
% Silica Concentrate                             -0.170989

                              Flotation Column 06 Level  \
% Iron Feed                                       0.003910
% Silica Feed                                    -0.020568
Starch Flow                                      -0.077105
Amina Flow                                       -0.209496
Ore Pulp Flow                                    -0.018051
Ore Pulp pH                                      -0.079157
Ore Pulp Density                                 -0.105346
Flotation Column 01 Air Flow                     -0.125196
```

```
Flotation Column 02 Air Flow                    -0.124478
Flotation Column 03 Air Flow                    -0.105673
Flotation Column 04 Air Flow                    -0.232610
Flotation Column 05 Air Flow                     0.043384
Flotation Column 06 Air Flow                    -0.285898
Flotation Column 07 Air Flow                    -0.262289
Flotation Column 01 Level                        0.316801
Flotation Column 02 Level                        0.302109
Flotation Column 03 Level                        0.317418
Flotation Column 04 Level                        0.544777
Flotation Column 05 Level                        0.605338
Flotation Column 06 Level                        1.000000
Flotation Column 07 Level                        0.633726
% Iron Concentrate                               0.086533
% Silica Concentrate                            -0.107805

                              Flotation Column 07 Level  % Iron Concentrate  \
% Iron Feed                                    0.026100            0.055526
% Silica Feed                                 -0.047079           -0.044292
Starch Flow                                   -0.039584            0.071303
Amina Flow                                    -0.171384           -0.111045
Ore Pulp Flow                                  0.017185            0.085098
Ore Pulp pH                                    0.019193            0.192705
Ore Pulp Density                              -0.087454            0.015116
Flotation Column 01 Air Flow                  -0.112074            0.099239
Flotation Column 02 Air Flow                  -0.114693            0.059709
Flotation Column 03 Air Flow                  -0.086448            0.100888
Flotation Column 04 Air Flow                  -0.197104           -0.068653
Flotation Column 05 Air Flow                   0.057331            0.090488
Flotation Column 06 Air Flow                  -0.272827           -0.036177
Flotation Column 07 Air Flow                  -0.244693           -0.001676
Flotation Column 01 Level                      0.309139           -0.014195
Flotation Column 02 Level                      0.239788           -0.026083
Flotation Column 03 Level                      0.268676           -0.022000
Flotation Column 04 Level                      0.641676            0.138780
Flotation Column 05 Level                      0.729287            0.162523
Flotation Column 06 Level                      0.633726            0.086533
Flotation Column 07 Level                      1.000000            0.148419
% Iron Concentrate                             0.148419            1.000000
% Silica Concentrate                          -0.143248           -0.800517

                              % Silica Concentrate
% Iron Feed                              -0.077108
% Silica Feed                             0.072778
Starch Flow                              -0.068008
Amina Flow                                0.157644
Ore Pulp Flow                             0.008519
```

```
Ore Pulp pH                                  -0.147550
Ore Pulp Density                              0.047413
Flotation Column 01 Air Flow                 -0.219698
Flotation Column 02 Air Flow                 -0.169029
Flotation Column 03 Air Flow                 -0.219593
Flotation Column 04 Air Flow                 -0.005890
Flotation Column 05 Air Flow                 -0.008821
Flotation Column 06 Air Flow                 -0.050805
Flotation Column 07 Air Flow                 -0.073707
Flotation Column 01 Level                     0.017315
Flotation Column 02 Level                     0.031137
Flotation Column 03 Level                     0.014250
Flotation Column 04 Level                    -0.151411
Flotation Column 05 Level                    -0.170989
Flotation Column 06 Level                    -0.107805
Flotation Column 07 Level                    -0.143248
% Iron Concentrate                           -0.800517
% Silica Concentrate                          1.000000

[23 rows x 23 columns]
```

[8]:
```
plt.figure(figsize = (20, 20))
sns.heatmap(mining_df.corr(), annot = True)
# From this pair plot, we can infer that there is a relationship between iron
 ↪feed and silica feed
# Also, a relationship between silica concentrate and iron concentrate.
```

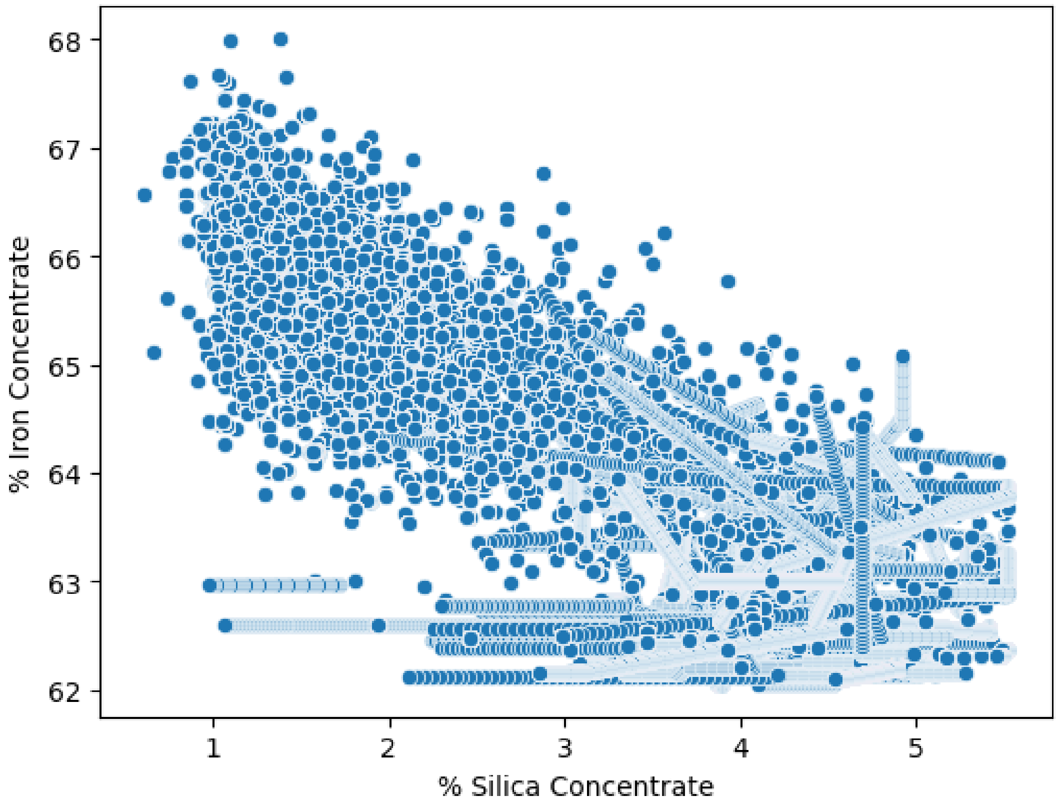
[8]: <AxesSubplot:>



- Plotting the scatterplot between % Silica Concentrate and Iron Concentrate and try to relate to the correlation matrix.

[9]:
```
sns.scatterplot(data = mining_df, x= '% Silica Concentrate', y = '% Iron Concentrate')
```

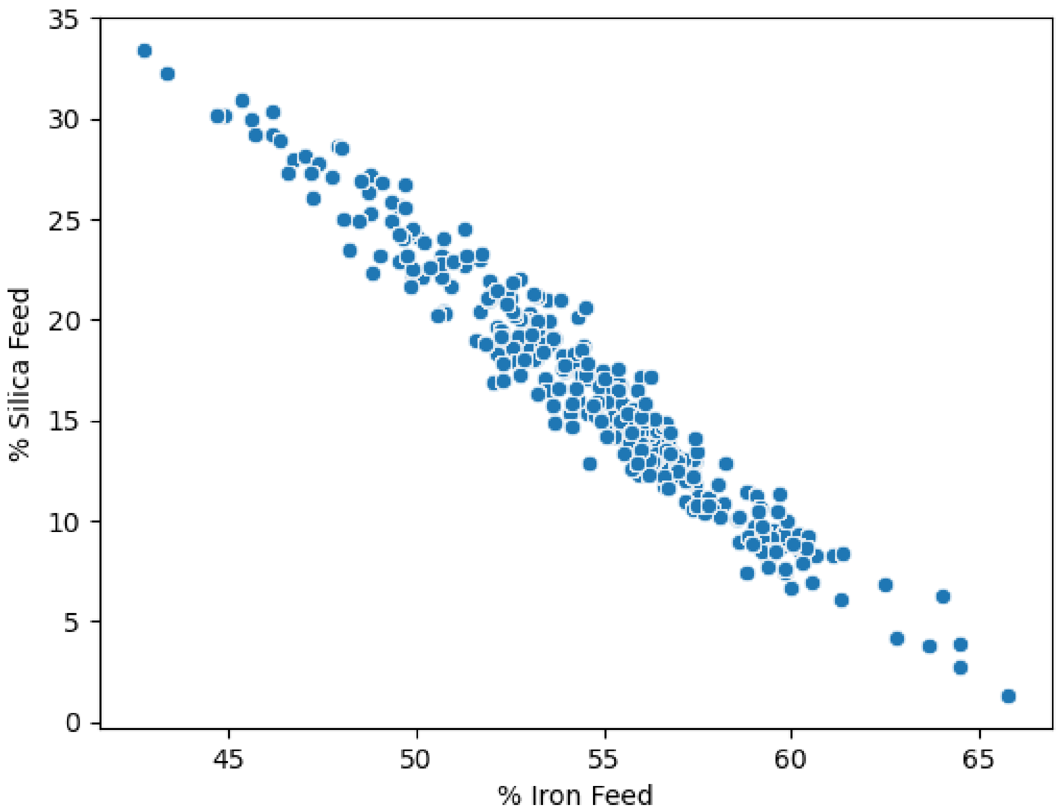
[9]: `<AxesSubplot:xlabel='% Silica Concentrate', ylabel='% Iron Concentrate'>`



- Plotting the scatterplot between % Iron Feed and % Silica Feed and try to relate to the correlation matrix.

[10]:
```
sns.scatterplot(data = mining_df, x = '% Iron Feed', y = '% Silica Feed')
```

[10]: <AxesSubplot:xlabel='% Iron Feed', ylabel='% Silica Feed'>



5 TASK #4: PREPARE THE DATA BEFORE MODEL TRAINING

[11]:
```
df_iron = mining_df.drop(columns = '% Silica Concentrate')
df_iron_target = mining_df['% Silica Concentrate']
```

[12]:
```
df_iron.shape
```

[12]: (245700, 22)

[13]:
```
df_iron_target.shape
```

[13]: (245700,)

[14]:
```
df_iron = np.array(df_iron)
df_iron_target = np.array(df_iron_target)
```

[15]:
```
# reshaping the array
df_iron_target = df_iron_target.reshape(-1,1)
df_iron_target.shape
```

[15]: (245700, 1)

[16]:
```
# scaling the data before feeding the model
from sklearn.preprocessing import StandardScaler, MinMaxScaler
scaler_x = StandardScaler()
X = scaler_x.fit_transform(df_iron)

scaler_y = StandardScaler()
y = scaler_y.fit_transform(df_iron_target)
```

[17]:
```
# spliting the data in to test and train sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
```

- The dataset was successful split successfully into train and test set

[18]:
```
X_train.shape
```

[18]: (196560, 22)

[19]:
```
X_test.shape
```

[19]: (49140, 22)

[20]:
```
y_train.shape
```

[20]: (196560, 1)

[21]:
```
y_test.shape
```

[21]: (49140, 1)

6 TASK #5: TRAIN AND EVALUATE A LINEAR REGRESSION MODEL

[22]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, accuracy_score
```

[23]:
```
LinearRegression_model = LinearRegression()
LinearRegression_model.fit(X_train, y_train)
```

[23]: LinearRegression()

[24]:
```
accuracy_LinearRegression = LinearRegression_model.score(X_test, y_test)
accuracy_LinearRegression
```

[24]: 0.6830372414644894

7 TASK #6: TRAIN AND EVALUATE A DECISION TREE AND RANDOM FOREST MODELS

8 Decision Tree Model

[25]:
```
# Decision tree builds regression or classification models in the form of a␣
↪tree structure.
# Decision tree breaks down a dataset into smaller subsets while at the same␣
↪time an associated decision tree is incrementally developed.
# The final result is a tree with decision nodes and leaf nodes.

from sklearn.tree import DecisionTreeRegressor

DecisionTree_model = DecisionTreeRegressor()
DecisionTree_model.fit(X_train, y_train)
```

[25]:
```
DecisionTreeRegressor()
```

[26]:
```
accuracy_DecisionTree = DecisionTree_model.score(X_test, y_test)
accuracy_DecisionTree
```

[26]:
```
0.9817471423741311
```

9 Random Forest Model(Ensemble Model)

[27]:
```
# Many decision Trees make up a random forest model which is an ensemble model.
# Predictions made by each decision tree are averaged to get the prediction of␣
↪random forest model.
# A random forest regressor fits a number of classifying decision trees on␣
↪various  sub-samples of the dataset and uses averaging to improve the␣
↪predictive accuracy and control over-fitting.
```

- Training a Random Forest Regressor Model with n_estimators = 100 and max_depth of 10

[28]:
```
from sklearn.ensemble import RandomForestRegressor

RandomForest_model = RandomForestRegressor(n_estimators = 100, max_depth = 10)
RandomForest_model.fit(X_train, y_train.ravel())

accuracy_RandomForest = RandomForest_model.score(X_test, y_test)
accuracy_RandomForest
```

[28]:
```
0.8922988079533989
```

10 TASK #7: TRAIN AN ARTIFICIAL NEURAL NETWORK TO PERFORM REGRESSION TASK

[29]:
```
pip install tensorflow
```

```
Requirement already satisfied: tensorflow in c:\users\abdul
mateen\anaconda3\lib\site-packages (2.11.0)
Requirement already satisfied: tensorflow-intel==2.11.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow) (2.11.0)
Requirement already satisfied: keras<2.12,>=2.11.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(2.11.0)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in
c:\users\abdul mateen\anaconda3\lib\site-packages (from tensorflow-
intel==2.11.0->tensorflow) (0.30.0)
Requirement already satisfied: protobuf<3.20,>=3.9.2 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(3.19.6)
Requirement already satisfied: tensorflow-estimator<2.12,>=2.11.0 in
c:\users\abdul mateen\anaconda3\lib\site-packages (from tensorflow-
intel==2.11.0->tensorflow) (2.11.0)
Requirement already satisfied: wrapt>=1.11.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.14.1)
Requirement already satisfied: flatbuffers>=2.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(23.1.21)
Requirement already satisfied: opt-einsum>=2.3.2 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(3.3.0)
Requirement already satisfied: tensorboard<2.12,>=2.11 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(2.11.2)
Requirement already satisfied: gast<=0.4.0,>=0.2.1 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(0.4.0)
Requirement already satisfied: setuptools in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(63.4.1)
Requirement already satisfied: termcolor>=1.1.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(2.2.0)
Requirement already satisfied: six>=1.12.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.16.0)
Requirement already satisfied: libclang>=13.0.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(15.0.6.1)
```

```
Requirement already satisfied: packaging in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(21.3)
Requirement already satisfied: google-pasta>=0.1.1 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(0.2.0)
Requirement already satisfied: typing-extensions>=3.6.6 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(4.3.0)
Requirement already satisfied: absl-py>=1.0.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.4.0)
Requirement already satisfied: h5py>=2.9.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(3.7.0)
Requirement already satisfied: numpy>=1.20 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.21.5)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.51.1)
Requirement already satisfied: astunparse>=1.6.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->tensorflow)
(1.6.3)
Requirement already satisfied: wheel<1.0,>=0.23.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from astunparse>=1.6.0->tensorflow-
intel==2.11.0->tensorflow) (0.37.1)
Requirement already satisfied: markdown>=2.6.8 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (3.3.4)
Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in
c:\users\abdul mateen\anaconda3\lib\site-packages (from
tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow) (0.6.1)
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (1.8.1)
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in
c:\users\abdul mateen\anaconda3\lib\site-packages (from
tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow) (0.4.6)
Requirement already satisfied: requests<3,>=2.21.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (2.28.1)
Requirement already satisfied: werkzeug>=1.0.1 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (2.0.3)
Requirement already satisfied: google-auth<3,>=1.6.3 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (2.16.0)
```

```
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from packaging->tensorflow-
intel==2.11.0->tensorflow) (3.0.9)
Requirement already satisfied: rsa<5,>=3.1.4 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from google-
auth<3,>=1.6.3->tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow)
(4.9)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from google-
auth<3,>=1.6.3->tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow)
(5.3.0)
Requirement already satisfied: pyasn1-modules>=0.2.1 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from google-
auth<3,>=1.6.3->tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow)
(0.2.8)
Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from google-auth-
oauthlib<0.5,>=0.4.1->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (1.3.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from
requests<3,>=2.21.0->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (1.26.11)
Requirement already satisfied: idna<4,>=2.5 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from
requests<3,>=2.21.0->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from
requests<3,>=2.21.0->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (2022.12.7)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from
requests<3,>=2.21.0->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (2.0.4)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from pyasn1-modules>=0.2.1->google-
auth<3,>=1.6.3->tensorboard<2.12,>=2.11->tensorflow-intel==2.11.0->tensorflow)
(0.4.8)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\abdul
mateen\anaconda3\lib\site-packages (from requests-oauthlib>=0.7.0->google-auth-
oauthlib<0.5,>=0.4.1->tensorboard<2.12,>=2.11->tensorflow-
intel==2.11.0->tensorflow) (3.2.2)
Note: you may need to restart the kernel to use updated packages.
```

[30]:
```python
import tensorflow as tf
from tensorflow import keras
```

```python
from tensorflow.keras.layers import Dense, Activation, Dropout
from tensorflow.keras.optimizers import Adam

optimizer = Adam(learning_rate=0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon =␣
 ↪1e-07, amsgrad = False)
ANN_model = keras.Sequential()
ANN_model.add(Dense(250, input_dim = 22,␣
 ↪kernel_initializer='normal',activation='relu'))
ANN_model.add(Dense(500,activation = 'relu'))
ANN_model.add(Dropout(0.1))
ANN_model.add(Dense(1000, activation = 'relu'))
ANN_model.add(Dropout(0.1))
ANN_model.add(Dense(1000, activation = 'relu'))
ANN_model.add(Dropout(0.1))
ANN_model.add(Dense(500, activation = 'relu'))
ANN_model.add(Dropout(0.1))
ANN_model.add(Dense(250, activation = 'relu'))
ANN_model.add(Dropout(0.1))
ANN_model.add(Dense(1, activation = 'linear'))
ANN_model.compile(loss = 'mse', optimizer = 'adam')
ANN_model.summary()
```

```
Model: "sequential"

_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 250)               5750

 dense_1 (Dense)             (None, 500)               125500

 dropout (Dropout)           (None, 500)               0

 dense_2 (Dense)             (None, 1000)              501000

 dropout_1 (Dropout)         (None, 1000)              0

 dense_3 (Dense)             (None, 1000)              1001000

 dropout_2 (Dropout)         (None, 1000)              0

 dense_4 (Dense)             (None, 500)               500500

 dropout_3 (Dropout)         (None, 500)               0

 dense_5 (Dense)             (None, 250)               125250

 dropout_4 (Dropout)         (None, 250)               0
```

```
 dense_6 (Dense)             (None, 1)                 251

=================================================================
Total params: 2,259,251
Trainable params: 2,259,251
Non-trainable params: 0
_________________________________________________________________
```

[31]:
```
history = ANN_model.fit(X_train, y_train, epochs = 5, validation_split = 0.2)
```

```
Epoch 1/5
4914/4914 [==============================] - 196s 40ms/step - loss: 0.2115 -
val_loss: 0.1590
Epoch 2/5
4914/4914 [==============================] - 193s 39ms/step - loss: 0.1460 -
val_loss: 0.1396
Epoch 3/5
4914/4914 [==============================] - 189s 38ms/step - loss: 0.1213 -
val_loss: 0.1019
Epoch 4/5
4914/4914 [==============================] - 175s 36ms/step - loss: 0.1066 -
val_loss: 0.0910
Epoch 5/5
4914/4914 [==============================] - 215s 44ms/step - loss: 0.0968 -
val_loss: 0.0809
```

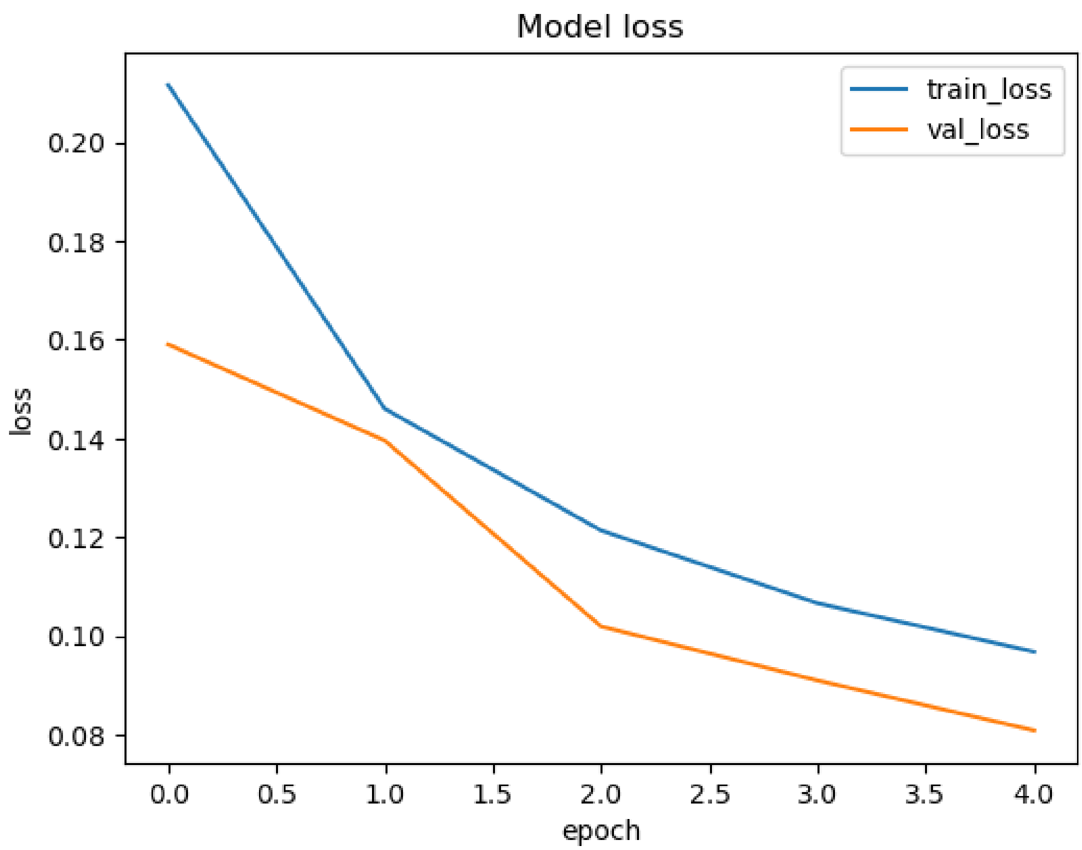
[32]:
```
result = ANN_model.evaluate(X_test, y_test)
accuracy_ANN = 1 - result
print("Accuracy : {}".format(accuracy_ANN))
```

```
1536/1536 [==============================] - 8s 5ms/step - loss: 0.0783
Accuracy : 0.9216900542378426
```

[33]:
```
history.history.keys()
```

[33]:
```
dict_keys(['loss', 'val_loss'])
```

[34]:
```
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train_loss','val_loss'], loc = 'upper right')
plt.show()
```

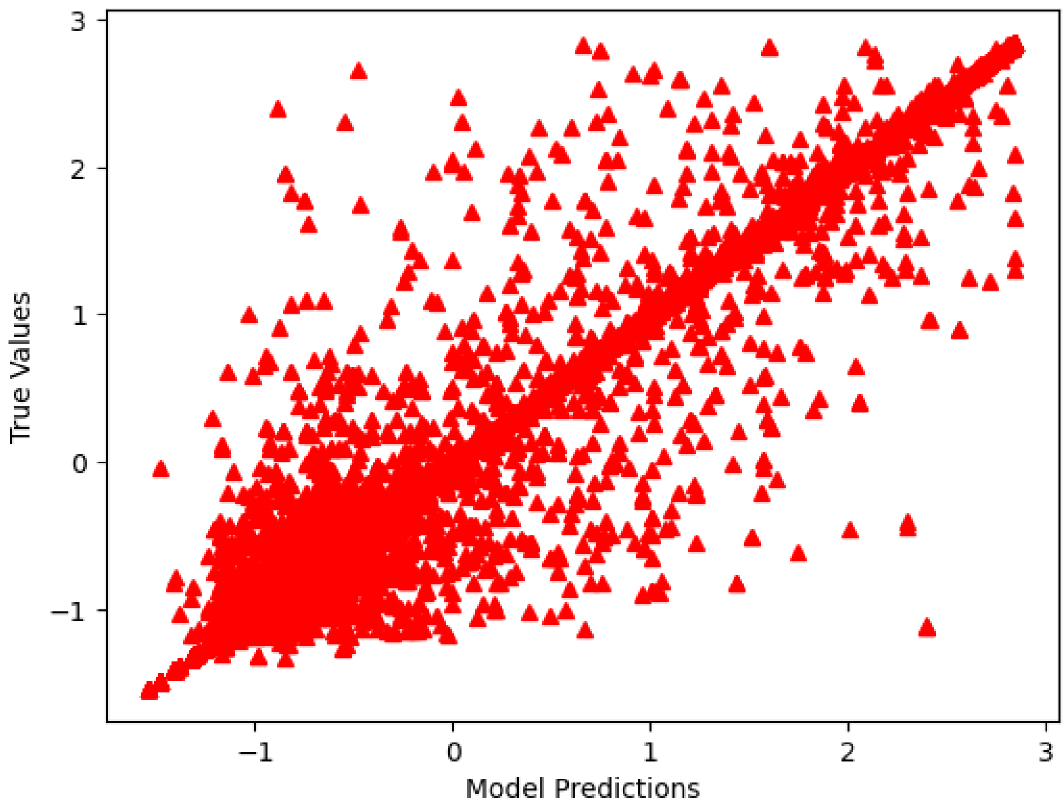


11 TASK #8: COMPARE MODELS AND CALCULATE REGRESSION KPIs

[35]:
```
# From the above results, it can be seen that, decision tree model out-performs␣
↪the other models.
```

[36]:
```
y_predict = DecisionTree_model.predict(X_test)
plt.plot(y_predict, y_test, '^', color = 'r')
plt.xlabel('Model Predictions')
plt.ylabel('True Values')
```

[36]: Text(0, 0.5, 'True Values')




```
[37]: y_test
```

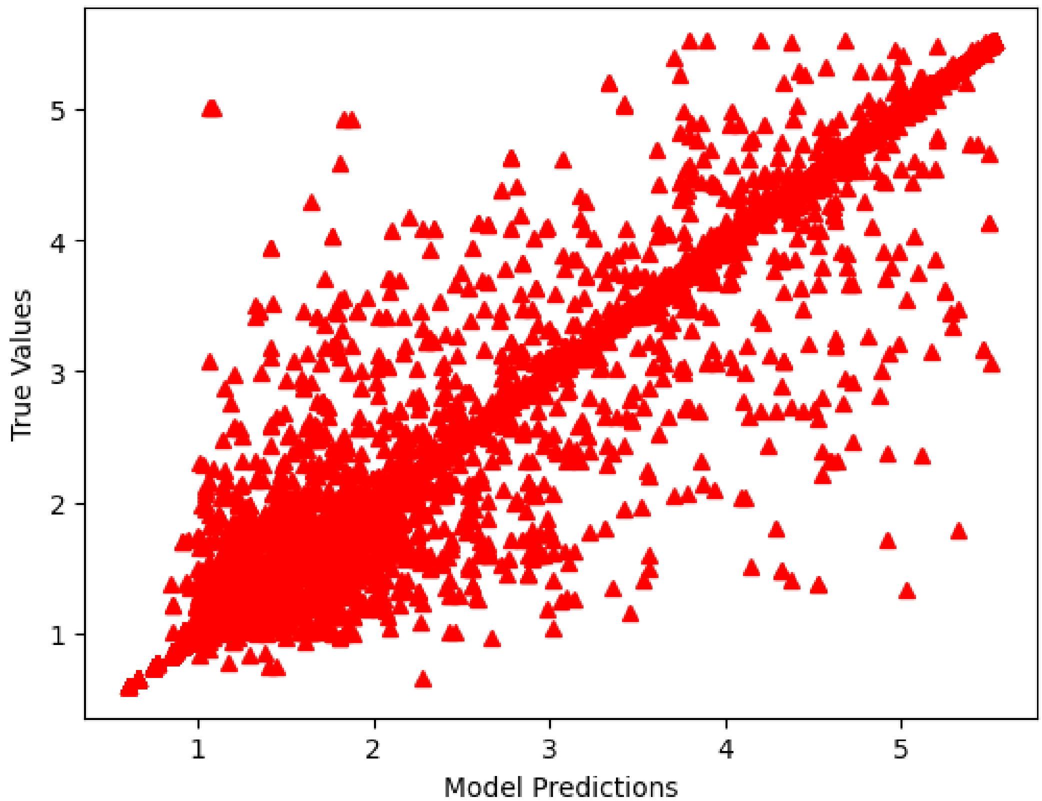
```
[37]: array([[ 2.37448641],
             [-0.17521735],
             [ 1.22845231],
             …,
             [ 0.30451785],
             [-0.44173691],
             [-0.77044437]])
```

```
[38]: y_predict =y_predict.reshape(-1,1)
      y_predict
```

```
[38]: array([[ 2.37448641],
             [-0.17521735],
             [ 1.22845231],
             …,
             [ 0.30451785],
             [-0.34401307],
             [-0.77044437]])
```

[39]:
```python
y_predict_orig = scaler_y.inverse_transform(y_predict)
y_test_orig = scaler_y.inverse_transform(y_test)
plt.plot(y_test_orig, y_predict_orig, "^", color = 'r')
plt.xlabel('Model Predictions')
plt.ylabel('True Values')
```

[39]: Text(0, 0.5, 'True Values')



[40]:
```python
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from math import sqrt

k = X_test.shape[1]
n = len(X_test)
RMSE = float(format(np.sqrt(mean_squared_error(y_test_orig, y_predict_orig)),'.
 ↪3f'))
MSE = mean_squared_error(y_test_orig, y_predict_orig)
MAE = mean_absolute_error(y_test_orig, y_predict_orig)
r2 = r2_score(y_test_orig, y_predict_orig)
adj_r2 = 1-(1-r2)*(n-1)/(n-k-1)
```

```python
print('RMSE =',RMSE, '\nMSE =',MSE, '\nMAE =',MAE, '\nR2 =', r2, '\nAdjusted R2␣
↪=', adj_r2)
```

```
RMSE = 0.152
MSE = 0.02313483450893113
MAE = 0.022792609651125197
R2 = 0.9817471423741311
Adjusted R2 = 0.9817389667349884
```

[ ]:

```python

```