Exploring Solubility Patterns: A Comparative Analysis of Linear Regression and Random Forest Regressor Models

Import padas library

In [65]:
```
import pandas as pd
```

Loading the data

In [66]:
```
df = pd.read_csv('https://raw.githubusercontent.com/dataprofessor/data/master/delaney_solubility_with_descriptors.csv')
```

In [67]:
```
df    # MollogP MolWt NumRotatableBonds AromaticProportion are independent variables        # logS is dependent variable
```

Out[67]:

	MolLogP	MolWt	NumRotatableBonds	AromaticProportion	logS
0	2.59540	167.850	0.0	0.000000	-2.180
1	2.37650	133.405	0.0	0.000000	-2.000
2	2.59380	167.850	1.0	0.000000	-1.740
3	2.02890	133.405	1.0	0.000000	-1.480
4	2.91890	187.375	1.0	0.000000	-3.040
...	...	...	...	...	...
1139	1.98820	287.343	8.0	0.000000	1.144
1140	3.42130	286.114	2.0	0.333333	-4.925
1141	3.60960	308.333	4.0	0.695652	-3.893
1142	2.56214	354.815	3.0	0.521739	-3.790
1143	2.02164	179.219	1.0	0.461538	-2.581

1144 rows × 5 columns

Data preperation as X and Y

In [68]:
```
y = df["logS"]
```

In [69]:
```
y
```

Out[69]:
```
0      -2.180
1      -2.000
2      -1.740
3      -1.480
4      -3.040
        ...
1139    1.144
1140   -4.925
1141   -3.893
1142   -3.790
1143   -2.581
Name: logS, Length: 1144, dtype: float64
```

Removing the 'logS' for x

In [70]:
```
x = df.drop('logS',axis=1)   # axis=1 allow the drop function to work with data to drop the column in column mode
                             # axis=0 allow the drop function to work with data to drop the row in row mode
```

In [71]:
```
x
```

Out[71]:

	MolLogP	MolWt	NumRotatableBonds	AromaticProportion
0	2.59540	167.850	0.0	0.000000
1	2.37650	133.405	0.0	0.000000
2	2.59380	167.850	1.0	0.000000
3	2.02890	133.405	1.0	0.000000
4	2.91890	187.375	1.0	0.000000
...	...	...	...	...
1139	1.98820	287.343	8.0	0.000000
1140	3.42130	286.114	2.0	0.333333
1141	3.60960	308.333	4.0	0.695652
1142	2.56214	354.815	3.0	0.521739
1143	2.02164	179.219	1.0	0.461538

1144 rows × 4 columns

Splitting the data for Training and Testing using Scikit learn library

In [72]:
```
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,random_state=100) # "random state =100" for getting the same data split every time we run the code
```

In [73]:
```
x_train # contains the 80% of the data from 1144 rows
```

Out[73]:

	MolLogP	MolWt	NumRotatableBonds	AromaticProportion
107	3.14280	112.216	5.0	0.000000
378	-2.07850	142.070	0.0	0.000000
529	-0.47730	168.152	0.0	0.000000
546	-0.86740	154.125	0.0	0.000000
320	1.62150	100.161	2.0	0.000000
...	...	...	...	...
802	3.00254	250.301	1.0	0.842105
53	2.13860	82.146	3.0	0.000000
350	5.76304	256.348	0.0	0.900000
79	3.89960	186.339	10.0	0.000000
792	2.52334	310.297	3.0	0.300000

915 rows × 4 columns

In [74]:
```
x_test # contains 20 % of the data from 1144 rows
```

Out[74]:

	MolLogP	MolWt	NumRotatableBonds	AromaticProportion
822	2.91000	172.268	7.0	0.000000
118	7.27400	360.882	1.0	0.666667
347	1.94040	145.161	0.0	0.909091
1123	1.98640	119.378	0.0	0.000000
924	1.70062	108.140	0.0	0.750000
...	...	...	...	...
1114	1.76210	478.513	4.0	0.000000
427	6.32820	276.338	0.0	1.000000
711	0.04430	218.205	5.0	0.000000
4	2.91890	187.375	1.0	0.000000
948	3.56010	318.328	2.0	0.750000

229 rows × 4 columns

Model building

Linear Regression

Importing the model

In [ ]:
```
from sklearn.linear_model import LinearRegression
```

Training the model

In [ ]:
```
lr = LinearRegression()
lr.fit(x_train,y_train) # x_train and y_train are the inputs for building the model
```

Applying the model to make prediction

In [76]:
```
y_lr_train_pred = lr.predict(x_train) # Applying the model to do prediction on the Training dataset (i.e) (X_train) # This is done to evalute tyhe perfomance of the algorithm
y_lr_test_pred = lr.predict(x_test)   # Applying the model to do prediction on the Test dataset (i.e) (X_test)
```

In [77]:
```
y_lr_train_pred
```

Out[77]:
```
array([-2.83668969,  0.84208401, -0.50930302, -0.13058053, -1.6269208 ,
       -2.86103037, -2.32208333, -4.16684323, -5.04235077, -2.29737009,
       -3.15082845, -4.33037208, -1.43700682, -2.59060064, -7.09025955,
       -2.32734557, -4.19073421, -2.63904005, -3.58104074, -3.75929058,
       -2.5501665 , -3.26968347, -2.78911367, -2.69251051, -0.77437547,
       -6.61955792, -2.99084623, -2.36975061, -7.18167736, -1.52816491,
       -4.81736265, -0.86996608, -7.09025955, -1.87404275, -2.58013247,
       -4.80953284, -4.17454084, -1.95444152, -3.50503339, -2.7411418 ,
       -1.87131835, -2.45270806, -3.07252433, -1.48988982, -3.98778754,
       -1.35216679, -0.16932277, -6.05237002, -3.29550658,  0.30513174,
       -3.25501463, -5.31993586, -0.4733928 , -4.79180882, -5.82486323,
       -2.70343539, -3.70838271, -2.4987341 , -2.12301018, -3.11393974,
       -2.79842053, -1.38274662, -4.3267103 , -3.57778946, -3.42356884,
       -9.15826626, -6.52588866, -1.75832085, -4.97842312, -1.59255758,
       -1.91317252, -2.81075595, -1.14056747, -1.6326702 , -1.6269208 ,
       -5.33227425, -3.88805074,  0.70432882, -3.99288096, -2.22595859,
       -2.63405871, -1.80643013, -4.18615661, -0.32535905, -3.36770115,
       -2.94105086, -3.02039211, -2.01312164, -2.97735566, -2.58917919,
       -6.78423811, -2.34521849, -3.11841945, -0.8177211 , -1.69000394,
       -1.20887457, -3.4996968 , -4.98602607, -4.36391839, -1.59023154,
       -1.36030606, -1.97047357, -2.63177729, -2.88308299, -5.15950486,
       -5.67559475, -3.04604495, -5.17135083, -1.38069437, -4.57622144,
       -3.72245121, -5.49591484, -3.39797976, -6.025302  , -3.80350106,
       -5.72200956, -0.20160041, -0.54441899, -1.4937231 , -4.17747099,
       -5.31993586, -5.30332414, -4.68554881, -2.54472004, -1.91317252,
       -5.0001379 , -1.29322696, -4.9018288 , -5.72200956, -1.63254894,
       -0.50844757, -2.79777831, -3.40409244, -3.75108143, -2.13430517,
       -2.3552617 , -4.47491213, -1.37020598, -2.98467788, -1.1097015 ,
       -2.65582153, -2.03469435, -3.82343254, -3.40409244, -3.86002039,
       -1.6454705 , -1.67610138, -2.25004084, -2.35293566, -3.97998511,
       -0.45142958, -3.70838271, -3.21804175, -3.26826004, -3.5011185 ,
       -0.69767632, -4.61318325, -2.53209123, -1.58352433,  0.11192491,
       -2.071356  , -2.52212098, -5.16155398, -3.79386016, -2.05580278,
       -3.0712261 , -5.04235077, -4.57622144, -0.59059425, -2.11738204,
       -0.14035732, -2.15201095, -3.90580754, -7.85208391, -1.77124241,
       -3.87977779, -1.17409032, -5.38735798, -0.38776428, -5.13905974,
       -6.52588866, -2.37600268, -3.67042165, -1.13694063, -1.52979384,
       -1.08806317, -2.17972634, -1.37566371, -1.69000394, -5.55433014,
       -4.38095601, -6.16274514, -5.75210573, -5.38139309, -2.1485872 ,
       -2.41257831, -1.21383508, -2.77794109, -2.77375743, -0.97332586,
       -0.07359075, -2.97947129, -3.43909806, -3.0639354 , -2.42856294,
       -6.51455759, -4.60929775, -3.36134518, -1.66451803, -0.34748753,
       -4.91887429, -1.5001642 , -2.55131925, -5.03116753, -1.1504684 ,
       -3.58407147, -7.22253774, -0.40922333, -3.44344971, -5.60855562,
       -8.54996931, -2.63926987, -1.35467792, -4.73905926, -1.92627702,
       -1.76204643, -2.04984044, -6.42989154, -5.4209013 , -5.4954495 ,
       -4.91209304, -1.84841838, -3.11841945, -1.19248532, -3.52050251,
       -1.6049836 , -3.31107991, -4.69327814,  0.6106115 , -3.60652062,
       -4.63133161, -2.5020505 , -1.04661836, -0.70671111, -2.77794109,
       -1.96622986, -1.24556874, -2.04067926, -0.47784193, -1.60070017,
       -1.73602998, -4.74777582, -2.35789126, -1.89320334, -4.94447725,
       -3.52116416, -1.03137202, -2.42722214, -2.41726714, -3.31243827,
       -2.45036038, -2.45036038, -2.09804761, -5.79050585, -2.12301018,
       -1.69652767, -5.04235077, -2.7183962 , -2.74034339, -3.08792242,
        1.16237365, -1.97390964, -2.36955435, -1.59518714, -2.64877661,
       -2.72328327, -1.69263351, -3.78676756, -2.70478582, -0.97951662,
       -6.17522371, -2.57848244, -4.23239293, -1.56313602, -4.14016489,
       -1.89020434, -0.42981761, -2.99158591, -4.47466933, -2.01820507,
       -2.27836725, -2.199028  , -4.05306078, -2.47347008, -1.21383508,
       -1.74165812, -1.03137202, -6.35437784, -2.13973631, -9.87954886,
       -1.51156575, -7.7015858 , -2.99084623, -2.90410699, -4.2232049 ,
       -2.01933871, -2.33668279, -2.81929845, -3.12088453, -4.42161212,
       -0.83133436, -2.63385972, -2.49648251, -0.79480376, -4.64013859,
       -4.39618439, -3.32058101, -2.61616401, -3.36987763, -1.51156575,
       -0.76911634, -2.06238448, -2.89331618, -4.57622144, -1.19907316,
       -2.72328327, -3.50414148, -3.10460597, -1.21005686, -4.38704198,
       -1.92623036, -4.17035059, -4.97728572, -1.79026429, -3.35283034,
       -3.59939381, -2.79154825, -3.25842453, -2.25768799, -2.14996491,
       -4.95790808, -1.4036531 , -4.64502179, -2.31758244, -3.01521178,
       -1.5577739 , -4.68504851, -2.98727379, -2.16925807, -2.65875355,
       -4.53388983, -3.39620209, -2.97542583, -2.63426164, -7.77785827,
       -2.34848929, -3.95395012, -3.14810449, -3.29620551, -1.99087234,
       -2.57905657, -5.72200956, -1.86861231, -1.48463069, -2.03428932,
       -7.44538031, -2.08246673, -3.2086445 , -2.137733  , -8.85737405,
       -3.15699268, -3.71109719, -0.99642961, -3.14546555, -4.88811295,
       -2.14179382, -4.15137001, -5.50933463, -4.69151986, -0.55120689,
       -2.24367474, -1.59518714, -4.86486127, -1.10853776, -0.06678132,
       -5.01101651, -3.95482844, -1.69000394, -4.69639167, -1.26852213,
       -4.30518407, -4.30377221, -2.23526305, -4.34616365, -0.23583339,
       -6.05831529, -3.63698515, -4.07880131, -3.2950999 , -1.75169752,
       -4.64981547, -2.68788426, -2.81766206, -4.38232674, -2.14927143,
       -5.44407214,  1.93550814, -0.12972508, -1.49498387, -2.50036024,
       -2.06396429, -5.93855312, -2.53403538, -2.89650533, -1.58868434,
       -5.73999593, -1.97390964, -4.64393496, -3.96695922, -1.30865188,
       -1.8711556 , -0.01555673, -4.75194249, -2.1532237 , -2.26674801,
       -4.60929775, -1.75320392, -1.16591868, -2.35789126, -2.73044354,
       -1.43453026, -2.29183886, -4.41564138, -3.88412944, -5.4954495 ,
       -1.64461506, -2.29089681, -2.32781082, -2.75074329, -3.88113542,
       -2.85196407, -1.59255758, -0.83248302, -4.41251959, -3.26968347,
       -1.83271568, -2.12301018, -0.45113096, -5.34942367, -0.69635755,
       -3.86581957, -3.82343254, -3.71474837, -3.36603645, -4.49530092,
       -7.77785827, -1.48607797, -4.66162003, -2.39395831, -3.75517868,
       -4.61318325, -4.17747099, -2.199028  , -3.01607829, -2.50036024,
       -1.88703307, -3.62102758, -3.51234807, -3.16279877, -7.09898181,
       -2.70478582, -3.37625597, -1.83271568, -4.87356837, -1.73865954,
       -2.88008616, -2.22977044, -6.70156561, -0.03717765, -3.53606465,
       -4.17657514, -5.01827423, -1.39251992, -1.36030606, -1.94448808,
       -4.00428682, -2.29865745, -1.3893508 , -0.39302341, -1.51156575,
       -2.34521849, -1.29089314, -0.97595542, -2.41038224, -0.4790451 ,
       -2.6818627 , -2.27035604, -2.46876907, -0.83629487, -2.00309292,
       -2.77794109, -2.70606495, -0.88802552,  0.33102924, -8.73475591,
       -6.08789325, -2.49648251, -0.84907925, -3.69568743, -4.37642963,
       -0.79254201,  1.98868187, -1.96622986, -2.49399292, -2.75566249,
       -2.80004662, -2.92698197, -2.29452458, -4.40478538, -3.41695228,
       -2.97735566, -4.53250784, -3.12440723, -1.48607797, -6.64951931,
       -1.29322696, -3.59615687, -4.84056475, -2.0637599 , -5.4954495 ,
       -1.13275935, -2.80728543, -4.40237156, -2.68666558, -3.44429049,
       -2.41959844, -4.95743275, -6.40483191, -1.13905974, -5.05708901,
       -1.66084991, -2.37974001, -2.05580278, -3.39656384, -1.42271129,
       -2.77786114, -2.30422683, -5.53938357, -3.36027875, -6.08708502,
       -1.53182046, -4.03471242, -3.35059383, -3.43858076, -5.34063365,
       -2.65303434, -0.32225969, -5.89030528, -1.66084991, -7.09025955,
       -1.66084991, -4.32045753, -3.2950999 , -4.00428682, -2.65971931,
       -2.61898292, -1.48081884, -3.31248863, -4.1285441 ,  0.81081357,
       -0.74032142, -0.55409165, -1.48988982, -2.2835894 , -2.52940291,
       -3.67078501, -3.11422447, -1.35216679, -2.33929584, -3.735629  ,
       -5.99133424,  1.68379176, -5.72200956, -1.1504684 , -3.78441114,
       -2.04833812, -4.23871436, -3.4723975 , -2.25259394, -3.75430163,
       -3.79386016, -4.33406929, -0.79254201, -2.72091671, -2.32584014,
       -3.04247237, -3.57422729, -2.05580278, -2.19679345, -3.21720962,
       -2.33668279, -3.18126918, -2.18323201, -1.1504684 , -2.82306619,
       -3.81016932, -0.62641139, -1.79284653, -3.0572287 , -2.68205673,
       -3.47393942, -1.35730748, -3.25283034, -3.50801673, -3.42242343,
       -2.33668279, -3.15515412, -0.81254078, -2.32180684, -7.23311502,
       -4.27331505, -0.41608127, -0.92992939, -2.85102139, -2.38452854,
       -1.35467792, -3.35250294, -4.61318325, -3.36770115, -4.54753068,
       -1.21120552, -1.41272408, -0.62894611, -5.46722693, -2.80332864,
       -4.70222762, -4.20128488, -3.75287911, -2.89485114, -4.44470271,
       -2.75272546, -6.40483191, -3.43858076, -1.33564461, -7.39909571,
       -2.3552617 , -1.97390964, -3.31913336, -0.99789515, -1.49137202,
       -2.54283347, -2.57935977, -1.5372532 , -1.46969589, -2.50256395,
       -2.84523685, -0.79050585, -0.72972899, -4.50417895, -3.41743324,
       -1.89701519, -3.04694269, -1.10327863, -1.48988982, -2.81732645,
       -6.11716924, -7.16934656, -1.96622986, -3.91516316, -3.61177065,
       -6.40483191, -4.95743275, -2.67078501, -1.54561991, -2.93180057,
       -3.27180065, -2.72328327, -2.53353851, -1.11363181, -2.90610261,
       -1.10853776, -1.49259585, -1.15572753, -4.70049144,  1.99249372,
       -1.80643013, -3.46728525, -2.32045753, -1.79366715, -2.25004084,
       -1.94336083, -1.87633593, -9.15826626, -3.65922507, -2.29683578,
       -0.5325366 , -7.5914365 , -2.15587374, -1.72389658, -1.78645244,
       -4.36890615, -1.91112045, -5.28950294, -4.70838271, -3.34373141,
       -1.82343254, -3.84290966, -1.75395277, -1.82944893, -1.92755736,
       -3.95353483, -2.51949142, -4.16972153, -4.47784193, -1.80004662,
       -2.08049675, -1.57187693, -3.79234971, -1.27341971, -1.10845148,
       -2.87977779, -1.08911411, -2.73044354, -1.76720276, -2.22965253,
       -1.56321848, -1.67044148, -1.51156575, -2.24983949, -1.72672119,
       -2.14897281, -4.05816764, -2.37140338, -1.897265  , -1.42797042,
       -3.08706826, -4.24521459, -6.52588866, -4.60929775, -4.67022068,
       -2.03122625, -3.11841945, -5.04235077,  0.60553271, -8.46728525,
       -4.23480474, -2.75684478, -2.14339849, -1.5231707 , -2.67580644,
       -3.59418929, -5.68837433, -5.02111613, -0.94099449, -3.51200413,
       -4.41666505, -1.9765392 , -2.67587664, -4.87753962, -3.73205568,
       -2.65599022, -2.46014399, -3.93774254, -4.46728525, -4.43381828,
       -2.81993282, -3.25038467, -5.64350181, -4.18419436, -6.47162256,
       -2.29452458, -3.70387135, -6.42099768, -5.53182846, -2.46049395,
       -1.91317252, -2.137733  , -1.31128145, -3.16496704, -2.22556914,
       -3.81016932, -2.44235234, -2.76845189, -2.73924332, -1.59715854,
       -2.90610261, -7.77785827, -2.74034339, -7.77785827, -6.53218341,
       -2.55678026, -2.59432621, -6.06602826, -2.73144347, -2.39319255,
       -3.79386016, -6.40483191, -1.32841939, -1.62955037, -4.07880131,
       -2.42494059, -6.04136471, -1.63103729, -5.2602532 , -3.86312053,
       -2.12601292, -4.28342693, -2.6178098 , -7.14847318, -1.53350725,
       -1.08911411, -4.24790282, -4.63837202, -3.79691901, -2.13813936,
       -3.86772233, -1.2281273 , -3.7672691 , -4.28538142, -1.43139099,
       -3.40980857, -2.6111225 , -1.41272408, -3.07730828, -3.90594959,
       -2.35283034, -0.7271857 , -4.19010154, -0.69374352, -1.10472591,
       -1.62166167, -3.90123327, -2.78518496, -3.43585411, -3.92684454,
       -2.54406861, -1.91317252, -1.77769503, -0.79480376, -4.23915408,
       -1.51148185, -1.14665655, -2.00855418, -7.03862139, -2.74305517,
       -2.11738204, -3.42356884, -2.49648251, -2.1372074 , -1.88151626,
       -1.62653948, -4.00428682, -2.04784256, -4.47943132, -2.61216962,
       -1.69531648, -5.54561991, -7.77785827, -2.13315833, -6.08238994,
       -0.73579117, -1.90396018, -3.66566374, -3.26337102, -3.5528319 ,
       -0.88717007, -4.50417895, -3.93774254, -2.31911758, -1.79407614,
       -2.73044354, -1.62464083, -3.36099002, -1.68087565, -4.17747099,
       -4.36658559, -2.68703435, -5.54667039, -2.40682831, -5.07481336,
       -3.80377549, -2.93847429, -3.69840797, -3.63698515, -2.77930842,
       -0.77056362, -2.37730267,  1.68379176, -3.2158222 , -3.70838271,
       -3.17308636, -4.36923501, -0.44844001, -5.24894879, -6.52225128,
       -1.80406335, -1.80643013, -1.90447849, -5.52193556, -2.07398557,
       -3.22132977, -6.70156561, -5.44962012, -1.25610101, -3.35223802,
       -3.98606913, -1.89546347, -6.08789325, -3.88711135, -3.78407726])
```

In [78]:
```
y_lr_test_pred
```

Out[78]:
```
array([-3.05722870e+00, -7.77785827e+00, -2.55016650e+00, -2.01523582e+00,
       -2.06375990e+00, -9.99672215e-01, -5.94603364e-01, -5.53626003e-01,
       -5.72200956e+00, -3.94006681e+00, -3.95496755e+00, -2.29737009e+00,
       -1.48980354e+00, -1.48988982e+00, -4.64510806e+00, -1.90396018e+00,
       -1.51566313e+00, -3.16424605e+00, -3.70863920e+00, -5.58105660e+00,
       -3.25038467e+00, -5.04235077e+00, -5.69194881e+00, -2.14339849e+00,
       -4.35689341e+00, -5.03964756e+00, -3.10383618e+00, -4.40286964e+00,
       -4.21276272e+00,  5.56508349e-01, -1.45537678e+00, -4.41027396e+00,
       -2.59668773e+00, -1.53336276e+00, -5.55749874e-01, -1.67111795e+00,
       -2.78163675e+00, -3.15395565e+00, -5.27083361e+00, -1.75321446e+00,
       -1.53350725e+00, -2.01255666e+00, -6.57559167e+00, -7.89433046e+00,
       -5.76437127e+00, -4.16422068e+00, -3.43694663e+00,  1.43834212e+00,
       -1.12679105e-02, -2.34521849e+00, -1.86480046e+00, -5.03964756e+00,
        8.55886378e-01, -3.17679292e+00, -5.06764094e+00, -1.99464442e+00,
       -7.77785827e+00, -1.21764693e+00, -9.09541075e-01, -5.04235077e+00,
       -2.43898748e+00, -2.84034045e+00, -2.53403538e+00, -2.36170311e+00,
       -1.63103729e+00, -1.53182046e+00, -3.23931568e+00, -2.88008616e+00,
       -1.88300518e+00, -3.21582220e+00, -3.40245202e+00, -9.01813905e-01,
       -4.82308940e+00, -7.69116343e-01, -7.12894308e+00, -1.05440427e+01,
       -1.95444152e+00, -3.50194744e+00, -7.18167736e+00, -6.01555673e+00,
       -2.08189806e+00, -2.31652280e+00, -3.44556948e+00, -2.05480142e+00,
       -6.01555673e+00, -2.88308299e+00, -4.84867198e+00, -3.51006495e-01,
       -3.54726250e+00, -1.21057919e+00, -4.36658559e+00, -4.21815903e-01,
       -1.63103729e+00, -2.51604291e+00, -2.16707077e+00, -1.48726025e+00,
       -3.20864450e+00, -1.51411141e+00, -1.65033691e+00, -3.66287663e+00,
       -3.26968347e+00, -3.94492313e+00, -4.22509088e+00, -3.68794650e+00,
       -5.98734972e+00, -1.43710934e+00, -1.97653920e+00, -1.85076729e+00,
       -1.14179382e+00, -3.07730828e+00, -4.84867198e+00, -2.19679345e+00,
       -1.68737438e+00, -2.20390218e+00,  1.89408269e+00, -3.61322115e+00,
       -2.79173430e+00, -2.41564138e+00, -7.53910534e-01, -8.54744860e-01,
       -9.20407401e-02, -6.14209981e+00, -3.79386016e+00, -7.77785827e+00,
       -1.79874130e+00, -2.50544035e+00, -3.77102985e+00, -2.25250766e+00,
       -2.57788713e+00, -2.06375990e+00, -3.33843958e+00, -1.03912484e+00,
       -6.68946164e+00, -1.91112045e+00, -2.58735850e+00, -2.19902800e+00,
       -1.90219551e+00, -2.81396751e+00, -4.16972153e+00, -5.72200956e+00,
       -1.60816482e+00, -3.68121117e+00, -4.60929775e+00, -2.45888480e+00,
        1.13185484e+00, -1.69279625e+00, -7.09025955e+00, -3.79386016e+00,
       -2.99712058e+00, -5.79600137e+00, -2.44845780e+00, -5.29399242e+00,
       -5.29389899e-01, -3.53652118e+00, -3.51200413e+00, -2.02419300e+00,
       -4.47466933e+00, -3.63836536e+00, -4.56596939e+00, -5.57842703e+00,
       -5.30676136e+00, -2.39225449e+00, -4.88290144e+00, -2.61359308e+00,
       -3.11841945e+00, -2.05580278e+00, -1.64987419e+00, -4.91881901e+00,
       -3.93774254e+00, -4.26411548e+00, -3.15082845e+00, -3.49352203e+00,
       -3.81768831e+00, -3.77197350e+00, -2.55016650e+00, -1.97653920e+00,
       -2.59432621e+00, -5.38480406e+00, -5.44932525e+00, -3.04107137e+00,
       -1.87252408e+00, -2.25124657e+00, -2.89215707e+00, -3.46087334e+00,
       -6.06861986e+00, -1.89916369e+00, -1.98035105e+00, -2.45036038e+00,
       -2.79393037e+00, -4.76010415e+00, -1.72379306e+00, -7.09025955e+00,
       -2.86880150e+00, -2.70674744e+00, -4.36825704e+00, -3.11841945e+00,
       -3.85805633e+00,  9.78662246e-03, -4.47466933e+00, -6.08708502e+00,
       -5.19970454e+00, -6.40483191e+00, -4.10155218e+00, -1.07044752e+00,
        1.99249372e+00, -3.63698515e+00, -8.50933009e-01, -3.26770298e+00,
       -4.78343575e+00, -1.48988982e+00, -2.24629276e+00, -4.13247222e+00,
       -4.36873484e+00, -1.89724815e+00, -1.50596465e+00, -1.16713539e-02,
       -1.73602998e+00, -2.34521849e+00, -4.54942814e-01, -4.18812419e+00,
       -1.87770440e+00, -3.70838271e+00, -1.59899899e+00, -3.26143822e+00,
       -4.17623614e+00, -6.67872053e+00, -1.23069039e+00, -3.14545964e+00,
       -4.79863925e+00])
```

Evaluate model perfomance

In [79]:
```
y_train    # Actual value
```

Out[79]:
```
107   -4.440
378   -1.250
529   -1.655
546   -1.886
320   -0.740
        ...
802   -2.925
53    -2.680
350   -7.020
79    -4.800
792   -3.240
Name: logS, Length: 915, dtype: float64
```

In [80]:
```
y_lr_train_pred  # Prediction using the (x_train) as an input to the model
```

Out[80]:

```
array([-2.83668969,  0.84208401, -0.50930302, -0.13058053, -1.6269208 ,
       -2.86103037, -2.32208333, -4.16684323, -5.04235077, -2.29737009,
       -3.15082845, -4.33037208, -1.43700682, -2.59060064, -7.09025955,
       -2.32734557, -4.19073421, -2.63904005, -3.58104074, -3.75929058,
       -2.5501665 , -3.26968347, -2.78911367, -2.69251051, -0.77437547,
       -6.61955792, -2.99084623, -2.36975061, -7.18167736, -1.52816491,
       -4.81736265, -0.86996608, -7.09025955, -1.87404275, -2.58013247,
       -4.80953284, -4.17454084, -1.95444152, -3.50503339, -2.7411418 ,
       -1.87131835, -2.45270806, -3.07252433, -1.48988982, -3.98778754,
       -1.35216679, -0.16932277, -6.05237002, -3.29550658,  0.30513174,
       -3.25501463, -5.31993586, -0.4733928 , -4.79180882, -5.82486323,
       -2.70343539, -3.70838271, -2.4987341 , -2.12301018, -3.11393974,
       -2.79842053, -1.38274662, -4.3267103 , -3.57778946, -3.42356884,
       -9.15826626, -6.52588866, -1.75832085, -4.97842312, -1.59255758,
       -1.91317252, -2.81075595, -1.14056747, -1.6326702 , -1.6269208 ,
       -5.33227425, -3.88805074,  0.70432882, -3.99288096, -2.22595859,
       -2.63405871, -1.80643013, -4.18615661, -0.32535905, -3.36770115,
       -2.94105086, -3.02039211, -2.01312164, -2.97735566, -2.58917919,
       -6.78423811, -2.34521849, -3.11841945, -0.8177211 , -1.69000394,
       -1.20887457, -3.4996968 , -4.98602607, -4.36391839, -1.59023154,
       -1.36030606, -1.97047357, -2.63177729, -2.88308299, -5.15950486,
       -5.67559475, -3.04604495, -5.17135083, -1.38069437, -4.57622144,
       -3.72245121, -5.49591484, -3.39797976, -6.025302  , -3.80350106,
       -5.72200956, -0.20160041, -0.54441899, -1.4937231 , -4.17747099,
       -5.31993586, -5.30332414, -4.68554881, -2.54472004, -1.91317252,
       -5.0001379 , -1.29322696, -4.9018288 , -5.72200956, -1.63254894,
       -0.50844757, -2.79777831, -3.40409244, -3.75108143, -2.13430517,
       -2.3552617 , -4.47491213, -1.37020598, -2.98467788, -1.1097015 ,
       -2.65582153, -2.03469435, -3.82343254, -3.40409244, -3.86002039,
       -1.6454705 , -1.67610138, -2.25004084, -2.35293566, -3.97998511,
       -0.45142958, -3.70838271, -3.21804175, -3.26826004, -3.5011185 ,
       -0.69767632, -4.61318325, -2.53209123, -1.58352433,  0.11192491,
       -2.071356  , -2.52212098, -5.16155398, -3.79386016, -2.05580278,
       -3.0712261 , -5.04235077, -4.57622144, -0.59059425, -2.11738204,
       -0.14035732, -2.15201095, -3.90580754, -7.85208391, -1.77124241,
       -3.87977779, -1.17409032, -5.38735798, -0.38776428, -5.13905974,
       -6.52588866, -2.37600268, -3.67042165, -1.13694063, -1.52979384,
       -1.08806317, -2.17972634, -1.37566371, -1.69000394, -5.55433014,
       -4.38095601, -6.16274514, -5.75210573, -5.38139309, -2.1485872 ,
       -2.41257831, -1.21383508, -2.77794109, -2.77375743, -0.97332586,
       -0.07359075, -2.97947129, -3.43909806, -3.0639354 , -2.42856294,
       -6.51455759, -4.60929775, -3.36134518, -1.66451803, -0.34748753,
       -4.91887429, -1.5001642 , -2.55131925, -5.03116753, -1.1504684 ,
       -3.58407147, -7.22253774, -0.40922333, -3.44344971, -5.60855562,
       -8.54996931, -2.63926987, -1.35467792, -4.73905926, -1.92627702,
       -1.76204643, -2.04984044, -6.42989154, -5.4209013 , -5.4954495 ,
       -4.91209304, -1.84841838, -3.11841945, -1.19248532, -3.52050251,
       -1.6049836 , -3.31107991, -4.69327814,  0.6106115 , -3.60652062,
       -4.63133161, -2.5020505 , -1.04661836, -0.70671111, -2.77794109,
       -1.96622986, -1.24556874, -2.04067926, -0.47784193, -1.60070017,
       -1.73602998, -4.74777582, -2.35789126, -1.89320334, -4.94447725,
       -3.52116416, -1.03137202, -2.42722214, -2.41726714, -3.31243827,
       -2.45036038, -2.45036038, -2.09804761, -5.79050585, -2.12301018,
       -1.69652767, -5.04235077, -2.7183962 , -2.74034339, -3.08792242,
        1.16237365, -1.97390964, -2.36955435, -1.59518714, -2.64877661,
       -2.72328327, -1.69263351, -3.78676756, -2.70478582, -0.97951662,
       -6.17522371, -2.57848244, -4.23239293, -1.56313602, -4.14016489,
       -1.89020434, -0.42981761, -2.99158591, -4.47466933, -2.01820507,
       -2.27836725, -2.199028  , -4.05306078, -2.47347008, -1.21383508,
       -1.74165812, -1.03137202, -6.35437784, -2.13973631, -9.87954886,
       -1.51156575, -7.7015858 , -2.99084623, -2.90410699, -4.2232049 ,
       -2.01933871, -2.33668279, -2.81929845, -3.12088453, -4.42161212,
       -0.83133436, -2.63385972, -2.49648251, -0.79480376, -4.64013859,
       -4.39618439, -3.32058101, -2.61616401, -3.36987763, -1.51156575,
       -0.76911634, -2.06238448, -2.89331618, -4.57622144, -1.19907316,
       -2.72328327, -3.50414148, -3.10460597, -1.21005686, -4.38704198,
       -1.92623036, -4.17035059, -4.97728572, -1.79026429, -3.35283034,
       -3.59939381, -2.79154825, -3.25842453, -2.25768799, -2.14996491,
       -4.95790808, -1.4036531 , -4.64502179, -2.31758244, -3.01521178,
       -1.5577739 , -4.68504851, -2.98727379, -2.16925807, -2.65875355,
       -4.53388983, -3.39620209, -2.97542583, -2.63426164, -7.77785827,
       -2.34848929, -3.95395012, -3.14810449, -3.29620551, -1.99087234,
       -2.57905657, -5.72200956, -1.86861231, -1.48463069, -2.03428932,
       -7.44538031, -2.08246673, -3.2086445 , -2.137733  , -8.85737405,
       -3.15699268, -3.71109719, -0.99642961, -3.14546555, -4.88811295,
       -2.14179382, -4.15137001, -5.50933463, -4.69151986, -0.55120689,
       -2.24367474, -1.59518714, -4.86486127, -1.10853776, -0.06678132,
       -5.01101651, -3.95482844, -1.69000394, -4.69639167, -1.26852213,
       -4.30518407, -4.30377221, -2.23526305, -4.34616365, -0.23583339,
       -6.05831529, -3.63698515, -4.07880131, -3.2950999 , -1.75169752,
       -4.64981547, -2.68788426, -2.81766206, -4.38232674, -2.14927143,
       -5.44407214,  1.93550814, -0.12972508, -1.49498387, -2.50036024,
       -2.06396429, -5.93855312, -2.53403538, -2.89650533, -1.58868434,
       -5.73999593, -1.97390964, -4.64393496, -3.96695922, -1.30865188,
       -1.8711556 , -0.01555673, -4.75194249, -2.1532237 , -2.26674801,
       -4.60929775, -1.75320392, -1.16591868, -2.35789126, -2.73044354,
       -1.43453026, -2.29183886, -4.41564138, -3.88412944, -5.4954495 ,
       -1.64461506, -2.29089681, -2.32781082, -2.75074329, -3.88113542,
       -2.85196407, -1.59255758, -0.83248302, -4.41251959, -3.26968347,
       -1.83271568, -2.12301018, -0.45113096, -5.34942367, -0.69635755,
       -3.86581957, -3.82343254, -3.71474837, -3.36603645, -4.49530092,
       -7.77785827, -1.48607797, -4.66162003, -2.39395831, -3.75517868,
       -4.61318325, -4.17747099, -2.199028  , -3.01607829, -2.50036024,
       -1.88703307, -3.62102758, -3.51234807, -3.16279877, -7.09898181,
       -2.70478582, -3.37625597, -1.83271568, -4.87356837, -1.73865954,
       -2.88008616, -2.22977044, -6.70156561, -0.03717765, -3.53606465,
       -4.17657514, -5.01827423, -1.39251992, -1.36030606, -1.94448808,
       -4.00428682, -2.29865745, -1.3893508 , -0.39302341, -1.51156575,
       -2.34521849, -1.29089314, -0.97595542, -2.41038224, -0.4790451 ,
       -2.6818627 , -2.27035604, -2.46876907, -0.83629487, -2.00309292,
       -2.77794109, -2.70606495, -0.88802552,  0.33102924, -8.73475591,
       -6.08789325, -2.49648251, -0.84907925, -3.69568743, -4.37642963,
       -0.79254201,  1.98868187, -1.96622986, -2.49399292, -2.75566249,
       -2.80004662, -2.92698197, -2.29452458, -4.40478538, -3.41695228,
       -2.97735566, -4.53250784, -3.12440723, -1.48607797, -6.64951931,
       -1.29322696, -3.59615687, -4.84056475, -2.0637599 , -5.4954495 ,
       -1.13275935, -2.80728543, -4.40237156, -2.68666558, -3.44429049,
       -2.41959844, -4.95743275, -6.40483191, -1.13905974, -5.05708901,
       -1.66084991, -2.37974001, -2.05580278, -3.39656384, -1.42271129,
       -2.77786114, -2.30422683, -5.53938357, -3.36027875, -6.08708502,
       -1.53182046, -4.03471242, -3.35059383, -3.43858076, -5.34063365,
       -2.65303434, -0.32225969, -5.89030528, -1.66084991, -7.09025955,
       -1.66084991, -4.32045753, -3.2950999 , -4.00428682, -2.65971931,
       -2.61898292, -1.48081884, -3.31248863, -4.1285441 ,  0.81081357,
       -0.74032142, -0.55409165, -1.48988982, -2.2835894 , -2.52940291,
       -3.67078501, -3.11422447, -1.35216679, -2.33929584, -3.735629  ,
       -5.99133424,  1.68379176, -5.72200956, -1.1504684 , -3.78441114,
       -2.04833812, -4.23871436, -3.4723975 , -2.25259394, -3.75430163,
       -3.79386016, -4.33406929, -0.79254201, -2.72091671, -2.32584014,
       -3.04247237, -3.57422729, -2.05580278, -2.19679345, -3.21720962,
       -2.33668279, -3.18126918, -2.18323201, -1.1504684 , -2.82306619,
       -3.81016932, -0.62641139, -1.79284653, -3.0572287 , -2.68205673,
       -3.47393942, -1.35730748, -3.25283034, -3.50801673, -3.42242343,
       -2.33668279, -3.15515412, -0.81254078, -2.32180684, -7.23311502,
       -4.27331505, -0.41608127, -0.92992939, -2.85102139, -2.38452854,
       -1.35467792, -3.35250294, -4.61318325, -3.36770115, -4.54753068,
       -1.21120552, -1.41272408, -0.62894611, -5.46722693, -2.80332864,
       -4.70222762, -4.20128488, -3.75287911, -2.89485114, -4.44470271,
       -2.75272546, -6.40483191, -3.43858076, -1.33564461, -7.39909571,
       -2.3552617 , -1.97390964, -3.31913336, -0.99789515, -1.49137202,
       -2.54283347, -2.57935977, -1.5372532 , -1.46969589, -2.50256395,
       -2.84523685, -0.79050585, -0.72972899, -4.50417895, -3.41743324,
       -1.89701519, -3.04694269, -1.10327863, -1.48988982, -2.81732645,
       -6.11716924, -7.16934656, -1.96622986, -3.91516316, -3.61177065,
       -6.40483191, -4.95743275, -2.67078501, -1.54561991, -2.93180057,
       -3.27180065, -2.72328327, -2.53353851, -1.11363181, -2.90610261,
       -1.10853776, -1.49259585, -1.15572753, -4.70049144,  1.99249372,
       -1.80643013, -3.46728525, -2.32045753, -1.79366715, -2.25004084,
       -1.94336083, -1.87633593, -9.15826626, -3.65922507, -2.29683578,
       -0.5325366 , -7.5914365 , -2.15587374, -1.72389658, -1.78645244,
       -4.36890615, -1.91112045, -5.28950294, -4.70838271, -3.34373141,
       -1.82343254, -3.84290966, -1.75395277, -1.82944893, -1.92755736,
       -3.95353483, -2.51949142, -4.16972153, -4.47784193, -1.80004662,
       -2.08049675, -1.57187693, -3.79234971, -1.27341971, -1.10845148,
       -2.87977779, -1.08911411, -2.73044354, -1.76720276, -2.22965253,
       -1.56321848, -1.67044148, -1.51156575, -2.24983949, -1.72672119,
       -2.14897281, -4.05816764, -2.37140338, -1.897265  , -1.42797042,
       -3.08706826, -4.24521459, -6.52588866, -4.60929775, -4.67022068,
       -2.03122625, -3.11841945, -5.04235077,  0.60553271, -8.46728525,
       -4.23480474, -2.75684478, -2.14339849, -1.5231707 , -2.67580644,
       -3.59418929, -5.68837433, -5.02111613, -0.94099449, -3.51200413,
       -4.41666505, -1.9765392 , -2.67587664, -4.87753962, -3.73205568,
       -2.65599022, -2.46014399, -3.93774254, -4.46728525, -4.43381828,
       -2.81993282, -3.25038467, -5.64350181, -4.18419436, -6.47162256,
       -2.29452458, -3.70387135, -6.42099768, -5.53182846, -2.46049395,
       -1.91317252, -2.137733  , -1.31128145, -3.16496704, -2.22556914,
       -3.81016932, -2.44235234, -2.76845189, -2.73924332, -1.59715854,
       -2.90610261, -7.77785827, -2.74034339, -7.77785827, -6.53218341,
       -2.55678026, -2.59432621, -6.06602826, -2.73144347, -2.39319255,
       -3.79386016, -6.40483191, -1.32841939, -1.62955037, -4.07880131,
       -2.42494059, -6.04136471, -1.63103729, -5.2602532 , -3.86312053,
       -2.12601292, -4.28342693, -2.6178098 , -7.14847318, -1.53350725,
       -1.08911411, -4.24790282, -4.63837202, -3.79691901, -2.13813936,
       -3.86772233, -1.2281273 , -3.7672691 , -4.28538142, -1.43139099,
       -3.40980857, -2.6111225 , -1.41272408, -3.07730828, -3.90594959,
       -2.35283034, -0.7271857 , -4.19010154, -0.69374352, -1.10472591,
       -1.62166167, -3.90123327, -2.78518496, -3.43585411, -3.92684454,
       -2.54406861, -1.91317252, -1.77769503, -0.79480376, -4.23915408,
       -1.51148185, -1.14665655, -2.00855418, -7.03862139, -2.74305517,
       -2.11738204, -3.42356884, -2.49648251, -2.1372074 , -1.88151626,
       -1.62653948, -4.00428682, -2.04784256, -4.47943132, -2.61216962,
       -1.69531648, -5.54561991, -7.77785827, -2.13315833, -6.08238994,
       -0.73579117, -1.90396018, -3.66566374, -3.26337102, -3.5528319 ,
       -0.88717007, -4.50417895, -3.93774254, -2.31911758, -1.79407614,
       -2.73044354, -1.62464083, -3.36099002, -1.68087565, -4.17747099,
       -4.36658559, -2.68703435, -5.54667039, -2.40682831, -5.07481336,
       -3.80377549, -2.93847429, -3.69840797, -3.63698515, -2.77930842,
       -0.77056362, -2.37730267,  1.68379176, -3.2158222 , -3.70838271,
       -3.17308636, -4.36923501, -0.44844001, -5.24894879, -6.52225128,
       -1.80406335, -1.80643013, -1.90447849, -5.52193556, -2.07398557,
       -3.22132977, -6.70156561, -5.44962012, -1.25610101, -3.35223802,
       -3.98606913, -1.89546347, -6.08789325, -3.88711135, -3.78407726])
```


Model perfomance evaluation

In [81]:
```
from sklearn.metrics import mean_squared_error, r2_score

lr_train_mse = mean_squared_error(y_train, y_lr_train_pred) #  y_lr_train_pred = x_train
lr_train_r2 = r2_score(y_train, y_lr_train_pred)

lr_test_mse= mean_squared_error(y_test, y_lr_test_pred)     # y_lr_test_pred = x_test
lr_test_r2 = r2_score(y_test, y_lr_test_pred)
```

In [82]:
```
print('LR MSE (train): ',lr_train_mse)
print( 'LR R2 (train): ',lr_train_r2)
print('LR MSE (test): ',lr_test_mse)
print( 'LR R2 (test): ',lr_test_r2)
```

```
LR MSE (train):  1.007536295109369
LR R2 (train):  0.7645051774663391
LR MSE (test):  1.0206953660861033
LR R2 (test):  0.7891616188563282
```

In [83]:
```
lr_results = pd.DataFrame(['Linear regression', lr_train_mse,lr_train_r2,lr_test_mse,lr_test_r2]).transpose()
lr_results.columns = ["method","training mse","training r2","test mse","test r2"]
```

In [84]:
```
lr_results
```

Out[84]:

	method	training mse	training r2	test mse	test r2
0	Linear regression	1.007536	0.764505	1.020695	0.789162

Random forest

Importing the model

In [98]:
```
from sklearn.ensemble import RandomForestRegressor # RandomForestRegressor is used here bcoz 'logS' (i.e) the 'y' variable is quantitative
                                                   # if the 'y' variable is categorical then calssification model would have been built
```

Model building

In [99]:
```
rf = RandomForestRegressor(max_depth=2, random_state=100)
```

Out[99]:
```
RandomForestRegressor
RandomForestRegressor(max_depth=2, random_state=100)
```

Training the model

In [ ]:
```
rf.fit(x_train, y_train)
```

Applying the model to make a prediction

In [88]:
```
y_rf_train_pred = rf.predict(x_train)
y_rf_test_pred = rf.predict(x_test)
```

Evaluate model perfomance

In [100…
```
from sklearn.metrics import mean_squared_error, r2_score

rf_train_mse = mean_squared_error(y_train, y_rf_train_pred)
rf_train_r2 = r2_score(y_train, y_rf_train_pred)

rf_test_mse= mean_squared_error(y_test, y_rf_test_pred)
rf_test_r2 = r2_score(y_test, y_rf_test_pred)
```

In [101…
```
rf_results = pd.DataFrame(['Random forest', rf_train_mse,rf_train_r2,rf_test_mse,rf_test_r2]).transpose()
rf_results.columns = ["method","training mse","training r2","test mse","test r2"]
```

In [106…
```
rf_results
```

Out[106…

	method	training mse	training r2	test mse	test r2
0	Random forest	1.028228	0.759669	1.407688	0.709223

Model comparison between Liner Regression and Random forest

In [107…
```
df_models = pd.concat([lr_results, rf_results],axis=0).reset_index(drop=True) # "axis=0" to combine as a row
df_models                                                                      # "axis=1" to combine as a column
```

Out[107…

	method	training mse	training r2	test mse	test r2
0	Linear regression	1.007536	0.764505	1.020695	0.789162
1	Random forest	1.028228	0.759669	1.407688	0.709223

In [108…
```
#df_models.reset_index(drop=True)
```

Data visulaisation of prediction results

Import matplot library and Numpy

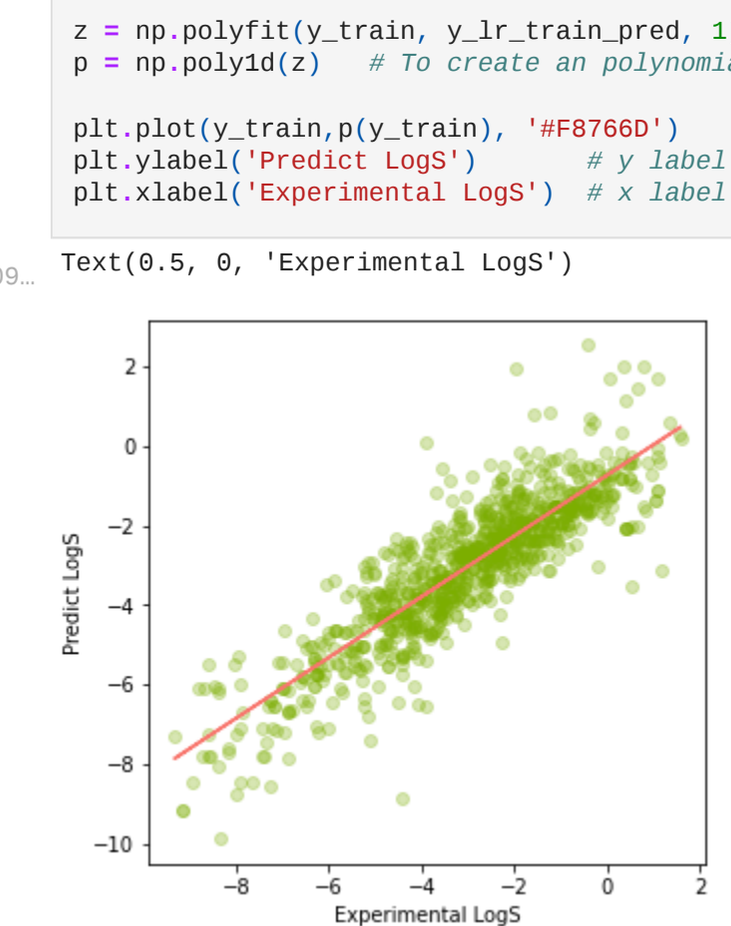
In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np    #To draw the trend line
```

In [109…
```
plt.figure(figsize=(5,5)) # size of the graph
plt.scatter(x=y_train,y=y_lr_train_pred, c="#7CAE00" ,alpha=0.3) # "C=#7CAE00" represents the Green colour

z = np.polyfit(y_train, y_lr_train_pred, 1)
p = np.poly1d(z)   # To create an polynomial object

plt.plot(y_train,p(y_train), '#F8766D')
plt.ylabel('Predict LogS')       # y label
plt.xlabel('Experimental LogS')  # x label
```

Out[109…
```
Text(0.5, 0, 'Experimental LogS')
```



Summary:

The dataset provided contains information on the solubility of molecules, which holds significant importance for chemists and biologists in determining the molecule's ability to dissolve in water or other solvents.

To address this problem, both Linear Regression and Random Forest Regressor models were employed.

The selection of the best model was based on evaluating performance metrics such as Mean Squared Error (MSE) and R2 (coefficient of determination).

Furthermore, visualizations were generated specifically for the best model to enhance the understanding and interpretation of the solubility patterns exhibited by the molecules.

These analyses and visualizations offer valuable insights to researchers, assisting them in making informed decisions regarding solubility behavior and solvent selection for various molecules