Application of ML to calculate spin

– Tarek Nabih

Table of contents

01 The Problem

Calculate the spin of molecules based on given parameters, plot the relation between a parameter and spin

Current method:

```
(base) tarek__nabih@Tareks-MacBook-Air round 2 % spinev Tarek
SpinEvolution 6.2.5 Double Precision Unlimited Edition
Experiment loaded
Using clust0 mode for splitting the computation
0: scan & ave dimensions point:   1
1: scan & ave dimensions point:   1
2: scan & ave dimensions point:   1
3: scan & ave dimensions point:   1
4: scan & ave dimensions point:   1
5: scan & ave dimensions point:   1
6: scan & ave dimensions point:   1
7: scan & ave dimensions point:   1

clock    40.11
```

- Not efficient
- Would take dozens of minutes to plot

02

Data collection

To train the machine learning model we had to collect big data sets

```
8
9   # Define the headers and their corresponding ranges
10  headers = ['H1_freq', 'spinning_freq' ,'power_1_1_1' ,'cs_beta_2', 'ss_iso_1_2',
11  ranges = [(50,1300),(0.05,80),(0,1500),(0,180),(-600e3,600e3),(-1200e3,1200e3),(0,
12
13  # Generate the input file
14  with open('input.txt', 'w') as f:
15      f.write(' '.join(headers) + '\n')
16      for _ in range(250000):
17          values = [str(np.random.uniform(low, high)) for low, high in ranges]
18          f.write(' '.join(values) + '\n')
19
20
21
```

For every parameter, a random was generated and the corresponding spin output was calculated via spin evolution

NYU’s supercomputer was used to generate the data

```
tmn8829@login2:/scratch/tmn8829$ squeue
      JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
    1894397   compute     file  tmn8829  R   11:51:16      9 cn[119,123-124,134-135,137-140]
    1894396   compute     file  tmn8829  R   11:51:28     16 cn[051,053-060,071-076,119]
    1894395   compute     file  tmn8829  R   11:52:18     14 cn[013-014,172-173,175-177,180-181,185-189]
    1894392   compute     file  tmn8829  R   11:54:13     18 cn[010-012,077-083,085-092]
tmn8829@login2:/scratch/tmn8829$
```

The concept of parallelism was useful in computing the results

data.csv

```
436.3192477658188,23.26414669994322,6.0448966787327 32,-6868.576772983535,-1486.001408637967,43.4119946494888086,282.2476508289824,68.982417857250 51,281249.75967253,5.6204896484653 12,0.934251408858687 40
990.9815116602596,38.527323269425395,34.965013160025 28,-296209.6981383927,-165423.27672481776,67.44149902155455,376.9299900687899,8.6385007848606 29,-968208.8545698595,8.000306889535167,0.9389069893448471 61
278.7788984925236,44.98923721248222,30.587104115061425,-122864.55355149484,-125560.49022218515,86.59835335884378,365.44925096156464,15.882300366926433,-777421.4496705733,5.837846814373912,237.75655768505348
434.76591955230845,4.280764415805153,8.243028488283217,-267628.8287672101,-52981.21990469168,44.13056410380519,161.1205839487131,20.460092030536 07,505952.99604593916,8.322429491724714,-49.138008771300214
1189.194272659794 7,4.826911551583743,69.16981417773978,-34370.82805281511,-87396.51871388828,78.1454269906144,796.8467357203529,5.304400331647421,-140433.42717899824,5.261094067083695,14.487285832043415
869.95428357043,37.29701033094893,19.87255838360951,-445182.4822897895,-113540.51984317088,83.01649478641055,26.191587639404545,109.446124025997 43,626201.7642448868,3.730434547298147,-7.0928607544148266
127.85716241229224,58.552317413873 98,7.540296168231202,-325027.87925621984,-201184.49128952302,17.928889146149825,863.2977227604606,26.588258065884425,-220518.55850851303,3.155338515194275,1.3477373235483401
537.6288797427577,52.15732185718731,57.459471711020 41,-352350.2177680484,-271663.0355868621,0.9209475245054488,936.1456930149047,46.010892633271226,-332701.390120312,2.5809178341921606,429.91370530290749
784.1339022460381,19.494401687020545,15.47054505014145,-450948.2418200813,-108714.47186376818,94.41438879298781,67.1309761274682 2,97.9575894725206,-20070.52268292161,2.0241160006721626,-5.2373760340388129
399.54588553814716,9.449580287715802,7.515462567176874,-435473.6490332576,-202988.60822354903,78.80393175369889,936.1066247467936,48.986685618885791,-399908.4037270277,4.157929721185631,-7.3856978782292213
1175.308621695018,30.805029745523214,37.07862609803881,-324374.52234036906,-251596.47338484725,87.2688663209328 2,360.486429638045,85.582019954161 96,167741.59672739287,4.076170230511611,-25.054935188133381
664.1423662496402,34.39358431803796,32.836528063137 41,-258826.43534896165,-149836.36708607626,50.981433333345016,984.6500056573985,52.140455936535 46,-205860.2336283481,4.368918751930385,41.452245087329075
1042.6568937754134,22.520862177726926,6.308672806686775,-167453.6285585775,-150552.21100620137,1.2885966888298173,956.4510061043424,43.200120518754986,943871.9523404001,4.696287365772063,0.6754798927993448 4
841.8447778555666,25.274894600616722,75.85460837339593,-205866.27692768205,-202666.11665439553,7.540599974316984,726.8860554844468,7.840356756967477,-408101.52969980845,3.196117825115142 5,446.40090801846947
1130.389196123648,14.276135115081578,25.08339864157755,-294065.4343884578,-36065.46231912123,96.58395199478036,158.6139883129301 6,18.982555906017513,-836482.1342784304,5.426375689672701,0.9613637671338306 5
244.99707991070449,55.349318482124026,25.292531209587956,-92762.12398846162,-276402.7175717408,105.64834505891845,337.7113460371773 5,80.025498550057 99,-180333.74470233428,8.886592344786287,18.641578798851327
753.9576555224428,62.51710752807864,87.08872968582293,-153671.95961280062,-74316.79087366717,33.003380985471,272.6363214172209,56.364215745752 93,-688960.5256070495,1.971503044049197,92.946358828710643
1182.729018806169,68.84944851061884,105.06094446546497,-130527.15329616441,-264315.21267266123,86.47228144769774,276.2486536950 77,47.325726704233 51,234502.8337117245,3.667304243079007 7,-26.016090955887769
474.4503191768674,34.886741445587 31,47.838627245176994,-123456.64210657205,-97880.38140768744,101.7751404983880 95,294.4118662858306,26.335085092318067,-957354.8290954754,2.030821543059265,1.0299044702830695
585.2476118597127,11.550612945794082,54.543665429397606,-203269.03917197615,-7683.150945073401,83.658045464721 68,606.0001418318349,88.025998539570 55,-761468.1589842224,6.963528705242721,181.807242914911907
1039.5452075630262,69.15589881097611,119.97186897823688,-152955.03561132163,-68207.54388455502,46.141307702466 88,409.0096084331517,64.901120710592 04,393115.5576916102,5.027283354596947,-84.1493964850399 29
886.2846620711,39.71546542913417,0.4999025690280678 4,-467666.7264370024,-68946.04177441765,24.525296926736 5,281.7889317400299 7,74.219586678128 01,341593.45193780307,7.831057214744827,-22.523574998172496
560.8785381161003,40.466259787361 58,99.713814862397 65,-215673.8206388519,-281802.60793184664,65.613839393747 11,784.2074411887138,38.962069096315 95,-851492.0539227733,3.637712319992841,93.956499504006530
850.0961597708 29,1.103199892905871,89.83821931669121,-484808.80205392407,-258771.7042191193,100.0770083639882,986.3685086079608,97.815453995805 19,360144.53467319603,8.923750108497899,-498.13897316385328
854.1057968562955,67.01518086726863,113.42255924321164,-226682.42679671606,-190923.4932750988,1.7830879217060724,660.3088138802409,94.491441321164 56,-582688.5693633421,1.555259802284847,317.25239149316542
616.7630397574602,64.29435264053879,49.9903615779865 76,-51086.470619775646,-86353.67668073677,26.852412274545824,422.6338897651440 3,45.131257576505874,540478.0660961296,9.686713075674771,-51.892996189135594
602.0413201556012,43.321194218188765,111.15559609344297,-4693.750869731477,-268100.5366327236,8.247709535794204,745.7921042232703,99.860333140400385,350755.91479639243,0.974503216887362,-74.303551617379341
830.682880902669,40.804655619526436,6.735869032360777,-391983.85670654435,-31505.64784608246,61.210070189805 97,77.1006119679446,87.435241677948 7,-389287.34805480647,1.417410231113 63,52.980226850990569
297.49391006601985,28.358598472043727,57.012289255317341,-460297.26691812894,-77990.46164926953,101.690199590385 79,493.619713621915,66.132502116584 05,986717.0315606829,6.680653685566795,0.866660412135485 39
1089.4227438138619,9.146512377002525,30.702613230972602,-340179.8148659336,-152107.993006690774,9.167292876260973,811.5808359587566,59.774210381728 27,-235006.52433796937,6.722060904205815,149.404107520355558
190.6842222897194,25.383824113450416,6.940628615992095,-341207.65315914317,-188488.7210203238,104.643589097499 07,113.020463629316 84,59.479684404873 236,653273.9066084786,9.698885514464935,0.982339281168067 24
1121.4335406814862,32.444143259864944,33.926722738691173,-224419.87129065522,-110823.52399344381,11.895474026923662,366.4928791464711,70.173323003230808,876596.2816192182,7.823377692987224,-35.384906154958436
364.478084059155,37.681620898244 02,31.296090768383813,-48470.21438525652,-62527.35849004163,43.126714437256 27,245.01844378351268,15.713317407216119,-254040.385479682,2.015436217914475,63.196923271904289
500.9798063847214,31.995284616067 45,93.311037366248,-137037.87702870247,-290637.34740600647,69.615856520795,611.9035845996995,3.372236723667829 7,493466.8762023151,3.9066720990553274,-389.24749897267338
459.2137687531369 5,16.713146870764 42,55.074480461235 87,-134502.85182531195,-113439.47975770014,56.601298698640 15,124.566786483555224,50.787136563084 44,158024.58255561348,9.203375989151764,-23.177709599143853
1151.5750264116484,45.010841583332 68,74.530431933798 94,-437050.08350364544,-255789.81819175478,70.551620106998 9,936.2293622836869,88.714537856976 66,790737.4493555541,3.205613799585769 6,-156.27851524373520
885.1561906856821,68.985849836493 64,38.581903203185 64,-448269.63430284365,-240112.1493313724,48.036608971679975,236.83123960681897,102.853857401372 55,-656176.6483691116,9.276641217983423,192.783797227704 70
869.0373814141503,39.649623329955 31,23.916576306714767,-304273.88750183966,-48102.33556092909,97.044503509342 46,122.2773728544213,57.849767710662206,45937.773089141934,7.526088703588705,-5.310289831464887 2
384.8692940067649,51.995782798875 55,12.595722864141976,-38611.90453935065,-148869.34233562145,84.696288264111 2,414.454452460315 1,106.375291842920 13,600072.3872706161,5.483681150169266,-30.597212099935284
140.4950812499398 3,50.929003978933 07,40.387629133695015,-492286.2250073751,-3500.903124342439 7,91.849736804877 2,796.1270117985678,16.716719019307092,-12732.174404127058,1.341340536949907,1.3152105594767838
228.2663399803922,15.315403720772354,105.51846875350074,-153667.04504779517,-113373.24097080165,53.704277910478034,111.227632088500 95,27.574330935937734,-45471.0994357159 36,8.366271234768 68,7.267409961198795
500.1217013921152,30.458398520419 06,92.775768298816 93,-280356.8339777464,-9985.426140448544,92.238642906052 9,134.516732302283 46,27.761884804184213,218958.86997908843,4.025115964491396 5,-238.71569031077942
993.3040584744273,26.430048674219723,22.459507443957 83,-132834.15664730337,-174836.55529200815,63.895941529 8233,739.1494203735574,75.324348096197 54,748572.2354854769,8.658205150253409,-189.43349385850337
546.3312057190427,25.768548950860 75,2.6962147795251745,-489286.70001311484,-286516.60959639 47,81.475850137053 84,550.4201489967429,81.803361201765 21,-536860.4061386102,4.521325135143831,164.19578480171438
1165.285063234959,23.054230293878863,95.25757251010988,-450929.6480954514,-166601.22326391737,55.925310336738 5,942.3936497411 5,28.120378750599975,-925776.7881537983,4.472259631990979,1.2222677993254578
1043.3855108880252,58.629931930568 96,86.825876722937 51,-323234.5372967329,-8009.906078073724,42.106515635046 56,38.380608799534436,115.820222684900883,-341943.7959575807,4.990365108001 04,32.583769207135838
319.2296382336560 3,20.562619982954182,95.765093324584 04,-410533.35488163965,-61526.654309211 47,8.195179064932173,7.626629472328505 57,-223286728114,311417.9829225221,8.653011203671742,1.0004130794740953
875.1322225213449,47.648447923164771,1.091371907082209,-57248.781011035026,-146740.693608907 34,56.408338993168 392,492.461083796309 8,107.264277940863 45,342629.47249424946,3.024876800191505,-4.162629904281743
121.60303966554162,40.032816127616 25,42.164290181905 88,-492194.39817737095,-270746.9232164269,8.224961555805512,599.3775136613221 3,16.669661287688132,958623.1189126323,2.358996015369205,-1.6455694238512724
162.102666676933227,7.664215333875956,73.166968292012 87,-459617.05396538426,-192385.860215800 79,2.881840942464713,944.080686893 9154,104.609492918271 44,-708485.7531085906,9.330955951441501,12.993407403060196
1042.1434267997306,19.789416401892 1,46.678298482243974,-187792.00647530484,-76178.16077637507,92.714531311494 05,112.479519184502 5,33.232718671573 82,626735.2057595882,3.412190756615324,-53.445812045618837
741.2294193058565,38.011655220205434,25.977433443602166,-356947.2509297988,-225306.3728012389,14.345256082721761,45.452575317070924,89.994243009552298,725807.8719694931,9.811103596702576,-25.338452242598777
290.5860532603283,2.5346489590471264,41.276641882670916,-4679.162075297034,-2235.097220431780 4,101.07366350730008,550.8873891181743,95.740819527279 37,631982.8110392459,5.506235493809 94,-95.852956002703061
771.9764476585893,29.351651080130914,23.120653829903056,-307504.9480066222,-161447.4600745454,77.537607339171 13,513.921935697190 7,17.157587962431 47,887838.5518434572,3.945782052010177,-81.592461629373602
1040.8458545180383,58.229024547262 51,63.398928953960635,-313240.57681789773,-110232.79918267694,5.251232850950855 5,808.6965365405875,94.681548290214 92,417849.8202368079,8.519466741086629,-346.624015165389 27
266.9383770896963,28.632479003780084,60.549239466196 78,-438051.1793440797,-215438.3338471552,96.730445217182 14,169.110650226908 52,119.022573274184 08,620940.3699859614,7.358244477218 23,0.992914365622141 41
105.8708377784041,39.329324535804 97,113.42964997226984,-360289.6575540488,-191777.04428200662,116.295812697381 73,816.6205497271669,101.248470586740 74,-701964.3982758381,8.741223172809 77,30.587770325845767
610.3455394055397,36.714999744947704,51.796114737361633,-376453.4922291436,-37615.492516633591 9,74705460514174 5,792.2080071294794,30.314293734811052,994256.427038928,8.470306283552903,-12.444227890940608
```

Around 250,000 lines of data was generated, preparing the first version of the model model to be trained.

03

Model training

A picture always reinforces the concept

Images reveal large amounts of data, so remember: use an image instead of a long text. Your audience will appreciate it

Models tried

```
# Split the data into a training set and a test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Define the models
models = {
    'Linear Regression': LinearRegression(),
    'Decision Tree': DecisionTreeRegressor(),
    'Random Forest': RandomForestRegressor(),
    'Gradient Boosting': GradientBoostingRegressor(),
    'Support Vector Regression': SVR()
}
```

Initially 5 models were tested on the same number of dataset.

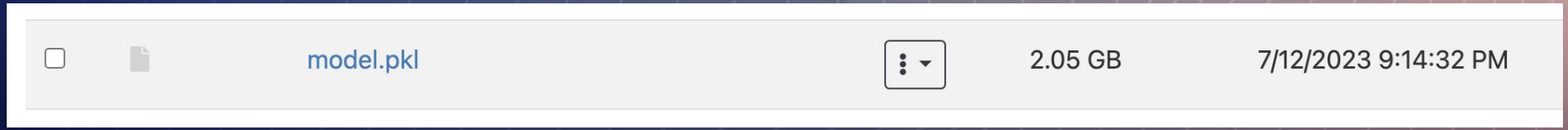
```
Linear Regression Mean Squared Error: 79624.36022025623
Decision Tree Mean Squared Error: 68327.38474590138
Random Forest Mean Squared Error: 31843.975350144592
Gradient Boosting Mean Squared Error: 42338.93848021941
Support Vector Regression Mean Squared Error: 46424.061255180524
best model is  Random Forest
```

Measuring accuracy

Since it’s a regression problem, percentage form couldn’t be used to depict accuracy, so instead Mean squared error was used as the loss function

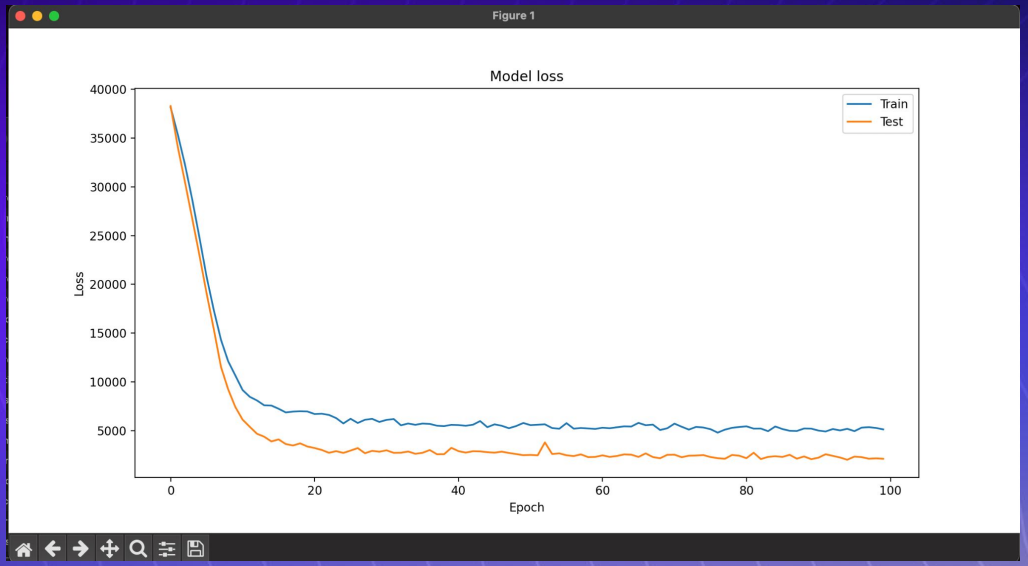
Random forest was used

Model’s size created problems

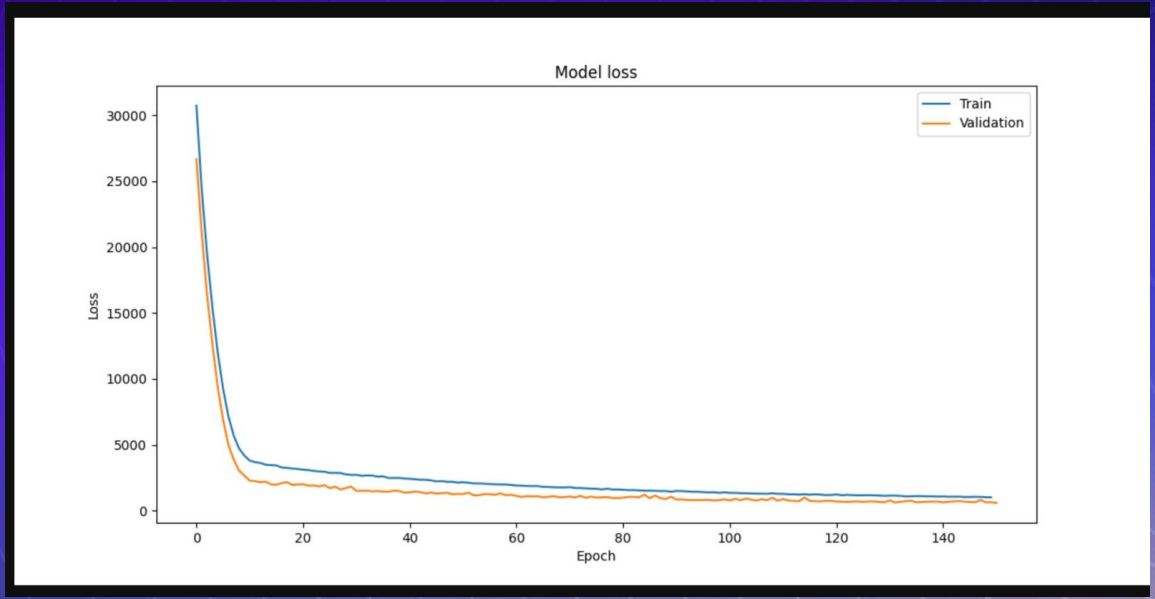


```
# Define the neural network model with L2 regularization and increased dropout
model = Sequential([
    Dense(32, activation='relu', input_shape=(10,), kernel_regularizer=l2(1)),
    BatchNormalization(),
    Dropout(0.05),
    Dense(64, activation='relu', kernel_regularizer=l2(1)),
    BatchNormalization(),
    Dropout(0.6),
    Dense(16, activation='relu', kernel_regularizer=l2(1)),
    BatchNormalization(),
    Dropout(0.05),
    Dense(1)
])
```

Instead a deep neural network was used



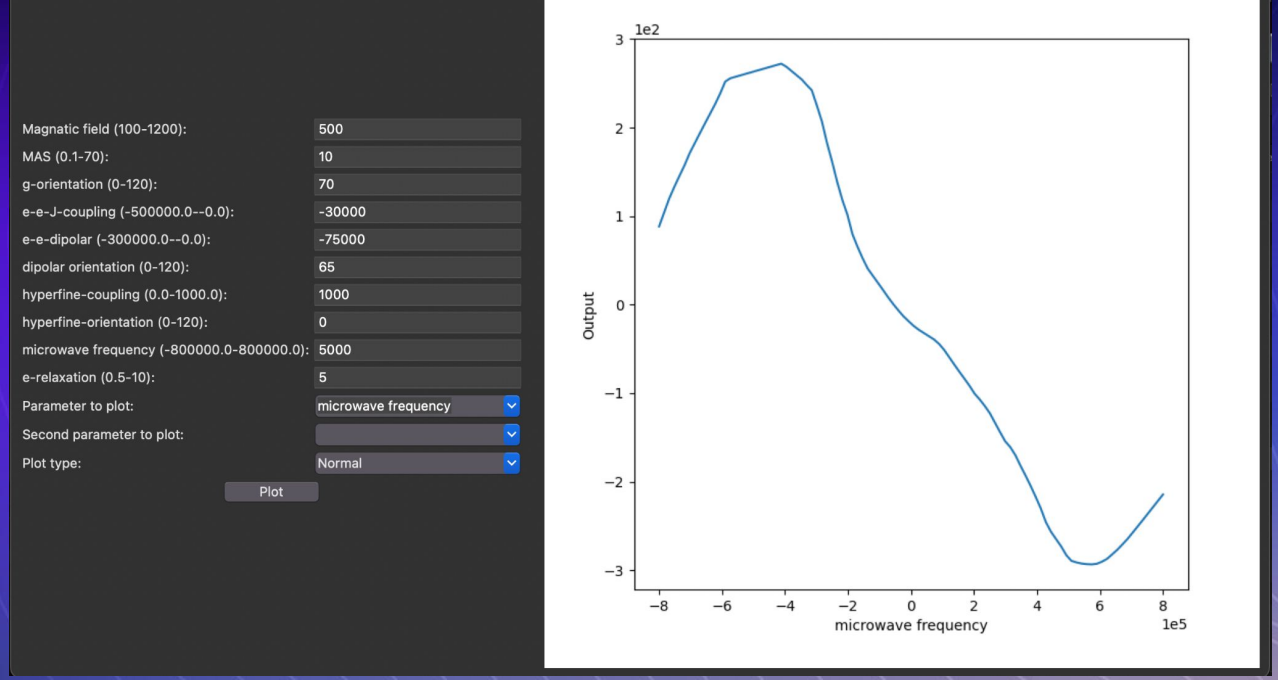
Achieved better results than other machine learning algorithms, achieving 2130 MSE



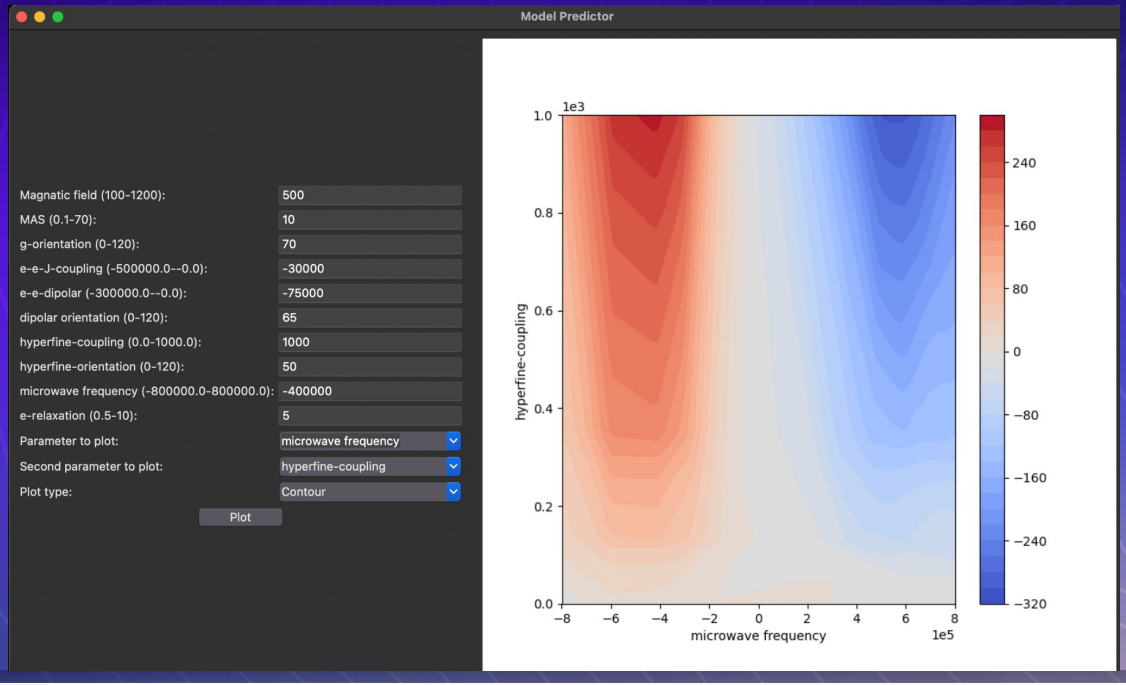
A lot of tweaks were made to achieve MSE of 331

04

Results



Using tkinter & matplot, UI was developed to plot parameters & spin



And later contour plot was implemented to depict relations

Current state?

Thanks!

Do you have any questions?
tmn8829
+971565340617