

Using Machine Learning to Develop a Supernatural Martian Substance

TASK DESCRIPTION

A visionary Company X is on a mission to build a self-sustaining city on the red planet. As part of this mission, scientists at Company X are developing a supernatural substance that can withstand Mars' extraterrestrial conditions. This substance will be the building block of this futuristic Martian city. Can Company X turn this idea into a reality? In this assignment, you will use machine learning to predict the quality of this substance based on historical data.

The task is formulated as a regression problem where you will predict a numerical metric that measures the quality of the Martian substance. You will be evaluated with the Mean Absolute Error (MAE) metric. Be aware that MAE is a negatively-oriented score, which means lower values are better!

DATASET DESCRIPTION

Number of instances: 4000

Number of attributes: 15

Target Variable: A numerical quality metric of the Martian substance that Company X is developing.

Attribute Information:

Name	Description	Type	Values
1	Chemical property A of the substance	Numerical	Positive and negative float
2	Chemical property B of the substance	Numerical	Positive and negative float
3	Chemical property C of the substance	Numerical	Positive and negative float
4	Chemical property D of the substance	Numerical	Positive and negative float
5	Chemical property E of the substance	Numerical	Positive and negative float
6	Chemical property F of the substance	Numerical	Positive and negative float
7	Chemical property G of the substance	Numerical	Positive and negative float
8	Whether or not a refraction test was performed on the substance	Categorical	0: Yes 1: No
9	Physical property A of the substance	Numerical	Positive and negative float
10	Whether or not a radioactivity test was performed on the substance	Categorical	0: Yes 1: No
11	Physical property B of the substance	Numerical	Positive and negative float
12	Physical property C of the substance	Numerical	Positive and negative float
13	Physical property D of the substance	Numerical	Positive and negative float
14	Physical property E of the substance	Numerical	Positive and negative float
15	Physical property F of the substance	Numerical	Positive and negative float

PREDICTION FILE SUBMISSION

You are kindly requested to strictly follow the described submission guidelines:

File Format: .csv

Filename: Student code (e.g. 123456.csv)

Column Format: 1 Column named *Prediction*

Row Format: Your predictions with the same number of rows and in the same order as the test set provided to you (see below)

Test Set													Prediction File		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Prediction
0.365235269938977	1.529.306.679.125.550	-11.908.325.762.181.000	-0.263735698263087	1.450.189.650.488.430	0.483842732979919	0	-16.082.445.485.643.800	1	-0.9042028428578651	0.944485812292873	0	4	-0.8018894404850599		
-0.70728978124798	-0.5630169524059513	-0.5630169524059513	-10.383.851.954.302.000	2	13.077.711.451.335.400	0.5326974361944942	0	0.22504050679075867	1	15.329.911.305.031.300	0	6	-0.7594260736227033	0	
0.791887817278419	-0.987884906401598	-0.294548853290141	-11.831.857.899.074.100	2	2.484.571.338.236.800	-54.434.639.950.812.500	0	-5.130.388.873.819.700	1	-0.548014322059049	-11.186.332.753.448.800	1	6	0.58411675933324	
-18.414.023.316.700.700	0.5843762635189957	-0.44542974588191	-11.215.808.852.813.300	2	-0.44032978918683196	0.5654039903754975	0	0.7775883532462854	0	12.640.684.488.741.800	0.14841113460102745	2	6	-0.350215033115721	
10.586.220.513.803.800	-0.3972468750292362	0.598326369590149	-0.18065915201813089	1	0.4420134639428853	-0.10632032369331388	0	-0.3098370330077032	1	-0.85405294502979709	0.2219592800272947	1	6	-11.909.328.736.422.800	
-0.8841884185029327	-1.175.028.415.513.880	-0.24746412571718943	0.872786358847538	3	0.5370940028853139	-0.4442215334897408	0	-0.4583377147048573	1	-0.1688322312778673	0.999744865243987	1	6	15.833.137.846.544.800	
-0.2514828678432137	-0.3873017812020503	-0.5178820887734943	0.48073820887150005	2	0.05088781632810596	0.5173762398848938	0	-0.44803035037159916	1	18.908.508.238.100.100	0.8489794415106852	1	6	0.6850265578331893	
-15.09.071.700.771.100	-13.285.759.871.132.000	-158.910.233.613.374	-0.1114391320506399	2	0.39665207772167777	-12.286.487.505.513.900	1	-0.8192172844378497	1	-0.6793708599098992	-0.2807788448623637	1	1	15.885.013.764.382.800	
0.28875128632702384	-13.487.801.441.423.000	0.328197047084938	-0.1681544020058158	1	-11.031.343.881.238.000	-0.639739113020049	0	0.848791146238342	0	-3.725841157898493	0.851987770060531	1	6	-102.838.016.411.807	
-0.001301918935988938	-0.914768884570248	-15.864.880.157.508.100	-12.881.050.700.471.800	2	-0.782894952584685	0.7394150604595858	1	10.607.047.158.728.100	0	-0.08283005301529053	-0.21308580670328187	0	6	0.1288697082297855	
0.3071029635368835	-0.3709627371049143	-0.7298158913960502	-0.12994823872288403	1	0.527637371899762	0.806.889.814.446.570	1	-11.702.234.763.314.800	1	0.489382131722673	-0.7698670813112125	0	6	-12.044.745.163.658.800	
-33.626.378.958.448.800	0.861603489151878	0.8611036981744702	0.481588463188013	1	-0.276368033740891	0.891028133888331	0	-0.4038187379040428	1	0.3503002044450289	15.818.448.427.178.800	1	6	0.368465071274213	
0.829744151050804	0.858854228068923	0.1561474135887276	0.44885442220568	1	-0.32729160878284217	16.327.817.872.288.300	0	0.679541838536523	0	0.6782882030500649	-0.1508235735157416	2	6	-0.23312488586820477	
0.863488418849887	0.815533493579153	0.138828918997153	0.1433347636344038	2	-11.248.407.408.273.400	0.848410198487543	0	-0.462938720881734	1	18.970.580.211.005.300	10.817.815.159.236.300	0	1	0.844803510714508	
-14.324.791.029.747.800	-0.27858883210982	-0.834308513727007	0.131098898722512	2	-682.020.274.880.380	-1.624.237.411.515.300	1	0.31548027440915103	0	-0.841898884845438	0.833530522255838	1	3	-2.131.885.880.580.200	
-16.887.100.245.538.800	17.284.628.658.045.100	1.059.501.483.678.230	-0.7889147299234826	2	0.06382147754202918	-0.17074278783088417	1	-10.783.972.194.581.200	1	0.6887243918070871	-15.335.855.851.727.800	1	6	-0.19063555979224614	
17.806.381.587.804.400	0.23889890388802711	0.02874201154460739	0.10874201154460739	1	15.843.881.259.035.200	18.928.824.391.886.100	0	-0.8609091891831862	1	0.11378802416728882	-0.839982321897369	1	6	-0.071.689.530.038.080	
13.876.887.580.537.000	0.00648773056719225	-10.248.570.058.880.100	-0.634410844501616	2	-0.696548180228884	-17.529.141.084.227.700	0	-0.845886073171077	0	0.3539114827280363	-0.88852584347459	1	6	0.8485110287551428	
0.21984367370104705	0.5809771854884844	-0.816887710643911	-1.484.916.325.313.290	1	0.4806849023887612	11.543.401.870.713.800	1	-0.7811781416791248	0	0.23894201521462187	-0.685006139118863	1	1	-21.834.873.178.558.800	
-0.150097085789543	-0.8028028187752672	-0.757038859520576	-0.878113348603307	0	0.231088845795007	-18.48.444.873.106.300	1	-12.13.182.787.430.800	0	-0.07888888888888887	0.491445321373648	1	4	-0.2772289842184158	
-0.684485943142113	-0.860302029144834	-0.720048347173051	12.530.720.270.583.800	2	0.3668813477199376	-0.4247749231938782	0	-0.826.881.823.798.800	1	-0.3545556543899115	-1.857.154.837.884.800	1	6	0.77757118684777	
-0.1831884311982837	0.518143030943118	-0.3420353072291	-0.8423527039041347	1	0.8447137403975882	0.58221338981313	0	0.18882676956582864	0	-0.0883873063441609	-0.11188827206340036	1	6	-0.788705630945958	
14.502.278.355.858.000	-11.152.281.231.788.700	-0.277316284022451	12.054.778.429.319.000	2	-0.583733905255578	2.203.426.252.788.710	0	-0.14788037803491802	0	-0.8337515591862967	0.01042080760161887	1	6	0.866723801402757	
-0.778989878835357	0.8781120035314846	-0.598157580514151	0.04235014393878695	1	0.203880008118977	-0.4285327257733995	24.801.154.643.083.800	-19.777.533.827.178.100	1	1.468.879.580.903.230	0	4	-0.697972714842318		
0.6984201882832378	17.881.041.981.882.800	2.182.853.813.048.400	-17.571.108.185.213.800	1	-0.130.728.563.158.800	-18.078.659.010.298.600	1	-18.086.875.482.147.100	1	-0.58632923293198418	0.34148638252643843	1	4	-22.845.327.286.484.100	
-0.865448371744143	12.838.134.558.170.300	-1.220.141.308.083.290	0.475349059186236	2	-0.20201190124474564	13.553.094.040.702.300	0	-0.6882148052344887	0	-0.4589703890002059	-0.208831172588749	2	6	-0.127382722841035	
0.518116143036657	-0.160410484702032	-0.158284327226564	-0.1680683163831756	2	0.6038344211349183	-10.047.387.128.428.800	0	-1.130.885.702.282.000	1	-0.294020086550213	0.8647386353518812	2	6	0.8518051350831981	
1.332.800.422.422.800	-6.816970320076151	-0.73151837854082	0.3788891410007215	2	0.388825318348704	-0.140908084071418	1	0.8341002640579838	1	11.117.017.451.038.000	-0.8612252631972743	1	4	0.306538963135771	
-0.259727705436854	-0.1844271888688338	3.878.204.038.183.180	0.84203200077142	0	0.3148027054843335	-19.690.839.739.340.000	0	0.02353436288765617	0	0.32438671677154287	-0.4803843883318073	1	4	-2.237.003.434.870.000	
-0.516165.086.365.000	-13.231.388.547.874.300	-0.2988607977858824	0.20248738156093827	1	-18.672.740.363.373.000	-1	-0.5356952621847803	0	-0.35183385344704015	-0.8301998512428323	1	0	0.0008993148800058		

PDF FILE SUBMISSION

Along with your predictions, you are asked to kindly submit the following supporting information:

1. A brief description of the step by step methodology (i.e. data cleaning, pre-processing, training, evaluation, etc.) that you have followed to do the assignment, with the aim of illustrating the motivation behind your selected approach.
2. The python code that you used to do the assignment, with a fair amount of comments within the code to ensure that they can be clearly understood.

File Format: .pdf

Filename: Student code (e.g. 123456.pdf)

ASSIGNMENT DEADLINES

- A. November 27th - The labeled dataset will be released on the BeeP platform;
- B. December 4th – The unlabeled dataset will be released on the BeeP platform;
- C. December 5th to December 7th - ASSIGNMENT DUE DATE: you are requested to upload both your PREDICTION FILE and the PDF FILE on the BeeP platform. The Assignment folder will be open from December 5th to December 7th at 19pm.