Prakriti Data Analytics

08.03.2020

Team Name - WAKANDA_FOR3V3R

Ayush Kumar   9661179290   17CS10007 IIT KHARAGPUR

Anshul Choudhary   8918500798   17CS10005 IIT KHARAGPUR

Prakhar Bindal   9593801201   17CS10036 IIT KHARAGPUR

Shubham Raj   9733331647   17CE10054 IIT KHARAGPUR

Overview

Plants require many mineral elements for their nutrition. These include the macronutrients nitrogen (N), potassium (K), calcium (Ca), magnesium (Mg), carbon (C) and sulphur (S) and the micronutrients iron (Fe), manganese (Mn), zinc (Zn), nickel (Ni), titanium(Ti), rubidium(Rb), silicon(Si) and strontium(Sr). Plant health highly depends upon the availability of these nutrients in the soil that can be estimated through several advanced techniques such as spectrometry, spectroradiometer, and PXRF(Portable X-ray Fluorescence) spectroscopy etc.

Goals:

Determination of RMSE score and accuracy of total nitrogen and total carbon percentage present in the soil on the basis of different micronutrient amounts and spectral data using different machine learning models.

Methods

Machine learning models are determined using the following steps:

1. Data Cleaning
2. Feature Engineering
3. Data Splitting - Training and Testing data
4. Model Training on the training data
5. Model testing, accuracy and RMSE determination
6. Evaluation

The following diagram represents the steps described.
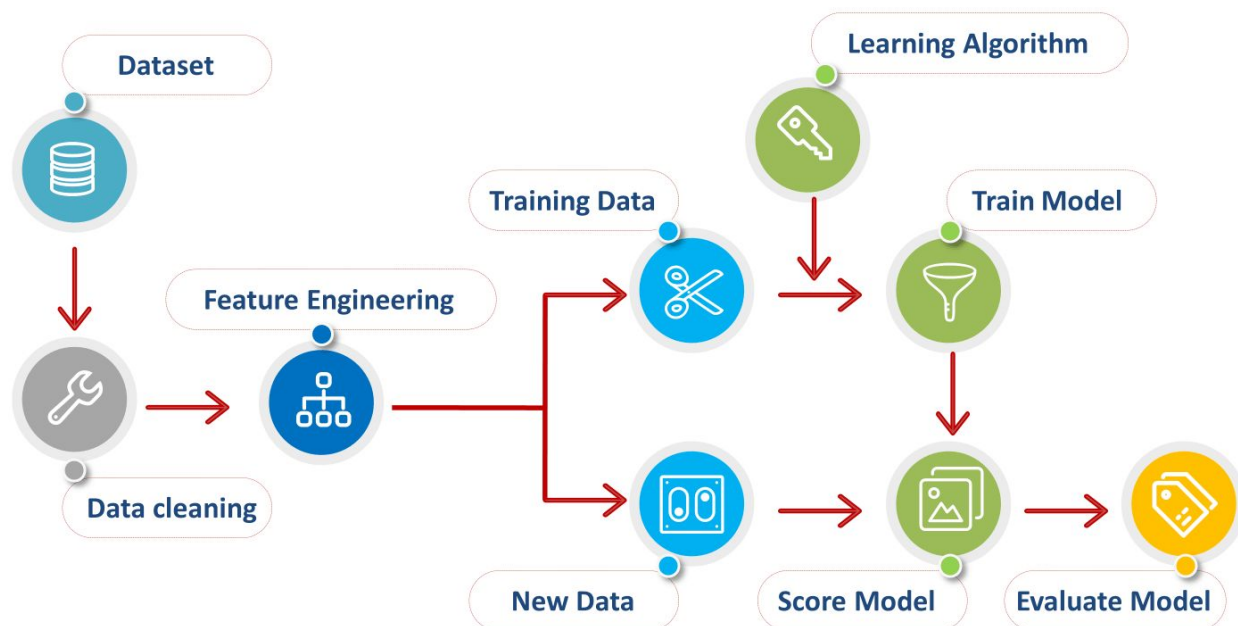
Data Cleaning

The data given contains data of TC(%) and TN(%) as n.d. and 0, as well as certain spectral values are missing. These data examples need to be taken out of consideration while we train and test the model.

Feature Engineering

Data is so sparse, having a lot of features. So first, we did Normalization and then extracted the influential features, who contributed more to the data set using random forest as base model. Influential variables data is shown in the appendix section.

Data Splitting

The data is split into training and testing data set as 80% of data is chosen randomly to be the training data and the rest is testing data.

Model Training

Since the number of training examples are very low compared to number of features. We avoid using Deep neural networks. We used different Machine Learning Algorithms like Random Forest, XGBoost, Support Vector Machine to predict the Total Carbon and Total Nitrogen from Elemental, Spectral and combined features. Random Forest Regressor gives promising results with a number of trees equal to 500.

Model Testing

The model is tested on the testing data with the following result shown below in the appendix section.

Appendix

MODEL - Random Forest Regressor (part A and B)

Random forest is an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.

Accuracy

vs	Spectral Data	Elemental Data	Combined Data
TC(%)	93.787	91.157	93.824
TN(%)	75.496	51.373	77.376

RMSE Score

vs	Spectral Data	Elemental Data	Combined Data
TC(%)	1.634	1.949	1.629
TN(%)	0.169	0.239	0.163

No. of influential variables

vs	Spectral Data (out of 2149)	Elemental Data (out of 11)	Combined Data (out of 2160)
TC(%)	135	3	97
TN(%)	356	3	325

Note : Our model uses random data splitting. So every time you run the model, the result may vary slightly from the above shown data.

Influential variables determined using correlation for each method (part C) :

1) Spectral Data to predict TC :

Influential Variables (wavelength) : [388, 402, 410, 411, 423, 434, 987, 991, 992, 1217, 1218, 1229, 1231, 1262, 1297, 1419, 1421, 1434, 1438, 1439, 1440, 1441, 1495, 1528, 1529, 1567, 1585, 1586, 1599, 1656, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1679, 1760, 1761, 1762, 1763, 1775, 1776, 1846, 1907, 2022, 2096, 2098, 2099, 2100, 2136, 2137, 2138, 2139, 2141, 2142, 2144, 2145, 2146, 2147, 2148, 2152, 2153, 2154, 2155, 2156, 2159, 2160, 2164, 2165, 2166, 2169, 2170, 2171, 2172, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2187, 2188, 2189, 2190, 2217, 2218, 2230, 2231, 2234, 2236, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2323, 2324, 2341, 2356, 2401, 2446, 2452, 2459, 2460, 2495]

2) Spectral Data to predict TN :

Influential Variables (wavelength) : [354, 355, 360, 362, 368, 375, 393, 398, 408, 409, 416, 428, 430, 433, 441, 445, 447, 450, 451, 453, 454, 455, 457, 458, 462, 465, 466, 468, 470, 540, 543, 546, 547, 548, 552, 553, 570, 574, 578, 580, 585, 586, 588, 589, 591, 592, 597, 599, 600, 601, 605, 607, 608, 609, 610, 612, 614, 616, 617, 619, 622, 623, 628, 635, 636, 637, 638, 639, 640, 641, 643, 644, 645, 646, 647, 648, 651, 653, 655, 657, 658, 659, 660, 661, 662, 664, 665, 667, 669, 670, 671, 672, 673, 943, 944, 945, 947, 950, 958, 965, 968, 970, 971, 973, 977, 978, 981, 985, 986, 987, 988, 989, 991, 993, 1023, 1025, 1027, 1034, 1112, 1285, 1298, 1301, 1302, 1304, 1306, 1307, 1308, 1312, 1313, 1315, 1317, 1318, 1319, 1320, 1321, 1327, 1330, 1342, 1443, 1444, 1445, 1449, 1452, 1516, 1526, 1530, 1531, 1533, 1534, 1541, 1545, 1546, 1547, 1549, 1550, 1564, 1566, 1567, 1569, 1591, 1593, 1596, 1612, 1614, 1616, 1618, 1619, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1632, 1634, 1677, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1700, 1702, 1703, 1704, 1705, 1714, 1721, 1728, 1738, 1740, 1741, 1742, 1743, 1744, 1745, 1748, 1750, 1755, 1759, 1760, 1761, 1771, 1782, 1783, 1785, 1793, 1797, 1798, 1799, 1800, 1813, 1814, 1817, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2064, 2071, 2077, 2081, 2082, 2083, 2084, 2086, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106,

```
2110, 2112, 2113, 2114, 2115, 2116, 2119, 2124, 2126, 2127,
2130, 2134, 2135, 2140, 2147, 2148, 2149, 2150, 2151, 2152,
2153, 2154, 2155, 2159, 2160, 2161, 2162, 2164, 2169, 2170,
2196, 2199, 2202, 2203, 2204, 2205, 2207, 2208, 2262, 2263,
2269, 2270, 2271, 2272, 2284, 2287, 2289, 2290, 2291, 2292,
2293, 2294, 2297, 2298, 2299, 2301, 2302, 2303, 2313, 2314,
2321, 2322, 2324, 2325, 2327, 2328, 2329, 2330, 2331, 2335,
2336, 2337, 2338, 2339, 2340, 2341, 2345, 2363, 2364, 2365,
2375, 2383, 2385, 2386, 2387, 2388, 2390, 2392, 2393, 2394,
2395, 2397, 2406, 2411, 2446, 2460, 2469, 2476, 2477, 2482,
2488, 2492]
```

3) Elemental Data to predict TC :

Influential Variables (element) : `['Rb', 'Al', 'Si']`

4) Elemental Data to predict TN :

Influential Variables (element) : `['S', 'Ca', 'Ti']`

5) Combined Data to predict TC :

Influential Variables : `['Ca', 'Rb', 'Si', 388, 991, 1219, 1419,`
```
1436, 1439, 1481, 1529, 1566, 1583, 1600, 1657, 1658, 1659,
1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669,
1670, 1761, 1762, 1763, 1776, 1940, 1941, 1948, 2096, 2097,
2098, 2137, 2138, 2139, 2145, 2147, 2148, 2149, 2154, 2155,
2156, 2157, 2158, 2164, 2165, 2166, 2167, 2168, 2169, 2170,
2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2188,
2189, 2194, 2212, 2219, 2234, 2242, 2244, 2245, 2246, 2247,
2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257,
2258, 2259, 2260, 2261, 2262, 2264, 2341, 2446, 2459, 2460]
```

6) Combined Data to predict TN :

Influential Variables : `['S', 'Ca', 'Ti', 'Mn', 'Si', 364, 368,`
```
372, 393, 398, 404, 408, 409, 416, 420, 421, 430, 431, 436,
441, 444, 446, 450, 453, 454, 455, 457, 458, 468, 526, 530,
538, 539, 546, 547, 548, 549, 556, 570, 574, 583, 584, 587,
588, 589, 591, 592, 597, 600, 603, 604, 609, 611, 617, 627,
636, 637, 638, 639, 640, 641, 643, 644, 645, 646, 648, 649,
653, 654, 657, 658, 659, 660, 661, 664, 665, 667, 668, 669,
670, 671, 672, 673, 674, 852, 947, 950, 958, 959, 969, 975,
977, 978, 981, 982, 986, 987, 988, 989, 991, 993, 1020,
1024, 1025, 1100, 1101, 1103, 1111, 1112, 1114, 1299, 1300,
1302, 1305, 1306, 1308, 1309, 1311, 1312, 1313, 1314, 1315,
```

1316, 1318, 1327, 1329, 1330, 1444, 1516, 1525, 1531, 1532,
1533, 1541, 1545, 1546, 1547, 1548, 1549, 1552, 1558, 1561,
1565, 1566, 1567, 1597, 1601, 1610, 1611, 1612, 1613, 1614,
1615, 1616, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625,
1626, 1627, 1628, 1629, 1630, 1631, 1632, 1634, 1635, 1685,
1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1695, 1703,
1704, 1705, 1706, 1721, 1728, 1729, 1733, 1738, 1740, 1741,
1742, 1744, 1748, 1760, 1762, 1771, 1783, 1784, 1785, 1791,
1797, 1798, 1799, 1800, 1810, 1812, 1813, 1814, 1815, 1817,
2002, 2007, 2008, 2009, 2010, 2011, 2063, 2068, 2071, 2072,
2077, 2082, 2084, 2085, 2086, 2093, 2094, 2095, 2096, 2097,
2098, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2113, 2114,
2115, 2116, 2126, 2130, 2135, 2148, 2149, 2150, 2151, 2152,
2153, 2154, 2160, 2161, 2162, 2170, 2198, 2200, 2201, 2202,
2203, 2204, 2205, 2206, 2208, 2209, 2268, 2269, 2270, 2271,
2272, 2273, 2274, 2280, 2283, 2288, 2289, 2290, 2291, 2292,
2293, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2308, 2313,
2314, 2323, 2325, 2327, 2331, 2336, 2337, 2338, 2339, 2340,
2341, 2346, 2364, 2369, 2379, 2383, 2387, 2388, 2389, 2392,
2393, 2394, 2395, 2396, 2397, 2460, 2470, 2475, 2476, 2477,
2479, 2481, 2485]

Thank You