The Greener, the Better

1.2 Summary

The purpose of this project is to apply different machine learning algorithms like classification and clustering models to identify the soil type, species of plants in various forest cover types which will help improve the efficiency of afforestation and sequestration of co2 from the atmosphere. The data set that we selected for our project is of Roosevelt National Forest in Northern Colorado. The idea was to identify the species based on various soil types that contribute the most in Co2 sequestration.

1.3 Project Description

The growing excitement around CO2 sequestration technologies could feed unrealistic expectations about how much we can rely on carbon removal, and thus how much nations and corporations can carry on emitting over the crucial coming decades. Market demands are also likely to steer attention toward cheaper solutions that are not as reliable or long-lasting.

The goal of this project is to anticipate different forms of forest cover in different provinces of the Roosevelt National Forest in Northern Colorado.

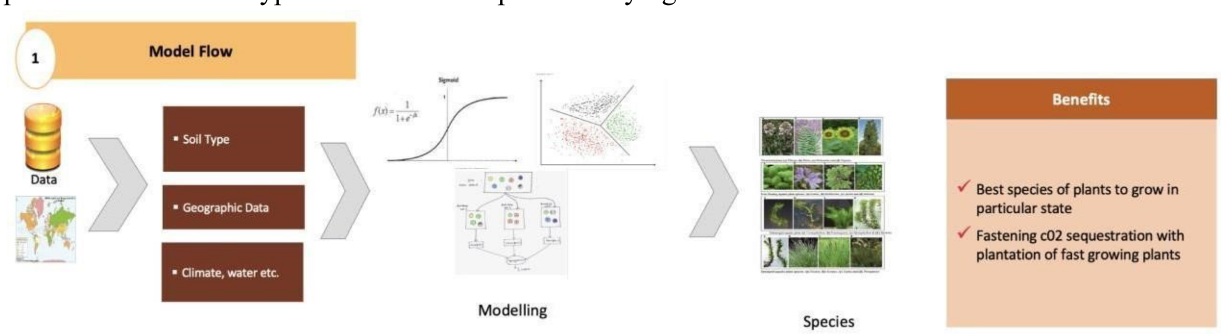
We identified the factors which affect the independent variable, searching for correlations between the different variables in the dataset. trying to ascertain the space between the forest and the water storage, as well as the wildfire sites to define different types of forest cover depending on soil type and conservation area.

Pursuing an all-of-the-above carbon removal portfolio in the United States would provide the most cumulative carbon removal at the lowest risk. It creates the most options for achieving the 2 Gt CO2 removal target by 2050, should any single pathway fail to realize its expected potential.

The objective of this project is to predict the cover type (species) which is more efficient for sequestration of co2 from the atmosphere with respect to each area(state) in order to fight against the adverse climate changes in the future keeping the greenhouse gas away from the atmosphere is the solution as trees are one of the major co2 sequestration agents and afforestation is one of the solutions.

It is important to choose the right species which helps in the sequestration the most. There are various factors involved such as soil type, climatic conditions, wilderness area, distance of the forest from the water reserve, growing time etc.

All species require different conditions to grow and hence need to be identified correctly. For example, a species which might grow faster in one state may not grow at all in another. We have predicted the same through our project for which we have used classification as well as clustering algorithms that will help in identifying the species based on their type and will also help in identifying the distance of the area from the available resources.



For this project to predict different forest cover types we are considering areas of the Roosevelt National Forest of Northern Colorado. We applied classification as well as clustering algorithms such as Naive Bayes, Knn and Decision Tree ,K Means and Gaussian to achieve our goal and then compared the accuracy of all the algorithms to identify the algorithms that identified the cover type most accurately.

1.3.1 Datasets:

The dataset that we have selected has a total 55 columns and 581012 rows. This dataset is suitable for our project as it has features such as soil type, wilderness area, cover type, distance from the water resource and fire points etc. The dataset has 4 types of wilderness areas that are :

1. Rawah
2. Neota
3. Comanche Peak
4. Cache la Poudre

The forest cover types that is given in the dataset are:

1. Spruce/Fir
2. Lodgepole Pine
3. Ponderosa Pine
4. Cottonwood/Willow
5. Aspen
6. Douglas-fir
7. Krummholz

1.3.3 Methodology and Results

1. To check for the correlation between the different variables in the dataset to identify the factors affecting the independent variable.
2. Identifying the distance of the forest area from the water reserve as well as from the firepoints.
3. To classify the forest cover type based on the soil type and wilderness area.

After performing basic cleaning of data we applied Pearson's coefficient correlation to identify the correlation between the variables.

A correlation matrix is a table showing correlation coefficients between variables/categories. Each cell in the table shows the correlation between two categories.

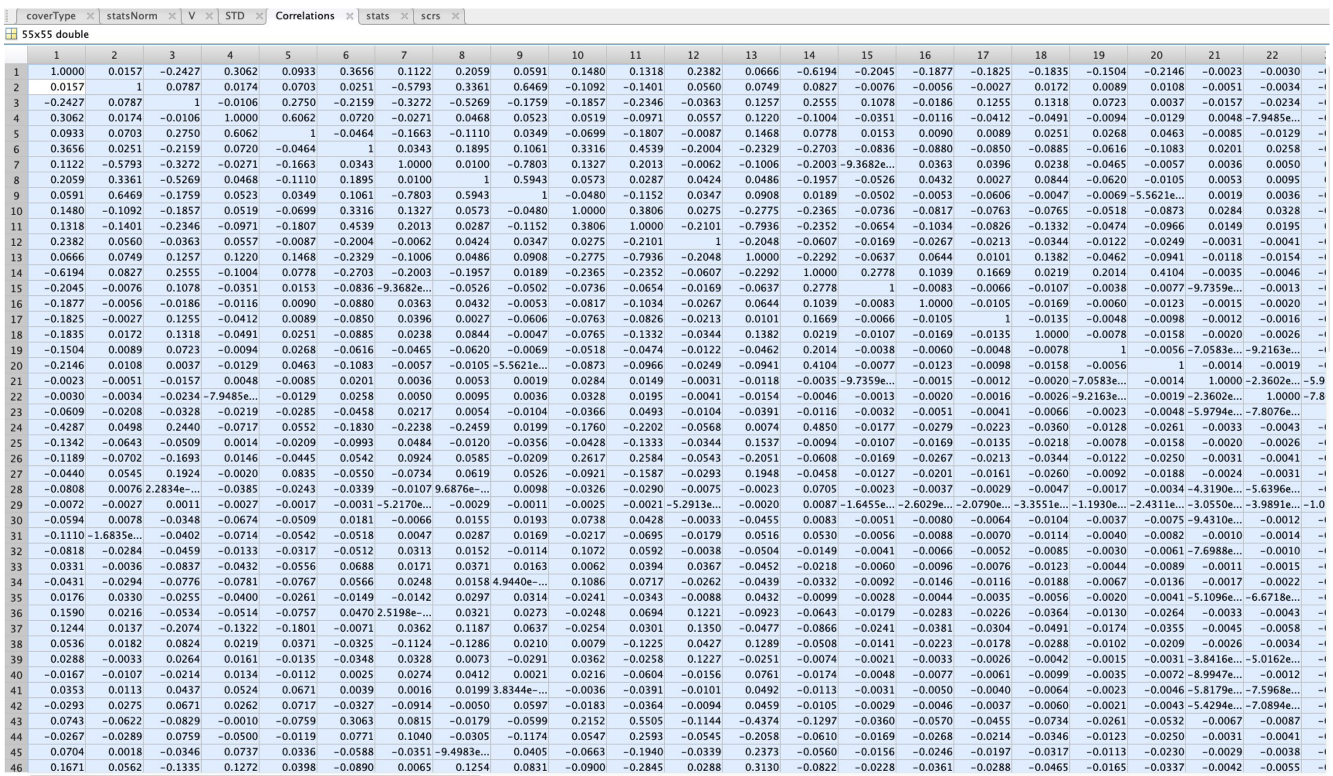
A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses

Category	Correlation
Cover_Type	1
Wilderness_Area4	0.323199554

Soil_Type10	0.243876302
Soil_Type38	0.160169609
Soil_Type39	0.155668261
Slope	0.148285405
Soil_Type40	0.1283513
Soil_Type2	0.11813526
Soil_Type6	0.112958278
Soil_Type4	0.099671864
Soil_Type1	0.090828152
Soil_Type17	0.090582304
Vertical_Distance_To_Hydrology	0.081664022
Soil_Type35	0.08031505
Soil_Type37	0.080271474
Soil_Type5	0.077889961
Soil_Type3	0.068064455
Wilderness_Area3	0.066845643
Soil_Type14	0.065562072
Soil_Type11	0.035378738
Soil_Type36	0.025396537
Soil_Type13	0.024403656
Aspect	0.017079802
Soil_Type16	0.009844345
Soil_Type18	0.007390382
Soil_Type15	0.006424845
Soil_Type34	0.004642623

Soil_Type26	-0.000374843
Soil_Type7	-0.000495517
Soil_Type28	-0.001702393
Soil_Type8	-0.003666794
Soil_Type9	-0.006109534
Soil_Type25	-0.006449047
Soil_Type30	-0.010436449
Soil_Type27	-0.01440654
Horizontal_Distance_To_Hydrology	-0.020316622
Soil_Type12	-0.023601134
Soil_Type21	-0.025400202
Soil_Type20	-0.02866471
Hillshade_9am	-0.035415004
Soil_Type19	-0.036451924
Wilderness_Area2	-0.04805895
Hillshade_3pm	-0.04828973
Soil_Type33	-0.062501747
Soil_Type31	-0.065347069
Soil_Type24	-0.068745879
Soil_Type32	-0.075562026
Hillshade_Noon	-0.096426002
Horizontal_Distance_To_Fire_Points	-0.108935536
Soil_Type29	-0.124932598
Soil_Type23	-0.135055171
Soil_Type22	-0.141746119

Horizontal_Distance_To_Roadways	-0.153449759
Wilderness_Area1	-0.203913214
Elevation	-0.269553778



coverType | statsNorm | V | STD | Correlations | stats | scrs

55x55 double

	1	2	3	4	5	6	7	8	9	10	11
1	1.0000	0.0157	-0.2427	0.3062	0.0933	0.3656	0.1122	0.2059	0.0591	0.1480	0.1318
2	0.0157	1	0.0787	0.0174	0.0703	0.0251	-0.5793	0.3361	0.6469	-0.1092	-0.1401
3	-0.2427	0.0787	1	-0.0106	0.2750	-0.2159	-0.3272	-0.5269	-0.1759	-0.1857	-0.2346
4	0.3062	0.0174	-0.0106	1.0000	0.6062	0.0720	-0.0271	0.0468	0.0523	0.0519	-0.0971
5	0.0933	0.0703	0.2750	0.6062	1	-0.0464	-0.1663	-0.1110	0.0349	-0.0699	-0.1807
6	0.3656	0.0251	-0.2159	0.0720	-0.0464	1	0.0343	0.1895	0.1061	0.3316	0.4539
7	0.1122	-0.5793	-0.3272	-0.0271	-0.1663	0.0343	1.0000	0.0100	-0.7803	0.1327	0.2013
8	0.2059	0.3361	-0.5269	0.0468	-0.1110	0.1895	0.0100	1	0.5943	0.0573	0.0287
9	0.0591	0.6469	-0.1759	0.0523	0.0349	0.1061	-0.7803	0.5943	1	-0.0480	-0.1152
10	0.1480	-0.1092	-0.1857	0.0519	-0.0699	0.3316	0.1327	0.0573	-0.0480	1.0000	0.3806
11	0.1318	-0.1401	-0.2346	-0.0971	-0.1807	0.4539	0.2013	0.0287	-0.1152	0.3806	1.0000
12	0.2382	0.0560	-0.0363	0.0557	-0.0087	-0.2004	-0.0062	0.0424	0.0347	0.0275	-0.2101
13	0.0666	0.0749	0.1257	0.1220	0.1468	-0.2329	-0.1006	0.0486	0.0908	-0.2775	-0.7936
14	-0.6194	0.0827	0.2555	-0.1004	0.0778	-0.2703	-0.2003	-0.1957	0.0189	-0.2365	-0.2352
15	-0.2045	-0.0076	0.1078	-0.0351	0.0153	-0.0836	-9.3682e...	-0.0526	-0.0502	-0.0736	-0.0654
16	-0.1877	-0.0056	-0.0186	-0.0116	0.0090	-0.0880	0.0363	0.0432	-0.0053	-0.0817	-0.1034
17	-0.1825	-0.0027	0.1255	-0.0412	0.0089	-0.0850	0.0396	0.0027	-0.0606	-0.0763	-0.0826
18	-0.1835	0.0172	0.1318	-0.0491	0.0251	-0.0885	0.0238	0.0844	-0.0047	-0.0765	-0.1332
19	-0.1504	0.0089	0.0723	-0.0094	0.0268	-0.0616	-0.0465	-0.0620	-0.0069	-0.0518	-0.0474
20	-0.2146	0.0108	0.0037	-0.0129	0.0463	-0.1083	-0.0057	-0.0105	-5.5621e...	-0.0873	-0.0966
21	-0.0023	-0.0051	-0.0157	0.0048	-0.0085	0.0201	0.0036	0.0053	0.0019	0.0284	0.0149
22	-0.0030	-0.0034	-0.0234	-7.9485e...	-0.0129	0.0258	0.0050	0.0095	0.0036	0.0328	0.0195
23	-0.0609	-0.0208	-0.0328	-0.0219	-0.0285	-0.0458	0.0217	0.0054	-0.0104	-0.0366	0.0493
24	-0.4287	0.0498	0.2440	-0.0717	0.0552	-0.1830	-0.2238	-0.2459	0.0199	-0.1760	-0.2202
25	-0.1342	-0.0643	-0.0509	0.0014	-0.0209	-0.0993	0.0484	-0.0120	-0.0356	-0.0428	-0.1333
26	-0.1189	-0.0702	-0.1693	0.0146	-0.0445	0.0542	0.0924	0.0585	-0.0209	0.2617	0.2584
27	-0.0440	0.0545	0.1924	-0.0020	0.0835	-0.0550	-0.0734	0.0619	0.0526	-0.0921	-0.1587
28	-0.0808	0.0076	2.2834e-...	-0.0385	-0.0243	-0.0339	-0.0107	9.6876e-...	0.0098	-0.0326	-0.0290
29	-0.0072	-0.0027	0.0011	-0.0027	-0.0017	-0.0031	-5.2170e...	-0.0029	-0.0011	-0.0025	-0.0021
30	-0.0594	0.0078	-0.0348	-0.0674	-0.0509	0.0181	-0.0066	0.0155	0.0193	0.0738	0.0428
31	-0.1110	-1.6835e...	-0.0402	-0.0714	-0.0542	-0.0518	0.0047	0.0287	0.0169	-0.0217	-0.0695
32	-0.0818	-0.0284	-0.0459	-0.0133	-0.0317	-0.0512	0.0313	0.0152	-0.0114	0.1072	0.0592
33	0.0331	-0.0036	-0.0837	-0.0432	-0.0556	0.0688	0.0171	0.0371	0.0163	0.0062	0.0394
34	-0.0431	-0.0294	-0.0776	-0.0781	-0.0767	0.0566	0.0248	0.0158	4.9440e-...	0.1086	0.0717
35	0.0176	0.0330	-0.0255	-0.0400	-0.0261	-0.0149	-0.0142	0.0297	0.0314	-0.0241	-0.0343
36	0.1590	0.0216	-0.0534	-0.0514	-0.0757	0.0470	2.5198e-...	0.0321	0.0273	-0.0248	0.0694
37	0.1244	0.0137	-0.2074	-0.1322	-0.1801	-0.0071	0.0362	0.1187	0.0637	-0.0254	0.0301
38	0.0536	0.0182	0.0824	0.0219	0.0371	-0.0325	-0.1124	-0.1286	0.0210	0.0079	-0.1225
39	0.0288	-0.0033	0.0264	0.0161	-0.0135	-0.0348	0.0328	0.0073	-0.0291	0.0362	-0.0258
40	-0.0167	-0.0107	-0.0214	0.0134	-0.0112	0.0025	0.0274	0.0412	0.0021	0.0216	-0.0604
41	0.0353	0.0113	0.0437	0.0524	0.0671	0.0039	0.0016	0.0199	3.8344e-...	-0.0036	-0.0391
42	-0.0293	0.0275	0.0671	0.0262	0.0717	-0.0327	-0.0914	-0.0050	0.0597	-0.0183	-0.0364
43	0.0743	-0.0622	-0.0829	-0.0010	-0.0759	0.3063	0.0815	-0.0179	-0.0599	0.2152	0.5505
44	-0.0267	-0.0289	0.0759	-0.0500	-0.0119	0.0771	0.1040	-0.0305	-0.1174	0.0547	0.2593
45	0.0704	0.0018	-0.0346	0.0737	0.0336	-0.0588	-0.0351	-9.4983e...	0.0405	-0.0663	-0.1940
46	0.1671	0.0562	-0.1335	0.1272	0.0398	-0.0890	0.0065	0.1254	0.0831	-0.0900	-0.2845

	12	13	14	15	16	17	18	19	20	21	22
1	0.2382	0.0666	-0.6194	-0.2045	-0.1877	-0.1825	-0.1835	-0.1504	-0.2146	-0.0023	-0.0030
2	0.0560	0.0749	0.0827	-0.0076	-0.0056	-0.0027	0.0172	0.0089	0.0108	-0.0051	-0.0034
3	-0.0363	0.1257	0.2555	0.1078	-0.0186	0.1255	0.1318	0.0723	0.0037	-0.0157	-0.0234
4	0.0557	0.1220	-0.1004	-0.0351	-0.0116	-0.0412	-0.0491	-0.0094	-0.0129	0.0048	-7.9485e...
5	-0.0087	0.1468	0.0778	0.0153	0.0090	0.0089	0.0251	0.0268	0.0463	-0.0085	-0.0129
6	-0.2004	-0.2329	-0.2703	-0.0836	-0.0880	-0.0850	-0.0885	-0.0616	-0.1083	0.0201	0.0258
7	-0.0062	-0.1006	-0.2003	-9.3682e...	0.0363	0.0396	0.0238	-0.0465	-0.0057	0.0036	0.0050
8	0.0424	0.0486	-0.1957	-0.0526	0.0432	0.0027	0.0844	-0.0620	-0.0105	0.0053	0.0095
9	0.0347	0.0908	0.0189	-0.0502	-0.0053	-0.0606	-0.0047	-0.0069	-5.5621e...	0.0019	0.0036
10	0.0275	-0.2775	-0.2365	-0.0736	-0.0817	-0.0763	-0.0765	-0.0518	-0.0873	0.0284	0.0328
11	-0.2101	-0.7936	-0.2352	-0.0654	-0.1034	-0.0826	-0.1332	-0.0474	-0.0966	0.0149	0.0195
12	1	-0.2048	-0.0607	-0.0169	-0.0267	-0.0213	-0.0344	-0.0122	-0.0249	-0.0031	-0.0041
13	-0.2048	1.0000	-0.2292	-0.0637	0.0644	0.0101	0.1382	-0.0462	-0.0941	-0.0118	-0.0154
14	-0.0607	-0.2292	1.0000	0.2778	0.1039	0.1669	0.0219	0.2014	0.4104	-0.0035	-0.0046
15	-0.0169	-0.0637	0.2778	1	-0.0083	-0.0066	-0.0107	-0.0038	-0.0077	-9.7359e...	-0.0013
16	-0.0267	0.0644	0.1039	-0.0083	1.0000	-0.0105	-0.0169	-0.0060	-0.0123	-0.0015	-0.0020
17	-0.0213	0.0101	0.1669	-0.0066	-0.0105	1	-0.0135	-0.0048	-0.0098	-0.0012	-0.0016
18	-0.0344	0.1382	0.0219	-0.0107	-0.0169	-0.0135	1.0000	-0.0078	-0.0158	-0.0020	-0.0026
19	-0.0122	-0.0462	0.2014	-0.0038	-0.0060	-0.0048	-0.0078	1	-0.0056	-7.0583e...	-9.2163e...
20	-0.0249	-0.0941	0.4104	-0.0077	-0.0123	-0.0098	-0.0158	-0.0056	1	-0.0014	-0.0019
21	-0.0031	-0.0118	-0.0035	-9.7359e...	-0.0015	-0.0012	-0.0020	-7.0583e...	-0.0014	1.0000	-2.3602e...
22	-0.0041	-0.0154	-0.0046	-0.0013	-0.0020	-0.0016	-0.0026	-9.2163e...	-0.0019	-2.3602e...	1.0000
23	-0.0104	-0.0391	-0.0116	-0.0032	-0.0051	-0.0041	-0.0066	-0.0023	-0.0048	-5.9794e...	-7.8076e...
24	-0.0568	0.0074	0.4850	-0.0177	-0.0279	-0.0223	-0.0360	-0.0128	-0.0261	-0.0033	-0.0043
25	-0.0344	0.1537	-0.0094	-0.0107	-0.0169	-0.0135	-0.0218	-0.0078	-0.0158	-0.0020	-0.0026
26	-0.0543	-0.2051	-0.0608	-0.0169	-0.0267	-0.0213	-0.0344	-0.0122	-0.0250	-0.0031	-0.0041
27	-0.0293	0.1948	-0.0458	-0.0127	-0.0201	-0.0161	-0.0260	-0.0092	-0.0188	-0.0024	-0.0031
28	-0.0075	-0.0023	0.0705	-0.0023	-0.0037	-0.0029	-0.0047	-0.0017	-0.0034	-4.3190e...	-5.6396e...
29	-5.2913e...	-0.0020	0.0087	-1.6455e...	-2.6029e...	-2.0790e...	-3.3551e...	-1.1930e...	-2.4311e...	-3.0550e...	-3.9891e...
30	-0.0033	-0.0455	0.0083	-0.0051	-0.0080	-0.0064	-0.0104	-0.0037	-0.0075	-9.4310e...	-0.0012
31	-0.0179	0.0516	0.0530	-0.0056	-0.0088	-0.0070	-0.0114	-0.0040	-0.0082	-0.0010	-0.0014
32	-0.0038	-0.0504	-0.0149	-0.0041	-0.0066	-0.0052	-0.0085	-0.0030	-0.0061	-7.6988e...	-0.0010
33	0.0367	-0.0452	-0.0218	-0.0060	-0.0096	-0.0076	-0.0123	-0.0044	-0.0089	-0.0011	-0.0015
34	-0.0262	-0.0439	-0.0332	-0.0092	-0.0146	-0.0116	-0.0188	-0.0067	-0.0136	-0.0017	-0.0022
35	-0.0088	0.0432	-0.0099	-0.0028	-0.0044	-0.0035	-0.0056	-0.0020	-0.0041	-5.1096e...	-6.6718e...
36	0.1221	-0.0923	-0.0643	-0.0179	-0.0283	-0.0226	-0.0364	-0.0130	-0.0264	-0.0033	-0.0043
37	0.1350	-0.0477	-0.0866	-0.0241	-0.0381	-0.0304	-0.0491	-0.0174	-0.0355	-0.0045	-0.0058
38	0.0427	0.1289	-0.0508	-0.0141	-0.0223	-0.0178	-0.0288	-0.0102	-0.0209	-0.0026	-0.0034
39	0.1227	-0.0251	-0.0074	-0.0021	-0.0033	-0.0026	-0.0042	-0.0015	-0.0031	-3.8416e...	-5.0162e...
40	-0.0156	0.0761	-0.0174	-0.0048	-0.0077	-0.0061	-0.0099	-0.0035	-0.0072	-8.9947e...	-0.0012
41	-0.0101	0.0492	-0.0113	-0.0031	-0.0050	-0.0040	-0.0064	-0.0023	-0.0046	-5.8179e...	-7.5968e...
42	-0.0094	0.0459	-0.0105	-0.0029	-0.0046	-0.0037	-0.0060	-0.0021	-0.0043	-5.4294e...	-7.0894e...
43	-0.1144	-0.4374	-0.1297	-0.0360	-0.0570	-0.0455	-0.0734	-0.0261	-0.0532	-0.0067	-0.0087
44	-0.0545	-0.2058	-0.0610	-0.0169	-0.0268	-0.0214	-0.0346	-0.0123	-0.0250	-0.0031	-0.0041
45	-0.0339	0.2373	-0.0560	-0.0156	-0.0246	-0.0197	-0.0317	-0.0113	-0.0230	-0.0029	-0.0038
46	0.0288	0.3130	-0.0822	-0.0228	-0.0361	-0.0288	-0.0465	-0.0165	-0.0337	-0.0042	-0.0055

Correlation Matrix

The next step was to divide the data into a training and testing set where 70% of the data was for training and 30% was for testing.

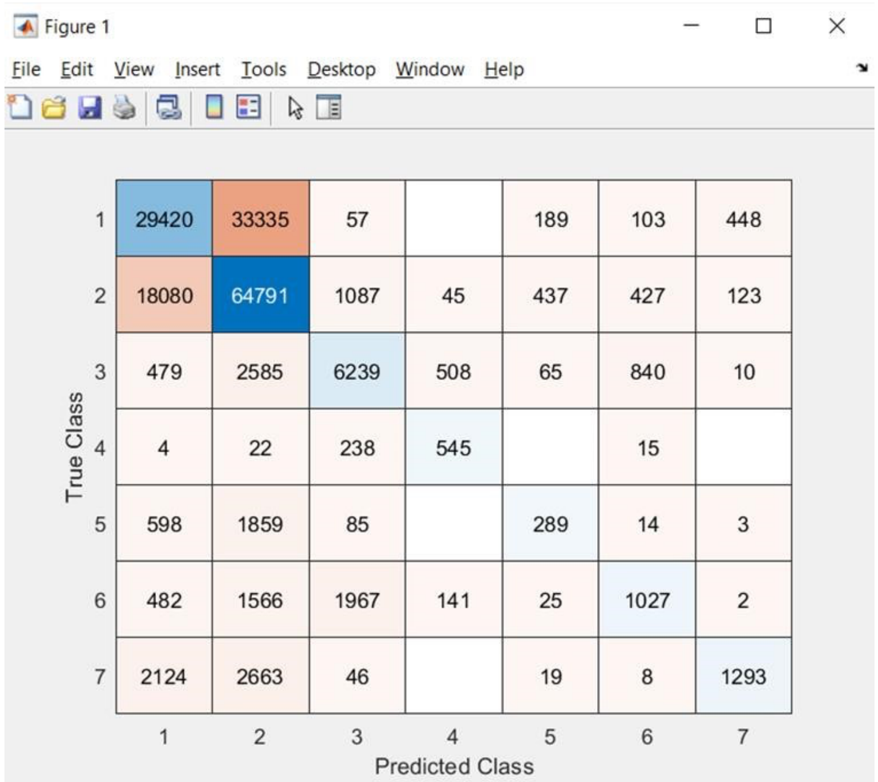
The data was trained using classification and clustering algorithms such as KNN, Decision Tree, Naïve Bayes , K Means and Gaussian Mixture Model that predicted the most species that were more prevalent in the specific wilderness area.

The models were then evaluated by calculating the different evaluation metrics such as accuracy, precision,recall using a confusion matrix which showed us the misclassified and well as correctly classified results.

The Machine learning algorithms along with their confusion matrices and the error rates are described below.

1. **KNN:**

   k-nearest neighbors algorithm is a non-parametric classification method used for classification and regression. In both cases, the input consists of the k closest training examples in a data set.

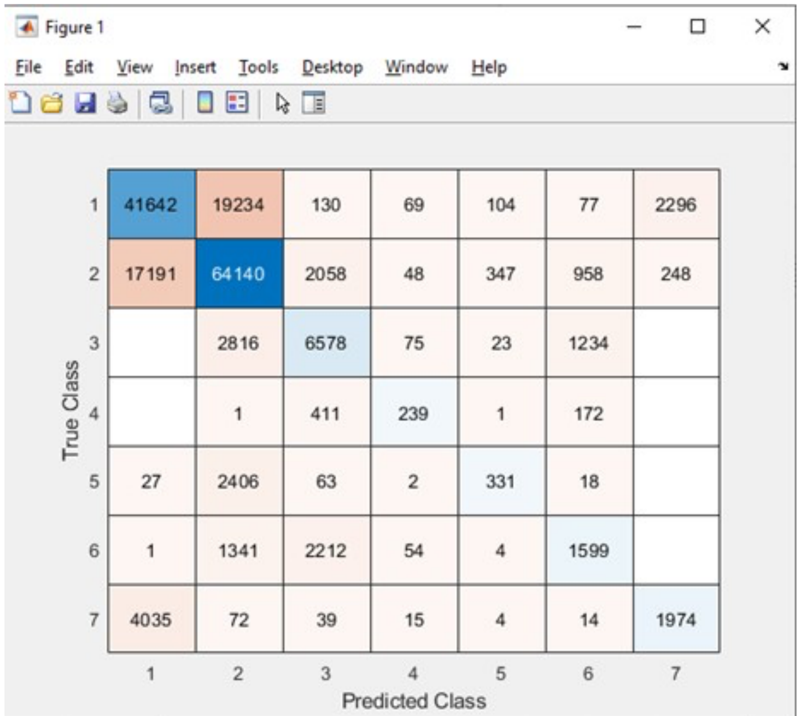


Confusion Matrix for KNN model

Above is the confusion matrix obtained after applying the KNN model on the dataset based on the forest covertype.
Misclassified Cover type : 18080
Misclassified Soil Type : 33335
The error rate = 0.4057

2. Decision Tree

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

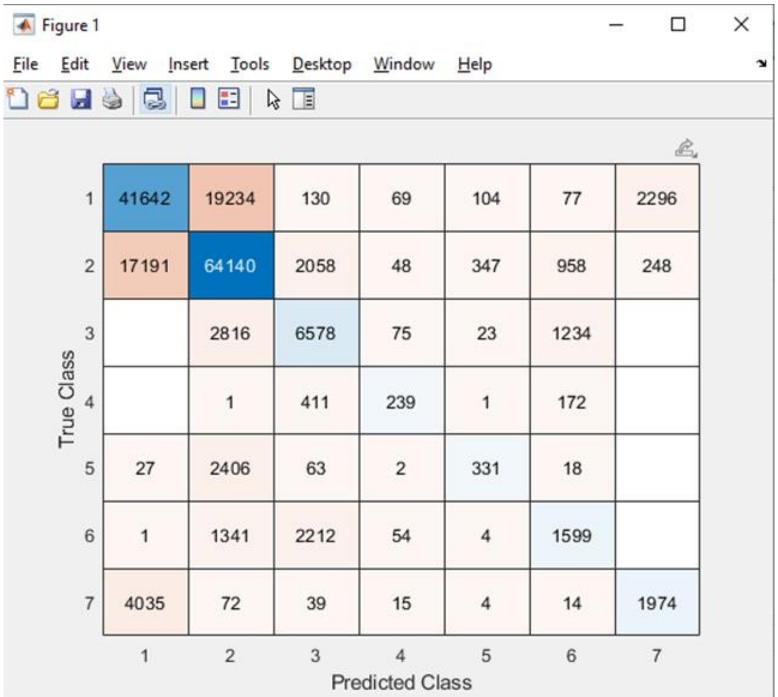


Confusion Matrix for Decision Tree

The confusion matrix that we got after applying the decision tree model shows the following results:
Misclassified Cover type : 17191
Misclassified Soil Type : 19234
The error rate = 0.3436

3. Naïve Bayes

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.



Confusion Matrix for Naive Bayes

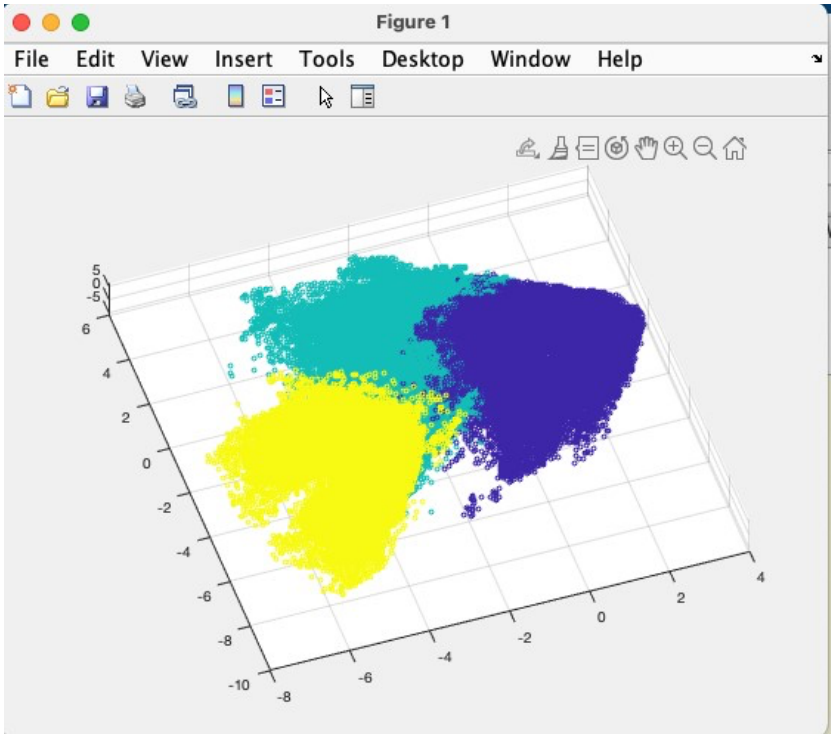
The confusion matrix that we got after applying the Naive Bayes model shows the following results:
Misclassified Cover type : 17191
Misclassified Soil Type : 19234
The error rate = 0.3316

4. Clustering techniques

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group.

K means clustering -

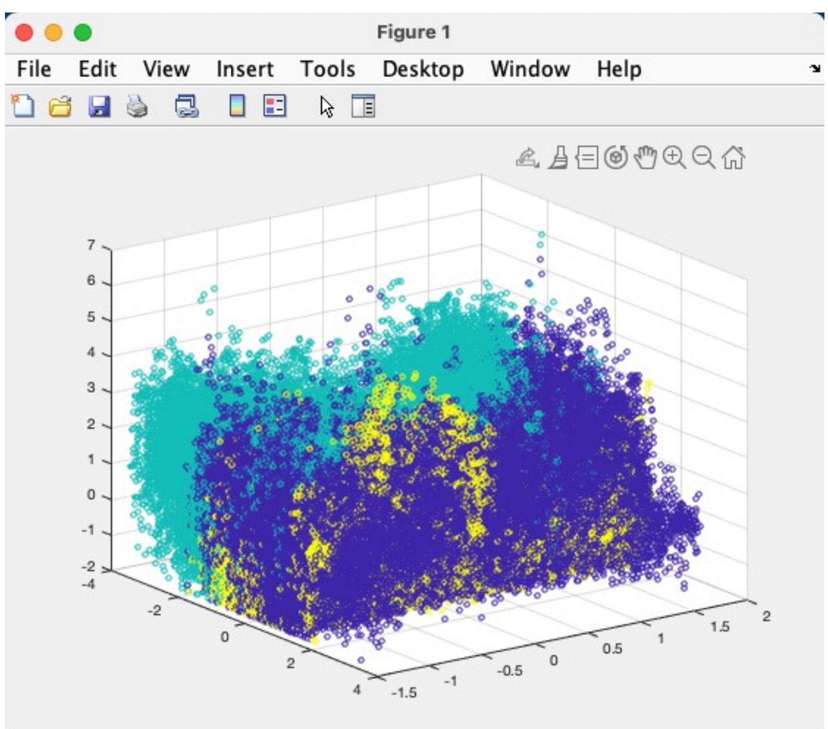
K means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.
K-Means requires much less time to discover and group the workloads into required number of clusters than required by GMM for corresponding number of Gaussian components



Scatter plot of K means Clustering

The data values are very close hence data points may belong to more than one cluster centroid.

Gaussian Mixture Model -
A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.
GMM will easily converge for local minimums. GMM have more flexible, elliptical decision boundaries. But, GMM is much slower as compared to clustering with K-Means.



Scatter plot of Gaussian Mixture Model

1.3.4 Final outcomes:

- Usage of clustering algorithms to group fertile lands with the same soil qualities to grow particular types of saplings and trees.
- Classification models like logistic regression, Random Forest are used to predict which type of trees work well with different combinations of soil, water, topographical location etc.

1.3.2 The Applied Technology:

Trees play a crucial role in the fight against climate change. In the United States, national forests alone absorb more than 50 million metric tons of carbon each year.[2] Approximately 2.6 billion tons of carbon dioxide, onethird of the CO2 released from burning fossil fuels, is absorbed by forests every year.[3] As per the researchers from the Swiss university ETH Zürich, forest restoration isn't just one of the climate change solutions, it is overwhelmingly the top one. International Union for Conservation of Nature – IUCN states that, nearly 2 billion hectares of degraded land across the world – have the potential for reforestation and this would capture nearly two-thirds of man-made carbon emissions.[3]

Below are a few successful forest reforestation programmes that this project draws inspiration from, 1. Bonn Challenge[4], launched by the government of Germany and IUCN is a global effort to bring 350 million hectares of deforested and degraded land under restoration by 2030. This could sequester up to 1.7 gigatonnes of carbon dioxide equivalent annually.

2. Teamtrees organization, under the reforestation program, aims to plant more than 20 million trees in the natural forest areas.

1.3.2 Impact of the Project :

A research article published on Global Forest Watch, states that the world's forests sequestered about twice as much carbon dioxide as they emitted between 2001 and 2019. In other words, forests provide a "carbon sink" that absorbs a net 7.6 billion metric tonnes of CO2 per year, 1.5 times more carbon than the United States emits annually.[4]

Carbon sequestration from the atmosphere in tropical rainforests is significantly more in comparison to the temperate or boreal forests and are the most important part of ecosystems for mitigating climate change. The world's three largest tropical rainforests are in the Amazon, Congo River basin and Southeast Asia.

Based on the above information, this project aims to improve and optimize the carbon sequestration process in forests by classifying the trees based on their potential to absorb carbon from the atmosphere. A particular type of tree grows better in one geographical location with respect to another. There are various factors that affect this such as soil type, climatic conditions, wilderness area, distance from the water reserve, growth time.

The goal of this project is to combine these various factors and datasets to find the best-fit optimal solution to find out which trees best suit a particular location. It also provides customized techniques for 2 different scenarios

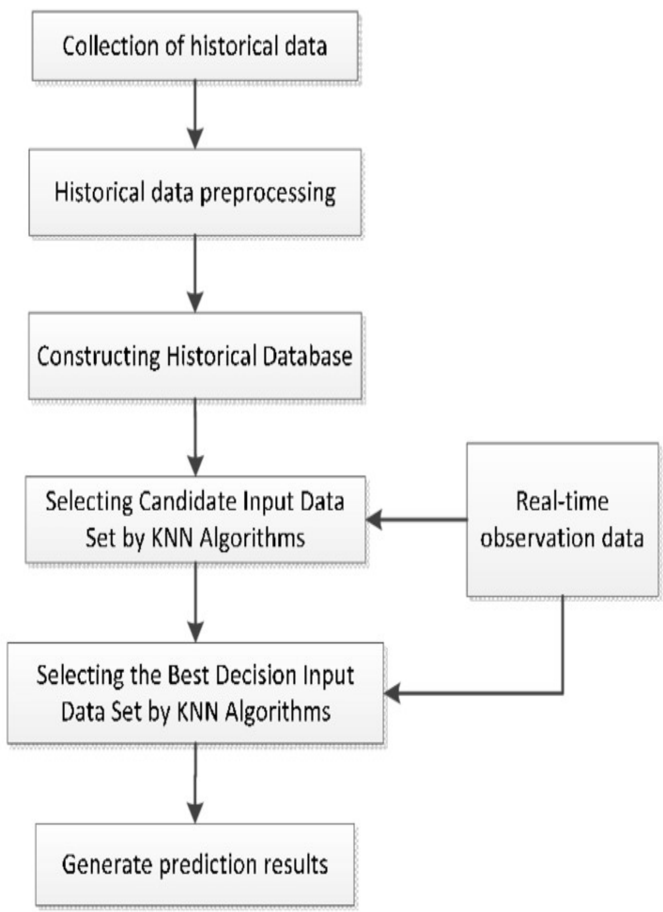
- Forest restoration(reforestation) – Replanting trees in the area where forest cover existed before and was lost due to deforestation or natural causes like wildfires.
- Afforestation - planting new trees where there were none before.

It provides a projection of future conditions given the ongoing and expected impacts of climate change and comparison of how increase in forests overall help combat the adverse effects of climate change. Carbon removal does not always demand full grown trees to contribute to the process, it starts early from the sapling stage. A silver maple sapling, for example, would sequester 400 pounds of carbon dioxide over 25 years. Therefore, plants, trees and full-grown forests all contribute to the goal of Carbon removal and the fight against climate change.

1.4 Project Drawings & Supporting Files

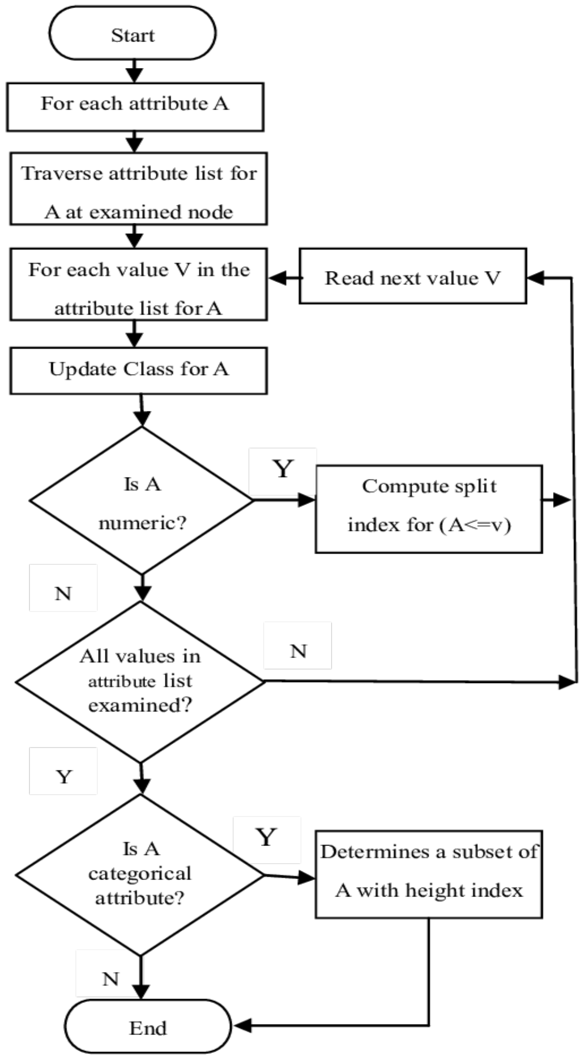
A schematic diagram of the applied models used can be seen in the following figure:

Flowchart for KNN:

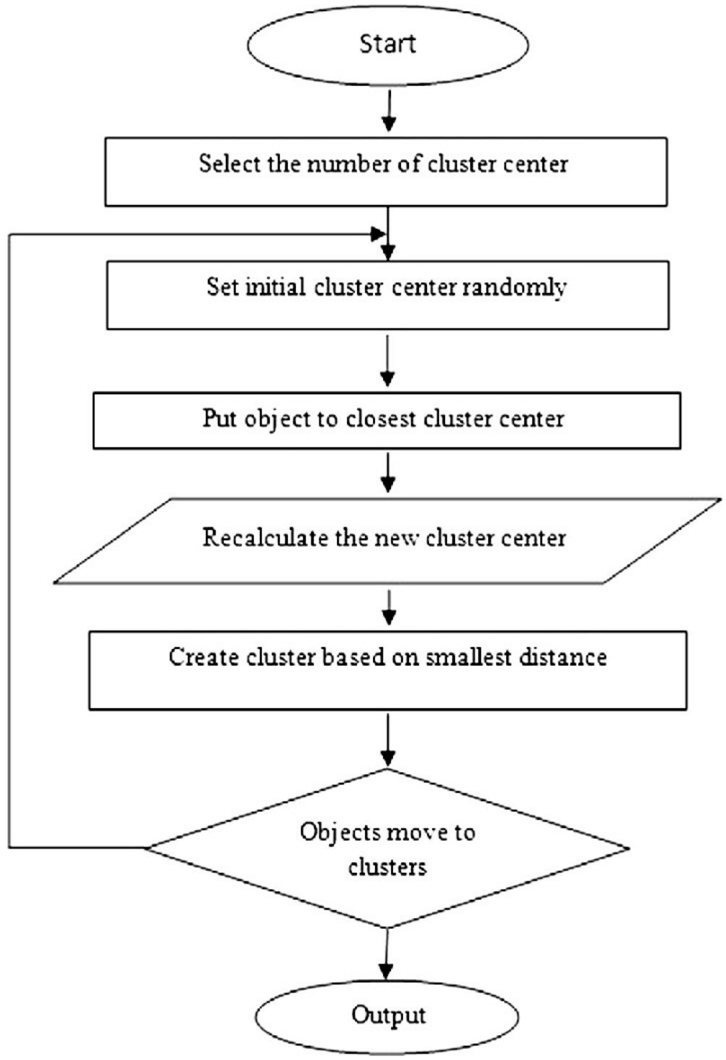


Flowchart for Decision Tree:

It is the most popular and the most efficient algorithm in decision tree-based approach. A decision tree algorithm creates a tree model by using values of only one attribute at a time.



Flowchart for Clustering Algorithm:



Flow chart for Naïve Bayesian classification:

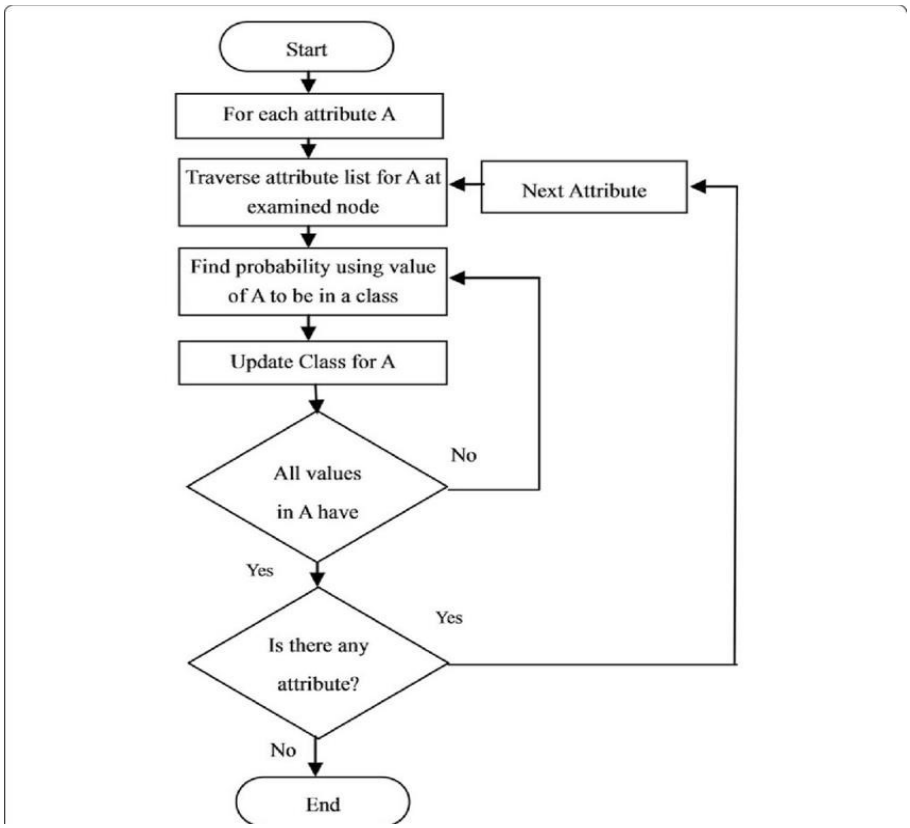
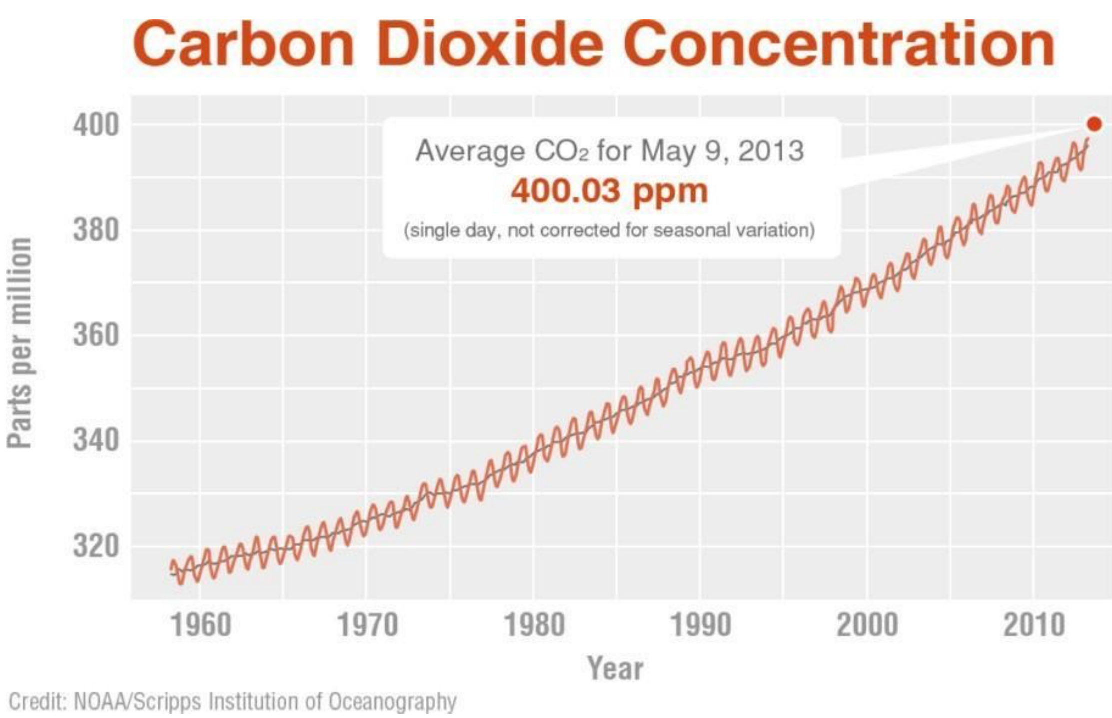


Diagram for monthly data for CO_2 increase over the years :



We can infer a few things from the chart above :

1. First, we notice that over the last six decades, carbon dioxide concentration has increased steadily, which is why the lines stack up nicely in a non-stacked chart type like this.

2. And the chart doesn't just show that CO2 concentration is going up, it even reveals that the pace of the increase is accelerating, as the vertical space between the decade lines is getting wider and wider. 3. We're also able to see the periodic pattern that follows the seasons. Carbon dioxide concentration hits a low around September, climbs up to a peak the next May, from where it then drops to the next low in September etc.

4. Finally, this form allows comparisons between the peaks and lows of different years. For instance, we can see that the "September low" of 2017 was already higher than the "May peak" of 2014**.**

1.4.2 Graphs and charts supporting the project:

The diagram below shows statistics of the CO2 reduction and releases over the 5-year interval.
The tree plantation also increases the climate cooling. Carbon dioxide reduction also reduces global warming.

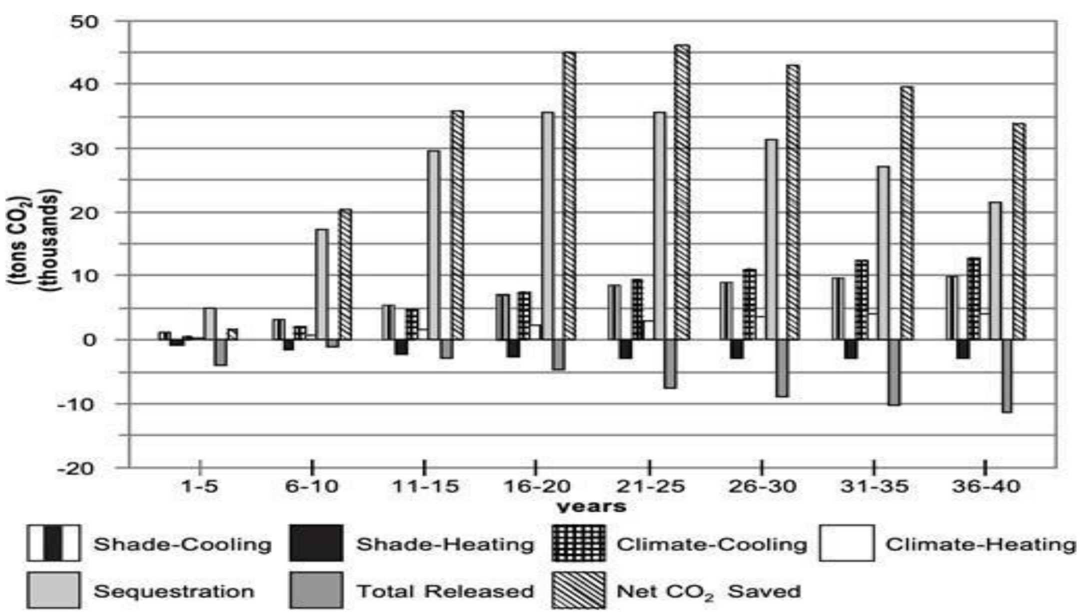
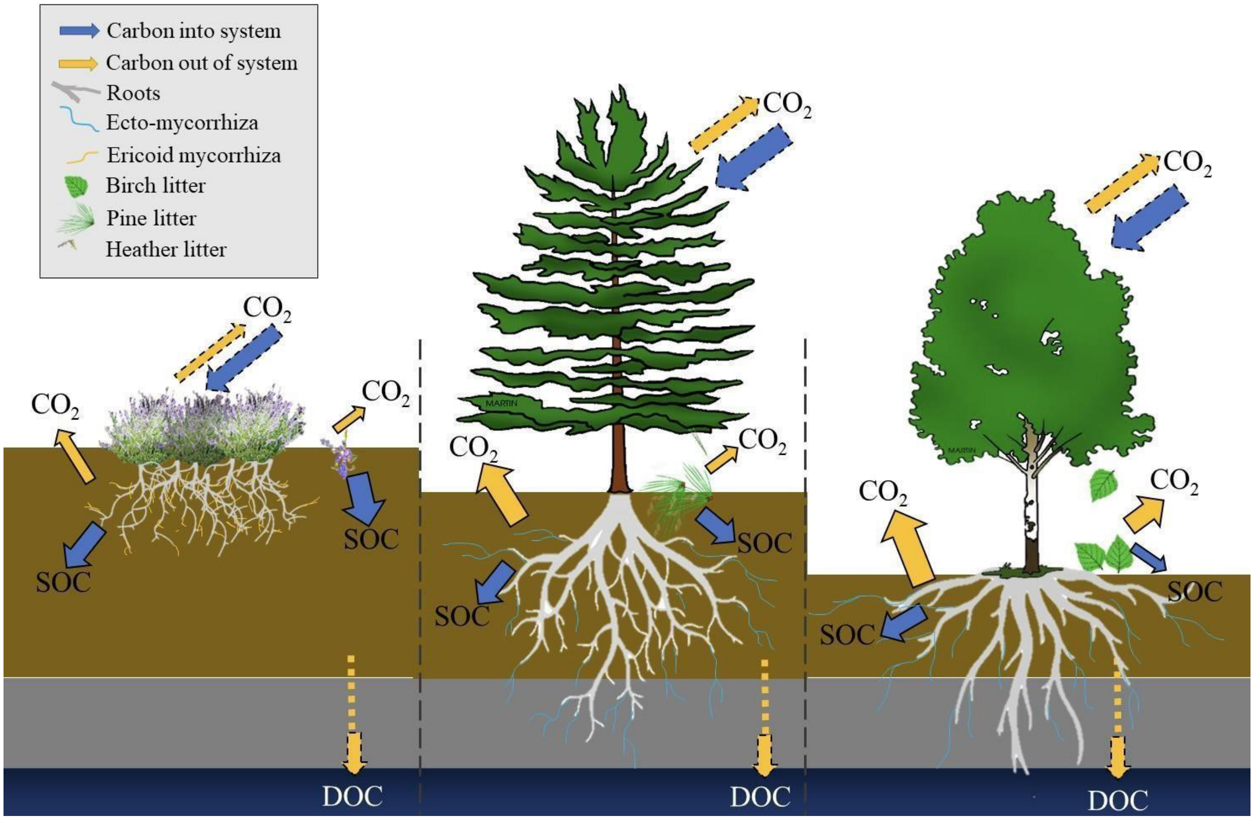


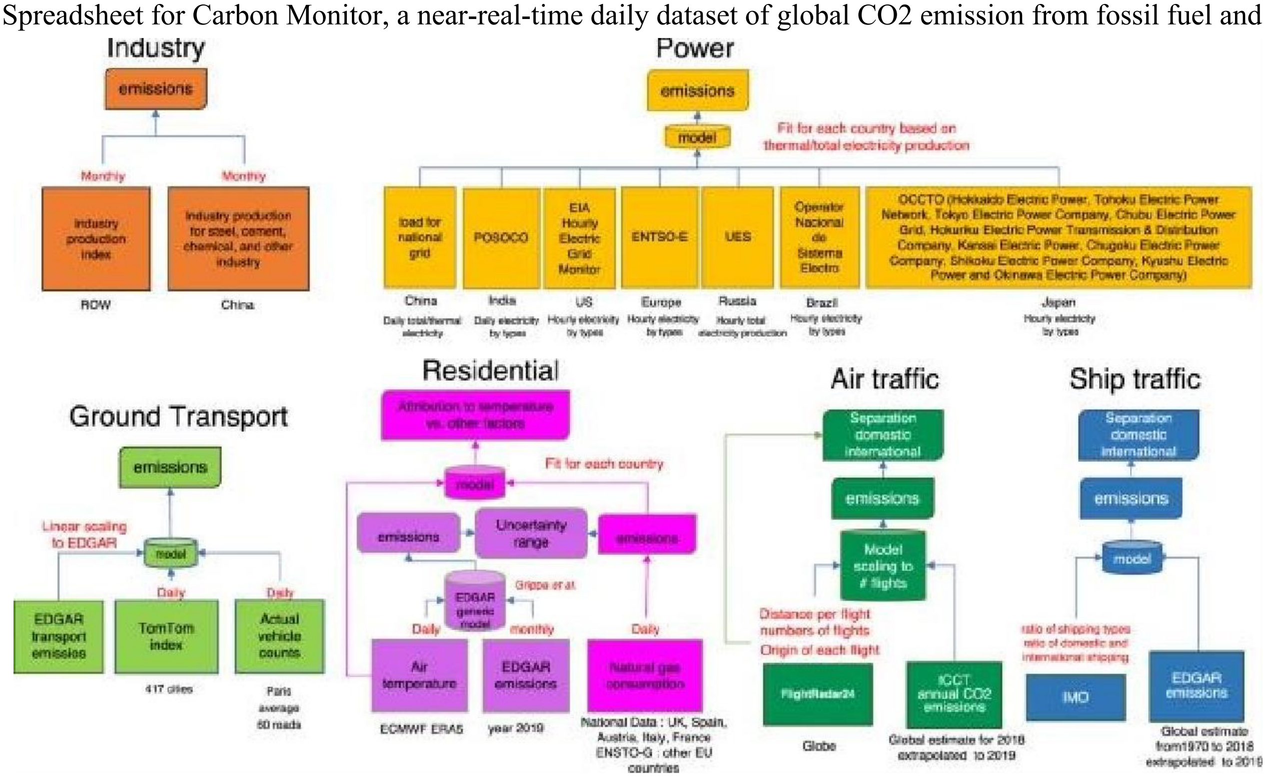
Figure 4- Tree plantation years Vs CO2 reduction [1]

"Although increased fossil emissions may not be fully responsible for the recent growth in methane levels, reducing fossil methane emissions are an important step toward mitigating climate change," said GML research chemist Ed Dlugokencky.

Engineering Drawing for Carbon Sequestration in tree biomass.

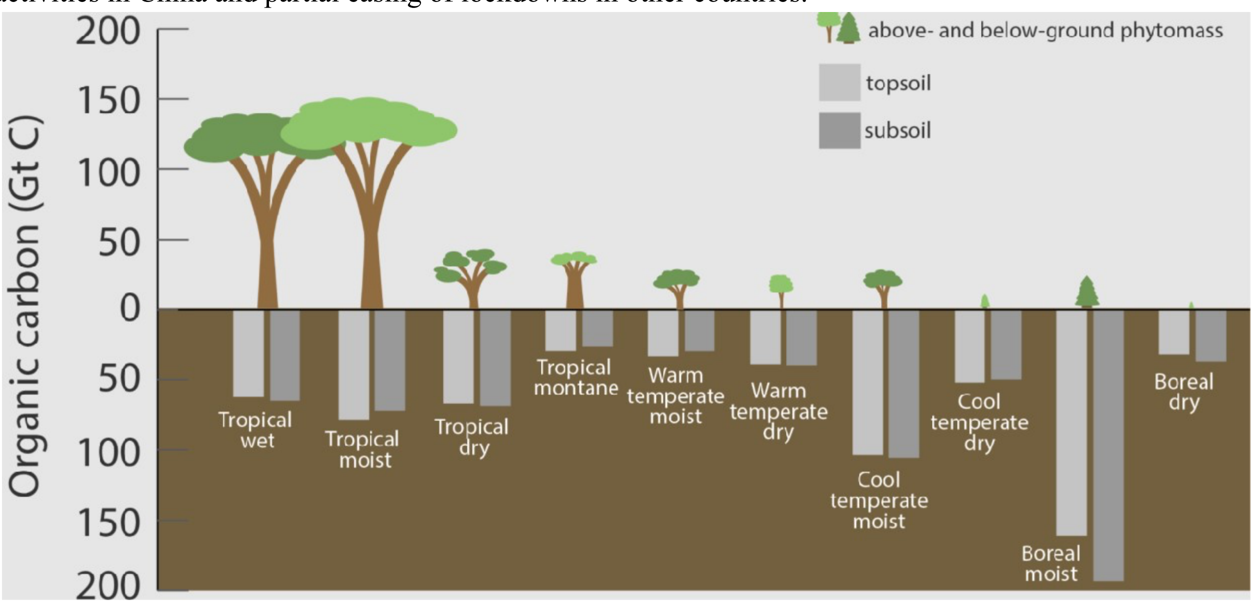


Spreadsheet for Carbon Monitor, a near-real-time daily dataset of global CO2 emission from fossil fuel and



production:

While daily emissions can be directly calculated using near-real-time activity data and emission factors for the electricity power sector, such an approach is difficult to apply to all sectors. For the industrial sector, emissions can be estimated monthly in some countries. For the other sectors, we used proxy data instead of daily real activity data to dynamically downscale the annual or monthly CO2 emissions totals to a daily basis. For instance, traffic indices in cities representative of each country were used instead of actual vehicle counts and categories, combined with annual national total sectoral emissions, to produce daily road transportation emissions.
As such, for the use of fuels in the road transportation, air transportation and residential sectors in most countries, we downscaled monthly or annual total emissions data in 2019 to calculate the daily CO2 emissions in that year. Subsequently, we scaled monthly totals of 2019 by daily proxies of activities to obtain daily CO2 emissions data for the first four months of 2020 during the unprecedented disruption of the COVID-19 pandemic. The Carbon Monitor near-real-time CO2 emission dataset shows a 8.8% decline in CO2 emissions globally from January 1st to June 30th in 2020 when compared with the same period in 2019 (Figure), and detects a regrowth of CO2 emissions by late April, which is mainly attributed to the recovery of economic activities in China and partial easing of lockdowns in other countries.



The organic carbon in all land-living species, both alive and dead, as well as carbon deposited in soils, is included in the terrestrial biosphere. Plants and other living organisms store about 500 gigatons of carbon above ground, while soil stores about 1,500 gigatons of carbon. Organic carbon makes up most of the carbon in the terrestrial biosphere, whereas inorganic carbon, such as calcium carbonate, makes up about a third of soil carbon. Organic carbon is found in abundance in all living things on the planet. Autotrophs collect carbon dioxide from the air and convert it to organic carbon, whereas heterotrophs obtain carbon through eating other species.

Carbon exits the terrestrial biosphere in a variety of ways and over a variety of timescales. Organic carbon is swiftly released into the atmosphere when it is burned or respired. It can also be transported into the ocean via rivers or stored as inert carbon in soils. [26] Soil carbon can be retained for thousands of years before being washed into rivers by erosion or released into the atmosphere by soil respiration.

The above figure states that carbon decomposition with plant and soil types

1.5 Literature to Support the approach:

Summarizing the findings of the research paper "A Review of the Role of Vegetal Ecosystems in CO_2 Capture" presented by Giuseppe Di Vita 1, Manuela Pilato 2,*Biagio Pecorino 1 ID , Filippo Brun 3 ID and Mario D'Amico 1 which illustrates that since the vegetal species have a high capacity to absorb exceeding carbon emission, they are more capable of CO_2 capture and storage.

The paper describes how their paper is all about identifying and examining the various forms of CO_2 sequestration made by plants and crops that are responsible in reducing the greenhouse gas (GHG) emission. The paper also discusses a recent study based on a control model that was used to identify the optimal level of afforestation that would be required for this purpose. The study states that plantation of the trees that grow fast on a less productive agricultural land is important for carbon absorption. But, a proactive approach is required which would involve public incentives for the strategy to work and so that the plantation can be done on a large scale.

The economic analysis on this approach shows that though the cost to implement such projects are lower than the other methods of CO_2 capture, the owners of the forest land that needs to be converted to agricultural lands for the purpose of reforestation may feel the cost as an obstacle.

The study also suggests that it is important to achieve a correct balance emission of carbon sequestration as it is necessary to evaluate the carbon that is stored from the forest and being considerate towards the issue of nonharvesting and including the forest biomass aimed at obtaining bioenergy.

A study on tape grass (Phalaris arundinacea L.) was conducted that helped in the evaluation of the absorption of carbon dioxide variations (CO_2 per the type of cultivation) that takes into consideration the physical-chemical characteristics of the soil (organic matrix, etc.) and its ecosystem. A special technique known as the eddy covariance; the total absorption of CO_2 was measured over four years in a tape grass cultivation on some peat bogs of Eastern Finland. This study has permitted the recognition that, during wet seasons, with moderate temperatures, with high humidity of the soil surface, and in the presence of low evaporation phenomena, a greater absorption of CO_2 is stimulated.

The findings from the same study also illustrate that the absorption of CO_2 appears positively correlated with soil humidity, air temperature, and inversely related with any deficiency in the vapor pressure. Thus, it summarizes that the total ecosystem respiration grows with an increase in soil temperature but declines with an increase in soil moisture. The study described in this paper highlights the current research topics involving CO_2 capture and other methods involved in Carbon removal and storage in farming, food, and energy sectors. This paper illustrates 56 works of potential relevance in the field of reforestation and CO_2 capturing techniques.

The focus of the paper is to identify and examine the different methods involving carbon sequestration and also the challenges that are faced during the process involved in reducing greenhouse gas emission. Through the review of "Literature Survey of Carbon Capture Technology" we came across the technology of Carbon capture and Storage.

Carbon capture and storage (CCS) involves the separation and capture of CO_2 from flue gas, or syngas in the case of IGCC. CS is a three-step procedure that includes the following steps:

1. CO_2 capture from electric power plants (or other industrial processes)

2. CO_2 is captured, compressed, and transported (usually in pipelines
3. Subsurface CO_2 injection and geologic sequestration (also known as storage) in deep underground rock formations.

These formations, which are often a mile or three below the surface, are made up of porous rock that traps CO2. Impermeable, non-porous layers of rock cover these formations, trapping CO2 and preventing it from flowing higher.

The study conducted at the UC Davis showcased the importance and the various types of Carbon sequestration. Carbon sequestration is the process of storing carbon dioxide in order to prevent it from entering the atmosphere. The goal is to keep carbon in solid and dissolved forms stable so that it does not warm the atmosphere. The method has a lot of promise for lowering people's "carbon footprint." Carbon sequestration can be divided into two categories: biological and geological. Carbon sequestration is a technique that has recently piqued the curiosity of many scientists. This method is directly linked to the Kyoto Protocol's GHG emission reduction order, which was established in 2004. Carbon sequestration is defined as the removal of carbon compounds from the environment, hence lowering carbon emissions.

1.6 Project Plan including Timeline, Budget, and Key Milestones

Topic	**Description**	**Date**
Milestone1	Data collection	Done
Milestone 2	Theoretical information gathering	Done
Milestone 3	Developing Algorithm	In progress (11/03/2021)
Milestone 4	Testing the algorithms on the dataset.	11/20/2021
Milestone 5	Final project submission	12/14/2021

Topic	Description	Date
Milestone 1	Data collection	1/11/21
Milestone 2	Theoretical information gathering	29/11/21
Milestone 3	Developing Algorithm	1/2/22

Milestone 4	Demo on practical working of the project	1/12/22
Milestone 5	Go Green campaign	1/1/23
Milestone 6	Tree Restoration	6/6/22
Milestone 7	Agricultural Soil management	1/9/22
Milestone 8	Expand Forest	1/1/23

Forests are one of the Earth's largest carbon sinks, decreasing the amount of carbon dioxide in the atmosphere. Despite their importance, however, systematic destruction by fires, deforestation, and other human activity has led to the loss of nearly half the world's trees, most within the last 100 years. So, we will be working on this idea where we can reduce the carbon through tree plantation. Once the project is approved, we will implement this idea of carbon reduction. Starting from Oct 30, 2021, we will start the data Collection Required for the Co2 reduction techniques, Plantation Data, Land Data available for tree plantation or agricultural data, and the areas which are affected more due to carbon emission. Once we have all the data then, we will start implementing the theoretical Information regarding how to implement the algorithm with the current challenges. Once the algorithm is co-related with data, we will define a process which will help in reducing carbon emission. Then we will create a demo project which will show us the practical data and we can compare it with the theoretical data. With respect to the process and the demo project we will decide the budget available. Once the final project is approved, we will start the campaign for GO GREEN. For this campaign we will need manpower, advertisement material, and convince more and more farmers to be a part of this Campaign. Soil carbon can increase soil health and crop yields and We will be Planting cover crops from September 2022. We will start tree restoration from June 2022 and in addition to that, We will try to Expand Forest in each year, starting from January 2023.

1.7 Conclusion:

This project gave us hands-on experience to use and apply most of the algorithms and models taught as a part of the coursework and analyse the results of each model along with providing a better understanding on comparison metrics.

The results after applying the machine learning algorithms on the dataset show that Spruce/Fir (labeled as 1), Lodgepole Pine(labeled as 2) and Krummholz(labeled as 7) are mostly spread out in Rawah, Neota and Comanche Peak Wilderness area ,Ponderosa Pine(labeled as 3) and cottonwood/willow(labeled as 4) are mostly found in the Cache la Poudre Wilderness Area, Aspen (labeled as 5) is mostly found in Rawah or Comanche whereas Douglas-Fir can be found to be spread across all the wilderness area as it is not very soil-specific species.

After applying the machine learning algorithms on the dataset we can conclude that the Naive Bayes algorithm was the most accurate with only 0.33% error rate and the lowest number of misclassified cases when compared to KNN algorithm. Though the Decision tree algorithms also gave the same results as Naive Bayes in terms of the confusion matrix, the error rate was higher for the same.

References:

1. https://www.wri.org/initiatives/carbon-removal

2. https://www.arborday.org/programs/replanting/

3. https://www.iucn.org/resources/issues-briefs/forests-and-climate-change

4. https://bonnchallenge.org/

5. https://www.mprnews.org/story/2019/09/27/climate-curious-what-are-the-best-carboncapturing-trees-to-plant

6. https://www.mprnews.org/story/2018/11/14/natural-climate-change-solutions-study

7. https://en.wikipedia.org/wiki/Carbon_cycle

8. https://www.wri.org/insights/forests-absorb-twice-much-carbon-they-emit-each-year

9. https://www.technologyreview.com/2021/07/08/1027908/carbon-removal-hype-is-a-dangerous-distraction- climate-change/

10. https://towardsdatascience.com/predicting-forest-cover-types-with-the-machine-learning-workflow-
11. 1f6f049bf4df
12. https://www.displayr.com/what-is-a-correlationmatrix/#:~:text=A%20correlation%20matrix%20is%20a,a%20diagnostic%20for%20advanced%20ana lyses.