**Role of Agriculture Scientist:**

There are many Role of agricultural scientist for soil analysis in project, are given below.

1. Contact to farmer and take soil sample from his field.
2. Test each sample in lab and evaluate the value of each attribute.
3. Make a sheet of datasets of each sample known as test data.
4. Study of particular soil and make a train data by stablishing a relationship between Y (nutrient power %) and X (each attribute of soil) as dependent and independent attribute of Train dataset.
5. With the help of train and test, he/she perform analysis and give final result of average nutrient power of soil samples.

**Attribute of soil samples:**

**Data Dictionary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **datatype** | **nullable** | **Description** |
| **Treatment** | Text | Not null | Each soil sample take as treatment name like T1, T2 and T3 ….. |
| **Replicate** | Text | Not null | Each treatment Ti have 6 replica as R1,R2,R3,R4,R5,R6 |
| **Depth** | Text | Not null | Depth determine depth of soil sample taking from field in cm. |
| **pH** | Value | Not null | It determine the acidic and basic nature of soil from 1 to 14 (<7 acidic , =7 neutral ,>7 basic ) . |
| **CEC** | Value | Null | Cation exchange capacity (CEC) is the total capacity of a soil to hold exchangeable cations.  CEC is an inherent soil characteristic and is difficult to alter significantly. It influences the soil's ability to hold onto essential nutrients and provides a buffer against soil acidification. |
| **Phosphorous** | Value | Null | **Phosphorus** is a component of the complex nucleic acid structure of plants, which regulates protein synthesis.  **Phosphorus** is, therefore, **important** in cell division and development of new tissue. **Phosphorus** is also associated with complex energy transformations in the plant. |
| **Potassium** | Value | Null | **Potassium** is an essential plant nutrient that plays a role in a wide range of physiological processes, from regulation of the stomata to enzyme activation. **Potassium** is held in the **soil** by the cation exchange capacity. |
| **Calcium** | Value | Null | **Calcium** plays a very important role in plant growth and nutrition, as well as in cell wall deposition.  The primary roles of **calcium**:  As a **soil** amendment, **calcium** helps to maintain chemical balance in the **soil**, reduces **soil** salinity, and improves water penetration. |
| **Magnesium** | Value | Null | Magnesium is an essential plant nutrient. It has a wide range of key roles in many plant functions.  One of the magnesium's well-known roles is in the photosynthesis process, as it is a building block of the Chlorophyll, which makes leaves appear green. |
| **Lime index** | Value | Null | Technically, reserve acidity is the acidity which is adsorbed on the surfaces of **soil** and organic matter particles.  As **lime index** decreases below 70, more **lime** is needed to bring **soil** pH up to the target level. **Soil** with a greater reserve acidity (lower **lime index** number) has more capacity to resist change in pH. |

**Way of analysis:**

1. **Tools and technology:**

**Technology:** machine learning, mathematical statistic, data interpretation

**Software:** R Studio (version 1.2.1335), M.S.Excel, M.S.Word

**Language:** R programming

**Operating System:** Windows 10

1. **Preparation of train data:**

We can study North coast soil to prepare train dataset because test data is also same as in which we can see that nutrient element are ranges in these format, are given below.

**Study:**

**Ph:** upto -7

**CEC:** upto-15

**Phosphorous:** upto-100

**Potassium:** upto-365

**Calcium:** upto-1400

**Magnesium:** upto-500

**Nitrogen:** upto-90

**Lime index:** upto-73

**CEC:** is part of soil test which is calculated from level of K, Mg, Ca, Na and H.

**CEC= (K/390+Mg/120+Ca/200+Na/230+H/1)**

On the basis of these study, we have prepared train data to build model to find unknown nutrient power of soil.

**Example of Bank policy to provide loan:**

A bank policy to provide loan is given below on the basis of percentage criteria:

1. Less than 23% of nutrient power -> 5000 rupees loan per 100 sq. feet
2. 24% to 52% of nutrient power ->­ 9200 rupees loan per 100 sq. feet
3. 53 to 72 % of nutrient power ­-> 14400 rupees loan per 100 sq. feet
4. Above than 72% of nutrient power -> 16000 rupees loan per 100 sq. feet

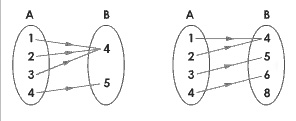
# A farmer field nutrient power is 71% then bank will provide loan 14400 rup. Loan per 100 sq. feet.

1. **Steps of analysis:**

**Degree of relationship:**

As per relationship among the attribute of train dataset (train\_soil.csv) Y (average nutrient power %) is dependent variable and xi (pH, CEC, phosphorous, potassium, calcium, magnesium, nitrogen and lime index) are independent variable.

On the basis of these variable, it is seen that there is many to one (1: M) relationship.



**Mathematical view:**

Y is dependent var. and Xi are independent var , so mathematics function is

Y=b+(b1\*X1)+(b2\*X2)+(b3\*X3)+(b4\*X4)+(b5\*X5)+…….

Where b is intercept and

b1, b2, b3….are regression coefficient.

**Machine learning and analysis:**

As per mathematical view of train data, we decide to perform one of the supervised learning method named as Multiple linear regression to make prediction for test data that have same concept of task.

First, we have to build a model on the basis of concept of multiple linear regression and decide to test data to test the nutrient power of soil samples.

Average nutrient power of soil samples is known as expected nutrient power of soil samples.

**Source code of R programming for data analysis:**

getwd()

setwd("F:/project")

dataset <- read.csv("train\_soil.csv")

head(dataset)

df1 <- as.data.frame(dataset)

str(dataset)

y=df1$Average.nutrient.power..

x1=df1$pH

x2=df1$CEC

x3=df1$phosphorous

x4=df1$potassium

x5=df1$calcium

x6=df1$magnessium

x7=df1$nitrogen

x8=df1$lime\_index

model1 <- lm(y~x1+x2+x3+x4+x5+x6+x7+x8,data = df1)

plot(model1)

summary(model1)

#plot(model1)

test\_data <- read.csv("soil.csv",as.is = 1,stringsAsFactors = FALSE)

summary(test\_data)

dim(test\_data)

head(test\_data)

str(test\_data)

names(dataset)

#x <-as.vector(as.numeric(test\_data[3:6,1]))

#x

x1 = as.numeric(test\_data[2:975,4])

x1

x2=as.numeric(test\_data[2:975,5])

x3= as.integer(test\_data[2:975,6])

x4= as.integer(test\_data[2:975,7])

x5= as.integer(test\_data[2:975,8])

x6= as.integer(test\_data[2:975,9])

x7= as.integer(test\_data[2:975,10])

x8= as.integer(test\_data[2:975,11])

class(x1)

df <- data.frame(x1,x2,x3,x4,x5,x6,x7,x8)

print(df)

str(df)

pred = predict(model1,newdata = df)

nut <- sum(pred)/nrow(test\_data)

print(nut)

**Output:**

getwd()

[1] "F:/project"

> setwd("F:/project")

> dataset <- read.csv("train\_soil.csv")

> head(dataset)

Average.nutrient.power.. pH CEC phosphorous potassium

1 18.20224 2.0 0.1397436 11 9

2 20.56548 2.5 0.4053846 13 19

3 27.45670 2.8 0.8970000 25 35

4 23.14233 3.0 0.6684615 17 28

5 25.73288 3.5 0.9315385 17 37

6 27.39168 3.6 1.1946154 18 46

calcium magnessium nitrogen lime\_index

1 10 8 4 70

2 38 20 6 69

3 51 25 7 45

4 66 32 8 68

5 94 44 10 70

6 122 56 12 70

> df1 <- as.data.frame(dataset)

> str(dataset)

'data.frame': 49 obs. of 9 variables:

$ Average.nutrient.power..: num 18.2 20.6 27.5 23.1 25.7 ...

$ pH : num 2 2.5 2.8 3 3.5 3.6 3.8 3.9 4.1 4.3 ...

$ CEC : num 0.14 0.405 0.897 0.668 0.932 ...

$ phosphorous : int 11 13 25 17 17 18 19 20 21 22 ...

$ potassium : int 9 19 35 28 37 46 54 54 63 72 ...

$ calcium : int 10 38 51 66 94 122 181 150 178 206 ...

$ magnessium : int 8 20 25 32 44 56 78 68 80 92 ...

$ nitrogen : int 4 6 7 8 10 12 67 14 16 18 ...

$ lime\_index : int 70 69 45 68 70 70 70 70 70 69 ...

> y=df1$Average.nutrient.power..

> x1=df1$pH

> x2=df1$CEC

> x3=df1$phosphorous

> x4=df1$potassium

> x5=df1$calcium

> x6=df1$magnessium

> x7=df1$nitrogen

> x8=df1$lime\_index

> model1 <- lm(y~x1+x2+x3+x4+x5+x6+x7+x8,data = df1)

> summary(model1)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8, data = df1)

Residuals:

Min 1Q Median 3Q Max

-18.6660 -2.1120 -0.0567 1.5936 12.5952

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -15.869602 11.005153 -1.442 0.157082

x1 4.702197 1.832032 2.567 0.014116 \*

x2 0.377562 0.469928 0.803 0.426467

x3 0.019969 0.164151 0.122 0.903785

x4 0.040057 0.064953 0.617 0.540920

x5 0.040247 0.008727 4.612 4.04e-05 \*\*\*

x6 -0.095735 0.024481 -3.911 0.000348 \*\*\*

x7 0.348126 0.074864 4.650 3.58e-05 \*\*\*

x8 0.305275 0.155452 1.964 0.056531 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.747 on 40 degrees of freedom

Multiple R-squared: 0.9482, Adjusted R-squared: 0.9378

F-statistic: 91.51 on 8 and 40 DF, p-value: < 2.2e-16

> test\_data <- read.csv("soil.csv",as.is = 1,stringsAsFactors = FALSE)

> summary(test\_data)

Treatment Replicate Depth pH

Length:975 R1:166 0-25cm:849 Min. :4.400

Class :character R2:166 0-5cm : 72 1st Qu.:6.000

Mode :character R3:166 5-25cm: 54 Median :6.300

R4:163 Mean :6.225

R5:157 3rd Qu.:6.500

R6:157 Max. :7.100

CEC phosphorus potassium calcium

Min. : 2.800 Min. : 10.00 Min. : 28.0 Min. : 295.0

1st Qu.: 6.000 1st Qu.: 25.00 1st Qu.: 85.0 1st Qu.: 762.0

Median : 6.680 Median : 33.00 Median :105.0 Median : 850.0

Mean : 6.733 Mean : 36.44 Mean :108.9 Mean : 835.3

3rd Qu.: 7.410 3rd Qu.: 44.00 3rd Qu.:126.0 3rd Qu.: 927.0

Max. :11.700 Max. :104.00 Max. :324.0 Max. :1263.0

magnesium Nitrogen lime\_index

Min. : 10.0 Min. : 6.00 Min. :63

1st Qu.:147.0 1st Qu.:22.00 1st Qu.:69

Median :176.0 Median :32.00 Median :70

Mean :168.6 Mean :34.78 Mean :69

3rd Qu.:198.0 3rd Qu.:45.50 3rd Qu.:70

Max. :286.0 Max. :76.00 Max. :71

> x2=as.numeric(test\_data[2:975,5])

> x3= as.integer(test\_data[2:975,6])

> x4= as.integer(test\_data[2:975,7])

> x5= as.integer(test\_data[2:975,8])

> x6= as.integer(test\_data[2:975,9])

> x7= as.integer(test\_data[2:975,10])

> x8= as.integer(test\_data[2:975,11])

> class(x1)

[1] "numeric"

> df <- data.frame(x1,x2,x3,x4,x5,x6,x7,x8)

|  |
| --- |
| pred = predict(model1,newdata = df)  > nut <- sum(pred)/nrow(test\_data)  > print(nut)  [1] 71.61433 |
|  |
| |  | | --- | | > | |

**Result of Analysis:**

After over all analysis of data with the help of test data (soil.csv), final average nutrient power percentage is calculated.

As per our test data of north coast soil, final average nutrient power is 71.61433 % of farmer field with 94.82 accuracy.