Classification of Agricultural Soil Contamination using Machine Learning

Giriraj Patil, Mahalaxmi, Chaitra, Kaveri, Ashwini Telang
Computer science and engineering, Guru Nanak Dev EngineeringCollege, Bidar

Abstract- **Soil contamination has severely increased over the last years, especially due to petroleum hydrocarbons, heavy metals and pesticides from industrial wastes and human activities. Even though in general soil quality research is facing an important technological challenge and several actions have been taken in order to assess, remediate and reduce the effects of contaminants on soils, suitable and standardized monitoring and remediation strategies of soil are required. In this sense, in the last decade there has been a growing emphasis on the utilization of residues and waste materials, coming from different industrial activities, in several remediation technologies (e.g., chemical degradation, photo-degradation) and bioremediation in order to clean up contaminated soils. The critical point regarding contaminated soil monitoring is the intrinsic difficulty in defining fixed monitoring variables and indicators as the establishment of any a priori criterion and threshold for soil quality can be still considered subjective.**

Index Terms— **Machine Learning, Degree of Contamination, Soil Degradation, Hydrocarbons etc.**

I. INTRODUCTION

Soil contamination refers to the destruction of land that could be used constructively by human activities either directly or indirectly.Soil contamination is a part of land degradation is caused by the presence of human made chemicals and alternation in natural soil environment. It is caused by industrial activity, agriculture chemicals, and improper disposal of waste, fertilizers, organic wastes. The chemicals may be hydrocarbons, solvents, pesticides, lead and other heavy metals.This project involves determining the degree of contamination in soil (DOC) and also provides information how it is affected and tells whether suitable for agriculture or not.

II. PROBLEM STATEMENT

It is first important to be able to identify whether site is contaminated before determiningsolution. It may arise due to both natural resources in soil or made. Hence these directly or indirectly affect plants, animals, & humans. The soil samples tested at particular site done in laboratory provides the composition of substances present in soil. Using this reports this software is used to investigate degree of contamination in soil.The goal is to find the degree of contamination (DOC), and to know whether it effects the soil and is that suitable for agriculture.

III. SYSTEM ANALYSIS

A) Parameters responsible for soil contamination: -

A.Mobility of heavy metals in soil:

The activity of elements like zinc-lead has very negative impact on environment, mainly because of accumulation of post mining and metallurgical waste which is long term leads to adverse transformation of natural environment due to migration of dust and metals to soil, surface water and ground water. The bioavailability is regulated by physical, chemical and biological processes and interactions between them. The method of binding heavy metals depends on several soil properties like organic matter content, pH value, sorption capacity, contents of macro and micro nutrients, activity of microorganisms, resistance of soil.Mechanical composition of soil is major factor determining the extent of soil contamination. Heavy soils due to large amount of suspended fraction have greater ability to retain metallic elements. Light soil does not have such sorption. Increasing the amount of organic matter in soil helps to minimize absorption of heavy metals by plants. Plants draw heavy metals from soil as the micro and macro nutrients from root system. Insufficient amount of nutrients results in

excessive accumulation of heavy metals. Causes caused by heavy metals is beginning of disturbance in metabolic activities, disturbance in collection, transport and accumulation of nutrients.

B. Industrial exposed inorganic toxic compounds:

Inorganic residues in industrial waste cause serious problems of disposal. They contain metals which have high potential of toxicity. They also emit large amount of industrial fluorides. Some of the fluorides are phosphoric acid, aluminium, ceramic etc. Sulphur dioxide emitted from factories make soil very acidic. These metals cause leaf injury and destroy vegetation. Copper, mercury, lead, nickel, arsenic are the elements which can accumulate in the soil Some of the fungicides due to copper and mercury also add to soil pollution. The toxicity can be minimized by building up soil organic matter, adding lime to soils and keeping the soil alkaline.

C.Organic wastes:

Domestic garbage, municipal sewage left in heaps or improperly disposed seriously effect health of human beings, plants and animals. Organic wastes contain borates, phosphates, detergents in large amount. If untreated they will affect the vegetative growth of plants. The main organic contaminants are phenols and coal. The radioactive metals like uranium, thorium cause a major part of soil contamination. Decontamination procedures may include continuous cropping and use of chelate amendments.

D.Sewage and sewage sludge:

Soil pollution is even caused by uncontrolled disposal of sewage and other liquid wastes resulting from domestic uses of water, industrial wastes containing a variety of pollutants, agricultural effluents from animal husbandry and drainage of irrigation water and urban runoff. Irrigation with sewage water causes profound changes in irrigation soil.The various changes that are brought about in soil are physical changes like leaching, changes in humus content, porosity etc., chemical changes like soil reaction, salinity. This led to phytoxicity of plants.

E.Organic pesticides:

Pesticides are quite frequently used to control several types of pests now a days. Pesticides exert harmful effects to microorganisms as a result of which plant growth may be affected. Pesticides which are not rapidly decomposed cause greater problem. Accumulation of pesticides is toxic for plants. Pesticides persistence in soil and movement into water streams may lead to their entry into foods and create health hazards.At present DDT and organochlorinecompounds are declared as harmful and banned in U.S.A. It is due to persistence of their residues in soil for longer time without losing their toxicity. This lead higher concentration in vegetation, flesh and milk.A major method of checking this is to increase the organic matter content of soil and choose such pesticides which are non-persistent and leave no harmful residue.

B) How to reduce soil contamination

A.Reforestation:

Most countries have policies that require its citizens to plant more trees where one has been cut. This is an effective method to curb soil erosion. Government should also take action against those who cut trees without a care in the world

B.Controlled farming practices:

Too much of anything is dangerous. The concept applies for farming practices in that they should be carried out in moderation. Practices such as overcropping and overgrazing should be avoided since they increase soil erosion.

C. Bioremediation:

This is the introduction to microorganisms into soil that break down contaminants. This is perfectly environmental approach since it allows nature to take its course thus restoring balance.

D.Reduce, Recycle and Reuse:

Items that can be used again should not be disposed of; things made up of peppergrasses should be recycled but excess use of plastic should be avoided then by all means reduce their use.

E.Use of biodegradable products:

Where possible, opt to use biodegradable products such as cartons for packing, if they would be disposed of, they would be easily broken down to become part of soil.

F.Reduce the use of pesticides and fertilizers:

Pesticides and fertilizers are major contributors to soil contamination so cutting own their usage could do good to soil.We have been on earth , if we completely destroy its surface, we will be the first to starve or poison ourselves. Redeeming land to its original state is impossible task. It will be involving altering its

properties which put our land into proper and responsible use

Effect of different elements on soil:

A.Effect of cobalt:

- Cobalt is the trace element which have a strong effect on growth and development of plants.
- Cobalt presence in plants allow the proper course of metabolic and growth processes to be maintained.
- To meet the nutritious requirements, cobalt should not be less than 0.08mg/kg.
- The highest dose of cobalt causes the pH, total exchangeable bases, cation exchange in the oxidation state of +2, capacity and degree of base saturation to decrease and the hydrolytic acidity of soil to increase.
- The accumulation rate for aquatic plants mainly depends on factors such as temperature, salinity and water oxygen concentration.
- The negative effect of soil contamination with cobalt on plants depend on other factors like soil reaction, content of organic matter, clay nutrients in soil.
- Cobalt toxicity varies depend on plant organ type. It is higher in leaves rather than other parts of vegetables.
- The toxic effect of cobalt contamination on plant development is smaller than in acid soil.
- The cobalt contamination also causes changes in soil properties.
- A decrease of soil pH can influence on increased cobalt uptake from soil by plants.
- The small amount of cobalt enters the soil from the airborne transport of particulate emissions and application of sewage sludge onto fields.
- A lack of cobalt in a form which plants or earth dwelling organisms are able to absorb can have major effects on the health of wildlife in an area.
- To meet the nutritious requirements cobalt content in the tissues should not be lower than 0.08mg/kg.
- Increased cobalt contents in soil can cause disturbances of physiological, biochemical and metabolic process.
- Photosynthesis is disturbed and intercellular spaces become smaller

B.Effect of nickel:

- Nickel is naturally occurring element in the earth crust, which is present in soil. It is the pollutant element.
- It is in the range of 50-100 mg/g is indicative of its toxicity in plants.
- The availability of nickel to plants is influenced by soil pH.
- The toxic symptoms produced by several heavy metals including nickel have usually chlorosis.
- The soil contamination with nickel have a negative impact on activity of soil enzymes.
- The combustion of fossil fuel results in considerable nickel pollution in some areas. It then reaches human body through contaminated food.
- Nickel may cause death of plants. Nickel content in terrestrial soils become more mobile when exposed to lower pH, water and this slight enrichment of nickel may be from terrestrial run off due to more acidic rainwater.
- Under acidic condition, nickel is more mobile in soil and might seep into ground water.
- It is the most toxic heavy metal.
- Nickel in soil appears most often on the second and third degree of oxidation.
- Nickel toxicity includes chlorosis, stunted root growth, and brown interval necrosis.

C.Effect of zinc:

- Zinc contaminated soils with negative impact on soil ecosystem are found around zinc smelters near zinc mining sites and under galvanized structures.
- Zinc levels as low as 100-200 mg/kg can have negative effects in plants but symptoms of zinc toxicity are not observed until zins levels in soil reaches 1000 mg/kg level.
- Plants often have zinc uptake that their systems can’t handle due to accumulation of zinc in soil.

- Zinc can interrupt the activity in soils as it negatively influences the activity of microorganisms. Hence retarding the breakdown of organic matter.
- On zinc rich soils only a limited number of plants have a chance of survival. That is why there is not much plant diversity near zinc disposing factories.
- Zinc is a serious threat to the productions of farmland.
- The function of zinc is to help plants produce chlorophyll.
- Leaves discolour when the soil is deficient of zinc and plant growth is stunted.
- Zinc deficiency causes a chlorosis, which causes the tissue between the veins to turn yellow while the veins remain green.

C. Effect of lead:

- Lead occur naturally in a soil at a concentration range from 10 to 50 mg/kg. Lead does not biodegradable or disappear over a time but remains in soil for thousand years.
- Soil lead is held tightly on the surfaces of fine clay and organic matter. Not all of the lead is available to plants. The availability depends on how tightly it is held by soil and solubility.
- Soils adjacent to heavy traffic in cities and busy roadways have highest concentration of lead along with leaded paint, chalking, leaching, scraping of leaded paint results in lead deposits in soil.
- The impact of exposure to lead contaminated soil onpb levels is also influenced by chemical and physical form of lead.
- Soils near highways, tree ways and smelting facilities have higher level of lead than soil in other areas.
- Soil is contaminated by lead by various sources, lead particles are deposited in soil from flaking lead paint from motor vehicles that use leaded gasoline.
- In general plants do not absorb or accumulate lead. It is known that lead does not accumulate in fruiting part of vegetables and fruiting crops (corn, tomato, strawberry) found in leafy vegetables and surface of root crops(carrot).
- The risk of lead poisoning of food chain increases as the soil lead level rises above 300ppm.

E. Effect of arsenic:

- Arsenic is found in soil in organic and inorganic forms.
- Arsenic is toxic component which pose high risk to soil and humans. 5mg/kg
- It ranges from 1 to 40 mg/kg, with a mean of 5mg/kgmuch higher level of arsenic occur in mining areas, waste sites from pesticides.
- Arsenic enters into farming systems through variety of means like natural geochemical process, use of as based pesticides, mining operations, irrigation with as contaminated ground water.
- The toxicity of arsenic depends on their oxidation states and chemical structures.
- Arsenic is non-essential to plants and generally toxic to plan. Roots are usually first tissue to be exposed to arsenic.
- Two types of arsenic AsV and AsIII disrupt plant metabolism, but through different mechanisms. These leads to imbalance in phosphate supply.
- Arsenic exposure generally induces production of reactive oxygen species that can lead to production of antioxidant metabolites and enzymes.
- Rice grain is found to be having more as content. Thus, elimination of as in rice is a major research goal.
- It can inhibit plant growth by slowing or arresting expansion and biomass accumulation, compromises plant reproductive capacity through loses in fertility, yield and fruit production.
- At high concentration with critical metabolic process can lead to death.
- Cellular membrane becomes damaged in plants due to presence of as causing electrolyte leakage.
- Plant transpiration intensity is reduced.

F.Effect of chromium:

- Chromium effects microorganisms in plants in the range from 67 to 5490 mg/kg of dry soil.

- The concentration of Cr in soil may vary according to natural composition of rocks and sediment that compose them.
- The level of chromium in soil may increase mainly through entheogenic deposition.
- Toxic effect of Cr on plant development includes alterations in germination process as well as in growth of roots, stems, leaves which may affect total dry matter production and yield.
- Chromium addition reduce soil pH with sludge soil pH.
- It gives deleterious effects on plant physiological process such as photosynthesis, water relations and mineral nutrition.
- Crops contain system that arrange chromium uptake to be low enough not to cause harm but when the amount of Cr increases in soil leads to higher concentration in crops. When concentration leads to certain value, negative effects can occur.
- The hazards depend on oxidation state of chromium.

G.Effect of copper:

- Copper is very common substance that naturally occurs in environment. Copper enter soil from the air during combustion of fossil fuels and settles on soil when it starts to rain. The copper compounds bound to water sediment or soil particles.
- Copper elements varies from 99mg/kg to 900 mg/kg. It ranges from 2 to 100ppm with average of 30ppm.
- The two factors commonly influence copper are soil pH and organic matter.
- Copper is highly toxic to microorganisms if present in access concentration, consequently changes soil biological equilibrium with adverse effect on both soil fertility, plant development and yield.
- It effects total yield, fruit number, dry root weight and plant height.When in access copper causes production of oxy radicals which has toxic effect in plant cells.
- Iteffects the root growth before effecting shoot production. And provides impair growth.
- On copper rich soils only limited number of plants has a survival. Therefore, copper is serious threat for production of farmlands.
- Copper can interrupt activity in soil as it negatively influences activity of microorganisms and earthworms. Thus, decomposition of organic matter may slow down.

IV. ALGORITHM

Multiple Linear Regression:

Multiple Linear Regression is an extension of the simple linear regression model in the way that there are multiple independent variables(features) that are used to predict the dependent variable.Suppose an analyst wants to know the price of the house then his linear equation will be the area of land as the independent variable and the price as the dependent variable. But in reality, more features affect the price of the house. The price of the house depends on other predictors like the floors in the house, Number of bedrooms, age of the house, etc. To understand the relationship in which multiple independent variables are involved, we use multiple linear regression.MLR is used to determine a mathematical relationship between the multiple independent variables. Once the factor or coefficient for each independent variable is determined then the information can be used to accurately predict the outcome. The relationship created by the model is informed of a linear (straight) line which best approximates all the individual data points.\

The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables.
- The independent variables are not too highly correlated with each other.y_i observations are selected independently and randomly from the population.

V. INPUT AND OUTPUT

The following some are the projects inputs and outputs.

Inputs:

Soil tested reports are collected from the labswhich includes different elements like Co, Ni, Zn, Pb, As, Cr, Cu.The input is sent manually to the ML model.

Outputs:

The model predicts the degree of contamination (DOC).Based on DOC, model classifies the soil suited for agriculture or not.

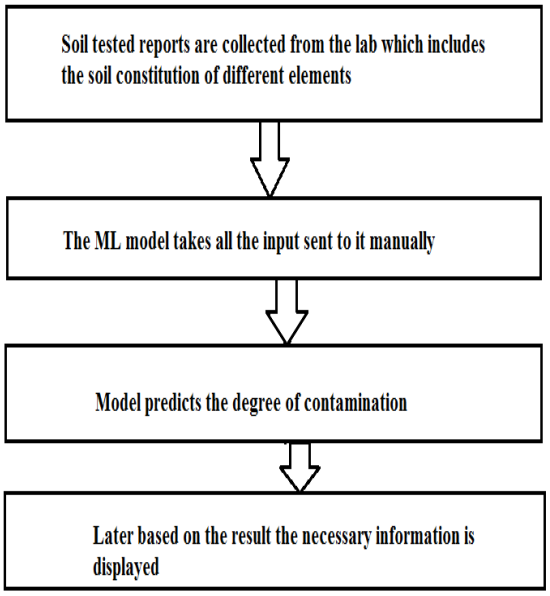
These may be classified in 4 outputs:

1.If soil is contaminated, bioremediation must be taken

2.If soil is less contaminated, suitable for agriculture with some remedies.

3.If soil is highly degradable, not suitable for agriculture.

4.If soil is least contaminated, suitable for agriculture.



This describes the actual and predicted values of contaminants of different elements. The actual and predicted values are approximately equal.

ELEMENT V/S DOC

Coefficient of different elements tells the degree of contamination that how much amount of particular element is present in soil and how does these elements effect the soil.Assessment includes describing the range of DOC which classifies the elements based on their DOC and how these elements effect soil based on different ranges of DOC.

- If DOC<4.5, soil is least contaminated.
- If DOC<6, soil is less contaminated.
- If DOC>6, soil is contaminated.
- If DOC>8.5, soil is highly contaminated.

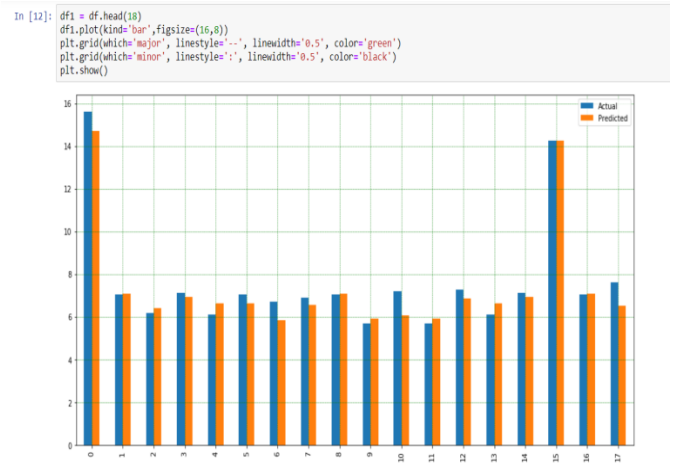


Fig:Actual v/s predicted values

The bar graph above represents the actual and predicted values.Blue bar represents actual values and orange bar represents predicted values.We can observe there is only a slight difference in actual and predicted values which means that model is predicting the values accurately.

VI.RESULT

- DOC is greater than 6. The soil is degraded and contaminated. Bioremediation must be done.
- DOC is less than 6. The soil is less contaminated. Soil is suitable for agriculture but with some measures.
- DOC is less than 2. The soil is least contaminated. Soil is suitable for agriculture without any remediation.
- DOC is greater than 8.5. The soil is highly degraded and contaminated. Soil is not suitable for agriculture

VII. CONCLUSION

This software gives the degree of contaminants of elements present insoil. This software tells whether the presence of these elements is suitable for agriculture or not.

REFERENCES

[1] Pradeep K Govil, Sujatha Dantu, Nitya Murthy, Keshav Krishna, "Soil Contamination of Heavy Metals in the Katedan Industrial Development Area, Hyderabad.

[2] Alyssa Ngu-Oanh Quanch, Lucie Tabor, Dany Dumont, Benoit Courceles,James-A.Goulet, "A Machine Learning Approach for Characterising Soil Contamination in the Presence of Physical Site Discontinuities and aggregated Samples", 2017.

[3] Faridon Ghadimi, "Prediction of Degree of Soil Contamination Baes on Support Vector Machine and K-Nearest Neighbour Methods", 5(4):345-353,2014.

[4] Chandan, Ritual Thakur,"Recent Trends of Machime Learning in Soil Classification", volume,08, September 2018.

[5] Krzysztof Loska, Danuta Wiechula, Irena Korus-Metal contamination of farming soils, Environmental International 30(2),159-165,2004

[6] KG Tiller-Urban soil contamination in Australia, Soil research 30(6) 937-957,1992

[7]AK Krishna, PK Govil-Assessment of heavy metal in soils around Chennai, Southern India, Environmental Geology 54(7),1465-1472,2008

[8] David G Rossiter-Classification of urban and industrial soils, Journal of soils and sediments 7(2),96-100,2007

[9] Jinwu, Yanguo Teng, Sijin Lu, Yeyao Wang-Evaluation of soil contamination indices in mining area of Jiangxi, China, Plos one 9(11),2014

[10] Wilhelm Kuhnelt-Soil biology with reference to animal kingdom, 1976