

# Classification of Soil Contamination

Mr.Karmegavarunan Kesavan  
*M.Sc Data Science and Business Analysis*  
*Rathinam College of Arts and Science, India*

Mr. Mervin George, M.Tech  
*Assistant Professor*  
Rathinam College of Arts and Science, India

**Abstract - All living things on Earth receive their nourishment from agriculture, which is a tree whose roots are the earth. It is somewhat to blame for our existence on this planet. Since soil is a non-renewable resource, as we have learned in both elementary and secondary education, the issue that needs to be answered is: How are we going to conserve it? It is clear from the soil data and graphs that soil pollution has reached an unacceptable level and that significant steps must be taken to reduce it. Although there are several methods available today to lessen soil pollution, none of them are optimal. Therefore, there has to be some solutions that farmers may also use in order to ensure that the quantity and quality of there is no impact on the produce. One area of computer science technology that can be very helpful in determining the best course of action for soil pollution is machine learning. The exciting part about experimenting with concepts in the field of machine learning is that it's an emerging technology. Our final year research, "Classification of Soil Contamination using Machine Learning," is a combination of these two themes. In this research, we will forecast if the soil entries we have examined are naturally polluted or artificially contaminated based on the state and metal proportions. The procedure has three primary modules, namely Data Pre-processing, Training, and Testing Modules. The ancillary goals we will incorporate are crop forecasting and An appropriate fertiliser is recommended.**

## **Keywords:**

**Soil contamination, agriculture, machine learning, forecast, soil pollution, crop, farmer.**

## I. INTRODUCTION

Given that environmental deterioration is now a major danger to the world's ecosystems, soil pollution is an important issue that has to be addressed. Conventional techniques of evaluation and cleanup are inadequate in providing timely and accurate solutions when soil contamination levels continue to grow. But the development of machine intelligence presents a viable way to transform our strategy for managing soil pollution. A novel framework for classifying soil pollution using machine learning algorithms is presented by the combination of computer science technology and soil science. This research aims to foresee and distinguish between naturally polluted and intentionally contaminated soil samples based on many variables, such as state and metal amounts, by utilising data analysis and predictive modelling. The objective of this project is to create strong classification models that can precisely identify the origins and level of soil contamination by integrating training, testing, and sophisticated data pre-processing modules. Furthermore, the suggested methodology's practical value is further enhanced by auxiliary goals like crop forecasting and fertiliser suggestion, which promote comprehensive solutions for sustainable agricultural practices. The ramifications of this research go well beyond science as we set out on our adventure at the nexus of environmental stewardship and technology. We not only lessen the current concerns to food security and agricultural output by using machine learning to address soil contamination, but we also provide the foundation for a more robust and sustainable future for future generations.

## 2. Related works

All living things on Earth receive their nourishment from agriculture, which is a tree whose roots are the earth. It is somewhat to blame for our existence on this planet. Since soil is a non-renewable resource, as we have learned in both elementary and secondary education, the issue that needs to be answered is: How are we going to conserve it? It is clear from the soil data and graphs that soil pollution has reached an unacceptable level and that significant steps must be taken to reduce it. Although there are several methods available today to lessen soil pollution, none of them are optimal. Therefore, there has to be some solutions that farmers may also use in order to ensure that the quantity and quality of there is no impact on the produce. One area of computer science technology that can be very helpful in determining the best course of action for soil pollution is machine learning.

### 3. Methodology

The methodology of the proposed system comprises of the following system is

- Data set collection
- Pre-processing
- Training and testing
- Prediction

#### 3.1 Dataset collection:

A number of procedures must be followed while gathering soil samples for machine learning applications in order to guarantee that the dataset is representative, diverse, and appropriate for the planned use. This is a broad overview of the procedure.

#### 3.1 Pre-processing

**NUMPY:** NumPy, often known as "Numerical Python" or "Numeric Python," is a machine learning module that is used to process large amounts of data. It facilitates quick mathematical computation on matrices and arrays for the user. Numpy's syntax is as follows:

```
>>import numpy as np
```

**PANDAS:** The most potent machine learning module or library that aids in providing a quick overview of the data set is called Pandas. It offers simple-to-use tools that lower the dimensionality and complexity of machine learning programs. Pandas is a great feature that makes it simple to experiment with various machine learning concepts.

Pandas syntax looks like this:

```
>>import pandas as pd
```

**MATPLOTLIB:** For the purpose of plotting graphs and correlation matrices, among other data visualization techniques, Matplotlib is an open-source tool. Because of its ability to visualize large data sets, users may select the right models and obtain the accuracy needed for their project with ease.

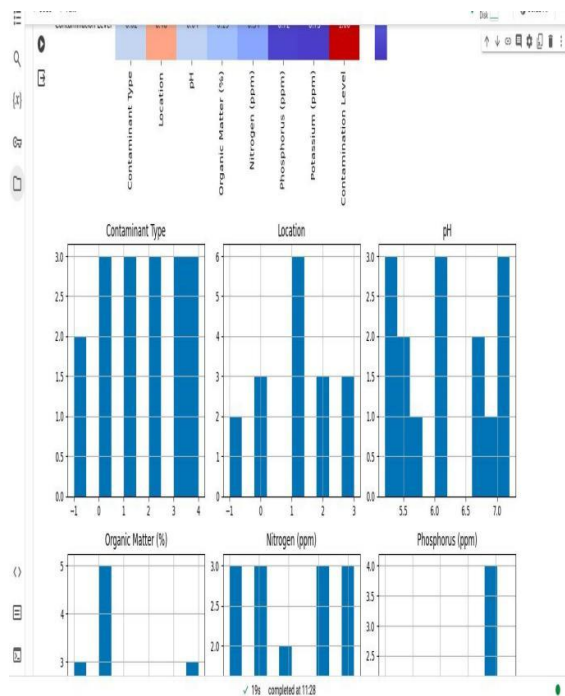
To import Matplotlib into Python, use:

```
>>import matplotlib.pyplot as plt
```

We have produced histograms, scatter plots, correlation matrices, grid correlation matrices, and other visualizations using matplotlib.

**SEABORN:** A matplotlib variant is called Seaborn. This technique for data visualization provides you with a high-level graphical user interface (GUI) to produce visually appealing and informative statistical visuals. Python can import Seaborn by using:

```
>>import seaborn as sns
```



Plotting categorical data and visualizing statistical relationships are two uses for Seaborn.

## 7. Conclusion

All things considered, CatBoost is an excellent option for categorization problems. Preprocessing is made simpler by the algorithm's natural capacity to handle categorical features. It performs well, attaining competitive accuracy and precision, and its clear feature importance analysis makes it easier to comprehend the key variables affecting predictions. CatBoost is effective and accessible due to its user-friendly interface, cross-validation support, and scalability via parallel and GPU training. The algorithm's better flexibility helped to reduce some of the challenges, like the requirement for hyperparameter tweaking. In the future, we plan to investigate ensemble approaches and gather more data to improve robustness. All things considered, CatBoost proves to be a flexible and strong instrument, proving its efficiency in classifying soil conditions and showing potential for wider machine learning uses.

## REFERENCES

- [1] M. Ito Possibility of automated soil quality judgment by machine learning, Proceedings of ZENCHIREN Technology Forum 2017 Asahikawa
- [2] S. Utsuki, M. Nakaya, R. Tsuruta Construction of a geological information management system using AI, CIM and image processing technology Journal of Japan Society of Engineering Geology, 58 (6) (2018), pp. 408-415
- [3] K. Koszela, J. Przybył, S. Kujawa, R.J. Kozłowski, K. Przybył, G. Niedbała, P. Idziaszek, P. Boniecki, M. Zaborowicz IT system for the identification and classification of soil valuation classes Proceedings of the Eighth International Conference on Digital Image Processing (ICDIP 2016), 10033 (2016), p. 100332J
- [4] M.C. Pegalajar, L.G.B. Ruiz, M. Sánchez-Marañón, L. Mansilla A munsell colour-based approach for soil classification using fuzzy logic and artificial neural networks Fuzzy Set Syst., 401 (15) (2020), pp. 38-54
- [5] S. Shojaei, M.A.H. Ardakani, H. Sodaiezhadeh, M. Jafari, S.F. Afzali Optimization using response surface method (RSM) to investigate the compaction of mulch Modeling Earth Systems and Environment, 5 (4) (2019), pp. 1553-1561
- [6] S. Shojaei, M.A.H. Ardakani, H. Sodaiezhadeh Optimization of parameters affecting organic mulch test to control erosion J. Environ. Manag., 249 (No. 1) (2019), p. 109414
- [7] Astel AM, Chepanova L, Simeonov V. Soil Contamination Interpretation by the Use of Monitoring Data Analysis. Water, Air, and Soil Pollution 2011;216(1-4):375-390. doi:10.1007/s11270-010-0539-1.
- [8] S. Shojaei, M.A.H. Ardakani, H. Sodaiezhadeh Simultaneous optimization of parameters influencing organic mulch test using response surface methodology Sci. Rep., 10 (1) (2020), pp. 1-11
- [9] W. Ng, B. Minasny, M. Montazerolghaem, J. Padarian, R. Ferguson, S. Bailey, A.B. McBratney Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra Geoderma, 352 (2019), pp. 251-267,
- [10] W. Ng, B. Minasny, A. McBratney Convolutional neural network for soil microplastic contamination screening using infrared spectroscopy Sci. Total Environ., 702 (2020), p. 134723,

- [11] Djahed, B., Taghavi, M., Farzadkia, M., Norzaee, S., and Miri, M. (2018). Stochastic exposure and health risk assessment of rice contamination to the heavy metals in the market of Iranshahr, Iran. *Food Chem. Toxicol.* 115, 405–412. doi:10.1016/j.fct.2018.03.040
- [12] Gautam, K., Sharma, P., Dwivedi, S., Singh, A., Gaur, V. K., Varjani, S., et al. (2023). A review on control and abatement of soil pollution by heavy metals: Emphasis on artificial intelligence in recovery of contaminated soil. *Environ. Res.* 115592, 115592. doi:10.1016/j.envres.2023.115592
- [13] Gustave, W., Yuan, Z., Liu, F., and Chen, Z. (2021). Mechanisms and challenges of microbial fuel cells for soil heavy metal (loid) s remediation. *Sci. Total Environ.* 756, 143865. doi:10.1016/j.scitotenv.2020.143865
- [14] Jiang, M., He, L., Niazi, N. K., Wang, H., Gustave, W., Vithanage, M., et al. (2023). Nanobiochar for the remediation of contaminated soil and water: Challenges and opportunities. *Biochar* 5 (1), 2. doi:10.1007/s42773-022-00201-x
- [15] Liu, L., Li, W., Song, W., and Guo, M. (2018). Remediation techniques for heavy metal-contaminated soils: Principles and applicability. *Sci. Total Environ.* 633, 206–219. doi:10.1016/j.scitotenv.2018.03.161
- [16] Peña, A. (2022). A comprehensive review of recent research concerning the role of low molecular weight organic acids on the fate of organic pollutants in soil. *J. Hazard. Mat.* 434, 128875. doi:10.1016/j.jhazmat.2022.128875
- [17] Singh, S. P., and Singh, M. K. (2020). “Soil pollution and human health,” in *Plant responses to soil pollution*. Editors P. Singh, S. K. Singh, and S. M. Prasad (Singapore: Springer).
- [18] Yuan, H., Wan, Q., Huang, Y., Chen, Z., He, X., Gustave, W., et al. (2021). Warming facilitates microbial reduction and release of arsenic in flooded paddy soil and arsenic accumulation in rice grains. *J. Hazard. Mat.* 408, 124913. doi:10.1016/j.jhazmat.2020.124913
- [19] Zhang, X., Jiang, M., He, L., Niazi, N. K., Vithanage, M., Li, B., et al. (2023). Pandemic COVID-19 ends but soil pollution increases: Impacts and a new approach for risk assessment. *Sci. Total Environ.* 890, 164070. doi:10.1016/j.scitotenv.2023.164070
- [20] Zhang, X., Li, B., Schillereff, D. N., Chiverrell, R. C., Tefsen, B., and Wells, M. (2022). Whole-cell biosensors for determination of bioavailable pollutants in soils and sediments: Theory and practice. *Sci. Total Environ.* 811, 152178. doi:10.1016/j.scitotenv.2021.152178
- [21] C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - *Journal of ELECTRICAL ENGINEERING*, Vol.63 (6), pp.365-372, Dec.2012.
- [22] C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’ - *Springer, Electrical Engineering*, Vol.93 (3), pp.167-178, September 2011.
- [23] C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’ - *Taylor & Francis, Electric Power Components and Systems*, Vol.39 (8), pp.780-793, May 2011.
- [24] C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’ - *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
- [25] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. “Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter” *Journal of VLSI Design Tools & Technology*. 2022; 12(2): 34–41p.
- [26] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya “Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay” *Asian Journal of Electrical Science*, Vol.11 No.1, pp: 1-8, 2022.
- [27] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash “Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM” *Suraj Punj Journal for Multidisciplinary Research*, 2021, Volume 11, Issue 4, pp:750-756
- [28] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, “Perfromance Investigation of T-Source Inverter fed with Solar Cell” *Suraj Punj Journal for Multidisciplinary Research*, 2021, Volume 11, Issue 4, pp:744-749
- [29] C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
- [30] M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, *Journal of Environmental Protection and Ecology*, Volume 23, Issue 2, pp: 520-530,2022
- [31] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, “Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District”, *International Research Journal of Multidisciplinary Technovation*, pp: 630-635, 2019