**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

# Categorization Of Micro-Plastic Concentration in Groundwater and Freshwater by Support Vector Machine Or SVMS

**M. Madhulatha [1], Farooq Sunar Mahammad[1*], P Sainiharika[2], B Sai Sri Harshitha[3], Bhuma Harshitha[4], B Sai Jyothi[5], K Sharana Priya[6]**

Department of Computer Science Engineering, Santhiram Engineering College, Nandyal
**Email:** madhulatha.cse@srecnandyal.edu.in

**ABSTRACT**

Among various models, the Support Vector Machine has been set up by having factors such as the minimum as well as the maximum size of the waste that is created by Micro-plastic and the level of concentration of micro-plastic in the groundwater measured in PPM. With the help of this study, we will be able to find out the patterns of the type of micro-plastic that are concentrated in which kind of water bodies. There are certain projected patterns that the concentration level of micro-plastic contamination in the seawater is lower due to possible deregulation. However, in Lake water, these concentration levels are much higher. Henceforth the above perceptions might or might not be justified by using the trained Support Vector Machine model which has been Stated.

**Keywords*:*** SVM, PPM, Machine learning

## 1. INTRODUCTION

For the survival of any living beings' water is one of the major as well as the most basic factors on this planet. The level of micro-plastic contamination has been a huge concern in current scenarios. In recent findings, it has been observed that the concentration of micro-plastic has been observed in drinking water and its various sources. This finding has established an argument on the implications of possible health issues on human beings. Although, there are various arguments regarding these research studies as there has been no such systematic sampling as well as data collection and identification of various methods for micro-plastics. In fresh water and drinking water, there is a significant concentration of micro-plastic which can be found in ten orders of magnitude ($1 \times 10^{-2}$ to $10^{8}$ #/$m^3$) that are found among various people as well as various types of water. Plastics that are less than 5 mm as well as they're in effects on human health have been researched on various water bodies such as rivers, seas, soils, and many more where various researchers are predicting the drastic impact on health does micro-plastics might have.

## 2. LITERATURE SURVEY

As per the argument presented in various research papers and the reference given in this study it can be observed that there has been a sufficient level of contamination of micro-plastic in freshwater as well as drinking water including their different sources. This level of concentration has triggered several arguments on the probability of its impact on human health. Although, there are various arguments regarding these research studies as there has been no such systematic sampling as well as data collection and identification of various methods for micro-plastics. In this study, we have studied various research papers on the presence of micro-plastic in freshwater as well as their possible sources. There are various sources from which data can be collected that find the presence of microplastic concentrations such as lake water, river water, groundwater, bottled water, and many more. Various samples have been collected from wastewater as well. In this study, we revealed as well as established the good practices for collecting samples, extract as well as identifying microplastics which includes providing qualitative and quality reviews on the study of the level of microplastic concentration. It has been established that

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

there is a requirement for a good quality of data to find out the occurrence of micro-plastic in freshwater to give and understand expected exposure and to focus on and highlight the risk to human health.[1]
Based on the review of the literature we propose an efficient software model using the SVM algorithm.

## 3. WORKING

**Proposed system:**

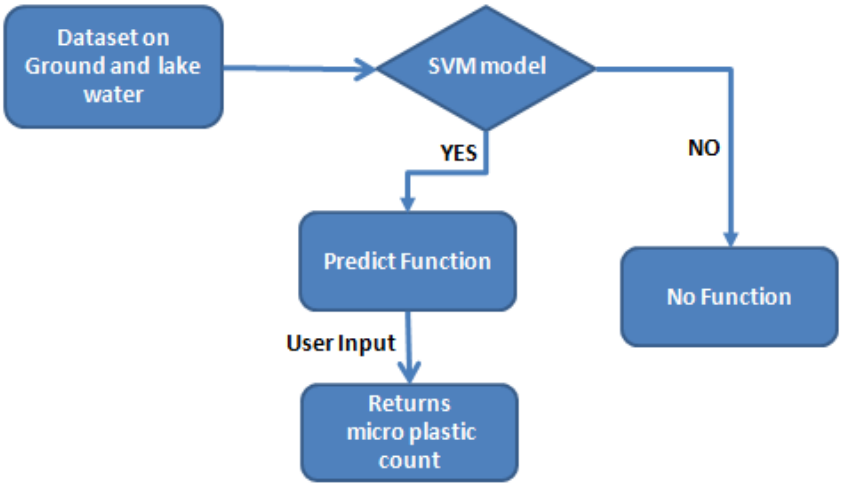


**Figure 1 Block Diagram**

This project employs an SVM algorithm to identify if a given water sample is a groundwater sample or a lake water sample. The x-axis consists of the TDS Level and the y-axis consists of the microplastic count. An SVM or Support Vector machine is termed as a categorized machine learning algorithm. Hence in this study two-dimensional SVM is deployed. That is, one input parameter and one output parameter are taken and the given dataset is plotted in a 2D graph. Two sets of points labeled as different groups are presented in the dataset. The output model of the Support Vector Machine is the best separation line, which is the line with the 'y=mx+c' equation, separating the two groups of given points of the dataset. This algorithm maps the segregation between the two sets of points.

**Support Vector Machine:**

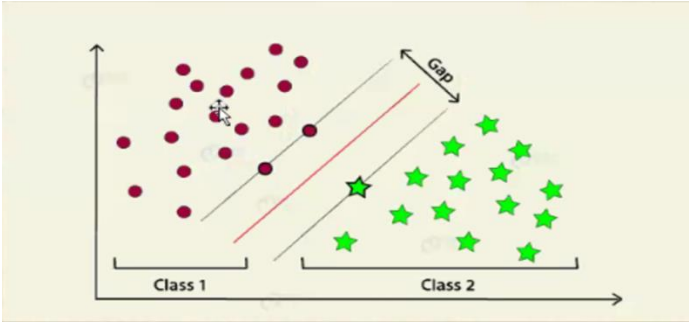


**Figure 2 SVM graph**

Support Vector Machine or SVM can be defined as a process of machine learning which mainly learns to assess data for regressive analysis as well as classification. The Support Vector Machine is a supervised learning process that observes data as well as splits it into mainly two categories. An output from an SVM is mainly a map that consists of sorted data along with

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

a margin that stretches as far as requirements. These SVMs have been used for recognizing text categorization, handwriting recognition, image segregation, in science, and many more. SVMs are also known as SVN Support Vector Network. SVM is called supervised learning of algorithms which sorts data into two main categories which are trained into a series of those data that are classified only into two major categories making the model primarily trained. the main requirement of Support Vector Machine algorithms is mainly to understand the category to which a new point of data belongs which makes SVM a non-binary linear classifier. the requirement of the SVM algorithm is not only just to project objects into different categories, it consists of the margins between them as wide as requires a graph. [2]

## 4. ARTIFICIAL INTELLIGENCE

The algorithmic flow presented in this paper falls under the field of Artificial Intelligence and Machine Learning. Presented below is an outline of those technologies.

AI or Artificial Intelligence can be termed as a type of intelligence presented by any machine, which is unlike any intelligence shown by any human beings as well as animals presenting both consciousnesses as well as emotions. The most powerful Artificial Intelligence is often presented as Artificial General Intelligence or AGI while on the other side, there are attempts to imitate various natural bits of intelligence which are termed Artificial Biological Intelligence or ABI. However, machines mirror the cognitive functions which humans often rely on with their human brain, for example, learning as well as problem-solving. [3]

As time passes, machines have become more and more capable and reliable and as a result, those tasks which can be considered to use intelligence are frequently bypassed as the definition of Artificial Intelligence, and this activity is often known as the AI effect. The capabilities of modern are often classified as Artificial intelligence that consists of a phenomenal knowledge of human speech, almost competing at the top level the most important game system, for example, chess as well as go, and on the other hand managing those games which are not perfect in information such as self-driving cars, pokers, simulations for the military as well as the process of intelligent routing in networks for delivering content.[3]

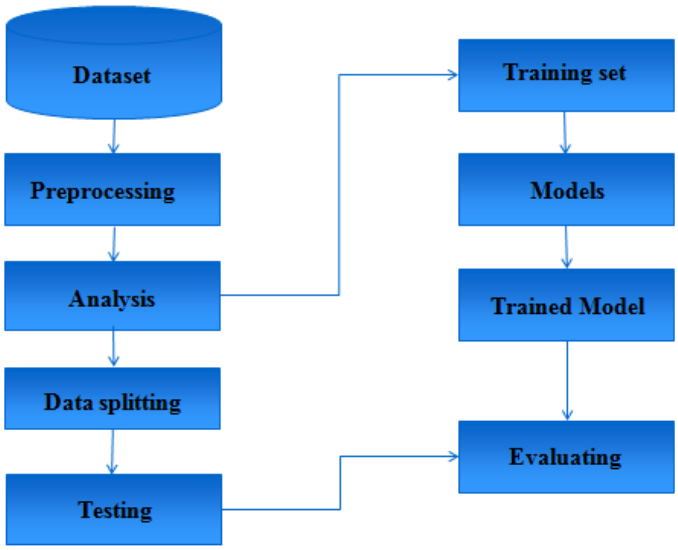


**Figure 3 Flow Chart**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

AI research has split into various sub-fields which frequently fail to communicate with one another. Various subfields are mainly focused on such considerations which are technical for example, machine learning as well as robotics, the use of certain tools, and many more. These are also based on certain social factors.[3]

During this twenty-first century has been an era of the emergence of Artificial Intelligence which has been witnessing huge advancements such as an increase in computer power, large databases as well as theoretical understanding. Artificial Intelligence has become a phenomenal part of technology which helps to solve many challenges. [3]

## 5. MACHINE LEARNING

ML or Machine learning is the study of algorithms that through experience as well as the use of data improves automatically and is also a part of Artificial Intelligence. ML algorithms are used in a variety of applications such as medicines, email filtering, speech recognition, and many more.[4]

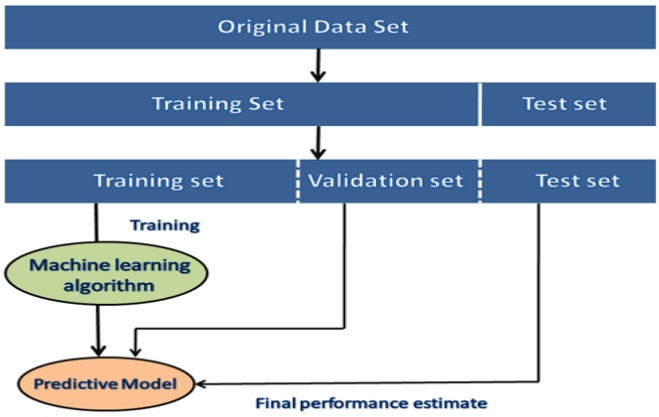


**Figure 4 Machine learning dataset**

ML or Machine learning or a subset of it is related to computational statistics for making future predictions. Acquiring knowledge on mathematical optimization delivers processes, theory as well as various applications that domain in machine learning. It is also considered a process of predictive analysis.[4]

## 5. INDUSTRY 4.0

Based on Wikipedia, industry 4.0, The Fourth Industrial Revolution reveals the progress of automation in various industries. The widespread machine-to-machine interaction is installed. Systematic communication, self-monitoring, and production as well as application of smarter machines that can perform huge tasks without any human interventions. huge amounts of tasks without any human interventions.[5]

Automation under Industry 4.0 has a particular schema or pattern at its outset. Presented below is how automation in the mass production industry, as well as consumer-level products, is built in today's technological era.

The schema presented in Figure 5 has a lot of other components involved but the generic outline of it stands justifiable for all kinds of automation today.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

The software automaton of the conventional automation model, which is the status quo, was built by a human expert or a team of human experts till now. With the advent of machine learning technology, the software automaton was not fully directly designed by human experts. The human experts build the machine learning software and give the real-world data set as training information. The machine learning software identifies the pattern between the input and the output parameters of the dataset in the form of a mathematical model. This mathematical model can be downloaded as a working software module to other electronic computing devices. This mathematical model is referred to as the 'trained machine learning module'. The software automaton of all the current digital embedded devices is a mathematical model that gives a numerical output for a numerical input based on arithmetic and logical conditions. This software automaton, as explained above can be either directly developed by a set of human experts by setting the boundary conditions themselves based on observation and requirements or can be downloaded as an executable module from machine learning training systems that are trained with the relevant dataset. In whatever way the software automaton is developed, it can be loaded onto the relevant embedded computing module that can be used for either sensor-based closed-loop automation or open-loop automation.

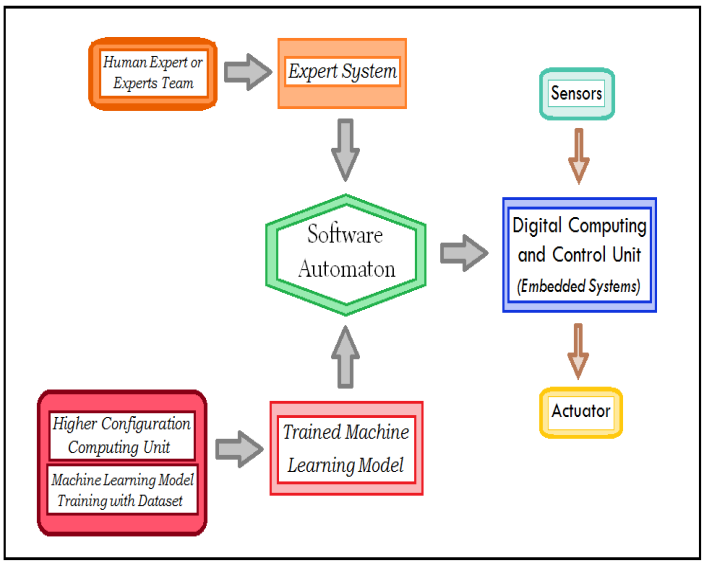


**Figure 5 Schema of Automation**

The technological components of Industry 4.0 include IoT, augmented reality, virtual reality, cloud computing, 3D printing, big data analytics, networking, data security, human-machine interaction, and others. IoT is a very effective way to collect real-world data. Sensors integrated with data acquisition and transmission systems can be placed anywhere and the collected data can be pre-processed if required and used as datasets to train machine learning models.

Cloud computing is employed for the optimized utilization of computing resources. There are many third-party vendors like Google and Amazon which are very reliable in terms of data security and speed of computation. These services offer companies and organizations a cheap and reliable way to harness the power of artificial intelligence and machine learning.

Big data analytics is the set of technological components involved with collecting, collating, and managing large quantities of data for analytics and decision making. When so much data is involved, especially with third-party service providers, data security plays an important role.

One of the paramount concerns about Industry 4.0 is the unemployment it can create due to powerful automation. The field of human-machine interactions and co-working has been a very developing field now to mitigate the above-mentioned problem.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

## 6. RESULTS AND DISCUSSION

| TDS | MPC | Water Type |
|---|---|---|
| 894 | 40 | Ground Water |
| 783 | 61 | Ground Water |
| 923 | 121 | Ground Water |
| 780 | 32 | Ground Water |
| 882 | 209 | Ground Water |
| 875 | 165 | Ground Water |
| 910 | 83 | Ground Water |
| 767 | 130 | Ground Water |
| 859 | 150 | Ground Water |
| 872 | 144 | Ground Water |
| 56 | 910 | Lake Water |
| 13 | 810 | Lake Water |
| 110 | 924 | Lake Water |
| 49 | 873 | Lake Water |
| 224 | 832 | Lake Water |
| 180 | 765 | Lake Water |
| 74 | 923 | Lake Water |
| 129 | 804 | Lake Water |
| 165 | 834 | Lake Water |
| 131 | 853 | Lake Water |

**Figure 6 Input Sample Dataset**

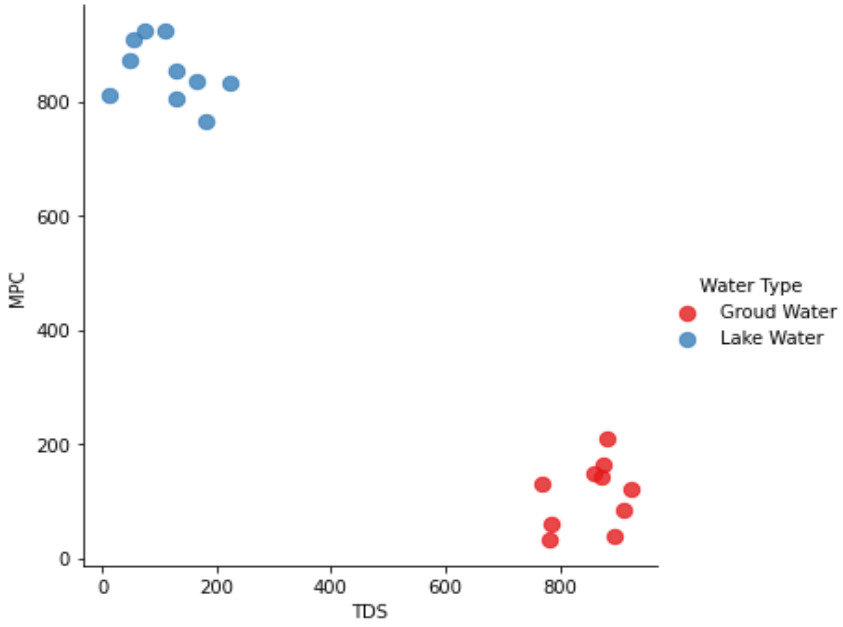


**Figure 7 Output graph 1**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 17 - 24
https://publishoa.com
ISSN: 1309-3452

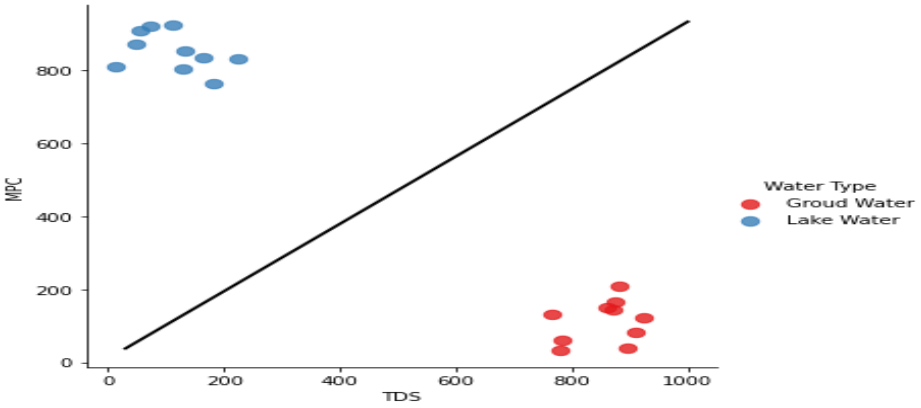


**Figure 8 Output graph 2**

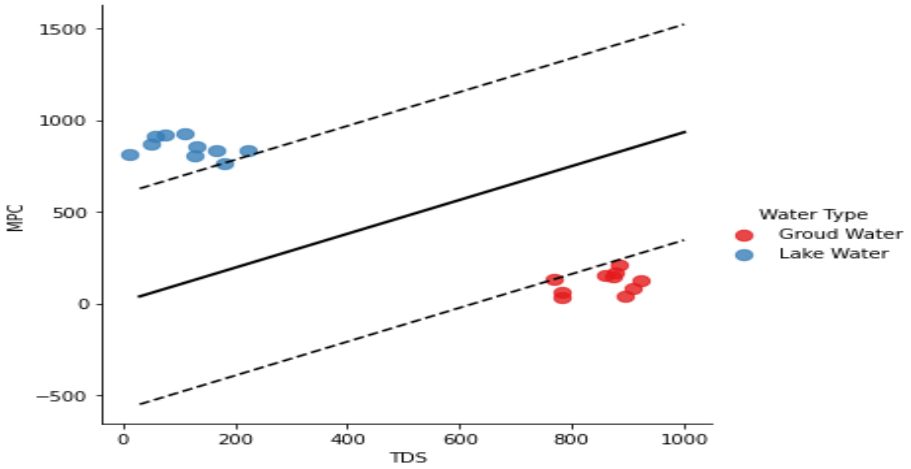


**Figure 9 Output graph 3**

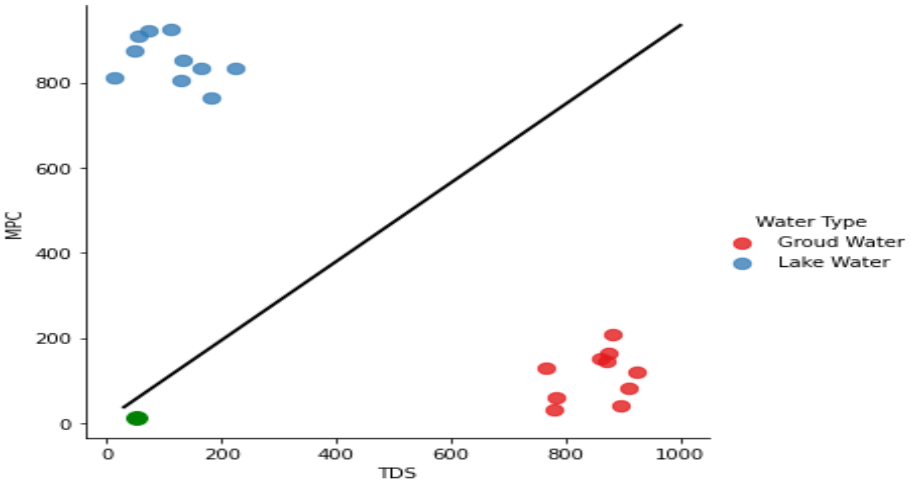


**Figure 10 Output graph 4**

This machine learning-based predictive model is implemented employing Python programming language. The relevant library files have been included for the execution of the code. The dataset was given as input and the predictive model was

obtained. The prediction model was tested and the results were satisfactory. The output graph of the machine learning model has been presented above. In this project, a Support Vector Machine algorithm is employed. The two parameters are the TDS level and microplastic count of the water sample. The Support Vector Machine model segregates and classifies the groundwater and lake water samples. This model can be used to either verify or disprove that groundwater has more microplastics than lake water.

## 7. CONCLUSION AND FUTURE WORK

This particular software automaton can be utilized in water research organizations. This particular software automaton can be used as a verification entity for microplastic distribution in groundwater and lake water. This model can be made better and better by constantly increasing the data size to be trained. The information on the microplastic count and the TDS can be obtained from IoT-based modules placed in various water sources. All this data is collected in a centralized server and analytics is performed on that. Automation under Industry 4.0 has a particular schema or pattern at its outset. The software automaton of the conventional automation model, which is the status quo, was built by a human expert or a team of human experts till now. With the advent of machine learning technology, the software automaton was not fully directly designed by human experts. The human experts build the machine learning software and give the real-world data set as training information. The machine learning software identifies the pattern between the input and the output parameters of the dataset in the form of a mathematical model. This mathematical model can be downloaded as a working software module to other electronic computing devices.

## REFERENCES

[1] Albert Koelmans, Nur Hazimah, Enya Hermsen, and Merel Kooi, “Microplastics in Freshwaters and Drinking Water: Critical Review and Assessment of Data Quality”, February 2019Water Research 155.
[2] https://www.techopedia.com/definition/30364/support-vector-machine-svm
[3] https://en.wikipedia.org/wiki/Artificial_intelligence
[4] https://en.wikipedia.org/wiki/Machine_learning
[5] https://en.wikipedia.org/wiki/Fourth_Industrial_Revolution