iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** V **Month of publication:** May 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.51998

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

# Classifying Data with Suggestive Causes and Flexible Solution

A.Micheal Clement[1], N. Daisy Deepika[2], K. Jenifer[3]
[1, 2, 3]*Sri Muthukumaran Institute of Technology*

## I. SDLC (SOFTWARE DEVELOPMENT LIFE CYCLE)

The Software Development Life Cycle is a systematic process for building software that ensures the quality and correctness of the software built. SDLC process aims to produce high-quality software which meets customer expectations. The software development should be completed within the pre-defined time frame and cost.

SDLC Phases

The entire SDLC process is divided into the following stages:




Phase 1: Requirement collection and analysis
Phase 2: A feasibility study
Phase 3: Design
Phase 4: Coding
Phase 5: Testing
Phase 6: Installation/Deployment
Phase 7: Maintenance

## II. PLATFORM KNOWLEDGE

### *A. Introduction to Java*

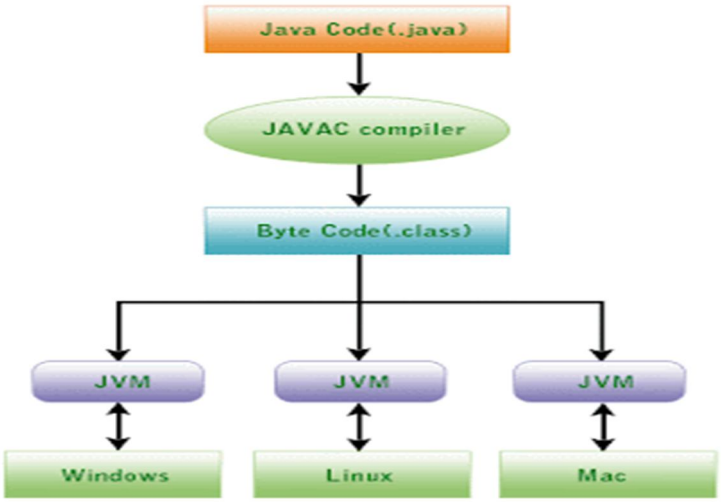
Java programming language was originally developed by Sun Microsystems which was initiated by James Gosling and released in 1995 as a core component of Sun Microsystems' Java platform. Initially, the language was called "Oak" but it was renamed "Java" in 1995. The primary motivation of this language was the need for a platform-independent language. Finally, Java is for Internet Programming whereas C was for System Programming.

### *B. Java Architecture*

Java is a high-level Object-oriented programming language. A program written in a high-level language cannot be run on any machine directly. First, it needs to be translated into that particular machine language.

©IJRASET: All Rights are Reserved 

The java compiler does this thing, it takes the java program (.java file containing source code) and translates it into machine code (referred to as byte code or .class file). Java Virtual Machine (JVM) is a virtual machine that resides in the real machine (your computer) and the machine language for JVM is byte code. JVM executes the byte code generated by the compiler and produces output. JVM is the one that makes the java platform independent.



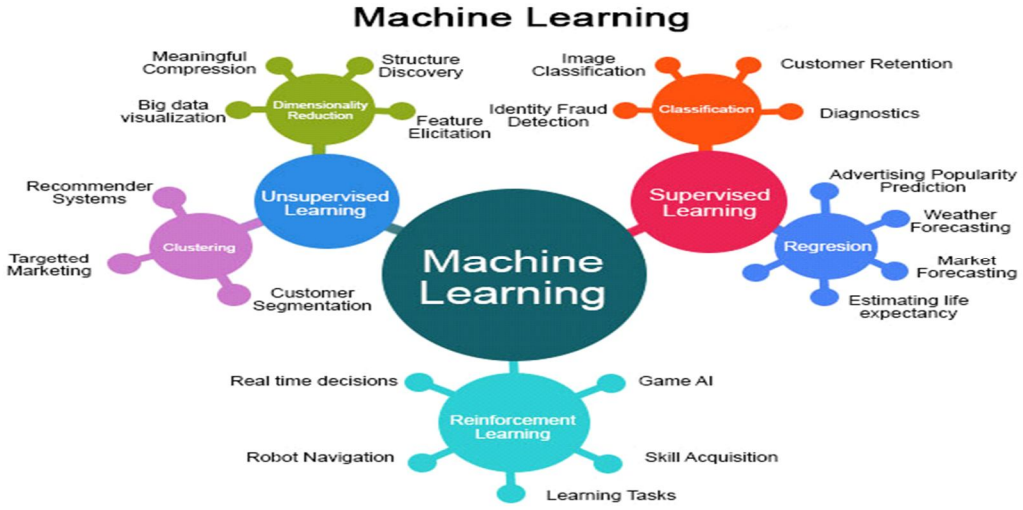

## III. DOMAIN KNOWLEDGE

### A. *AI – (Aritificial Intelligence)*

Artificial intelligence (AI) is intelligence perceiving, synthesizing, and inferring information demonstrated by machines, as opposed to intelligence displayed by animals and humans.AI applications include advanced web search engines (e.g., Google), recommendation systems (used by YouTube, Amazon and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Tesla), automated decision-making and competing at the highest level in strategic game systems (such as chess and Go). As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology.

### B. *Machine Learning*

Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computer to learn automatically without human intervention or assistance and adjust actions accordingly.



*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

*C. Machine Learning Methods*

Some of the methods of Machine Learning algorithm are categorized as

*D. Supervised Learning*

A Supervised learning algorithm learns from labelled training data, helps you to predict outcomes for unforeseen data. It is highly accurate and trustworthy method.

*E. Unsupervised Learning*

Unsupervised learning algorithm is the type of self - organized with the help of previously unknown patterns in dataset without pre-existing labels.

*F. Semi-Supervised Learning*

Semi-supervised learning is the combination of both supervised and unsupervised which means labelled and unlabelled data.

*G. Reinforcement Machine Learning*

Reinforcement machine learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

*H. Applications Of Machine Learning*

*1)* Video Surveillance
*2)* Social Media Services
*3)* Email Spam and Malware Filtering
*4)* Financial Services
*5)* Health Care
*6)* Retail
*7)* Transportation

Classifying data with suggestive causes and flexible solution

## IV. ABOUT THE PROJECT

*A. Abstract*

In this project we have proposed efficiently classifying the bulk of the marine species data on certain requirements and certain conditions also. After that we have to finding the which are all the species are endangered species or threatened species based on the requirement. Here not only classifying the bulk of data and also we have give the causes and solution to the classifying data. So that we have to use two algorithms, one is decision tree algorithm and another one is logistic regression algorithm. The decision tree algorithm is used for classifying the bulk of data based on the requirement and certain condition. The logistic regression algorithm is used for given causes and solution to that in a prospective way of approach with well defined methodology. This application has proposed a method for effectively researching and analysing the bulk of data based on the requirement and condition. We use decision trees also to understand classification, segregation, and arrive at a numerical output or regression. This is a supervised learning algorithm that is used for classifying problems.

In this algorithm, we split the bulk of data into two or more homogeneous sets based on the most significant attributes and conditions. In an automated process, we use a set of algorithms and tools to do the actual process of decision making and branching based on the attributes of the data.

The originally unsorted data at least according to our needs must be analyzed based on a variety of attributes in multiple steps and segregated to reach lower randomness or achieve lower entropy.

It aids in the development of quick machine learning models capable of making accurate predictions. To estimate discrete values (often binary values like 0/1) from a set of independent variables, logistic regression is utilized. By adjusting the logic function to the data, it aids in predicting the likelihood of an event. These algorithms are effectively works on this application. We refer to logistic regression as a binary classifier, since there are only two outcomes. Here also it will gives the causes and solution to that .

©IJRASET: All Rights are Reserved 

*B. Scope Of The Project*

The project's purpose is to predict whether the endangered species or threatened species based on the requirement given by the clients. The bulk of data are manipulated, searching or filtering by the human is too difficult by the humans because it will take lot of time and the final output of the report is not accurate. Because of lot of data is not handled manually by humans. So that situation how to overcome these aspect here we will applied the machine learning concept to resolving these problems. Simply we will give the bulk of data as input to the machine it will give the accurate output. Which is broadly defined as the capability of a machine to imitate intelligent human behaviour. ML systems are used to perform complex tasks in a way that is similar to how humans solve problems. After classification based on the requirement then report will be generate in the pdf format. After that it will we be send to the analyst for analysing the report for how these species are in endangered or threatened. After the analysing then final report will be generated and send to the client.

*C. Existing System*

In the existing system, there is very difficult to identify the which is endangered and which is threatened from the bulk of data. These data are taken by the oceanographers. Which was taken in certain area in ocean in certain period of time? By taking the species data over a period of time there is lot of species are recorded . The count also massive, after that recorded data are compared with the previously recorded data the manipulation of taken the endangered species or threatened species is very difficult. At the same time it takes lot time also the accuracy of the manipulated data is questionable. There are 2 million known species are in the ocean . These are handled by manually is not an easy job . They can face lot of issues like avoidable errors , lot of time consumption, inconsistency, high cost of training, lag of accuracy. The inaccuracy may leads to be biggest disaster for the research.

*1) Disadvantages*

*a)* The manual process is undertaken by humans who are not infallible in the performance of repetitive tasks.
*b)* There is no way a man can complete with the machine in term of processing speed.
*c)* In an environment where consistency of data is crucial to the success of the system.
*d)* Manual data manipulation service is always a disadvantage as maintaining consistency for humans is a challenging task that must be avoided.
*e)* Classified information may leak, and sensitive data may develop legs and walk away and thus compromise the entire research.

*D. Proposed System*

In the proposed system, we have implemented that each client gets their own ID and password. . Based on the company name, the upload and retrieve of the data is very convenient for the clients. This system which has the enhancement of security as well as the fast data manipulation. Improving accuracy: Machine learning algorithms can achieve much higher accuracy than humans when making predictions or classifying labelled data. This improved accuracy can lead to better business outcomes and increased profits. Here we have implemented the bulk of species data as input and it will find the output as endangered species data or threatened species data as per the requirements. Not only find out the endangered species data and threatened species data also gives the causes of those species are going to be endangered or threatened.

*1) Advantages*

*a)* Big data analysts are looking at machine learning as the most effective source for precise data prediction.
*b)* It will give high accuracy on data manipulation.
*c)* High accuracy will helps the good research results.
*d)* This method can handle an unlimited amount of data, assess them and provide a proper analysis for the same.
*e)* Besides being cost-effective and time-saving, it is also easy to operate.

## V. BOTTOM LINE AND FUTURE ENHANCEMENT

Our proposed version has applied the decision tree algorithm and logistic regression algorithm. The two types of algorithms are more effectively works on the application. For the purpose of classifying the bulk of data we have use the decision tree algorithm. And the purpose of finding the causes and given solution to the causes we have use the logistic regression algorithm. The classification process will find out which species are endangered or threatened based on the given requirements. Then collect the water sampling lab test details from the oceanographers.

Because identifying why those species are going to be endangered or threatened. These sampling details are collected certain period of time. Thus our proposed version makes the top notch effect and satisfies required want in research .Main blessings of this method is to give good accuracy and correctly at the time. In future it has been enhanced and applied with experimented for an effective needed situations.

## VI. HARDWARE AND SOFTWARE REQUIREMENTS

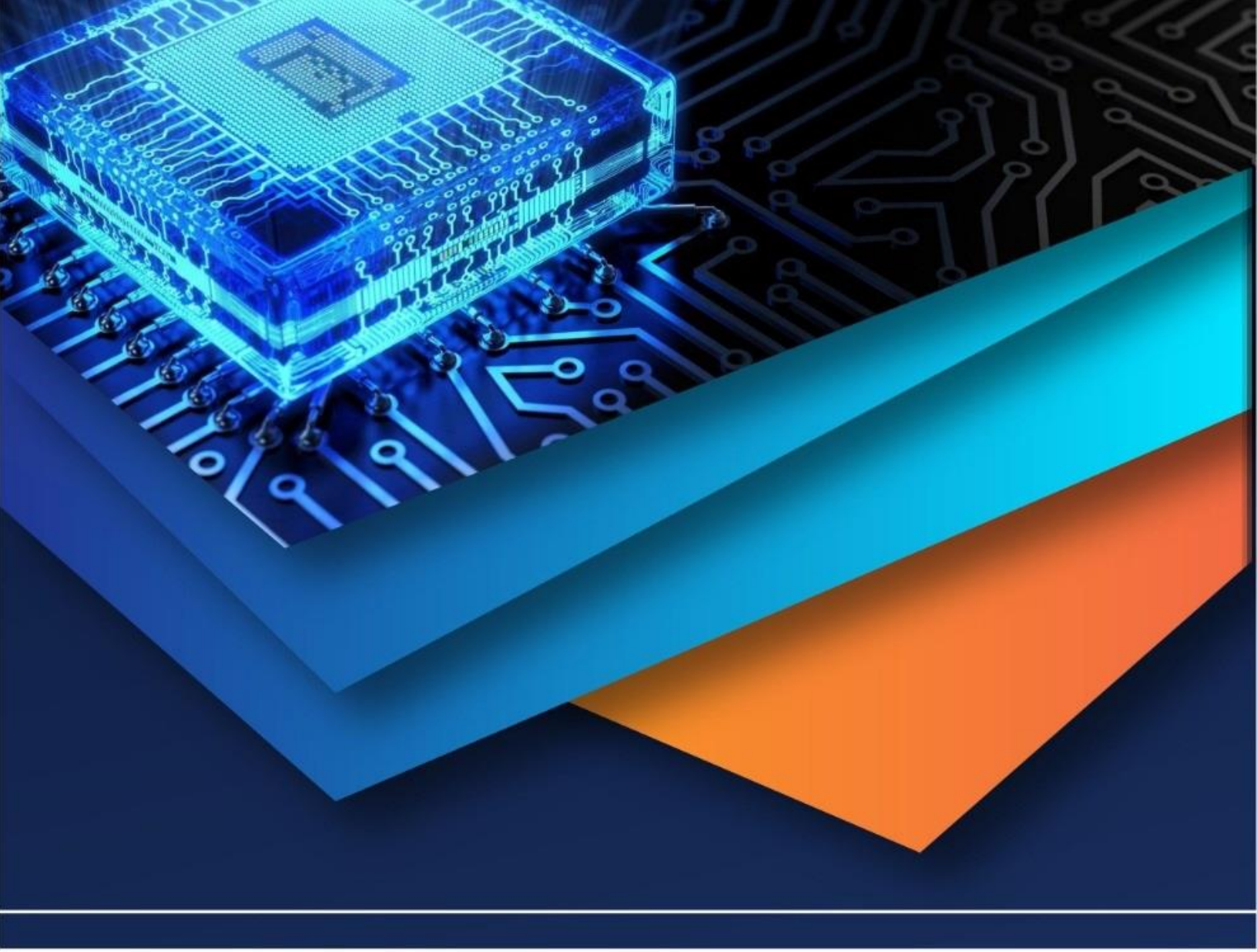
### A. *Hardware Requirements*

| | | |
|---|---|---|
| *1)* | Processor | : Intel (R) Pentium (R) |
| *2)* | Speed | : 1.6 GHz and Above |
| *3)* | RAM | : 4 GB and Above |
| *4)* | Hard Disk | : 120 GB |
| *5)* | Monitor | : 15'' LED SVGA |
| *6)* | Input Devices | : Keyboard, Mouse |

### B. *Software Requirements*

| | | |
|---|---|---|
| *1)* | Operating system | : Windows 7 / 8 / 8.1 / 10 |
| *2)* | Coding Language | : JAVA / J2EE |
| *3)* | Java Version | : JDK v8 |
| *4)* | IDE | : Eclipse Oxygen / Neon |
| *5)* | Database | : MySQL v5.1 |
| *6)* | Database Tool | : HeidiSQL v11.0 |
| *7)* | Application Server | : Apache Tomcat 8.x / 9.x |

©IJRASET: All Rights are Reserved 



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)