**METHODOLOGY**

**KEY FACTORS IN METHOD ASPECTS :**

**DATA TYPE :** Finding the data type whether it’s theoretical or technical data we had collected .

**RESOURCE AVAILABILITY :** Before collecting the data , we need to check the avail resource related to the data is present or not.

**VALIDITIY AND RELIABILITY :** Selecting the data which give precise result and more accurate results .

**ANALYSE DATA :** After the data collection done , we need to analyze the given metrics we would use .

**EVALUATION :** Machine Learning accuracy gives us the correct results.

**OBJECTIVE :** The study of AI based Mental HealthCare System .

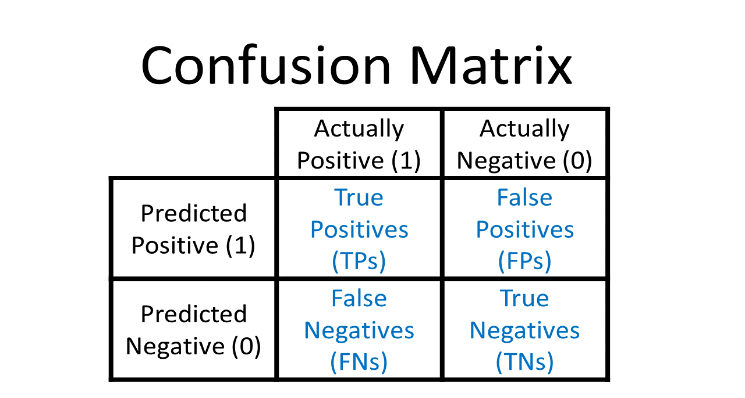
**METHOD :** Dataset and Machine Learning algorithm and Evaluation Metrics .

**EVALUATION METRICS :** These are the quantitative measures used to access the performance and efficiency of the given machine learning model . It helps to compare the effectiveness of different model accuracy .

Accuracy= No. of correct predictions​ / Total number of input samples

**CONFUSION MATRIX :** A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known.The key classification metrics : , Accuracy , Recall , Precision and F1 Score . The table is divided into four terminologies, which are as follows:

1. True Positive(TP): In this case, the prediction outcome is true, and it is true in reality, also.
2. True Negative(TN): in this case, the prediction outcome is false, and it is false in reality, also.
3. False Positive(FP): In this case, prediction outcomes are true, but they are false in actuality.
4. False Negative(FN): In this case, predictions are false, and they are true in actuality.



**RECALL :** It aims to calculate the proportion of actual positive that was identified incorrectly. It can be calculated as True Positive or predictions that are actually true to the total number of positives.

**PRECISION :** It is used to overcome the limitation of accuracy . It determines the proportion of positive prediction that are correct.

**F1 SCORE :** F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class. It is calculated with the help of Precision and Recall. It is a type of single score that represents both Precision and Recall.

**AUC ROC :**ROC represents a graph to show the performance of a classification model at different threshold levels. The curve is plotted between two parameters, which are: True Positive rate and False Positive rate .

**LOGARITHMIC LOSS :** It is also known as log loss . It basically works for finding the false positive classification of the model .

**NLP :** Natural language processing (NLP) is a machine learning (ML) technology that enables computers to understand human language. NLP is a subfield of artificial intelligence (AI). NLP uses computational linguistics, which is the study of how language works, and various models based on statistics, machine learning, and deep learning. These technologies allow computers to analyze and process text or voice data, and to grasp their full meaning, including the speaker’s or writer’s intentions and emotions.

**CONCLUSION :** Among the above evaluation metrics and NLP , confusion matrix gives more accuracy regarding the above mentioned objective for studying the machine learning model.

**OBJECTIVE :** Testing of AI based mental healthcare system .

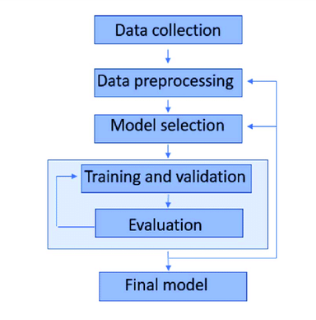
**METHOD :** Data Processing, Decision Tree and evaluation metrics.

**DATA PROCESSING :** Data Processing is the task of converting data from a given form to a more usable and desired form i.e. making it more meaningful and informative. The main steps required for data processing are the data collection , preprocessing , analysis, interpretation , storage and management , visualization and reporting .

**Mathematical formula :**

[where y is the target variable, x1, x2, ... xn are the scaled features, b is the bias, and w1, w2, ... wn are the model weights]

**Work Flow Diagram :**

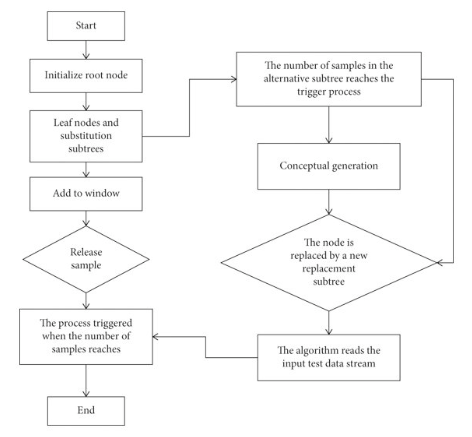


**DECISION TREE :** A decision tree is a part of supervised learning of machine learning that models decision based on input features .It forms a tree-like structure where each internal node represents a decision based on an attribute, leading to leaf nodes representing outcomes. Decision trees aid decision-making by representing complex choices in a hierarchical structure. Each node tests specific attributes, guiding decisions based on data values. Leaf nodes provide final outcomes, offering a clear and interpretable path for decision analysis in machine learning.

**Mathematical formula :**

[where "S" is a set of data points, and "π" is the probability of class "i" within that set.]

**Work Flow Diagram :**



**EVALUATION METRICS :** Evaluation metrics in machine learning are used to assess the performance of a model. They help determine how well a model makes predictions or classifications on data. It helps to find out the type of metric classification we can choose for the data evaluation thing . There are two types of evaluation metrics are as follows : Classification Metrics and Regression Metrics.

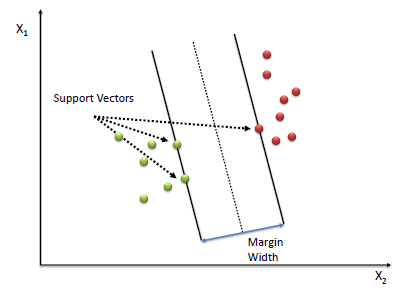
**CONCLUSION :** From the above method used such as data processing , decision tree and evaluation metrics , the most constructed conclusion of the method gives us on the decision tree part . As it helps to identify the most accurate result of dataset.

**OBJECTIVE :** Management Implication in the AI based mental health care study .

**METHOD :** Data Collection , Data Analysis, Ethical.

In data analysis , SVM , RANDOM FOREST and logistic regression are the part of the method used .

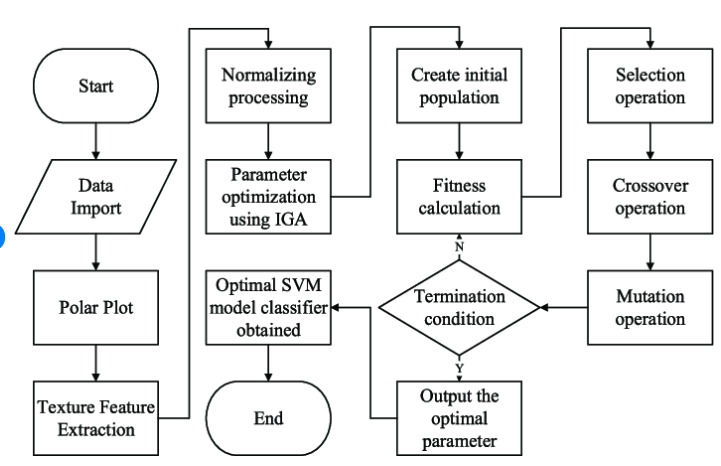
**SVM :** SVM(Support Vector Machine) is a part of supervised machine learning and is used for classification and regression tasks. The goal of SVM is to identify a hyperplane that effectively divides a dataset into distinct classes. This division maximizes the margin, defined as the distance between the closest points from both classes to the hyperplane.



**Mathematical Formula :**

[where K is the kernel function, x is a data point, and x' is another data point.]

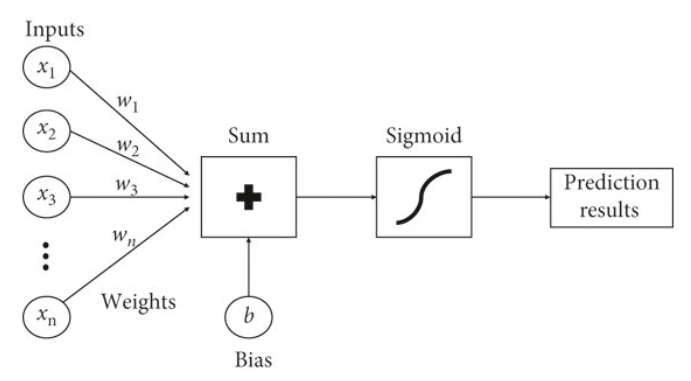
**Work Flow Diagram :**



**LOGISTIC REGRESSION :** Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. In Logistic regression, instead of fitting a regression line, we fit an “S” shaped logistic function, which predicts two maximum values (0 or 1).

**Mathematical Formula :**

**Work Flow Diagram :**



**CONCLUSION :** Both the methods are equally potential in their respective accuracy task , but keeping in mind the accuracy and more near extend of the object the logistic regression is more accurate as it values are near to the equation .

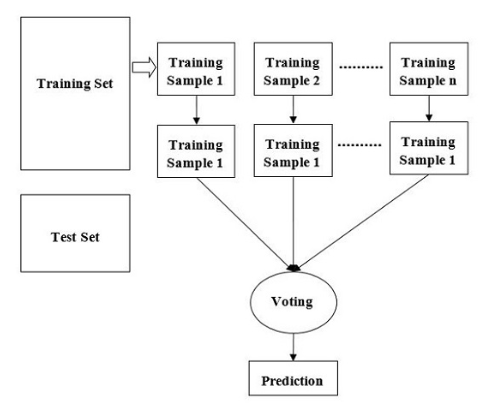
**OBJECTIVE :** The use of AI in mental healthcare and neurobiological research.

**METHOD :** Random forest , k-means clustering and Genetic Algorithm.

**RANDOM FOREST :** It is a collection of many decision trees where it only helps for prediction and it’s a part of the supervised machine learning part and is based on ensemble learning . Multiple Decision Trees are created from the training data. Each tree is trained on a random subset of the data (with replacement) and a random subset of features. This process is known as bagging or bootstrap aggregating. Random Forest builds many decision trees, each on a random subset of data and features. It then combines their predictions for a more accurate and robust result.The randomness makes it less prone to overfitting.

Mathematical formula : Accuracy , Recall ,Precision and F1 Score

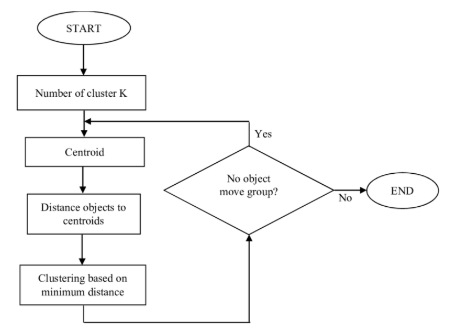
Work Flow Diagram :



**K- MEANS CLUSTERING :** K-means clustering is a machine learning algorithm that groups similar data points into clusters. It's an unsupervised learning technique, which means it works with unlabelled data.  technique used to organize data into groups based on their similarity.

**Mathematical Formula :**

**Work Flow Diagram :**



**GENETIC ALGORITHM :** In machine learning, "GA" stands for "Genetic Algorithm," which is a heuristic search optimization technique inspired by natural selection principles, used to find near-optimal solutions to complex problems by iteratively improving candidate solutions through operations like selection, crossover, and mutation, essentially mimicking biological evolution to explore large solution spaces efficiently.

* **Population:** A set of potential solutions (individuals) are represented as "chromosomes" with "genes" representing parameters or features.
* **Fitness function:** Each solution is evaluated based on a fitness function that determines how well it performs on the problem.
* **Selection:** Individuals with higher fitness are more likely to be chosen to "breed" and produce offspring.
* **Crossover:** Genetic material from selected parents is combined to create new offspring solutions.
* **Mutation:** Random changes are introduced to the offspring to maintain diversity in the population.
* **Iteration:** This process of selection, crossover, and mutation repeats over multiple generations until a satisfactory solution is found.

**Mathematical formula :**

**1. Population Representation :** A population PtP\_t at generation tt consists of multiple individuals (candidate solutions):

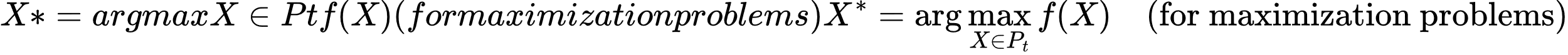
C:/Users/Debanjan Das/AppData/Local/Temp/wps.gAPVQmwps

Each individual XitX\_i^t is typically represented as a set of parameters, often in binary, real, or symbolic form.

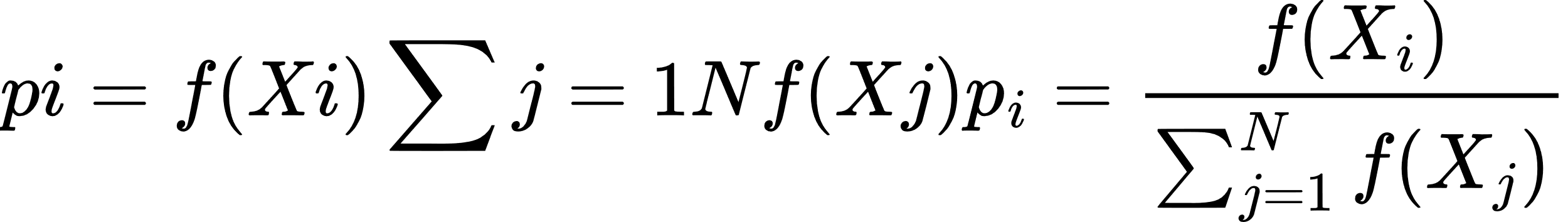
**2. Fitness Function**

Each candidate solution is evaluated using a function f(X)f(X) that measures how well it performs:

The goal of GA is to maximize or minimize this function:



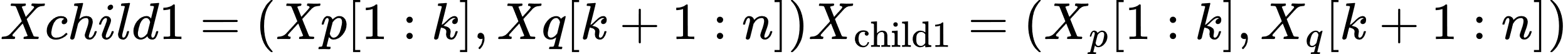
**3. Selection Mechanism**

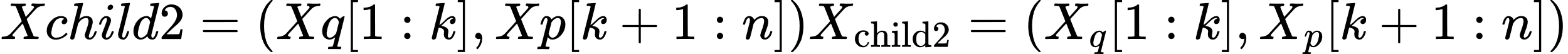
The probability of choosing a particular solution XiX\_i is typically based on its fitness value. Using **Roulette Wheel Selection**, the selection probability is: 

Alternative selection strategies include **tournament selection** and **rank-based selection**.

**4. Crossover Operation**

Crossover combines information from two parent solutions to produce offspring. For single-point crossover, if kk is the crossover point, the new solutions are formed as:





**5. Mutation Process**

Mutation ensures genetic diversity by introducing small random changes in an individual. In a real-valued representation, mutation can be expressed as:

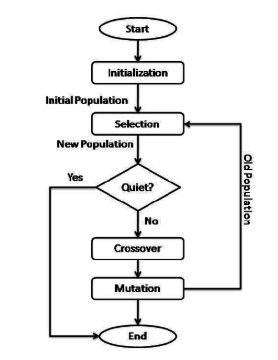
where δ\delta represents a small random variation.

**6. Stopping Criteria**

The algorithm terminates when one of the following occurs:

* A solution meets a predefined fitness threshold.
* The maximum number of generations is reached.

**Work Flow Diagram :**



**CONCLUSION :** Among the three above methods K-means Clustering is more accurate than random forest and genetic algorithm .