# Disease Prognosis System Using Machine Learning
# (Due to Covid 19)

**MAJOR PROJECT REPORT**

**Submitted for the partial fulfillment of the requirement for the award of Degree**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted By:**                                                      **Guided By:**

**Abhiraj Barsaiya -0101CS191003**                      **Dr. Rajeev Pandey**

**Deevyarshe Verma -0101CS191043**                   **Dr. Raju Baraskar**
**Kaoshlendra Singh -0101CS191057**
**Lalit Silawat -0101CS191061**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**UNIVERSITY INSTITUTE OF TECHNOLOGY**

**RAJIV GANDHI PROUDYOGIKI VISHWAVIDALAYA**

**BHOPAL-462033**

**SESSION 2019- 2023**


**RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**


# CERTIFICATE

This is to certify that **Abhiraj Barsaiya, Deevyarshe Verma, Kaoshlendra Singh, Lalit Silawat** of B.TECH. Fourth Year, Computer Science & Engineering have completed their Major Project entitled "Disease Prognosis System Using Machine Learning" during the year 2022-23 under our guidance and supervision.

We approve the project for the submission for the partial fulfillment of the requirement for the award of degree of B.TECH in Computer Science & Engineering.



**Dr. Rajeev Pandey**                          **Dr. Raju Baraskar**

Project guide                                        Project guide

# DECLARATION BY CANDIDATE

We, hereby declare that the work which is presented in the major project, entitled "Disease Prognosis System Using Machine Learning" submitted in partial fulfillment of the requirement for the award of Bachelor degree in Computer Science and Engineering has been carried out at University Institute of Technology RGPV , Bhopal and is an authentic record of our work carried out under the guidance of **Dr. Rajeev Pandey** (Project Guide) **and Dr. Raju Barskar** (Project Guide) ,Department of Computer Science and Engineering, UIT RGPV, Bhopal.

The matter in this project has not been submitted by us for the award of any other degree

Abhiraj Barsaiya
[0101CS191003]

Deevyarshe Verma
[0101CS191043]

Kaoshlendra Singh
[0101CS191057]

Lalit  Silawat [0101CS191061]

# ACKNOWLEDGEMENT

After the completion of major project work, words are not enough to express our feelings about all those who helped us to reach our goal, feeling  above all this is our indebtedness to the almighty for providing us this moment in life.

First and foremost, we take this opportunity to express our deep regards and heartfelt gratitude to our project guide **Dr. Rajeev Pandey and Dr. Raju Baraskar of Computer Science and Engineering Department, RGPV Bhopal** for their inspiring guidance and timely suggestions in carrying out our project successfully. They have also been a constant source of inspiration for us.

We are extremely thankful to **,  Prof. Uday Chaurasia Head, Computer Science and Engineering Department, RGPV Bhopal** for his cooperation and motivation during the project. We would also like to thank all the teachers of our department for providing invaluable support and motivation. We are also grateful to our friends and colleagues for their help and cooperation throughout this work.

 

Abhiraj Barsaiya [0101CS191003]

Deevyarshe Verma [0101CS191043]

Kaoshlendra Singh [0101CS191057]

Lalit  Silawat [0101CS191061]

# Table of Contents

# CHAPTER 1

## INTRODUCTION

### What is Machine Learning

Machine learning is a branch of <u>artificial intelligence (AI)</u> and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix's recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

As machine learning technology has developed, it has certainly made our lives easier. However, implementing machine learning in businesses has also raised a number of ethical concerns about AI technologies.

The machine learning process begins with observations or data, such as examples, direct experience or instruction. It looks for patterns in data so it can later make inferences based on the examples provided. The primary aim of ML is to allow computers to learn autonomously without human intervention or assistance and adjust actions accordingly.

### Benefits of Machine Learning

- Easily identifies trends and patterns

- No human intervention needed (automation)

- Handling multi-dimensional and multi-variety data

- Rapid Analysis Prediction and Processing

# CHAPTER 2

# LITERATURE SURVEY

## Existing problem

Different vaccines such as covi-shield and co-win were applied to applied to covid patients and others for treating covid or preventing it. But these vaccines come with side effects on lung, liver and hearts. Our major projects deal with the prediction of risk a person has after treatment with that certain vaccine. Datasets and Machine Learning Models are used for this purpose.

Which is why we decided to present a prediction model for everyone to be ready beforehand.

## Our approach

We have proposed the solution as to make the prediction model using Machine Learning.
Prediction in machine learning refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome.

Prediction in machine learning is commonly used for security, marketing, operations, risk, and fraud detection.
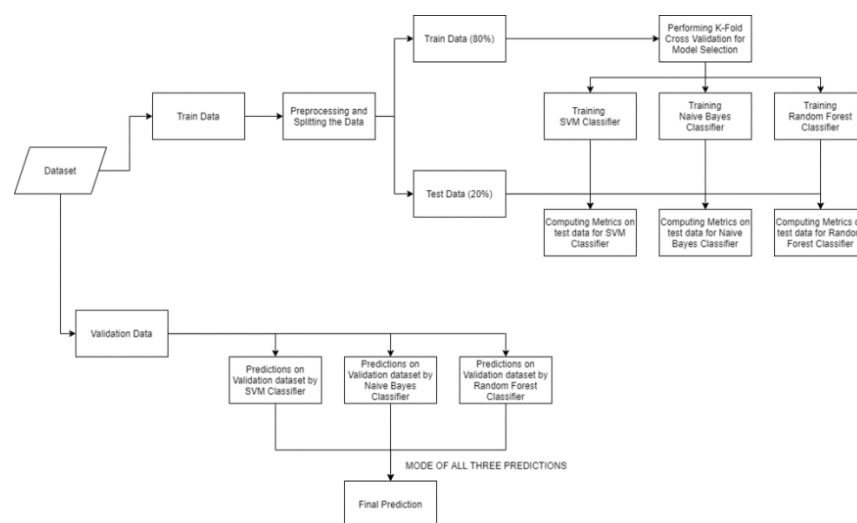
**Figure : 2.1 :- Work Flow for Implementation**

# CHAPTER 3

## NEED FOR MACHINE LEARNING

### What is Machine Learning?

Machine learning is a branch of <u>artificial intelligence (AI)</u> and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix's recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

### Benefits of Machine Learning

- Easily identifies trends and patterns

- No human intervention needed (automation)

- Handling multi-dimensional and multi-variety data

- Rapid Analysis Prediction and Processing

### What is Prediction in Machine Learning?

"Prediction" refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be.

Here are just a few examples of how predictive analytics is utilized in different industries:

### Banking and Financial Services

Predictive analytics in the banking and financial services industry are used in conjunction to detect and reduce fraud, measure market risk, and identify new business opportunities.

### Security

Predictive analytics and machine learning play a critical role in security. Security institutions typically use predictive analytics to improve services and performance, but also to detect anomalies, fraud, understand consumer behavior and enhance data security.

### Retail

Predictive analytics and machine learning in allow retailers to better understand consumer behavior, such as who will buy what and at what store? These questions can be readily answered with the right predictive models and data sets, helping retailers to plan ahead and stock items based on seasonality and consumer trends.

### Why are Predictions Important?

Predictions in machine learning allow businesses to make an accurate assumption as to the likely outcome of a question based on historical data. These predictions give businesses insights that result in tangible business value. For example, with churn, if a model predicts a customer is likely to churn, the business can target them with specific communications and outreach that can help prevent the loss of that customer.

# CHAPTER 4

# PROPOSED WORK

## How Prediction is Done

A prediction, in the context of machine learning, is an information output that comes from entering some data and running an algorithm.
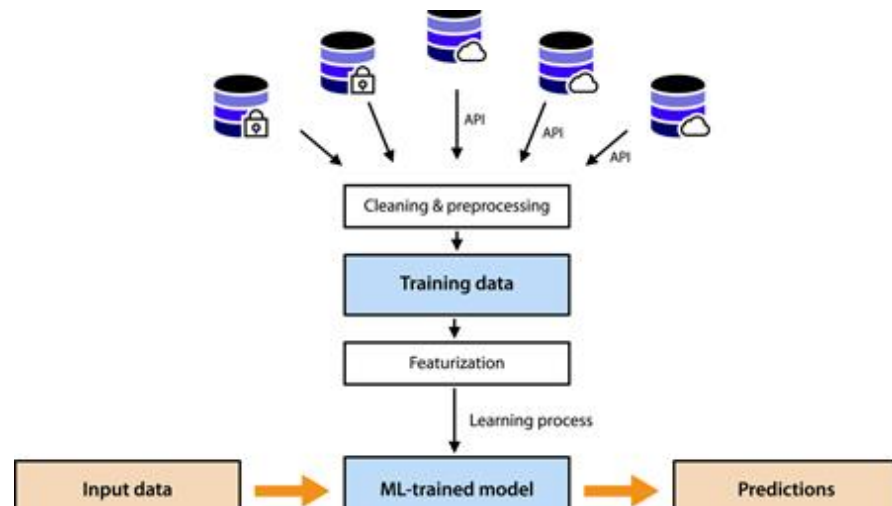


**Figure 4.2 : - Machine Learning Model**

The key challenge with any prediction process is that training data—the inputs you need in order to start getting reasonable outcomes—has to be either created (by, say, hiring experts to classify things) or procured from existing sources (say, health records). Some kinds of data are easy to acquire from public sources (think of weather and map information).

## Models

A machine learning model is defined as a mathematical representation of the output of the training process. Machine learning is the study of different algorithms that can improve automatically through experience & old data and build the model. A machine learning model is similar to computer software designed to recognize patterns or behaviors based on previous experience or data. The learning algorithm discovers patterns

within the training data, and it outputs an ML model which captures these patterns and makes predictions on new data.

Machine Learning models can be understood as a program that has been trained to find patterns within new data and make predictions. These models are represented as a mathematical function that takes requests in the form of input data, makes predictions on input data, and then provides an output in response. First, these models are trained over a set of data, and then they are provided an algorithm to reason over data, extract the pattern from feed data and learn from those data. Once these models get trained, they can be used to predict the unseen dataset



**Figure : 4.3 :- Models**

## Classifications of Machine Learning Models

Based on different business goals and data sets, there are three learning models for algorithms. Each machine learning algorithm settles into one of the three models:

o Supervised Learning
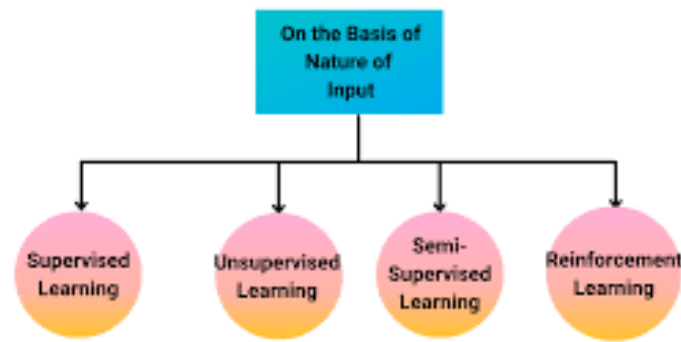
o Unsupervised Learning

o Reinforcement Learning

Figure 4.3 :- Classification of Models

## Models

## Decision Tree

o Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

o In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

o The decisions or the test are performed on the basis of features of the given dataset.
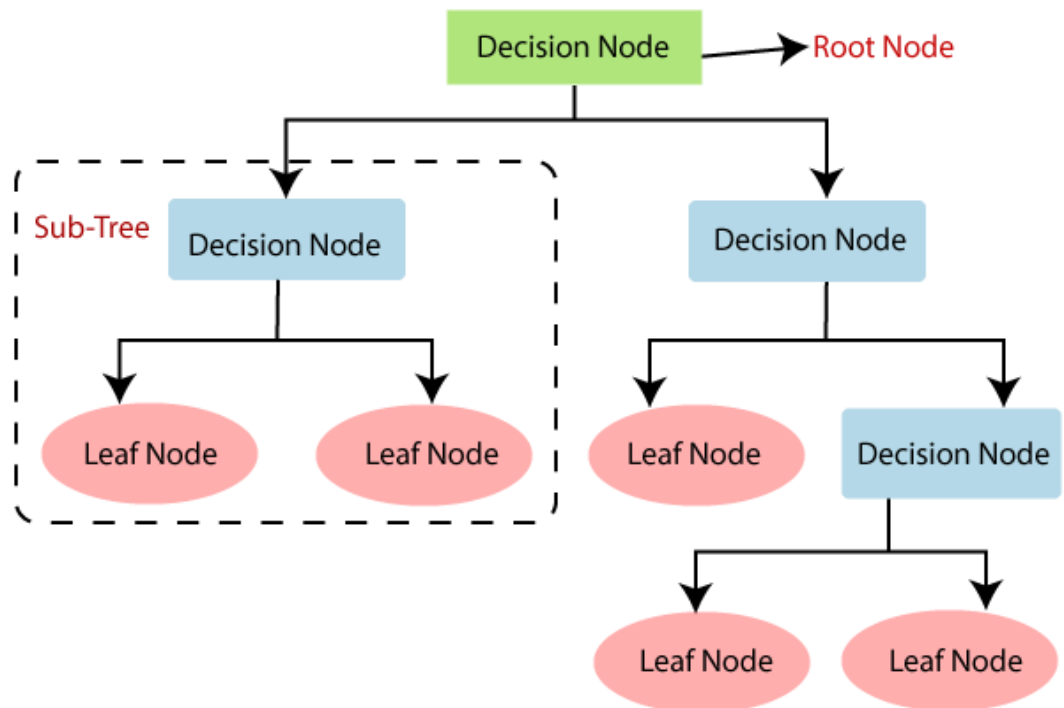
**Figure 4.4 : - Decision Tree**

## XGBoost

XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

It's vital to an understanding of XGBoost to first grasp the machine learning concepts and algorithms that XGBoost builds upon: supervised machine learning, decision trees, ensemble learning, and gradient boosting.

Supervised machine learning uses algorithms to train a model to find patterns in a dataset with labels and features and then uses the trained model to predict the labels on a new dataset's features.
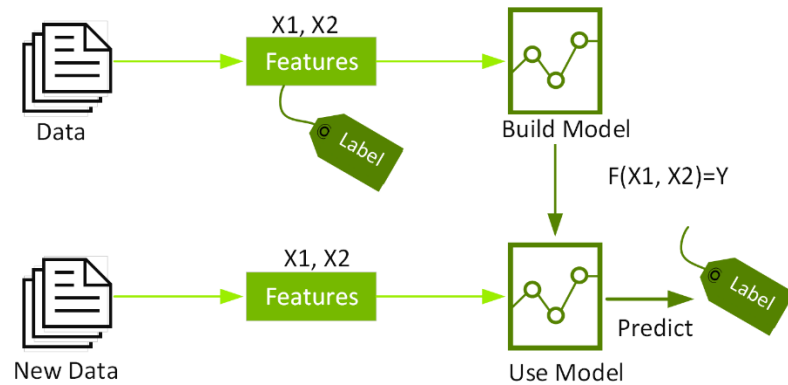
Figure : 4.5 :- XGBoost

# Random Forest

The random forest algorithm is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. Feature randomness, also known as feature bagging or "the random subspace method"generates a random subset of features, which ensures low correlation among decision trees. This is a key difference between decision trees and random forests. While decision trees consider all the possible feature splits, random forests only select a subset of those features.

Decision trees run the risk of overfitting as they tend to tightly fit all the samples within training data. However, when there's a robust number of decision trees in a random forest, the classifier won't overfit the model since the averaging of uncorrelated trees lowers the overall variance and prediction error.
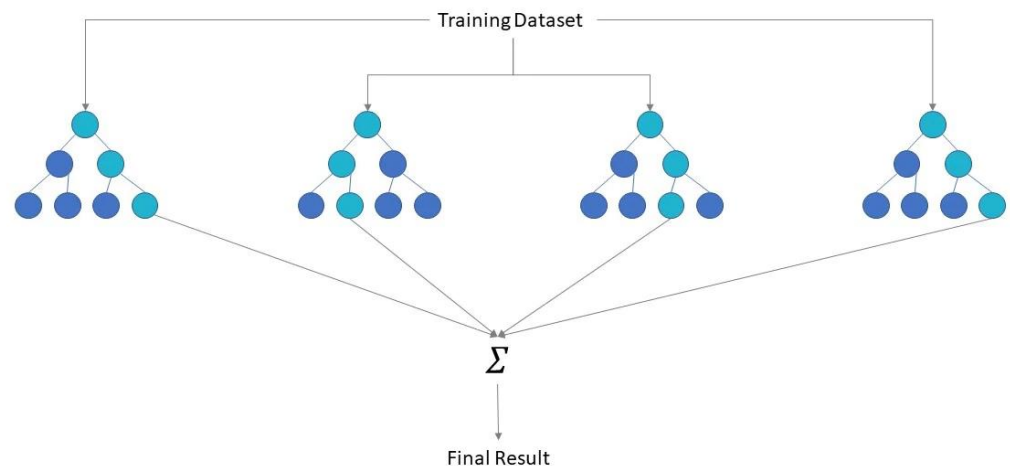


**Figure : 4.6 :- Random Forest**

# CHAPTER 5

## Conclusion

Our major projects deals with the prediction of risk a person has after treatment with that certain vaccine. Datasets and Machine Learning Models are used for this purpose. Different vaccines such as covi-shield and co-win were applied to applied to covid patients and others for treating covid or preventing it. But these vaccines comes with side effects on lung, liver and hearts.

To test our Model we will be creating our own dataset(dummy data) based on various medicines which were given to covid affected patients and others and taking notes the number of doses applied and keeping in mind their side effects. And to check our model's efficiency we will be testing it on datasets gathered from various organizations.

In this Major Project we will be engaging with the fields of data science, machine learning, data collection, data cleaning and data analysis. We will be collecting data, creating our data sets, cleaning them and then with the machine learning models, creating a predictive algorithm with highest accuracy possible to predict the risk a person carries due to covid vaccine or treatment he used.

# CHAPTER 6

## References

1. https://hbr.org/2020/09/how-to-win-with-machine-learning

2. https://lilianweng.github.io/posts/2017-08-01-interpretation/

3. https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a

4. https://www.tutorialspoint.com/machine_learning/machine_learning_implementing.htm

5. https://www.kaggle.com/datasets/

6. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov

7. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/expect/after.html