

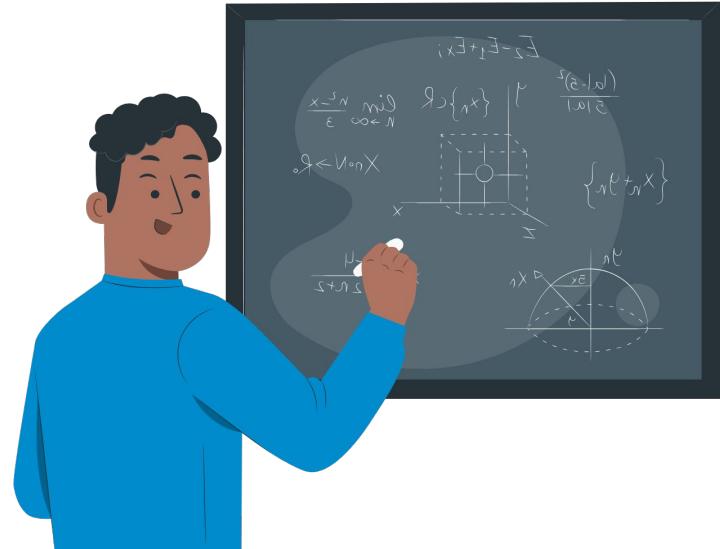


Introduction to Predictive Modeling

Introduction to Predictive Modeling

Predictive analytics is the process of using data, statistical algorithms, and machine learning techniques to analyze current and historical data to make predictions about future events or behaviors

How Does Predictive Analytics Work?



Step 1: Data Collection

Step 2: Data Cleaning

Step 3: Data Analysis

Step 4: Deployment

Step 5: Model Monitoring

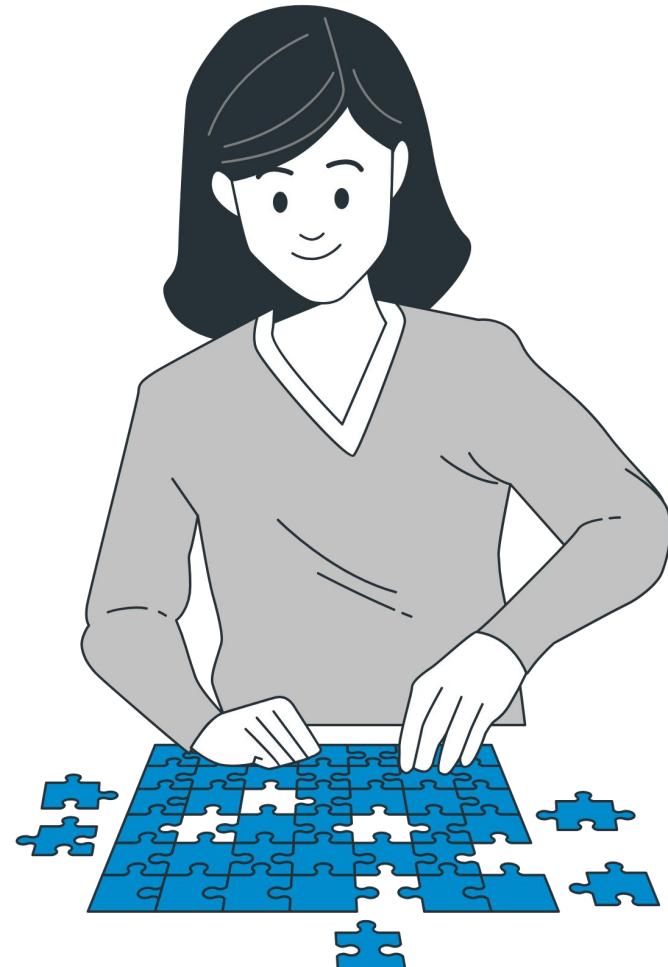
Applications of Predictive Analytics

- **Business:** Forecasting sales, optimizing marketing campaigns
- **Healthcare:** Predicting disease outbreaks, personalizing treatment plans
- **Finance:** Assessing credit risks, detecting fraudulent transactions
- **Sports:** Enhancing team performance, predicting injury risks

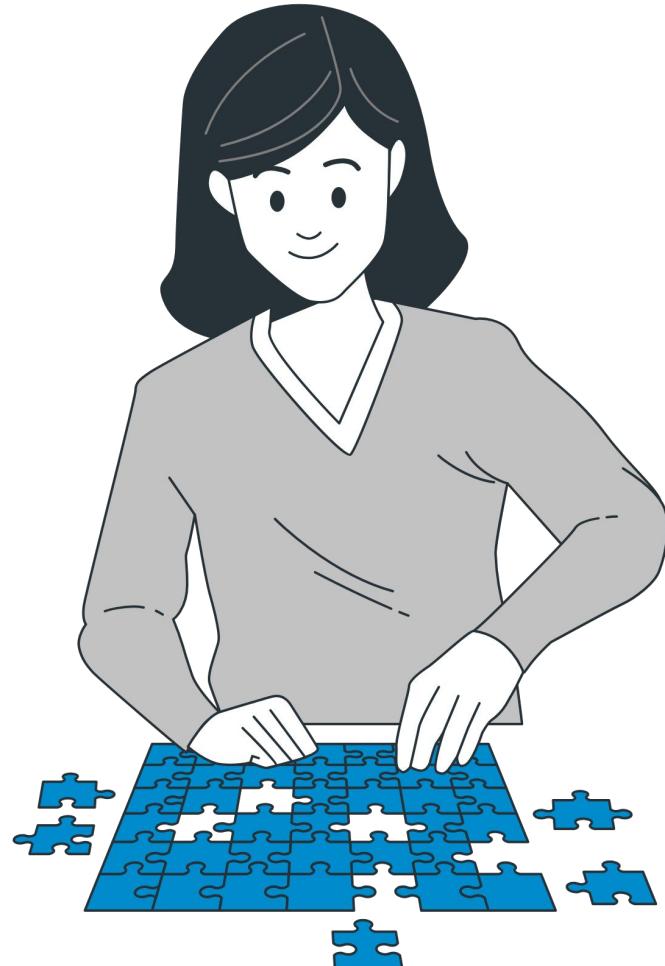
Benefits of Predictive Analytics

- Improved Decision Making: More data-driven and informed decisions
- Increased Efficiency: Streamline processes and reduce waste
- Enhanced Risk Management: Identify and mitigate potential risks early
- Competitive Advantage: Stay ahead by predicting trends before they happen

Types of Analytics

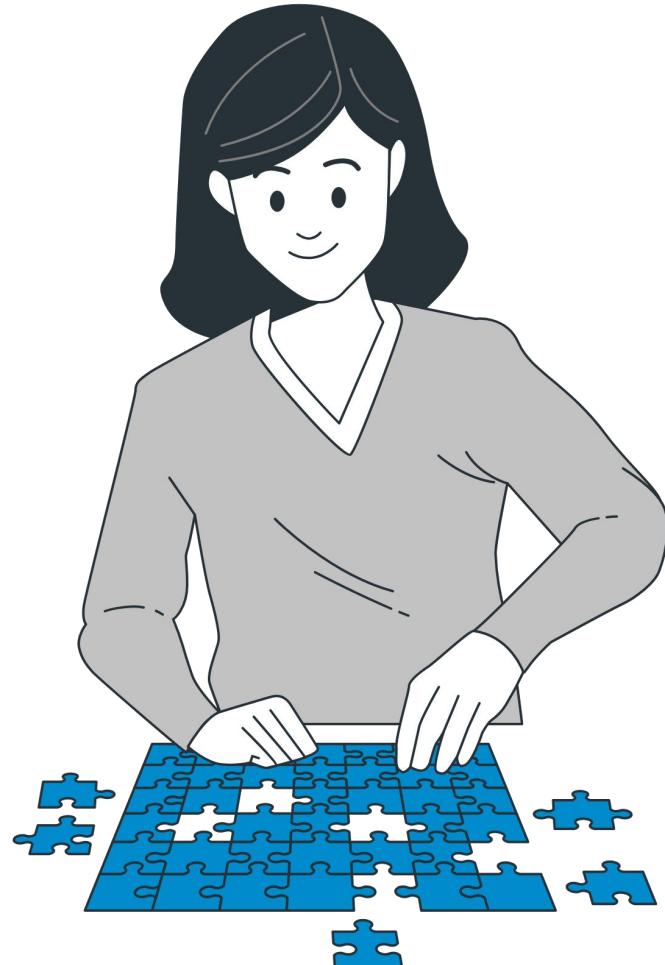


Descriptive Analytics



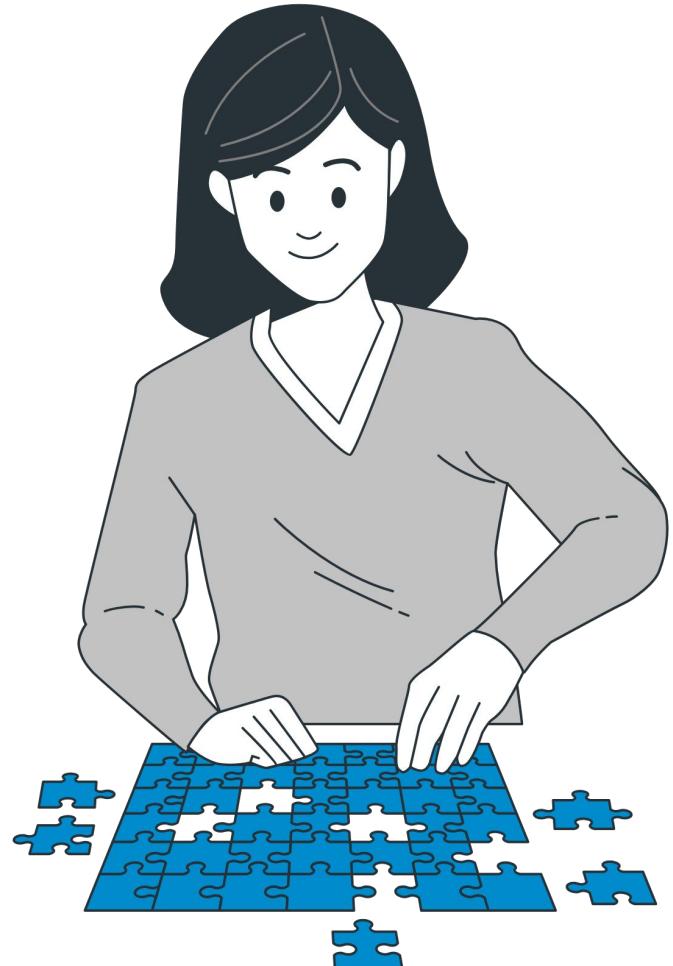
- Analyzes historical data to understand changes that have occurred in a business.
- To provide insight into past performance and understand the reasons behind success or failure.
- Basic Business Intelligence (BI), dashboards, and data visualization techniques.
- Example: Monthly sales reports that show historical sales trends.

Predictive Analytics



- Uses statistical models and forecast techniques to understand the future and predict outcomes based on historical data.
- To forecast future probabilities and trends.
- Statistical algorithms, machine learning models.
- Example: Predicting customer churn based on engagement metrics and historical behavior patterns.

Prescriptive Analytics



- Goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the implications of each decision.
- To advise on possible outcomes and recommend actions based on predictive and descriptive analytics.
- Simulation algorithms, optimization, and real-time decision-making tools.
- Example: A logistics company uses prescriptive analytics to determine the best routes and schedules for deliveries that minimize costs and maximize efficiency



How do we make Predictions?

How do we make Predictions?

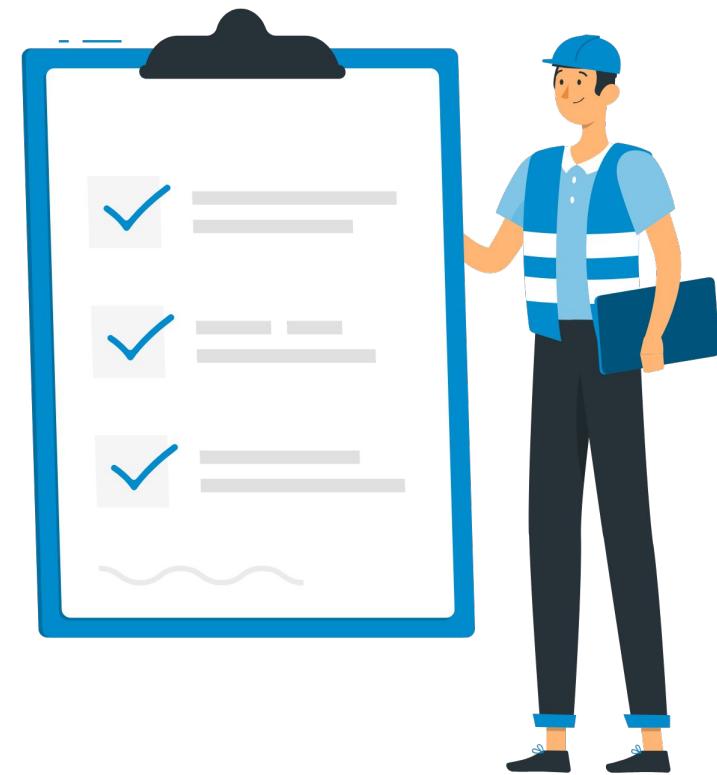


Benchmark Model

Selecting Benchmark Models

Some common ways to build a prediction model include:

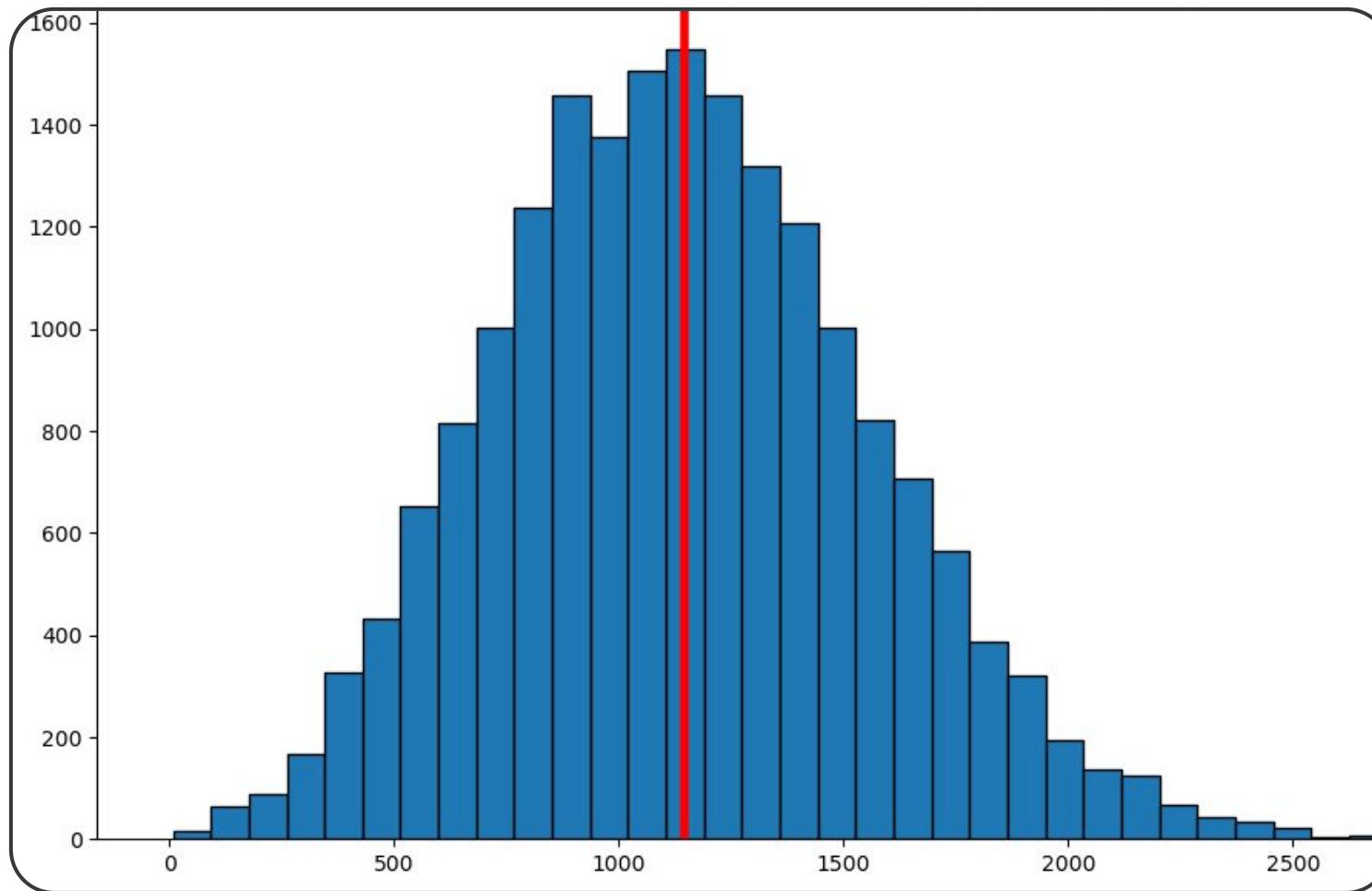
- Mean/Median Based Models
- Rule Based Models
- Machine Learning Models



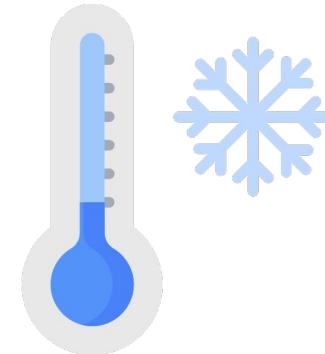
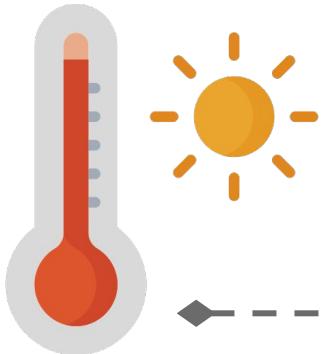


Mean/Median Based Models

Mean/Median



Mean/Median



Different median temperatures



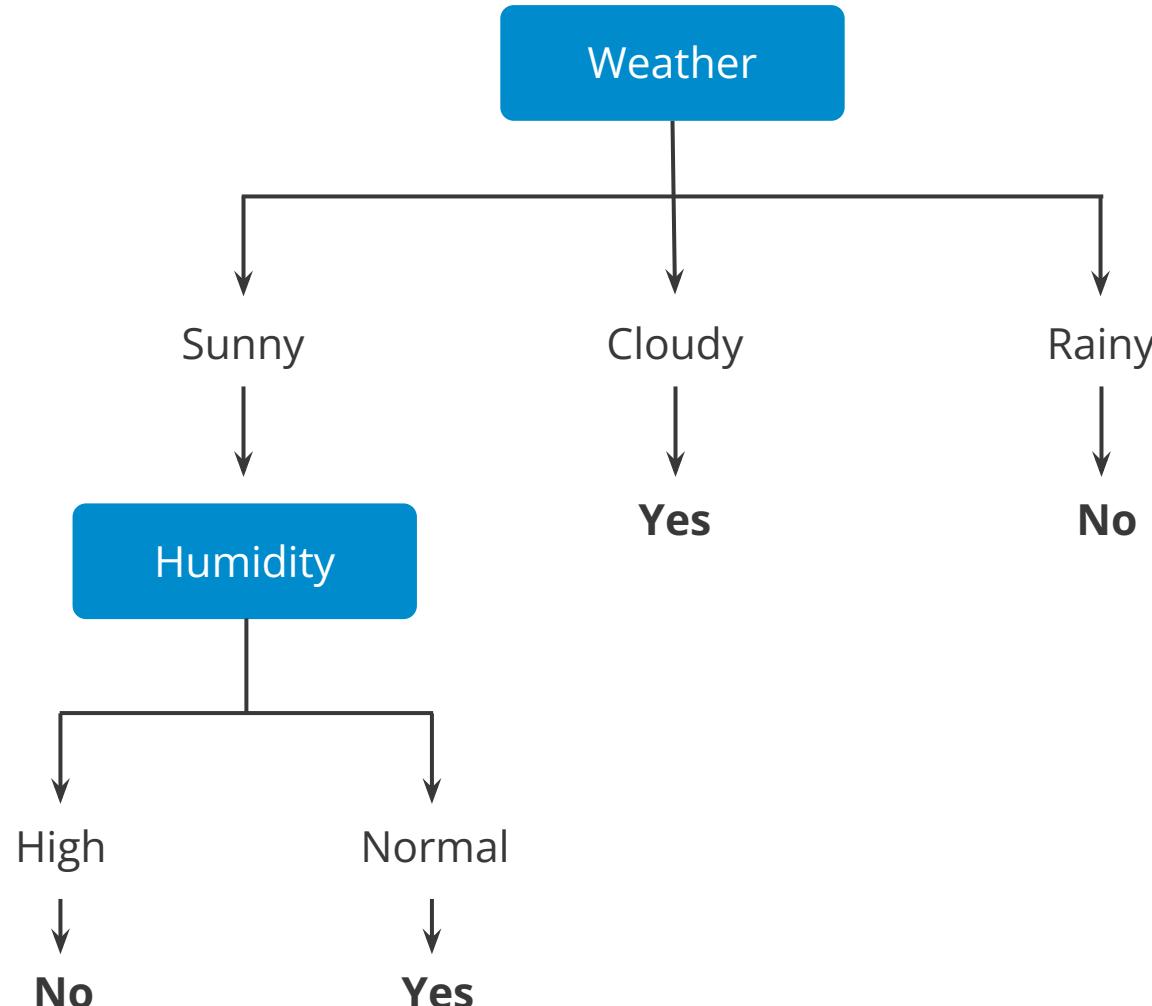
Rule Based Models

Rule Based Models

Creating a set of rules to make predictions based on detailed data analysis and business understanding.



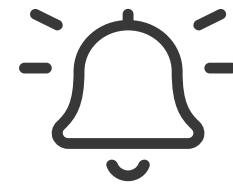
Rule Based Models



Machine Learning



Machine learning is an intelligent assistant that can learn from the data, identify hidden patterns, and automatically build complex rules.

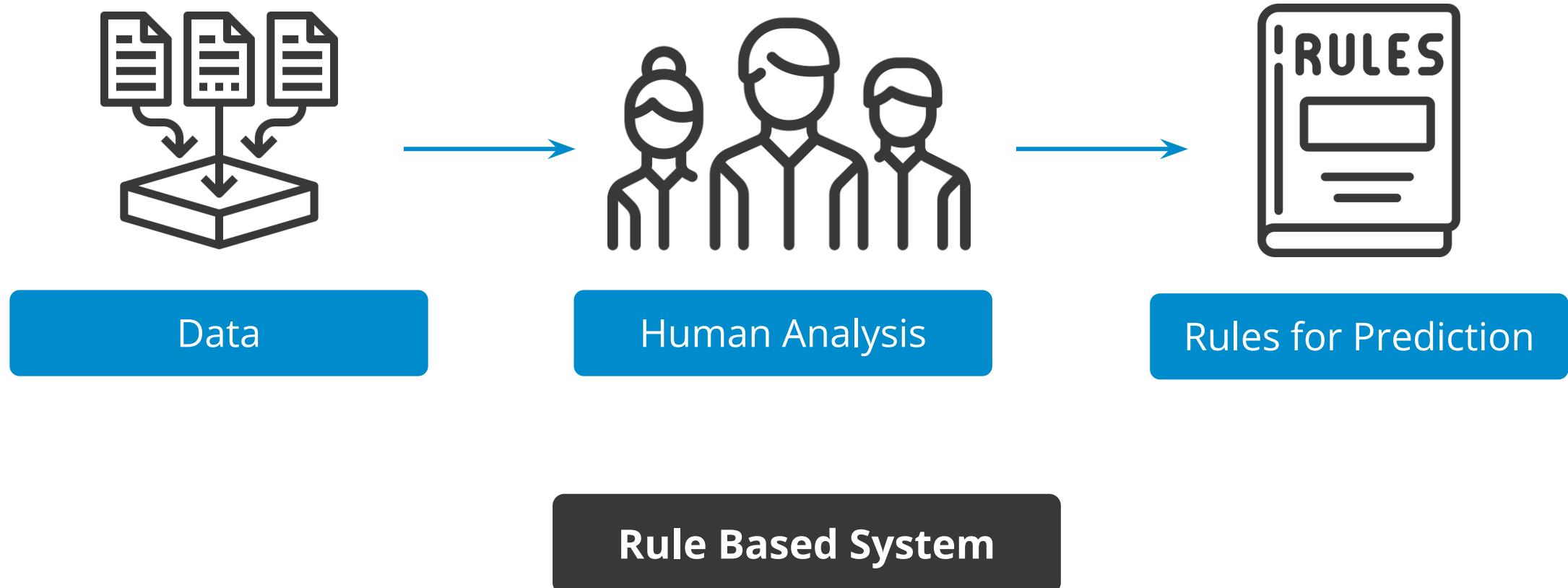


Machine Learning

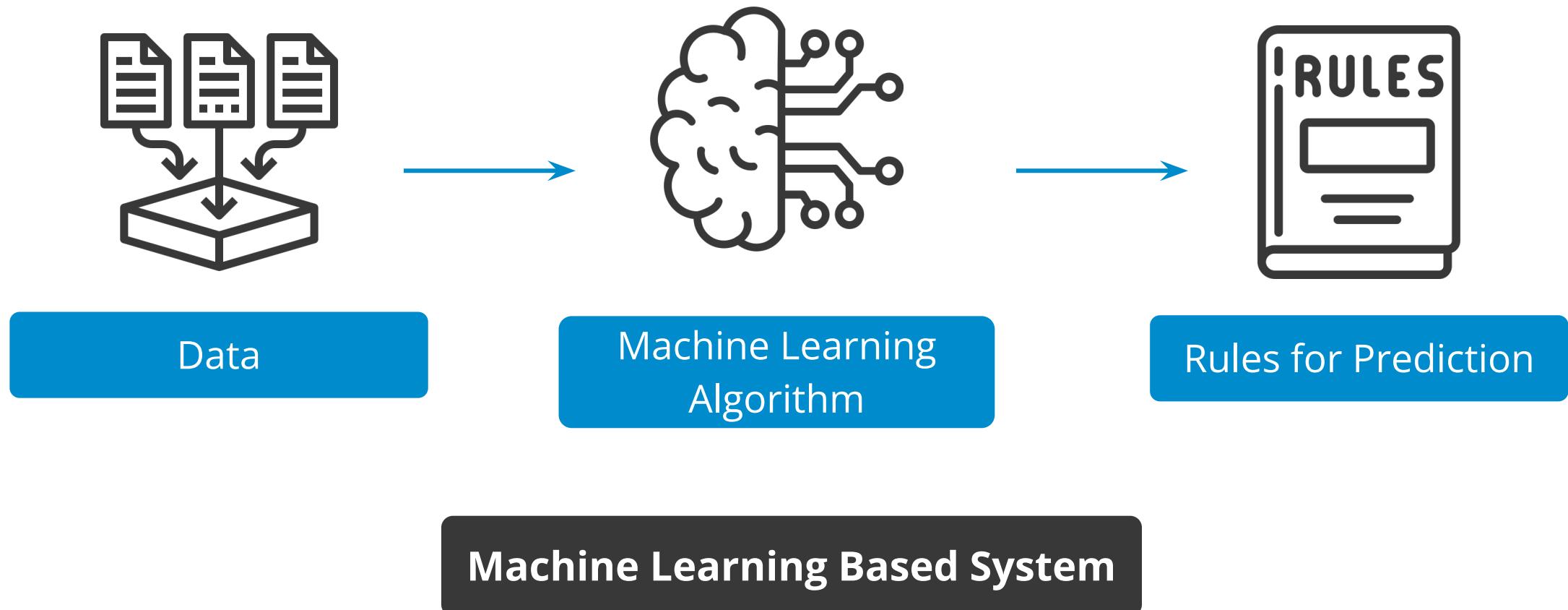


Machine Learning v/s Rule Based Predictions

Rule Based System v/s Machine Learning

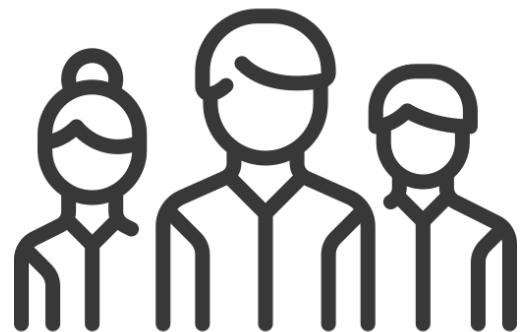


Rule Based System v/s Machine Learning



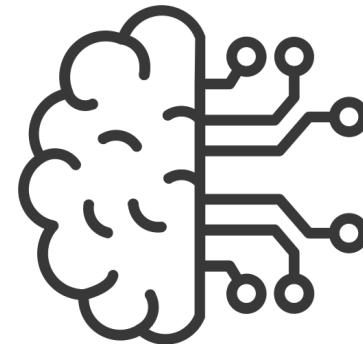
Rule Based System v/s Machine Learning

Simple Problems



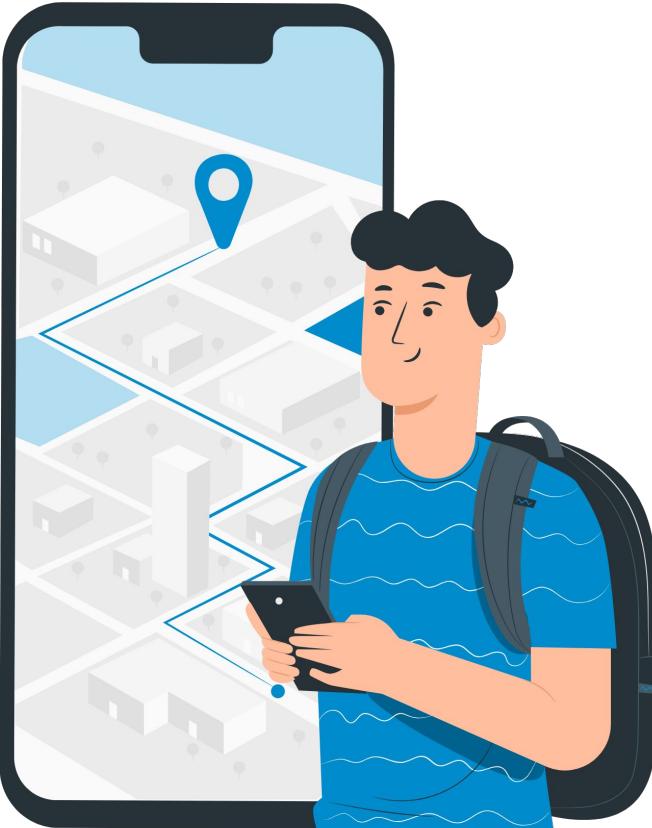
Humans can make good predictions

Complex Problems



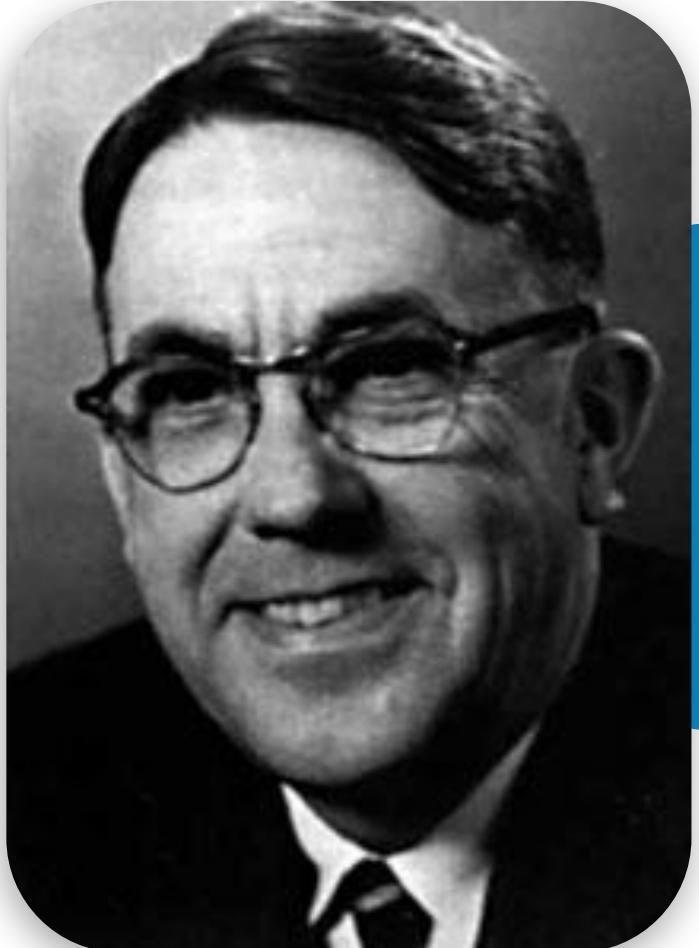
Machine Learning algorithms are more efficient

Rule Based System v/s Machine Learning



We generally use our memory to navigate familiar places but rely on GPS for longer or more complex routes.

Rule Based System v/s Machine Learning

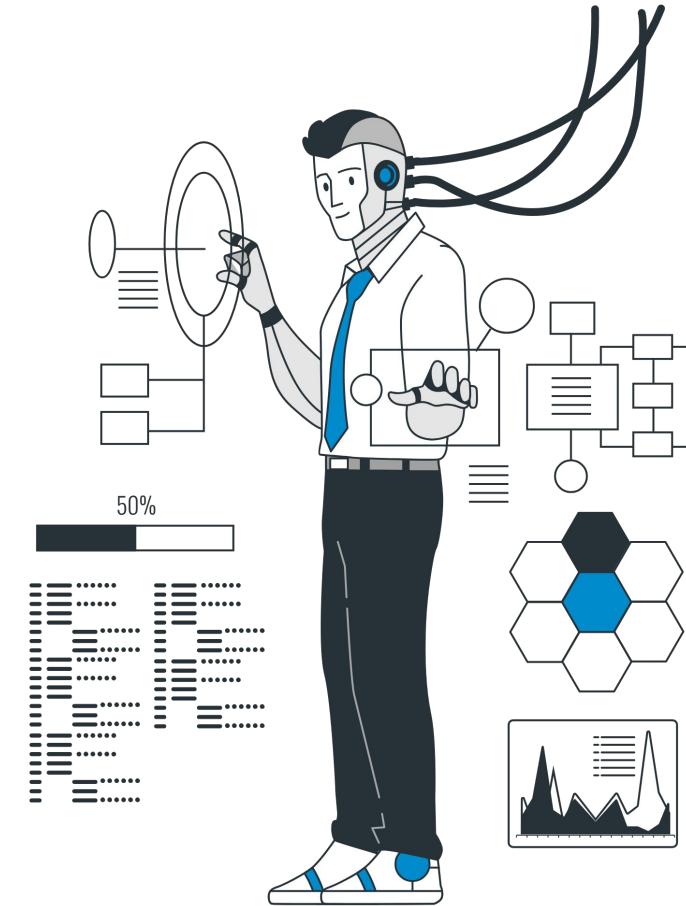


“Machine learning is a field of study, that gives computers the ability to learn without being explicitly programmed.”

- Arthur Samuel, 1959

Rule Based System v/s Machine Learning

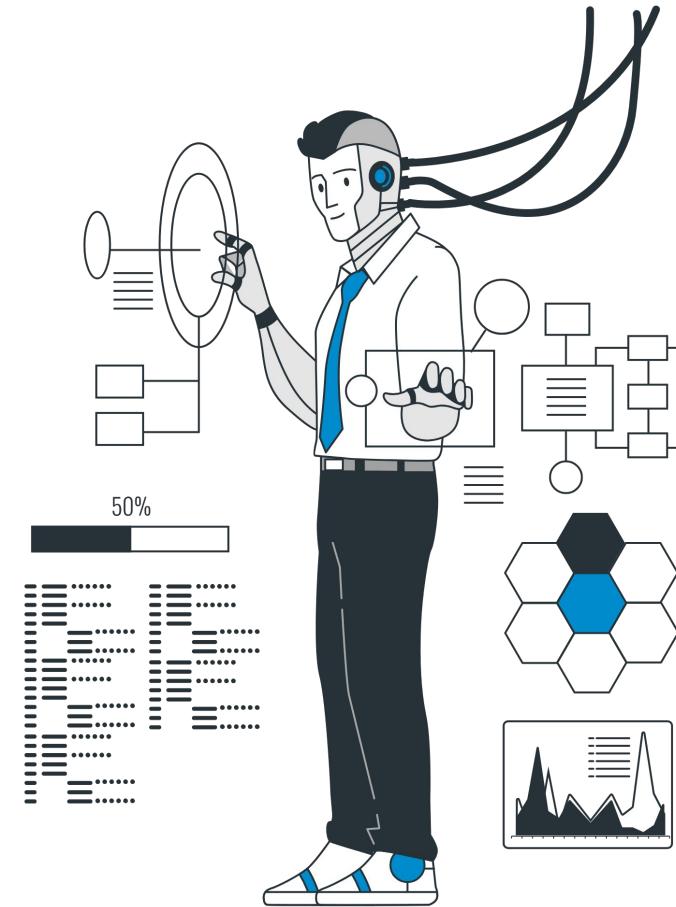
The Machine Learning algorithms learn the rules to make predictions on data by **learning from different instances of data** with being given any hard-coded guidelines.



Rule Based System v/s Machine Learning

The Machine Learning algorithms learn the rules to make predictions on data by **learning from different instances of data** with being given any hard-coded guidelines.

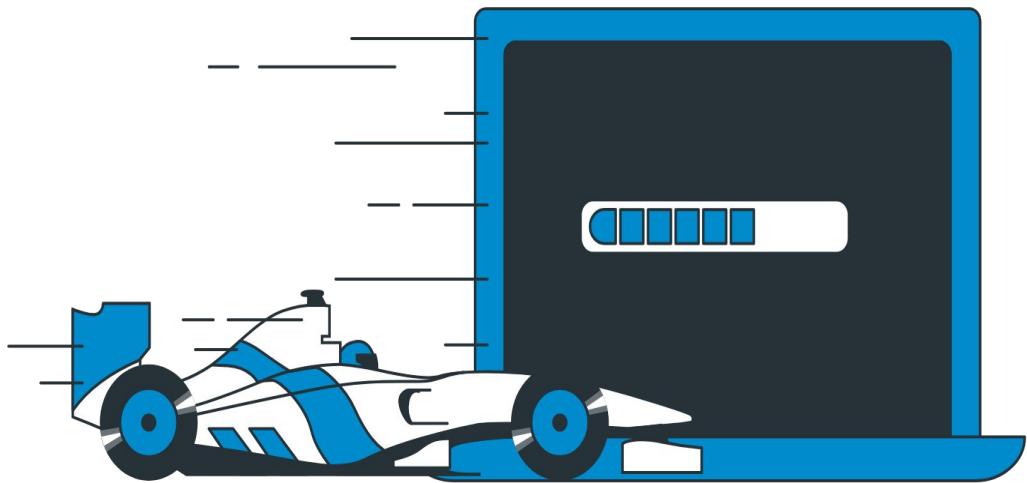
To reduce error in data provided



Advantages of Machine Learning

Adaptability and Speed

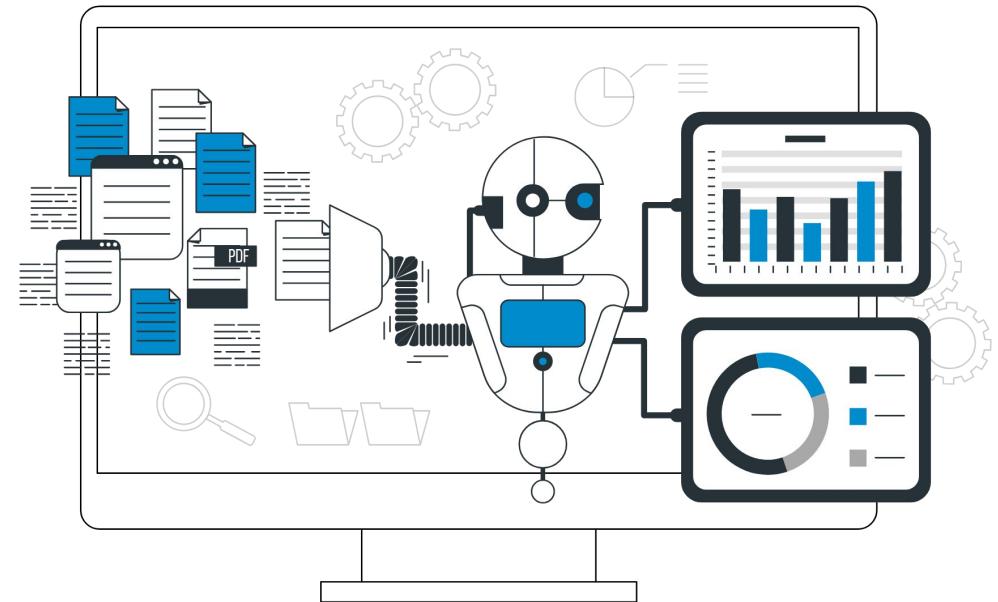
Machine learning algorithms learn quickly analyzing large amounts of data, while humans may struggle to with extensive datasets.



Advantages of Machine Learning

Scalability

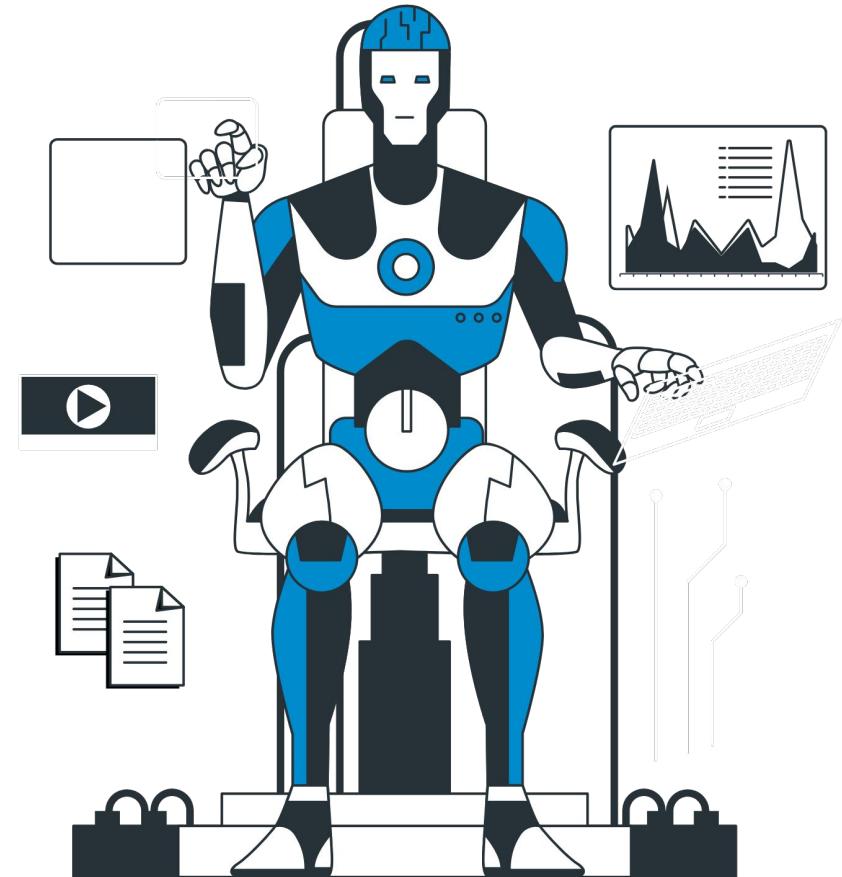
Machine Learning algorithms efficiently process and analyze the information as the dataset grows.



Advantages of Machine Learning

Consistency and Objectivity

Machine Learning algorithms do not tire while learning the rules for predictions or making predictions based on the rules and do not have any biases unlike humans.



Application of Machine Learning

5

Smart Navigation



Google Maps



Apple Maps

ML looks at real-time traffic data and predicts when you'll arrive. It can even suggest alternate routes to avoid traffic.

Application of Machine Learning

7

Finance

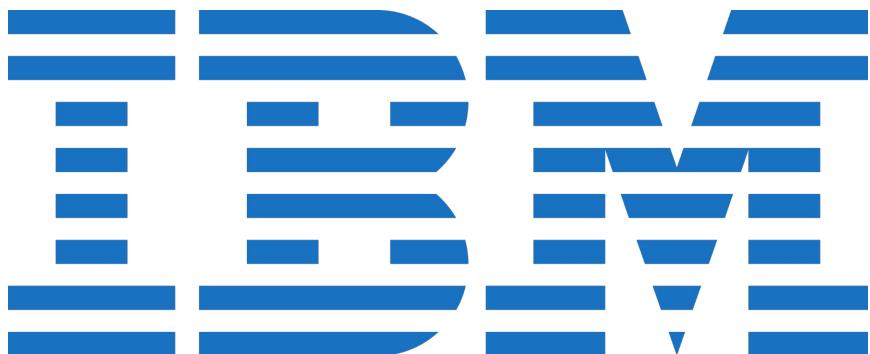


ML is used to detect suspicious activity, fraud, manage risk and predict creditworthiness of a customer.

Application of Machine Learning

8

Marketing



ML is used to analyze the customer data to improve marketing

Application of Machine Learning

12

Education



Khan Academy



Growth
School

ML is used to personalize learning experiences by assessing students' strengths and weaknesses.

Application of Machine Learning

13

Agriculture



Farmers

ML is used to optimizes crop yields and reduces waste. It predicts crop health and recommends actions.

Application of Machine Learning

14

E-commerce

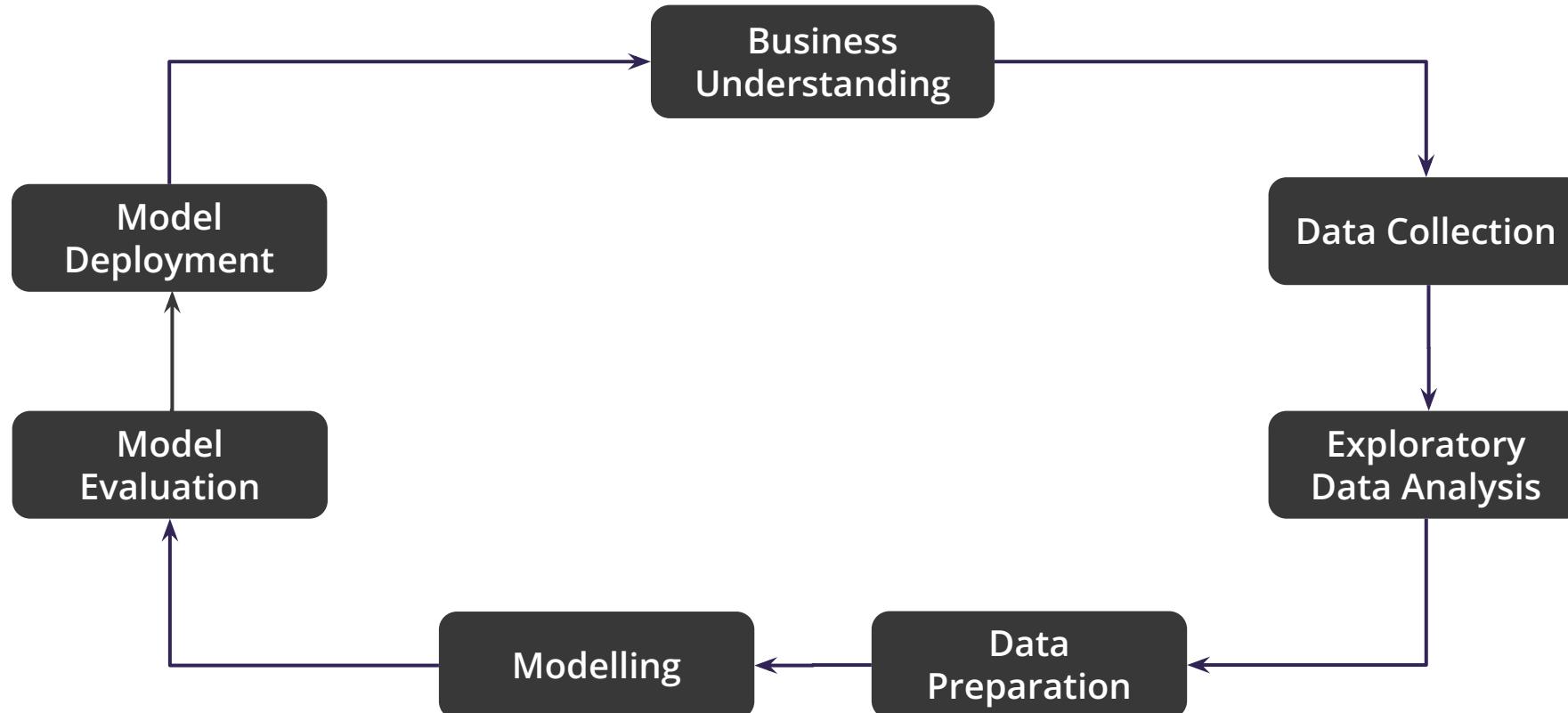


ML is used to enhance customer experiences and streamlines operations to personalize suggestions.



Machine Learning Workflow

Machine Learning Workflow



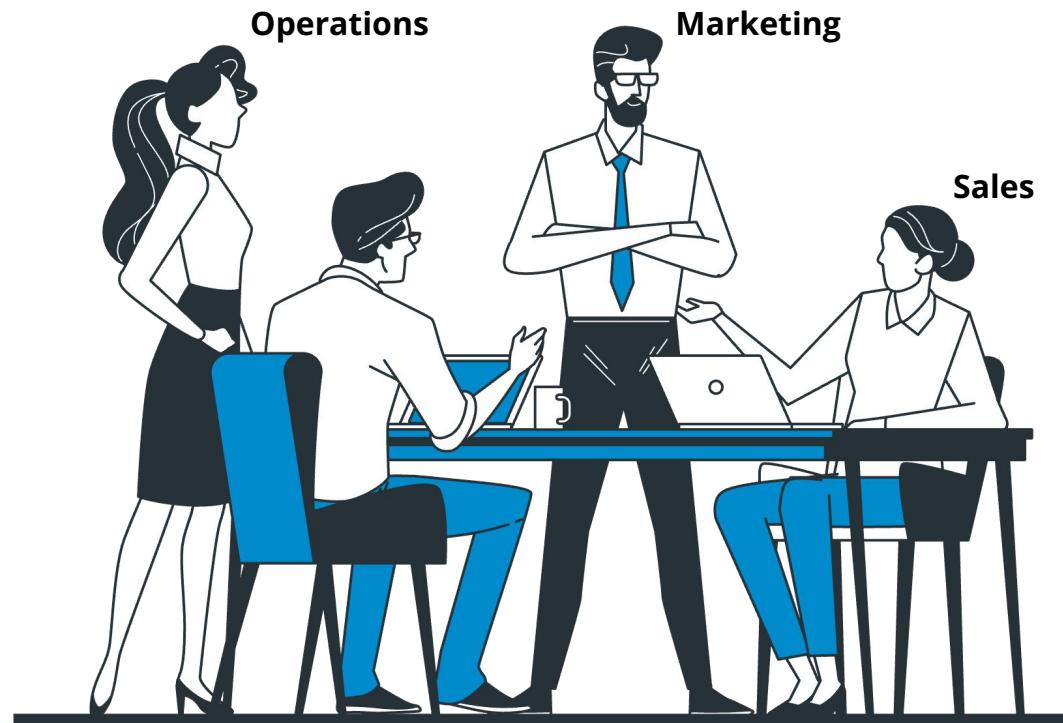
Business Understanding



Problem Statements

1. Identify the factors to predict units sold for products
2. Identify which products will sell more than 1000 units
3. Create a new-age bundled online marketing strategy

Data Collection



Stakeholder Meeting

Understand the factors and map required data to solve the problem

Data Collection

POS_Data

Online_Data

Offline_Data

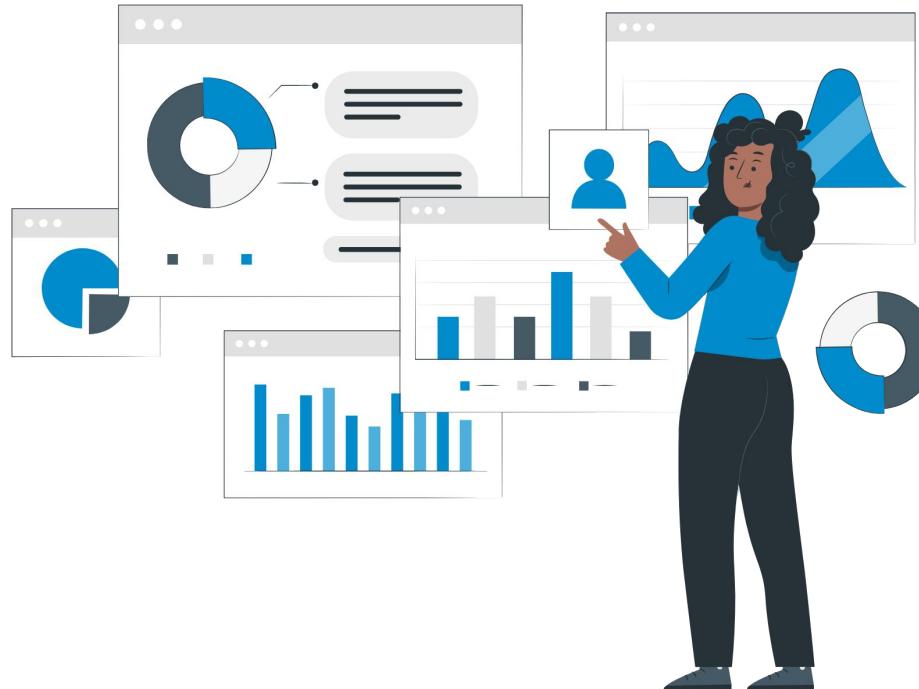
Product_Attribute_Data

VPC_Data

Search_Rank_Data

Synergix Solutions has collected data in 6 different tables.
Let's take a look at what each of these tables contain.

Exploratory Data Analysis



In EDA, data is **analyzed** and **visualized** using various techniques by the user to obtain relevant insights.

Exploratory Data Analysis

1

Generate relevant insights for understanding



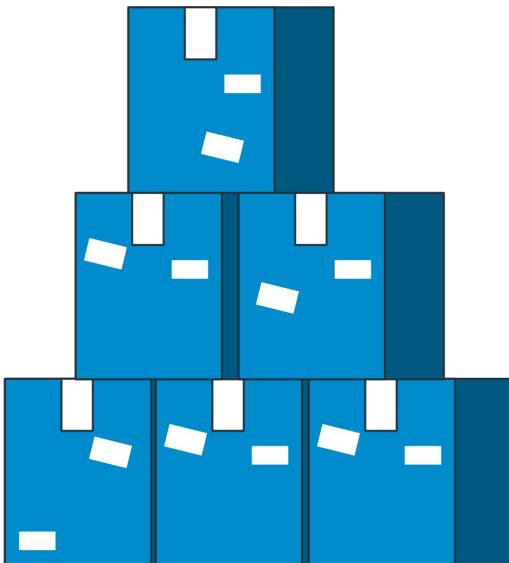
Generate insights on:

- A) What has been happening?
- B) What is working?
- C) What needs improvement

Exploratory Data Analysis

2

Identify important features for problem solving



In our case:

To find which factors impact the Units_sold.

Data Preparation



This step ensures that the data is accurate, consistent and ready for modelling

Data Preparation

1

Data Division

2

Data Cleaning

3

Data Restructuring

4

Data Transformation

5

Feature Engineering

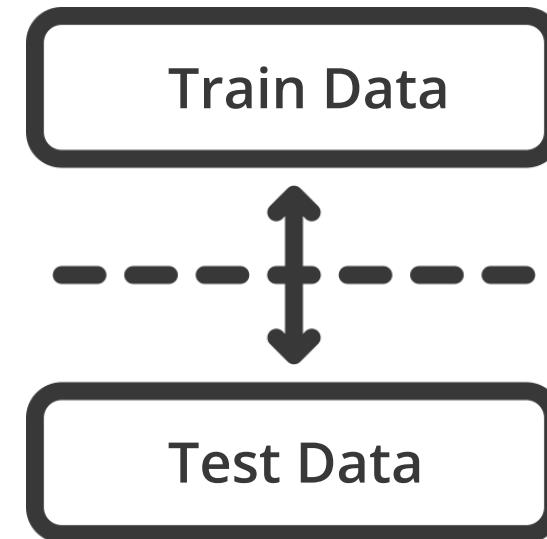
Data Preparation

1

Data Division



This process improves the reliability of the model



Data Preparation

2

Data Cleaning



Removing invalid values



Imputing missing values



Treating outliers



Example: Imputing missing values
for Page_traffic in POS_Data with
mean of overall traffic

Data Preparation

3

Data Transformation



Converting data from one format to another



Example: Converting the segment column in pos_data into integers

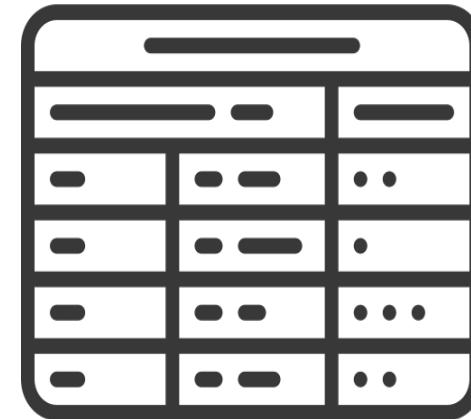
Data Preparation

4

Data Restructuring



Combining data from multiple sources to easily analyze and work with it



Example: Combine multiple tables for Synergix Solutions to analyze it easily.

Data Preparation

5

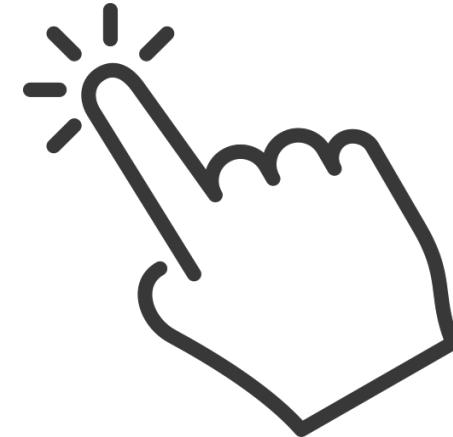
Feature Engineering



Create new features by combining existing ones



Requires domain knowledge



Example: Create a new feature called click through rate

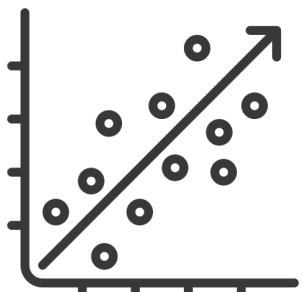
Modelling



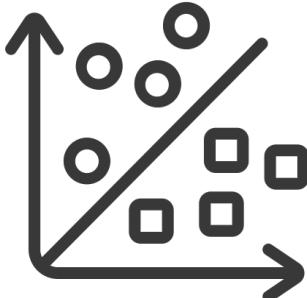
Try different ML algorithms and train the model using the prepared dataset

Modelling

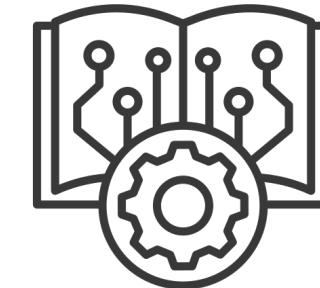
3 Problem Statements



Regression Model

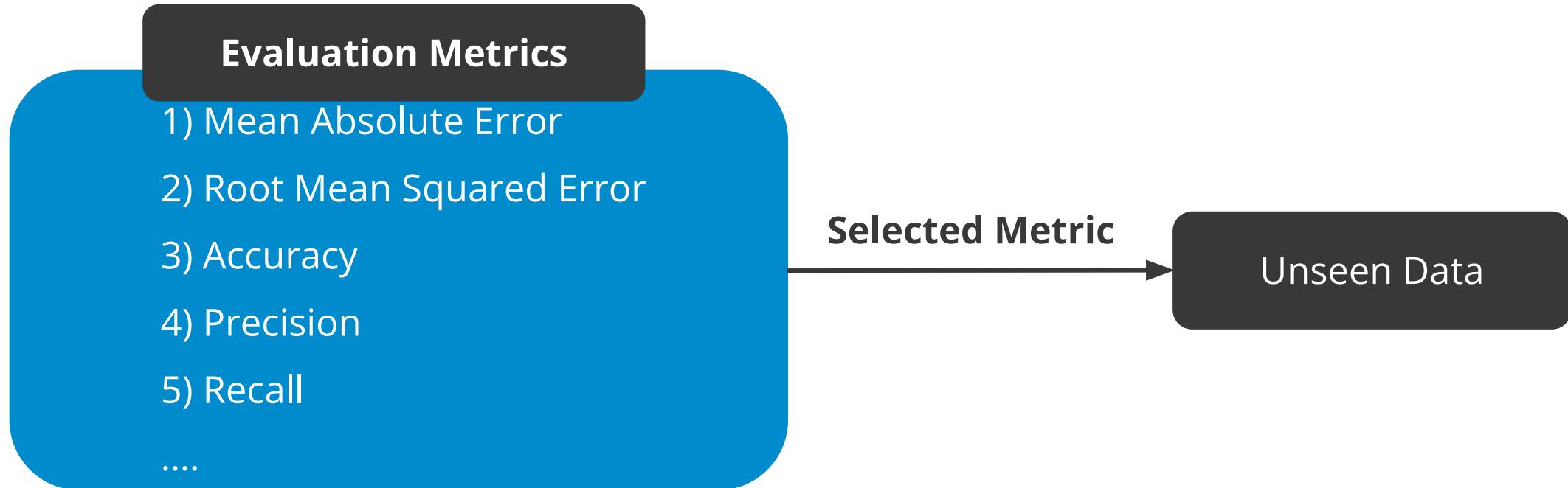


Classification Model

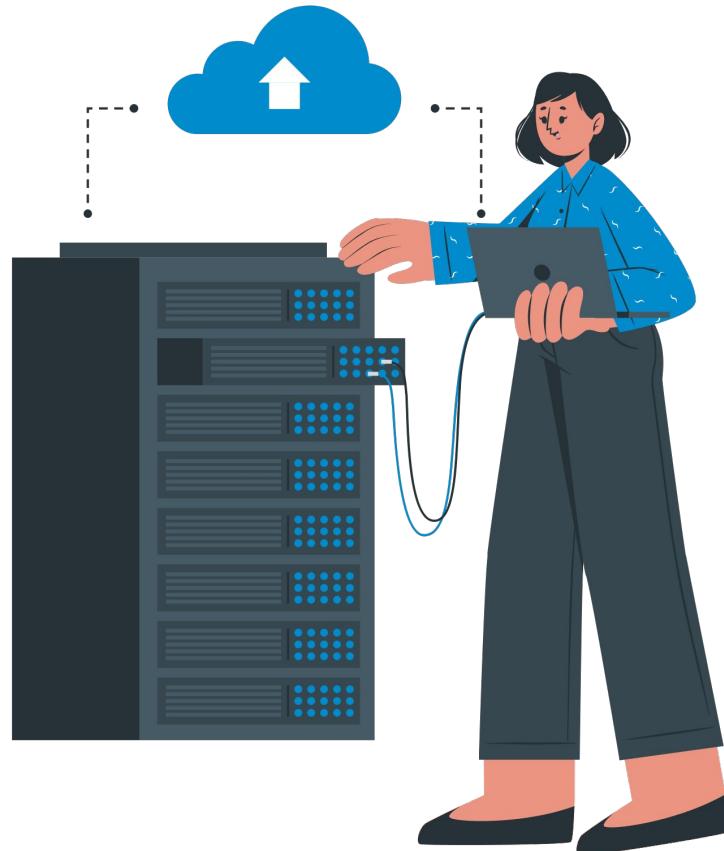


Unsupervised ML Model

Machine Learning Workflow



Model Deployment



In this step the trained and evaluated model is deployed or integrated into the eco-system.