

# DATA SCIENCE : INTRODUCTION

Data science is a field that combines domain expertise, programming skills, and statistical knowledge to extract meaningful insights and knowledge from data. It's about transforming raw data into actionable information that can drive better decision-making.

## APPLICATION:

- Healthcare: Analyzing patient data to predict disease outbreaks, personalize treatment plans, and improve patient outcomes
- Finance: Detecting fraud, assessing risk, and developing personalized financial products
- E-commerce: Personalizing product recommendations, optimizing pricing strategies, and improving inventory management
- Technology: Developing and improving AI algorithms, optimizing search engine results, and enhancing user interfaces.

## WORKFLOW

A typical data science workflow involves several stages, including:

- 1. Problem Definition:** Clearly defining the problem that needs to be solved and formulating relevant questions.
- 2. Data Collection:** Gathering data from various sources, both structured and unstructured.
- 3. Data Cleaning and Preparation:** Cleaning, transforming, and preparing the data for analysis by handling missing values, outliers, and inconsistencies.
- 4. Exploratory Data Analysis (EDA):** Exploring the data to understand its characteristics, identify patterns, and generate hypotheses.
- 5. Model Building:** Selecting and training appropriate machine learning models to address the defined problem.
- 6. Model Evaluation:** Evaluating the performance of the trained model using appropriate metrics and techniques.
- 7. Model Deployment:** Deploying the chosen model to a production environment and integrating it into a system or application.

## SKILLS REQUIRED :

- 1. Programming:** Proficiency in languages like Python, R, and SQL is essential for data manipulation, analysis, and model building.
- 2. Mathematics and Statistics:** Strong understanding of statistical concepts, probability, linear algebra, and calculus is crucial for model development and interpretation.

**3.Domain Expertise:** Understanding the specific industry or domain where the data science project is being conducted

**4.Phython library:** learn to use numpy , pandas for data manipulation and saeborn for data visulization.

**MUTABLE / IMMUTABLE**