**Task – 1 : Data Science Applications Brainstorm**

**Shopping**

Data Science plays a crucial role in gathering the purchase data from the customer and helps to analyse the frequently or most purchased product that the customer likes and helps to improve customer satisfaction

**Education**

Helps to store students data in college to track and allocate the students for the needed fields like placements, higher studies etc, Helps to analyse the student’s capability by storing his activity as data.

**Sports**

Helps to store the players data to analyse and categorize players according to the matches to ensure the strength and winning rate and also helps to improve their practice and precision.

**Medical**

Helps to store and predict the disease and problems in the patients and helps to track their health continuously.

**Military**

Data science is very essential in the field of Military to store all kind of data and maintain threats in a controlled manner.

**Task -2 : Create a visual diagram illustrating the relationship between AI, Machine Learning, and Deep Learning. Include a short explanation.**

Artificial intelligence (AI) is a technology that allows computers to perform tasks that are usually done by humans.

Deep learning is a machine learning technique that uses artificial neural networks to learn from data.

Machine learning (ML) is a type of artificial intelligence (AI) that allows computers to learn and improve without needing to be explicitly programmed.

DL

ML

AI

**TASK -3 : Data Scientist Roles Research**

* **Data Collection and Cleaning:**

This involves gathering data from various sources, cleaning it from inconsistencies or errors, and preparing it for analysis.

* **Data Analysis and Modeling:**

Applying statistical techniques and machine learning algorithms to identify patterns, trends, and relationships within the data to build predictive models.

* **Communication and Visualization:**

Presenting findings and insights from the data analysis through clear visualizations like charts and graphs, and communicating them effectively to stakeholders to inform

**Source :** <https://graduate.northeastern.edu/knowledge-hub/what-does-a-data-scientist-do/>

**TASK -4 : Data Science Lifecycle Example**

* **Problem Definition :**

Super Market’s Main goal is to attract customers and to improve their marketing sale. The problem here is to optimize its inventory and improve customer satisfaction by personalizing recommendations for customers. The goal could be to predict which products a customer is likely to buy based on their previous shopping behavior.

* **Data Collection**
* **Sources of Data :**

**Transaction Data**: This includes data on items purchased, quantities, prices, and the total amount spent by customers.

**Customer Data**: Demographic data (age, gender, location, etc.), loyalty program membership, and browsing history.

**Product Data**: Details about products, such as categories, shelf locations, stock levels, and promotions.

**External Data**: Weather data, local events, and holidays that can influence buying behavior.

* **Data Cleaning & Preprocessing :**

 **Handling Missing Values**: Filling in or removing missing data points from transaction records or customer profiles.

 **Data Transformation**: Converting categorical data (e.g., product categories) into numerical format (e.g., one-hot encoding).

 **Outlier Detection**: Identifying and handling outliers (e.g., abnormally large transaction amounts).

 **Feature Engineering**: Creating new variables that might be useful, such as calculating customer lifetime value (CLV) or customer segmentation based on purchase frequency.

 **Normalization**: Scaling numerical data (e.g., sales data) for machine learning models to improve accuracy.

* **Model Training & Evaluation**

\* Split data into training and testing sets.

\* Train models on the training set (e.g., using product purchase history or customer demographics).

**For Recommendations**: Precision, recall, or mean squared error (MSE).

**For Segmentation**: Cluster purity, or customer retention rates.

* **Deployment & Monitoring**

**Recommendation System**: Integrate the recommendation model into the supermarket’s website or mobile app so customers receive personalized product suggestions.

**Model Performance**: Continuously monitor the model’s performance (e.g., accuracy of product recommendations or sales forecasts) and retrain the model periodically as new data is collected.

**Task -5 : Data Science Ethics Discussion**

One big ethical issue in data science is **bias in algorithms**. This happens when an algorithm makes unfair decisions because it's trained on biased data. For example, if an algorithm used for job hiring is trained on past hiring data that favored one gender or race, it might keep making those same biased decisions, even though we want the process to be fair. This is important because it can cause real harm by discriminating against certain groups of people, even if the algorithm isn't intentionally trying to do that. It’s really important for data scientists to make sure their models are fair and don’t unintentionally make things worse for certain people.

**Bonus : Data Science Career Paths**

**Career Paths :**

* Data Analyst
* Data Scientist
* Machine Learning Engineer
* Data Engineer
* Data architect
* Business Intelligence Analyst

**Career Paths :** Data analyst, Machine Learning Engineer