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

1. **Shopping**Data Science enables retailers to collect and analyze customer purchase patterns, helping them recommend personalized products, optimize inventory management, and enhance customer satisfaction through targeted marketing strategies.
2. **Education**Educational institutions leverage Data Science to track student performance, predict academic success, and allocate resources effectively. It aids in personalized learning experiences, career guidance, and placement assistance based on student data analysis.
3. **Sports**Data Science is instrumental in evaluating player performance, injury prevention, and game strategy optimization. By analyzing real-time match statistics and training data, teams can improve their chances of success and refine individual player skills.
4. **Healthcare**In the medical field, Data Science helps in disease prediction, patient monitoring, and treatment optimization. AI-driven models assist in diagnosing illnesses early, recommending treatments, and improving overall healthcare efficiency.
5. **Military**Data Science plays a crucial role in defense by analyzing security threats, optimizing resource allocation, and predicting potential risks. It aids in intelligence gathering, cybersecurity, and strategic decision-making for national security.

**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.

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

**ARTIFICIAL INTELLIGENCE**

**MACHINE LEARNING**

**DEEP**

**LEARNING**

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

**TASK -3 : Roles of a Data Scientist**

1. **Data Acquisition and Preprocessing:**Data Scientists collect data from multiple sources, ensuring its quality by cleaning and transforming it to remove inconsistencies, missing values, and errors, making it ready for analysis**.**
2. **Exploratory Data Analysis and Modeling:**They apply statistical techniques and machine learning algorithms to uncover hidden patterns, trends, and correlations. This helps in developing predictive models that drive data-driven decision-making**.**
3. **Insight Communication and Visualization:**Using visual tools like charts, graphs, and dashboards, Data Scientists present their findings in an easily understandable way. Effective storytelling ensures stakeholders can make informed business decisions**.**

**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**

**Data Science Ethics Discussion**

Data Science has transformed industries by enabling data-driven decision-making, but ethical considerations are crucial to ensure responsible use. Key ethical concerns include:

1. **Privacy and Data Security**  
   Handling sensitive user data requires strict security measures to prevent breaches and unauthorized access. Organizations must comply with data protection regulations like GDPR and CCPA.
2. **Bias and Fairness**  
   Machine learning models can inherit biases from training data, leading to unfair outcomes. Ethical Data Science involves identifying and mitigating bias to ensure fairness across different demographic groups.
3. **Transparency and Accountability**  
   AI models should be explainable and transparent, allowing users to understand how decisions are made. Organizations must take responsibility for errors or unintended consequences of their algorithms.
4. **Informed Consent and Data Ownership**  
   Users should have control over their data and be informed about how it is collected, stored, and used. Ethical practices include obtaining proper consent and respecting data ownership rights.
5. **Social Impact and Responsibility**  
   Data Science applications should aim for positive societal impact, avoiding harm or misuse, such as surveillance abuse, misinformation, or unethical business practices.

**Bonus : Data Science Career Paths**

Data Analyst – Focuses on data interpretation, reporting, and visualization (SQL, Excel, Tableau).

Data Scientist – Builds predictive models using machine learning and statistics (Python, R, AI).

Machine Learning Engineer – Develops and deploys AI models for production (TensorFlow, PyTorch).

Data Engineer – Manages large-scale data pipelines and infrastructure (SQL, Hadoop, Cloud).

BI Developer – Creates dashboards and business insights (Power BI, Tableau, Data Warehousing).

AI Research Scientist – Works on advanced AI and deep learning research (Neural Networks, Reinforcement Learning).

Data Privacy & Ethics Consultant – Ensures responsible AI practices and data security (GDPR, Ethical AI).