Module :1 – introduction to data science

**Task 1 :: brief report on the evolution of data science**

Introduction :

The term “Data Science” was created in the early 1960s to describe a new profession that would support the understanding and interpretation of the large amounts of data which was being amassed at the time.

**— What is data science ?**

Data science is the process of studying and analyzing past data and giving predictions and insights related to specific data.

**i) early stages :**

In early 1960s , data analytics began with the development of statistics in the 18th and 19th centuries.

People used paper based records and manual calculations as tools for data analysis.

**ii) business intelligence era:**

In 1990s , there was emergence of data warehouse to store massive amounts of structured data.

**iii) Modern data science:**

Developers are using python, jupyter notebook, Tensorflow as tools for data science.

Machine learning and AI are widely used for prediction and automating.

**Task 3 : how generative AI is transforming data science**

Generative AI is a type of artificial intelligence that can create new content such as text, images, audio, video and even code.

Generative AI is revolutionizing data science by automating tasks, enhancing model accuracy, and driving innovation in problem-solving.

**Role of generative AI in data science**

* **Synthetic data generation**

In situations where real-world data is limited, sensitive or expensive to collect, generative AI can create synthetic data. This data helps to train machine learning models without risking privacy.

* **Enhanced decision making**

By generating predictive scenarios, summarizing reports and providing automated recommandation , AI helps business make faster and more accurate decisions.

**• Enhanced model accuracy and optimization**

Generating AI can be used to optimize model parameters and generate new features leading to improved model accuracy and performance.

**Applications and use cases**

* **Healthcare**

AI generated medical images are used to train diagnostic models without compromising patient privacy.Virtual patient simulations help in medical research.

* **Finance**

AI automates the creation of financial reports, detects fraud by simulating fraudulent activities . It also helps to forecasts stock movements in stock markets.

* **Marketing**

AI creates customer personas, designs marketing campaigns, writes ad content, predicts customer behaviour patterns.

**Benefits of generative AI in data science**

• Saves time by automating repetitive and time consuming tasks.

• Improves the accuracy and diversity of model by providing large scale synthetic data.

• Supports creativity

• Generates multiple solutions

**PracticalTasks:**

**Task 1: Case Study Analysis on Data-Driven Decision Making**

**AMAZON**

Amazon is one of the biggest e-commerce industry in the world. It uses data for sales performance, customer recommendation, most selling products, regional sales etc.

**Problem :**

**Personalized recommendation problem :**

How can amazon recommend the most relevant products to each customer?

**Data collection for the problem :**

1. Customer’s Browsing history
2. Recently purchased product
3. Items added or removed from cart
4. Products which was saved for later by customer
5. What customer searches on amazon
6. Products which was rated by customer
7. Specific region , age group etc.

**Techniques used to solve the problem :**

* Cleaning of data
* Filterazation of data
* Machine learning

**Impact on business :**

– personalized suggestions lead to higher purchase rates.

– by selling more products, they can create more revenue.

– customer can find easily whatever they want.

– satisfaction in customers

– higher customer retention

**Task 2 :Identifying Real-World Data Science Problems**

**Chosen industry : E-commerce**

Potential problems and their solutions

1. Demand forecasting :

* Problem :

How to predict that how much stock of product will be needed to reach the public demand ?

* Solution :

Predictive analytics models analyze past sales data and seasonable sales to forecast demand.

1. Fraud detection problem :

* Problem :

How to detect fraudulent transactions or fake reviews ?

* Solution :

Machine learning algorithms detect unusual patterns in payments, order frequency and user behaviour.

1. Customer churn prediction :

* Problem :

How to predict that which customer is likely to stop using the platform ?

* Solution :

Machine learning models analyze purchase frequency , complaints, engagement to predict churn.

1. Delivery cost and time :

* Problem :

How to deliver products cheaper and faster ?

* Solution :

Algorithm uses traffic patterns, delivery locations ,fuel cost etc. to find the fastest and shortest route.

1. Marketing problem :

* Problem :

How to send the right marketing message to right user ?

* Solution :

Data science uses customer data like behaviour, interests to create personalized ads and promotions.