Name: Sathvik B R

Class: BCA - A

Roll No: 2411021240027

Subject: introduction to data science

GitHub link: https://github.com/Sathvik0007/IDS-Assignment.git

## **Assignment 1 Part 1: Theoretical Understanding**

1. Define Data Science: What is Data Science? Discuss its key components and the CRISP-DM process.

Ans Data Science is the study of data to extract meaningful insights using scientific methods, algorithms, and technology. It combines statistics, machine learning, and domain expertise to analyse complex data for decision-making.

Key Components of Data Science

- **Data Collection** Gathering raw data from various sources.
- **Data Cleaning –** Removing errors and inconsistencies.
- **Exploratory Data Analysis (EDA)** Identifying patterns and trends.
- **Feature Engineering** Enhancing data for better model performance.
- ♦ Modelling- Applying machine learning to build prediction models. ♦ Evaluation- Measuring the accuracy of the model.
- **Deployment-** Application of the model in real time
- CRISP-DM Process
- CRISP-DM, or Cross-Industry Standard Process for Data Mining, is an approach to managing data science projects in a structured way:

Business Understanding- Defines the objectives of the project.

Data Understanding- Exploring patterns in data.

Data Preparation- Cleaning and transformation of data Modelling-

Applying machine learning techniques.

Evaluation- Measuring the performance of the model.

**Deployment-** Usage of the model in real applications.

This process ensures efficiency and reliability in data-driven solutions.

2.Explain how the CRISP-DM framework is applied in solving realworld problems (e.g., predicting customer churn or recommending movies).

Ans Applying the CRISP-DM Framework to Real-World Problems

The CRISP-DM framework is a systematic way to solve problems in data science. Below is how it can be applied to two real-world scenarios: predicting customer churn and recommending movies.

## 1. Predicting Customer Churn

Customer churn prediction will help a business identify potential customers to leave and take proactive steps in retaining them.

**Business Understanding:** 

- Define objective: Identify customers at risk of churning.
- Understand business requirements and major churn drivers (e.g., service issues, pricing).
- Data Understanding:
- Collect customer information like demographics, transaction history, and service usage.
- ❖ Analyse trends such as the number of complaints and inactive periods.
- **❖** Data Preparation:
- Clean missing or inconsistent data.
- Convert categorical variables like subscription plans into numerical form.
- Modelling:
- ❖ Use machine learning models like logistic regression, decision trees, or neural networks to classify churn risk. ❖ Evaluation:

Evaluate the model's accuracy using precision, recall, and F1-score.

Compare various models to choose the best one.

Deployment:

Implement the model in a CRM system to give churn alerts.

Use the insights to create retention strategies, such as personalized offers.

## 2. Movie Recommendation System

Recommender systems help platforms like Netflix and Amazon suggest content based on user preferences.

#### **Business Understanding:**

Define the goal: Increase user engagement by providing personalized movie recommendations.

Understand user behaviour and preferences.

## Data Understanding:

- Gather ratings, watch history, and genre preferences of the users.
- Determine the trends in movie popularity and user interaction.
- Data Preprocessing
- Remove missing ratings and normalize formats.
- Convert data into a structured format, such as a user-movie matrix.
- Modelling
- ❖ Apply collaborative filtering or content-based filtering algorithms.
- Train the model on past interactions to predict future preferences.
- Evaluation
- Measure the performance using metrics like Mean Absolute Error (MAE) and precisionrecall.
- Optimize algorithms for better recommendations.
- Deployment Integrate the system with the streaming platform.
- Always evolve recommendations based on user activity.
- By using the CRISP-DM approach, businesses can construct effective data-driven solutions that drive customer experience.

# 2. Case Study Questions: From the case studies in the "Module 1 Case Studies" file, answer the following:

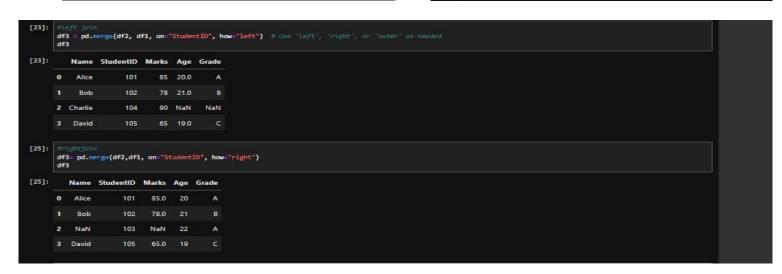
What is the main business objective of the Netflix Recommendation System?

Ans The main business objective of the Netflix Recommendation System is to increase user engagement and retention by providing personalized content suggestions. Based on user preferences, watch history, and behaviour patterns, Netflix aims to recommend movies and TV shows in order to:

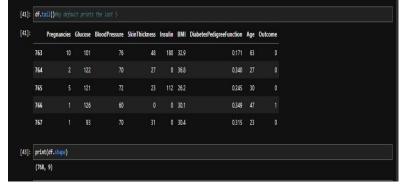
>>Reduce search time for enhancing user satisfaction.

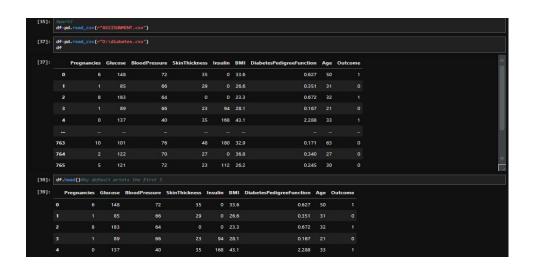
- >>Improve customer retention by making viewers engage.
- >>Optimise content discovery and increase watch time.
- >>Reduce churn by ensuring that users like what they're seeing.

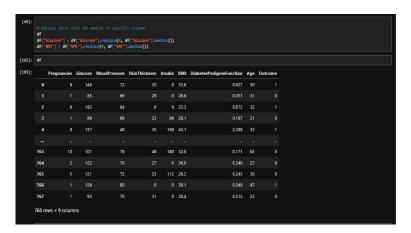
The recommendation system of Netflix ensures that the company remains competitive by delivering an extremely personalized viewing experience. and decision-making.

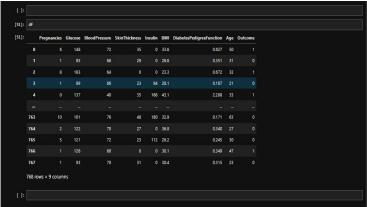












l:	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	ВМІ	DiabetesPedigreeFunction	Age	Outcome	
763	10	101	76	48	180	32.9	0.171	63	0	
764		122	70	27		36.8	0.340	27		
765		121	72	23	112	26.2	0.245	30	0	
766		126	60			30.1	0.349	47		
767		93	70	31		30.4	0.315	23	0	
: pri	nt(df.shape)									