Case Study

Netflix Recommendation System

Data Science Process

Step 1) **Business understanding ->** Netflix has thousand of movies and TV Shows for which user find difficulty to choose and decide to what to watch. The user choose the main goal is to recommends & personalized movies & Shows to each user based on their past viewing and based on their rating and if user does not get good suggestion they may have waste of internet aswell as they leave the platform

Step 2) **Data understanding ->** It basically used to collect data from given Source. and Column of dataset which I was already provide in question here we have Movie lens dataset its public dataset we use for recommendation.

Columns of dataset we have

1) User ID -> 123 (Identify each uses)

2) Movie ID -> Identifies each movie)

3) Rating (1-5) -> 4 (Shows how much user likes movie)

4) Timestamp -> When rating was given

5) Movie Meta data -> Krish, 2004, Drama (title, year of release, genres)

Step 3) **Data preparation ->** Before modeling we need to clean and transform the data

1) Missing values in movie detail like (release year) are fixed or removed

2) Timestamp converted into human understanding language like dates

3) genres are converted to numerical format like 0's for

Action and 1 for Drama this is called one hot encoding

Now data is ready for analysis and modeling

Step4) **Modeling →** As you saw the question it was given that there

are 2 types of approaches uses

1 Collaborative filtering → if user have Similar trend to watch and like/want for movies we can identify using SVD (Singular Value Decomposition) technique which is used for find pattern between user and movies

Eg → if there are two users A and B both have Similar movie taste and user A Watch movie that User B have not Seen then it will go in recommendation of that movie to users

2) Content Based filtering

1) uses movies metadata (like genres, release date)

2) if user like action movie then the System recommends another action movie to the user

Step5) **Evaluation →** it basically how user watch movie by Collaborating filtering (80%) and (20%) by Content based filtering

Step6) **Deployment →** Final Step user can login in Netflix and System check personalized taste of movie, rating and viewing history to and the System regularly updated as new rating and user activity

This Way Netflix ensures every user gets a unique engaging experience

Case Study

Predicting Customer churn in telecom

Data Science Process

Step1) **Business understanding ->** Telecom Company face big issue Customer leaves there companies because of higher rate resolve the share price and increasing price in mobile recharge and other DTH recharge because the Company in going in loss - find genuine customers and take action early give then some offers, discounts to does not leave telecom Service and improve overall revenue

Step2) **Data understanding ->** we use telco Customer Churn Dataset (from kaggle)

1. Customer id -> unique for each customer
2. Demographics -> find which age group of people uses this Service
3. Service detail -> internet service type, monthly charges
4. Churn -> basically customer want to left or Stayed

Step3) **Data preparation ->** missing values are filled with monthly charges Categorical variable

Eg-> gender, contact type are converted into numeric best encoding - monthly are charges are same or little bit less than other to company to perform better

Step4) **Modeling ->** There are two models

1. logistic regression -> Simple, interpretable
2. Random Forest -> more powerful, handle complex relationship Contact type (one year, two year), how long customer with company.

Step 5) **Evaluation →** to check how model is read if we given Matrix

Accuracy → Overall correctress

Precision → How many are actually left

Recall → How many real verified uses

F1 Score → balance between precision and recall

Eg-> Accuracy → 80%

Precision → 75%

Recall → 70%

F1 Score → 72%

Step 6) **Deployment →** Once the model is trained and face tested it Can be used in real life

1) Customer Service teams get alerts for high risk Customers

2) Special reptation action like discount, Special offer or personalized plans

3) The system can be updated regularly with new data to improve predictions

Case Study

Predicting diabetes in Health Care

Data Science Process

Step1) **Business understanding →** Diabetes is a serious disease that affects millions of people worldwide early detection can help in starting treatment which improve health

Step2) **Data understanding →** We use PIMA Indians diabetes dataset from kaggle

Columns in Dataset

1 Age

2 Body Mass Index (BMI)

3 Blood pressure

4 glucose level

5 Insulin level

6 Diabetes outcome 0 = No, 1 = yes

Higher glucose level are strongly linked to diabets

Patient have higher BMI (overweight) are more likely diabetes

Step3) **Data Preparation**

1) Missing value, BMI = 0 which means not real, these are replaced with median value

2) Normalisation → BMI and glucose are normalized similar scale

Step4) **Modeling →** 2 types of Model

1 Logistic regression → simple and interpretable help understand doctor to diabetes risk

2 Randon forest → more powerful model and improve prediction , Accuracy

Step5) **Evaluation →** to measure how good model is :-

1 Accuracy → overall correctness

2 AUC / ROC → Identify diabetic & non diabetic patient

3 Confusion Matrix → Show true positive

Step6) **Deployment →** once the model is built and validate it can used in healthcare system

1) Integrated in hospital management system

2) Doctor receive alert for patient with high risk of diabetes

3) patient give proper treatment plan (diet, excercise)

4) The Model can be updated as more patient data is available