1. A healthcare startup wants to build a model to predict the onset of diabetes. The dataset has 50 features, but the team wants to use only the most relevant ones. How can they apply feature selection effectively?
    - Collect the data
    - Preprocess the data if there is any missing values.
    - Using correlation method remove the columns which are highly correlated.
    - Split the train and test data.
    - Apply selectKbest or RFE (with Logistic Regression Algorithm since it is classification problem) to reduce the features with relevant ones.
    - Train the model using selected features
    - Evaluate Accuracy for the model.

2. You're creating a new project for managing student data. The project name is student_portal. What is the step-by-step command to start it?write logic for this django project

    - Open the command prompt
    - Go to the path where the project to be
    - django-admin startproject student_portal
    - student_portal folder is created in the path.

3. A company is analyzing customer purchase patterns with 200+ behavioral features. How can they reduce dimensionality without losing predictive power?

    - Collect the data
    - Preprocess the data if there is any missing values.
    - Use PCA for dimensionality reduction.
    - Train the model using top features
    - Evaluate Accuracy or RMSE for the model.

4.A digital library wants to recommend books to readers based on what similar readers liked. How should they design this system?

    Collaborative recommendation system. (User or item based)

    - Create the pivot table with user and the books Id and the rating of the books.
    - Find the cosine similarity score between the users or the between the items.
    - Create prediction table if user reads the books.
    - List the similar users read the books.
    - Remove the books from the user list if user already read the books.
    - For the remaining books, select the books which has more rating in the prediction table.
    - Recommend those books to the users.

5. A bank wants to assess the risk level of credit applicants using only the most important financial indicators. How can they reduce the number of features?

    - Collect the data
    - Preprocess the data if there is any missing values.
    - Using correlation method remove the columns which are highly correlated.

- Split the train and test data.
- Apply RFE (with Logistic Regression Algorithm since it is classification problem) to reduce the features with relevant ones.
- Train the model using selected features
- Evaluate Accuracy for the model.

6. A news app wants to recommend articles based on both article similarity and user reading history. How can they implement a hybrid system?

Content Based Recommendation

- Create the pivot table for user and the items.
- Create a vector list for whole user with to item vector
- Create the content vector using user and item vectors
- Evaluate the content vector using rating

7. You're building a spam detection model and have thousands of text features from emails. How do you identify the most useful ones?

- Collect the data
- Preprocess the data if there is any missing values.
- Using correlation method remove the columns which are highly correlated.
- Split the train and test data.
- Apply RFE (with Logistic Regression Algorithm since it is classification problem) to reduce the features with relevant ones.
- Train the model using selected features
- Evaluate Accuracy for the model.

8. An ed-tech platform wants to recommend courses based on what similar learners have enrolled in. What steps would you take?

Collaborative recommendation system. (User or item based)

- Create the pivot table with user and the course Id and the rating of the course.
- Find the cosine similarity score between the users or the between the items.
- Create prediction table if user learn the course.
- List the similar users learn the course.
- Remove the course from the user list if user already learn the course.
- For the remaining courses, select the course which has more rating in the prediction table.
- Recommend those course to the users.

9.You’re developing a car price prediction tool. With 100+ features (e.g., brand, mileage, engine type), how do you reduce complexity?

- Collect the data
- Preprocess the data if there is any missing values.

- Using correlation method remove the columns which are highly correlated.
- Split the train and test data.
- Apply selectKbest to reduce no of the features with relevant ones.
- Train the model using selected features
- Evaluate Accuracy for the model.

10. How do you recommend products to new users who haven't interacted with anything yet?

Popularity based recommendation

- Create the pivot table for user and the items.
- Create a vector list for whole user with to item vector
- Create the content vector using user and item vectors
- From the content vectors select the top 5 predicted values. For the top5 values select the products.