AIML | MODULE PROJECT

greatlearning
Power Ahead

5

Takeaways

greatlearning
Power Ahead

1 AIML module projects are designed to have a detailed hands on to integrate theoretical knowledge with actual practical implementations.

2 AIML module projects are designed to enable you as a learner to work on realtime industry scenarios, problems and datasets.

3 AIML module projects are designed to enable you simulating the designed solution using AIML techniques onto python technology platform.

4 AIML module projects are designed to be scored using a predefined rubric based system.

5 AIML module projects are designed to enhance your learning above and beyond. Hence, it might require you to experiment, research, self learn and implement.

greatlearning
Power Ahead

AIML | MODULE
PROJECT

RECOMMENDATION
SYSTEMS

AIML module project consists of industry based problems framed as detailed questions which can be solved using recommendation model.

TOTAL
SCORE | 60

PROJECT **BASED**

- **DOMAIN:** Smartphone, Electronics
- **CONTEXT:** India is the second largest market globally for smartphones after China. About 134 million smartphones were sold across India in the year 2017 and is estimated to increase to about 442 million in 2022. India ranked second in the average time spent on mobile web by smartphone users across Asia Pacific. The combination of very high sales volumes and the average smartphone consumer behaviour has made India a very attractive market for foreign vendors. As per Consumer behaviour, 97% of consumers turn to a search engine when they are buying a product vs. 15% who turn to social media. If a seller succeeds to publish smartphones based on user's behaviour/choice at the right place, there are 90% chances that user will enquire for the same. This Case Study is targeted to build a recommendation system based on individual consumer's behaviour or choice.
- **DATA DESCRIPTION:**
    - **author** : name of the person who gave the rating
    - **country** : country the person who gave the rating belongs to
    - **data** : date of the rating
    - **domain**: website from which the rating was taken from
    - **extract**: rating content
    - **language**: language in which the rating was given
    - **product**: name of the product/mobile phone for which the rating was given
    - **score**: average rating for the phone
    - **score_max**: highest rating given for the phone
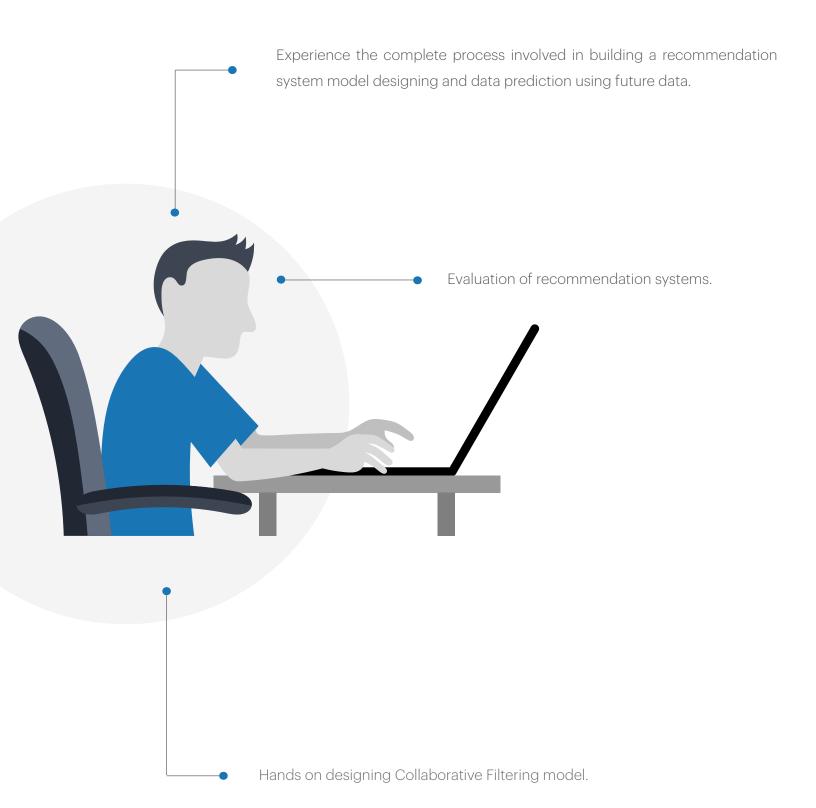   - **source**: source from where the rating was taken
        *Data source:
- **PROJECT OBJECTIVE:** We will build a recommendation system using popularity based and collaborative filtering methods to recommend mobile phones to a user which are most popular and personalised respectively..

Steps and tasks: **[Total Score: 60 points]**

1. Import the necessary libraries and read the provided CSVs as a data frame and perform the below steps.
    - Merge the provided CSVs into one data-frame.
    - Check a few observations and shape of the data-frame.
    - Round off scores to the nearest integers.
    - Check for missing values. Impute the missing values if there is any.
    - Check for duplicate values and remove them if there is any.
    - Keep only 1000000 data samples. Use random state=612.
    - Drop irrelevant features. Keep features like Author, Product, and Score.
2. Answer the following questions
    - Identify the most rated features.
    - Identify the users with most number of reviews.
    - Select the data with products having more than 50 ratings and users who have given more than 50 ratings. Report the shape of the final dataset.
3. Build a popularity based model and recommend top 5 mobile phones.
4. Build a collaborative filtering model using SVD. You can use SVD from surprise or build it from scratch(Note: Incase you're building it from scratch you can limit your data points to 5000 samples if you face memory issues). Build a collaborative filtering model using kNNWithMeans from surprise. You can try both user-based and item-based model.
5. Evaluate the collaborative model. Print RMSE value.
6. Predict score (average rating) for test users.
7. Report your findings and inferences.
8. Try and recommend top 5 products for test users.
9. Try cross validation techniques to get better results.
10. In what business scenario you should use popularity based Recommendation Systems ?
11. In what business scenario you should use CF based Recommendation Systems ?
12. What other possible methods can you think of which can further improve the recommendation for different users ?

LEARNING
OUTCOME

Experience the complete process involved in building a recommendation system model designing and data prediction using future data.

Evaluation of recommendation systems.

Hands on designing Collaborative Filtering model.

" *Put yourself in the shoes of an actual* "

DATA SCIENTIST

THAT's **YOU**

Assume that you are working at the company which has received the above problem statement from internal/external client. Finding the best solution for the problem statement will enhance the business/ operations for your organisation/project. You are responsible for the complete delivery. Put your best analytical thinking hat to squeeze the raw data into relevant insights and later into an AIML working model.

PLEASE **NOTE**

Designing a data driven decision product typically traces the following process:

1. Data and insights:
   Warehouse the relevant data. Clean and validate the data as per the the functional requirements of the problem statement. Capture and validate all possible insights from the data as per the the functional requirements of the problem statement. Please remember there will be numerous ways to achieve this. Sticking to relevance is of utmost importance. Pre-process the data which can be used for relevant AIML model.

2. AIML training:
   Use the data to train and test a relevant AIML model. Tune the model to achieve the best possible learnings out of the data. This is an iterative process where your knowledge on the above data can help to debug and improvise. Different AIML models react differently and perform depending on quality of the data. Baseline your best performing model and store the learnings for future usage.
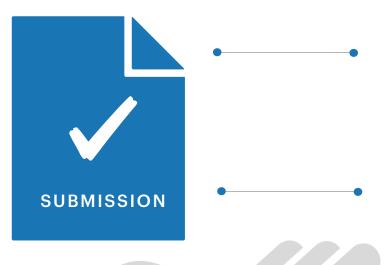
3. AIML end product:
   Design a trigger or user interface for the business to use the designed AIML model for future usage. Maintain, support and keep the model/ product updated by continuous improvement/training. These are generally triggered by time, business or change in data.

IMPORTANT
POINTERS

Project should be submitted as a single ".html" and ".ipynb" file. Follow the below best practices where your submission should be:

- ".html" and ".ipynb" files should be an exact match.
- Pre-run codes with all outputs intact.
- Error free & machine independent i.e. run on any machine without adding any extra code.
- Well commented for clarity on code designed, assumptions made, approach taken, insights found and results obtained.

SUBMISSION

Project should be submitted on or before the deadline given by the program office.

Project submission should be an original work from you as a learner. If any percentage of plagiarism found in the submission, the project will not be evaluated and no score will be given.

greatlearning
Power Ahead

HAPPY
LEARNING