Walk in the shoes of a Data Scientist (the most sought after job of the century) by using innovative approaches, statistical analysis and cutting-edge machine learning & deep learning algorithms to get the best predictions to top the real-time leaderboard.

This year the modelling competition deals with '*Bibliotheca E-Book Subscription Prediction*'. The competition has been designed for the participants to explore machine learning algorithms for this recommendation problem and provide a platform to showcase their machine learning skills.

Problem Statement

Bibliotheca is a platform that gives its users an exclusive chance to subscribe to E-books at discounted price. Users can also get a preview of the books by reading a sample of pages even before they buy subscription for any specific book. Each purchase for the subscription gives users access to the relevant E-book for a fixed period.

We have information about all the books that a User has previewed/purchased subscription for; on the platform in the past and the problem statement asks you to predict the ordered set 10 most probable books from the test set, the user is going to buy in the coming weeks.

To participate in the competition please register your team on the Dlabs Registration Page.

Evaluation

For this competition, you need to predict the right books that User is most likely to buy out of the books lying in test-dataset, and every correct prediction for the user rewards you; your overall task is to Maximize the rewards point to win.

For i[th] User who has bought 'm' books in the coming weeks where $P_i(k)$ is fraction of correct books identified in top-k books prediction; with 'n' books (here n=10) recommended for the user:

$$Reward_i = \begin{cases} 0 \; when \; m = 0 \\ \frac{1}{min(m,10)} \sum_{k=1}^{\min(n,10)} P_i(k) \; , otherwise \end{cases}$$

$$MODELSCORE = \frac{1}{Number\_of\_Users} \sum_{i=1}^{For\ all\ Users} Reward_i$$

Winners will be decided for each of the location based on highest total Model Score achieved by the team. In case of a tie (for top 3 ranks), the following criteria (in the same order) will be used:

(1.) Number of Users with a Predicted E-Book
(2.) Submission provided earlier will get priority

Both the above criteria will be used only in case of tie and in the given order while awarding a better rank.

Submission Rules

(1.) Each team can make 5 submissions on the competition website per day. The submission with the best performance on the public data will be visible on the leaderboard.
(2.) There should be 10 unique BookIDs recommended for each of the 15,000 users in the database. Thus meaning that submission file should have 15,001 rows(15,000 users + 1 header)
(3.) Till 7[th] April 2019 11:59 PM, teams can make submissions like the sample submission file which will be provided in the DATA section.
(4.) By 14[th] April 2019 11:59 PM, all teams need to submit their final scored submission file and code used to generate those scores along with any data cleaning, model training, parameter selection, hyperparameter tuning as a zipped file named as Location_TEAMNAME_ModelingCompetition.zip. This single file with a brief README with instructions on running the codes needs to be sent at usianalyticssummitcompetitions@deloitte.com.
(5.) The public leaderboard will be visible on the competition website and will keep on updating on a real-time basis with new

th

submissions until 7 April 2019.
(6.) The submission file for the competition should consist of the following columns:

"USERID" : User-ID
"PURCHASEDBOOKID" : Comma separated list of all the BOOKIDs recommended

Presentation Round:

For each location, 2 winners will be declared;
--one team with best performance on the modeling score leaderboard and,
--one team with most innovative solution

The innovation quotient of a team will be scored based on the following four high-level criteria:
--Innovative use of unstructured data, image data and external data
--Innovative machine learning algorithms/techniques used
--Explore the key drivers of the target variable, feature engineering and insight creation
--Quality of the solution documentation (visualizations encouraged)

The teams securing first position in their respective locations will be invited to Hyderabad on 8th May 2019 to compete at the Grand Summit and stand a chance to win the overall Modeling award at the Deloitte USI Analytics Summit 2019.

Competition Timeline

Event	***Date***	***Description***
Competition Launch	5th March 2019	Data Download links become active
Evaluation Start	6th March 2019	Submissions acceptance starts – **Leaderboard updated after each submission**
Final Evaluation: Submission Deadline	7th April 2019 11:59PM	**Final date for submissions**.
Announcement of teams shortlisted for Presentation	9th April 2019	The leaderboard will be updated to reflect the final scores (on the updated test data) at this point.

Presentation on modeling approach and insights	TBD	Top 5 teams from each location may be invited to present their modeling approach to a PPD panel, followed by a Q&A session

NOTE: Google Chrome and Firefox are the preferred browsers for this website.

Data
The training data will include the information about the Books, Users and their history. The books that users have previewed/subscribed in past has been made available as part of the visiting history dataset. The visit data will not be available for the test period and all the books recommendations has to be created for the books in the test-set only.

Filename
Remark
BOOKSCATALOGUE.csv
This dataset has information about which all geographies the book is active and available on

BOOKSMASTERTRAIN.csv
Information about all the books in the train-set and you'll have information about the visit/purchase history for books belonging to this set

BOOKSMASTERTEST.csv
Information about all the books in the test-set. The predicted books must only belong to this set.

BOOKSPURCHHISTORY.csv
All the timestamp and purchase history captured for the users. You may find more purchases in this dataset than in visiting history (Not all the purchases may be tracked using existing tracking-framework at Bibliotheca)

BOOKSVISITHISTORY.csv
Historical data for each book that users have previewed or subscribed. Column 'SUBSCRIBED' is the target for this modeling competition and participants needs to recommend what are the possible set of top 10 books that user is going to subscribe next.

USERMASTER.csv
Users dataset

SAMPLESUBMISSION.csv
Sample Submission File

If you have already registered your team for the competition. You can download the competiton data here. (Please use VPN for downloading dataset)