Problem Statement   Discussions   Our Submissions   About Hackathon

Identifying Healthcare Professionals (HCP) and their specialization

The ad server logs contain valuable information such as browser details, IP addresses, geographic locations, search patterns, site urls, and other relevant data. The objective of the hackathon is to build a robust model that can accurately predict whether a user belongs to the HCP category and its specialization id/taxonomy.

Input: Ad server logs containing information on user behavior and other features
Output: Given a userid(userplatformuid) and ad server log, classifies if the user is an HCP and its specialization (taxonomy).

The input contains the following type of data:

1. HCP has been identified but the taxonomy/specialization is not identified - IS_HCP is true but taxonomy is not populated.

2. Both HCP and taxonomy/specialization are identified - IS_HCP is true and taxonomy is also populated.

3. The remaining rows pertain to data where there is no information about HCP. It may correspond to HCP or may not.

Data Description:
The data fields are described below:

Field	Description	Example	Data Type
ID	Serial number	1001, 1002, etc.	Integer
DEVICETYPE	The type of device used by the user, such as desktop, mobile, tablet, etc.	desktop, mobile, tablet, smartphone	String
PLATFORM_ID	Unique identifier associated with the platform used by the user.	12345, PLATFORM_XYZ, 987654321	Numeric
BIDREQUESTIP	IP address from where the request originated.	192.168.0.1, 10.0.0.2, 172.16.0.100	String
USERPLATFORMUID	Unique identifier of the user on a platform.	user123, platform_user_456, 789user	String
USERCITY	City of the user.	New York, London, Tokyo	String
USERZIPCODE	User's ZIP code.	12345, 54321, 98765	String
USERAGENT	User's browser agent, providing information about the browser and operating system.	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36	String
PLATFORMTYPE	Type of platform being used, such as Electronic Health Record (EHR), e-prescribing, medical journal, etc.	EHR, e-prescribing, medical journal, telemedicine	String
CHANNELTYPE	The channel through which the user interacts with the platform, such as a website or email.	website, email, mobile app, SMS	String
URL	URL from which the request originated.	https://example.com	String
KEYWORDS	Keywords appear in URLs	healthcare, doctor, medicine, patient care	String
TAXONOMY	Code for specialization, which could refer to a specific medical field or category.	cardiology, dermatology, pediatrics, oncology	String
IS_HCP	Indicates whether the user is a Healthcare Professional (HCP) or not.	1, 0	Boolean

Click here for Sample PPT(Mandatory)-PPT submission is mandatory.

Note: The sample submission file contains random data.

Data Set Download Data Set

File Name	Description	Format	Size

Doceree-HCP_Train.csv	Train_Data	csv	
Doceree-HCP-Test.csv	Test_Data	csv	
Doceree_HCP_Sample_Submission_1.csv	Sample_Submission	csv	

Data Dictionary

Here's a brief version of what you'll find in the data description file.

Variable	Description
ID	key
TAXONOMY	Code for specialization, which could refer to a specific medical field or category
IS_HCP	Indicates whether the user is a Healthcare Professional (HCP) or not.

Submission

1. Output file generated by the model.
2. Source Code file or .ipynb file or Report
3. Demonstration Video (If Possible)

Evaluation Metric

1. Correctness of code / Score Received
2. Source or .ipynb file Submitted
3. Presentation with model explanation (mandatory)