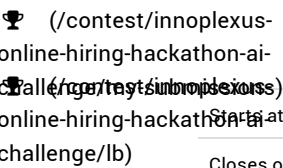


₹  
?  
⌘  
✓



Starts at	Sat Aug 11 2018 00:00:00 GMT+0530 (IST)
Closes on	Sun Aug 12 2018 23:59:00 GMT+0530 (IST)
Mode	Online
Fee	Free
Location	Online
# Participants	1699
Prizes	Google Pixel2 and Interview opportunities with Innoplexus

## Feedback

## Innoplexus Online Hiring Hackathon: Take the Artificial Intelligence (AI) Challenge

The best participants showcasing their skills will be shortlisted for interviews with Innoplexus and additionally the winners will receive exciting prizes.

At Innoplexus we are helping organisations move to continuous decision-making by generating insights from structured and unstructured private and public data leveraging cutting edge Artificial Intelligence and Deep Learning technologies. We aim to transform batch decision making in enterprises into a continuum. We continuously develop our products to enable broader, deeper and faster insights and best-found decision-making.

Our Data as a Service (DaaS) and Continuous Analytics as a Service (CaaS) solutions are helping Fortune 500 companies drive business outcomes. We automate the collection, curation, aggregation and analysis, of billions of data points from thousands of data sources, using machine learning, network analysis, ontologies, computer vision and entity normalization.

Next



Learn more at [www.innoplexus.com](http://www.innoplexus.com) (<http://www.innoplexus.com/>)

Registration: Free

Prizes:

1. An AI powered smartphone - Google Pixel 2
2. Interview opportunities with Innoplexus
3. Other smart AI powered prizes

We are hiring for:

Innoplexus is a place where people with varied industry and functional backgrounds and experience join to do their best. To strengthen our team in Pune, we are looking for Data Scientists at various levels.

You are the right person in our team if you are:

- Enthusiastic about developing scalable, intuitive product
- An out-of-the-box thinker
- Strong problem solver
- Extremely good in data structures
- A PhD, Master or Bachelor in Computer Science or Engineering

Role and responsibilities:

- Develop solutions for real-world noisy data, large-scale problems.
- Develop highly scalable deep learning algorithms to improve our products.
- Develop state-of-the-art machine learning and neural network methodologies to improve our intelligence platform.

To excel in this job you must have:

- Hands-on experience in machine learning.
- Strong track record in AI / ML publications in renowned scientific journals or conferences.
- Experience in any of the following: Computer Vision, Image Processing, Speech Recognition, Natural Language Understanding, Machine Learning, Deep Learning, HCI, Text Mining, Computational Genomics, Bioinformatics, other Machine Intelligence/Artificial Intelligence related areas.
- Programming experience in one or more of the following: C, C++, Python.

You will earn brownie points if you have experience in:

- Working with life sciences or healthcare datasets.
- Solving real world ML problems.
- Either of the major NoSQL products.

We believe in the power of collaboration. You will work alongside industry's best creative thinkers and can grow with your team and turn great ideas into reality. We support you to develop your skill with training courses, knowledge sharing and learn while you work.

Feedback

## Registration Fee

Free

## Problem Statement

Classification of Web page content is vital to many tasks in Web information retrieval such as maintaining Web directories and focused crawling. The uncontrolled nature of Web content presents additional challenges to Web page classification as compared to traditional text classification, however the interconnected nature of hypertext also provides features that can assist the process.

Here the task is to classify the web pages to the respective classes it belongs to, in a single label classification setup (Each webpage can belong to only 1 class).

Basically given the complete html and url, predict the tag a web page belongs to out of 9 predefined tags as given below:

- 1) People profile
- 2) Conferences/Congress
- 3) Forums
- 4) News article
- 5) Clinical trials
- 6) Publication
- 7) Thesis
- 8) Guidelines



## 9) Others

### Data Dictionary

train.zip contains 2 csvs

#### 1. train.csv: Train set

Variable	Definition
Webpage_id	Unique ID for the Web page
Domain	Domain
Url	Complete Url
Tag	(Target) Tag (Class) of the Web page

#### 2. html\_data.csv: Contains web page data in HTML for both train and test web pages

Variable	Definition
Webpage_id	Unique ID for the Web page
Html	Web page data in HTML

#### test.csv: Test Set

Variable	Definition
Webpage_id	Unique ID for the Web page
Domain	Domain
Url	Complete Url

#### sample\_submission.csv: Submission format

Variable	Definition
Webpage_id	Unique ID for the Web page
Tag	(Target) Tag (Class) of the Web page

### Train-Test Split

The train and test data split is done based on Domain-Tag combination. For example, suppose we want to split the following sample of 16 URLs into train and test set.

Domain	Url	Tag
1	1.1	news
2	2.1	news
3	3.1	news
1	1.2	news
2	2.2	news
3	3.2	news
4	4.1	news
5	5.2	others
6	6.1	others
7	7.1	others
8	8.1	news
9	9.1	news
10	10.1	others
11	11.1	news
11	11.2	news
11	11.3	others

- First the overall dataset is split into subsets by Tag as shown below:

Domain	Url	Tag
1	1.1	news
2	2.1	news
3	3.1	news
1	1.2	news
2	2.2	news
3	3.2	news
4	4.1	news
8	8.1	news
9	9.1	news
11	11.1	news
11	11.2	news

Domain	Url	Tag
5	5.2	others
6	6.1	others
7	7.1	others
10	10.1	others
11	11.3	others

- Now for each subset(Tag) we store all unique domains and randomly shuffle them, so in this case lets say we have:



Tag	All Domains (after random shuffling)
news	(2,1,4,3,11,9,8)
others	(5,6,11,10,7)

- Next, every third domain (3<sup>rd</sup>, 6<sup>th</sup>, 9<sup>th</sup> and so on) in the all domain sequence is assigned to the test and the rest (1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 7<sup>th</sup> and so on) are assigned to train as shown in the following table:

Tag	All Domains (after random shuffling)	Train domains	Test domains
news	(2,1,4,3,11,9,8)	(2,1,3,11,8)	(4,9)
others	(5,6,11,10,7)	(5,6,10,7)	(11)

- Final train and test set would be:

Train Set			Test Set		
Domain	Url	Tag	Domain	Url	Tag
1	1.1	news	4	4.1	news
2	2.1	news	9	9.1	news
3	3.1	news	11	11.3	others
1	1.2	news			
2	2.2	news			
3	3.2	news			
5	5.2	others			
6	6.1	others			
7	7.1	others			
8	8.1	news			
10	10.1	others			
11	11.1	news			
11	11.2	news			

## Evaluation Metric

The evaluation metric for this competition is weighted F1 score.

## Public and Private Split

Test data is further randomly divided into Public (40%) and Private (60%) data.

- Your initial responses will be checked and scored on the Public data.
- The final rankings would be based on your private score which will be published once the competition is over.

## Innoplexus Hackathon Rules

1. Entries submitted after the contest is closed will not be considered.
2. Since this is a hiring hack, you are expected to solve the problem on your own.
3. Use of external dataset is strictly prohibited.
4. Use of Webpage\_id as a feature is not allowed.
5. Participation is free-of-charge.
6. Participant must update their profile details and upload their latest CV.

## User Privacy

The profile of the user as updated at time of registering for the contest along with their CV and Analytics Vidhya profile will be shared with the sponsor of the hackathon for purposes of hiring.

## Solution Checker

1. You are free to use solution checker as many times as you want.
2. Adding comments is mandatory for the use of solution checker
3. Comments will help you to refer to a particular solution at a later point in time.

## Final Submission

1. Setting final submission is mandatory. Without a final submission, your entry will not be considered.

Feedback



2. Code file is mandatory while sending final submission. For GUI based tools, please upload a zip file of snapshots of steps taken by you, else upload code file.
3. The code file uploaded should be pertaining to your final submission.

## Rules of Conduct

1. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.
2. Slack Live Chat admins hold the right to block any participant found to use foul / disrespectful language.
3. Analytics Vidhya and Innoplexus hold the right to disqualify any participant at any stage of competition, if participant(s) are deemed to be acting fraudulently.

## How to make a submission?

How to make a submission



Feedback

## How to set a Final Submission?

How to set a Final Submission on DataHack?



## Data



*Note :- The datasets in this competitions is solely meant to be used for this competition only. You cannot use it for*

any other purpose.


- 📄 Test File (/contest/innoplexus-online-hiring-hackathon-ai-challenge/download/test-file)
- 📄 Train File (/contest/innoplexus-online-hiring-hackathon-ai-challenge/download/train-file)
- 📄 Sample Submissions (/contest/innoplexus-online-hiring-hackathon-ai-challenge/download/sample-submission)

## Solution Checker

Code File	<input type="button" value="Browse..."/> No file selected.
Solution File* (.csv, .zip only)	<input type="button" value="Browse..."/> No file selected.
Solution Description (max : 180 chars)*	<input type="text" value="Solution Description (for your interest)"/>
Do you want to show your code on leaderboard?*	Yes <input type="radio"/> No <input type="radio"/>
<input type="button" value="Add Solution"/>	

[Join Slack Live Chat](#)

(HTTPS://TWITTER.COM  
/SHARE?UTM\_SOURCE=TWITTER\_SHARE&  
UTM\_MEDIUM=DH\_SOCIAL&  
URL\_HTTPS://DATAHACK.ANALYTICSVIDHYA.COM+  
/CONTEST/INNOPEXUS-  
ONLINE-HIRING-  
HACKATHON-AI-  
CHALLENGE  
&TEXT=HEY! I AM  
PARTICIPATING IN AN  
EXCITING PROBLEM BY  
@ANALYTICSVIDHYA  
INNOPEXUS ONLINE  
HIRING HACKATHON  
COMPETE AND LEARN.  
#DATAHACK)



**FLASH SALE**

**GET 60% OFF**

**OFFER ENDS**  
10TH AUG 11:59PM IST

**COURSE ON**

**COMPUTER VISION  
USING DEEP LEARNING**

**USE CODE CV10160**

**AVAIL THE COURSE AT JUST ₹ 12,000**

[Feedback](#)

([https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+CVDL101+CVDL101\\_T1/about?utm\\_source=datahackcombannerflash60](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+CVDL101+CVDL101_T1/about?utm_source=datahackcombannerflash60))

