

# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Participants kits

Thank you for your participation in the largest AI Hack Tunisia in Tunisia and the MENA region. By your participation in AI Hack Tunisia, you're not just making an impact on your life, but you are impacting your society, your country, and changing the status quo. **Together let's Build the Next** and make it happen.

Please check the [General Info Kits](#) for the event layout, schedule, and seating info.

Most competitors will be staying at the venue during the whole events where they be places to rest. If you are coming from outside Tunis on Thursday we will provide accommodations.

## Individual CHALLENGES

Participants can develop innovations revolving one of the four projects:

- Tunisian dialect Natural Language Understanding (NLU)
- Predictive Analytics: Ministry of Finance Tax Fraud Detection
- Predictive Analytics: STEG Electricity and Gas Consumption Fraud Detection
- Computer Vision: Tunisian Vehicle License Detection and Recognition

# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Tunisian Dialect NLU challenge

Multi-Turn Response Selection with the Tunisian Dialect

**Problem statement:** The Tunisian dialect is very complicated when it comes to understanding its written forms. Despite the complex aspect of this dialect, it remains the language used in the everyday communications and in most Tunisian social media. Humans languages understanding relies on semantic and functional dependencies, and the relationship between dialogue elements and their context. One important task based on this approach is the “Multi-Turn Response Selection” task. It aims to select the best-matched response from a set of candidates given the context of a conversation.

**Objective of the challenge:** The aim of the competition is to create a model that can understand the Tunisian dialect using the TunisianDialect dataset. Train the model on the “Multi-Turn Response Selection” task and deliver significant test results revealing how far the model can deal with this dialect.

**Relevance of the solution:** Proposing an NLP solution, inspired by the recent advances in NLP, that solves the explained problematic. The solutions to this challenge are the first steps towards the field of Tunisian language processing models and similar areas. We look forward to taking this journey with you!

**Evaluation Metrics:** The model is asked to select the  $k$  best-matched responses from  $n$  available candidate responses ‘ $r$ ’ for the given context ‘ $c$ ’. So, we calculate the recall  $Rn@k$  of the true positive replies among the  $k$  selected ones. The Recall expression is given below:

$Rn@k = \sum_{i=1..k} y_i / \sum_{i=1..n} y_i$  where  $y_i$  are the labels.

More specifically, we choose to work with  $R2@1$ ,  $R10@1$ ,  $R10@2$  and  $R10@5$ ,

**Data:** InstaDeep created a Tunisian corpus, the TunisianDialect dataset, for the “Multi-turn Response Selection” task. The dataset contains more than 120K conversations extracted from the “Ooredoo Tunisie”, a Tunisian telecommunication company, Facebook page. The collection period is from July 2018 until March 2019.

# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Ministry of Finance Tax Fraud Detection

### Predictive Analytics Challenge

**Problem statement:** Tax fraud is the intentional act of lying on a tax return form with intent to lower one's tax liability. Under-reporting is one of the most common types of tax frauds. It consists of filling a tax return form with a lesser tax base. As a result of this act, fiscal revenues are reduced, undermining public investment.

**Objective of the challenge:** The objective of the challenge is to detect tax fraud. As it is one of the main priorities of local tax authorities which are required to develop cost-efficient strategies to tackle this problem.

**Relevance of the solution:** Apply a supervised machine learning technique to detect the potential fraudulent tax payers, hence, increasing the operational efficiency in the tax supervision process using historical data.

**Evaluation Metrics:** The error metric for this competition is the Root Mean Squared Error.

**Data:** The dataset provided by the Tunisian Ministry of Finance includes variables about tax analysis, taxpayer inspection and VAT returns. The training dataset provided here is a subset of over 25,000 samples aggregated by year.



République Tunisienne

**Ministère des Finances**

# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## STEG Fraud Detection in Electricity & Gas Consumption

Predictive Analytics Challenge

**Problem statement:** The Tunisian Company of Electricity and Gas (STEG) is a public and a non-administrative company, it is responsible for delivering electricity and gas across Tunisia. The company suffered tremendous losses in the order of 200 million Tunisian Dinars due to fraudulent manipulations of meters by consumers.

**Objective of the challenge:** The aim of the challenge is to detect and recognize all clients involved in fraudulent activities using the client's billing history.

**Relevance of the solution:** The solution will enhance the company's revenues and reduce the losses caused by all fraudulent activities.

**Evaluation Metrics:** The metric to be used is AUC (Area Under Curve)

**Data:** The data provided by STEG is composed of two files. The first one is comprised of client data and the second one contains billing history since 2005.



الشركة التونسية  
لل كهرباء والغاز

Société Tunisienne  
de l'Electricité et du Gaz



# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Tunisian Vehicles License Plates Detection and Recognition

Computer Vision Challenge

**Problem statement:** Vehicle License plate detection and recognition is a well known challenge that has been tackled by many computer vision labs and companies. However, each country has its specific license plates formats. This challenge is targeting regular Tunisian license plates.

**Objective of the challenge:** The objective of this challenge is to detect the vehicles license plates then recognize the characters in each license plate.

**Relevance of the solution:** This challenge was designed by InstaDeep in partnership with the Ministry of Interior, specifically for the AI Hack Tunisia 2019.

**Evaluation Metrics:** The metric used for this challenge is the Multi class Log Loss. This metric is used to evaluate the error for each number in each license plate.

**Data:** The data provided for this challenge is composed of two datasets: A set of vehicle images (900 images) taken from the internet and annotated manually. The annotations are the coordinates of the bounding box containing the license plate. A set of license plate images (900 images) where the annotations are the text written in the license plate. In the test set, we provide 213 images of cars where each image contains only one car and one license plate. The submission file will have N times 7 rows, where N is the number of images and multiplied by seven because the license plate is composed of two numbers the first one contains at most 3 digits and the second one at most 4 digits. If the first number on the image for example contains only two digits then the first digit should be filled with zero.



# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Bills Classification

Computer Vision Challenge

**Problem statement:** Expensya is a Cloud based multi platform expense management software. The web and mobile solution covers the entire travel & expense journey: before, during and after each business trip.

One of the main features of Expensya is the automatic extraction of information out of the receipts. Currently, this is done with OCR+.

**Objective of the challenge:** The objective of this challenge is to create a deep learning model to classify images of receipts by the category of the expenses.

**Relevance of the solution:** With the emergence of relevant deep learning techniques, Expensya would like to explore the performance of Deep Learning to improve one of the main features of the application, which is the automatic classification of the receipts into different categories, right after snapping a picture of the receipt.

**Evaluation Metrics:** The evaluation metric for this challenge is the multi class Log Loss.

**Data:** The data is composed of a set of receipt images that belong to 4 categories: restaurants, parking, fuel and transport.

Each category contains images of receipts in different languages (English, French, German, Italian and Spanish)



# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Flight Delay Prediction

Predictive Analytics Challenge

**Problem statement:** Flight delays hurt airlines, airports, and passengers. Their prediction is crucial during the decision-making process for all players of commercial aviation. So, the development of accurate prediction models for flight delays will be very beneficial to prevent delays in such cases based on the airlines flight data.

**Objective of the challenge:** This competition aims to predict the type of the delay from five existent ones: Type 1, Type 2 ... Type 5.

**Relevance of the solution:** Proposing solution approaches to build flight delay prediction models from the Data Science perspective.

**Evaluation Metrics:** The evaluation metric for this challenge is the Log Loss.

**Data:** The data is composed of a set of flight details like scheduled departure time, scheduled arrival time, ID of the plane, ID of the flight,...



# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## Team / Startup CHALLENGES

Participants can develop innovations using Machine Learning & AI revolving around (but not limited to) one of these areas:

- Consumer Products / Services
- Tourism / Travel / Transport
- Healthcare / Nutrition / Wellness / Environment
- Education / E-learning / Training / Jobs
- Finance / Payments / eCommerce
- Telecommunications / Cybersecurity / High Tech

## WHAT SHOULD I BRING?

Everything you think is necessary to present your project

- Laptop + charger
- Smart phone
- And any extra tech you might need
- Passion!

**P.S. Please note that all external hardware brought to the event is under the participants' responsibility.**

## KEY DEADLINES

Individual Track: Friday 11pm first submission. Saturday ~5am final submission

Team Deadline 2 - Saturday 21st ~15:00 - idea is locked and documented

Team Deadline 3 - Sunday 21st 8:00 - Projects idea & code is "locked"

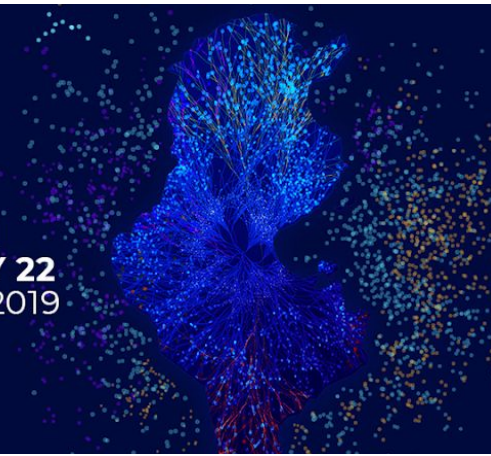
## ATTENDANCE

- Only participants are allowed to attend
- Not Allowed to bring friends or family members



# AI HACK TUNISIA 2019

FRIDAY 20,  
SATURDAY 21  
AND SUNDAY 22  
SEPTEMBER 2019



## [FREQUENTLY ASKED QUESTIONS - FAQs](#)

### PARTICIPANT RIGHTS & DUTIES

#### EXPLOITATION RIGHTS AND INTELLECTUAL PROPERTY

- Exploitation and property rights to prototypes developed during the Hackathon are reserved to Participants.
- During the Hackathon, the Participants shall only use elements which are clear of any rights.
- Any third-party elements included in the prototype, including open source software, must be clearly identified with their version, as well as the terms of the applicable license and any other details regarding their use.
- The Participant understands that this information is taken into account in the assessment for the allocation of prizes.
- The prototypes shall not infringe any intellectual property rights or images right.
- Each Participant warrants to the Organizer that their creation does not infringe in any way any third-party rights, and that he/she obtained, where necessary, the authorization of any third party.
- The Participants may be filmed or photographed during the event. By joining the Hackathon, the Participants agree to the use and the distribution of their image by the Organizer, including promotional events occurring after the Hackathon.
- The production and distribution of films and photographs of the event will not lead to any compensation of the Participants.
- We reserve the right to remove any rewards if we discover forgery or mis-representation of data even if this happen after the event

**Please reread the participation agreement which is attached in the email sent to you and you'll be signing it upon your arrival.**