

Moscow Institute of Physics and Technology sacré champion du monde étudiant de la Data Science lors du Data Science Game

Après une ultime épreuve d'hackathon, l'équipe **Russian Data Mafia** de MIPT termine en première position et décroche le titre de champion du monde étudiant de la data science !



La finale du Data Science Game, compétition internationale étudiante dédiée aux Big Data s'est tenue les 10 et 11 septembre derniers, au Château Les Fontaines, près de Paris. 80 étudiants provenant des 20 meilleures formations mondiales en Data Science se sont affrontés dans une course au code effrénée pour bâtir le meilleur algorithme. Le challenge consistait à mettre les données et modèles de machine learning au service d'une problématique métier de l'assurance.

Le Deep Learning à l'honneur

Pour marquer l'ouverture de la finale, les participants venus du monde entier ont d'abord été accueillis à Paris, au sein des locaux de Microsoft France, lors d'un meetup de la communauté internationale de Data Science organisé pour l'occasion. Cet événement a trouvé son apogée lors d'une table ronde inédite consacrée au *feature engineering* et au *deep learning*. Le débat,

animé par Isabelle Guyon, présidente de ChaLearn, opposait des ténors des branches analytics des trois principaux partenaires du Data Science Game : Capgemini Consulting, Microsoft, et Axa.

A l'issue de conférence, étudiants et partenaires ont pu librement échanger sur le sujet autour d'un cocktail.

Le Big Data et le machine learning au service des exigences métier

Grâce à un accès libre aux puissants clusters de calcul Azure de Microsoft, et aux conseils avisés des mentors des partenaires, les meilleures conditions étaient réunies pour résoudre le problème posé cette année par AXA. Après un marathon de 30 heures presque ininterrompues, les compétiteurs sont parvenus à des solutions tout aussi efficaces qu'innovantes.

A partir d'un jeu de données volumineux sur les devis d'assurance en ligne, le challenge consistait à élaborer un modèle de concrétisation des offres.

A l'issue des 30h de compétition, les équipes ont été classées par leur score de prédiction. La mesure de performance utilisée était le log-loss, une mesure qui pénalise fortement les prédictions à la fois certaines (probabilité proche de 0 ou de 1) et fausses.

Un finaliste précise :

“Nous nous sommes sentis démunis après avoir entraîné un réseau de neurones pendant des heures et remarqué à quel point le feature engineering était important dans ce challenge.”

MIPT couronné

L'équipe **Russian Data Mafia**, du Moscow Institute of Physics and Technology, en Russie, se classe en première position, avec un score en log-loss de 0,008344.

6 équipes ont été distinguées à l'issue de la compétition :

- Prix Capgemini, pour la 1e place : Russian Data Mafia, Moscow Institute of Physics and Technology, Russie
- Prix AXA, pour la 2e place : Cantab, Cambridge University, Royaume-Uni
- Prix Microsoft, pour la 3e place : We just want our name to be the longest one, Skoltech University, Russie
- Prix Numberly, pour la 4e place : Jonquille, Université Pierre et Marie Curie, France
- Prix Milliman, pour la 5e place : Brosio2BeWild, University of Padova, Italie
- Prix QuantCube Technology, pour le prix du Jury : ml_noobs, Moscow State University, Russie

Les pays représentés dans cette finale étaient : la France, les Pays-Bas, la Russie, l'Allemagne, le Royaume-Uni, Singapour, les Etats-Unis, le Japon, l'Inde et l'Italie.

A propos

Le Data Science Game est une association qui promouvoit le développement de la data science et des compétences associées dans les domaines de l'informatique, des statistiques, de l'ingénierie, et des mathématiques appliquées. Notre équipe est composée principalement de data scientists bénévoles qui travaillent toute l'année pour faire du Data Science Game un challenge unique et mémorable pour les étudiants et les partenaires.