



Communiqué de presse

Le 30 juin 2015

Franc succès pour la première édition du Data Science Game

Les 20 et 21 juin 2015 a eu lieu la phase finale de la première compétition internationale étudiante autour des Big Data (ou mega-données) : le Data Science Game. Cette compétition est une initiative conjointe de ParisTech et trois de ses écoles : ENSAE ParisTech, Télécom ParisTech et ENSTA ParisTech. Elle vise à encourager les étudiants à mesurer leurs capacités et démontrer l'importance de ce domaine pour ParisTech et ses écoles sur la scène internationale. La compétition, qui vient de s'achever, a été un grand succès, tant auprès des étudiants que des partenaires.

Montée en partenariat avec Google France et Capgemini, l'un des leaders mondiaux du conseil, des services informatiques et de l'infogérance et soutenue par Sanofi-Aventis, Datasciences.net, Ekimetrics et la Fondation ParisTech, cette première édition du Data Science Game a remporté un franc succès en réunissant 20 équipes internationales d'étudiants (France, Allemagne, Italie, Russie, Inde, Pays-Bas, Royaume-Uni, Irlande) dans l'objectif de trouver des solutions à un problème complexe en s'appuyant sur l'analyse de données massives, à travers l'élaboration d'algorithmes de traitement capables de gérer et de comprendre ces données.

La compétition s'est déroulée en deux parties : chaque phase comportait un sujet spécifique avec son propre jeu de données.

Le Data Science Game 2015 en chiffres :

- 20 équipes de 4 étudiants.
- 8 pays.
- 13 nationalités.
- 11 Grandes Écoles et Universités françaises et 9 Universités étrangères
- 5 Écoles de ParisTech : MINES ParisTech, École Polytechnique, ENSTA ParisTech, ENSAE ParisTech, Télécom ParisTech
- 5000 € de prix
- 5 pays lauréats
- 6 sponsors : Google, Capgemini, Sanofi, Datascience.net, Ekimetrics, Fondation ParisTech

La première partie, non compétitive, s'est déroulée du 15 mai au 15 juin, via la plateforme datascience.net. Cette phase amicale devait permettre aux étudiants de se familiariser avec les données.

Durant cette première épreuve, les étudiants ont reçu un jeu de données composé d'extraits de livres. Ce jeu de données, ou jeu d'apprentissage, était scindé en 2 colonnes avec, d'un côté, les textes et de l'autre les auteurs correspondants. Les équipes devaient construire un modèle et des algorithmes pour identifier les auteurs de ces textes parmi William Shakespeare, Mark Twain, Oscar Wilde, Edgar Allan Poe, Jane Austen et Arthur Conan Doyle. L'efficacité de leurs modèles était ensuite évaluée sur datascience.net à l'aide d'un jeu de test. Le score ainsi obtenu correspondait au taux de bonnes réponses du modèle.

12, villa de Lourcine - 75014 Paris - France
Tél. (+33) 01 45 81 84 34 - Siret 13000368400037

www.paristech.fr



La deuxième partie, compétitive, s'est déroulée les 20 et 21 juin 2015, au château des Fontaines, le centre international de formation de Capgemini situé près de Chantilly. Cette fois-ci, il a été demandé aux participants de classer des vidéos Youtube en 15 catégories, chaque vidéo étant associée à une seule catégorie. Leur but : utiliser l'ensemble des données à leur disposition pour classer ces vidéos (titres, descriptions, durées, commentaires, dates, formats, etc.). La difficulté reposant sur des problèmes de syntaxe, d'orthographe, de langue, etc.

Pour ce faire, les équipes avaient à leur disposition un jeu d'entraînement de 240 000 vidéos. Chaque vidéo étant décrite par une quinzaine de variables et les catégories associées sur le modèle des données de la première phase. Comme lors de la phase amicale, les équipes devaient élaborer des modèles à partir de ce jeu d'apprentissage puis les soumettre au site datascience.net pour vérifier leur justesse à l'aide d'un jeu de données test.

À la fin de la compétition, les équipes ont choisi l'un de leurs modèles pour une ultime soumission à l'aide d'un nouveau jeu de données inconnu de tous et dont les résultats ont permis l'élaboration du palmarès.

C'est l'équipe Russe de l'Université d'état de Moscou (MSU) qui a remporté cette première édition avec un score de plus de 75% :

1. **Russie** : Université de Moscou
2. **Italie** : Université de Rome
3. **France** : Télécom ParisTech
4. **Pays-Bas** : Université d'Amsterdam
5. **Royaume-Uni** : Imperial College

Le succès de cette première édition constitue une base solide qui va permettre de bâtir les éditions suivantes et faire du Data Science Game une compétition de référence à l'échelle mondiale dans le domaine des Big Data.

À propos de ParisTech

ParisTech est une structure de coopération inter-écoles, œuvrant à l'élaboration de projets communs articulés autour de la formation, de la recherche, et de l'innovation dans le domaine des sciences, des technologies et du management. L'Institut des Sciences et Technologies de Paris, dénommé ParisTech, est un Établissement public de coopération scientifique (EPCS) depuis 2007. La marque ParisTech s'appuie sur deux socles: la formation Grande École à la française et la pluridisciplinarité au service de l'entreprise. ParisTech est constitué de 12 Écoles membres et dispose d'une forte interaction au niveau international au travers de nombreux accords de partenariat.

CONTACT PRESSE : Jacques BRINGUEZ – 01.45.81.84.39 – jacques.bringuez@paristech.fr