



## Ligue des Champions

Et si 2 élèves ingénieurs connaissaient depuis quelques semaines le futur vainqueur grâce à la Data Science ?



Cyril Tso et Tsiory Razafindrazaka sont deux étudiants en dernière année à l'[ESME Sudria](#). Depuis l'année dernière, ils se sont spécialisés dans l'univers des Data Science en choisissant la majeure Big Data & IoT. Dans le cadre de l'[ESME SPEED](#), le pitch des meilleurs projets de fin d'étude des étudiants de l'ESME Sudria, ils ont présenté un projet pour tenter de **prédire l'identité du club de football qui soulèvera le trophée de la Ligue des Champions** en juin prochain. Ce fut pour eux, un défi particulièrement amusant à relever !

### Pourquoi ce choix de projet ?

Tsiory et Cyril sont **passionnés de foot** et ils avaient envie de travailler sur un projet de fin d'études mêlant leur passion aux compétences acquises sur le traitement de données.

### Leur prédiction ?

**Manchester City**, une prédiction qui n'est finalement pas très éloignée de celle des pronostiqueurs !

### Comment s'y sont-ils pris ?

Ils ont mis en place une démarche en 4 étapes :

- **Le Data Mining** pour extraire les données issues de sites internet de statistiques footballistiques français et internationaux
- **Le Dataframing** pour la transformation et le stockage de ces données
- Le nettoyage et l'exploration des données pour s'assurer de leur fiabilité
- Pour finir, du **Machine learning**, sous-domaine de l'**intelligence artificielle**, pour mettre en place la solution sur la base des anciennes éditions.

### Comment se sont-ils assurés de la justesse de leur modèle ?

Cyril et Tsiory ont testé différents modèles comme le KNN (pour k-nearest neighbors), Random Forest, EdgeBoost mais ils ont aussi entraîné leur modèle sur les anciennes éditions... Pour obtenir un résultat cohérent.

Leur plus grande difficulté ? La collecte de données. En effet, à l'inverse du basket il n'y avait pas un grand historique de données normalisées sur le foot car, avant 2011-2012, le football n'était pas **data-driven**. Cyril et Tsiory ont donc **scrapé\*** trois grands sites de statistiques : **WhoScored.com**, **FootyStats.org** et **ESPN.com**.

\* Extraire des données d'une page Web à l'aide d'un programme automatisé

Retrouvez l'interview complète de Cyril et Tsiory en cliquant sur ce [lien](#).

**Contact média**

Emmanuelle de Coudenhove (Directrice Marketing et Communication ESME Sudria)

[emmanuelle.de-coudenhove@esme.fr](mailto:emmanuelle.de-coudenhove@esme.fr)

06.73.64.27.20

**ESME Sudria**

40 rue du Docteur Roux

75015 Paris

**À propos de l'ESME Sudria :**

Fondée en 1905, l'ESME Sudria forme en 5 ans des ingénieurs pluridisciplinaires, prêts à relever les défis technologiques du XXI<sup>e</sup> siècle : la transition énergétique, les véhicules autonomes, les réseaux intelligents, les villes connectées, la cyber sécurité, les biotechnologies, la santé et l'Intelligence Artificielle. L'école est présente à Paris, Bordeaux, Lille et Lyon. Trois composantes font la modernité de sa pédagogie : l'importance de l'esprit d'innovation ; l'omniprésence du projet et de l'initiative ; une très large ouverture internationale, humaine et culturelle.

Retrouvez l'actualité de l'ESME Sudria sur sa [Newsroom](#).

<http://www.esme.fr/>

