



Qui gagnera le match?

Prédiction de l'équipe gagnante lors d'un match de football.

Johann Hospice et Jérémy Phalente



Dataset

- Source: kaggle.com/hugomathien/soccer
- Une base de donnée de football pour l'analyse de donnée
- 25k+ matchs, joueurs et équipes pour le football européen professionnel





Nettoyage

Parsing

- Un parser python utilisant sqlite3
- Enregistrement des données dans un .csv

Nettoyage

- Filtrage des données erronées +300 / +25k

Encoding

- Peu ou inexistant (simple cast)

Mise en forme

```
Equipe: {  
    buildUpPlaySpeed,  
    buildUpPlayPassing,  
    chanceCreationPassing,  
    chanceCreationCrossing,  
    chanceCreationShooting,  
    defencePressure,  
    defenceAggression,  
    defenceTeamWidth  
}  
Match: {butE1, butE2, (E1), (E2)}  
Y: butE1 > butE2
```

Algo / Modele

```
def distance(data1, data2):  
    """ distance euclidienne entre 2 matchs  
    """  
    return math.sqrt(sum([math.pow(data1[i] - data2[i], 2) for i in range(16)]))  
  
def k_nearest_neighbors(x, match, dist_function, k):  
    """Retourne la liste des matchs les plus proches  
    """  
    distance_and_indices = [(dist_function(match[i], x), i) for i in range(len(match))]  
    distance_and_indices.sort()  
    return [di[1] for di in distance_and_indices[:k]]  
  
def match_result_knn(x, train_x, train_y, dist_function, k):  
    """predit le resultat d'un match  
    """  
    neigh = k_nearest_neighbors(x, train_x, dist_function, k)  
    num_win = sum([1 for i in neigh if train_y[i]])  
    return num_win > k - num_win
```

- Méthode des K plus proches voisins
- Distance euclidienne sur des points en 16D



Train / Validation

- Utilisation de cross validation pour pouvoir entraîner nos ensembles de tests
- Module StratifiedKFold de python
- Séparation des données en ensemble d'entraînement et de test

0.41 % d'erreurs

424 matchs de foot

59 voisins



Discussion

- 8 attributs décrivant une équipe
- 16 attributs décrivant un matchs
- 1 sortie booléenne
- Gagner ou perdre le match

Pour aller plus loin: **Quel sera le score du match?**

Merci de votre attention!

Jérémy Phalente et Johann Hospice