



# MLG : Football



Agoulif Youssef  
Ferroni Sandro  
Mouti Abdelhamid  
Moyo-Kamdem Auren

# Présentation des données

Librairie : **statsbomb**

- compétitions
- plusieurs saisons
- plusieurs matchs
- stats des matchs

Choisi : Compétition Liga, Saison 2015/2016, Equipe Barcelona





# 1. Prédire la victoire ou la défaite en fonction de graphe de passe

## 1. Données Utilisées :

- Matches de La Liga 2015-2016
- Événements de passe (passeur, récepteur, positions, type, résultat)

## 2. Construction des Graphes de Passes :

- Joueurs comme nœuds
- Passes comme arêtes

## 3. Mesures Clés :

- Centralité de degré
- Centralité de proximité
- Centralité d'intermédierité
- PageRank



# 1. Prédire la victoire ou la défaite en fonction de graphe de passe

## 4. Extraction des Features :

- Centralité et positions des joueurs

## 5. Modélisation :

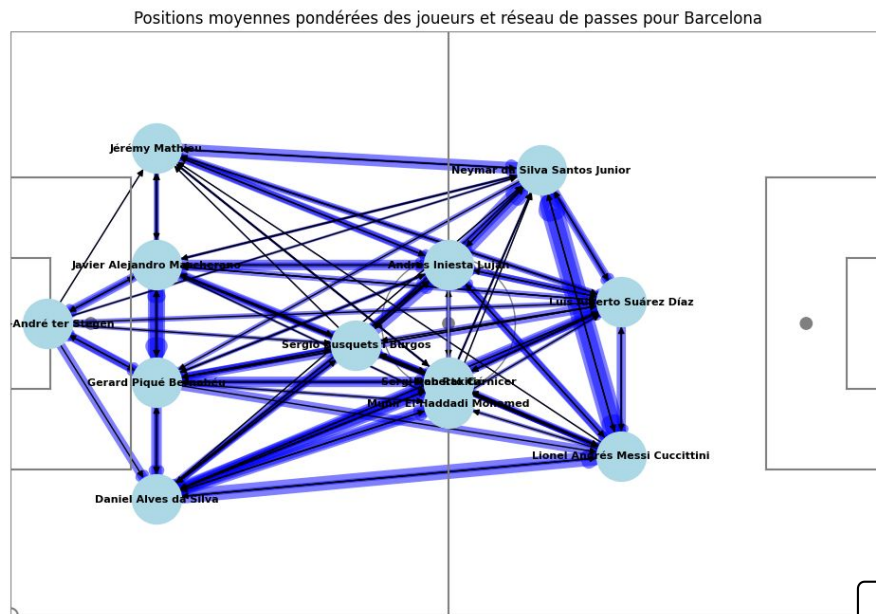
- RandomForestClassifier

## 6. Évaluation :

- Précision du modèle (accuracy)
- Prédiction des victoires, défaites, nuls

## 2. Identification de la composition d'équipe lors de victoires

1. Récupération des matchs de Barcelone sur la Liga 2015/2016
  - a. 38 matches
2. Pour chaque match: construction d'un graphe avec la position moyenne des postes, pondéré par le temps passé
  - a. position X et position Y





## 2. Identification de la composition d'équipe lors de victoires

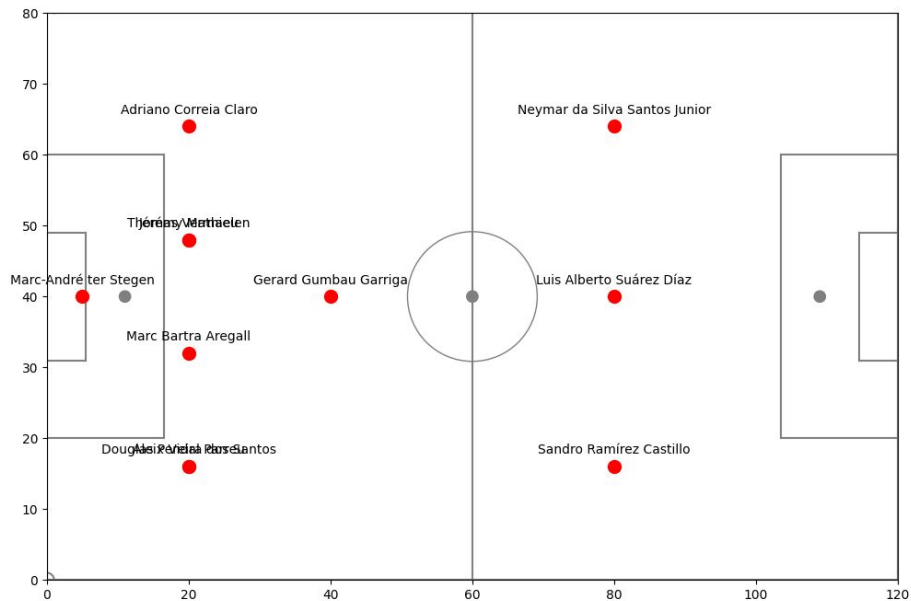
3. Transformation des graphes en tableau
4. Séparation en Trainset et Testset (70%, 30%)
5. Entraînement sur plusieurs modèles avec GridSearch
  - RandomForest
  - SVM
  - KNN
  - DecisionTree

Modèle	Accuracy	Precision	Recall
RandomForest	0.67	0.44	0.67
SVM	0.67	0.44	0.67
KNN	0.67	0.48	0.67
DecisionTree	0.58	0.47	0.58



## 2. Identification de la composition d'équipe lors de victoires

6. Affichage des joueurs ayant le plus d'influence sur la victoire

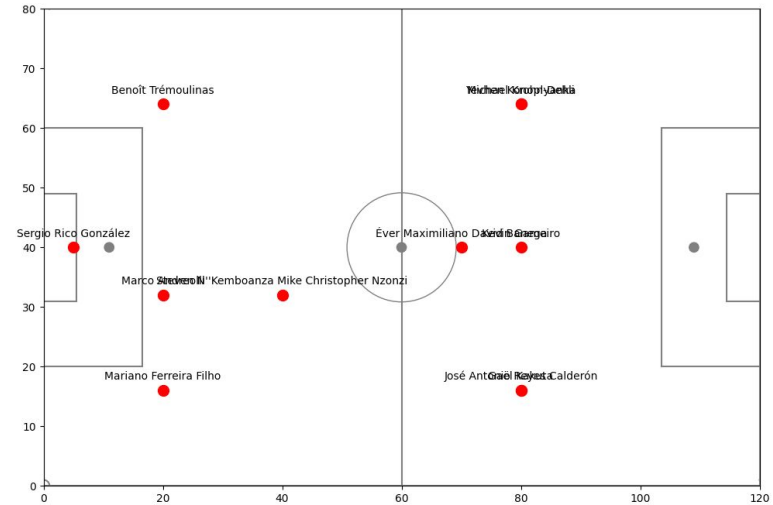




## 2. Identification de la composition d'équipe lors de victoires

### 7. Essai sur Sevilla : statistiques plus équilibré

Modèle	Accuracy	Precision	Recall
RandomForest	0.50	0.55	0.50
SVM	0.25	0.57	0.25
KNN	0.25	0.27	0.25
DecisionTree	0.50	0.63	0.50

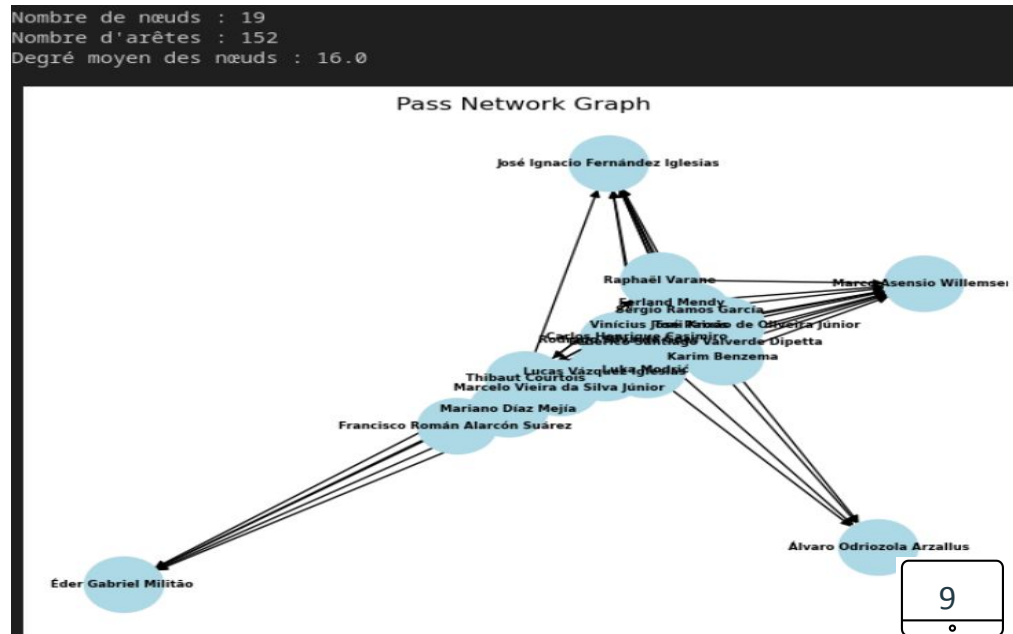






### 3. Identification et prédiction des joueurs majeurs dans une équipe d'une saison à l'autre

1. Utilisation des données d'events pour filtrer les passes effectuées par une équipe pendant des intervalles de temps spécifiques.
2. Construction d'un graphe dirigé où les nœuds représentent les joueurs et les arêtes représentent les passes entre les joueurs.





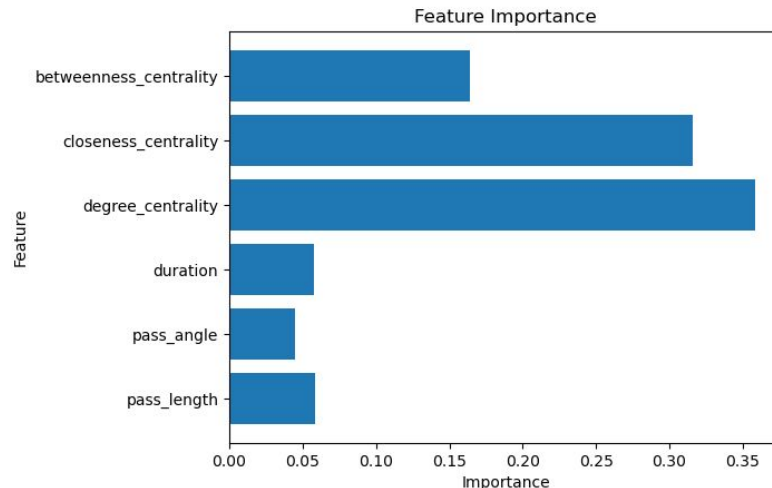
### 3. Identification et prédiction des joueurs majeurs dans une équipe d'une saison à l'autre

1. Utilisation du graphe créé pour créer un réseau de passe et calculer les mesures de centralité telles que la centralité degré, la centralité de proximité et la centralité d'intermédiarité pour chaque joueur.
2. Fusion des caractéristiques des passes et des mesures de centralité dans un seul dataframe pour chaque saison.
3. Entraînement d'un modèle random Forest pour prédire les joueurs majeurs basés sur les caractéristiques des passes et les centralités.

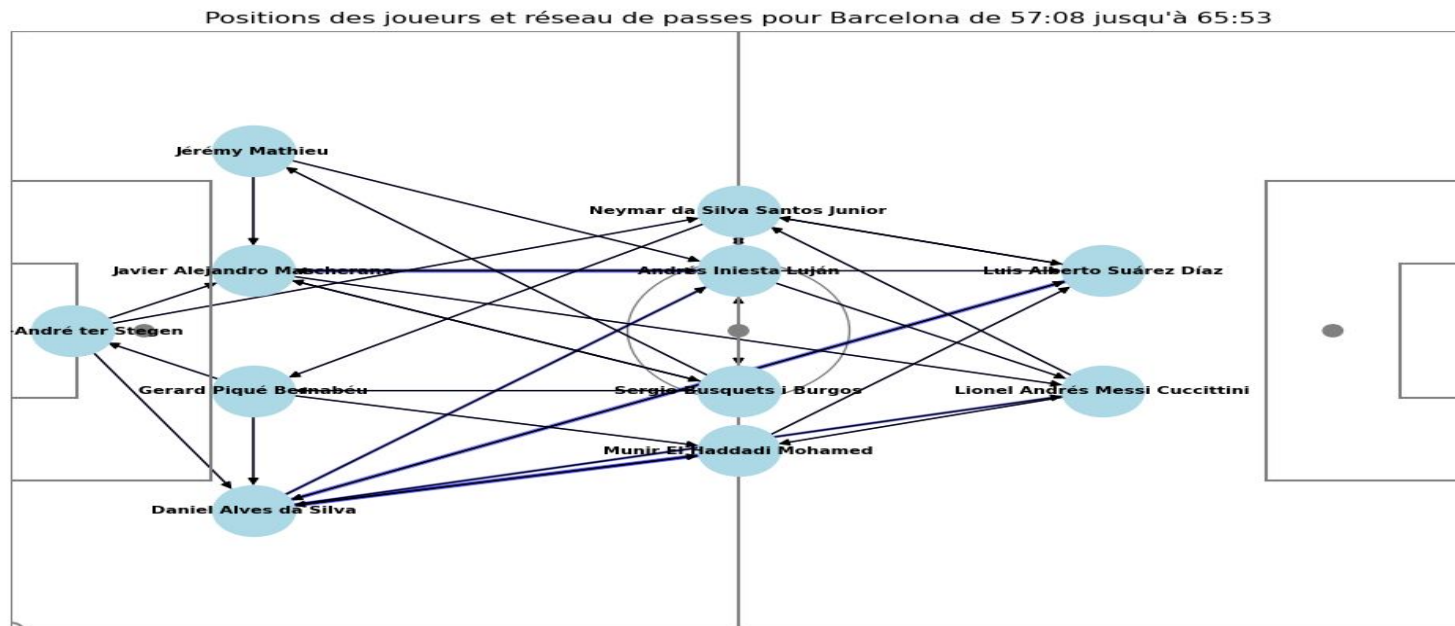
	precision	recall	f1-score	support
0	0.79	1.00	0.88	11
1	1.00	0.25	0.40	4
accuracy			0.80	15
macro avg	0.89	0.62	0.64	15
weighted avg	0.84	0.80	0.75	15

### 3. Identification et prédiction des joueurs majeurs dans une équipe d'une saison à l'autre

1. Utilisation du graphe créé pour créer un réseau de passe et calculer les mesures de centralité telles que la centralité degré, la centralité de proximité et la centralité d'intermédiarité pour chaque joueur.
2. Fusion des caractéristiques des passes et des mesures de centralité dans un seul dataframe pour chaque saison.
3. Entraînement d'un modèle random Forest pour prédire les joueurs majeurs basés sur les caractéristiques des passes et les centralités.



### 3. Identification et prédiction des joueurs majeurs dans une équipe d'une saison à l'autre





## 4. Prédire si un tire va marquer

- 1 saison, 38 matchs, équipe: Barcelona
- 604 tirs en total
- 109 buts marqués
- prédiction de buts marqués à partir de tirs (classification binaire, supervised)

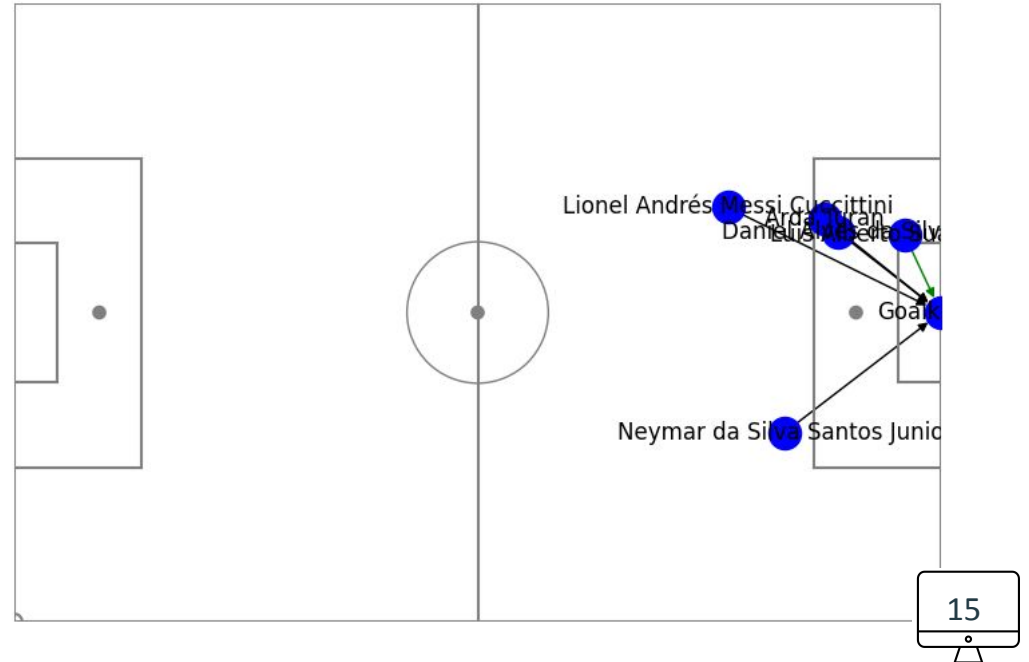


## 4. Prédire si un tire va marquer

	shot_id	match_id	player_name	shot_time	shot_position	player_shot_count	team_shot_count	shot_body_part	shot_technique	shot_type	is_goal
0	4ade6c91-08c8-41f9-bdaf-e6dc5287a2a8	3825660	Luis Alberto Suárez Díaz	00:06:36.134	[99.6, 53.3]	0	0	Right Foot	Half Volley	Open Play	False
1	61e1070a-d214-4411-83c7-d50275d14d5a	3825660	Luis Alberto Suárez Díaz	00:07:58.135	[104.1, 30.9]	1	0	Left Foot	Half Volley	Open Play	False
2	8657b856-7061-4cea-912e-36670704f22c	3825660	Daniel Alves da Silva	00:18:10.272	[95.7, 44.3]	0	2	Right Foot	Volley	Open Play	False
3	7e075db9-7f56-440d-87b6-02a01cf92881	3825660	Jérémy Mathieu	00:21:52.413	[109.9, 35.9]	0	3	Head	Normal	Open Play	False
4	def121ed-b09c-4e8b-a1a5-51882665a6fc	3825660	Andrés Iniesta Luján	00:23:58.109	[114.9, 27.9]	0	4	Left Foot	Normal	Open Play	False
...	...	...	...	...	...	...	...	...	...	...	...
599	b5c5023a-330b-4cfe-af52-07651c543e35	266557	Neymar da Silva Santos Junior	00:24:54.700	[107.2, 22.0]	3	14	Left Foot	Normal	Open Play	False
600	b0f16225-8f81-4553-8d92-ed0c79a2c410	266557	Ivan Rakitić	00:29:16.944	[105.8, 40.0]	2	16	Left Foot	Normal	Open Play	False
601	297bd1b8-04ab-478a-9714-86b945f9ec40	266557	Neymar da Silva Santos Junior	00:38:37.558	[110.3, 34.4]	4	15	Right Foot	Half Volley	Open Play	False
602	e5965a61-bbfe-4746-a037-a804c6423379	266557	Gerard Piqué Bernabéu	00:39:19.887	[106.2, 40.1]	0	20	Right Foot	Normal	Open Play	False
603	86e27aef-55ee-4709-b0ba-3faffca05ddc	266557	Gerard Piqué Bernabéu	00:43:49.008	[111.5, 47.0]	1	20	Right Foot	Normal	Open Play	False

# Affichage des tirs en graph

- Affichage des tirs, dans 38 matches.
- 1 graph par match
- arcs (noirs: raté, vert: marqué)
- Exemple:





# Extraction de features

```
[250] features_df
```

	num_nodes	num_edges	avg_degree	shot_distance	player_shot_count	team_shot_count	shot_body_part_Left Foot	shot_body_part_Other	shot_body_part_Right Foot	shot_technique_Half Volley	shot_tech
0	8	7	1.750000	24.352618	2	6	0.0	0.0	1.0	1.0	
1	8	7	1.750000	24.352618	1	0	1.0	0.0	0.0	1.0	
2	8	7	1.750000	8.626703	0	2	0.0	0.0	1.0	0.0	
3	8	7	1.750000	18.314202	0	3	0.0	0.0	0.0	0.0	
4	8	7	1.750000	11.926860	0	4	1.0	0.0	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...
599	7	6	1.714286	9.689685	3	14	1.0	0.0	0.0	0.0	
600	7	6	1.714286	14.200000	2	16	1.0	0.0	0.0	0.0	
601	7	6	1.714286	9.689685	4	15	0.0	0.0	1.0	1.0	
602	7	6	1.714286	24.090247	2	9	1.0	0.0	0.0	0.0	



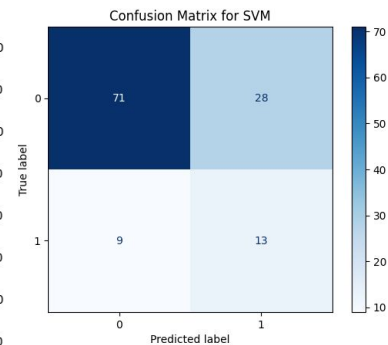
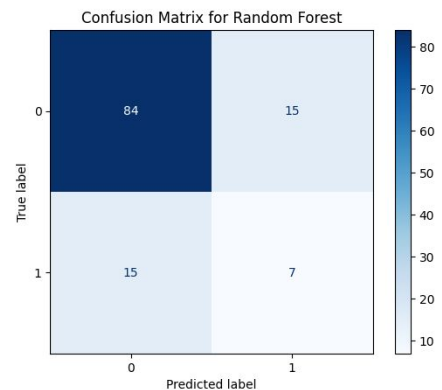
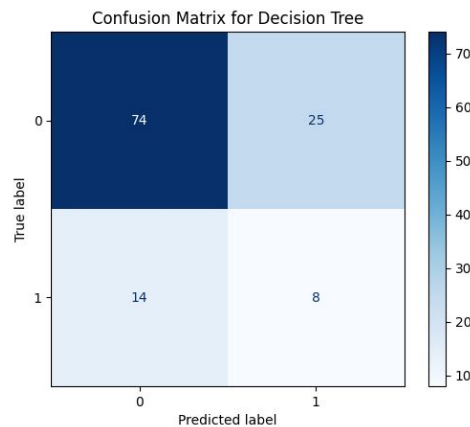
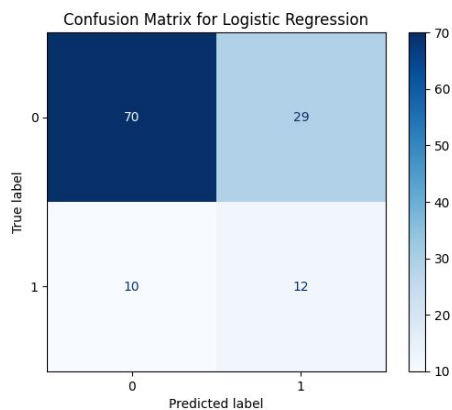


# Application en Machine Learning

- Data split en train et test (80% et 20%)
- Entraîner les données sur des modèles de classification (logistic regression, decision tree, random forest, svm)
- Meilleurs hyper-paramètres avec grid search.



# Résultats





# Résultats

	Model	Accuracy	Precision	Recall	F1 Score	Best Params
0	Logistic Regression	0.677686	0.292683	0.545455	0.380952	{'C': 10, 'solver': 'liblinear'}
1	Decision Tree	0.677686	0.242424	0.363636	0.290909	{'max_depth': 5, 'min_samples_split': 5}
2	Random Forest	0.752066	0.318182	0.318182	0.318182	{'max_depth': 5, 'min_samples_split': 10, 'n_e...
3	SVM	0.694215	0.317073	0.590909	0.412698	{'C': 10, 'kernel': 'linear'}

Best model is: Random Forest with f1 score of 0.3181818181818182, accuracy of 0.7520661157024794 and best parameters: {'max\_depth': 5, 'min\_samples\_split': 10, 'n\_estimators': 200}



# M<sub>(LG)</sub>erci !

Agoulif Youssef

Ferroni Sandro

Mouti Abdelhamid

Moyo-Kamdem Auren

