



LES PREDICTIONS SPORTIVES DANS L'OPEN DATA

DEVLAMINCK Johan
DA SILVA Yoann
ANGELI kevin



SOMMAIRE

- Introduction
- L'Open Data pour les prédictions sportives
- L'exploration des données dans le paris Sportifs
- Nos démarches & Notre Application

INTRODUCTION

- Ouverture des données :
 - Libre disponibilité et accessibilité
 - Réutilisation et redistribution
 - Une participation universelle
- 1ere Analyse de données sportives :
 - *Percentage Baseball*, Ernshaw Cook 1964
- Utilisation de l'analyse prédictive :
 - Les tigres de Leicester





L'OPEN DATA pour les prédictions sportives

PROVENANCE DES SOURCES

- Manque de données liées aux paris sportifs en France
- Données liées aux paris sportifs disponible sur des sites anglais
- Pas de sources “officielles”
- Données dynamiques

LES UTILISATEURS

→ Parieurs amateur / professionnels



→ Cabinet de conseil sportifs



→ Analystes



EXEMPLES D'APPLICATION

BetGPS

Outil d'analyse détaillé
d'un jeu de donnée
sportif d'une ligue de
football

Figure 1 : page d'accueil de l'outil

<http://www.betgps.com/betting-data.php>
Le 4/11/2018

England Championship 2013/2014 - LEAGUE SUMMARY

Completeness		Matches played		Remaining matches	
Total match count		Count	%	Count	%
552		552	100.0	0	0.0

GOALS	FULL TIME		First half		Second half	
	Count	Avg. per match	Count	Avg. per match	Count	Avg. per match
Home goals scored	793	1.44	345	0.63	448	0.81
Away goals scored	641	1.16	298	0.54	343	0.62
Total goals scored	1 434	2.60	643	1.16	791	1.43

FULL TIME RESULT	Match count	%	Average odds		Average winning odds		Fair odds
			Pinnacle	Bet365	Pinnacle	Bet365	
Home win	228	41.3	2.41	2.40	2.20	2.19	2.42
Draw	156	28.3	3.60	3.59	3.57	3.56	3.54
Away win	168	30.4	3.76	3.72	3.19	3.17	3.29
Overround, %			95.8	96.4	104.9	105.2	100.0

FIRST HALF RESULT	Match count	%	Fair odds
Home win	175	31.7	3.15
Draw	234	42.4	2.36
Away win	143	25.9	3.86

SECOND HALF RESULT	Match count	%	Fair odds
Home win	213	38.6	2.59
Draw	188	34.1	2.94
Away win	151	27.4	3.66

EXEMPLES D'APPLICATION

FiveThirtyEight

Probabilité d'obtenir le bon vainqueur d'un match ou un score nul

Figure 2 : vue générale de tous les matchs prévus

TIME (ET)	LEAGUE	MATCH	WIN	DRAW	LOSS	QUALITY The harmonic mean of the teams' SPI ratings	IMPORTANCE How much the result will alter the league projections	MATCH RATING The average of quality and importance
Sunday, Nov. 4 10:00 a.m.	La Liga 2 Spain	Oviedo vs Mallorca	43%	28%	29%	29	24	27
Sunday, Nov. 4 10:00 a.m.	Premier League England	Man. City vs Southampton	90%	8%	1%	78	52	65
Sunday, Nov. 4 10:00 a.m.	Primeira Liga Portugal	Feirense vs Tondela	44%	31%	25%	40	38	39
Sunday, Nov. 4 10:00 a.m.	Super League Switzerland	Basel vs Lugano	66%	19%	15%	50	41	45
Sunday, Nov. 4 10:00 a.m.	Super League Switzerland	Luzern vs Zürich	38%	27%	35%	49	32	41
Sunday, Nov. 4 10:00 a.m.	Super League Switzerland	Sion vs St. Gallen	54%	21%	24%	43	16	30
Sunday, Nov. 4 10:15 a.m.	La Liga Spain	Villarreal vs Levante	56%	22%	22%	71	27	49
Sunday, Nov. 4	Super League	Giannina	52%	26%		34	68	41

<https://projects.fivethirtyeight.com/soccer-predictions/matches/>

Le 4/11/2018

EXEMPLES D'APPLICATION

FiveThirtyEight

Probabilité qu'une
équipe remporte la
ligue, joue une phase
finale, ...

Figure 3 : vue générale des
différentes ligues et
championnats européens



UEFA CHAMPIONS LEAGUE 2018-19 season

TEAM	SPI RATING	MAKE FINAL	WIN FINAL
Man. City 6 pts	94 . 3	37%	21%
Barcelona 9 pts	91 . 7	25%	13%
Bayern Munich 7 pts	91 . 4	22%	12%

See all for UEFA Champions League



UEFA EUROPA LEAGUE 2018-19 season

TEAM	SPI RATING	MAKE FINAL	WIN FINAL
Chelsea 9 pts	87 . 6	24%	14%
PSG	90 . 1	14%	8%
Arsenal 9 pts	81 . 6	13%	7%

See all for UEFA Europa League



PREMIER LEAGUE 2018-19 season

ENGLAND

TEAM	SPI RATING	MAKE UCL	WIN LEAGUE
Man. City 26 pts	94 . 3	>99%	67%
Liverpool 27 pts	90 . 8	94%	22%
Chelsea 24 pts	87 . 6	81%	8%

See all for Premier League



LA LIGA 2018-19 season

SPAIN

TEAM	SPI RATING	MAKE UCL	WIN LEAGUE
Barcelona 24 pts	91 . 7	97%	68%
Real Madrid 17 pts	90 . 5	85%	17%
Atletico Madrid 20 pts	86 . 7	63%	7%

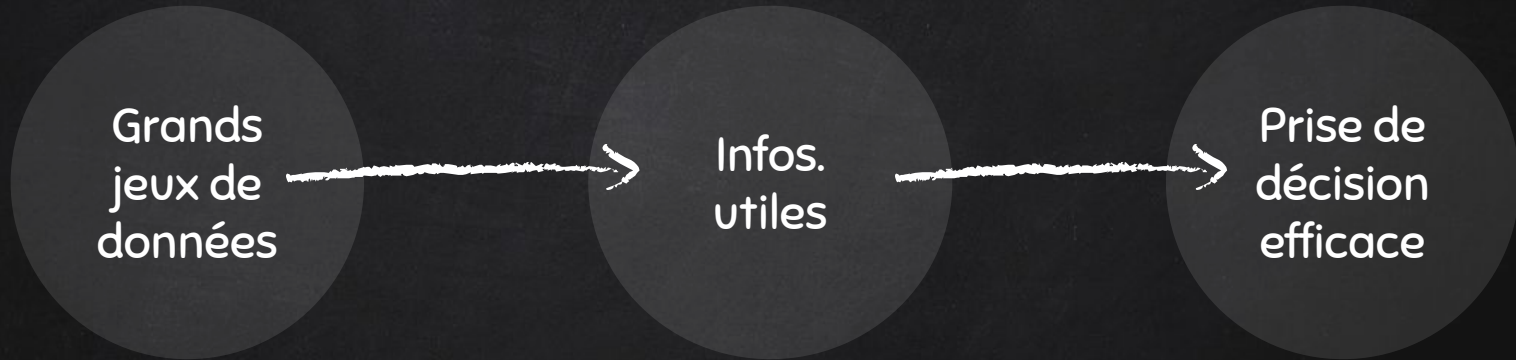
See all for La Liga



EXPLORATION DES DONNÉES DANS LES PARIS SPORTIFS



DATA MINING



But recherché dans les prédictions sportives :
-Un point de vue statistique



CAS D'UTILISATION DES STATISTIQUES : VALUE BET

Victoire Equipe 1	Résultat	Victoire Equipe 2
Victoire de Rennes	Nul	Victoire de Nantes
1,9	1,4	1,5

Estimation:

Victoire Nantes 75%

Annonce
Bookmaker :

Victoire Nantes
66,67%

Figure 4 : Cote Annonce Bookmaker

Cote = $1/p$
 $1/1,5 = 66,67\%$
 $1/75\% = 1,33$



DES DONNÉES FIABLES ?

Statistique

- Opposé aux probabilités
- Part d'aléatoire
- Basé sur des données récentes

Data Mining

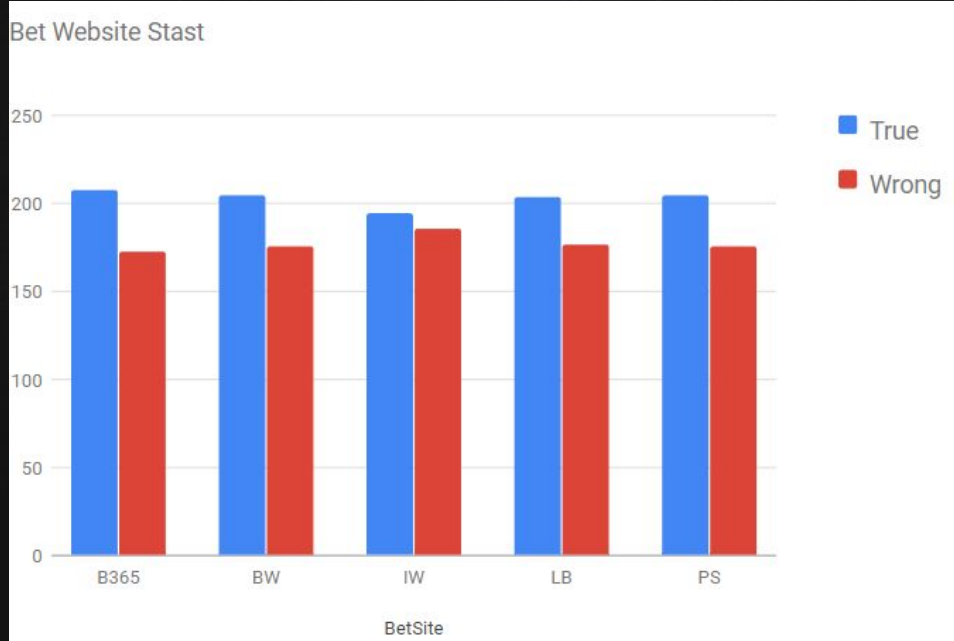
- Mauvaises interprétations
- Difficultés qualités des données
- Traitement de données : ETL et Algorithme



NOTRE APPLICATION



FONCTIONNEMENT DE L'APPLICATION



Ligue 1 saison 2017/2018

Total des matchs sur la saison : 381

Total des Victoire à Domicile : 173

Total des Matches Nuls: 96

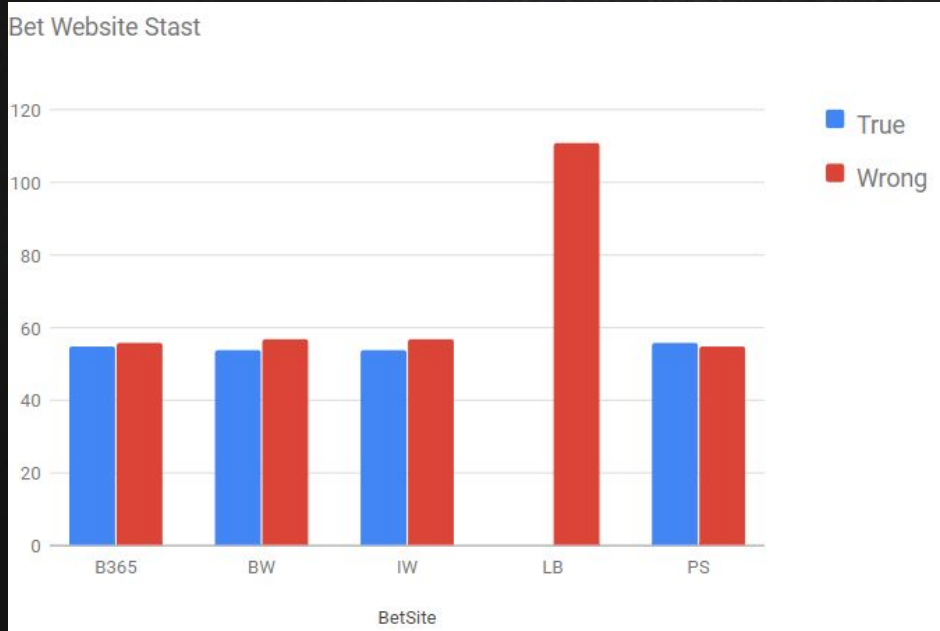
Total des Victoire à l'extérieur : 111

Total des cotes justes : 208

Figure 5 : Statistiques des Sites de Paris Sportif



FONCTIONNEMENT DE L'APPLICATION



Ligue 1 saison 2018/2019

Total des matchs sur la saison : 111

Total des Victoire à Domicile : 46

Total des Matches Nuls: 30

Total des Victoire à l'extérieur : 34

Total des cotes justes : 55

Figure 6 : Statistiques des Sites de Paris Sportif



UN CAS PARTICULIER

Réussite de Bet365 : 73%

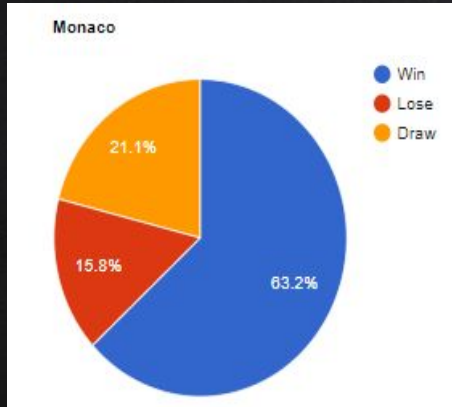


Figure 7 : Monaco 2017/2018

Réussite de Bet365 : 36%

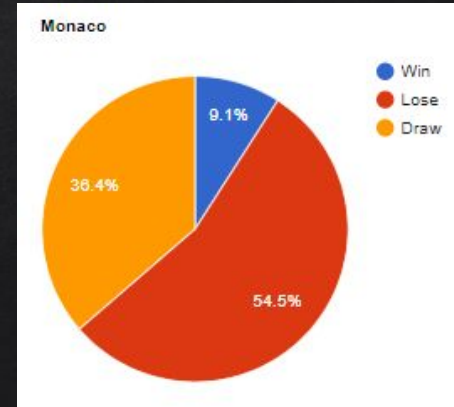


Figure 8 : Monaco 2018/2019



LES LIMITES & LES AMÉLIORATIONS POSSIBLES

- Améliorations
 - Ajout de données
 - Exploitation de données Mathématiques
- Les Limites
 - Technique
 - Accessibilité des données

CONCLUSION

- Utilisateurs spécifiques
- Application Dynamique en général
- Données basées sur le Data Mining
- Probabilités non négligeables
- Sources Limitées en France



Sources

Document csv:

<http://www.football-data.co.uk/francem.php>

Value bet :

<https://www.pinnacle.com/fr/betting-articles/educational/data-mining-in-sports-betting/4FC29M3VJY59Y4Q3>

Data et le sport :

<https://portail-ie.fr/analysis/1595/big-data-et-sport-de-la-donnee-a-la-performance>



MERCI DE VOTRE
ATTENTION!

Avez-vous des questions?