

Projet final Skillcraft



Thomas FROPO
Bastien GERARD

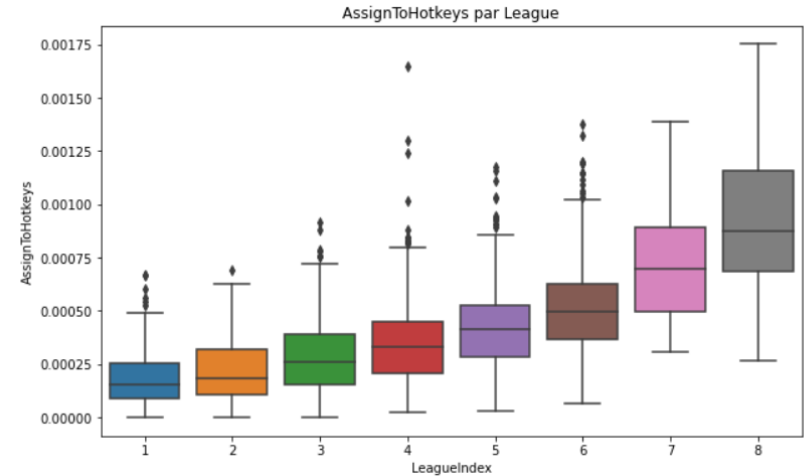
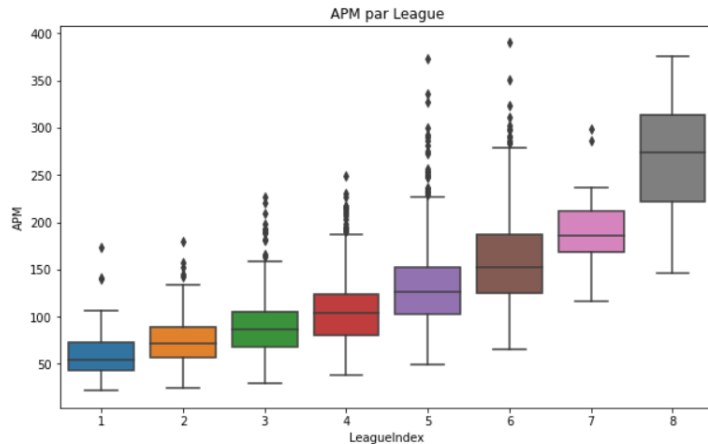
Description du projet

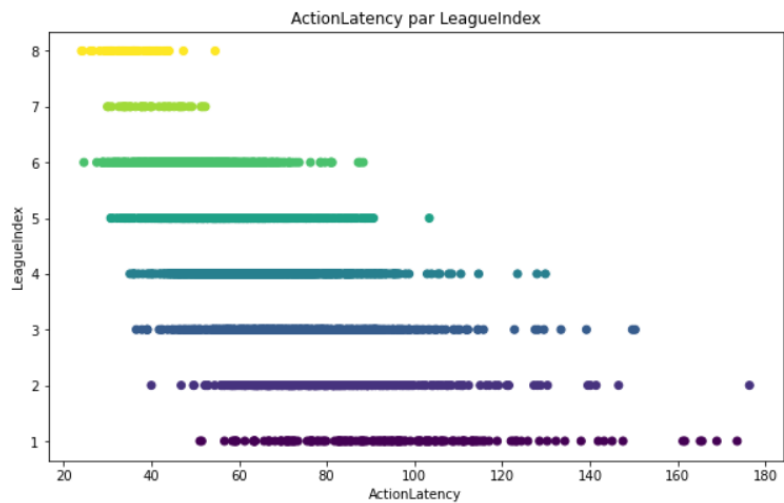
Tout d'abord ce dataset est composé de plus de 3000 données regroupées sur des joueurs de Starcraft 2 allant de bronze à professionnel. Le but de cette analyse sera de prédire le rang des joueurs grâce aux autres données.

Nous connaissons tous les deux le jeu Starcraft 2 d'un point de vue gameplay et compétition. Nous connaissons donc quelles peuvent être les données intéressantes à utiliser pour différencier les professionnels des joueurs casual. Comme par exemple la maîtrise de la micro et macro. La micro (micro gestion) est le fait de gérer son armée pour maximiser les dégâts faits à l'adversaire et des colonnes du dataset s'y prêtent comme les APM (actions par minute) ou ActionLatency. La macro (macro gestion) est le fait de gérer son évolution au cours de la partie. Nous pensons donc que ces colonnes influenceront beaucoup sur le fait d'avoir un rang élevé.

Exploration des données

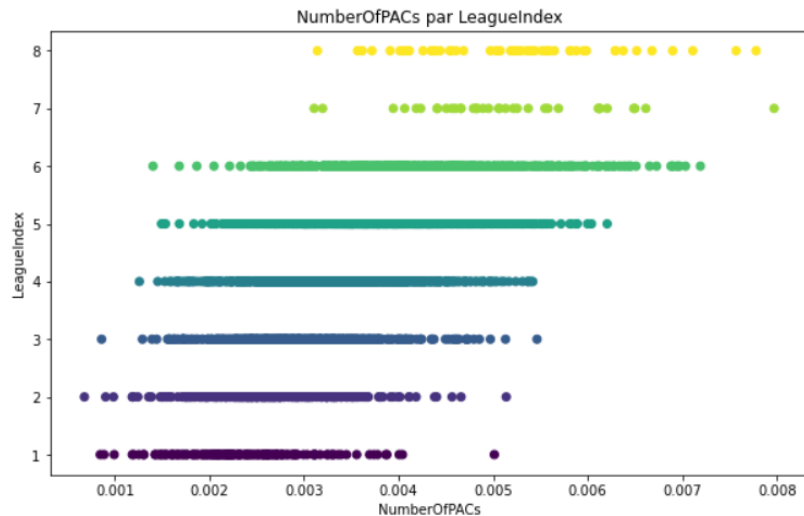
Nous avons eu des données manquantes dans 3 colonnes, sans grand impact étant donnée que ces 3 colonnes seront enlevé par manque de corrélation avec le LeagueIndex (colonnes qui donne les rangs des joueurs). Voici quelques exemples de données interprétées.





Matrice de corrélation

GameID	1	0.1	0.089	-0.025	0.0042	0.16	0.16	0.14	0.12	0.074	0.047	0.09	-0.1	-0.087	0.047	0.032	-0.012	0.021	-0.027	0.0032
LeagueIndex	0.1	1	-0.13	0.22	0.024	0.66	0.49	0.53	0.35	0.31	0.23	0.61	-0.55	-0.67	0.14	0.22	0.3	0.13	0.15	0.15
Age	0.089	-0.13	1	-0.18	-0.017	-0.21	-0.13	-0.11	0.015	0.043	-0.02	-0.2	0.11	0.24	-0.045	-0.025	-0.092	0.023	-0.08	-0.066
HoursPerWeek	-0.025	0.22	-0.18	1	0.024	0.25	0.21	0.16	0.07	0.084	0.049	0.17	-0.13	-0.19	0.095	0.065	0.051	0.039	0.059	0.075
TotalHours	-0.0042	0.024	-0.017	0.024	1	0.073	0.082	0.042	0.0093	0.00087	0.0077	0.04	-0.021	-0.036	0.011	0.02	0.015	-0.0024	-0.0072	-0.0063
APM	0.16	0.66	-0.21	0.25	0.073	1	0.84	0.58	0.36	0.27	0.32	0.65	-0.58	-0.72	0.38	0.22	0.35	0.1	0.14	0.13
SelectByHotkeys	0.16	0.49	-0.13	0.21	0.082	0.84	1	0.5	0.3	0.2	0.14	0.4	-0.3	-0.42	0.16	0.083	0.14	0.0061	0.044	0.061



Nous avons trouvé plusieurs colonnes qui auraient une grande corrélation avec le « LeagueIndex ». Comme nous le pensions au début, APM et ActionLatency en font partie. Mais il y a aussi NumberOfPACs et GapbetweenPAC. Nous ne connaissions pas ce qu'était le PAC (Perception Action Cycle) et après avoir appris ce que c'était, ça nous paraissait logique qu'un pro fasse plus de PAC qu'un joueur casual et que son temps entre chaque PAC soit plus petit. Pour rappel un PAC est le fait de savoir ce qu'il faut faire selon certaine circonstance (par exemple le fait de bouger sa caméra à un endroit et agir en conséquence)

Test de plusieurs modèles

Après avoir retiré les colonnes qui ne nous intéressaient pas ("GameID", "Age", "TotalHours", "HoursPerWeek", "UniqueUnitsMade", "ComplexUnitsMade", "ComplexAbilitiesUsed"), nous avons commencé à tester différents modèles au nombre de 8.

Nous avons pu constater une faible précision dans les tests faits, sûrement lié au fait que nous ayons peu de donnée à exploiter au final.

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis()
lda.fit(X_train1, y_train)
print("Précision de LDA sur l'ensemble train : ", round(lda.score(X_train1, y_train),2))
print("Précision de LDA sur l'ensemble test : " ,round(lda.score(X_test1, y_test),2))
```

Précision de LDA sur l'ensemble train : 0.4

Précision de LDA sur l'ensemble test : 0.35

```
pred = rf.predict(X_test1)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

```
[[12 10  8  2  0  0  0  0]
 [ 8 20 24 18  3  0  0  0]
 [ 1 14 38 48 17  4  0  0]
 [ 2  7 29 64 44  9  0  0]
 [ 0  2 11 50 57 39  0  0]
 [ 0  0  3 13 34 68  0  3]
 [ 0  0  0  0  0  6  0  0]
 [ 0  0  0  0  0  6  0  5]]
```

	precision	recall	f1-score	support
1	0.52	0.38	0.44	32
2	0.38	0.27	0.32	73
3	0.34	0.31	0.32	122
4	0.33	0.41	0.37	155
5	0.37	0.36	0.36	159
6	0.52	0.56	0.54	121
7	0.00	0.00	0.00	6
8	0.62	0.45	0.53	11
accuracy			0.39	679
macro avg	0.38	0.34	0.36	679
weighted avg	0.39	0.39	0.39	679

Nous avons choisis le RandomForest comme modèle parmi tous les autres car c'est celui qui avait la précision la plus haute.

On peut voir le peu de donnée pour trouver si un joueur appartient au rang 7 ou 8 (grand master et pro)

Conclusion

Nous pouvons voir une précision assez faible due au manque de données (un peu plus de 3000) surtout en ce qui concerne les pro et grand master. Un autre moyen de l'améliorer aurait été de plus manipuler certaines variable et utilisé des hyperparamètres.

