

Projet MRR Groupe 16 - Belhomme, Di Martino, Fraise

Jeu de données : SkillCraft1 Master Table Dataset

Contexte :

Les données recueillies concernent le jeu Stracraft 2, c'est un RTS : un jeu de stratégie en temps réel. Ce jeu a eu une scène e-sportive très riche depuis 2010, avec de nombreux joueurs professionnels et des prix importants.

Enjeu :

On souhaite savoir si, à partir des informations contenues dans une seule partie, il est possible d'estimer le classement du joueur, et avec quelle précision.

Moyen de récupération des données :

La majorité des données ont été récupérées à l'aide d'un sondage auprès de communautés Starcraft en ligne. Ce sondage a récupéré à la fois des informations subjectives (âge, temps de jeu) et des informations objectives (enregistrement de la partie par le jeu). Les réponses ont ensuite été filtrées de manière à correspondre aux critères voulus (partie humain vs humain, joueurs existants, parties pas trop courtes, etc).

Par manque de réponses de joueurs de rang professionnel, les données sur ce rang ont été récupérées via des enregistrements de parties sur des sites tiers.

Description des variables :

- GameID : id unique de la partie envoyée par le joueur ;
- 4 variables concernant le joueur : LeagueIndex (Ligue du joueur, variable que l'on souhaite modéliser), Age, HoursPerWeek (nombre d'heures de jeu par semaine) et TotalHours (nombre total d'heures de jeu);
- APM : Nombre d'actions par minute ;
- 11 variables concernant le nombre de fois où une certaine action a été effectuée au cours d'une unité de temps "timestamp" (1 seconde = 88.5 timestamps) : SelectByHotkeys, AssignToHotkeys, UniqueHotkeys, MinimapAttacks, MinimapRightClicks, NumberOfPACs, TotalMapExplored, WorkersMade, UniqueUnitsMade, ComplexUnitsMade, ComplexAbilitiesUsed ;
- GapBetweenPACs : durée moyenne (en ms) entre deux PACs ;
- ActionLatency : durée moyenne (en ms) entre le début d'un PAC et la première action du joueur ;
- ActionsInPAC : nombre moyen d'actions dans chaque PAC.

PAC (Percepted Action Cycle) : Un PAC correspond à un intervalle de temps au cours duquel le joueur a réalisé au moins une action sans se déplacer significativement.

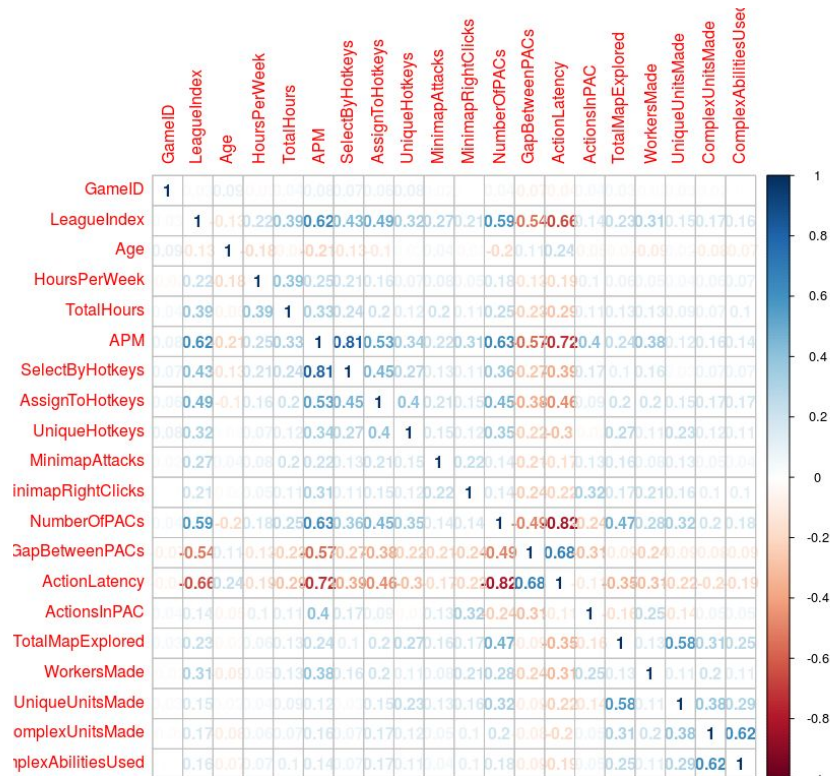
Premier traitement des données :

Nous avons remarqué qu'aucune des lignes concernant les joueurs professionnels ne contenait d'information sur le joueur. De plus les données concernant le temps de jeu des joueurs ont été fournies par les joueurs eux-mêmes et nous avons remarqué que certains avaient largement exagéré leur temps de jeu.

Ces lignes et les lignes présentant des données manquantes représentaient moins de 2% du jeu de données, nous avons pris la décision de les retirer avant de procéder à ce qui suit.

Matrice de corrélation : (goo.gl/qLuiHo)

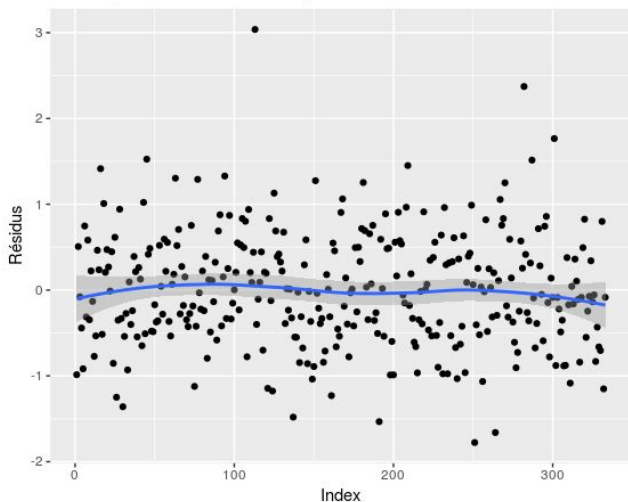
Les variables très corrélées ou anti-corrélées entre elles sont SelectByHotkeys et APM, ActionLatency et APM, ActionLatency et NumberOfPACs.



Régression sur tout le jeu de données :

Une régression basique semble nous indiquer que les variables ComplexAbilitiesUsed, MinimapRightClicks et APM n'ont pas un grand rôle à jouer dans la modélisation. Cette régression a également un R-squared bas (de l'ordre de 0.5). Nous sommes en présence de variables d'une significativité extrême (de l'ordre de 10^{-15} et 10^{-16}) pour les variables SelectByHotkeys, AssignToHotkeys, MinimapAttacks et ActionLatency. Ces valeurs si faibles nous font douter de la viabilité des données

Affichage des résidus sur le jeu d'entraînement



Prochaines étapes :

Essayer de sélectionner des variables, essayer d'autre types de régression.

Cependant, l'article et nos premiers tests semblent indiquer que le poids des variables varie en fonction du rang du joueur, il va peut être falloir que nous regroupions les ligues entre elles pour avoir de meilleurs résultats.