Rapport RLD TME1

Christopher Heim et Thomas Floquet

	Random	**Greedy (10 init)**	**Optimale**	**UCB (1 init)**	**UCB (10 init)**	**linUCB**
Reward	408	489	1532	993	1210	1431

Random : Si on choisit l'annonceur aléatoirement, alors le reward cumulé est de 408

Greedy : Si on décide de toujours rester sur le même bras (i.e. annonceur) après avoir estimé lequel rapporterait le plus grand reward à partir d'une exploration des 10 premiers articles, alors le reward cumulé est de 489, ce qui est déjà une meilleure solution que le random

Optimale : si on considère qu'on connaît déjà tous les rewards, alors en choisissant le meilleur annonceur pour chaque article le reward maximal qu'on peut obtenir est donc 1532

UCB (1 initialisation) : En appliquant la stratégie UCB avec une exploration limitée (on regarde les taux de clics des 10 annonceurs pour le premier article uniquement), on obtient un reward cumulé de 993, ce qui est déjà bien meilleur que les méthodes random et greedy. Cela s'explique par le fait qu'on choisit le bras qui serait le meilleur si les valeurs des bras étaient les meilleures possibles selon l'intervalle de confiance

UCB (10 initialisations) : En augmentant l'initialisation du premier aux dix premiers bras, le temps de calcul est un peu plus long car la phase d'exploration est plus importante mais on obtient logiquement de meilleurs résultats car l'estimation est plus précise (reward cumulé de 1210)

linUCB : Enfin, sans tricher (contrairement à la stratégie Optimale où on connaît déjà tous les taux de clics), la meilleure stratégie est linUCB avec un reward de 1431, car elle prend en compte les contextes (ici les descriptions que l'on a sur les articles)