

Projet RCP209 : Prediction of daily stock movements on the US market

Jerome Petit

4 avril 2019

1 INTRODUCTION

Dans le cadre du cours RCP209 du master TRIED, nous avons pu voir différentes méthodes pour l'analyse et la prédiction de données. L'objectif de ce projet est d'utiliser les différentes notions abordées au cours de cemode dans un cas réel. J'ai décidé de travailler sur le jeu de données issue du challenge CFM : prédiction des mouvements journalier des actions US. Le but est de prédire le signe du rendement du stock sur la période 15h30-16H connaissant l'évolution par tranche de 5min sur la période 9h-15H30. Par mouvement, l'organisateur du challenge entend le mouvement relatif à un certain benchmark qui n'est pas communiqué. C'est pour la donnée n'est pas un prix mais une difference de ratio , ainsi une valeur nulle signifie que l'action a évolué de la même manière que le benchmark. L'utilisation d'un benchmark est une technique courant dans la gestion de fond. Le benchmark peut être le CAC40, SP500 et l'objectif du gérant est de surperformer ce benchmark.

Dans ce rapport je présente les différentes approches que j'ai pu tester au cours de ce challenge ainsi que le résultat obtenu pour chacune d'elle

2 ANALYSE DE DONNÉES

2.1 Première analyse

Pour ce projet nous disposons de 52918217 données d'entraînement. Parmi ces données d'entraînement il y a 347646 données manquantes soit 0.65% de l'échantillon disponible. Il y en a tout 745327 séries temporelles de longueur 71. Le nombre 71 correspond au nombre de tranches de 5min comprises entre 9h et 15h30. Ces séries temporelles sont issues de 680 actions pour 1511 dates différentes. Nous n'avons pas d'info sur les dates et les titres. Les dates s'étendent sur plusieurs années et donc certainement permettent d'avoir des séries temporelles sur différentes situations économiques.