# Big Data Analytics Assignment 2

# Implement and train a decision tree

Hailun Zhu
Id: hailunz

1. Features
   I have changed some features used in this assignment.
   The current features are:
   [bid range(bidMax – bidMin), bid average, spread, direction compare to previous time, EURUSD direction, GBPUSD direction], label[predicted direction]

   As the first three features are not in a binary form, I use the average value of each one as the threshold to reprocess the feature vectors. That is if the value is less than the threshold, then set it to 0, otherwise 1.

   The reason that I do not use the bidMax and bidMin is that the trends of the bidMax, bidMin are usually the same. So I instead use bid range.

2. Statistics
   The total data I use are 15266 lines, including the data of the year 2013, 2014 and 2015.
   80% of them are training data, and 20% of them are test data.

   The correctness I got from my decision tree is: 0.5648330058939096.