1. **How did you preprocess this dataset?**
   除了原始 Dataset 裡面包含的 Open Price、Close Price、High Price、Low Price、Volume 之外，我還計算了前後兩天 Open Price 與 Close Price 的漲跌幅度。

```
# Define Predictor/Independent Variables
train_data['Open - Open'] = train_data['Open Price'] - train_data['Open Price'].shift(1)
train_data['Close - Close'] = train_data['Close Price'] - train_data['Close Price'].shift(1)
train_data = train_data.dropna()
x_train = train_data
x_train.head()
```

Date	Open Price	Close Price	High Price	Low Price	Volume	Open - Open	Close - Close
05-Jan-2009	929.17	927.45	936.63	919.53	5413910016	26.18	-4.35
06-Jan-2009	931.17	934.70	943.85	927.28	5392620032	2.00	7.25
07-Jan-2009	927.45	906.65	927.45	902.37	4704940032	-3.72	-28.05
08-Jan-2009	905.73	909.73	910.00	896.81	4991549952	-21.72	3.08
09-Jan-2009	909.91	890.35	911.93	888.31	4716499968	4.18	-19.38

2. **Which classifier reaches the highest classification accuracy in this dataset?**
   Logistic Regression、Support Vector Machine、Neural Network 的準確率相同，都是 0.51。（見 HW3_LR.ipynb、HW3_SVM.ipynb、HW3_NN.ipynb）
   (1) Why?
   有可能是因為 Dataset 是 Linearly separable 的，使預測變得容易，所以三種 Classifier 的準確率才會一樣。
   (2) Can this result remain if the dataset is different?
   如果換了一個 Linear separability 比較低的 Dataset，結果就有可能會改變。以上課提到過的 SAheart.data 為例，Logistic Regression、Support Vector Machine、Neural Network 的準確率分別為 0.74、0.64、0.63。（見 HW3_DifferentDataset.ipynb）
3. **How did you improve your classifiers?**
   要增加 Classifier 準確率的話，除了增加 Training 的資料量之外，也可以對 Dataset 做標準化，或是設定 class_weight 之類的參數去解決 Class Imbalance 的問題。