Background.
In predicting stock prices, you collect data over some period of time - day, week, month, etc. But you cannot take advantage of data from a time period until the next increment of the time period. For example, assume you collect data daily. When Monday is over you have all of the data for that day. However, you can invest on Monday, because you don't get the data until the end of the day. You can use the data from Monday to invest on Tuesday.
In our research[1] each record (row) is data for a week. Each record also has the percentage of return that stock has in the following week (*percent_change_next_weeks_price*). Ideally, you want to determine which stock will produce the greatest rate of return in the following week. This can help you train and test your algorithm. Some of these attributes might not be use used in your research. They were originally added to our database to perform calculations. Brown, Pelosi & Dirska (2013) used *percent_change_price*, *percent_change_volume_over_last_wk, days_to_next_dividend*, and *percent_return_next_dividend*. We left the other attributes in the dataset in case you wanted to use any of them. Of course, what you want to maximize is *percent_change_next_weeks_price*.

Attribute Information:

- quarter: the yearly quarter (1 = Jan-Mar; 2 = Apr=Jun).
- stock: the stock symbol.
- date: the last business day of the work (this is typically a Friday)
- open: the price of the stock at the beginning of the week
- high: the highest price of the stock during the week
- low: the lowest price of the stock during the week
- close: the price of the stock at the end of the week
- volume: the number of shares of stock that traded hands in the week
- percent_change_price: the percentage change in price throughout the week
- percent_chagne_volume_over_last_wek: the percentage change in the number of shares of
- stock that traded hands for this week compared to the previous week
- previous_weeks_volume: the number of shares of stock that traded hands in the previous week
- next_weeks_open: the opening price of the stock in the following week
- next_weeks_close: the closing price of the stock in the following week
- percent_change_next_weeks_price: the percentage change in price of the stock in the
- following week days_to_next_dividend: the number of days until the next dividend
- percent_return_next_dividend: the percentage of return on the next dividend

Training data vs Test data:
Use quarter 1 (Jan-Mar) data for training and quarter 2 (Apr-Jun) data for testing.
Task: Build models to predict stock prices and evaluate risks of stocks.
Try different models (LM, Decision Trees/SVR) to test for accuracy. Discuss appropriateness of model and insights from findings. Discuss prediction accuracy. Discuss risks of different stocks using CAPM. **Hint: Try lagged variables for modeling?**

Interesting data points:
If you use quarter 2 data for testing, you will notice something interesting in the week ending 5/27/2011 every Dow Jones Index stock lost money.

---

[1] Brown, M. S., Pelosi, M. & Dirska, H. (2013). Dynamic-radius Species-conserving Genetic Algorithm for the Financial Forecasting of Dow Jones Index Stocks. Machine Learning and Data Mining in Pattern Recognition, 7988, 27-41.