1. **Installation**

   ```
   # set-up anaconda environment
   # launch anaconda prompt
   conda update -n base -c defaults conda
   conda create -n datasci python=3.8
   conda activate datasci
   conda install -y -q -c conda-forge/label/cf2020003 numpy
   conda install -y -q -c anaconda pandas
   conda install -y -q -c conda-forge/label/cf202003 matplotlib
   conda install -y -q -c anaconda notebook
   conda install -y -q -c anaconda seaborn
   conda install -y -q -c anaconda scikit-learn
   conda install -y -q -c conda-forge/label/cf202003 imbalanced-learn
   conda update -y -q --all
   python -m ipykernel install --user --name datasci --display-name "Python (datasci)"
   # navigate to folder with datafiles and notebook
   Jupyter Notebook
   ```

2. **Project Motivation**
   This is a Udacity Data Science nanodegree project.
   There's much financial data available that minimizes wrangling and maximizes modeling.

3. **File Descriptions**
   "GSPC.csv" is a Yahoo Finance download of S&P 500 index data.
   "IBM.csv" is a Yahoo Finance download of IBM stock data.
   "stock_data_build_r1.ipynb" is a Jupyter Notebook of this data study.
   "Imperfect Models for Stock Trades.pdf" project blog writeup for laymen.

4. **Project Interaction**
   A. The test_train_split function can be run for other validation years and spans.
   B. The rest of the notebook may be run sequentially to ascertain improvements.

5. Licensing, Acknowledgments
   A. References:
      Yahoo Finance Downloads to CSV files for imports:
      https://finance.yahoo.com/quote/IBM/history
      https://finance.yahoo.com/quote/%5EGSPC/history

      Hyperparameter Tuning the Random Forest in Python

      https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74

B. License, https://opensource.org/licenses/MIT