

##DAND - Udacity Course
Machine Learning Final Project by Teresa Aysan
March 4, 2018

TEXT FILE LISTING MY REFERENCES

All of the following references were used while doing the Machine Learning Project.

I have cited the origins of any parts of the submission that were taken from that origin. The numbers at the start of the reference cross-reference back to the superscript number in my code and/or in my "Documentation of My Work".

1<https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475461/lessons/3174288624239847/concepts/31726388830923> and <https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475460/lessons/2948958577/concepts/30368385820923>

2<https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475460/lessons/2948958577/concepts/30368385820923>

3<https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475460/lessons/2291728537/concepts/30100186330923> - Size of Training Set. More data is always better.

4<https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475460/lessons/2254358555/concepts/24196685390923>

<https://stackoverflow.com/questions/5904969/how-to-print-a-dictionarys-key>

<https://stackoverflow.com/questions/5844672/delete-an-item-from-a-dictionary>

<https://stackoverflow.com/questions/30345574/scatter-a-2d-numpy-array-in-matplotlib>

5<https://stackoverflow.com/questions/27516849/how-to-convert-list-of-numpy-arrays-into-single-numpy-array>

<https://stackoverflow.com/questions/7118276/how-to-remove-specific-element-in-an-array-using-python>

6<https://stackoverflow.com/questions/10996140/how-to-remove-specific-elements-in-a-numpy-array>

<https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.concatenate.html>

<https://stackoverflow.com/questions/13634349/plot-2-dimensional-numpy-array-using-specific-columns>

<https://blogs.sas.com/content/iml/2016/07/18/color-markers-third-variable-sas.html>

<https://stackoverflow.com/questions/31685078/change-color-of-2d-plot-line-depending-on-3rd-value>

7<https://stackoverflow.com/questions/26754626/how-to-make-2d-scatter-plot-in-matlab-with-colors-depending-on-values>

<https://www.mathworks.com/matlabcentral/answers/354423-how-to-access-gscatter-property-and-update-data?requestedDomain=www.mathworks.com>

<https://stackoverflow.com/questions/35313489/python-matplotlib-legend-based-on-a-conditional-variable>

<https://www.mathworks.com/help/matlab/ref/scatter.html>

8<https://stackoverflow.com/questions/8486294/how-to-add-an-extra-column-to-an-numpy-array>

9<https://stackoverflow.com/questions/45709743/how-to-replace-a-value-in-array-in-numpy-python>

10<https://stackoverflow.com/questions/4700614/how-to-put-the-legend-out-of-the-plot>

11<https://stackoverflow.com/questions/39500265/manually-add-legend-items-python-matplotlib>

<https://stackoverflow.com/questions/16826711/is-it-possible-to-add-a-string-as-a-legend-item-in-matplotlib>

<https://stackoverflow.com/questions/19125722/adding-a-legend-to-pyplot-in-matplotlib-in-the-most-simple-manner-possible>

<https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.chararray.html>

https://www.google.ca/search?dcr=0&ei=-uAWWrGNE4aJ0gLOgbnwCg&q=matplotlib+scatter&oq=matplotlib&gs_l=psy-ab.1.3.0i131i67k1l2j0i67k1l2j0i131i67k1j0i67k1l5.1253682.1257529.0.1264127.10.4.0.6.6.0.362.651.0j2j0j1.3.0....0...1c.1.64.psy-ab..1.9.703...0j0i131k1.0.03c4XKiuU-o

https://matplotlib.org/devdocs/api/_as_gen/matplotlib.pyplot.scatter.html

https://matplotlib.org/devdocs/gallery/lines_bars_and_markers/scatter_masked.html#sphx-glr-gallery-lines-bars-and-markers-scatter-masked-py

https://matplotlib.org/devdocs/_downloads/scatter_masked.py

https://matplotlib.org/devdocs/gallery/lines_bars_and_markers/scatter_symbol.html#sphx-glr-gallery-lines-bars-and-markers-scatter-symbol-py

<http://anh.cs.luc.edu/python/hands-on/3.1/handsonHtml/ifstatements.html>

<https://www.mathworks.com/help/matlab/ref/scatter.html>

https://www.mathworks.com/help/matlab/ref/scatter.html#btrli6o-1_1

<https://www.mathworks.com/help/matlab/ref/scatter.html#btrli6o-1>

<https://stackoverflow.com/questions/17411940/matplotlib-scatter-plot-legend>

12<https://stackoverflow.com/questions/7561017/get-the-column-names-of-a-python-numpy-ndarray>

13<https://stackoverflow.com/questions/8598163/split-title-of-a-figure-in-matplotlib>

<https://docs.scipy.org/doc/numpy-1.13.0/reference/routines.char.html>

14<https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.zeros.html#numpy.zeros>

<https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.divide.html>

<https://stackoverflow.com/questions/3877209/how-to-convert-an-array-of-strings-to-an-array-of-floats-in-numpy>

<https://discussions.udacity.com/t/outlier-removal-with-a-classification-algorithm/16573/18> Convert to pandas and boxplot

<https://stackoverflow.com/questions/31695360/convert-multidimensional-array-string-to-float-in-python>

<https://stackoverflow.com/questions/27317517/make-division-by-zero-equal-to-zero>

<https://stackoverflow.com/questions/176918/finding-the-index-of-an-item-given-a-list-containing-it-in-python>

15<https://stackoverflow.com/questions/27350048/python-return-index-of-an-element-of-an-array-given-specific-conditions>

16<https://stackoverflow.com/questions/18079029/index-of-element-in-numpy-array>

17<https://stackoverflow.com/questions/3877491/deleting-rows-in-numpy-array>

<https://stackoverflow.com/questions/28124971/removing-rows-from-a-multi-dimensional-numpy-array>

18<https://stackoverflow.com/questions/43152529/python-matplotlib-adding-regression-line-to-a-plot>

<https://stackoverflow.com/questions/37234163/how-to-add-a-line-of-best-fit-to-scatter-plot>

<https://stackoverflow.com/questions/12981696/how-to-draw-line-inside-a-scatter-plot> Shows dashed red line

19<https://stackoverflow.com/questions/35099130/change-spacing-of-dashes-in-dashed-line-in-matplotlib>

https://matplotlib.org/1.3.1/examples/pylab_examples/line_styles.html

https://matplotlib.org/devdocs/gallery/linesBars_and_markers/line_demo_dash_control.html

http://matplotlib.org/api/pyplot_api.html#matplotlib.pyplot.plot

20<https://stackoverflow.com/questions/20335290/matplotlib-plot-set-x-ticks> Use of ax and how to set ticks

<https://stackoverflow.com/questions/12608788/changing-the-tick-frequency-on-x-or-y-axis-in-matplotlib>

21<https://stackoverflow.com/questions/4455076/how-to-access-the-ith-column-of-a-numpy-multidimensional-array>

<https://docs.python.org/2/library/time.html>

<https://docs.python.org/2/library/functions.html#round>

<https://stackoverflow.com/questions/997797/what-does-s-mean-in-python>

<https://stackoverflow.com/questions/6839542/how-to-use-more-than-one-condition-in-python-for-loop>

<https://stackoverflow.com/questions/19122345/to-convert-string-to-variable-name>

<https://stackoverflow.com/questions/8028708/dynamically-set-local-variable>

<https://stackoverflow.com/questions/11354214/creating-dynamically-named-variables-from-user-input>

<https://stackoverflow.com/questions/5036700/how-can-you-dynamically-create-variables-via-a-while-loop>

<https://stackoverflow.com/questions/1373164/how-do-i-create-a-variable-number-of-variables>

<https://stackoverflow.com/questions/402504/how-to-determine-a-python-variables-type>

<https://stackoverflow.com/questions/30937667/naivebayes-classifier-handling-different-data-types-in-python>

http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.DictVectorizer.html

<https://stackoverflow.com/questions/6774980/is-there-a-way-to-add-multiple-conditions-in-a-for-loop>

<https://stackoverflow.com/questions/4323946/loop-with-conditions-in-python>

<https://stackoverflow.com/questions/18648626/for-loop-with-two-variables>

<https://stackoverflow.com/questions/2146419/how-to-do-while-loops-with-multiple-conditions>

<https://stackoverflow.com/questions/3052793/python-output-from-functions>

22 Code from Final Project given by Udacity in Tools\feature_format.py

23<https://stackoverflow.com/questions/1024847/add-new-keys-to-a-dictionary>

<https://stackoverflow.com/questions/12453799/how-to-create-a-table-using-python>

24<https://stackoverflow.com/questions/38622975/typeerror-object-of-type-file-has-no-len>

<https://stackoverflow.com/questions/354883/how-do-you-return-multiple-values-in-python>

25<https://stackoverflow.com/questions/419163/what-does-if-name-main-do>

26<https://stackoverflow.com/questions/40801789/using-return-value-inside-another-function>

27<https://stackoverflow.com/questions/455612/limiting-floats-to-two-decimal-points>

28From Udacity Feature_Format.py imported featureFormat() and targetFeatureSplit()

29Based on Udacity Regression lesson on how to code and plot regressions.

<https://machinelearningmastery.com/rescaling-data-for-machine-learning-in-python-with-scikit-learn/>

30<https://stackoverflow.com/questions/20073639/python-syntaxerror-invalid-syntax-end>

<https://github.com/spyder-ide/spyder/issues/64>

<https://github.com/JulianEberius/SublimeRope/issues/82>

<https://mail.google.com/mail/u/0/#inbox>

<https://github.com/SublimeLinter/SublimeLinter-for-ST2/issues/561#issuecomment-31094858>

<https://github.com/SublimeLinter/SublimeLinter-for-ST2/issues/561>

https://www.codecademy.com/en/forum_questions/514c4578172ea3daa90018f4

https://en.wikibooks.org/wiki/Python_Programming/Print_version

<https://www.cse.msu.edu/~cse231/Online/debugging.pdf>

<https://stackoverflow.com/tags/spyder/hot?filter=all>

<https://stackoverflow.com/questions/5951135/how-to-save-a-list-as-numpy-array-in-python>

<https://stackoverflow.com/questions/23668509/dictionary-keys-and-values-to-separate-numpy-arrays>

<https://stackoverflow.com/questions/24168569/how-to-add-names-to-a-numpy-array-without-changing-its-dimension>

<https://stackoverflow.com/questions/26850355/converting-list-to-numpy-array>

<https://stackoverflow.com/questions/24168569/how-to-add-names-to-a-numpy-array-without-changing-its-dimension>

<https://github.com/numpy/numpy/issues/5134>

<https://stackoverflow.com/questions/3405715/elegant-way-to-remove-fields-from-nested-dictionaries>

<https://stackoverflow.com/questions/5064822/how-to-add-items-into-a-numpy-array>

<https://stackoverflow.com/questions/23668509/dictionary-keys-and-values-to-separate-numpy-arrays>

<https://stackoverflow.com/questions/5951135/how-to-save-a-list-as-numpy-array-in-python>

<https://stackoverflow.com/questions/21016684/delete-rows-from-numpy-array-depending-on-a-condition-on-a-single-cell>

<https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.delete.html>

<https://stackoverflow.com/questions/3877491/deleting-rows-in-numpy-array>

<https://stackoverflow.com/questions/25380255/how-do-i-remove-rows-from-a-numpy-array-based-on-multiple-conditions>

<https://stackoverflow.com/questions/19561857/remove-rows-in-3d-numpy-array>

<https://stackoverflow.com/questions/28124971/removing-rows-from-a-multi-dimensional-numpy-array>

<https://stackoverflow.com/questions/21016684/delete-rows-from-numpy-array-depending-on-a-condition-on-a-single-cell>

<https://stackoverflow.com/questions/27532503/remove-nan-row-from-x-array-and-also-the-corresponding-row-in-y>

<https://stackoverflow.com/questions/30340628/delete-rows-at-select-indexes-from-a-numpy-array>

<https://www.dataquest.io/blog/numpy-tutorial-python/>

<https://stackoverflow.com/questions/971678/iterating-through-a-multidimensional-array-in-python>

<https://discussions.udacity.com/t/final-project-doubts/248874/13>

31http://scikit-learn.org/stable/modules/generated/sklearn.cross_validation.StratifiedShuffleSplit.html

32Final Project: tester.py

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.2085&rep=rep1&type=pdf>

<https://stats.stackexchange.com/questions/18815/increasing-number-of-features-results-in-accuracy-drop-but-prec-recall-increase>

<https://machinelearningmastery.com/better-naive-bayes/>

http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

http://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes

http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB

ab.12...488549.488549.0.491068.1.1.0.0.0.0.558.558.5-
1.1.0....0...1c.1.64.psy-ab..0.0.0....0.A7F7BviJzDo

<https://stackoverflow.com/questions/176918/finding-the-index-of-an-item-given-a-list-containing-it-in-python>

<https://stackoverflow.com/questions/5735841/typeerror-list-object-is-not-callable-while-trying-to-access-a-list>

<https://stackoverflow.com/questions/31087111/typeerror-list-object-is-not-callable-in-python/31087151>

<https://stackoverflow.com/questions/176918/finding-the-index-of-an-item-given-a-list-containing-it-in-python>

<https://stackoverflow.com/questions/10645444/unbound-method-with-instance-as-first-argument-got-string-but-requires-something>

https://www.google.ca/search?dcr=0&ei=PWJfWtfxI9HejwPforBw&q=isinstance+python+example&oq=isinstance&gs_l=psy-ab.1.2.016j0i10k1j0l3.171026203.171029376.0.171033069.20.14.0.0.0.0.508.2404.0j5j3j1j0j1.11.0....0...1c.1.64.psy-ab..9.10.2398.0..0i67k1j0i131i67k1j0i131k1j0i10i67k1.135.vHlgkIM5e-A

<https://discussions.udacity.com/t/difference-in-precision-and-recall-in-final-project/250667>

37<http://benalexkeen.com/feature-scaling-with-scikit-learn/>

36<https://stackoverflow.com/questions/35595766/matplotlib-line-magic-causes-syntaxerror-in-python-script>

<http://scikit-learn.org/stable/modules/preprocessing.html>

<https://community.modeanalytics.com/python/tutorial/pandas-dataframe/#selecting-columns-in-a-dataframe>

<https://github.com/pandas-dev/pandas/blob/v0.22.0/pandas/core/frame.py#L1906-L1909>

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.transpose.html>

<https://stackoverflow.com/questions/31658183/how-to-switch-columns-rows-in-a-pandas-dataframe>

<https://stackoverflow.com/questions/12555323/adding-new-column-to-existing-dataframe-in-python-pandas>

<https://stackoverflow.com/questions/14940743/selecting-excluding-sets-of-columns-in-pandas>

https://chrisalbon.com/python/data_wrangling/pandas_dropping_column_and_rows/

<https://stackoverflow.com/questions/22392786/excluding-rows-from-a-pandas-dataframe-based-on-column-value-and-not-index-value>

<https://stackoverflow.com/questions/13851535/how-to-delete-rows-from-a-pandas-dataframe-based-on-a-conditional-expression>

<https://stackoverflow.com/questions/13295735/how-can-i-replace-all-the-nan-values-with-zeros-in-a-column-of-a-pandas-datafra>

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html>

<https://stackoverflow.com/questions/22470690/get-list-of-pandas-dataframe-columns-based-on-data-type>

<https://stackoverflow.com/questions/1841565/valueerror-invalid-literal-for-int-with-base-10>

<https://pandas.pydata.org/pandas-docs/stable/merging.html>

https://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjtzPf90_jYAhUN1WMKHarcBd4QFjAAegQIARAB&url=https%3A%2F%2Fstackoverflow.com%2Fquestions%2F11346283%2Frenaming-columns-in-pandas&usg=AOvVaw1bFLs5KQ4Gk7krL7GN9clQ

<https://pandas-docs.github.io/pandas-docs-travis/dsintro.html>

<https://stackoverflow.com/questions/19712264/what-is-record-array-in-pandas>

<https://stackoverflow.com/questions/26716616/convert-a-pandas-dataframe-to-a-dictionary>

<https://docs.scipy.org/doc/numpy/reference/generated/numpy.recarray.html>

<https://stackoverflow.com/questions/46837472/converting-pandas-dataframe-to-structured-arrays>

<http://pandas.pydata.org/pandas-docs/version/0.18/generated/pandas.DataFrame.html>

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.as_matrix.html

<https://stackoverflow.com/questions/13187778/convert-pandas-dataframe-to-numpy-array-preserving-index>

<https://stackoverflow.com/questions/11346283/renaming-columns-in-pandas>

<https://stackoverflow.com/questions/39050539/adding-multiple-columns-to-pandas-simultaneously>

<http://pandas.pydata.org/pandas-docs/stable/indexing.html#basics>

<https://stackoverflow.com/questions/12555323/adding-new-column-to-existing-dataframe-in-python-pandas>

<http://scikit-learn.org/stable/modules/ensemble.html>

http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

http://scikit-learn.org/stable/auto_examples/plot_compare_reduction.html

<https://discussions.udacity.com/t/dont-know-how-to-validate/201521/4>

<http://scikit-learn.org/stable/modules/pipeline.html>

38<https://discussions.udacity.com/t/error-using-pipeline/171750/4>

<https://discussions.udacity.com/t/project-fear-struggling-with-machine-learning-project/198529/47>

<https://discussions.udacity.com/t/dont-know-how-to-validate/201521/19>

<https://discussions.udacity.com/t/error-using-pipeline/171750/10>

http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html

http://scikit-learn.org/stable/auto_examples/plot_compare_reduction.html#sphx-glr-auto-examples-plot-compare-reduction-py

<http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

http://scikit-learn.org/stable/modules/grid_search.html

<https://stackoverflow.com/questions/39839112/the-easiest-way-for-getting-feature-names-after-running-selectkbest-in-scikit-le>

39: <https://stackoverflow.com/questions/39839112/the-easiest-way-for-getting-feature-names-after-running-selectkbest-in-scikit-le>

<https://stackoverflow.com/questions/46902367/numpy-array-typeerror-only-integer-scalar-arrays-can-be-converted-to-a-scalar-i>

40<https://discussions.udacity.com/t/selectpercentile-how-to-fill-the-parameters/174780/15>

<https://stackoverflow.com/questions/3776275/how-to-add-key-value-pair-to-dictionary/28380174>

<https://stackoverflow.com/questions/11371691/adding-a-new-key-to-a-nested-dictionary-in-python>

<https://www.sitepoint.com/community/t/adding-a-new-key-to-a-nested-dictionary-in-python/247766/2>

<https://mail.python.org/pipermail/tutor/2010-May/076238.html>

<http://www.pythonforbeginners.com/dictionary/dictionary-manipulation-in-python>

41<https://www.quora.com/In-Python-can-we-add-a-dictionary-inside-a-dictionary-If-yes-how-can-we-access-the-inner-dictionary-using-the-key-in-the-primary-dictionary>

http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

http://scikit-learn.org/stable/auto_examples/calibration/plot_calibration.html#sphx-glr-auto-examples-calibration-plot-calibration-py

http://scikit-learn.org/stable/auto_examples/hetero_feature_union.html#sphx-glr-auto-examples-hetero-feature-union-py

<https://github.com/scikit-learn/scikit-learn/blob/a24c8b46/sklearn/base.py#L213>

http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB.set_params

<http://scikit-learn.org/stable/modules/generated/sklearn.base.BaseEstimator.html>

<https://discussions.udacity.com/t/gaussiannb-parameters/22982>

<https://discussions.udacity.com/t/implementation-the-optimal-selection-of-the-k-value-within-the-gridcv-process/224205/14>

<https://discussions.udacity.com/t/using-gaussiannb-when-gridsearchcv/198156>

42<https://stackoverflow.com/questions/22903267/what-is-tuning-in-machine-learning>

43https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-9863-7_233

44<https://stats.stackexchange.com/questions/19048/what-is-the-difference-between-test-set-and-validation-set>

45<https://rapidminer.com/resource/correct-model-validation/> This is a
very good white paper... but it is also a marketing tool for rapidminder
46 [https://machinelearningmastery.com/overfitting-and-underfitting-with-
machine-learning-algorithms/](https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/) Very good. Simple to read re overfitting.
<https://www.casact.org/pubs/monographs/papers/06-Kunce-Chatterjee.pdf>

47[http://scikit-
learn.org/stable/modules/generated/sklearn.metrics.precision_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html)

48[https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-
tuning-xgboost-with-codes-python/](https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/)

49[https://gallery.azure.ai/Experiment/Evaluating-and-Parameter-Tuning-a-
Decision-Tree-Model-1](https://gallery.azure.ai/Experiment/Evaluating-and-Parameter-Tuning-a-Decision-Tree-Model-1)

50<http://scikit-learn.org/stable/modules/tree.html>

51[http://scikit-
learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.ht
ml](http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html)

52Course Lesson 4 Decision Trees, 13. Data Impurity and Entropy
[https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/
317428862475460/lessons/2258728540/concepts/24488285540923](https://classroom.udacity.com/nanodegrees/nd002/parts/0021345409/modules/317428862475460/lessons/2258728540/concepts/24488285540923)

53[http://scikit-
learn.org/stable/auto_examples/model_selection/plot_precision_recall.html](http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html)