# Assignment 2

# CSCI6906 - Spec. Grad. Topics in Computer Science

# Visual Analytics

Python Version used is 2.7.6

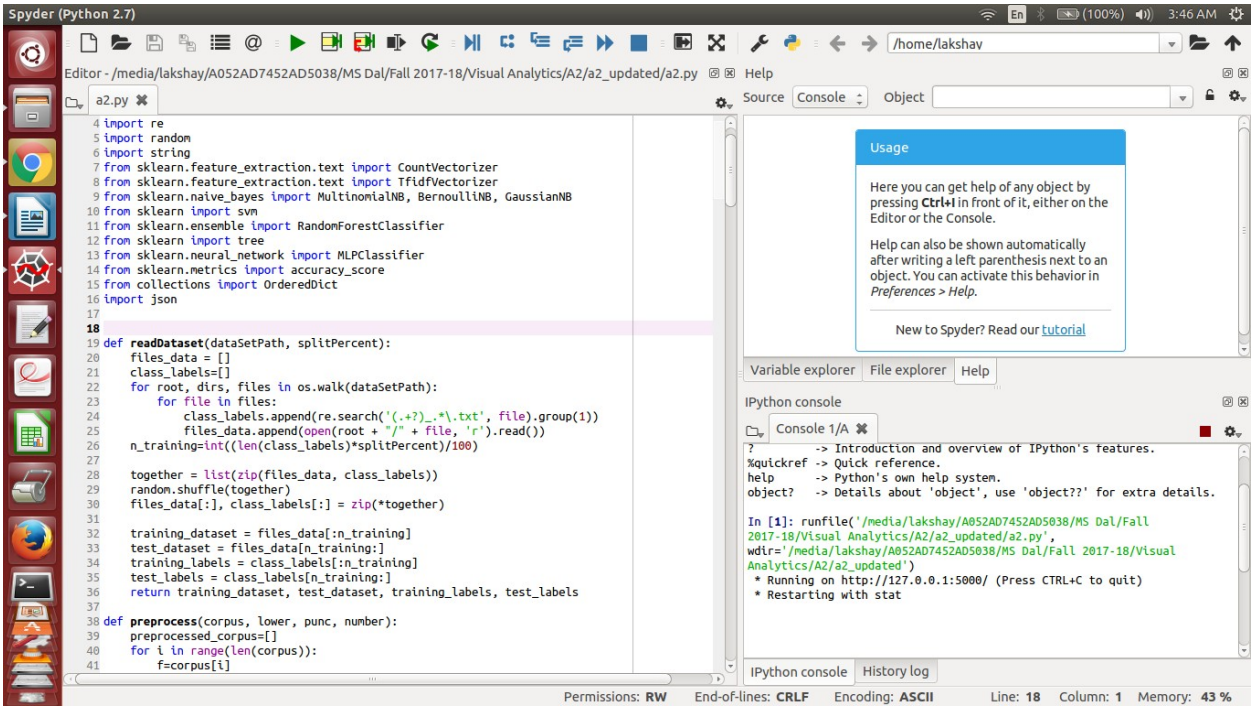
Operating System – Ubuntu 14.04

Flask version used is 0.12.2

Before installing flask, Java has to be installed on the system.

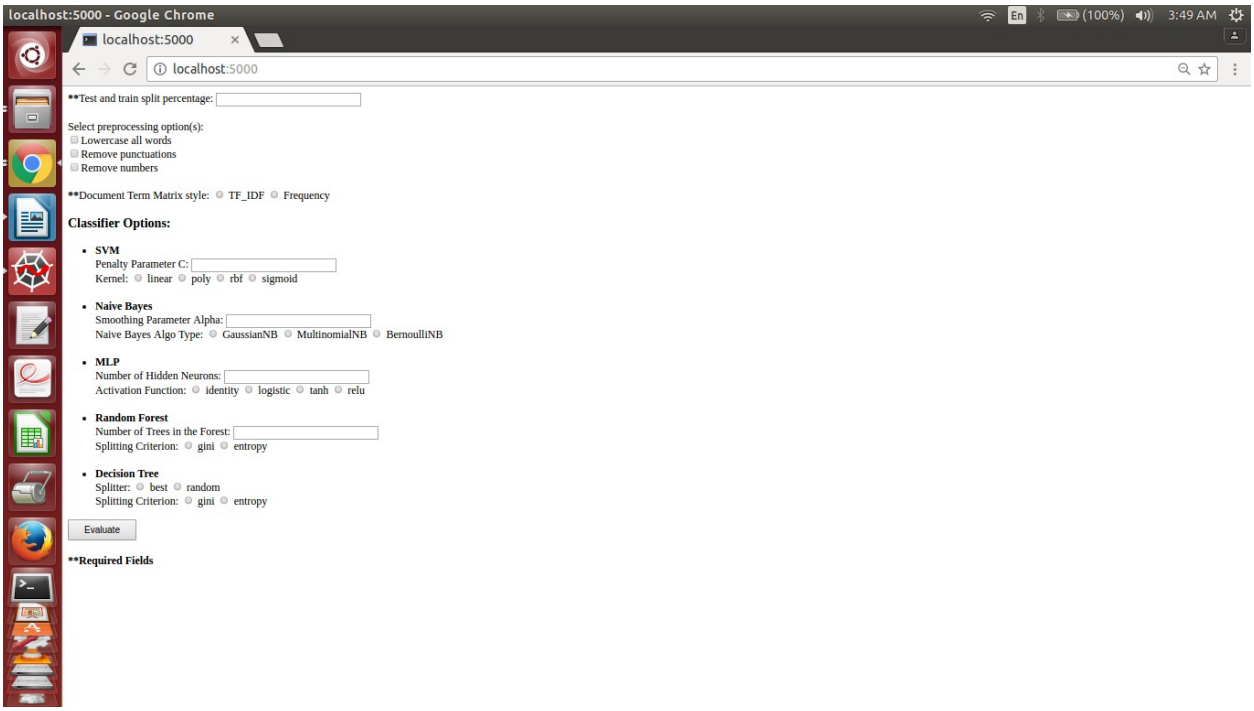
Connection between front-end and back-end is done using Flask. The front-end code is saved in the file a2.html and the back-end code is saved in a2.py file. The folder containing the a2.py (flask code file) also contains the folder named **templates**, which contains the front-end a2.html file. Saving the a2.html file inside templates folder is a requirement of flask in order to work.

The below screenshot shows a running server at localhost:5000.

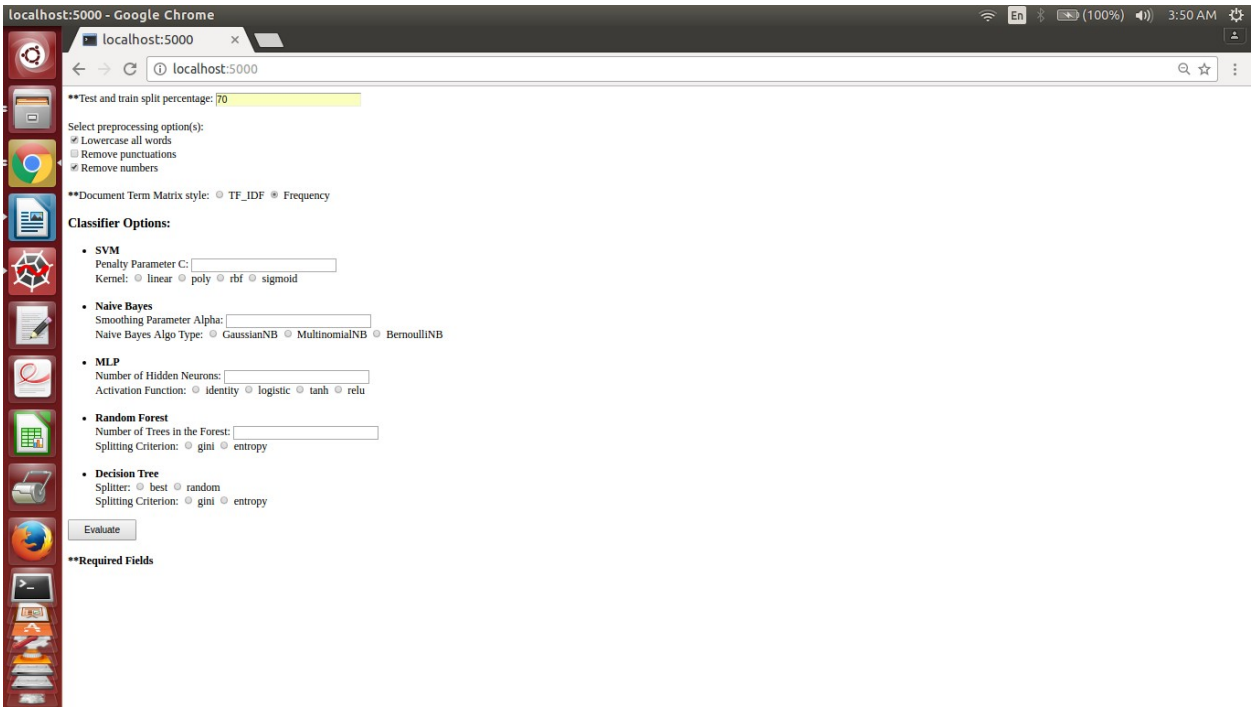


Apart from the basic options like splitpercentage, preprocessing options and Document-Term Matrix style, various other bonus options are also provided for each classifier. Data is gathered on the frontend using the form shown below in the screenshot, and this data is passed onto the backend (Python part).

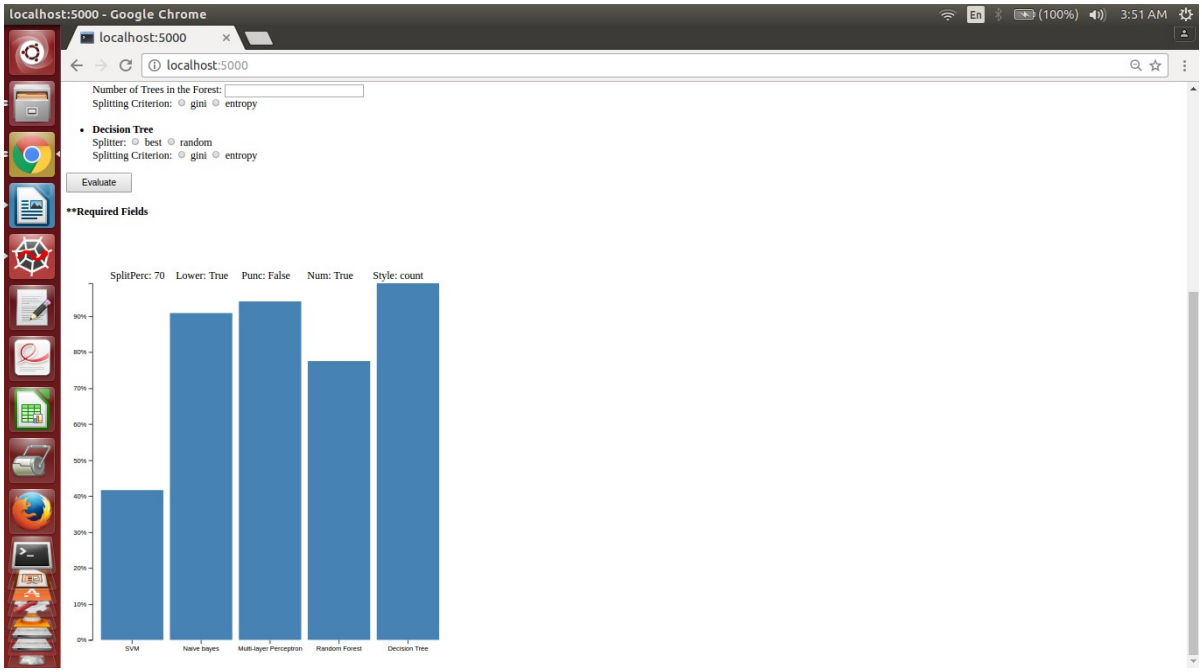
Options which are mandatory (i.e. SplitPercetage and Document-Term Matrix Style) are marked with two bold asterisk** symbols. This is done to make it easy for any user to identify which fields are required to proceed further. All other fields except asterisk symbol are optional.



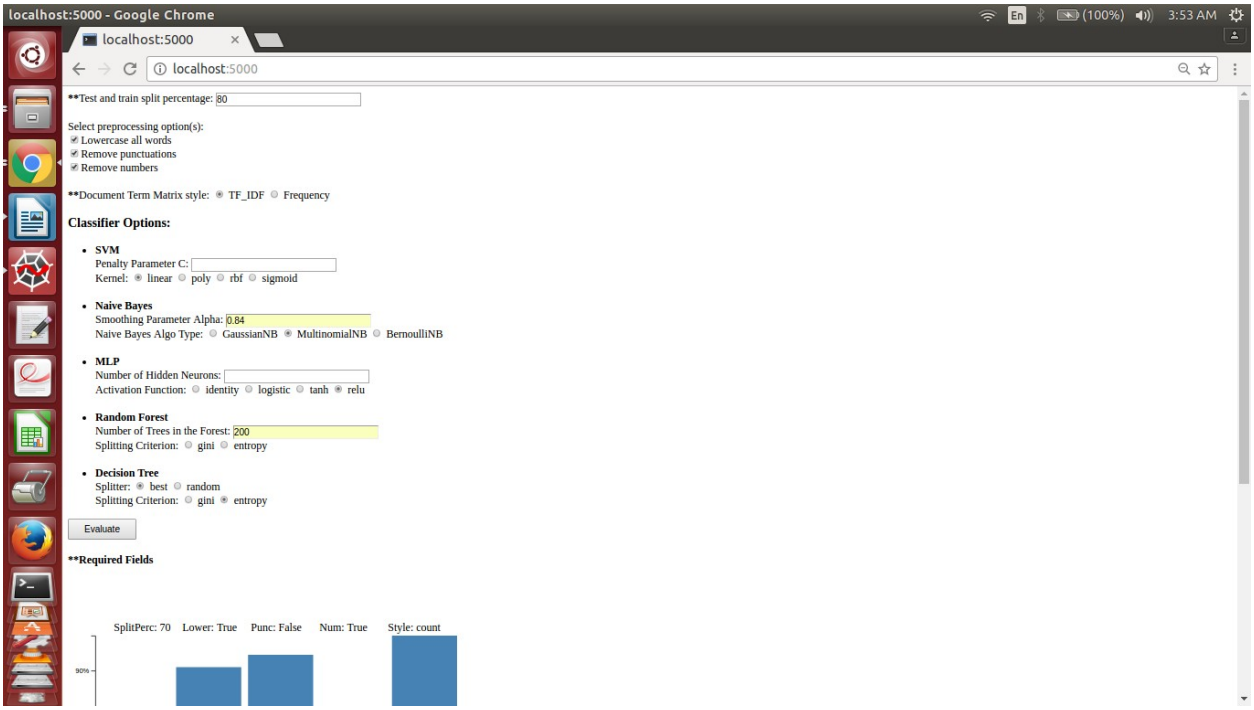

Suppose the below values are filled in the form before evaluate button is clicked -



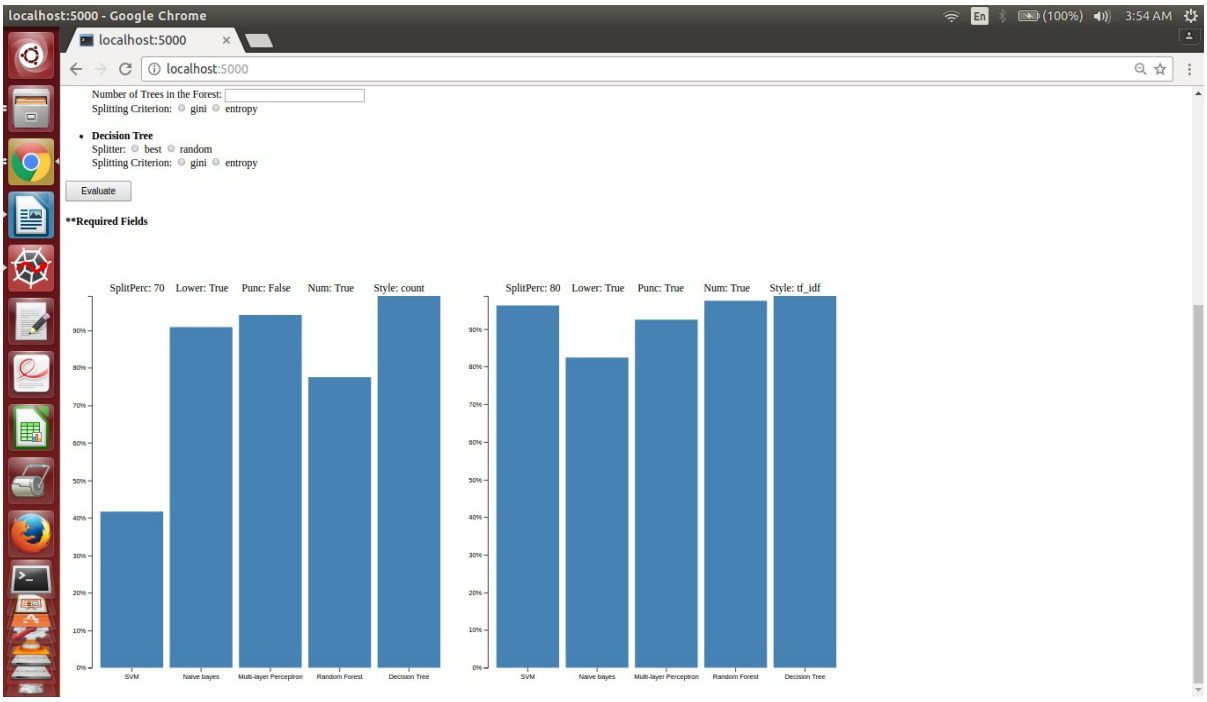

Once evaluate button is clicked, we get a bar chart representing all 5 classifier names along the x-axis with their accuracies on the y-axis. Legend shows the splitpercentage, preprocessing options like lowercase, punctuation and numbers removal, the chosen term-document matrix style. It was also possible to add all the bonus options as well in the legend, but this will make the legend and overall bar chart look dirty and unclear to a user. Hence only the basic fields in the form are considered to be represented in the legend. The below bar chart shows the classifier names and their accuracies as per the form fields submitted in the screenshot above.



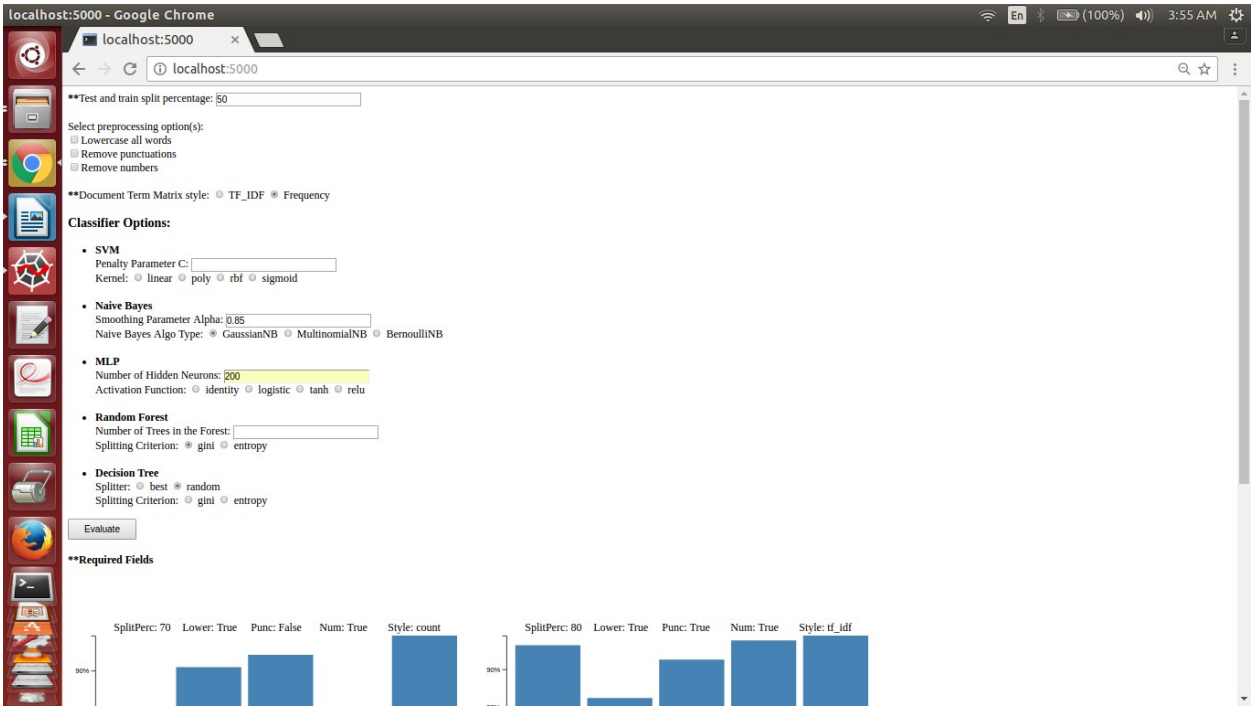
Further, we can fill the form multiple times and it will keep on adding a new bar chart on the page, upon clicking on the evaluate button. This time, form is filled with some different fields and then evaluate button is clicked. The below screenshot shows the filled form.

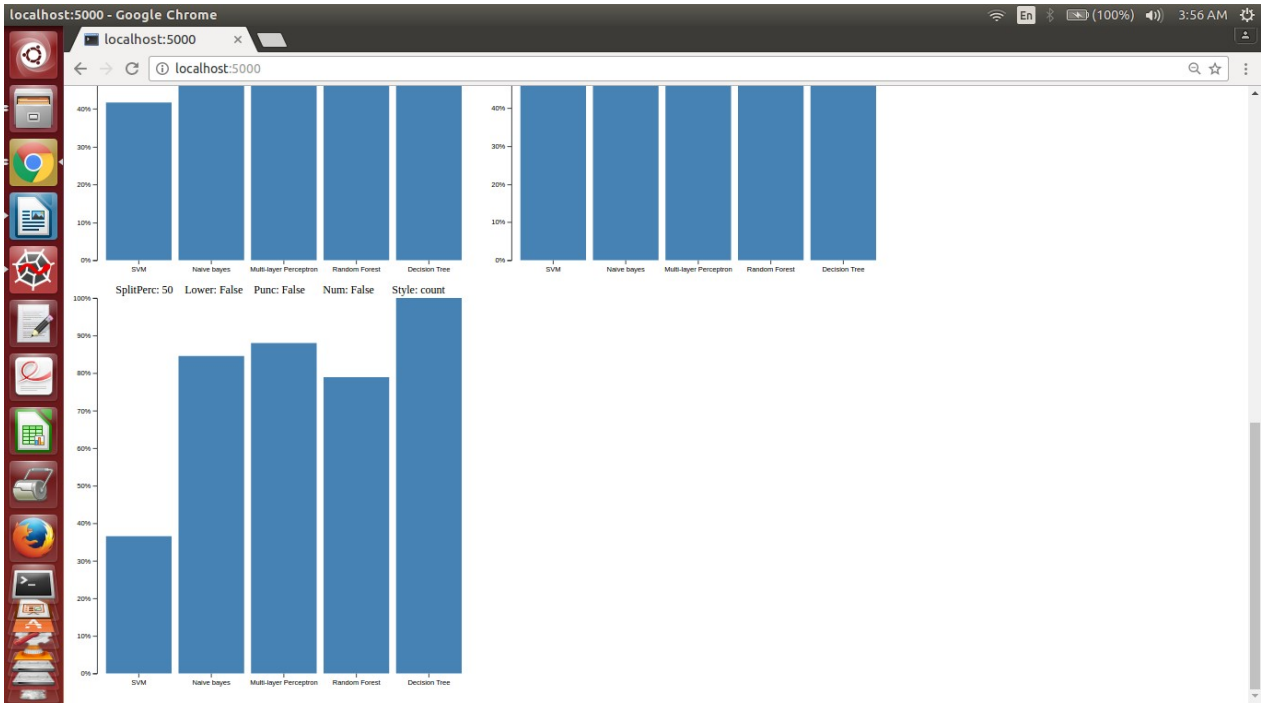


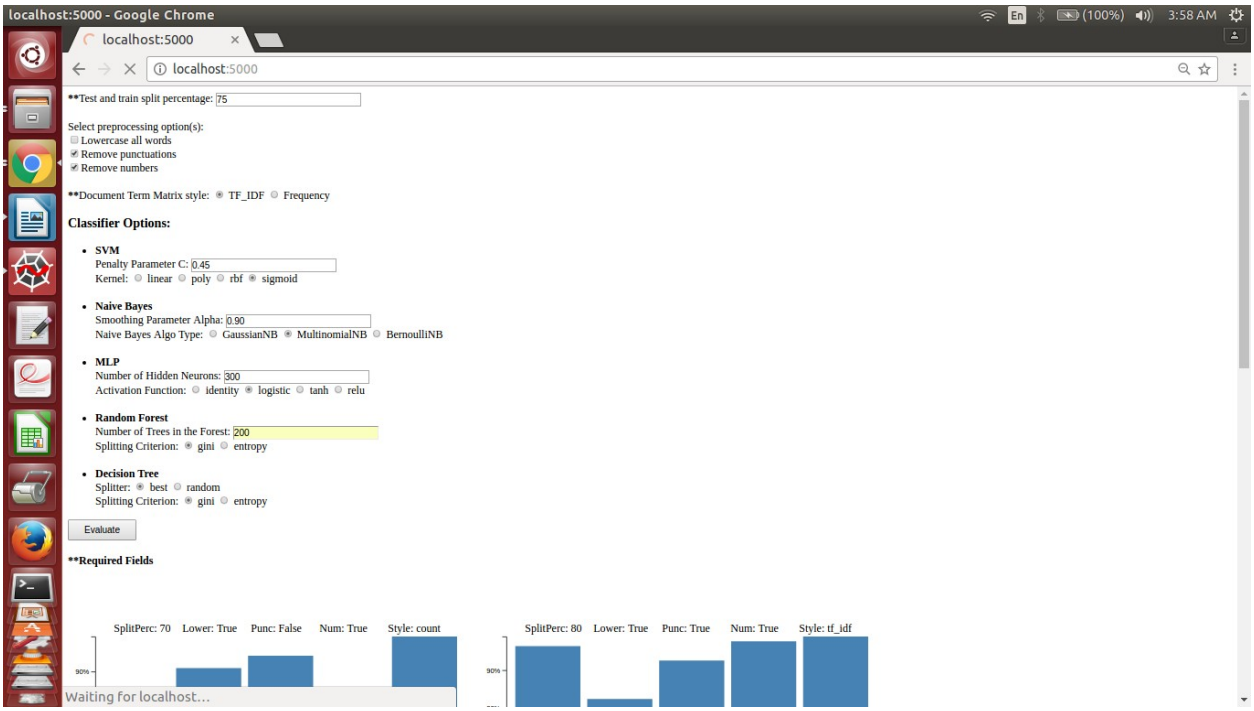

After clicking the evaluate button, a new bar chart gets added to the webpage showing classifier names and their accuracies, along with the corresponding legend, as per the form fields filled in the above screenshot.

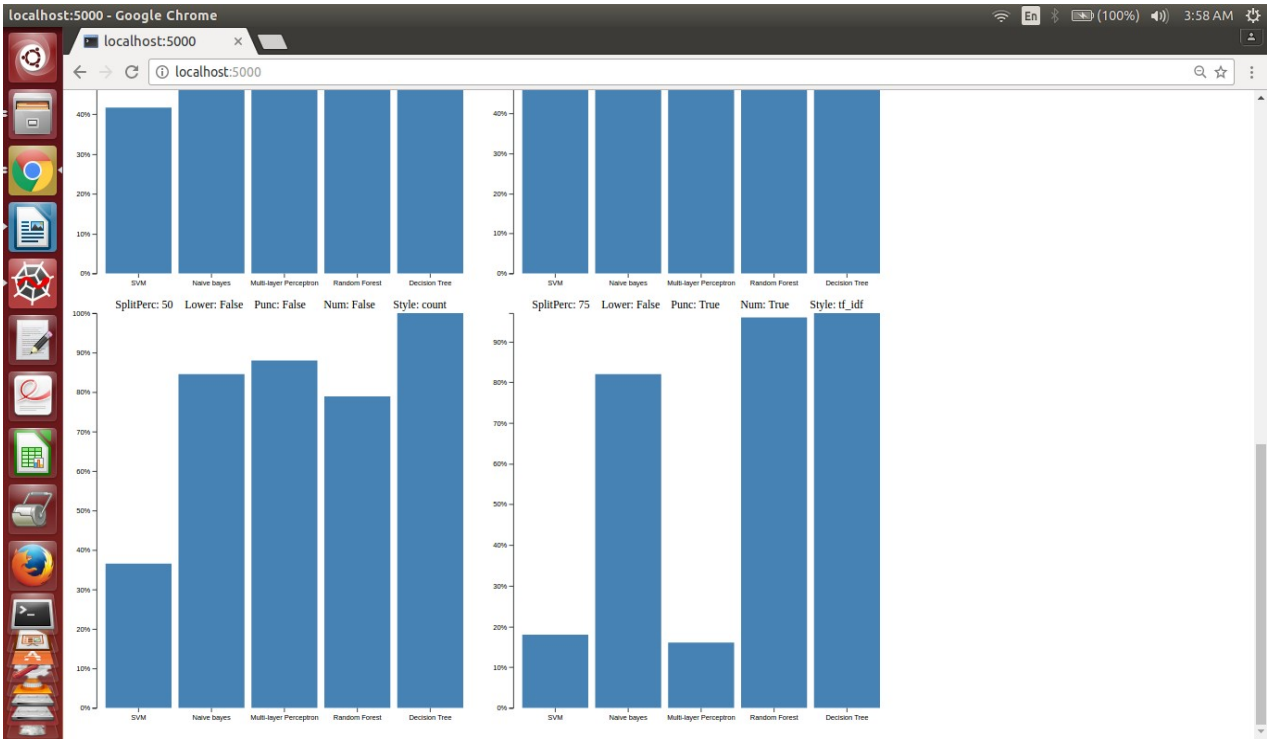


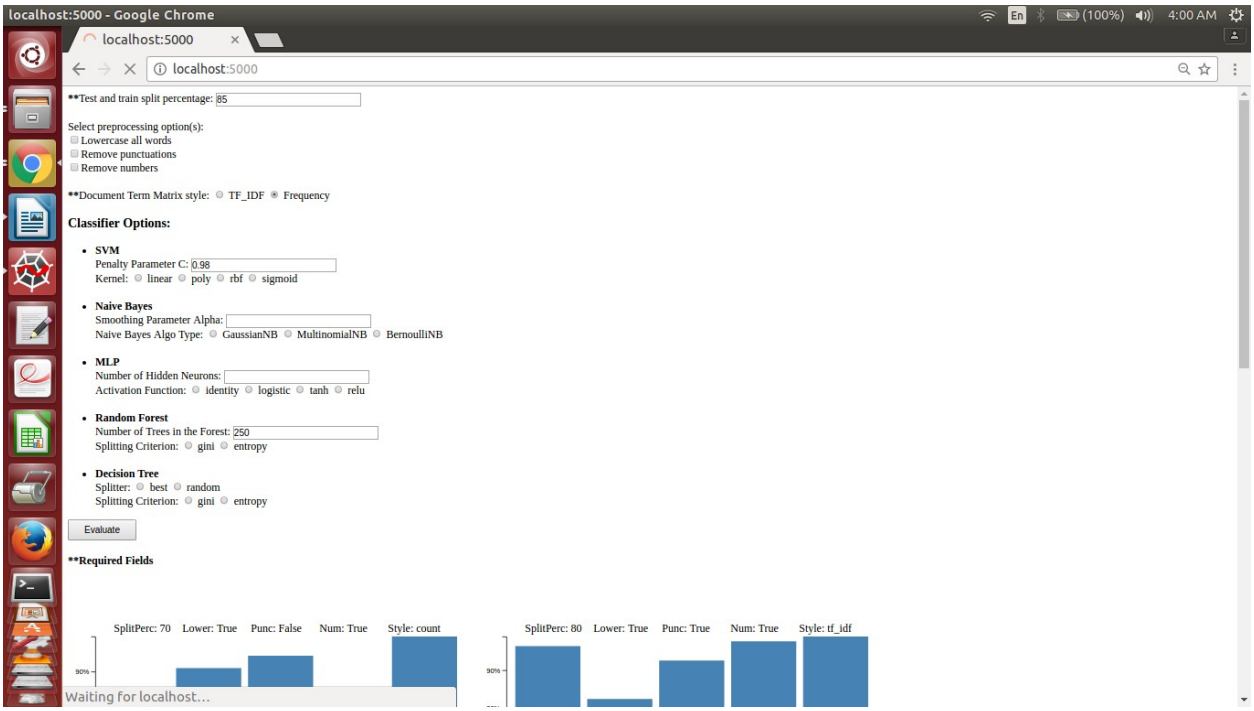
A similar process is repeated few times, and all the screenshots of the filled form and the bar chart added in the webpage are shown below. The bar charts will keep on getting added in the webpage until the connection between front-end and back-end is re-established.

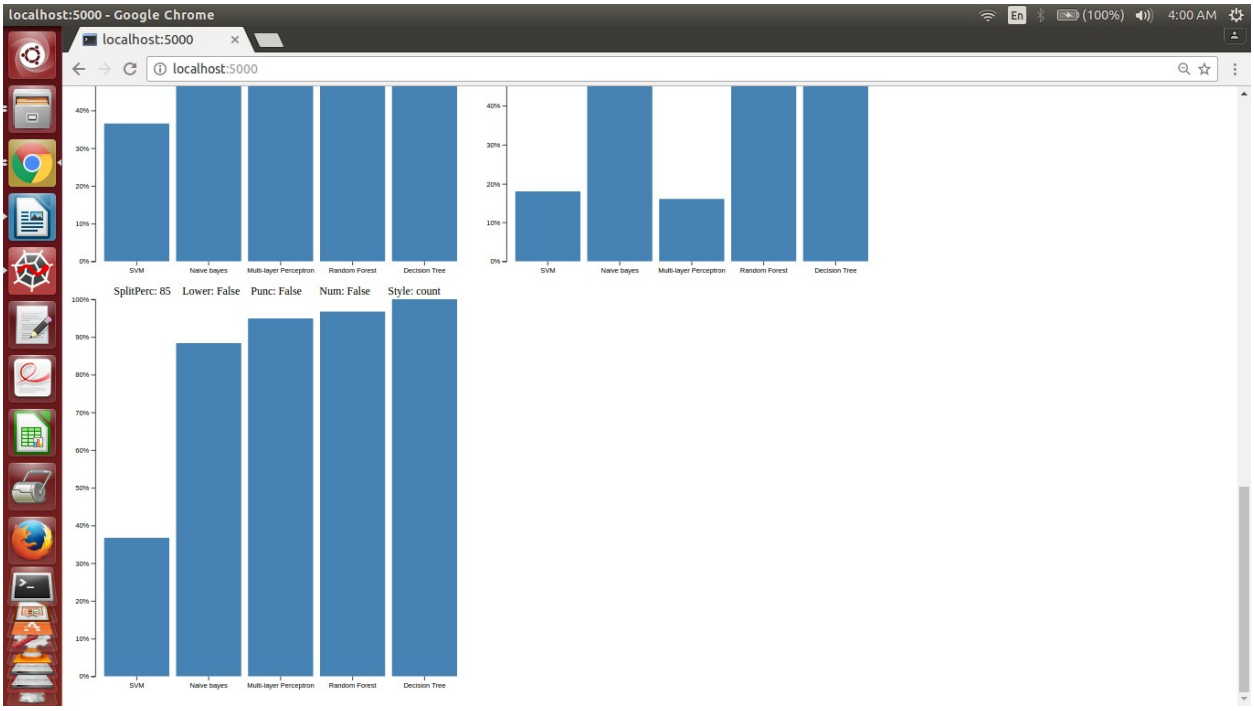



localhost:5000 - Google Chrome
localhost:5000
localhost:5000
**Test and train split percentage: 75
Select preprocessing option(s):
Lowercase all words
Remove punctuations
Remove numbers
**Document Term Matrix style: TF_IDF Frequency
Classifier Options:
SVM
Penalty Parameter C: 0.45
Kernel: linear poly rbf sigmoid
Naive Bayes
Smoothing Parameter Alpha: 0.90
Naive Bayes Algo Type: GaussianNB MultinomialNB BernoulliNB
MLP
Number of Hidden Neurons: 300
Activation Function: identity logistic tanh relu
Random Forest
Number of Trees in the Forest: 200
Splitting Criterion: gini entropy
Decision Tree
Splitter: best random
Splitting Criterion: gini entropy
Evaluate
**Required Fields
SplitPerc: 70 Lower: True Punc: False Num: True Style: count
SplitPerc: 80 Lower: True Punc: True Num: True Style: tf_idf
Waiting for localhost...


localhost:5000 - Google Chrome
localhost:5000
localhost:5000
SplitPerc: 50 Lower: False Punc: False Num: False Style: count
SplitPerc: 75 Lower: False Punc: True Num: True Style: tf_idf
SVM
Naive bayes
Multi-layer Perceptron
Random Forest
Decision Tree




In the back-end, we have considered an extra attribute in the function trainModel(), the name of the attribute is bonus_options, which includes a list of values of all the bonus options. Data is gathered from the form data using flask and this data is passed over to the machine learning part of the python code to evaluate the

accuracies for each classifier considering the filled form options. This data about accuracies is passed back to the webpage on localhost:5000. D3 js library helps in the representation of this passed over json data to be shown as a bar chart.

References –

[1]H. directory?, "How do I list all files of a directory?", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/3207219/how-do-i-list-all-files-of-a-directory

[2]P. matches, "Python extract pattern matches", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/15340582/python-extract-pattern-matches

[3]r. python, "randomizing two lists and maintaining order in python", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/13343347/randomizing-two-lists-and-maintaining-order-in-python

[4]S. python, "Strip spaces/tabs/newlines - python", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/10711116/strip-spaces-tabs-newlines-python

[5]B. Python, "Best way to strip punctuation from a string in Python", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/265960/best-way-to-strip-punctuation-from-a-string-in-python

[6]R. [closed], "Removing numbers from string", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/12851791/removing-numbers-from-string

[7]h. python?, "how to replace punctuation in a string python?", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/12437667/how-to-replace-punctuation-in-a-string-python

[8]N. .toarray(), "NumPy and SciPy - Difference between .todense() and .toarray()", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/30416695/numpy-and-scipy-difference-between-todense-and-toarray

[9]Blog.christianperone.com, 2017. [Online]. Available: http://blog.christianperone.com/2011/09/machine-learning-text-feature-extraction-tf-idf-part-i/

[10]"sklearn.metrics.accuracy_score — scikit-learn 0.19.1 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html

[11]"sklearn.tree.DecisionTreeClassifier — scikit-learn 0.19.1 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-

learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier.predict

[12]"3.2.4.3.1. sklearn.ensemble.RandomForestClassifier — scikit-learn 0.19.1 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[13]"1.4. Support Vector Machines — scikit-learn 0.19.1 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/svm.html

[14]"1.17. Neural network models (supervised) — scikit-learn 0.19.1 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/neural_networks_supervised.html

[15]M. Python, "Map two lists into a dictionary in Python", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/209840/map-two-lists-into-a-dictionary-in-python

[16]P. dict, "Preserve ordering when consolidating two lists into a dict", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/15372949/preserve-ordering-when-consolidating-two-lists-into-a-dict

[17]"18.2. json — JSON encoder and decoder — Python 2.7.14 documentation", Docs.python.org, 2017. [Online]. Available: https://docs.python.org/2/library/json.html

[18]"Bar Chart", Bl.ocks.org, 2017. [Online]. Available: https://bl.ocks.org/mbostock/3885304

[19]P. table, "Python dictionary in to html table", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/14652325/python-dictionary-in-to-html-table

[20]d. chart, "d3js Create legend for bar chart", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/16178710/d3js-create-legend-for-bar-chart

[21]f. [duplicate], "flask: error passing variables to javascript", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/41880725/flask-error-passing-variables-to-javascript

[22]H. [duplicate], "How to obtain values of request variables using Python and Flask", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/13279399/how-to-obtain-values-of-request-variables-using-python-and-flask

[23]C. int, "Cast Flask form value to int", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/12551526/cast-flask-form-value-to-int

[24]C. FLASK?, "Changing values using POST method in FLASK?", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/40182215/changing-values-using-post-method-in-flask

[25]C. application, "Changing a global variable from outside in a Flask based Python web application", Stackoverflow.com, 2017. [Online]. Available: https://stackoverflow.com/questions/16055024/changing-a-global-variable-from-outside-in-a-flask-based-python-web-application