**CS5590 APS - Python Programming**

**LAB2**

**Deadline: 10/10/2018**

The following assignment focus on to make one familiar with Python and Machine Learning

**Lab Assignment:**

1. Pick any dataset from the dataset sheet in the class sheet

a. plot how many of each category is available in your dataset (you can use seaborn library or matplotlib)

b. create one prediction model based on Naïve Bayes Classification and evaluate your model

2. Implement Support Vector Machine classification:

a. Apply SVC with kernel “poly” degree =4

b. Apply SVC with “rbf” kernel

c. change gamma and C parameters in the model to see how the result may change

d. Report the accuracy of the model on both models separately and with which parameters you got better result.

3. Write a program in which take an Input file. Use the simple approach below to summarize a text file:

a. Read a file

b. Apply lemmatization on the words

c. Apply the bigram on the text

d. Calculate the word frequency (bi-gram frequency) of the words (bi-grams)

f. Choose top five bi-grams that have been repeated most

g. Go through the original text that you had in the file

h. Find all the sentences with those most repeated bi-grams

i. Extract those sentences and concatenate

j. Enjoy the summarization

4. Report your views on the k nearest neighbor algorithm when we change the K how it will affect the accuracy. Provide a good justification for the changes of the accuracy when we change the amount of K.

For example: compare the accuracy when K=1 and K is a big number like 50, why the accuracy will change

**LAB Submission Guidelines (for both In Class and Online students):**

1. LAB submission is in pairs of two students.

2. Submit your source code and documentation to GitHub and represent the work through wiki page properly (submit your screenshots as well. The screenshot should have both the code and the output)

3. Comment your code appropriately

4. Video Submission (2 – 3 min video showing the demo of the LAB, with a brief voice over on the code explanation)

5. Submit **only** report at Turnitin in UMKC blackboard

6. Remember that similarity score should be less than **15%**

7. Use this link to submit your LAB#: <https://docs.google.com/forms/d/e/1FAIpQLScjwMFgNiEOKz3A06nb4jlfFvUaDT1HGa9FQs32GyBprr9A8g/viewform>

8. The report should include below details

I. Introduction

II. Objectives

III. Approaches/Methods

IV. Workflow

V. Datasets (if applicable)

VI. Parameters

VII. Evaluation & Discussion

VIII. Conclusion

**LAB Evaluation Criteria:**

1. Report similarly score (should be less than **15%**)

2. Report Quality (check the below example reports for reference)

3. Time (should submit before due time)

4. Wiki page

**Example Reports:**

<https://github.com/stratospark/food-101-keras>

<https://github.com/matterport/Mask_RCNN>

<http://blog.stratospark.com/deep-learning-applied-food-classification-deep-learning-keras.html>

**University Policy on Student Conduct:**

Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: <https://catalog.umkc.edu/special-notices/academic-honesty/>

<https://www.umsystem.edu/ums/rules/collected_rules/programs/ch200/200.010_standard_of_conduct>