EXAMINATION QUESTION PAPER

PG-DAI September 2021

Module Name : Natural Language Processing & Computer Vision

EXAM Type : Main Exam **SET:-**

DATE: 05/03/2022 **Duration: 2hrs.** **Max. Marks: 40**

Instructions:

- All questions are compulsory.
- Question 1 is of 15 marks [5 Strategy + 10 Implementation]
- Question 3 is of 20 marks [5 Strategy + 15 Implementation]
- Write all programs in Python language and use appropriate python libraries. Follow the coding standards (proper method/function and variable name, comments etc.) for best practice.
- **Suggested Python Libraries:** Numpy, SpaCy, Scipy, Sklearn, NLTK, matplotlib, PIL, Pandas, CV2 etc.
- Kindly ensure before creating zip file, make sure all required files should be enclosed to avoid any corruption of zip file.
- Zip file should be renamed with last 3 digit of PRN followed by name of the candidate and exam name (for example: 063-Pratik Gaikwad-Python.zip)

Q.1. Performed text classification or implement sentiment analysis using any classification machine leaning algorithm on provided dataset link (15 Marks)

https://archive.ics.uci.edu/ml/machine-learning-databases/00331/

I. Dataset cleaning
II. Vectorisation
III. Divide dataset into test & train
IV. Classification Algorithm
V. Performed evaluation of model
VI. Accuracy estimate of trained model
VII. Inferencing of trained model on unseen dataset

Q.2. Performed following activities using any NLP library on following text: (05 Marks)

"Once upon a time there was an old mother pig who had three little pigs and not enough food to feed them. So when they were old enough, she sent them out into the world to seek their fortunes.
The first little pig was very lazy. He didn't want to work at all and he built his house out of straw. The second little pig worked a little bit harder but he was somewhat lazy too and he built his house out of sticks. Then, they sang and danced and played together the rest of the day.
The third little pig worked hard all day and built his house with bricks. It was a sturdy house complete with a fine fireplace and chimney. It looked like it could withstand the strongest winds."

I. Tokenization, find & plot of words/token frequency
II. Part of Speech analysis
III. Printing tokens and Boolean values stored in different attributes (stop-words, punctuations and digits)
IV. Remove Stop-words and punctuations
V. Lemmatizing & Stemming the tokens of following sentences

"Such an analysis can reveal features that are not easily visible from the variations in the individual genes and can lead to a picture of expression that is more biologically transparent and accessible to interpretation"

Q.3. Perform Non-Mask (0) vs Mask (1) classification using deep learning architectures. (20 Marks)

The data-set is provided at below link. Both non mask vs mask comprised each of 5000 images. Divide accordingly into train (80%) & test (20%).

https://drive.google.com/file/d/1NGgMO-qd1gYy0uPXY0BX53s5DXUmRgwZ/view?usp=sharing

I. Train from scratch using any Res-Net flavor (preferably 18 to start with).
II. Try two experiments with input size i.e.
   - The input to the architecture should be a fixed size [Batch, Channel, W, H] i.e. Width & Height should be fixed.
   - The input to the architecture should be a arbitrary size [Batch, Channel, W, H] i.e. Width & Height should be variable.
III. The output layer i.e. soft-max will have 2 classes, 0 for non-mask; 1 for mask.

----------ALL THE BEST!!!!! ----------