Logo

Description automatically generated **FACULTY OF COMPUTER SCIENCE AND ENGINEERING**

**Time: 1 hours AI 361 Lab Marks: 50**

**Instructor: Muhammad Muneeb Baig Lab #04**

**Dear AI 2020,**

**NOTE!**

**I would like to remind you that plagiarism is a serious academic offense that can have severe consequences on your academic career. Plagiarism is the act of taking someone else's work and presenting it as your own, without giving proper credit to the original author.**

**To prevent instances of plagiarism, I ask that you record all of your coding activity (via screen recording) and upload the video to a YouTube channel. (**I have to check how you are searching on internet, how you install a new library, how you learn and followa new library.) **Additionally, I ask that you upload the code to your GitHub profile, which will serve as a way to verify your work and ensure that it is original.**

**To upload files on GitHub, please follow these steps:**

1. **Create a GitHub account and sign in.**
2. **Create a new repository by clicking on the "+" icon in the top right corner of the dashboard.**
3. **Give your repository a name and add a description if you like.**
4. **Click on the "Create repository" button.**
5. **On the next screen, you will see an option to add files. Click on "Add files."**
6. **You can either drag and drop your files into the repository or click on "choose your files" to select the files from your computer.**
7. **Once you have added your files, add a commit message to describe the changes you have made.**
8. **Click on the "Commit changes" button to save your files to the repository.**

**Remember to always give credit where credit is due and to create original work. We wish you all the best in your academic pursuits.**

**Task 01:**

**Write a code for N-grams, bi-grams and tri-grams from scratch also find the frequency of each N-gram and plot its bar chart.**

1. **Once you are done with the scratch programming find out POS tags of each grams.**
2. **Now measure the frequency of two grams repeating frequently for example the frequency of (NNP, VVP).**
3. **Plot the graphs of each POS tags.**
4. **Load any corpus from nltk and apply all previous steps on this corpus.**
5. **Now compare your results with Sklearn CountVectorizer results**