Nanyang Technological University

IS6751 Text and Web Mining
Lab: Text Pre-processing and Text Classification

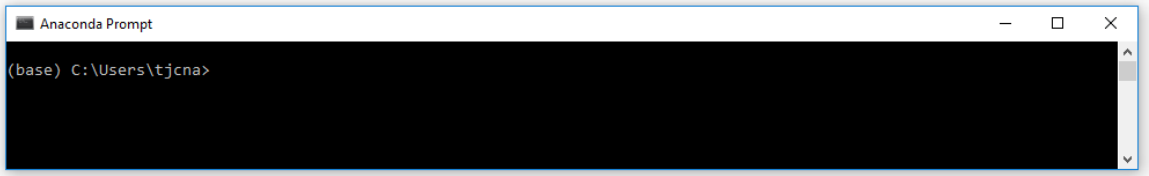
In this lab, you will learn the basic steps of text pre-processing and text classification with Naïve Bayesian algorithms.

1. Install Python tools on a Local Machine.

- **Anaconda** (https://www.continuum.io/downloads) has been installed on lab machines.
  - Windows 64-bit Graphical Installer: **Anaconda3-x.x.x-Windows-x86_64.exe**
- If you cannot run Jupyter Notebook, reinstall Anaconda:
  - Uninstall anaconda by executing E (or D):\Anaconda3**Uninstall-Anaconda.exe**
  - And reinstall anaconda using **Anaconda3-x.x.x-Windows-x86_64.exe** with the following options:
    - Just Me (recommended)
    - Destination folder: E (or D):\Anaconda3

2. Install Python Packages

2.1. Start | Anaconda3 | Anaconda Prompt



2.2. Type the following commands (conda command reference: https://docs.conda.io/projects/conda/en/latest/commands.html)

>>> conda install inflect
>>> pip install textblob
>>> python -m textblob.download_corpora
>>> conda config --add channels conda-forge
>>> pip install contractions

pip install inflect
pip install textblob
python -m textblob.download_corpora
pip install contractions

3. Running the Jupyter Notebook – text-preprocessing-v1.ipynb

3.1. Copy the **notebook files** downloaded from the class website to **C:\Users*youraccountname*\documents\python\IS6751**

3.2 Open a Jupyter Notebook

- File – Open …, **text-preprocessing-v1.ipynb in C:\Users***youraccountname* **\documents\python\IS6751\wk2-text-preprocessing-classification**
- Run the notebook document step-by-step (one cell a time)

4. Running the Jupyter Notebook – text-classification-v1.ipynb

4.1 Open a Jupyter Notebook

- File – Open …, **text-classification-v1.ipynb in C:\Users***youraccountname* **\documents\python\IS6751\wk2-text-preprocessing-classification**
- Run the notebook document step-by-step (one cell a time)