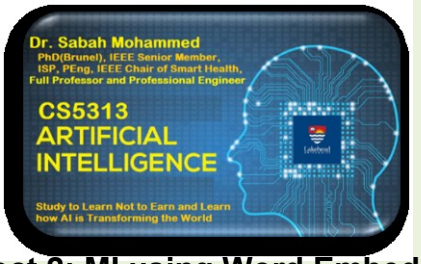
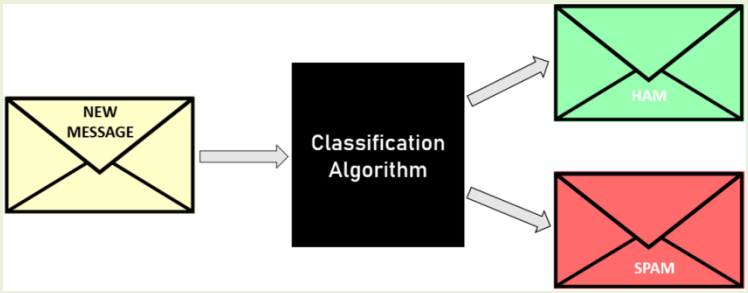
COMP5313 Artificial Intelligence

Department of Computer Science



Project 2: MI using Word Embedding



In this project you are going to use variety of machine learning techniques to classify messages (SMS or Tweets) into Spam or Ham. This project is an important exercise for text classification which is one of the most common natural language processing tasks. You are requested to use at the following two methods of classifications:

1. Use Traditional Machine Learning Classifiers from SKLearn:

```
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
```

2. Use Word Embedding Layers with Deep Learning. For word embedding I recommend to use Glove model:
https://nlp.stanford.edu/projects/glove/
https://medium.com/analytics-vidhya/basics-of-using-pre-trained-glove-vectors-in-python-d38905f356db

You are requested also to evaluate the predictions of both sets of classifiers.

HINTs:
1. Use the UCI Dataset:
https://archive.ics.uci.edu/ml/datasets/Spambase
2. For the GloVe model use:
https://www.kaggle.com/danielwillgeorge/glove6b100dtxt

Submission Details:

1. One ZIP file (Other compression types like RAR are **NOT** Acceptable) containing the source file (YourName_Project2.py) + **ReadMe.pdf** (MS Word is **NOT** acceptable) describing the idea of your program + Screen Shot of the outputs + the Jupyter File (IPYNB)
2. Submit to D2L only before due date (**One hr delay take 1 Mark up to three hours**).
3. It must be your individual work.
4. Double submissions are **not** allowed.
5.

Important Note: **Respecting the student behaviour code is highly appreciated and any submission found with high similarity with other students solutions or from solutions over the Internet will be rejected.**
https://www.lakeheadu.ca/faculty-and-staff/policies/student-related/code-of-student-behaviour-and-disciplinary-procedures