# Assignment-8: Decision Tree Classifier

## Course: Artificial Intelligence (CS571)

(Read all the instructions carefully & adhere to them.)

Date: 17-10-2019 Deadline: 25-10-2019

Write a Python program that implements Question classification using Decision Tree classifier.

**Example**

**Question**: What is the temperature at the center of the earth ?

**Class**: NUM, which refers to the question that looks for the numeric type answer.

**Dataset**

Training Set: http://cogcomp.org/Data/QA/QC/train_5500.label

Test Set: http://cogcomp.org/Data/QA/QC/TREC_10.label.

Use only the coarse grained class label to build your model. For more details about the dataset follow these paper: https://goo.gl/jAJFKQ

**Features**

(a) Length of the question

(b) Lexical Features: Word n-gram.

(c) Syntactic Features: Parts of speech tag unigrams.

Implement n-gram ( n=1,2 and 3) features for each question instance. You may choose only the most frequent n-grams to provide as a features for your model. For n=1, use 500 most frequent 1-gram, similarly use 300 and 200 most frequent n-grams, for n=2 and 3 respectively. For parts of speech tag unigrams, first you need to get POS tag for each question instance. Use can use any library like Stanford POS tag-ger see https://nlp.stanford.edu/software/tagger.shtml NLTK POS tagger see http://www.nltk.org/book/ch05.html etc. Similar to lexical feature use 500 most frequent 1-gram to build the model. For more details about the features, see section 2.3 in the following paper https://goo.gl/X7X7ox.

### Result and Evaluation

- Report the 10-fold cross-validation results in terms of precision, recall, and F-score.
- Report results of feature ablation study and state which feature has contributed most towards correctly predicting a particular class.

### Instructions

- Please submit your assignment here: http://tiny.cc/6t3oez
- The submission file should be as follows:

**Group-NUMBER Assignment-NUMBER.zip**