CSE 3024: Web Mining **Slot: L51 + L52**

Online Submission Deadline: 26th September 2020

Web Structure Mining, Supervised Learning

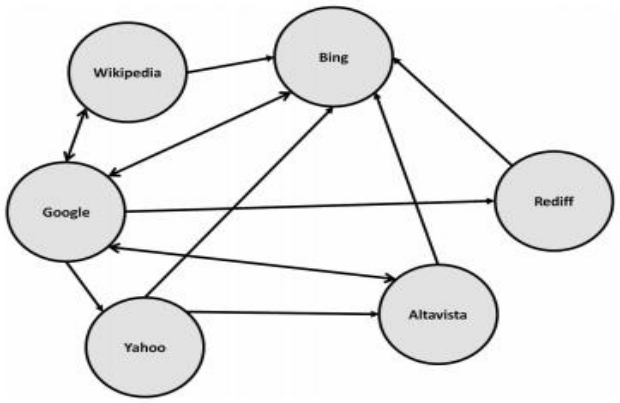
[3 + 3 + 4]

- **Upload your code and result as a single PDF file in VTOP and MOODLE**
- **File should contain**
  - **Question**
  - **Code**
  - **Result / Output screen (including contents of all generated files)**

---

1. Write a python program to calculate the degree prestige, proximity prestige and rank prestige using a graph dataset given in the following link.
   http://snap.stanford.edu/data/wiki-Vote.txt.gz

2. Write a python program to show the implementation HITS algorithm for the following graph and display the authority as well as hub score for all the nodes. (**stopping criteria:- ε = 0.04 for both hub and authority)**



3. Write a python program to show the implementation of Decision Tree and Naïve-Bayes techniques using the below mentioned dataset..
   - Handle missing values, If any
   - Use 5-fold cross validation technique
   - Prepare the confusion matrix, find out the precision, recall value, F-measure and prediction accuracy.
   - Prepare ROC and AUC curve based on the result obtained.
   - Compare the results obtained using these two techniques in order to assess their performance for the considered dataset.

   https://drive.google.com/open?id=14rXh1RUXBa2bWl5nmb5UiQM7tx1EeS8f

   The detailed description of the dataset is given in the below link:
   https://archive.ics.uci.edu/ml/datasets/Credit+Approval