Computer Processing of Natural Language (CSCE771-001-FALL-2020)

By Avineet Kumar Singh

**Project Title**: Question Answering Application using FAQs.  
**Description:** Building a model which can retrieve or predict answers from a given datasets. This can be further be implemented upon user manuals or huge set of documents.

Note: May add document summarization and Knowledge graphs.

**Illustrative testcases:**

Input: A question from the user.

Output: Nearest answer based on the available dataset(FAQs).

**Data Sources:**

1) <https://www.kaggle.com/stanfordu/stanford-question-answering-dataset>

Note: Will add more.

**Techniques and Tools to use:**

1) BERT embeddings for prediction.

2) genism for text processing.

**Metric of measuring output:**

1) Objective evaluation: label the data and check number of times we get correct answers.

2) BLEURT proposed by Google Research.

**How will you collect results?**

Results would be in form of the evaluation metrics, as to how the model performed.

**Format of Report:**

PDF document including code examples.

**Time Schedule:**

Till Sep 20: Clear out the objectives and the possible techniques which can be used.

Till Sep 28: Work on data collection.

Till Oct 2: Text-Processing

Till Oct 15: Train the model.

Till Oct 26: Literature Review.