AMANUL HAQUE

https://ahaque2.github.io/ | ahaque2@ncsu.edu |(919)-946-6094

EDUCATION

Doctor of Philosophy (PhD) in Computer Science — *GPA - **4.0/4*** — *Expected May 2023*
North Carolina State University, Raleigh, NC
With a focus on: Artificial Intelligence, Machine Learning and Natural Language Processing

Master of Science in Computer Science — *CGPA - **4.0/4*** — *May 2019*
North Carolina State University, Raleigh, NC

Bachelor of Engineering in Information Science — *CGPA - **8.76/10*** — *May 2015*
PES University, Bangalore, India

RESEARCH AND PROJECTS

- ***Identifying Preferred Entity in User Generated Text*** (Github) — ***Python, Tensorflow***
  Created Pixie, a gold standard dataset of ~9k labeled sentences to identify preferred entity in user generated comparative sentences from app reviews. Our dataset includes implicit and indirect comparisons, something previous works have neglected. We achieved an F1-score of 84% by fine-tuning BERT with segment embeddings to demarcate compared entities, while SOTA achieved 74% F1-score on Pixie (*under review at ACL 2022*)
- ***Emotional Affects of COVID-19 News Coverage*** (Github) — ***Python, Tensorflow, Gensim***
  Analyzed the emotional response to COVID-19 News on Twitter by popular English news media from 12 countries. Curated a dataset of Covid-19 related News Tweets and user responses to them (~20M tweets). Used RoBerta model fine-tuned for emotion recognition task and LDA to identify discussion topics *(Accepted at IEEE Access)*
- ***Understanding echo chambers via social simulation (Github)*** — ***Python, Mesa***
  Designed a multiagent social simulation with independent agents having political predisposition. We use this to understand the factors contributing to political polarization and formation of echo chambers as a result of user sharing behavior. We also analyze the effect of filter bubbles created by collaborative filtering algorithms.
- ***From Collaboration Characteristics to Code Quality (Github)*** — ***Python, Gensim, SonarQube***
  Analyzed collaboration characteristics and team performance of Software Development teams based on an empirical study 168 Github Projects. Proposed metrics to measure collaboration characteristics using linguistic features from online interactions among team members and measured team performance based on code quality.
- ***Prioritizing student discussion posts in Online Courses (Masters' Thesis)*** — ***Python, Gensim, Scikit-Learn***
  Designed a semi-supervised model to identify posts needing immediate instructor attention on online student discussion forums in MOOCs to reduce response time. The model uses linguistic features to categorize posts based on content, like, questions, answers, socializing posts, software issues, etc and identifies students needing help.

PROFESSIONAL EXPERIENCE

Seagate, Longmont, Colorado, Machine Learning Summer Intern — ***May 2020 - Aug 2020***
- Designed a graph based unsupervised abstractive multi-document text summarizer for a social listening tool to identify trending online topics and summarize related documents.
- Implemented an unsupervised aspect based sentiment analyzer for online user reviews.

Lenovo, Morrisville, NC, Computer Science Summer Intern — ***May 2018 - Aug 2018***
- Designed modules to automate test plan generation using Requirement Management Kanban (RMK) and historical test results to reduce test suite execution time.
- Designed an information extraction model to identify executable commands from unstructured text in RMK.

Oracle, Bangalore, India, Software Developer — ***July 2015 - June 2017***
- Collaborated with a global team to design and develop modules for Service Deployment Infrastructure (SDI) which governs the provisioning flow for all Oracle Public Cloud (OPC) subscription life-cycle.
- Implemented modules for a data center level load balancer and a loosely coupled execution mode to improve runtime and increase parallelism in execution.

SKILLS

- *Programming Languages*: Python, Java, Shell, SQL, PHP, HTML, JavaScript
- *API and Libraries:* PyTorch, Tensorflow, Keras, Scikit Learn, Gensim, Numpy, Pandas, REST