AKANKSHA DARA

Stony Brook University | Apple | BITS Pilani

Stony Brook, New York   @ akankshadara7@gmail.com   ☎ +1-631-710-1468   in akanksha-dara   AkankshaDara

EDUCATION

MS in Computer Science

Stony Brook University

2020-2021   Stony Brook, NY

- Graduate Teaching Assistant for the course CSE-216 (Programming Abstractions).)
- Graduate NLP researcher under (*Prof. Ritwik Bannerjee*)

B.E.(Hons.) in Computer Science

Birla Institute of Technology and Science, Pilani (BITS Pilani)

2014-2018   Pilani, India

- **Teaching Assistant** for the course CS-F111 (Computer Programming)
- **Undergraduate Research Thesis** at the University of Melbourne, Australia Fall 2017 — (Grade: 10/10)
- **CGPA:** 7.80/10

EXPERIENCE

SOFTWARE DEVELOPMENT ENGINEER

Apple Inc.

October 2018 – January 2020   Hyderabad, India

- **Java Development:** Implemented a **Solr cluster** solution to facilitate efficient management of the test cases. Delivered business logic in back-end APIs and an interactive web application.
- Automated the process of publishing code coverage reports using JaCoCo (Java Code Coverage) as an agent across all the servers by generating the build files dynamically; analysed the coverage for reflection classes in the source code. Used these results to optimise the test cases by identifying redundancy in code execution.

RESEARCH ASSISTANT

Memory and Language Lab, University of Melbourne
(Advisor: Prof. Simon Dennis)

July 2017 - December 2017   Melbourne, Australia

- Built a computational model of language processing based on mathematical model of human memory using the stochastic gradient descent algorithm.
- The model characterized sentence processing and learning as an interaction of three memory systems (lexical, syntactic, and relational) that operate on distributed instance-based knowledge representations.
- Conducted several experiments to optimize the model with different input feature sets, objective functions, corpora and tokenization schemes. This helped in gaining insight into the operation of the model, in particular the impact of interference on memory based algorithms.
- Achieved a perplexity of 42 on a vocabulary size of 65,536.

ACADEMIC PROJECTS

Sarcasm Detection on Twitter Data: Using Behavioral Modelling Approach

- Implemented sarcasm detection as a binary classification problem using the User's behavior modelling approach. Identified sarcasm as three different forms of expressions.
- This was achieved by leveraging users' historical information of past tweets and by identifying sarcasm as a contrast of sentiments, as a means of conveying emotion and as a function of familiarity, which I further translated into feature sets for training various supervised learning algorithms; made some more enhancements such as handling the use of emoticons and complex hashtags effectively to improvise the results.

HMV - Medical Decision Support Framework

- Implemented an ensemble framework using hierarchical majority voting and multi-layer classification for disease classification and prediction using data mining techniques.
- The model overcomes the limitations of conventional performance bottlenecks by utilizing an ensemble of seven heterogeneous classifiers.

Text Summarization using Audio Retieval

- Generated a textual summary for a noise-free audio dataset using audio frequency and amplitude.
- Modified LexRank algorithm to visualize sentences as vertices of a graph while taking into account the term frequencies of keywords, and the frequencies of words spoken as appearing in the audio waveform generated.
- Implemented the seq2seq model in Tensorflow and trained it on the CNN Daily Mail data set to obtain a Rouge-1 score of 0.39 and compared it's performance to that of the LexRank algorithm (Rouge-1 score of 0.59).

Compiler Construction

- Implemented a compiler for the language ERPLAG. The project was done in a pipelined manner with its various stages being: Lexer - Parser - Abstract Syntax Tree generation - Type-checking, Semantic checking and Assembly Code Generation.

SKILLS

- **Over 50,000 lines:** C/C++
- **Over 5,000 lines:** python, java
- **Machine Learning:** scikit-learn, TensorFlow, openCV, pandas, NLTK
- **Search and Databases:** Solr, MongoDB, mySQL
- **Operating system** : Windows, Linux, macOS
- **Web:** HTML/CSS/JS, Spring, Flask, D3.js