AAKASH BHATIA

✉ 1 Coed Lane, Stony Brook, NY, 11790   ☎ +1 (408)-668-5501   @ asbhatia@cs.stonybrook.edu   in aakashsbhatia

○ aakashsbhatia2   🌐 https://aakashsbhatia2.github.io/

PROFESSIONAL EXPERIENCE

Stony Brook University   **Stony Brook, NY**
Graduate Researcher - Natural Language Processing and Machine Learning   *Jan 2020 to present*
- Advised by Dr. Ritwik Banerjee
- Research area: Study the spread of misinformation in medical news.
- Master's Thesis: Data-set creation and Machine Learning model development to study textual entailment and perform classification of medical news.

Ernst & Young LLP   **Mumbai, India**
Associate Consultant   *June 2016 to July 2019*
- Performed anomaly detection on client systems in the Financial Services sector and developed applications to automate audit procedures, reducing weekly effort by over 75%.
- Led a team of 6 to conduct an I.T. General and Application Controls for a multinational Investment Bank. My team was awarded the EY Spotlight Award for delivering exceptional quality on this engagement.

SKILLS

Programming Languages: Python, SQL, JavaScript, C, C++, Java

Data Science Libraries: Pandas, Numpy, Scikit-Learn, NLTK, TensorFlow, PyTorch, d3.js

Deep Learning Frameworks: DAN, CNN, RNN/ LSTM, GRU, Transformers, Google BERT

Technologies: Docker, Apache Hadoop, Apache Spark, IBM DB2, Linux Operating System, Git

PROJECTS

Word2Vec - Skip Gram Model   **Stony Brook University**
Course project: CSE 538 - Natural Language Processing   *Fall 2020*
- Developed a Word2Vec - Skip Gram model to train word embeddings using the Text8 (Wikipedia) data corpus by Matt Mahoney.
- Implementation was performed using cross-entropy and noise contrastive estimation loss functions.
- Technologies Used: Python, TensorFlow

Classification of Deceptive Hotel Reviews   **Stony Brook University**
Course project: CSE 512 - Machine Learning   *Spring 2020*
- Re-implementation of an NLP Research paper to perform classification of truthful and deceptive reviews for the top 25 hotels in Chicago obtained from Trip Advisor. Classification was performed using Naive Bayes classifier and Support Vector Machines (SVM).
- Key Finding: Using the SVM classifier along with trigram encoding and tf-idf, we were able to obtain an accuracy of 89.75% and an F1 score of 89.94%.
- Technologies Used: Python, Scikit-Learn, NLTK

COVID-19 Analytics using Hadoop and Spark   **Stony Brook University**
Course project: CSE 532 - Theory of Database Systems   *Spring 2020*
- Studied the spread of COVID-19 from Dec '19 to Mar '20 using MapReduce.
- Statistics derived: No. of cases per million, No. of cases for a given time-period, No. of cases per country, city and zip-code.
- Technologies Used: Docker, Python, Java, Apache Hadoop, Apache Spark

Interactive Dashboard for Road Accident Analysis   **Stony Brook University**
Course project: CSE 564 - Visualisation and Visual Analytics   *Spring 2020*
- Developed a single screen, interactive dashboard to study the effects of weather conditions on road accidents from 2014-2019.
- Performed dimension reduction using Principal Component Analysis (PCA), stratified sampling using K-means clustering and visualised over 1 million data-points.
- Key Findings: Using a combination of a PCA scatter-plot and parallel co-ordinate chart, we identified distinct clusters of data-points. These clusters represented the weather conditions responsible for accidents based on their severity.
- Technologies Used: Python, d3.js, Flask

EDUCATION

Stony Brook University   **Stony Brook, NY**
M.S. Computer Science - Data Science and Engineering   *Expected Graduation: May 2021*
Relevant Coursework: Machine Learning, Natural Language Processing, Artificial Intelligence, Data Visualisation and Visual Analytics, Theory of Database Systems, Analysis of Algorithms

University of Mumbai   **Mumbai, India**
B.E. Computer Engineering   *Aug 2012 to May 2016*