

# AAKASH BHATIA

✉ 1 Coed Lane, Stony Brook, NY, 11790    ☎ +1 (408)-668-5501    @ asbhatia@cs.stonybrook.edu    in aakashsbhatia  
🌐 aakashsbhatia2    🌐 https://aakashsbhatia2.github.io/

## EDUCATION

### Stony Brook University

M.S. Computer Science (Thesis) - Data Science and Engineering

Stony Brook, NY

Expected Graduation: May 2021

CGPA: 3.51/4.0

**Relevant Coursework:** Machine Learning, Natural Language Processing, Artificial Intelligence, Data Visualisation and Visual Analytics, Theory of Database Systems, Analysis of Algorithms

### University of Mumbai

B.E. Computer Engineering

Mumbai, India

Aug 2012 to May 2016

CGPA: 3.50/4.0

## SKILLS

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

**Machine Learning Models:** Perceptron, SVM, Naive Bayes, Regression, Decision Trees and Random Forests, Clustering, Boosting, Bagging, DAN, CNN, RNN, LSTM, GRU, Transformers, BERT

**Data Science:** Statistics, Probability, Hypothesis Testing, PCA, Pandas, Numpy, Scikit-Learn, NLTK, TensorFlow, PyTorch, d3.js, MLlib

**Technologies:** Docker, Amazon Web Services (AWS), Ajax, Apache Hadoop, PySpark, Apache Spark, DB2, MSSQL, MySQL, Linux, Git

## WORK EXPERIENCE

### Ernst & Young LLP

Risk Analytics - Information Technology Risk Consultant

Mumbai, India

June 2016 to July 2019

- Used a data-driven approach to analyze the design and operating effectiveness of application access management, software change management and application infrastructure management for 8 clients.
- Experienced in working with cross-functional teams and presenting findings to technical and non-technical stakeholders.
- Increased efficiency by 60% by developing data-driven applications using Python and SQL to automate technology audit analytics.
- Managed the EY India team of 6 people to deliver a global risk analytics project spanning 5 countries. My team was awarded the EY Spotlight Award for delivering exceptional quality for this project.

## ACADEMIC RESEARCH

### Stony Brook University

Master's thesis (Natural Language Processing and Machine Learning) - Advised by Dr. Ritwik Banerjee

Stony Brook, NY

Jan 2020 to present

- Introducing a novel data-set with 10,000+ unique news-claims to study semantic change in news.
- Crowd-sourcing a labelled dataset via AWS MTurk to generate 60,000+ semantically changed news-claims.
- Classifying the dataset using standard machine learning models and state-of-the-art NLP models. Performing a quantitative analysis of the results to highlight gaps in the learning task.

## PROJECTS

### Author's sentiment prediction

Stony Brook University - Fall 2020

- Inferred an author's sentiment towards the main entity in a news article using a Deep Averaging Network and BERT encoders.
- The model obtained an F-1 score of 0.48. Analysis of the results showed that the model found it difficult to distinguish between author specific sentiments and the general sentiment of an article.
- Python, PyTorch, Huggingface

### Deep learning models for natural language processing

Stony Brook University - Fall 2020

- Implemented a neural-network based transition parsing (arc-standard algorithm) model with a custom cube activation function. Carried out a comparative study between the cubic, tanh and sigmoid activation functions. The cubic activation function improved the Unlabelled Attachment Score (UAS) of the model by 2% as compared to the sigmoid activation function.
- Performed entity-relation extraction using a bi-directional GRU with custom attention layer. The model achieved an F1-score of 0.57.
- Python, TensorFlow

### Classification of deceptive hotel reviews

Stony Brook University - Spring 2020

- Classified truthful and deceptive reviews from Trip Advisor using Naive Bayes classifier and Support Vector Machines (SVM).
- SVM classifier using tri-grams and tf-idf gave the best results. Accuracy of 0.8975 and F1 score of 0.8994 was achieved.
- Python, Scikit-Learn, NLTK

### COVID-19 analytics using Hadoop and Spark

Stony Brook University - Spring 2020

- Generated statistics related to the spread of COVID -19 from Dec '19 to Mar '20 using MapReduce. The statistics that were derived were - number of cases per million, number of cases for a given time-period and number of cases per country, city and zip-code.
- Docker, Python, Java, Apache Hadoop, Apache Spark