

AAKASH BHATIA

I am an aspiring Data Scientist with a keen interest in Natural Language Processing and Machine Learning. I am actively seeking full-time opportunities as a Data Scientist/ Machine Learning Engineer.

📍 Stony Brook, NY @ asbhatia@cs.stonybrook.edu ☎ +1 (408) 668 5501
🌐 aakashsbhatia2.github.io

in aakashsbhatia 🔗 aakashsbhatia2

PROFESSIONAL EXPERIENCE

Stony Brook University

Graduate Researcher, Advised by Dr. Ritwik Banerjee

📅 January 2020 - Present 📍 Stony Brook, NY

- Study the spread of misinformation in medical news.
- Data-set creation and Machine Learning model development to study textual entailment and perform classification of medical news.

Ernst & Young LLP

Associate Consultant - Information Technology Risk and Assurance

📅 June 2016 - July 2019 📍 Mumbai, India

- Performed anomaly detection on client systems 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.

PROJECTS

Classification of Deceptive Hotel Reviews

📅 Spring 2020

- Re-implementation of an NLP Research paper as a course project for CSE 512 - Machine Learning.
- Performed classification of truthful and deceptive reviews for the top 25 hotels in Chicago from Trip Advisor 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

📅 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

📅 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

M.S. Computer Science

Focus: Data Science and Engineering

Stony Brook University 🎓 GPA: 3.42/4.0

📅 Expected Graduation: May '21

B.E. Computer Engineering

University of Mumbai 🎓 GPA: 3.40/4.0

📅 May '16

SKILLS

Programming Languages

Python SQL JavaScript C/ C++

Java

Data Science Libraries

Pandas Numpy Scikit-Learn

NLTK TensorFlow Pytorch d3.js

Technologies

Docker Hadoop Spark DB2

Linux Operating System Git

Professional Skills

Collaboration Critical Thinking

Leadership

RELEVANT COURSES

Machine Learning

- PAC Learning, Supervised ML Models: Perceptron, Linear and Logistic Regression, Boosting, SVM, Decision Trees
- Unsupervised ML Models: Clustering
- Model Selection, Loss functions, Neural Networks

Artificial Intelligence

- Search Algorithms: DFS, BFS, A*, UCS
- Game Theory: Mini-max, Expecti-max, Pruning (Alpha-Beta Pruning)
- First Order Logic, Bayesian Networks, Expectation Maximization

Data Visualisation and Visual Analytics

- Create interactive visualisations using standard and non-standard Visualisation techniques
- Principal Component Analysis, Multi-Dimensional Scaling, Clustering, Sampling techniques