# Saurabh Raut

sraut@seas.upenn.edu | (+1)-267-902-6290 | Philadelphia, USA

## EDUCATION

**University of Pennsylvania (UPenn)**  Philadelphia, USA
*Master of Science in Engineering - Computer and Information Science*  *May 2023*

**Veermata Jijabai Technological Institute (VJTI)**  Mumbai, India
*Bachelor of Technology in Computer and Information Science*  *May 2021*

## INTERNSHIPS

**Amazon Music (Software Development Engineer)**  May 2022 – August 2022
- Designed, implemented and deployed a caching framework for Amazon Music's 3P evaluation framework
- Extensively used CDK and AWS tools like Lambda, S3, SQS and DynamoDB for implementing the cache

**Barclays Investment Bank (Software Development Engineer)**  June 2020 – July 2020
- Automated testing of Barclays trade application - BATMAN by using Barclays Automation Test Harness System(BATHS) and testNG framework

**Silicon Veins (Android Developer)**  May 2019 – June 2019
- Built custom Share View replacing the built-in Android chooser for BizAnalyst app. This feature lets users directly share app data to mobile numbers on the ledger through Whatsapp without saving the number in their contacts
- Set up passcode activity, floating action button and other UI related features for the app

## SELECTED PROJECTS

**Penn Search - Distributed Search Engine** | *Java, AWS tools*
- Built a completely distributed search engine consisting of Web Crawler, Indexer, Page Ranker, & an interactive UI
- About 1 million web pages were crawled using StormLite, a lightweight version of Apache Storm, running on EC2 instances. SQS was used as a crawl frontier and the crawled pages were stored in a S3 bucket
- The pages were indexed and ranked as Hadoop MapReduce tasks and search queries from user were handled by self-built version of Apache Spark

**Penn Cloud** | *C++, HTML, CSS*
- Built a cloud platform that supports web-mail service and storage service analogous to Gmail and Google Drive
- Storage was based on distributed key-value store (similar to Bigtable) and factors like load balancing, fault tolerance, replication, consistency were taken into consideration as well

**Chat Server** | *C++*
- Built a replicated chat server that uses User Datagram Protocol (UDP) to multicast chat messages
- The server supports three different ordering modes: unordered, FIFO, and total ordered multicast

**Emotion Analysis of Movie Characters** | *Scikit-learn, PyTorch, Pandas*
- Performed scene wise emotion analysis of movie characters based on their dialogues in the script
- Trained and tested a few linear and non-linear models and found that a simple MLP (Multi-Layer Perceptron) model with BERT embeddings generated best results

## RELEVANT COURSEWORK

Software Systems, Internet and Web Systems, Applied Machine Learning, Principles of Deep Learning, Linear Algebra and Optimization, Database and Information Systems, Analysis of Algorithms, Engineering Negotiation, Data Structures & Algorithms, Discrete mathematics

## TECHNICAL SKILLS

**Languages:** Java, C/C++, Python, TypeScript, SQL, Latex
**Frameworks & Software:** MongoDB, Neo4j, TestNG, Scikit-learn, Numpy, Pandas, Hadoop MapReduce, Apache Spark
**Development Tools:** CDK, AWS (RDS, DynamoDB, S3, SQS, EC2, Lambda), Android Studio, Git

## PUBLICATIONS

**Content Labelling of Hidden Services With Keyword Extraction**  Jan 2021 – May 2021
*Dark Web Pattern Recognition and Crime Analysis Using Machine Intelligence - IGI Global*
- Classified the darknet data into topics using two different approaches
- Approach 1: Topic modelling using LDA
- Approach 2: Generated Graph of Words of darknet data and used centrality measures for topic classification