

EDUCATION

New York University, Tandon School of Engineering, Brooklyn, New York	May 2019
Master of Science in Computer Science, GPA-3.9, Received the Academic Excellence Award	
Narsee Monjee Institute of Management Studies, MPSTME, Mumbai, India	May 2017
Bachelor of Technology in Computer Engineering, GPA-3.7	

TECHNICAL SKILLS

Languages: Python, SQL, Scala, C++, Java, Bash, HTML, CSS, JavaScript, R

Big Data Technologies: PySpark, Amazon S3, DynamoDB, SparkSQL, Pig, Hadoop, Kafka

Tools: Bitbucket, Jupyter, Zeppelin, JIRA, Git, Visual Studio Code, MySQL, PostgreSQL, Tableau, Docker, Postman, AWS Lambda

PROFESSIONAL EXPERIENCE

BXB Digital (Brambles), Santa Clara, Data Scientist **August 2019 - Present**

- Working on mapping Wi-Fi mac addresses to their Geolocations to locate and track assets easily. For each mac address, the observations and locations (provided by various API services) are combined with the RSSI values to determine its coordinates
- Performed ETL operations on master location data to assign locations to the mac addresses using Geohashing
- Developed an algorithm in Spark to determine dwells in the asset journey from the Wi-Fi observations

King Street Capital Management, New York, Machine Learning Intern **February 2019 – May 2019**

- Researched gradient boosting algorithms like XGBoost, LightGBM and compared their performance. Worked on enhancing the performance of LightGBM algorithm in C++ by handling sparsity and missing values better
- Extracted and analyzed data to evaluate Universities based on their incoming student's qualifications for hiring interns

Kimmel Media Services, New York, Media Student Assistant **November 2017 – December 2018**

- Responsible for providing technical media support services to the NYU community including equipment repair and installation, event support, and equipment delivery and retrieval

BXB Digital (Brambles), Santa Clara, Data Science Intern **May 2018 – August 2018**

- Developed a perceptual hashing algorithm for image verification of pallets. Implemented block mean value hashing to compute rotation invariant hash to find matches, compute cycle time and damage rates. Built a Siamese network model using TensorFlow to perform one shot classification for similar pallet images
- Worked on anomaly detection using rolling median and standard deviation in asset tracking data involving location via GPS, WIFI or Cellular to detect errors in location pings because of fluctuating signal strengths with the cell towers

Tradeshift, San Francisco, Software Engineer Intern **January 2018**

- Worked on Pulse, a project that created a live map with invoice data that were being generated and sent from one company to another around the world at the exact moment
- Performed data extraction using Akka streams and Spark on AWS data. Parsed through UBL file streams to obtain the details of transactions including sender, receiver, amount and volume
- Designed module in Scala to find the location (latitude, longitude) from address obtained

Persistent Systems, India, Intern **June 2016 – July 2016**

- Designed interactive live-streaming dashboard using Apache Spark on the ShareInsights platform to provide weather forecast and analyzed various aspects such as temperature and humidity using Apache Kafka and Hadoop

PERSONAL AND ACADEMIC PROJECTS

Smart Class (Python/JavaScript/Amazon Web Services) **January 2019**

- Designing a web application to stream lecture videos while also enabling the user to search videos based on content and view the lecture from corresponding timestamp of the content
- The application is built using AWS services like S3 to host the content and web application, AWS CDN for content delivery, AWS Lex and Lambda to interact with the user, AWS Transcribe to convert the video speech to text, AWS Comprehend to perform NLP on the video transcript and AWS Elasticsearch to search the content in videos

Radiation & Nuclear Data Analysis (Python/R) **December 2017**

- Cleaned and processed Radiation data from NASA and Safecast APIs by leveraging tools like Spark and Hadoop.
- Analyzed the streaming data and developed insights which were then visualized using various packages in R to show co-relation between temperature, altitude and radiation level

Designing a Custom SVM Kernel for improving accuracy in Drug Classification (R/Python) **April 2017**

- Trained the SVM and tested various kernels (RBF kernel, Polynomial kernel, Gaussian kernel) for accuracy
- Presented a technical paper in NCCEE, IETE Cynosure proposing designing a custom kernel for SVM to improve the accuracy of drug classification. Won the Young Researcher's Award for the presentation at the conference