Jay Kakkad

Stony Brook, NY | +(1) 864 412 9724 | jkakkad1997@gmail.com | linkedin.com/in/jay-kakkad | github.com/jay-kakkad

EDUCATION

Master of Science, Computer Science, Stony Brook University, New York – GPA: **3.7/4.0** **Aug '19 - Dec' 20**

Courses: Big Data, Machine Learning, Statistics, Algorithms, Network Security, Database management

Teaching Assistant: Object Oriented Programming - Java (CSE 114)

Bachelor of Technology, Computer Engineering, University of Mumbai, Mumbai, India – GPA: **7.7/10** **July '15 - June '19**

Courses: Data Structures, Operating System, Networking, Algorithms, Software Engineering

Technical Skills

- **Languages & Database**: Python, Java, SQL, C, JavaScript, HTML5, CSS3, MySQL, Bigtable, DynamoDB, IBM DB2, Oracle SQL, AWS S3
- **Tools & Technologies:** AWS lambda, AWS EMR, AWS Athena, Apache Spark, Hadoop, TensorFlow, PyTorch, Node.js, React, Redux, Express, Spring, GCP, Azure,Databricks, Git, Docker, Linux, Windows.

EXPERIENCE

Amazon - People's Engine, Seattle **Software Developer** **Feb '21 – Present**

- Spearheading pilot project (WRAP) to create visual analytics platform for human resource management by leveraging statistical and machine learning models to forecast project timelines, and normalize work load on any given resource, while generating effective analytics report for appropriate stakeholders.
- Efficiently developed large-scale data processing pipeline using spark and AWS EMR to perform anomaly detection on People Soft's Integration data, reducing processing time from over 10 mins to under 120 seconds

Brookhaven National Laboratory, NY **Graduate Researcher** **Feb '20 – Dec'20**

- Developing few-shot learning based visual scene recognition model using graph neural network, in **PyTorch.**
- Reviewed and Implemented Large scale visual relationship Understanding by developing visual relation detection module using Graph Neural Networks and semantic module using node2vec embeddings, achieving **59.87%** accuracy in SOR recognition.

Mozilla Builders (Open Source), Remote **Summer Developer** **July '20 - Sep '20**

- Developed full stack social media platform to facilitate civilized engagement using **React, Node.js, Express,** and **MongoDB**
- Integrated microservice for hate speech detection based on pretrained **BERT** model to determine toxicity of text, achieved **90%** recall accuracy, in **Python Flask** and deployed using **Docker** containers on **Google Cloud Platform.**

Barclays Bank PLC, Pune, India **Business Analyst Intern** **June '18 - Aug '18**

- Instrumented development of health check monitoring dashboard, using **JavaScript, Selenium**, **Spring**, **MySQL**, **HTML** and **CSS,** for multiple applications & database environments with CRON job and log report generation capabilities for internal stakeholders.

Innefu Labs, Delhi, India **Machine Learning Intern** **Dec '17 - Jan '18**

- Designed web scraping tool for Twitter posts in **Python** and facilitated redesigning of bilingual multi-class hate speech classification model using N-grams and Support Vector Machine (SVM) algorithm, increasing recall accuracy by **12%.**
- Built object detection model using image vectorization and pretrained YOLO to identify consumer graded objects in satellite images, and improved classification accuracy by **10%,** using **TensorFlow** and **Python.**

PROJECTS

New York metro travel planner for Covid-19 - Ranked 3rd in MTA Back on Track Hackathon

- Analyzed passenger travel patterns and train occupancy between Jan & July '20, applied regression modelling and time series analysis using PySpark while achieving MAE score of **0.7** & deployed microservice on Azure to forecast train & platform occupancy.

Income growth Analytics - Big Data Analytics | Spark | Map reduce | Python

- Efficiently processed large scale data using Spark, formulated hypothesis testing for economic growth among demographics in US and employed regression models to predict economic growth for year 2020, achieving **MAE of less than 1.**

Reverse Website Fingerprinting - Ruby on Rails | MySQL | Python

- Analyzed security hygiene of WordPress Websites by identifying plugin signatures via DOM and scraped 10k plugins in 1.5k active websites over IPv4 space resulting in **25%** websites with over 5 documented critical plugin vulnerability.

Covid-19 Analytics - Python | Time-Series Analysis | Hypothesis testing

- Discovered correlation between Covid-19 impact and increase in President Trump's approval rating by performing hypothesis testing on COVID-19 US data. Predicted number of Covid-19 deaths through time-series analysis using EWMA and Auto-Regression

Stock Price Trend Prediction - Python | Regression Modelling | Sentiment Analysis

- Forecasted stock trend based on Support Vector Regression algorithm by analyzing historic stock returns and performing sentiment analysis on industry and company news. Achieved **75%** binary classification on 2-week live simulation.