BLESSY ANTONY

EDUCATION

Virginia Polytechnic Institute and State University (Virginia Tech)

Doctor of Philosophy (Ph.D.), Masters (MS) – Computer Science; GPA: **3.96/4.0** Research areas: Computational Biology, Bioinformatics, Graph Machine Learning

Fr.Conceicao Rodrigues College of Engineering (Fr.CRCE),

Bachelor of Engineering - Computer; CGPA: 9.36/10, Class Rank: 1/75

- Topped Mumbai University in Semester 4 with CGPA: 10/10
- Won the Tata Consultancy Services Best Student of Fr.CRCE Award across all departments for excellence in academics and co-curricular activities.
- Won the Dr.S.S.S. Award for securing 1st rank in Applied Mathematics across all departments through all 4 years of engineering.
- Awarded the Sir Ratan Tata Trust Scholarship and Sir Dorabji Tata Trust Scholarship for excellence in academics in 2013 and 2014 respectively.

EXPERIENCE

JP Morgan Chase & Co., Software Engineer - Associate Software Engineer - Analyst

- Developed a pipeline of containerized microservices that collate, parse, and enrich application logs to determine the availability of applications and reduce issue triaging time by 20%.
- Implemented a framework using JMeter, Elasticsearch-Logstash-Kibana, and NodeJS to automate performance testing and enhance report generation. The system saved ~\$100,000 annually by reducing manual effort by 70%.
- Awarded the Quarterly Global Excellence Award in December 2020 for revamping and migrating Post Trade Reporting services from legacy to new, regulated datacenters, and containerized platform (Kubernetes) thus reducing the infrastructure cost by 50%. Re-implemented and converted monolith applications to microservices to improve resiliency, stability, and efficiency.
- Volunteered twice in the year-long 'Force for Good' initiative to build customized data management and process automation systems for the NPOs 'Katalyst' (2017) and 'Muktangan' (2020) to decrease the process workflow time from 2 hours to under 5 minutes.
- Conceptualized and materialized a prototype of a Virtual Reality system to analyze bank transactional data and establish customer expenditure pattern, during the global JPMC Hackathon 2018 as a part of a 2-member team. The project was selected for the final round globally.
- Received the 'Excellence in Performance Award' for the month of April 2017 across all levels in the department.

PROJECTS & RESEARCH

Predictive Models of Long COVID

Prediction of post-acute sequelae of SARS-CoV-2 infection (PASC, or long COVID) based on information derived from the electronic health records (EHRs) of acute COVID-19 infection available in the <u>National COVID Cohort Collaborative</u> (N3C) — largest harmonized repository of EHRs of COVID-19 patients in the US. (Manuscript status: submitted.)

- Trained Logistic Regression and Random Forest models using features such as symptoms experienced by the COVID-19 patients, drugs ordered or administered to them, measures of COVID-19 treatment, comorbidities, and demographic information. Identified long COVID patients using U09.9 ICD10-CM code in EHRs.
- Leveraged Boruta method for feature selection and *k*-fold cross-validation based hyperparameter search for optimal model configuration yielding mean AUROC of 0.70 in disease prediction.
- Computed the local interpretation of predictions using SHAP method to identify risk-factors in COVID-19 patients.
- Performed novel cross-site analysis to evaluate the impact of different data sources on prediction models.
- Implemented reusable and easily adaptable classification pipeline in N3C Enclave for disease prognosis using EHRs.

A time evolving online social network generation algorithm

Developed a community-based, time-evolving online social network generation algorithm to generate synthetic online

Blacksburg, Virginia, USA linkedin.com/in/blessyantony9/

Blacksburg, USA Jan 2021 – present Expected graduation: Dec 2025

> Mumbai, India June 2012 – May 2016

Mumbai, India Jan 2019 – Dec 2020 July 2016 – Jan 2019

Sep 2021 – Present

Aug 2021 – Dec 2022

social networks for downstream social network analysis tasks. Published in Scientific Reports (February 2023).

- Generated networks paralleling real-world social networks in terms of network properties such as clustering coefficient. • degree distribution, and community structure.
- Designed the algorithm to emulate evolution of nodes, edges, and communities in the network at different time steps. •
- Devised tunable stochastic parameters based on compartmental modeling from epidemiology (for node evolution) and . communities (for edge evolution) to allow generation of varied, realistic synthetic networks.
- Generated sequences of directed and undirected synthetic networks with 100,000 nodes over 100 time steps and • observed, over time, densification law in number of nodes, long-tailed degree distribution with friendship paradox, and community structure with a large number of small communities.

Evaluation of Code Representation Learning Models

Designed a standardized pipeline of downstream tasks for evaluation of code representational learning models - CodeBERT (uses contextual information only) and GraphCodeBERT (uses both contextual and structural information).

- Used BigCloneBench and CodeSearchNet datasets to benchmark the performance of models in code summarization and clone detection tasks respectively.
- Fine-tuned pre-trained models from Hugging Face Transformers, evaluated them across ten randomly sampled test • datasets, and computed and compared performance metrics.
- Observed the importance of structural information in learning code embeddings with GraphCodeBERT yielding higher • BLEU score in code summarization, and higher AUPRC, F1, Precision, and Recall in clone detection.

Social Media Sentiment Analyzer for COVID-19

Built a model that stratifies Tweets pertaining COVID-19 into 3 sentiment categories to identify posts propagating antivaccine misinformation through social media.

- Implemented Logistic Regression, Support Vector Machine (SVM), Long Short-Term Memory Recurrent Neural • Network (LSTM-RNN) multi-class classification models using TweetsCOV19 dataset.
- Performed parameter tuning and 5-fold cross validation for all models, and evaluated them using Accuracy, Precision, • Recall and F1 scores. Highest accuracy of 98.2% achieved by LSTM-RNN model.
- Added functionality to analyze and classify real-time data by streaming Tweets using Twitter API v2. •

SARS-CoV-2 Interactors Identification using Network-Based Label Propagation

Implemented Random Walk with Restarts (RWR) algorithm to discover human proteins interacting with the SARS-CoV-2 virus proteins directly or indirectly.

- Propagated the labels from known human protein interactors in a protein-protein interaction network STRING and identified unknown virus interactors.
- Compared with external experimentally derived results using Fisher's Exact Test and Jaccard coefficient. •

LEADERSHIP

June 2019, June 2018 Subject Matter Expert, Code For Good, JP Morgan Chase & Co. Mentored high school students and helped them build solutions for social good during a 24-hour hackathon.

Editor-In-Chief, Newsletter Committee-India & Philippines, JP Morgan Chase & Co. *Apr* 2017 – *Mar* 2018

Led a cross-country 20-member team to collectively compose and roll out the firm's internal quarterly e-newsletter.

Second Year Representative, Institute of Electrical and Electronics Engineers – Women June 2013 – May 2014 in Engineering (IEEE - WIE), Fr.CRCE Student Chapter

- Planned and organized Technomania 2014 the national level Project and TPP competition with 150+ participants.
- Represented the college in the IEEE Student Congress-2013 conducted by IEEE-Bombay Section Student Network. •

| SKILLS | |
|----------------|---|
| Programming: | Python, Java, C, Javascript, GNU Octave, XML, SQL, HTML-CSS |
| Platforms: | Linux - RedHat and Ubuntu, Windows XP, Windows7 & 8, Mac OS X |
| Frameworks: | Numpy, nltk, NetworkX, Pandas, PyTorch, SciKit, Spring, SpringBoot, Node JS, Express JS |
| Concepts: | Cloud Computing (Cloud Foundry), Containerization (Docker, Kubernetes, Apache Mesos), |
| | Message Queue (Apache Kafka) |
| Miscellaneous: | Git, LaTeX, Microsoft Office Suite, MySQL, Jenkins, Joget, Firebase |

Jan 2021 – May 2021

Jan 2021 – May 2021

Jan 2022 – May 2022