Md Arafat Hossain Khan

+1 (775) - 338 - 9105
arafat.math@gmail.com
linkedin.com/in/mdarafathossainkhan/

Summary

Data scientist and **pure mathematician** with professional experience in **statistics** and **artificial intelligence**. 9 years of experience in team leading and managing projects for software development, machine learning and data science solutions for business problems. 11 years of research experience with grant/proposal writing and managing. Hands-on experience with big data analysis and implementing in cloud platforms (e.g. AWS). Submitted several patents, published papers in renowned mathematical journal and multiple others are in the pipeline.

Skills

Machine Learning / Data Science /

Hands on experience with deep Learning techniques including CNN, RNN, LSTM and embedding models (e.g., BERT), Deep Generative Models, Variational Autoencoder (VAE), Generative Adversarial Networks (GAN), Progressive Growing of GANs, CycleGAN domain transformation, Deep Reinforcement Learning, Graph Neural Networks etc. Deploying machine learning and natural language processing models on AWS. Professional experience with ML-Ops. Possess strong foundation on advanced concepts of Neurosymbolic AI, Topological data analysis, quantum computing, ontological concepts, knowledge graph embedding etc.

Optimization, Statistical Learning and Mathematics

Linear, integer and mixed integer programming, convex/non-convex optimization, parametric/non-parametric models, statistical inference, sampling, frequentist and Bayesian statistics, hypothesis testing. KNN, Multiple Regression, LDA/QDA, Logistic Regression, Ridge/Lasso, Bootstrap/Cross Validation, Clustering, Decision Tree, Bagging/Boosting/Random Forest, SVC/SVM with different-kernels, Spline, PCA, PCR/PLS/Best Subset Selection etc. Expert in solving abstract and applied mathematical problems (e.g. advanced linear algebra, group theory, graph theory, topology, combinatorial optimization and so on). Hands on experience with Gurobi optimizer, CPLEX Optimization Studio.

Languages/ DB/ Technologies

Python, Java, MATLAB, R, Scala, Scikit-learn/TensorFlow/Keras, PyTorch, ASP.NET Core with C#, SQL Server, NoSQL, MongoDB, Cassandra, unstructured/semi-structured/structured data handling, Data crawling, cleansing, warehousing; Android and desktop applications with Java/C#. Familiar with Agile and DevOps.

Advanced Algorithms

Network Flow problems, Graph Algorithms, Dynamic Programming, Algorithms of Topological Graphs, Homotopy and Homology based Algorithms on graphs, topological graph theory, 3D spatial data analysis with persistent homology and higher dimensional hypersurface embedding, surface normal estimation, Deep Learning based models for 3D Reconstruction, vehicle tracking system using OpenCV with C++, filtering methods like Kalman filter etc.

Projects and Grants

Written and led projects and grants including a World Bank funded (USD 236,714, from 2011 to 2013) project, experienced in dealing with small/large scale industrial and government funded software/web applications for different clients including Bangladesh Army.

Education

Doctor of Philosophy — Pure Mathematics
The University of Texas at Dallas — *2020*
Dissertation: Left Orderability of Dehn Surgery on Odd Classical Pretzel Knots.

Master of Science — Mathematics (Concentration: Data Science)
The University of Texas at Dallas — *2019*

Graduate Degree Certificate — Data Science
The University of Texas at Dallas — *2018*

Master of Science — Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology — *2013*

Bachelor of Science — Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology — *2010*

Professional Data Science Projects (Selected)

- **Ontology for Data Echosystem and Customer Intents:** Developed an ontological infrastructure for data scientists and other data consumers to discover the appropriate data of their needs which essentially reduces many person hour and reducing costs. This data centric approach has numerous downstream benefits. Created an ontological stucture for the intent labelers so that the label noise can be reduced. This essentially has direct effect on natual language models and recommendation engine.

- **Banking Chatbot Design:** Designed a chatbot for satisfying customer's need through digital interaction with almost no need of typing. Understanding customer's need from their recent activities, transaction related data and static account features before even customer click on the chatbot using deep learning techniques. Once the customer clicks on the chatbot the guided conversation will match with customer's expectation and thereby serving them with their needs. Moreover, designing a robust and scalable solution of this digital customer agent in a generalized model that not only predicts customer needs but also achieves their goal with as less typing as possible. Currently this model is in production behind the chatbot called ENO for 50M+ Capital One customers defeating the performance all previous models. This is a state-of-art model will multiple components of machine learning, natural language processing, heuristic algorithms etc.

- **Targeted Marketing Strategy:** Analyzing the customers (ethnicity, ability, need, social circumstances etc) to offer promotions/deals of new/existing products. My role is designing the entire solution and lead the technical team. Utilized Deep Learning, Language Models (e.g BERT), Bayesian Statistics using TensorFlow/Keras, python on AWS over multiple different unstructured/semi-structured datasets. First phase of this project successfully went through the production and is being used by the real estate branch of my current company.

- **Recommendation system:** Suggesting knowledge seekers (institutions) the best match of the Subject matter Expert based on collaborative filtering, matrix factorization and deep learning Techniques. Scaled with Apache Spark, Amazon DSSTNE, AWS SageMaker for massive datasets. Used TensorFlow/Keras, Python to build the model. Main application is developed using ASP.NET core MVC with SQL Server.

- **Rigged election detection:** Analyzed 2014, 2018 Bangladeshi general elections with Citizens for Good Governance which was published in national newspapers and the government responded to that. Used Python, R, TensorFlow, Excel VBA. Built novel algorithms and designed the architecture.

Work Experience

Principal Data Scientist — 5 April 2021 – Present
Capital One Financial Corporation — *8020 Towers Crescent Dr, Vienna, VA-22182*

Working in Recommendation Engine and Natural Language Processing (NLP) teams called servicing intelligence. The servicing intelligence DS team researches and develops machine/deep learning (ML/DL), natural language processing (NLP) models leveraging state-of-the-art architectures, packages and techniques. The team also closely partners with engineering team to build end-to-end solutions, and deploy in scalable and resilient production system that serve 50+ million customers.

Responsibilities:

- End to end coding, testing, validating and monitoring fully functional machine learning applications.

- Creating/improving deep learning based architecture for natural language processing related tasks, e.g. chatbot for customer service.
- Improving and extending code and models using more efficient algorithms and more robust/generalized code.
- Understanding production needs and respond to changing production requirement with upgraded architecture/code.
- Working closely with production and tech teams to fulfill customer needs.
- Technical documentation and presentation, whitepaper writing.
- Mathematical and statistical analysis of the machine learning based products.
- Advanced research on deep learning and general machine learning to take the product one step ahead of the market competition.

Projects:

- **Conversational Platform:** Working on customer facing applications like ENO Chatbot which serves 50+ millions of customers to solve their banking problems. I was able to start from writing the entire codebase to productionalize the chatbot including monitoring performance on real time data in two months. Moreover, the new version that I created showed much better performance than the previous version of the chatbot. Other than model centric approach a data centric approach was taken and ontological structure was introduced to incorporate much cleaner data.
- **Smarter NLP:** Improving the existing NLP infrastructure and providing more robust solutions. Multiple components of NLP and rule-based algorithms identify customer needs from their utterances.

Technology: PyTorch, TensorFlow/Keras, SQL, Deep learning, Natural Language Processing.

Software Application Developer October 2020 – 4 April 2021

Tasacom Technologies Inc. *Dallas, TX*

Responsibilities:

- Handling large databases (unstructured/semi-structured) to infer knowledge, mining data if not directly available (crawling if legally and technically possible), cleaning and pre-processing the data, reporting and presentation using tools like Tableau, GIS, designing long-term solutions for real-world business problems.
- Based on business needs preparing solution using AWS AI Services, Platforms and Frameworks.
- Building and maintaining applications using ASP.NET Core MVC C#, SQL Server.
- Work with Big data app development strategies, NoSQL Data Modeling, build models with deep learning frameworks like Keras/Pytorch/Tensorflow, cloud deployments in AWS.
- Theoretical/applied concepts of statistical learning techniques to analyze data and solving complicated problems and build mathematical models if needed. Interpret solutions in terms of data science and production environment.

Projects' Detail:

- Targeted market campaign for real estate business.
- Automation system of collaboration between different subject matter experts and institutions seeking for expertise across the world based on the demand.
- Ethnicity classification for ethnic product campaign.

Technology: NoSQL, MongoDB, Apache Cassandra, Python, R, Java, Deep learning, Natural Language Processing, TensorFlow/Keras, ASP.NET MVC Core C#, SQL Server.

Graduate Teaching Assistant Sep. 2015 – Aug. 2020

Department of Mathematical Sciences *The University of Texas at Dallas*

Responsibilities: Teaching mathematics courses to undergraduate students of various departments.

Courses Taught: Linear Algebra, Topology, Differential Calculus, Calculus of Several Variables, Abstract Algebra, Integral Calculus, Theoretical Concepts of Calculus, Differential Geometry, Mathematical Analysis.

Assistant Professor (on leave for PhD) Jan. 2014 – Aug. 2020

Department of Computer Science and Engineering *Jagannath University, Bangladesh*

Projects and Grants: I was one of the founders of the *Reengineering of the Information Technology Research and Resource Center (ITRRC)* at Jagannath University, which is a platform for innovative computer science research in Bangladesh. This project was funded jointly by the *World Bank* and *Bangladesh Government*.

Administrative: I was the member of planning, procurement and recruitment committees. I also worked to make courses curriculum and syllabus in conjunction with different universities.

Used Technologies: SQL Server, MySQL, ASP.NET, C#, Verilog, MATLAB, EMU8086 Assembly.

Graduate Research Assistant Sep. 2013 – July 2015

Department of Civil and Environmental Engineering *The University of Nevada, Reno*

Research: Worked with intelligent transportation system. Did the project of On-ramp queue estimation and optimization, developed guidelines for transportation engineers with California Department of Transportation. Later transferred to the University of Texas at Dallas changing major to mathematics due to the interest of solving more mathematically challenging problems.

Used Technologies: Excel VBA, Java, Mono for Android, Computer Vision, OpenCV with C++.

Lecturer (was partially on leave for higher studies) May 2011 – Jan. 2014

Department of Computer Science and Engineering *Jagannath University, Bandladesh*

Responsibilities other than teaching: Worked in exam and admission committees, written grants and project proposals. Designed and built electrical circuit and digital system design laboratories.

Used Technologies: MATLAB, Assembly, C, C++, PSpice, hSpice.

Lecturer Dec. 2009 – May 2011

Department of Electronics and Telecommunication Engineering *The People's University of Bangladesh*

Responsibilities other than teaching: Built circuit and machine laboratories. Supervising undergraduate thesis. Managed and improved the official website of People's University using ASP.NET C#.

Manager March 2008 – Dec. 2009

Project Management Division *Vintage IT Limited, Dhaka, BD*

Responsibilities:

- Writing project proposals for various large scale professional projects. Requirement analysis from both client side and translating it to technical teams. Creating the timeline and work-flow diagram. Maintaining the client and technical team coordination. Distributing works among team members.
- Financial analysis of projects, feasibility studies and SWOT analysis, milestone preparation. Ensuring best practices and future extendibility of projects. Managing subprojects with other third-party companies. Conducting training programs for newly appointed software developers.

Leading Software Engineer June 2004 – March 2008

Project Management Division *Vintage IT Limited, Dhaka, BD*

Responsibilities:

- Designing and maintaining ASP.NET based web and desktop applications, SQL Server and MySQL Database design and implementation, creating documentations of Sequence Diagram, Entity Relationship Diagram, Relational Schema, SQL Commands, PL-SQL Commands, Triggers, Stored Procedures, LINQ query for SQL and XML, Normalization of Databases - BCNF/4NF/5NF/DKNF/EKNF. ASP.NET Web User Control, ASPX webform, Sitemap designing, working with caching and cookies, Telerik UI for ASP.NET.
- Worked with GIS Data Display on web maps, ArcGIS, QGIS spatial drawing based Web applications, Ajax, JavaScript, JQuery, ActionScript, HTML and CSS, ASP.NET Security, Cross-site scripting, JavaScript disabling, SQL injection, Cross-site request forgery (CSRF/XSRF), Storing secrets, Encrypting Web.cong settings, Password hashing, authenticating in the app, Securing cookies and sessions.

Environment: ASP.NET Platform 2.0/3.0/3.5, ASP.NET Security, ASP.NET MVC C#, ASP.NET Web APIs, ASP.NET Web User Control, ASPX webform, ASCX User Controls, Telerik UI, Sitemap, Cache, Cookie, Entity Framework, JavaScript, JQuery, ActionScript, HTML and CSS, SQL Server, MySQL, PL/SQL, XML, LINQ, IIS.

Patents

Several patents (proprietary to Capital One) have been submitted and multiple others are in the pipeline.

Selected Publications

[1] Arafat Khan and Anh T. Tran. "Classical pretzel knots and left orderability". In: *International Journal of Mathematics* (2021).

[2] M. Helal Uddin, M. Arafat H. Khan, and Charles J. Coronella. "3-D face-masking detection and tracking algorithm for bubble dynamics: Method and validation for gas–solid fluidized beds". In: *Powder Technology* 313 (2017), pp. 88–98. ISSN: 0032-5910.

[3] M Helal Uddin et al. "Effects of baffle on bubble break-up: A numerical simulation of a pilot scale bubbling fluidized bed with Geldart B particles". In: Nov. 2015.

[4] M. A. H. Khan et al. “Multiple polynomial regression for modeling a MOSFET in saturation to validate the Early voltage”. In: *2011 IEEE Symposium on Industrial Electronics and Applications*. 2011, pp. 261–266.