Md Arafat Hossain Khan

+1 (775) - 338 - 9105 ✉ arafat@simplisolve.us linkedin.com/in/mdarafathossainkhan/

Summary

- **Data scientist** and **Pure Mathematician** with professional experience in **Statistics** and **Artificial Intelligence**.
- **Hands on** experience of building machine learning solutions (time series analysis, targetted market campaign, speech signal processing, deepfake detection, conversational ai etc.) with big data and deploying in cloud platforms (e.g. AWS).
- Large Language Model (**LLM**) based banking **chatbot** design with **Recommendation Engine** and natural language processing for Capital One currently serving more than **100 million+** customers.
- Developed **Deep Forecasting** model and deployed **Trade bot** in AWS for **Cryptocurrency** market generating high ROI. Extensive experience working with **ERCOT energy market** data. Also dealt with Oil and other energy market price prediction and got very promising results in back-testing report. Building bidding strategies.
- Extensive experience working with **Large Speech Models** and **Generative AI** for **DeepFake Detection** for voice authentication system. This includes generation of sysnthetic customer calls, schemes to detect the synthetic calls and deploying the system into production. This is serving multiple credit unions at this moment increasing customer satisfaction and company cost and risk minimization.
- **10 years** of experience in **team leading** and **managing projects** for software development, machine learning and data science solutions for various business problems. Worked in collaborative team environments. Presented **ideas to business leaders**. **Mentored** associates to succeed. Developed **quarterly plan** to accomplish business goals and managed multiple projects.
- **11 years** of research experience with large scale **grant/proposal writing and managing**. Submitted several **patents**, published **papers** in renowned mathematical and computer science **journals** and multiple others are in the pipeline.

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*

Skills

Machine Learning / Data Science

Hands on experience with deep Learning techniques including CNN, RNN, LSTM, GNN, GAN, hierarchical attention network (HAN), neural factorization machine (NFM), deep reinforcement learning etc. 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, Classical Machine 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.

Professional Data Science Projects (Selected)

- Data Governance
  - **Generic Data Monitoring Scheme:** Developed a generic monitoring scheme and put in into production that can monitor the customer data and performance drift of all the customer facing and experimental models. This helps the data scientist to understand not only the model performance shift but also to track exact cause of the this. I used PySpark, SQL, Python, Snowflake and Tableau for building the application and productionalized it in a batch environment on AWS.
  - **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 hours and company costs. I use Python and SQL for this.
- Machine Learning Model Building
  - **Time Series Modeling for Energy Market Bidding:** Built a state of art machine learning model to predict day ahead and hour ahead electricity price (per MWh) prediction for ercot. I used TensorFlow to build the model.
  - **Banking Chatbot Design:** Understanding real time customer's need from their current and past activities, transaction related data and static account features for chatbot using deep learning techniques. Currently this model is in production behind the chatbot called ENO for 100M+ Capital One customers. This state-of-art model multiple modal machine learning, NLP, heuristic algorithms etc. Used Python, PySpark, SQL, Pytorch etc to build this.
  - **Targeted Marketing Strategy:** Analyzing the customers and offer promotions/deals of new/existing products. Utilized Deep Learning, Language Models (e.g BERT), Bayesian Statistics using TensorFlow/Keras, python on AWS over multiple different unstructured/semi-structured datasets.

- **Recommendation system:** Suggesting knowledge seekers (institutions) the best match of the Subject matter Expert based on collaborative filtering, matrix factorization and deep learning techniques. I 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 national 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

Founder & CEO 29 March 2023 – July 15, 2024

SimpliSolve Inc *13151 Emily Rd Ste 100-C Dallas, TX 75240-8966*

SimpliSolve is a team of academic and industry experts from diverse fileds to develop innovative AI algorithms and models that transform complex data into actionable insights.

Responsibilities:

- Business strategy, budget allocation, client meeting etc.
- Research leading, solution design and deployment.

Technology: Advanced AI, Cloud platforms.

Chief AI Scientist 20 Feb 2023 – Present

Black Creek Digital *80 State Street, 425 Market Street Suite 3300, Albany, NY 12207*

Working with predictive algorithms for different markets. Building trading strategies and doing end to end deployment.

Responsibilities:

- Business strategy design and implementing trading bot.
- Reaserch on advanced predictive algorithms combining multi-disciplinary fields.
- End to end deployment of the real time predictive algorithms.
- Leading the AI research and figuring out the best possible strategies for market.

Technology: Deep learning, Mathematics, Python, AWS.

Lead Research Scientist 14 April 2023 – Present

Illuma Labs Inc. *7700 Windrose Ave. Ste G300, Plano, TX 75024*

Working with voice biometric authetication system to prevent frauds. Recent advancements of generative AI leads towards synthetically generated voices which has the potential to break the authentication system. Worked in building security models to defreat AI attack as well.

Responsibilities:

- Reaserch on building robust powerful authentication and fraud prevention systems.
- End to end deployment of the system into client's server and local development environments.
- Continuous improvement and understanding the business needs.
- Leading the team of AI research.

Technology: SQL, Deep learning, Digital Signal Processing, Generative AI, MATLAB, Python, AWS.

Principal Data Scientist 5 April 2021 – 14 April 2023

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 100+ 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 100+ 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. Also to reduce the noise in data label I came up with an ontology based data labeling scheme.

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. Worked with guided conversations with recommendations.
- Automation system of collaboration between different subject matter experts and institutions seeking for expertise across the world based on the demand. Recommendation for on knowledge seekers to connect with SMEs.
- Ethnicity classification for ethnic product campaign. Recommending products based on ethnicity.

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

1. (US20230419046A1) Systems and methods for generating real-time dynamic conversational responses during conversational interactions using machine learning models.
2. (US11811708B2) Systems and methods for generating dynamic conversational responses using cluster-level collaborative filtering matrices.
3. (US20230108855A1) Systems and methods for generating training data for sequential conversational responses.
4. (WO2023056195A1) Systems and methods for generating dynamic conversational responses based on predicted user intents using artificial intelligence models.

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.

Sweet Achievement

In Bangladesh, 10th and 12th grade exams are conducted nationwide and around 3 million students perticipate in these extremely competitive exams. I scored highest in both my 10th and 12th grade exams ranking #1 in the country which made me a teenage celebrity. I never heard of anyone ranking #1 in both the exams.