

Md Arafat Hossain Khan

Contact: 7780 McCallum Blvd, Apt-26221
Dallas, Texas-75252, ☎ +1-775-338-9105

in <https://www.linkedin.com/in/mdarafathossainkhan/>
🔗 <https://github.com/MdArafatHKhan>
🌐 <http://www.utdallas.edu/~arafat/>
✉ arafat@utdallas.edu

Summary

Data Scientist with PhD in pure mathematics having proficiency in solving abstract mathematical problems of pure mathematics as well as applied knowledge of machine learning using different frameworks.

Research Interest

Knot theory, computational topology, topological data analysis, quantitative finance and machine learning.

List of Experiences

(5 Years) ASP.NET C#, SQL Server database, JavaScript, JQuery, CSS, HTML.

Worked in Vintage IT Limited as a professional software programmer.

(10 Years) MATLAB

Did research works, published papers. Used Matlab and C programming in same application extensively.

(5 Years) Mathematica, R

Used extensively for scientific calculation, machine learning projects.

(4 Years) Data Analysis, Statistical Learning and Machine Learning

Used Python and R for numerous projects, worked with different libraries and also coding from scratch.

(2 Years) Excel VBA

Used VBA to model on ramp queue with California Department of transportation.

(1 Year) Scala, MongoDB and AWS

Worked with Big data using Scala and AWS. Did projects on document based database.

(1 Year) OpenCV using C++

Built application for video editing and did computer vision related projects.

(1 Year) Android Application using Java

Built accident tracker system for university of Nevada, Reno.

(1 Year) Grant writing experience with successful accomplishment of the projects.

As a faculty in Jagannath University I did several funded projects and written grants.

Education

University of Texas at Dallas

2015 – 2020

PhD, Pure Mathematics

Research Topic: Conjectured equivalence of topological, geometric and algebraic characterization of 3-manifolds obtained by Dehn surgery on links and knots..

University of Texas at Dallas

May 2020

MS, Data Science

Skill Obtained: Machine learning, statistical learning, big data, bio-informatics, mathematics for data science, multiple programming languages and frameworks.

University of Texas at Dallas

May 2020

Graduate Certificates, Data Science

I have done numerous projects on Big Data and Machine Learning while doing this certification.

University of Nevada, Reno

2013-2015

PhD (incomplete - transferred to the University of Texas at Dallas), Civil and Environmental Engineering
Started PhD in Civil and Environmental Engineering and worked with On-ramp queue estimation and optimization and developed guidelines for transportation engineers. Later transferred to the University of Texas at Dallas changing major to mathematics due to the interest of solving harder critical problems.

Bangladesh University of Engineering and Technology

2009-2013

MS, Electrical and Electronic Engineering

Research Topic: Simultaneous adjustment of chirality and width produces interesting and useful bandstructures for circuit application. A rigorous quantum mechanical bottom-up analysis was done to model all-Graphene Nanoribbon circuits and the model was simulated by MATLAB.

Bangladesh University of Engineering and Technology

2004-2009

BS, Electrical and Electronic Engineering

Thesis Description: Multiple polynomial regression method was used to predict MOSFET characteristics in saturation and existing Early voltage approximation was found invalid. For simulation Cadence Virtuoso and HSPICE along with MATLAB were used.

Skills

Abstract Mathematics Knot Theory, Differential Geometry, Hyperbolic Geometry, Algebraic Geometry, Topology

Big Data Apache Spark, Hadoop, AWS

Machine Learning Spark MLlib in Scala, scikit-learn in Python

Statistical Learning Parametric and non-parametric, linear and non-linear methods, hypothesis testing

Programming Languages Assembly, C, C++, C#, Java, Visual Basic, Python, Scala

Web Developing ASP.NET, HTML, CSS, JavaScript, JQuery, XML

Mobile Application Android application using eclipse and Mono for Android

Database NoSQL, MongoDB, SQL Server, MySQL, SQLite

Technical Programming Verilog, VHDL, MATLAB, R, LABVIEW

Simulation Software Cadence Virtuoso, HSpice, PSpice, Quartus II, AVR Studio, Proteus, VISSIM, Synchro, Vistro

Skill Related Projects (Selected)

Abstract Mathematics Projects

- > Bi-orderability of fundamental group and generalized torsion element.
- > Left orderability of Dehn Surgery on odd classical Pretzel knots using Representation Theory.

Statistical and Machine Learning Projects

- > Analyzing how PSA level (Serum prostate-specific antigen level) is related to prostate cancer in adult men using regression based model.
- > Analyzing how GRE, GMAT and undergraduate GPA influences the admission decision in USA using LDA, QDA, KNN, support vector classifier and support vector machine with different kernels.
- > Analyzing bankrupt and nonbankrupt firms based on their cash flow, total debt, net income, total assets, liabilities, net sales etc using Logistic regression, LDA and QDA.
- > Analyzing the purchase of orange juice of different brands based on their features using LDA, QDA, KNN and applying k-fold cross-validation and bootstrap.
- > Analyzing mpg of different automobiles based of various features using multiple linear regression model, best-subset selection, forward stepwise selection, backward stepwise selection.

- Analyzing whether or not a given individual will purchase a caravan insurance policy based on many predictors using penalized maximum likelihood method with ridge and lasso penalty, decision tree, bagging, boosting, random forest, KNN, logistic regression, support vector classifier, support vector machine with different kernels.
- Survey as part of a study to assess options for enhancing food security through the sustainable use of natural resources in the Sikasso region of Mali (West Africa) applying dimensionality reduction and PCA
- Survey as part of a study to assess options for enhancing food security through the sustainable use of natural resources in the Sikasso region of Mali (West Africa) applying dimensionality reduction, best-subset selection, ridge regression, lasso, PCA, PCR and PLS.
- Clustering choice of breakfast cereals into groups using hierarchically cluster, complete linkage and different distance measures and also K-means clustering.
- Predicting salary of hitters using decision tree.
- Light detection and ranging (LIDAR) experiment on two variables - range (indicating distance travelled before the light is reflected back to its source) and logratio (indicating logarithm of the ratio of received light from two laser sources) using cubic regression spline, natural cubic regression spline and smoothing spline.
- Relationship between probability of gain versus expiration of SPY contracts by ensemble methods using SVM with Random forest (prediction from 8 years of news data) and LSTM (prediction from 26 years of stock price data).

Big Data Projects

- Analysis of the words in in large Wikipedia data using Spark Scala on AWS and Databricks.
- Analysis of routes of airlines based on large airport data using GraphX Spark Scala on AWS.
- Breast Cancer Prediction with multiple classification methods and hyperparameter tuning using Spark Scala on AWS.

Professional Web, Desktop and Mobile Applications (Few of many)

- [Education Management System.](#)
- [Human Resource Management System.](#)
- [Digital Library Management System.](#)
- [Professional Knowledge Development Assistant Software.](#)
- [Accident Tracker System application for Android.](#)

Personal Projects

- My target was to forecast SPY stock price movement for its high liquidity on contracts for option trading. I tried to predict the lower bound of the time span for choosing the best expiration date for a bullish call. A stacking based ensemble model was created using support vector machine achieved an f1-score of 57% on the testing data. I used LSTM (variation of Recurrent Neural Networks) for predicting stock price from previous price data, and Random Forest for news sentiment analysis and SVM for ensembling.
- Analysis of voting for 11th parliamentary election of Bangladesh and city corporation election 2015 of Bangladesh collaborating with Citizens for Good Governance (SHUJAN). The purpose of this project is to determine from the data provided by election commission of Bangladesh to determine if the election was manipulated and the data was falsified.
- Impact of the advent of artificial intelligence on US job market: Job vulnerability with respect to the advent of artificial intelligence and economic cycle was analyzed. Spark Scala and Python were used.
- Building portfolio management system for US stock market.

Funded Research Projects

Caltrans-Queue Storage and Acceleration Lane Length Design at Metered On-Ramps in California

Graduate Research Assistant, University of Nevada, Reno

- › Role: Research Assistant.
- › Duration: March 2013 to February 2015
- › Project Goal: Developing guidelines for practicing transportation engineers.
- › Location: California Department of Transportation
- › Total Cost: USD 249,310
- › Funding Agency: Department of Transportation (Caltrans)
- › Description: On-Ramp queue storage modeling is very important for new ramp design. From maximum and 95th percentile queue storage we can get the idea whether the queue will spillback. We can also find out the optimum stop line position if we know the maximum queue length. If the maximum queue length is very close to the ramp capacity there are chances for queue spillback.

Reengineering of the Information Technology Research and Resource Center (ITRRC)

Associate Researcher, Jagannath University

- › Role: Project Leader.
- › Duration: May 2010 to Dec 2013
- › Project Goal: Developing an advanced IT research infrastructure for present and future studies.
- › Location: Jagannath University, Dhaka, Bangladesh
- › Total Cost: USD 236,714
- › Funding Agency: World Bank and Bangladesh Government
- › Description: The purpose of building this center is to conduct various activities (research, seminars, workshops, remote and on-class lectures, presentations etc.) through a research network called 'ITRRC Research Fellows'.

Work Experience

Graduate Teaching Assistant

Department of Mathematical Sciences

Sept. 2015 to Present
University of Texas at Dallas

Graduate Research Assistant

Department of Civil and Environmental Engineering

Sept. 2013 to Sept. 2015
University of Nevada-Reno

Assistant Professor

Department of Computer Science and Engineering

Jan 4, 2014 to Present
Jagannath University (On-Leave)

Lecturer

Department of Computer Science and Engineering

May 2, 2011 to Jan. 4, 2014
Jagannath University (On-Leave)

Lecturer

Department of Electronics and Telecommunication Engineering

Dec. 2009 to May 2011
The People's University of Bangladesh

Leading Software Engineer (Part-Time)

Project Management Division

June, 2004 to Dec. 2009
Vintage IT Limited, Dhaka

Publications

Effects of Baffle in Bubble Break-up: A Numerical Simulation of Pilot Scale Bubbling Fluidized Bed with Geldart B Particles

- › M Helal Uddin, M Arafat H. Khan, Charles J. Coronella, Marisa Zuzga, American Institute of Chemical Engineers Annual Meeting November 11, 2015.

Active Contour Tracking of Individual Bubbles in CFD Simulation of Fluidized Beds

- › M Helal Uddin, M Arafat H. Khan, Charles J. Coronella, American Institute of Chemical Engineers Annual Meeting November 10, 2015.

A Computational Drug Designing From Active Product of Herbal Plant *Ochna Squarrosa* to Relieve Menstrual Complexities

- Aubhishek Zaman, Md. Arafat Hossain Khan, International Journal of Biometric and Bioinformatics (IJBB), 2012.

Multiple Polynomial Regression for Modeling a MOSFET in Saturation to Validate the Early Voltage

- Md. Arafat Hossain Khan, A. S. M. Zahidur Rahman, Tanvir Muntasir, Uzzal Kumar Acharjee, Md. Abu Layek, IEEE Symposium on Industrial Electronics and Applications, 2011.

Detection of Some Major Heart Diseases Using Fractal Analysis

- Nahina Islam, Nafiz Imtiaz Bin Hamid, Adnan Mahmud, Sk. M. Rahman, International Journal of Biometric and Bioinformatics (IJBB), 2010.

Poster Presentations

Shortcut Game on Non Semitransitive Graphs

- Prosper Akrobotu, Md Arafat H Khan, Mahadi Hasan, 3rd Annual Innovation and Technology Summit 2018.

Attended Certification Programs

National Training Course (NTC) on LabVIEW

- Atomic Energy Centre, Dhaka (AECD).

Industrial Training on Power System

- Power Grid Company of Bangladesh Limited (PGCB).

Graduate Courses Taken (Selected)

University of Texas at Dallas

- Probability and Statistics for Data Science and Bioinformatics, Statistical and Machine Learning, Algorithm Analysis and Data Structures, Machine Learning, Mathematical Foundations for Data Science, Introduction to Big Data Management and Analytics for non-CS Majors, Real Analysis, Functional Analysis, Complex Analysis, Ordinary Differential Equations, Topology, Differential Geometry, Abstract Algebra, Combinatorics and Graph Theory, Numerical Analysis of Partial Differential Equations, Principles and Techniques of Applied Math, Topological Degree Theory, Optimization, Theory of Relativity, Classical Mechanics.

University of Nevada, Reno

- Mathematical Physics, Pattern Recognition, Numerical Method II, Transportation Systems, Math in Transportation, Abstract Real Analysis.

Bangladesh University of Engineering & Technology - Dhaka, Bangladesh

- Advanced Telecommunication Engineering, MOS Devices, VLSI Technology and Device Modeling, Advanced Digital Signal Processing, Broadband Wireless Communication, Power Semiconductor Circuits.

Taught Undergraduate Courses

As a Teaching Assistant at the University of Texas at Dallas

- Topology, Applied Calculus, Differential Calculus, Calculus of Several Variables, Abstract Algebra, College Algebra for the Non-Scientist, Integral Calculus, Theoretical Concepts of Calculus, Differential Geometry, Linear Algebra, Mathematical Analysis.

As a full-time and part time instructor in Bangladesh

- Power Electronics , VLSI Design , Digital Logic Design , Numerical Analysis , Optoelectronics , Digital Signal Processing , Digital Communication , Processing & Fabrication , Wireless Communication , Microprocessor & Interfacing , Web Database Programming , Electronic Devices & Circuits , Neural Networks , Control Systems Engineering , Electrical Circuit Analysis.

Supervisory Experience

Undergraduate Thesis Supervision

- Hypothetical Quantum wire of Hydrogen - The People's University of Bangladesh.
- Design and Implementation of a Web-Based Teamwork Management System - The People's University of Bangladesh.

Training Supervision

- Training on Computer Interfacing - November, 2012, Jagannath University, Dhaka.
- Training on Web- Technology - January, 2013, Jagannath University, Dhaka.
- Training on Wireless Mobile Application - February, 2013, Jagannath University, Dhaka.
- Training on MATLAB - March, 2013, Jagannath University, Dhaka.

References Available upon request.