Md Muhtasim Billah

LinkedIn: in/mmb039
Github: github.com/mmbillah
Portfolio: mdmuhtasimbillah.netlify.app
Kaggle: kaggle.com/mdmuhtasimbillah

1650 NE Valley Road, Pullman, WA 99163
Mobile: 509-330-6287
Email: mdmuhtasim.billah@wsu.edu
Medium: mmbillah.medium.com
Google Scholar: scholar.google.com/citations

Experience

- **Paylocity** — Schaumberg, IL
  Data Science Intern — *May 2022 - August 2022*
  Worked on modeling, benchmarking and unit testing NLP tasks. Contributed towards building, testing and deploying a REST API with multiple NLP endpoints hosted on AWS. Also worked on several other projects that are in use across company's data practice org.
- **Washington State University** — Pullman, WA
  Graduate Researcher — *August 2018 - Present*
  Applied ML techniques (i.e. physics informed neural network (PINN) and Bayesian inference) for heat transfer problems using Python. Monte Carlo based probabilistic modeling in C++ and Fortran for studying bioparticle transport across blood-brain barrier (BBB).

Education

- **Washington State University** — Pullman, WA
  Ph.D. Candidate in Mechanical Engineering — *Expected Fall 2023*
  Courses: *Statistical Computing, Neural Networks, Big Data and Cloud Computing, Data Mining and Analysis, Numerical Methods*
- **Washington State University** — Pullman, WA
  M.S. in Statistics — *Expected Fall 2022*
  Courses: *Applied Linear Models, Design of Experiments, Statistical Theory, Multivariate Analysis, Genomic Data Analysis*
- **Bangladesh University of Engineering and Technology** — Dhaka, Bangladesh
  B.S. in Mechanical Engineering — *May 2012 - Feb 2017*

Skills Summary

- **Languages** Python, R, SQL, Matlab, SAS, C++, C, Fortran, Bash
- **Database** MySQL, Microsoft SQL Server, PostgreSQL
- **Frameworks** Scikit-learn, TensorFlow, PyTorch, Keras, Transformers, Spark, FastAPI, dplyr, caret, ggplot2
- **Tools** Git, Docker, AWS, Azure, Pulumi, Databricks, Teamcity, Postman, Octopus Deploy, Tableau

Projects

- **Multilabel Classification of Drug from Mechanism of Action (MoA):** Employed several FFNN architectures with hyperparameter tuning and multilabel stratified k-fold cross validation. Created model ensemble to further minimize the cross entropy loss. Acquired bronze medal in the associated Kaggle competition launched by LISH at Harvard (2020). Link
- **Predictive Modeling of Trading Decisions for Global Stock Exchange:** Trained RNN, LSTM models on time series data (real stock market data consisting 2.39M tradings and 130 features) to maximize AUC. Achieved high utility score (weighted ROI) by testing on future data. Received bronze medal in the associated Kaggle competition lauched by Jane Street (2021). Link
- **End-to-end Recommender Systems for Amazon Products:** Used Apache Spark to handle large Amazon datasets (233M reviews). Wrote Python and SQL scripts to parse and import data into MySQL. Applied memory based (user and item based) and model based (SVD, ALS matrix factorization) collaborative filtering methods. Harnessed fast cloud computing environment on AWS EC2. Link
- **Sentiment Analysis of COVID-19 Vaccines Based on Twitter Data:** Mined 42,478 unique tweets filtered by few predetermined keywords. Processed tweet metadata by tokenizing, stemming and lemmatizing. Determined positive, negative and neutral tweets from the polarity scores. Visualized findings with geolocation filters in Tableau dashboard. Link1 Link2
- **Cancer Classification and Clustering from Gene Expression Monitoring:** Performed PCA on 7,123 human genes (found from microarrays data). 85% of the total variance was found to be explained by top 50 genes. Applied k-means clustering for analyzing cancer classes AML and ALL. Employed Elbow and Silhouette Score methods for the selection of k. Link
- **Stock Portfolio Optimization and Diversification:** Performed Monte Carlo simulation on stock prices of top tech companies based on Modern Portfolio Theory (MPT) using Sharpe Ratio as metric. Generated insightful plots from the simulation results to suggest optimum diversification of user portfolio to maximize return and minimize volatility. Link
- **Female Employment Against Socioeconomic Factors in Bangladesh:** Processed World Bank data on Bangladesh spanning over 30 years. Utilized statistical testing and diagnostic plots for checking model assumptions, possible outliers, multicollinearity and autocorrelations. Multivariate regression achieved an adjusted R-squared value of 0.99. Link

Publications

- **Bayesian Method for Parameter Estimation in Transient Heat Transfer Problem,** A.I. Khan, **M.M. Billah**, C. Ying, J. Liu, P. Dutta, *International Journal of Heat and Mass Transfer,* Feb 2021. Link
- **Physics Informed Deep Neural Network for Solving Inverse Heat Transfer Problem, M.M. Billah**, A.I. Khan, J. Liu, P. Dutta, *International Journal of Heat and Mass Transfer,* Submitted Sep 2022.

Honors and Awards

- **Kaggle Competition Expert** - Received Kaggle competition expert badge (ranked within the top 1%), 2021.
- **Kaggle Bronze Medals** - Jane Street Market Prediction (top 8%), 2021 Link & MoA Detection (top 9%), 2020. Link
- **Best Project** - Best project (1st of 15 teams) award, CptS 415: Big Data, WSU, Fall 2020.
- **Dean's List Scholarships** - Awarded by Faculty of Mechanical Engineering, BUET, 2017 & and 2016.

Certifications

- **Data Scientist with Python, DataCamp** - Issued: April 2021 Link
- **Deep Learning Specialization, Deeplearning.ai** - Credential ID: BZABKY6TK8F3, Issued: Nov 2020. Link
- **Machine Learning, Stanford University** - Credential ID: 7QYVHH69P4ZL, Issued: July 2020 Link