Umnoon Binta Ali

Address: Sector 18, Uttara, Dhaka
E-mail: umnoonbintaali@gmail.com ✻ *Phone:* +880 1627 097 553
LinkedIn: linkedin.com/in/umnoonbintaali/
Portfolio: https://umnoon.github.io/

Objective

Highly motivated Data Scientist with a strong foundation in statistical analysis, machine learning and artificial intelligence. Possess experience in designing and implementing data pipelines, building models, leveraging cloud platforms and communicating insights to stakeholders.

Work experience

Astha.IT (Bangladesh) *September 2023 - Current*
Data Scientist *Dhaka, Bangladesh*

- Developed and implemented PySpark transformations within a drag-and-drop ETL tool, optimizing data processing efficiency for a seamless user experience.
- Led the fine-tuning of a Keras model in a gaming analytics project, enhancing the accuracy of event detection in real-time gameplay data.
- Spearheaded the fine-tuning of a Keras model in a gaming analytics project, enhancing the accuracy of event detection in real-time gameplay data.
- Collaborated with cross-functional teams to design and implement scalable and efficient PySpark workflows for data processing, ensuring high-performance data pipelines.
- Applied statistical methodologies like empirical Bayesian estimation, to derive meaningful patterns and trends from large datasets, contributing to data-driven decision-making.
- Played a pivotal role in the end-to-end machine learning model deployment, leveraging cloud computing platforms such as Amazon SageMaker for scalability and efficiency.
- Conducted thorough data analysis to identify opportunities for optimization and enhancement, leading to actionable recommendations for game feature improvements.
- Collaborated with stakeholders to gather and understand requirements, ensuring the successful translation of business needs into effective data science solutions.
- Stayed abreast of industry trends and emerging technologies, incorporating new tools and methodologies into the workflow to enhance overall project efficiency and effectiveness.

Astha.IT (Bangladesh) *July 2023 - September 2023*
Intern *Dhaka, Bangladesh*

- Collaborated with the data engineering team to troubleshoot and resolve issues related to data pipeline performance, ensuring the reliable and timely delivery of high-quality data to downstream applications.
- Contributed to the development and optimization of PySpark SQL and ML workflows, enhancing data processing efficiency and supporting the integration of machine learning models into the data pipeline.
- Implemented and maintained ETL pipelines using Apache Airflow, ensuring seamless data flow and optimal scheduling for data processing tasks.
- Documented and maintained comprehensive documentation for ETL processes, data schemas, and system configurations, facilitating knowledge transfer and providing a valuable resource for future development and troubleshooting.

North South University *April 2022 - July 2022*
Research Assistant *Dhaka, Bangladesh*

- Analyzed data, prepared datasets, used Transfer Learning, Knowledge Distillation.
- Bangla NLP scopes and future.

Portfolio of most relevant projects

Quantalyze.ai *Pilot Phase*
Astha.IT

- Spearheaded the development of Quantalyze.ai, an ETL tool, by designing and implementing critical features that enable users to effortlessly extract, transform, and load data.
- Leveraged PySpark capabilities to create efficient and scalable data transformation modules within Quantalyze.ai, ensuring optimal performance for diverse data processing tasks.
- Collaborated closely with cross-functional teams to gather requirements and align development efforts with user needs, ensuring Quantalyze.ai met industry standards. Additionally, played a key role in documenting technical processes for the tool, facilitating ongoing maintenance and future development.

DataCoach Rocket League *On Going*
Astha.IT

- Collaborated closely with cross-functional teams to gather requirements and align development efforts with user needs, ensuring Quantalyze.ai met industry standards. Additionally, played a key role in documenting technical processes for the tool, facilitating ongoing maintenance and future development.
- Utilized PySpark on Databricks to conduct comprehensive analytics for DataCoach, generating user-specific answers and actionable recommendations for game improvement strategies based on the analysis of large-scale gaming data.
- Played a crucial role in optimizing PySpark workflows for game analytics, ensuring efficient data processing and facilitating the seamless integration of machine learning insights into the overall DataCoach system.

Technical skills

Languages	Python, R, C++, Java, PHP, Latex
Machine Learning Frameworks	Keras, Pytorch, TensorFlow, Scikit Learn
Data Analysis and Visualization	Pandas, Matplotlib, Seaborn, Plotly
Computer Vision	OpenCv
Database Management	PostgreSQL, MySQL, SQLite
Big Data Tools	PySpark, Hadoop, Spark, Databricks, Apache Airflow
Cloud Services	Amazon S3, Amazon SageMaker, AWS Lambda
Web Frameworks	Django, Flask
Data Visualization Tools	PowerBI, Tableau, Streamlit
Version Control	Git
Operating System	Windows, Linux
Office Tools	MS Office, Google Workspace

Education

Bachelor's in Computer Science and Engineering — *Dhaka, Bangladesh*
North South University — *January 2017 - December 2022*
CGPA: 3.60

Higher Secondary School Certificate — *Dhaka, Bangladesh*
Rajuk Uttara Model College — *2016*
GPA: 5.00

Secondary School Certificate — *Dhaka, Bangladesh*
Rajuk Uttara Model College — *2014*
GPA: 5.00

Academic Projects

Transfer Learning for Speaker Diarization on Bangla Audio Dataset
CSE499 Graduate Dissertation
In this research, explored transfer learning for speaker diarization on a Bangla audio dataset. Experimented with different clustering algorithms and embedding techniques to reduce Diarization Error Rate (DER) for noisy data.

Contrastive Learning for Text-to-Image on Paraphrasing Captions
CSE498 CO-OP Research Project
Presented a contrastive learning strategy to enhance the quality and semantic consistency of synthetic images by paraphrasing captions from the CUB dataset.

A Deep Convolutional Neural Network for Bangla Handwritten Numeral Recognition
CSE465 Pattern Recognition and Neural Network
Studied the NumtaDB dataset for classifying images of isolated Bangla numerals. Augmented synthetic data using the Keras library and employed the ResNet34 CNN architecture to improve model performance.

Tweet Emotion Recognition with TensorFlow
Coursera
Developed an RNN for recognizing emotions in tweets, using a dataset from Hugging Face with thousands of tweets classified into one of six emotions.

Outbreak Prediction and Visualization with PowerBI: Covid-19
CSE299 Junior Design Project
Utilized GOARN data to predict the time of the highest spread of Covid-19 and created time-series visualizations with Power BI.

Data Science Web App for Motor Vehicle Collisions in NYC
Coursera
Built a data science web app with Streamlit analyzing motor vehicle collisions in NYC. Utilized libraries such as NumPy, Pandas, Pydeck, and Plotly.

Calculator with JavaScript
Coursera
Developed a calculator using JavaScript, incorporating ES6 classes, and utilizing CSS Grid and Flexbox for improved organization and UI syncing.

Licenses and Certifications

- Hyperparameter Tuning with Keras Tuner
- Deep Learning with Pytorch: GAN

- Multivariate Calculus for Machine Learning
- Data Science Math Skills
- Introduction to Artificial Intelligence

  Linear Algebra for Machine Learning
- Introduction to AR and ARCore
- Introduction to Python
- Introduction to R
- Intermediate R

Language proficiencies

English	Professional
Bengali	Native Speaker

References

Excellent references are available upon request

✻ *I hereby declare that all the details mentioned above are in accordance with the truth and I hold the responsibility for the correctness of the above-mentioned particulars.*