

# RAKSHAK KUNCHUM

Boston, MA | (857) 230-2483 | [krakshak7698@gmail.com](mailto:krakshak7698@gmail.com) | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

**Northeastern University**, Boston, MA

January 2023 – Present

Khoury College of Computer Science

Expected Graduation: December 2024

Master of Science in Data Science, **GPA: 4.0/4.0**

Related Courses: Supervised Machine Learning, Data Mining, Data Processing and Management, Algorithms, Deep Learning.

**BMS College of Engineering**, Bangalore, India

September 2016 - August 2020

Bachelor of Engineering in Information Science Engineering, **GPA: 8.8/10**

Related Courses: Machine Learning, Statistics, Relational Database Management Systems, Data Science Foundations, Programming.

## TECHNICAL SKILLS

- **Programming Languages:** Python, R programming, SQL, Java.
- **Databases:** MySQL (Server and Workbench), MongoDB.
- **Cloud Computing Platforms:** AWS (EC2, S3, Athena, Redshift).
- **Toolkits/Software:** Selenium, Airflow, Kafka, Docker, Tableau, Git, Excel, Shell Scripting (Unix/Linux), Web Scraping.
- **Data Science Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Plotly, Seaborn, Tensorflow, Pytorch, Dask, OpenCV.
- **Data Science Techniques:** Regression, Decision Trees, Random Forest, Boosting, SVM, Ensemble Models, Clustering, PCA, Neural Networks (CNN, RNN, LSTM, GAN), Inferential Statistics, Hypothesis Testing, Statistical Forecasting.
- **Soft Skills:** Effective Communication, Analytical Skills, Problem-Solving, Leadership-Oriented, Proactive, Self-Motivated.

## PROFESSIONAL EXPERIENCE

**Dataweave (Infoweave Analytics Pvt Ltd), Bangalore, India**

July 2020 - October 2022

Data Engineer

- Collaboratively automated more than 10 end-to-end data science pipelines using **Apache Airflow, Python, and AWS** to integrate web crawlers, data sources, APIs and internal ETL frameworks resulting in enhanced project outcomes.
- Programmed complex **SQL** queries to handle large datasets using **AWS Athena and S3** to generate business data for reports and interactive dashboards, empowering data-driven decision-making and enhancing business value for clients.
- Attained 50% reduction in text/image **batch processing** in the company's clustering algorithm by implementing **dask** library.
- Developed and deployed over 15 **Selenium browser automation** data crawlers and 200+ **Python-based deep crawler bots** to efficiently mine data and insights from e-commerce websites, bolstering the company's product offerings.

**Dataweave (Infoweave Analytics Pvt Ltd), Bangalore, India**

January 2020 - June 2020

Data Engineering Intern

- Performed **data analysis** and **brand analysis** to identify counterfeit products, ensuring brand protection and compliance with price benchmarks for enhanced market competitiveness.
- Led the successful delivery of structured data by employing **data-wrangling** methods that unlocked new business for clients.

## ACADEMIC PROJECTS

**US Air Pollution Time Series Analysis** | Python/Deep Learning/Time Series

- Applied advanced time-series analysis and forecasting approaches (SARIMAX, LSTM) to predict future US air pollution trends, providing valuable data-driven insights for environmental monitoring and strategic decision-making.

**Forecasting Credit Card Expenditure** | Python/Machine Learning/Boosting/Bagging

- Engineered and optimized a regression-based machine learning model to predict credit card spending using feature selection techniques, bagging, boosting, and grid search, enabling data-informed credit card limit decisions for the banking industry.

**Trending News Prediction using Machine Learning and Web Scraping** | Python/Web Scraping/Machine Learning/NLP

- Developed a web scraper to collect news articles, utilized NLP techniques for data processing, and applied multiple ML models (logistic regression, k-NN, decision tree, random forest, xgboost) to predict and analyze news trends.

**Deep Learning Approaches to Detect Pneumonia** | Python/Deep Learning/Computer Vision

- Ran a comparative analysis of three machine learning models (CNN, U-Net, Mask-RCNN) for pneumonia detection.
- Achieved precise localization of pneumonia infection regions by leveraging the performance of the Mask-RCNN model.

## EXTRACURRICULARS

Data Science Hub, Northeastern Graduate Student Government, Hiking, Volleyball, Video Gaming.