

# PRAGYA RAGHUVANSHI

Tel: +1 984-683-3369 | [pr158@duke.edu](mailto:pr158@duke.edu) | [linkedin.com/in/pragyaraghuvanshi](https://www.linkedin.com/in/pragyaraghuvanshi) | [github.com/pr-124](https://github.com/pr-124)

## EDUCATION

**Duke University**, Durham, NC  
*Master of Science (Data Science)*

May 2024

- Relevant Courses: Computer Vision, Deep Learning, Machine Learning, Causal Inference, Natural Language Processing, Data Engineering, Deep Reinforcement Learning, Biostatistics, Statistical Modeling, Data Visualization, Practical Data Science Tools

**Indian Institute of Technology**, Roorkee, India  
*Master of Technology (Power System Engineering)*

May 2016

- Relevant Courses: MATLAB, Computer Aided Power System Analysis, Machine Learning (Thesis)

**Uttar Pradesh Technical University**, Lucknow, India  
*Bachelor of Technology (Electrical Engineering)*

May 2013

- Relevant Courses: Linear Algebra, Probability and Statistics, Computer Programming in C, Calculus, Discrete Mathematics

## SKILLS

- Languages:** Python, R, SQL, Bash, C, MATLAB
- Database, Big Data & Visualization:** Pandas, GeoPandas, MySQL, PostgreSQL, MongoDB, Spark, Hadoop, Matplotlib, Seaborn, Tableau
- Software Tools & Cloud:** AWS, Azure, Databricks, GCP, Big Query, Dask, Docker, Snowflake, Git, Airflow, Kafka, Redshift, R Shiny
- Machine Learning Framework:** NumPy, PyTorch, TensorFlow, Keras, NLTK, Scikit-learn, SpaCy, HuggingFace, Langchain

## WORK EXPERIENCE

**Energy Data Analytics Lab, Duke University**, *Research Assistant*

August 2023 – January 2024

- Instrumental in developing a high-resolution(1km<sup>2</sup>) deep learning model using satellite-derived features to estimate global greenhouse emissions.
- Implemented a data pipeline for processing 600M+ geospatial data points using multi-processing techniques with Spark, Dask, and GeoPandas.

**SunFi**, *Data Science Intern*

May 2023 – August 2023

- Improved the prediction of customers' Probability of Default (PD) by ~40% by deploying a credit risk model pipeline, collaborating effectively with cross-functional teams. Utilized a range of techniques like cost sensitive predictive algorithms, SMOTE and PCA, which resulted in a boost in model performance by 20% and more reliable risk assessments.
- Analyzed Shapley Additive explanations (SHAP) values for interpretable clustering analysis, to gain insights into shared financial behavior.
- Identified anomalous transactions and potential fraud by implementing outlier detection methods like Isolation Forest. Enhanced preprocessing by fine-tuning a pre-trained NER model on bank statements to extract key entities, thereby improving the accuracy of fraud detection.

**KK Footwear**, *Data Scientist*

November 2017 – March 2022

- Streamlined operations by implementing demand forecasting using statistical techniques (ARIMA, ETS) and machine learning models (e.g., LightGBM) capturing seasonality, trends, and sales variability. Reduced stock-outs by 20% and quadrupled annual inventory turnover.
- Expanded brand portfolio (1 to 5 brands) and retail partnerships (19 to 124) by implementing ML models for market trend analysis and product recommendations, leveraging sentiment analysis, topic modeling, and NER on customer reviews and social media data.
- Optimized pricing using models like GAMs leveraging historical sales data to adjust pricing in real-time. This approach improved profit margins by identifying optimal price points, timing discounts effectively, leading to a 5% increase in revenue and enhanced customer satisfaction.
- Designed and implemented 10+ ETL pipelines for data warehousing to operationalize data for improved data quality and extraction of KPIs from sales data. This data-driven approach resulted in a 30% increase in Average Order Value (AOV) and an additional INR 13.5 million in revenue.

**Larsen & Toubro-Sargent & Lundy Limited**, *Power System Engineer*

August 2016 – July 2017

- Engineered comprehensive system reliability studies, integrating machine learning for predictive short circuit analysis for routine management.

## RESEARCH & PROJECTS

**A Deep Learning Approach to Correcting Limb Deformity** *Computer Vision, Transfer Learning, Image Preprocessing*

August 2023 – Present

- Fine-tuned a state-of-the-art ResNet-101 computer vision model for the autonomous identification of anatomic key-points from hip and knee radiographs, reducing limb deformity following orthopedic trauma; abstract accepted at the Orthopedic Trauma Association (OTA) 2024.

**Assessing Look-ahead bias in LLMs** *Generative AI, LLM, Prompt Engineering*

January 2024 – Present

- Evaluate the extent of familiarity bias and look-ahead bias in LLMs when analyzing earnings call transcripts for downstream tasks. Develop mitigation strategies using a combination of company anonymization and prompt engineering.

**Achieving Efficiency in Large Language Models** *Generative AI, LLM, Model Compression*

August 2023 – December 2023

- Implemented quantization and pruning techniques to reduce the size of LLMs (3B to 7B range) and improve their inference speed by at least 50% with minimal loss in performance.

**Debiasing Machine Learning models for Financial Services** *Fairness, Debiasing, Machine Learning*

August 2023 – May 2024

- Collaborated with 2OS, to improve fairness, using techniques like matching to control for confounding variables and algorithms like reweighing, in a Gradient Boosting (XGBoost) model, to improve the Disparate Impact across race and gender by 40% without substantial loss in accuracy.

**Amend Class Imbalance using Data Augmentation** *Deep Learning, Data Augmentation, Generative models*

March 2023 – April 2023

- Performed data augmentation by adding synthetic images generated by DCGANs (Generative Adversarial Networks) and Stylistic GAN to amend class imbalance on facial emotion dataset to improve emotion recognition accuracy of CNN and Resnet-50 Classifier by 2-4%.