PRAGYAN JYOTI BORTHAKUR

• (480) 953-5995 • pborthak@asu.edu • linkedin.com/in/pragyan555 • pragyanborthakur.online

EDUCATION

Master of Science in Data Science, Analytics and Engineering **Expected May 2026**
Arizona State University, Tempe, AZ **3.66/4 GPA**

SKILLS

Programming Languages: Python (NumPy, Pandas, Matplotlib), R, Java, Next.js, JavaScript, HTML, CSS
Tools: SQL, Power BI, Spark, PyTorch, Scikit-Learn, Git, Jira, Azure, Databricks, AWS, CRM, Excel, Word,
Areas of Expertise: Data Science, Data Analytics, Business Analytics, Machine Learning, Predictive Modeling, Time Series Analysis, Data Transformation Pipelines, Data Warehousing, Predictive Analytics, Regression Analysis

PROFESSIONAL EXPERIENCE

The Boeing Company, Bengaluru: Data Engineer **August 2023 – August 2024**

- Entrusted with managing and streamlining ingestion of diverse obstacle data (CSV, PDF, XML) from 15+ global sources into **Databricks,** executing precise risk evaluations, and implementing workflows boosting data processing efficiency.
- Led a team of three engineers to develop an automated **ETL** pipeline with **PySpark** and **Python**. Conducted process mapping and gap analysis to identify inefficiencies in data workflows, resulting in enhanced data management practices.
- Created interactive **Power BI** dashboards to communicate insights to cross-functional teams, enhancing data accuracy by 2**0%** and aligning project requirements with business objectives.
- Recognized with *'Boeing Recognition Award'* for processing complex PDF data with **99%** accuracy, cutting down data extraction time by **90%**.

Skill Vertex, Bengaluru: Cloud Intern **March 2023 – May 2023**

- Took initiative to design a scalable and innovative serverless text transfer system, improving application responsiveness and user experience beyond original project scope.
- Developed a serverless text transfer system with **AWS Lambda** and **DynamoDB**, enabling real-time data exchange, decreasing latency by **40%,** and ensuring scalability for a cloud-based traffic platform.

PROJECTS & PUBLICATIONS

Sustainable Innovation Web Application, *Personal Project* **February 2025**

- Crafted a **full-stack Next.js** application championing sustainable waste management by harnessing Google's **Gemini** API to reimagine discarded items as valuable resources through inventive upcycling, repurposing, and recycling ideas.
- Engineered secure user authentication with **NextAuth.js**, incorporated **MongoDB** for reliable data storage, and assembled a modern UI using **Radix** UI components featuring customizable profiles and avatar systems, showcasing advanced expertise in current web development practices.

Adversarial Robustness Evaluation for LLMs*, Personal Project* **April 2024**

- Built an adversarial robustness evaluation pipeline leveraging **PyTorch**, **Transformers**, and **Hugging Face** to test GPT-2 on 600+ prompt–attack–defense scenarios; implemented six defense strategies, including a classifier (**99.06%** accuracy).
- Assessed LLM alignment and safety through **NLP**-based metrics, sentence embeddings, and **SQuAD**-derived prompts; generated reproducible analysis reports and visualized results for model performance benchmarking.

Multivariate S&P 500 Forecasting Using News Sentiment and Global Indices, *Personal Project* **March 2025**

- Engineered an **LSTM**-based multivariate forecasting model integrating historical index prices (S&P 500, Nikkei, Hang Seng, Shanghai) and FinBERT-derived sentiment from financial news to predict market movements.
- Achieved out-of-sample correlations of **0.966** (daily), **0.982** (weekly), and **0.997** (monthly), demonstrating strong predictive accuracy for macroeconomic trend analysis and investment decision support.

Detection And Classification of Oral Cancer, *Class Project* **April 2023**

- Built a DenseNet network architecture with **Scikit-Learn** and **TensorFlow.** Applied machine learning techniques, achieving **95.24%** accuracy in oral cancer classification while optimizing training time through transfer learning**.**
- Utilized advanced data preprocessing techniques, including feature extraction, image augmentation (rotation, flipping, and scaling), and noise reduction. These methods improved model accuracy from **70.71%** to **95.24%.**
- **Published** a paper titled "Deep Learning Techniques for the Detection and Classification of Oral Cancer Using Histopathological Images" at the 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT). DOI: 10.1109/ICCPCT58313.2023.10244890.