Ayush Baral

601-398-7530 | ayushbrt73@gmail.com | linkedin.com/in/ayush-baral

SKILLS

- Programming Languages & Tools: Python, R, C++, JMP, SQL
- Data & Visualization: Tableau, Power BI, R Shiny, ggplot2, Seaborn, Plotly Dash
- Libraries & Frameworks: NumPy, Pandas, Scikit-Learn, statsmodels, TensorFlow, Keras, PyTorch, OpenCV, NLTK, spaCy
- Cloud & DevOps: Azure (Web Apps, Blob), Docker, GitHub Actions, CI/CD
- LLM & NLP: Claude 3.5, LangChain, BERT, spaCy NER, Prompt Engineering

WORK EXPERIENCE

Apollo MIS: White Plains, New York
Data Scientist April 2024 – Present

- Led full-stack development at an early-stage startup, building an AI-powered voice bot for real-time sports analytics. Delivered end-to-end features across frontend (React Native), backend (Flask), and DevOps, taking the product from concept to production.
- Developed prompt engineering strategies using Claude 3.5, achieving greater than 85% SQL query accuracy. Replaced greater than 95% of pattern search logic with BERT-based similarity search and enhanced data quality with spaCy-based NER.
- Engineered a dynamic revision pipeline for LLM-generated SQL, leveraging execution feedback, query intent reasoning, and schema metadata to refine queries and reduce manual corrections.
- Deployed the chatbot using Azure Web Apps with containerization and built a CI/CD pipeline for seamless updates. Ensured scalable, low-latency interactions via React Native frontend and Flask backend hosted in a reliable Azure environment.

First Solar, Inc.: Perrysburg, Ohio
Associate Data Scientist January 2023 – April 2024

- Developed an automated defect detection pipeline for solar panels using computer vision and machine learning, leveraging OpenCV, SVM, and XGBoost for multi-label classification; Attained high F1-score - reducing manual annotation time by over 3 hours per batch.
- Applied unsupervised learning techniques, including dimensionality reduction (SVD, PCA) and clustering, to optimize experimental design by identifying multicollinearity and creating orthogonal encodings for categorical variables, improving model interpretability.
- Designed predictive models using regression analysis and statistical inference to analyze the impact of manufacturing parameters on panel efficiency, enabling data-driven optimization of production processes.

BRE Hotels & Resorts (Blackstone Real Estate): New York City, New York
Data, Analytics and Research Intern June 2022 – August 2022

- Developed and executed an extensive Tableau dashboard to monitor key performance indicators (KPIs) specific to the hospitality industry.
- Managed the full business intelligence lifecycle, including data modeling and the smooth deployment of the production-ready dashboard.

Credit Suisse: Morrisville, North Carolina
Technology Analyst Intern July 2020 - August 2020

- Experimented with AR, MA, ARIMA, LSTM, and SARIMAX, for forecasting trade volumes of different asset classes.
- Strategically utilized these models to proactively target anomalies and achieve MAPE of ~12%.
- Built a UI utilizing Python Dash via whereby the analysts had immediate insight to the current metrics.

Cree, Inc.: Durham, North Carolina
Information Security Intern May 2019 - August 2019

- Built an interactive Tableau dashboard to show various trends of vulnerabilities present in the company's IT infrastructure.
- Built a scheduled pipeline to clean the data from Tenable API and store it in a MS-SQL database which was later used for visualization.
- Analyzed trends leading to increased visibility of asset owners and respective unpatched vulnerabilities.

EDUCATION

Columbia University New York City, New York
Master of Science in Data Science Graduated: December 2022
Courses: Algorithms, Theoretical Machine Learning, Applied Machine Learning, Time Series Modeling, Statistical Inference

Alcorn State University Lorman, Mississippi
Bachelor of Science in Computer Science | Summa Cum Laude Graduated: May 2021
Courses: Data Structures & Algorithms, Computer Networks, Linear Algebra, Discrete Mathematics, Numerical Analysis, Operating System

TECHNICAL PROJECTS

- ***US Equity Arbitrage Project:*** Implemented the strategies outlined in the research paper *'Statistical Arbitrage in the US Equity Market'* to develop a robust trading model for analyzing and exploiting market inefficiencies. Github.
- ***Johnson & Johnson Capstone Project:*** Prediction of commercial insurance payments for surgical procedure using Machine Learning in collaboration with Columbia University and Johnson & Johnson. Poster.
- ***Stellar Blockchain – Gas price Prediction:*** Developed a predictive model using XGBoost, LightGBM, Gradient Boosting, HistGradient Boosting, and ExtraTreeRegressor to accurately estimate gas prices. Achieved 3rd place in the competition. Certification.