

Sujay Shrivastava

📍 Buffalo, New York | 📞 716-617-1691 | ✉️ sujay8192@gmail.com

🌐 [linkedin.com/in/sujayshrivastava](https://www.linkedin.com/in/sujayshrivastava) | 🌐 [jayshrivastava0.github.io/portfolio](https://github.com/jayshrivastava0) | 🌐 github.com/jayshrivastava0

Skills

Programming Languages: Python, SQL, R, MATLAB, NoSQL

Core Competencies: Machine Learning, Computer Vision, Deep Learning, Clustering, Reinforcement Learning, Data Visualization, Data Analytics, Big Data Technologies, Statistics, Regression Analysis, Data Storytelling, Natural Language Processing, Time Series Analysis

Frameworks: NumPy, Pandas, scikit-learn, SciPy, TensorFlow, Pytorch, Keras, Hadoop, PySpark, Gemini, GPT, BERT, Matplotlib

Tools: Google Cloud Platform (Compute, Vertex AI, Big Query), GitHub, Power BI, Tableau, Docker, Unix/Linux, Microsoft Office Excel

Soft Skills: Agile Project Management, Stakeholder Management, Cross Functional Collaboration, Teamwork

Education

State University of New York at Buffalo | Master's in Data Science

Jan 2023 - May 2024

Experience

AI Engineer | Reality AI

Mar 2024 - Present

- Enhanced educators' efficiency by **40%** by developing and deploying **Gen AI - driven tools**, including a Rubric Generator and Flashcard Creator, leveraging **LangChain** and **RAG** for context-aware automation.
- Increased **platform interaction by 25%** by engineering **predictive analytics models** using **PySpark**, **Vertex AI**, and **LangSmith** to optimize user engagement through **tailored content recommendations**.
- Improved **user satisfaction by 18%** by conducting **statistical analyses** with **Python** and **TensorFlow**, refining LangChain-powered educational tools based on insights into AI-generated **content effectiveness**.
- Streamlined decision-making processes by **20%** by designing interactive **Power BI dashboards** for real-time monitoring of **AI tool performance** and user engagement metrics.

Data Scientist – Research Assistant | University at Buffalo

Sep 2023 - May 2024

- Revived a **6-month stagnant** project saving **\$3,000** by identifying and resolving critical data preprocessing issues performing **root cause analysis** and agile problem-solving techniques.
- Optimized **MLOps** efficiency by **75%**, **reducing training time** from **2 hours to 30 minutes**, through developing a custom CNN pipeline using **few-shot learning**, and **containerization** for streamlined model deployment on GCP.
- Improved **Intersection over Union (IoU)** metric by **20%**, through **hyperparameter optimization** and fine-tuning of **open-source pre-trained models** by Meta and AllenAI, enabling better adaptation to our specific use case.
- Streamlined model prediction by **10%**, **reducing latency** and handling **1000 concurrent requests**, with automated testing and deployment via **Vertex AI Endpoints**, ensuring reliable **real-time inference** and improved model serving.

Data Analyst | Curelink

May 2022 - Aug 2022







- Saved **\$500k** from **projected losses**, and enabled strategic business restructuring, by leading a **Scrum project** that analyzed departmental forecasts using **predictive modeling**, resulting in the **closure of a non-profitable department**.
- Created a **powerBI dashboard** to visualize inside sales performance, tracking call durations and successful conversions, which led to a **performance-based incentive** program and a **12% increase** in sales team **productivity**.
- Engineered **ETL pipelines** to clean data with **complex SQL queries**, integrating it into **PowerBI** and **reducing redundancy** by **8%** creating various dashboards to track Key Performance Indicators (**KPI**).
- Improved decision-making efficiency by **12%** by developing **Power BI** dashboard to measure **recovery time** of expenditure of **marketing campaign**, providing insights to **cross-functional** teams, resulting in targeted perks for **customer retention**.

Data Scientist | IIT Kharagpur

Jan 2022 - Mar 2022

- Architected a **dynamic meta-learner ensemble** combining **LightGBM**, **CatBoost**, and **Neural Networks** for airfare prediction, achieving **92% accuracy** in price forecasting, resulting in a **25% improvement** in revenue optimization for route planning.
- Leveraged **Bayesian optimization** for **hyperparameter tuning** across 200+ parameters, resulting in **28% lower RMSE** compared to baseline models and identifying price elasticity patterns that **increased booking conversion by 22%**.
- Implemented **automated anomaly detection** using isolation forests to identify pricing irregularities across **50+ routes**, leading to the detection of **seasonal pricing patterns** that **increased profit** margins by **15%**.
- Orchestrated **A/B testing** framework to **validate model** performance across different market segments, leading to identification of **price-sensitive routes** with **85% confidence intervals**.

Kaggle Competitions and Projects

- Kaggle Regression of Used Car Prices (Rank: 428)** | Regression Analysis, Feature Engineering 
- Kaggle Jane Street Real-Time Market Data Forecasting (Rank: 488)*** | Stock Market Forecast, Time Series 
- Hill Climb Racing AI Bot** | Artificial Intelligence, Computer Vision, Object Detection, Reinforcement Learning 
- US Accident Analysis** | Data Wrangling, Data Visualization, PCA, Sampling Techniques 
- Land Mines Clustering and Oil Sales Forecasting** | Anomaly Detection, Trend Forecasting and Decomposition 
- Telecom Customer Churn Prediction** | Sampling Techniques, Churn Prediction, Hyperparameter Tuning 

Publications and Certifications

- GeoAI** in resource-constrained environments, at **IEEE AI**. 
- Microsoft:** Azure AI Engineer Associate (AI - 102) 