Dhruv Sood

New York | (203) 428-8995 | ds7403@nyu.edu | LinkedIn | Github | Portfolio

Summary

Data Scientist with strong experience deploying LLM-driven solutions and predictive models in enterprise settings. Improved decision-making speed and efficiency at SQOR.ai and developed a cost-saving predictive maintenance model at Dassault Systèmes. Skilled in building data pipelines and implementing advanced machine learning across cloud platforms. Looking to leverage these strengths to drive impactful, data-driven results.

EDUCATION

New York University **Sep 2023 - May 2025**
Master of Science, Information Systems *New York, United States*
- **Coursework:** Data Science, Data Mining, Database Management Systems, Applied Statistics, Business Analytics, Data Visualization

Vellore Institute of Technology **Jul 2019 - Jun 2023**
Bachelor of Technology, Computer Science *Vellore, India*
- **Coursework:** Data Structures & Algorithms, Artificial Intelligence, Machine Learning, Software Engineering, Probability & Statistics

TECHNICAL SKILLS

- **Programming & Tools**: Python, SQL, Tableau, Power BI, GCP, Microsoft Azure, AWS, Git, Flask, FastAPI
- **Libraries & Frameworks**: TensorFlow, PyTorch, Keras, NumPy, Pandas, Darknet, Spark, scikit-learn, Matplotlib, Seaborn
- **Techniques**: Statistical Analysis, Predictive Modeling, Data Mining, Data Analytics, Deep Learning, ETL Processes
- **Generative AI & LLMs**: Hugging Face, LangChain, OpenAI API, Prompt Engineering, Large Language Models, NLP Pipelines

PROFESSIONAL EXPERIENCE

SQOR.ai **Sep 2024 - Present**
Data Scientist
- Build and deploy LLM-driven autonomous agents for KPI forecasting, anomaly detection, and real-time risk alerts, improving decision-making speed by 40 percent and driving 15 percent higher operational efficiency across enterprise datasets
- Designed data pipelines integrating 70+ sources to unify 500+ KPIs, reducing manual reporting by 60%
- Developed RAG-based querying and reporting modules, reducing request time by 75% and enabling contextual business insights

Dassault Systèmes **Jun 2024 - Aug 2024**
Data Science Intern
- Reduced downtime and maintenance costs of automotive welding systems by 15%, saving an estimated $200K annually, through a predictive maintenance model using time-series sensor data and failure analysis
- Delivered 99% accuracy by constructing and fine-tuning an XGBoost classification model with SMOTE rebalancing, hyperparameter tuning, and A/B testing
- Transformed raw sensor logs into actionable insights using SQL and Pipeline Pilot for correlation and trend analysis, uncovering hidden failure drivers that boosted predictive model accuracy by 20%
- Built and deployed ETL pipelines in Python and SQL on AWS to ingest, clean, and transform over 5 million sensor records, enabling scalable model training and real-time analytics

Ernst & Young **Sep 2022 - Apr 2023**
Data Science Intern
- Increased vehicle safety by 30% through timely detection of dashboard warnings using an ML-based vision system
- Devised a Python/TensorFlow model using the transformer-based YOLOv7 algorithm to detect warning signal lights, achieving 92% detection accuracy
- Collaborated with cross-functional teams in Agile sprints to integrate the ML model into a mobile app, using AWS EC2 for model hosting, resulting in a 30% improvement in real-time warning detection accuracy.
- Enhanced accuracy and robustness of the image detection model by applying bagging and backbone variations, achieving 98% accuracy while cutting retraining time by 35% and saving $50K in compute resources.

Vellore Institute of Technology **Aug 2021 - Feb 2022**
Research Intern
- Achieved 98.6% accuracy in detecting deepfake media by developing an ensemble machine learning model in Python
- Implemented multi-scale training, optimized hyperparameters, and applied advanced image processing techniques with varied backbone architectures to enhance robustness and reduce false positives
- Constructed a custom dataset of 20,000+ images and 1,000 videos, annotated via Roboflow, which improved the model's classification accuracy for real versus manipulated media

PROJECT

Predicting Trends in Ethical Consumerism **Aug 2024 - Dec 2024**
- Leveraged NLP techniques and Large Language Models (LLMs) to analyze sentiment in Reddit discussions, utilizing APIs and advanced sentiment analysis tools (VADER, TextBlob) to map stock trends
- Built a Python pipeline on NYU HPC to process sentiment data and align it with stock prices via the yfinance API
- Streamlined data collection by managing API rate limits and refining filtering techniques for sentiment-driven stock trends

CERTIFICATIONS

- **Google Data Analytics Professional Certificate: Google**
- **Microsoft Power BI Data Analyst Professional Certificate: Microsoft**