

Yash Yennam

Email: yashyennam@gmail.com | linkedin.com/in/yash-yennam
Phone: +918369235878
Mumbai, India

Data Scientist — Machine Learning Engineer — Data Engineer — Marketing Analyst — Data Analyst

Summary:

- Over 8+ years of experience in Data Science, specializing in Machine Learning (ML), Artificial Intelligence (AI), and Data Engineering.
- Expertise in developing and deploying predictive analytics models, focusing on credit risk assessment and Marketing Analytics.
- Strong knowledge of cloud platforms such as AWS, including hands-on experience with cloud-based data pipelines and automation tools.
- Proficient in leading teams of data scientists and engineers, with significant experience in managing and delivering end-to-end machine learning projects.
- Skilled in implementing scalable ML models using Python, TensorFlow, PyTorch, LightGBM, and XGBoost for various business applications.
- Understanding of Big Data technologies like Hadoop and Spark with experience in handling large datasets using PySpark and SparkSQL.
- Strong analytical and problem-solving capabilities, with experience in advanced algorithms for classification, regression, clustering, and anomaly detection.
- Developed and optimized real-time data pipelines and ETL processes, improving efficiency and scalability across industries like finance and telecom.
- Proven ability to drive business impact through data-driven decision-making, including financial inclusion initiatives for MSMEs and improving operational efficiency.
- Experienced in data visualization and reporting using Power BI, Tableau, and Matplotlib to communicate insights to stakeholders and drive strategic actions.

Technical Skills: Programming Languages: Python, Java, C++
Machine Learning Frameworks: XGBoost, Random Forest, Logistic Regression, NLP, TensorFlow, RNN, Py- Torch, LightGBM, LLM, GenAI, Langchain, Deeplearning
Data Analysis: Pandas, Sklearn, MapReduce, SparkSQL, PySpark, MS Excel
Data Engineering: AWS, Google Cloud, Data Pipelines, ELT, Docker, Kubernetes
Big Data Technologies: Hadoop, Spark, Hive, Pig
Data Visualization: Power BI, Tableau, Matplotlib, Seaborn
Tools: MLFlow, SQL, Tableau, REST APIs, Dataflow, Automation Tools

Education:	
Masters in Science, Machine Learning and Artificial Intelligence Liverpool John Moores University	Jan 2023 - Jan 2024
Postgraduate Degree, Machine Learning and Artificial Intelligence International Institute of Information Technology Bangalore	Oct 2021 - Jan 2023
B.E. Electronics Engineering University of Mumbai	Jul 2013 - May 2018

Certifications:

- Executive Programme in Algorithmic Trading (EPAT) - Quantitative Analysis
- Advanced Python
- Pretraining LLMs - Deeplearning.ai
- Building RESTful Web APIs with Django

Work Experience:

Sr. Data Scientist Trafigura, Mumbai

Sept 2024 - Present

- Founding member of the Digital Transformation Team driving cloud modernization initiatives in the Metals, Oil, and Gas trading sectors by migrating legacy systems to cloud-native platforms including **AWS** and **Azure**.
- **Conducted extensive data analysis** on historical and real-time market data using **Pandas, NumPy, and SQL**, extracting insights to optimize trading models and enhance predictive accuracy.
- Led strategic migration of complex, business-critical **data pipelines** supporting commodity trading, ensuring scalability and system optimization through close **collaboration** with cross-functional stakeholders.
- Applied **business intelligence** best practices to architect cloud-based solutions that improved system efficiency, operational insights, and decision-making capabilities.
- Spearheaded the transformation of legacy **SQL** logic and **QlikView** reports into Python applications, deployed via **GitLab CI/CD to AWS ECS and Lambda**, improving maintainability and runtime performance.
- Used Dremio platform to Optimize and deliver faster ETL pipelines using Dremio's reflection building techniques.
- Leveraged **computer science** principles and clean code standards to build scalable and reliable data solutions aligned with business and technical goals.
- Promoted adoption of **cloud best practices** for infrastructure management, data engineering, and continuous delivery processes.
- Contributed to the long-term **vision** and **strategy** for digital transformation by aligning technical initiatives with core **business objectives** and ROI-focused outcomes.

Tech stack: Python, SQL, MySQL, Postgres, AWS Lambda, AWS Code commit, AWS S3, AWS EC2, Pandas, Numpy, Advance MS Excel, Openpyxl, XlsxWriter, JSON, XML, ETL, Linux, Cronjob. GenAI, Dremio

Sr. Data Scientist - Deputy Vice President U GRO Capital, Mumbai

Jan 2022 - Sept 2024

- Lead a dynamic team of 10 Data Scientists and 2 Data Engineers, driving innovation and excellence in the development of AI/ML models for credit risk assessment and financial analytics.
- Fostered a culture of continuous learning and development within the team, mentoring junior data scientists and engineers to achieve high performance and deliver exceptional results.
- Spearheaded the development of a real-time **credit risk scoring system** using AWS Lambda, enabling sub-minute customer scoring by leveraging data from multiple sources such as credit bureaus, bank statements, and GST returns. Scaled **machine learning models for real-time prediction analysis**, enhancing UGRO Capital's operational efficiency, mathematically express data in risk matrix and enhance customer service capabilities.
- Managed the deployment of machine learning models in production, ensuring rigorous testing, validation, and scalability to meet the needs of financial disbursements and risk management.
- Utilized **Graph Theory** for analyzing the supply chain network of sales and purchases across different sectors, improving **Marketing Analytics** and efficiency and driving supply chain optimization.
- Oversaw the successful disbursement of over **INR 8000+ Cr (USD 1 Billion)** in loans by deploying an internal credit risk scoring ML model forecasting early mortality accounts.
- Designed and maintained data pipelines for internal consumption, enhancing data accessibility and processing efficiency across the organization.
- Developed a first-of-its-kind proprietary credit scoring model based on GST data, revolutionizing financial assessment and risk evaluation in the MSME sector.
- Implemented an entity recognition system using **NLP to extract counterparty names** from customer bank statements, improving data accuracy, Metrics and reliability using GenAI LLM Model.
- Delivered data-driven insights and recommendations to senior management, helping to align data initiatives with business objectives and long-term growth strategies.
- Collaborated with cross-functional teams, including IT, product, Accounting, and sales, to ensure the successful deployment of machine learning models across the organization.
- Managed the end-to-end lifecycle of data science projects, from ideation and prototyping to deployment and performance monitoring, ensuring timely and impactful deliveries.
- Implemented scalable data Schema solutions using **cloud technologies such as AWS**, enabling the company to handle large datasets with ease and improve data processing times.
- Pioneered the development of a GST-based proprietary model for creditworthiness evaluation, setting a new industry standard and contributing to UGRO Capital's leadership in financial innovation.

- Worked closely with business stakeholders to translate complex data problems into actionable solutions, enabling UGRO Capital to achieve significant growth in its Asset Under Management (AUM).
- Played a pivotal role in automating data workflows and processes, utilizing open-source tools and cloud services to streamline operations and reduce manual effort.

Tech stack: Python, SQL, Statistics, MySQL, Postgres, AWS Lambda, AWS Code commit, AWS S3, AWS EC2, Pandas, Numpy, Scikit-learn, Logistics Regression, Q-GIS, NLP, NetworkX, Credit Risk, Advance MS Excel, Openpyxl, XlsxWriter, JSON, XML, ETL, Linux, Cronjob, GenAI.

Software Development Engineer - Python Data Engineer
Bank of America, Mumbai

Jul 2021 - Jan 2022

- Developed **ETL pipelines** for market risk analysis, ensuring robust risk assessments across multiple business lines, enhancing the efficiency of risk management processes.
- Collaborated with the Global Agile team to develop and deploy internal software solutions for market risk and data modeling, ensuring effective and timely delivery.
- Built and maintained data pipelines for a risk monitoring system, providing risk managers with real-time insights to make informed market risk decisions.
- Designed market risk software that integrated feeds from multiple lines of business (credit, trade, currency, equity, mortgage) to generate comprehensive risk assessments.
- Utilized advanced analytics tools like **Pandas and Numpy** for data manipulation, enabling efficient statistical analysis and improving risk prediction accuracy.
- Optimized and Validate ETL processes to reduce data processing time allowing quicker response to market changes and improving operational efficiency.
- Automated **data ingestion workflows**, Roadmap,minimizing manual intervention and ensuring continuous data availability for critical risk management tasks.
- Conducted in-depth analysis to identify and resolve data discrepancies, improving overall data accuracy, integrity, and consistency across business lines.
- Played a key role in **Product Analytics and lifecycle management**, ensuring that market risk software was designed, tested, and deployed according to business and compliance requirements.
- Worked closely with **risk management teams** to develop risk monitoring **dashboards**, providing visual insights and facilitating better decision-making across departments.

Tech stack: Python, SQL, MySQL, Postgres, Pandas, Numpy, Credit Risk, Market Risk, Advance MS Excel, Openpyxl, XlsxWriter, JSON, XML, ETL, Linux, Cronjob

Senior Software Development Engineer - Python Spatial Data Engineer
Reliance Jio, Mumbai

Jan 2019 - Jul 2021

- Conducted **geographic spatial data analysis** using **arcpy and Geo-Spark**, improving telecom network integration and optimization across multiple regions.
- Developed automated **ETL data processing pipelines** using Windows Task Scheduler and Python, generating actionable insights and delivering reports via automated email notifications.
- Spearheaded the migration and modernization of the **data warehouse infrastructure**, implementing **ArcGIS Server** for optimized data processing and real-time data flow monitoring.
- Built robust REST APIs using Django REST Framework and Oracle Database, providing seamless data integration for web and mobile platforms, increasing system efficiency by 30%.
- Collaborated with a research team to design and implement machine learning models for real-time data processing, significantly enhancing application performance and predictive outcomes.
- Leveraged Big Data technologies such as **Apache Spark and Geo-Spark** to optimize telecom networks Use Cases, improving overall network performance and reducing operational costs.
- Led cross-functional teams in implementing scalable solutions, ensuring the successful integration of data analytics modules into existing workflows and platforms.
- Played a pivotal role in building a high-performance REST API architecture, improving data accessibility and allowing real-time data consumption by web and mobile applications.
- Developed and deployed **Python-based data analytics modules**, enabling seamless data integration and efficient consumption by client-facing applications, ensuring reliable and consistent data access.
- Recognized for exceptional performance and innovative solutions in delivering high-quality, scalable data engineering projects, driving significant improvements in system performance and usability.
- **Automated several manual processes** by designing custom Python scripts, coding, reducing operational workloads and improving OKRs, enabling the team to focus on more critical tasks.

- Conducted in-depth **spatial data analysis for network expansion** projects, contributing to strategic decisions and improved geographic coverage for the telecom network.
- Experiment & Improved the accuracy and reliability of geographic data insights through the development of robust data validation and verification procedures.

Tech stack: Python, SQL, MySQL, Postgres, Oracle, BASIC, Pandas, Numpy, Scikit-learn, R, Logistics Regression, Q-GIS, ArcGIS, NLP, Advance MS Excel, Openpyxl, XlsxWriter, JSON, XML, Django, Flask, REST APIs, ETL, Linux, Cronjob.

Projects:

Marketing Analytics at UGRO Capital

Jan 2022 - Aug 2024

- Developed a **cross-sell lead generation** engine for the call center by analyzing existing customers' CIBIL reports to identify cross-sell opportunities — such as pitching secured loan balance transfers to unsecured loan holders and vice versa — improving targeting precision for outbound campaigns.
- Built a counterparty **profiling pipeline leveraging customers'** GST data to extract counterparty PAN and GSTIN numbers, followed by automated CIBIL scrapes to generate enriched customer profiles and qualify high-intent leads for the call center.
- Designed and deployed **propensity models** using bank statement analysis and CIBIL scores to predict an individual's likelihood of taking a loan, enabling the call center to prioritize outreach toward high-propensity prospects.
- Improved call center conversion efficiency by ~18.3% by delivering pre-qualified, data-enriched leads backed by predictive scoring, reducing time spent on low-intent prospects and increasing the overall quality of the sales pipeline.
- Integrated analytics-driven lead generation with fintech data sources including GST filings, bureau data, and transaction history to build a scalable, end-to-end marketing intelligence system supporting both cross-sell and new customer acquisition strategies.

Lead Developer for GRO Score Model at UGRO Capital

Jan 2022 - Aug 2024

- Spearheaded the development of GRO Score, an **AI/ML-driven credit scoring system**, analyzing over 25,000 data features from credit bureau records, bank statements, and GST data.
- Improved risk assessment processes, significantly enhancing loan disbursement efficiency and accuracy for MSME borrowers.
- Impacted over **80,000 customers across India** and drove growth in Asset Under Management (AUM) by over **INR 8000+ Cr (USD 1 Billion)** in loans, supporting financial inclusion.
- Collaborated with cross-functional teams to scale the model, integrating it with multiple banking partners and optimizing **financial operations. ex. Mastercard**

Official [Link](#) of Mastercard collaborating Gro Score Model using GST Data.

Freelance - Python for Finance - Technical Reviewer, BPB Publications

Feb 2024 - Apr 2024

- Reviewed and provided technical insights for "Python for Finance," by Dmytro Zherlitsyn, published by BPB Publications (Asia's biggest technical textbook Publisher) covering **time-series analysis, portfolio management, and Market risk assessment**.
- Ensured accuracy and practical implementation of Python in financial Data models, contributing to high-quality content for the book.
- Evaluated financial analysis methodologies using Python to bridge the gap between **finance** and data science.
- Contributed to the development of resources empowering professionals to leverage Python for financial decision-making and risk evaluation.

Amazon [Link](#) to the Text Book, Official [Link](#) of the Publisher

Individual Contributor for Jio-bp (British Petroleum) EV Charging Station Location Analysis and Deployment Jan 2021 - Jun 2021

- Led the analysis of geographic dataset using **GIS** to identify optimal locations for EV charging stations across India.
- Developed a location identification service to support strategic EV infrastructure deployment for Jio-bp, maximizing accessibility.
- Leveraged advanced GIS analytics and spatial data management to optimize site selection, ensuring comprehensive coverage.
- Contributed to the expansion of Jio-bp's EV network, supporting sustainability initiatives and low-carbon mobility solutions.

Official Media Article for Jio-BP EV [Link](#)