

Summary

- Data Scientist with 8 years of experience building **production-grade machine learning and forecasting solutions** across healthcare and financial domains, delivering measurable business impact on large-scale datasets. Strong expertise in Python, SQL, and Snowflake, with a proven track record of processing multi-million record datasets and improving decision-making through data-driven insights.
- Experienced across the **full data science lifecycle**, including problem framing, feature engineering, model development, evaluation, and monitoring, with a focus on delivering scalable and reliable analytical solutions. Proficient in the Python ecosystem (Pandas, NumPy, SciPy, Dask, Scikit-learn) for large-scale data processing, statistical modeling, and predictive analytics.
- Specialized in working with **claims, EOBs, enrollment, and financial datasets**, performing complex multi-source analysis, data reconciliation, and large-scale SQL transformations. Enabled consistent and accurate analytical outputs by integrating diverse data sources and improving data quality across pipelines.
- Strong background in **model evaluation and experimentation**, applying metrics such as RMSE, MAE, precision, recall, and F1-score along with cross-validation and hyperparameter tuning. Ensured robust model performance, reduced overfitting, and improved reliability across real-world business scenarios.
- Hands-on expertise in **time series forecasting, classification, regression, and clustering techniques**, selecting appropriate modeling approaches based on data characteristics and business requirements. Delivered consistent improvements in prediction accuracy and operational efficiency across multiple use cases.
- Aligned with modern Data Science advancements, applying a **data-centric approach to improve data quality, feature engineering, and model robustness**, rather than relying solely on model complexity. Incorporated experimentation-driven workflows and statistical rigor to enhance model performance and generalization.
- Experienced in translating business problems into **decision science frameworks**, defining KPIs, evaluating trade-offs, and ensuring models drive measurable outcomes. Emphasizes interpretability, explainability, and stakeholder communication to ensure models are trusted and actionable in production environments.
- Demonstrates working knowledge of emerging **Generative AI capabilities**, including foundational understanding of Large Language Models (LLMs), embeddings, and NLP evolution, with the ability to evaluate their applicability for analytical and decision-support use cases. Maintains a strong focus on traditional machine learning and statistical rigor while staying aligned with advancements in AI-driven systems.
- Proven ability to build **scalable and sustainable data science solutions**, incorporating model monitoring, retraining strategies, and performance tracking over time. Optimized models and data workflows to achieve **20–40% improvements in performance, accuracy, and efficiency** across healthcare and financial applications.

Education

Master of Professional Studies, Data Sciences and Applications

01/2022 - 06/2023

University at Buffalo, The State University of New York (GPA: 3.967/4.0)

Relevant Coursework: Machine Learning, Probability & Data Analysis, Statistical Data Mining, Numerical Analysis, Data Structures & Algorithms, Database Management Systems, Cybersecurity, Python Programming

Experience

Data Scientist, CareHigh, Virginia, USA

09/2023 - Present

- Led the design and implementation of a **scalable predictive analytics platform** for healthcare financial forecasting, processing **multi-million record datasets** across claims, EOBs, and clinical systems. Delivered accurate monthly projections across hospitals, improving planning accuracy by **30%** and enabling data-driven decision-making aligned with financial and operational KPIs.
- Developed robust **end-to-end data pipelines using Python and SQL**, handling ingestion, cleansing, transformation, and feature preparation from heterogeneous healthcare sources including EHR/EMR, insurance, and financial systems. Reduced manual preprocessing effort by **40%** while improving data consistency and supporting scalable modeling and forecasting workflows.
- Framed complex healthcare business problems into **data science solutions**, defining modeling objectives, success metrics, and KPIs aligned with financial and operational outcomes. Ensured modeling approaches directly supported decision-making and measurable business impact across multiple hospitals.

- Engineered advanced **feature sets from structured and semi-structured healthcare data**, including temporal utilization patterns, treatment trends, and financial indicators. Improved model stability and predictive performance by **25%**, enabling more accurate forecasting and stronger downstream analytics.
- Built and validated **supervised and unsupervised machine learning models (classification, regression, clustering)** using Scikit-learn, applying cross-validation, hyperparameter tuning, and evaluation metrics. Achieved **20–30% improvement in prediction accuracy** across forecasting and risk modeling use cases.
- Designed a **probabilistic forecasting framework** combining event frequency and financial severity modeling to generate scenario-based projections including baseline, optimistic, and conservative cases. Improved forecast robustness and enabled stakeholders to evaluate trade-offs and plan budgets effectively.
- Implemented **time series modeling strategies** incorporating historical trends, seasonal variations, and lag-based features to enhance temporal consistency. Improved prediction reliability across rolling monthly forecasts and reduced variance in model outputs across time periods.
- Performed comprehensive **data-centric preprocessing and anomaly handling**, including outlier detection, normalization, and missing value imputation on large healthcare datasets. Improved data quality and reduced noise, enhancing model generalization performance by **27%**.
- Built scalable **data validation and reconciliation frameworks** to compare transformed datasets against source systems and ensure data consistency across pipelines. Reduced discrepancies by **30%**, improving trust in analytics outputs and enabling reliable reporting.
- Optimized and executed **complex SQL queries (joins, aggregations, CTEs, window functions)** across large healthcare datasets including claims, enrollments, and provider records. Improved query performance by **35%**, enabling faster data access for modeling, analytics, and reporting workflows.
- Integrated datasets into **Snowflake data warehouse**, enabling efficient storage, transformation, and querying of large-scale healthcare data. Reduced data retrieval latency by **25%** and supported scalable analytics across multiple teams and decision-making use cases.
- Developed **automated reporting pipelines and dashboards**, generating monthly summaries, trend analyses, and performance insights for stakeholders. Reduced manual reporting effort by **40%** while improving visibility into key business metrics and supporting decision science workflows.
- Translated model outputs into **actionable business insights**, supporting financial planning, operational optimization, and program management decisions. Improved stakeholder adoption of data-driven strategies by **25%** through clear communication and interpretability of results.
- Designed the solution as a **modular and reusable ML pipeline architecture**, enabling scalability, maintainability, and integration into enterprise workflows. Reduced redevelopment effort and improved reusability for future data science and forecasting applications.
- Prepared pipelines for **production readiness**, incorporating experiment tracking, version control, and reproducible workflows to support consistent deployment. Ensured alignment with model lifecycle best practices and enabled smooth transition into production environments.
- Established **model monitoring and lifecycle management strategies**, tracking key performance metrics, identifying model drift, and implementing periodic retraining. Ensured long-term model stability and alignment with evolving data distributions and business conditions.
- Applied **model explainability techniques**, including feature importance analysis and interpretability methods, to ensure transparency and trust in predictions. Enabled stakeholders to understand key drivers influencing model outputs in a regulated healthcare environment.
- Leveraged the **Python ecosystem (Pandas, NumPy, SciPy, Dask, Scikit-learn)** for large-scale data processing, statistical modeling, and machine learning development. Enabled efficient handling of high-volume datasets and improved computational performance across workflows.
- Applied **NLP techniques using spaCy and NLTK** to process and standardize unstructured healthcare text fields such as clinical notes and descriptions. Improved data usability and consistency by **20%**, enabling more comprehensive analytics and feature extraction.
- Evaluated model performance using **RMSE, MAE, precision, recall, and F1-score**, ensuring alignment with healthcare business objectives and forecasting accuracy requirements. Compared multiple models and improved selection consistency across multi-year datasets.
- Conducted structured **experimentation workflows including baseline modeling, cross-validation, and hyperparameter tuning**, improving model robustness and reducing overfitting. Established repeatable experimentation practices aligned with modern data science standards.

- Maintained awareness of **emerging Generative AI capabilities**, including Large Language Models (LLMs) and embeddings, evaluating their applicability for healthcare analytics. Ensured alignment with evolving AI advancements while maintaining strong statistical and modeling rigor.
- Supported **model deployment and integration workflows by exposing prediction outputs into reporting pipelines and downstream systems**, ensuring models were consumable by business stakeholders. Established periodic retraining and monitoring strategies to maintain model performance and prevent degradation over time.

Data Scientist, Accenture Solutions Pvt Ltd, Bangalore, India

05/2017 - 12/2021

- Led data science initiatives supporting Procure-to-Pay and Order-to-Cash financial workflows by developing analytical solutions on **multi-million record transactional datasets** across procurement, invoicing, and payment processes. Improved reporting turnaround time by **30%** and enabled data-driven decision-making aligned with financial KPIs and operational objectives.
- Developed and maintained large-scale **SQL Server-based analytical data models**, restructuring normalized and denormalized schemas to support efficient querying, reporting, and downstream analytics. Improved accessibility and reliability of financial data across multiple systems, enabling consistent analysis and supporting business-critical reporting needs.
- Built structured **data preparation and transformation workflows using Python (Pandas, NumPy)** to process high-volume financial transactions and accounting records for analytics and modeling use cases. Reduced manual data preparation effort by **35%** while improving data quality, consistency, and readiness for downstream statistical analysis.
- Performed detailed **exploratory data analysis (EDA)** on procurement, invoicing, and payment datasets to identify trends, anomalies, seasonality patterns, and operational bottlenecks across financial workflows. Improved business understanding of data behavior and informed feature engineering, model selection, and reporting strategies for downstream analytical solutions.
- Translated financial business requirements into **data science use cases**, including payment delay prediction and transaction risk scoring, by defining modeling objectives, success criteria, and evaluation metrics aligned with KPIs. Ensured analytical solutions addressed real business problems and supported measurable improvements in financial decision-making outcomes.
- Developed **machine learning models (classification, regression, clustering)** using Scikit-learn to support financial forecasting, segmentation, and risk analysis across structured transactional datasets. Improved prediction accuracy by **25%** and enabled early adoption of predictive analytics to support operational planning and financial process optimization.
- Performed advanced **feature engineering on structured financial datasets**, incorporating transaction timing, vendor attributes, payment history, and behavioral patterns to strengthen analytical and predictive use cases. Improved model stability, interpretability, and alignment with real-world financial decisions by enhancing the quality and relevance of predictive inputs.
- Applied **time series analysis techniques** to study payment trends, invoice cycles, and financial flows using historical patterns and temporal dependencies within transaction data. Supported forecasting use cases and improved consistency in financial trend analysis, enabling better planning for cash flow and reporting cycles.
- Optimized and executed **complex SQL queries (joins, aggregations, CTEs, window functions)** across multiple financial systems to support analytics, reporting, and structured modeling workflows. Improved query performance by **30%**, enabling faster data extraction and more reliable access to business-critical financial information at scale.
- Applied **statistical analysis using SciPy and NumPy** to identify trends, correlations, and anomalies in financial datasets across vendor payments, procurement records, and transaction histories. Delivered actionable insights that improved operational efficiency, reduced inconsistencies, and strengthened confidence in reporting and analytical outputs.
- Developed **unsupervised learning models (clustering and segmentation)** to categorize vendors and transaction behaviors based on historical payment patterns, financial characteristics, and operational attributes. Improved segmentation accuracy by **20%** and enabled more targeted financial analysis, exception handling, and reporting strategies across business workflows.
- Implemented **data validation and reconciliation processes** to ensure consistency between transformed datasets and source systems across procurement, vendor, and payment records within financial operations. Reduced discrepancies by **25%**, strengthening trust in analytics outputs and improving reliability of downstream reporting and modeling processes.
- Integrated datasets into **cloud-based data warehouses including Snowflake**, enabling scalable storage, transformation, and efficient querying for financial analytics and business intelligence use cases. Improved data availability and reduced latency for high-volume analysis, supporting broader access to financial insights across teams and stakeholders.
- Performed **multi-source data integration and large dataset joins**, combining transactional, vendor, and financial records across enterprise systems to build unified analytical views of financial processes. Enabled comprehensive reporting and improved visibility into end-to-end Procure-to-Pay and Order-to-Cash workflows for operational and management teams.
- Utilized **parallel processing techniques (Dask)** to efficiently handle large datasets beyond memory constraints and support higher-volume analytical processing across financial records and reporting workloads. Improved data processing performance by **30%**, enabling more scalable computation for large transaction-driven datasets and repeated analytical tasks.

- Automated repetitive workflows using **Python scripting**, reducing manual effort by **20%** across recurring data processing, reconciliation, and reporting activities within financial analytics operations. Improved consistency, efficiency, and turnaround time for analytical deliverables while reducing reliance on manual interventions across routine reporting cycles.
- Designed and executed **unit and integration tests using PyTest**, ensuring correctness of data transformations, analytical calculations, and model outputs across structured financial data science workflows. Improved reliability and stability of analytical systems supporting reporting and modeling, while reducing defects introduced during ongoing enhancements and releases.
- Developed **data visualization and reporting solutions**, creating dashboards, trend summaries, and analytical views to communicate financial insights clearly to finance teams and business stakeholders. Improved accessibility of insights and strengthened data-driven decision-making by presenting complex analytical findings in a concise and actionable format.
- Collaborated with cross-functional teams including finance stakeholders, analysts, and engineering partners to align **data science solutions with business requirements, KPIs, and reporting priorities**. Improved adoption of analytics-driven decisions and strengthened collaboration by translating technical outputs into clear business value and measurable operational impact.
- Evaluated model performance using **RMSE, precision, recall, and F1-score** to ensure predictive models aligned with financial objectives and real-world analytical use cases across transaction and risk datasets. Improved consistency in model evaluation and selection by comparing multiple approaches against defined business and technical success metrics.
- Conducted structured **experimentation workflows including baseline modeling, cross-validation, and hyperparameter tuning** to improve robustness, reduce overfitting, and strengthen reliability of financial analytics models. Established repeatable experimentation practices that supported more disciplined model development and more consistent performance across multiple business scenarios.
- Applied **basic model monitoring and periodic evaluation practices** to review predictive performance over time and ensure models remained aligned with changing financial data patterns and operational behavior. Supported early-stage model lifecycle management within analytics workflows by identifying performance shifts and informing appropriate refinement decisions.
- Applied **model explainability techniques**, including feature importance analysis and interpretation of predictive drivers, to improve transparency of model outputs and strengthen stakeholder confidence in analytical results. Enabled business users to better understand factors influencing financial predictions, improving trust and adoption of data science recommendations.

Skills

- **Programming Languages:** Python, SQL
- **Machine Learning & Data Science:** Supervised Learning, Unsupervised Learning, Predictive Modeling, Classification, Regression, Clustering, Ensemble Methods (Random Forest, Gradient Boosting), Time Series Forecasting, Feature Engineering, Model Validation, Hyperparameter Optimization, Anomaly Detection
- **Python Libraries:** Pandas, NumPy, SciPy, Scikit-learn, Dask
- **Model Evaluation & Experimentation:** Cross-Validation, RMSE, MAE, Precision, Recall, F1-score, Model Comparison, Bias-Variance Tradeoff, Experiment Design
- **Statistical Analysis:** Hypothesis Testing, Statistical Inference, Probability Distributions, Regression Analysis, Variance Analysis
- **Data Analysis & Processing:** Exploratory Data Analysis (EDA), Data Cleaning, Data Transformation, Missing Value Handling, Outlier Detection, Data Profiling
- **Data Systems & Querying:** Relational Databases, SQL Queries, Joins, CTEs, Window Functions, Aggregations
- **Data Platforms:** Snowflake, Data Warehousing
- **Cloud Environment:** AWS (S3, Lambda, EC2), Azure (ADF, ADLS, Azure SQL), Google Cloud Platform (GCP)
- **Model Deployment & Lifecycle:** Model Integration, Batch Scoring, Model Monitoring, Performance Tracking, Model Retraining
- **Visualization & Communication:** Data Visualization, Insight Generation, Business Reporting, KPI Analysis, Decision Support
- **Tools:** Git, JIRA
- **Soft Skills:** Analytical Thinking, Problem Solving, Stakeholder Communication, Collaboration, Adaptability

Certifications

- Microsoft Certified: Azure Data Scientist Associate
- IBM Data Science Professional Certificate
- Google: Data Analytics Capstone
- HackerRank: Advanced SQL