AKHILESH CHOUDHARY

Yorba Linda Blvd, Fullerton, Apt G07, CA, 92831

657-799-0115 | akhileshchoudary1994@gmail.com | LinkedIn | Portfolio Website

PROFESSIONAL SUMMARY

- **Data Scientist & Engineer** with **5+ years** of experience in data analytics, modeling, and engineering, with a focus on machine learning, statistical analysis, predictive modeling, and systems programming to improve patient outcomes and operational efficiency.
- Proficient in **Python**, **SQL**, C**++,** and **AWS**, with a strong track record of building **ETL pipelines**, processing large-scale datasets, and optimizing performance using object-oriented design principles.
- **Optimized data pipeline** performance and scalability using object-oriented design in C++, while implementing **CI/CD** practices with Git and automation tools for seamless, low-downtime deployments.
- Well-versed in medical record systems (EHR, EMR), HL7 standards, HIPAA compliance, and core computer science concepts including **data structures**, **algorithms**, and **object-oriented design**.
- Advanced knowledge of **statistical techniques** such as regression, hypothesis testing, and forecasting, with a focus on applying **machine learning** to streamline clinical and operational workflows.
- Strong communicator with proven success in **collaborating with cross-functional teams**, healthcare professionals, and IT stakeholders to deliver tailored analytical solutions and drive continuous improvement.

TECHNICAL SKILLS

Programming	Python, R, SQL, C++
Data Science Libraries	NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow
Machine Learning	Regression Analysis, Bayesian Methods, Decision Trees, Random Forests, Support Vector Machines, Neural Networks, K-Means, KNN, Reinforcement Learning, Naive Bayes, NLP, LLM, CNN, XGBoost, Deep Learning, Predictive Modeling, Statistical Modeling
Data Engineering	Data Warehousing, ETL (Informatica, Talend), Data Integration, Data Modeling (ER Diagrams, Dimensional Modeling, Star Schema, Data Vault), Object-Oriented Design
Technical Skills	Object-Oriented Design, Systems Programming, Linux (explored), Compiler Optimization, Data Structures, Algorithms, CI/CD Pipelines
Data Warehousing	Snowflake, Google Big Query, Apache Hive
Cloud Technologies	AWS (EC2, RDS, S3, Lambda, Redshift, Athena, Glue), Azure (Cloud Services, Databricks), Google Big Query
Big Data Tools	Apache Hive, Pyspark, Snowflake
Visualization Tools	Tableau, Power BI, QuickSight
Development Tools	Git, CI/CD Pipelines
Project Management	Agile/Scrum, Kanban, Waterfall, Jira, Confluence
Other	SAS, Advanced Excel (VBA, Pivot Tables), Healthcare Standards (HIPAA, EHR, HL7)

EDUCATION

Master's in **Information Systems – Business Analytics** **| Jan 2022 - Jan 2024**
California State University, Fullerton, CA

Bachelor of Technology **| Jun 2012 - Aug 2016**
R.V.R & J.C College, INDIA

EXPERIENCE

Cigna Healthcare Group, USA
Data Engineer *| Feb 2024 – Present*

- Developed and automated ETL pipelines using Python and SQL to integrate diverse healthcare data sources into a centralized SQL Server database, ensuring data consistency and synchronization.
- Utilized object-oriented design principles and systems programming concepts with C++ to optimize data processing scripts, enhancing performance and scalability of data pipelines.
- Established continuous integration and deployment (CI/CD) practices for data pipelines using Git and automation tools, ensuring seamless updates and maintenance with minimal downtime.
- Engineered key features from integrated datasets to enhance predictive model performance, deploying models within the claims processing workflow for actionable insights.
- Applied statistical tools like SAS and R to develop machine learning models (e.g., regression, gradient boosting machines),

predicting claim processing times and identifying error-prone claims, contributing to a 30% reduction in processing times and a 20% decrease in errors.

- Conducted exploratory data analysis (EDA) using Python with Pandas, NumPy, and Seaborn to uncover insights into operational efficiencies and device performance, informing data-driven optimization strategies.
- Designed and optimized SQL Server databases to streamline data collection, querying, and reporting processes, ensuring data integrity and supporting the development of economic models and cost calculators for health economics analysis.
- Developed and maintained interactive dashboards using Power BI to effectively communicate complex healthcare information, track key performance indicators (KPIs), and monitor healthcare quality metrics.
- Ensured compliance with healthcare regulations (HIPAA, HL7) during the implementation and integration of EMR software, maintaining data privacy and security standards.
- Collaborated with cross-functional teams, including data scientists, IT specialists, and clinical staff, to deliver tailored analytical solutions and support continuous improvement initiatives.

ValueLabs, India
Data Engineer *| Mar 2019 – Dec 2021*

- Led a cross-functional team in applying Agile methodology to ensure efficient project execution, collaboration, and iterative delivery of healthcare data solutions.
- Optimized Python scripts to automate data extraction, cleaning, and transformation, reducing manual effort by 25% and increasing operational efficiency.
- Utilized SAS for advanced statistical analysis and predictive modeling of patient outcomes, providing accurate forecasts to inform strategic decisions related to patient care and resource allocation.
- Performed data mapping and logical data modeling, created diagrams, and utilized SQL queries to filter and analyze data, enhancing data accuracy and operational efficiency.
- Delivered Proof of Concepts (POCs) using AWS services to demonstrate innovative visualization solutions tailored to business requirements and objectives.
- Processed charge postings in EHR systems, handling billing for prescriptions and medical visits using ICD-10 and CPT codes for cancer patients.
- Developed and maintained data pipelines using ETL tools like Informatica and Talend to ensure seamless data flow from source systems to analytical platforms.
- Implemented data warehousing solutions using Snowflake and Google Big Query to store and manage large-scale healthcare datasets, enabling efficient querying and reporting.
- Applied object-oriented programming principles with C++ to develop efficient data processing modules, optimizing performance-critical components of the data pipeline.
- Established CI/CD pipelines using Git and automation tools to streamline deployment and testing of data engineering workflows, ensuring reliability in production environments.

Hexaware, India
Data Scientist *| Jan 2018 – Feb 2019*

- Conducted data analysis and profiling using SQL on Oracle and Teradata databases to identify trends and patterns for business intelligence solutions.
- Developed predictive models using Python (e.g., scikit-learn, NumPy) to forecast business outcomes, improving decision-making accuracy by 10%.
- Designed and deployed interactive Tableau dashboards to visualize key metrics, incorporating Custom SQL queries from multiple data sources for comprehensive insights.
- Collaborated with stakeholders to gather requirements and create Source-to-Target Mapping documents, ensuring alignment of data transformations with business needs.
- Leveraged ETL tools (e.g., Informatica, Talend) to extract and load data into AWS Redshift, enabling scalable analytics and reporting.
- Integrated Python scripts with Tableau for advanced statistical analysis, enhancing visualization capabilities and user engagement.
- Implemented CI/CD pipelines to streamline deployment of data models and dashboards, reducing delivery time by 15%.
- Leveraged cloud-based data warehousing solutions such as AWS Redshift and Google Big Query to store and analyze large datasets, enabling scalable and cost-effective data processing.
- Implemented data pipeline automation using CI/CD tools and Git to streamline the deployment of data models and visualizations, reducing time-to-market for new features and updates.
- Collaborated with data engineers and developers to optimize data workflows and ensure seamless integration of data sources into visualization platforms.