

KEERTHI PONNAM

Data Analyst

Email: ponnamkeerthi7743@gmail.com | Phone: +1(470)238-9818 | Location: GA

SUMMARY

- Accomplished as a Data Analyst having 3+ years of experience in Python, R, and SQL for automating workflows in building ETL pipelines, and extracting data using Flask/Django for backend scripting.
- Engineered ETL workflows (Informatica, Apache Nifi) to process datasets, reducing latency by 2x via Hadoop, Spark, Kafka optimizations across PostgreSQL, MySQL, and NoSQL databases.
- Proficient in migrating legacy systems to AWS, deploying scalable pipelines and integrating CNN/LSTM models forecasting.
- Expertise in handling databases with SQL Server, Oracle, MongoDB, Cassandra, Snowflake and ML like CNN, LSTM.

SKILLS

• Programming & Scripting:	Python, R, SQL, Shell Scripting
• Version Control Tools:	Git, GitHub
• Methodologies:	Agile, Waterfall
• Big Data & Machine Learning:	Hadoop, Spark, Kafka, CNN, LSTM
• ETL Tools:	Informatica, Apache Nifi, ETL Automation
• Data Visualization:	Power BI, Tableau, Microsoft Excel, QuickSight
• Data Analysis:	Data Mining, Reporting, A/B Testing, Data Warehousing
• Databases:	SQL Server, Oracle, MongoDB, NoSQL, Cassandra, Snowflake
• Libraries & Frameworks:	NumPy, Pandas, SciPy, PySpark, Matplotlib, Flask, Django, Bootstrap
• Cloud Platforms:	AWS (Kinesis, Lambda, Athena, SageMaker, DynamoDB), Azure (Data Factory, Databricks)

PROFESSIONAL EXPERIENCE

BlueCrossBlueShield	Aug2024-Current
Data Analyst	GA, USA
- Harnessed Python to analyze electronic health records (EHR) and claims data, uncovering workflow inefficiencies that reduced weekly claim processing time by 8 hours, boosting operational efficiency in healthcare settings by 15%. - Developed Power BI dashboards integrating data, billing platforms, and other healthcare sources, providing real-time insights into key performance indicators such as wait times and emergency room overcrowding. - Integrated ERP data from multiple sources into Power BI dashboards, providing insights into resource utilization and financial performance, enhancing cost optimization by 15%. - Developed SQL-based ETL pipelines to extract and transform data from ERP systems, improving decision-making for operations and reducing manual data entry errors by 30%. - Integrated MongoDB to store and manage unstructured data, elevating data accessibility and supporting analytics. - Applied data mining techniques and CNN models to extract actionable patterns from 1000+ medical histories and imaging records, enabling predictive health strategies and improving patient outcomes through targeted interventions. - Implemented data governance policies ensuring compliance with HIPAA and healthcare regulations, reducing data privacy violations by 40%. - Designed data security frameworks using AWS security tools, ensuring encrypted access control for sensitive patient records and enhancing audit compliance. - Employed AWS Kinesis to automate real-time ETL workflows, streaming and processing healthcare data to decrease manual data processing time from 12 hours to 4 hours per week, improving throughput efficiency by 66%. - Collaborated within an Agile framework to deliver healthcare analytics solutions, using Git for version control to maintain project integrity and streamline team coordination. - Designed KPI dashboards in Tableau and Power BI to track patient wait times, resource utilization, and claim processing efficiency, leading to a 20% improvement in service delivery. - Developed custom KPI models for claims processing performance, helping identify inefficiencies and reducing claim settlement time by 25%. - Ensured HIPAA compliance by securing over 250 reports and claims data within AWS storage systems, implementing encryption and privacy safeguards to meet regulatory standards and protect sensitive healthcare information.	

- Conducted an in-depth research project titled "Predicting Autistic Spectrum Disorder using WEKA Modeler", analyzing 200+ data points to develop a machine learning model for early ASD detection.
- Employed the WEKA platform to build and optimize predictive models, achieving 85%+ accuracy by integrating data preprocessing, feature selection, and advanced classification techniques on ASD-related datasets.
- Utilized Python, R, and SQL for exploratory data analysis (EDA) and feature engineering, improving the accuracy of predictive healthcare models by 18%.
- Conducted A/B testing and hypothesis analysis using statistical modeling tools, providing data-driven insights to improve patient outcomes.
- Investigated a dataset encompassing over 150 attributes and 500 instances, drawing from diverse sources such as demographic profiles, behavioral observations, and medical histories to uncover meaningful patterns.
- Tested and evaluated machine learning classifiers, including Random Forest, Naïve Bayes, & Random Tree, achieving an accuracy of up to 72% in predicting ASD based on behavioral and demographic data samples.
- Investigated public health implications of ASD prediction models, demonstrating a improvement in early detection rates, which could lead to a 25% increase in timely interventions and improved patient outcomes.

Cipla**Dec2021 -Jul2022**

Data Analyst

India

- Created advanced SQL queries to manage and transform large-scale healthcare datasets, enabling analysis of hospital archives and supporting clinical decisions with optimized performance, minimizing execution time by 15 seconds.
- Conducted statistical modeling in R, applying predictive analytics to assess patient outcomes and optimize resource allocation, improving decision-making for healthcare providers.
- Migrated over 1TB of patient records and clinical data from on-premise databases to AWS Redshift and S3, using AWS Glue to automate data transformations. This reduced query execution time from 45 seconds to under 10 seconds, allowing faster access to insights for doctors and analysts.
- Implemented data-driven supply chain analytics using Power BI to monitor procurement, inventory levels, and supplier performance, leading to a 12% reduction in operational costs.
- Optimized supply chain workflows by integrating real-time ETL processes using Apache NiFi, improving efficiency in drug distribution and reducing inventory discrepancies by 20%.
- Applied the Waterfall methodology to structure and execute 3 healthcare analytics projects, delivering 6 well-defined phases (requirements gathering, design, deployment) with milestones and alignment across 2 stakeholders per project.
- Designed and automated data ingestion workflows using Apache NiFi, streamlining the integration of real-time patient records from sources, reducing data processing latency by 30%.
- Migrated 1TB of healthcare data to Snowflake, enhancing scalability and query performance, enabling seamless data access for analytics users and generating 8 reports monthly.
- Leveraged Matplotlib to create data-driven visual reports, illustrating trends in patient admissions, treatment success rates, and hospital performance metrics for healthcare stakeholders.
- Conducted A/B testing on healthcare interventions, measuring treatment effectiveness across patient cohorts, leading to a 5% improvement in recovery rates based on data-driven recommendations.
- Employed Databricks to execute large-scale healthcare analytics, processing millions of patient records, uncovering insights optimizing hospital operations and patient care strategies.
- Executed 10 data integration workflows using Informatica, consolidating 2 disparate healthcare data sources into unified datasets with 2 million structured reports to support cost analysis and population health insights.

Cipla**Jun2021-Nov2021**

Data Analyst Intern

India

- Generated ETL pipelines to automate data integration between clinical trial databases and healthcare supply chain systems, saving 4 hours of manual processing per dataset and ensuring seamless data availability for care analysis.
- Applied MySQL queries to analyze anonymized archives, uncovering drug efficacy trends to guide healthcare teams in refining treatment protocols across medical programs.
- Designed and developed BI dashboards using Tableau and Power BI, enabling executives to analyze hospital revenue trends and optimize resource allocation.
- Automated data extraction from ERP and hospital management systems into BI reporting tools, reducing manual reporting efforts by 60%.
- Integrated DynamoDB to store and manage structured and semi-structured healthcare records, reducing query latency by 50% and enabling real-time data retrieval for high-volume transactional data.
- Enforced 4 Oracle databases to store and organize healthcare information, ensuring high availability and integrity of data entries for 12 critical downstream analytics processes and 5 compliance reports annually.

- Optimized MySQL queries to enhance reporting efficiency, improving data clarity and accessibility for at least 6 stakeholders, including healthcare regulators and hospital supply chain coordinators.
- Devised Excel-based reporting workflows to deliver precise and timely insights for healthcare partners, ensuring compliance with monthly audits for safety and operational standards.
- Created 4 Tableau dashboards to visualize healthcare trends and key performance indicators, empowering 12 stakeholders with actionable insights derived from datasets containing medical history and operational metrics.
- Built reporting dashboards to visualize real-time healthcare metrics, such as medication sales and adherence rates, empowering medical staff to address inefficiencies in treatment delivery swiftly.

EDUCATION

Bradley University
Masters in Computer Science

Aug 2022 – May 2024
Peoria, IL

R.V.R. & J.C. College of Engineering
B.Tech in Electronics and Communication Engineering

Jul 2018 – May 2022
India

PROJECTS

Real-Time Data Pipeline for E-Commerce Analytics

- Designed and deployed a real-time data pipeline to process and analyze e-commerce transactions using Apache Kafka, Spark Streaming, and AWS Kinesis, enabling instant insights into sales trends and customer behavior.

Predicted autistic spectrum disorder using WEKA modeler

- Developed a predictive model for Autism Spectrum Disorder (ASD) using machine learning algorithms in WEKA. ● Pre-processed and analyzed ASD screening datasets to identify key behavioral and demographic indicators.

Healthcare Analytics Dashboard for Real-Time Insights

- Develop interactive dashboards for real-time healthcare monitoring and integrated billing systems and hospital databases for comprehensive data analysis by using Power BI, Tableau. **Business Intelligence (BI) System for IT Industries**
- Integrated data from databases, APIs, and cloud storage, for real-time reporting and insights and implemented ETL (Extract, Transform, Load) pipelines to clean, process, and store data in a structured format.

CERTIFICATIONS

-
- Joy of Computing using Python by NPTEL
 - Tableau for Data Scientists, Issued by LinkedIn Learning
 - SQL and Relational Databases 101, DB0101EN, provided by IBM
 - Anaconda Python for Data Science Professional by LinkedIn Learning
 - AWS Certified Solutions Architect – Associate certification, Issued by Amazon Web Services