

# PRITHEEP K

**Email:** kaliannanpritheep@gmail.com

**GitHub:** <https://github.com/PritheepK>

**Mobile No:** 8667266433

**Linkedin:** [www.linkedin.com/in/pritheep-k](http://www.linkedin.com/in/pritheep-k)

## PROFILE SUMMARY:

Motivated AWS Cloud enthusiast with hands-on experience in AWS services including S3, EC2, Lambda, Glue and Redshift. Skilled in Python, SQL, and PySpark for building serverless ETL pipelines and cloud-based data solutions. Strong understanding of cloud architecture, data processing and automation workflows. Eager to contribute to cloud projects, optimize infrastructure and deliver scalable and secure solutions.

## TECHNICAL SKILLS:

**Programming Languages:** PySpark, SQL

**Big Data Technologies:** AWS Glue, Amazon EMR

**Cloud Services:** AWS (S3, Redshift, Athena, Lambda, IAM, CloudWatch, SNS, Step Functions)

**Data Warehousing:** Amazon Redshift

**Data Streaming:** Amazon Kinesis

**Databases:** Oracle

**Data Visualization:** AWS QuickSight

**Operating Systems:** Windows, Linux

## PROJECT DETAILS:

### Project-1: Healthcare Data Lake & ETL Pipeline on AWS

A cloud-based project to build a data lake that processes and analyzes patient recovery data using AWS services.

- Used AWS S3 to store raw patient data and AWS Glue (PySpark) to clean and process it.
- Applied AWS Glue Crawler to detect data structure and Amazon Athena to run queries.
- Loaded final data into Amazon Redshift and created QuickSight dashboards for insights.
- **Technologies Used:** AWS S3, AWS Glue, PySpark, Amazon Athena, Amazon Redshift, Amazon QuickSight, SQL

### Project-2: Stroke Risk Analysis Using PySpark

Developed a big data pipeline to predict stroke risk using medical features from large-scale healthcare datasets.

- Cleaned and processed health data (demographics, smoking, BMI) using PySpark DataFrame API.
- Trained and evaluated ML models (Logistic Regression, Decision Trees) using Apache Spark MLlib.
- Visualized feature importance and model performance using ROC curves and classification metrics.
- **Technologies Used:** PySpark (RDD & DataFrame), Python, MLlib, Pandas, Matplotlib, Seaborn

## EXPERIENCE SUMMARY:

### AWS Cloud Data Engineer – Intern

Trinethra Tech Solutions,

May 2025 – Nov-2025

- Supported the development of **scalable data pipelines using AWS Glue and PySpark**.
- Assisted in transforming **healthcare datasets** for analytics using Apache Spark and Amazon Redshift.
- Collaborated with senior engineers to monitor ETL workflows using AWS CloudWatch and Step Functions.
- Gained hands-on exposure to cloud-native data tools (S3, Lambda, Athena) in a real-time environment.

## CERTIFICATIONS & ACHIEVEMENTS:

- **AWS Cloud Practitioner Essentials** – AWS Skill Builder (Jul 2025)
- **Presented and Published research paper on Stroke Prediction using Deep Learning** at AMATHE 2025 (IEEE co-sponsored) International Conference
- **Salesforce Administrator Virtual Internship** – SmartInternz (Nov 2023 – Jan 2024).

## EDUCATION DETAILS:

- **Bachelor of Engineering (B.E.) in Biomedical Engineering** from K.S.R. Institute for Engineering and Technology with **CGPA: 7.57**, Namakkal, Tamil Nadu Passed out: 2025.
- **Higher Secondary Certificate (HSC)** from Sri Vinayaga Higher Secondary School with **86.6%**, Namakkal, Tamil Nadu Passed out: 2021.
- **Secondary School Leaving Certificate (SSLC)** from Government Higher Secondary School with **79.6%**, Velagoundampatti, Tamil Nadu Passed out: 2019.

## STRENGTHS:

- Quick learner with strong analytical and problem-solving skills.
- Good communication and teamwork abilities.
- Eager to adapt to new tools, technologies, and industry best practices.

## REFERENCE:

1. Mr. I. Sathishkumar  
Cloud Data Engineer,  
Trinethra Tech Solutions  
Email: sathishmail57@gmail.com  
Phone: +91 98456 34242
2. Mr. Karthick Ravichandran  
Big Data Consultant,  
Trinethra Tech Solutions  
Email: karthick.ravichandran.msc@gmail.com  
Phone: +91 94449 99041