

# Yalini Sathiya R

Erode, Tamil Nadu | [yalinisathiya3@gmail.com](mailto:yalinisathiya3@gmail.com) | +919865954970 [LinkedIn](#) | [GitHub](#)



---

## SUMMARY

Certified Databricks Associate Data Engineer with strong training in building and optimizing scalable data pipelines using Apache Spark, SQL, and Delta Lake. Experienced in transforming raw data into reliable, analytics-ready datasets and applying modern data engineering best practices on the Databricks platform. Possesses a solid foundation in programming, data systems, and analytics, with a strong eagerness to contribute to data-driven decision-making and grow in data engineering, machine learning, and analytics. Seeking an entry-level data engineering role to deliver high-quality, impactful data solutions.

---

## CERTIFICATIONS

- NPTEL Certification: Foundation of Cloud IoT and Edge ML. • Mastermind Technologies: Python with AI Tamil Nadu
- Government (Naan Mudhalvan): Data Analytics with Tableau (Awarded on 13th July 2023).
- GUVI: Python Programming
- PAPER PRESENTATION: Presented a paper on Data Science using Mathematical Algorithms at Paavai Engineering College, Namakkal.
- Published a paper "HemoDetect: An AI-Powered Web Platform for Automated Leukemia Detection and Clinical Management" on International Journal of Progressive Research in Engineering Management and Science
- DATABRICKS DATA ENGINEER ASSOCIATE CERTIFICATE

---

## PROJECTS

### 1.HemoDetect: An AI-Powered Web Platform for Automated Leukemia Detection and Clinical Management

**Technology Used:** Machine Learning, Python, Web Development.

**Description:** Developed an AI-based web platform to automate leukemia detection by analyzing medical data and generating diagnostic predictions. The system preprocesses patient data, applies machine learning models for classification, and provides a secure interface for result visualization and clinical data management.

### 2. Social Media Insights & Sentiment Analysis Platform

**Technology Used:** Databricks, Apache Spark (PySpark), Spark NLP, Python, Google Colab, Databricks SQL.

**Description:** Developed an end-to-end data analytics platform to extract insights from public social media data (Twitter/X, Reddit). The system ingests and processes large-scale data using Databricks Bronze-Silver-Gold architecture, applies pretrained Spark NLP models for sentiment and emotion analysis, and visualizes trends, sentiment distribution, and hashtag analytics through interactive SQL dashboards.

---

## TECHNICAL SKILLS

**Programming:** Python, C, C++ **Web Development:**

HTML, CSS, JavaScript

**Tools:** Git, GitHub, Tableau, VS Code **Cloud & IoT:** Basics of Cloud Computing, IoT, Edge ML

**Data Science:** NumPy, Matplotlib,

**Database:** MySQL, MongoDB

---

## **EDUCATION**

### **MASTER OF COMPUTER APPLICATIONS (2023-2025)**

VIVEKANANDHA INSTITUTE OF INFORMATION AND MANAGEMENT STUDIES

CGPA: 9.0

### **BACHELOR OF SCIENCE (B.SC.) IN MATHEMATICS (2021-2023)**

VIVEKANANDHA COLLEGE FOR WOMEN-AGGREGATE: 84%

### **HIGHER SECONDARY CERTIFICATE (HSC) (2020-2021)**

SREE JANANI MATRICULATION HR. SEC. SCHOOL-AGGREGATE 82.5%

### **SECONDARY SCHOOL LEAVING CERTIFICATE (SSLC) (2019-2020)**

SREE JANANI MATRICULATION HR. SEC. SCHOOL-AGGREGATE: 95.2%

---

## **AREA OF INTEREST**

- Data Engineering (ETL pipelines, data modeling, Spark, SQL, Lakehouse concepts)

---

## **DECLARATION**

I Yalini Sathiya.R, hereby declare that the above written particulars are true to the best of my knowledge.