# SPARSHIKA AJMAAN DINESH KUMAR

+1 (919)-923-8528 | LinkedIn | sparshikaajmaan707@gmail.com | GitHub | Chicago, USA

## EDUCATION

**DePaul University**                                                                                   **Chicago, IL, USA**
Master's in Data Science.                                                                   (GPA – 3.75) Novemebr2025

**P.E.S College of Engineering**                                                              **Mandya, KA, India**
Bachelor's in Industrial and Production Engineering                       (CGPA – 7.5) June 2022

## EXPERIENCE

**Accenture**                                                                                        **Bengaluru, KA, India**
**Associate Software Engineer – iDRAC Server Project**              **October 2022 – November 2023**

- Performed Build Verification Testing (BVT) and automation validation on Dell iDRAC servers, ensuring firmware stability and compliance.
- Conducted comprehensive server validation including BIOS, CPLD, and license updates, achieving 100% compliance with iDRAC build standards.
- Built Python automation scripts for testing and reporting, cutting manual effort by 30%.
- Conducted server validation for BIOS, CPLD, and licensing, achieving 100% build compliance.
- Collaborated across teams for deployment, troubleshooting, and test execution to improve reliability.
- Completed Spring Boot and Advanced Java training, gaining hands-on experience in REST APIs and backend integration.

## PROJECTS

**Scalable Titanic Survival Prediction Using AWS & Apache Spark**                                    **Link**

- Built a distributed AWS EMR pipeline for 1M+ passenger records, improving query performance by 90% via Parquet optimization.
- Automated Spark-based ETL workflows, reducing processing time by 60% and enabling real-time analytics with Athena.

**Health Disease Prediction Using Python, Streamlit, Scikit-learn**                                 **Link**

- Designed and deployed an interactive ML web app that predicts disease risk (e.g., diabetes) from patient health inputs.
- Trained and optimized Random Forest and Logistic Regression models, achieving ~90% accuracy through robust feature engineering and EDA.
- Built a real-time Streamlit interface for dynamic user input, enabling instant, personalized health risk predictions.

**Automobile Sales Insights with RFM, PCA & Clustering**                                             **Link**

- Built an ETL process using AWS Glue to cleanse and transform automotive sales data for segmentation and PCA-based clustering.
- Applied RFM metrics to categorize customers into value tiers and exported structured results to Tableau dashboards for sales analytics.
- Streamlined monthly report generation workflows, reducing manual effort by 50% through automated Glue jobs.

**Road Traffic Accident Severity Data Pipeline Using AWS & Apache Spark**                           **Link**

- Engineered a real-time data pipeline integrating multi-source accident datasets into AWS S3 using Spark and Lambda, enabling continuous ingestion and near-instant analytics.
- Optimized transformations and feature extraction with Spark SQL and Pandas, improving pipeline throughput by 65% and enabling seamless data preparation for ML-based severity prediction.

**Stock Movement Prediction Using Twitter Sentiment Analysis (FinBERT + LSTM)** (In Progress)

- Built a data pipeline integrating tweets and stock for sentiment-based stock movement prediction.
- Used FinBERT and LSTM models for sentiment extraction and time-series forecasting, achieving 85 percent + accuracy.
- Delivered actionable insights on how sentiment affects daily stock returns — aligning with market analytics strategy.

## TECHNICAL SKILLS

**Programming:** Python, c, R, SQL, HTML/CSS, Object-Oriented Programming.
**Machine Learning:** Ensemble Methods - Random Forest, XGBoost, CNN, Transfer Learning, Adversarial Training, Feature Engineering, PCA, SMOTE
**Data Engineering:** Apache Spark, Hadoop, Hive, ETL Pipelines, Data Cleaning, Data Preprocessing
**Cloud & DevOps:** AWS - S3, EMR, Glue, Lambda, EC2, SageMaker, Athena, Kinesis, Redshift, QuickSight, DynamoDB, IAM, Bedrock, Docker, Jenkins, Git, GitHub, CI/CD, Agile, JIRA
**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn, Excel, StreamLit
**Libraries & Frameworks:** Scikit-learn, Pandas, NumPy, TorchVision, TensorFlow
**Tools**: Git, GitHub, Agile, Excel, ETL processes, Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn, TorchVision, Git, Jenkins, JIRA, Postman, AWS (S3, Lambda, EC2), Docker

## CERTIFICATIONS & ACHIEVEMENTS

- Spring Boot & Advanced Java Stream Training — Accenture (2023)                                    link
- GenAI Data & Analytics Academy Specialization — Starweaver (2025)                                 Link
- AWS Certified Cloud Practitioner (In Progress)