

Sudev Das P

Data Scientist

✉ mail4me.sudev@gmail.com ☎ 8848510413 📍 Malappuram, India 🌐 Github 🔗 LinkedIn

Data Scientist with 3+ years of expertise in building scalable data pipelines, data analytics, and machine learning. Proficient in Apache NiFi, Hive, PySpark, and Hadoop, with a strong background in machine learning and python. Passionate about data-driven problem-solving and delivering business insights.

Professional Experience

nStore Technologies

2024 – present | Thane

Data Scientist

- Engineered and deployed an automated ETL pipeline using Apache NiFi, streamlining real-time transaction data ingestion from PostgreSQL to HDFS, enhancing data availability by 90%.
- Optimized and automated Hive queries to structure and manage transactional data, cutting fraud analysis processing time by 90% and accelerating query performance by 25%.
- Converted predefined fraud detection rules into scalable Hive queries and PySpark scripts, improving fraud detection accuracy by 20% while minimizing manual intervention by 90%.
- Automated fraud analytics reporting, delivering real-time insights on suspicious transactions and boosting reporting efficiency by 70%.
- Integrated and fine-tuned NiFi, Hadoop, Hive, and PySpark workflows, enhancing system performance by 40% and enabling scalable fraud detection processes.

Infosys

2021 – 2023 | Bengaluru

System Engineer

- Designed and optimized SQL queries for data extraction, transformation, and reporting, improving query performance by 30%.
- Developed and maintained relational database schemas to ensure data consistency and integrity across enterprise applications.

Skills

Functional Skills

- Domain Knowledge: Fintech
- Data Cleaning and Preprocessing
- Data Analysis and Interpretation
- Statistical Analysis and Pattern Recognition
- Machine Learning and Predictive Analytics

Soft Skills

- Excellent Communication skills
- Proven ability to work collaboratively in teams
- Critical Thinking and Strong Problem Solving
- Time Management and Organization

Technical Skills

- Programming languages: Python | SQL
- ETL & Data Integration: Apache NiFi, Apache Spark
- Data analysis tools: PySpark | Pandas | NumPy
- Databases & Data Warehousing: PostgreSQL, Hive
- Data visualization tools: MS Power BI | Matplotlib | Seaborn
- AI: Machine Learning | Deep Learning | NLP
- Gen AI | ChatGPT

Projects

Customer Conversion Prediction - [Python | Pandas | NumPy | Matplotlib | Seaborn | Scikit]

- Engineered a high-accuracy predictive model achieving a 0.9 AUROC score, effectively identifying potential insurance policy subscribers.
- Employed various classification models (logistic regression, decision trees, XGBoost) to assess performance.
- Conducted thorough exploratory data analysis (EDA), unveiling crucial insights for feature selection leading to a

Resume Analyser and Job Finder - [Python | Open AI | LangChain | Selenium | Pandas]

- Developed an application that achieves a great accuracy in automatically extracting crucial information from PDF
- Utilized natural language models and AI, including OpenAI's GPT-3.5, to process and evaluate resumes, strengths, weaknesses, and actionable tips for resume enhancement, increasing candidates' job prospects.

- Integrated web scraping techniques and Selenium to search for and present job opportunities on LinkedIn that match

Certificates

- Guvi Master Data Science Certificate [↗](#)
- Supervised Machine Learning: Regression and Classification | Deeplearning.ai [↗](#)
- Machine Learning Specialization | Deeplearning.AI [↗](#)

Languages

English | Malayalam | Hindi

Education

B.Tech in Mechanical Engineering

College of Engineering Trivandrum

CGPA : 8.8

2017 – 2021