Albin Manuvel

+91 8884983769 albinmanuvel31@gmail.com *Bangalore,India*
github.com/Albinmanuvel www.linkedin.com/in/albin-manuvel-6607a9221

Summary

A recent graduate of St. Joseph's College with Bachelors in Computer Applications and data analytics, I bring a strong proficiency in SQL, R, Python, Power Bi , Machine Learning, and Big Data Technologies to the table. My educational journey has been complemented by hands-on internships at Suvidha Foundation and Neramlabs, where I honed my skills and tackled complex data challenges.

Education

St. Joseph's College , Bengaluru
Bachelors in Computer Applications (Data Analytics)
Jul, 2021 - Jul, 2024

Technical Skills

Python (Django , Pandas , NumPy , SciPy , Matplotlib , Seaborn , Scikit-learn ,Selenium ,TensorFlow , PyTorch), Hadoop Ecosystem (HDFS, Hive, Pig, HBase , PySpark , Kafka), AWS, core java, SQL , MongoDB , Neo4J, PowerBI, Linux

Soft Skills

Collaboration , Leadership , Problem-solving, Communication, Time management, Result-oriented

Additional Skills

Public Speaking, Writing, Research

Languages

English, Hindi, Kannada, Malayalam, Tamil, German

Work Experience

Neram Labs Private Limited
Data Science Intern
Apr, 2023 - Aug, 2023
Early stage start-up with under 50 employee , developed Grapp app for private tutors to find and students .

- Developed Python scripts to automate image noise-reduction, improving analysis time by 40%.
- Participated in 20+ team meetings to brainstorm and implement innovative solutionsand maintained comprehensive documentation of all processes, models, and analyses.

Suvidha Mahila Mandal NGO
Machine Learning Intern
Feb, 2023 - Mar, 2023
A NGO working for empowering communities through education, healthcare, and social care.

- Increased by 35% the reach of users for social causes in the platform using ML model trained with data collected and preprocessed 50,000+ data entries using Pandas and NumPy over the last 8 years.
- Developed 6 predictive models using Scikit-learn, including Logistic Regression and Random Forest and created 11 detailed reports and visualizations to communicate insights improving strategic decision-making by 25% through data-driven insights.

Projects

Project 1
E-commerce Data Analysis :

- Conducted an in-depth analysis of an e-commerce platform's transaction data using Hadoop and Spark, processing over 1 lakh transactions. Reduced data processing time by 30% through optimization techniques.
- Implemented data preprocessing, transformation, and analysis pipelines to derive insights on customer behavior, sales trends, and product performance.
- Developed interactive dashboards and visualizations using Power BI, resulting in a 25% increase in data-driven decision-making.

Project 2
Real-time Data Processing :

- Designed and implemented a real-time data processing pipeline using Apache Kafka and Spark Streaming, handling over 50,000 events per second.
- Processed and analyzed streaming data from IoT devices to monitor and predict equipment failures, improving predictive maintenance accuracy by 40%.

Certifications

- IBM Data Engineering Essentials course
- Data Structures and Algorithms using Java (IIT Kharagpur)
- Introduction to Artificial Intelligence (Infosys)
- Hands-on Artificial Intelligence with TensorFlow (Infosys)
- NCC, B grade in B Certificate.

- Reduced system downtime by 20% through timely identification of potential failures.

Project 3

Sentiment Analysis on Social Media Data :

- Utilized Hadoop and NLP techniques to analyze sentiment trends in social media data related to various brands, processing over 1.25 lakh social media posts.
- Extracted sentiment scores and generated visual reports using PowerBI, covering sentiment analysis of 50 brands.
- Achieved 85% accuracy in sentiment classification, providing actionable insights that led to a 15% improvement in brand sentiment management.

Project 4

Data Lake Implementation for Financial Services :

- Led the design and implementation of a Data Lake solution for a financial services company using Hadoop and HDFS, managing over 10 TB of data from 20 sources. Improved data retrieval times by 35%.
- Developed and optimized 50+ complex queries using Hive and PySpark for data analysis and reporting, reducing query execution time by 40%.
- Employed DevOps practices for automated deployment, decreasing deployment time by 25% and increasing system reliability.