Vinit Ladse

Hyderabad vinitladse5@gmail.com 9923441421 github.com/Vinitladse

SUMMARY

Highly skilled and results-driven Data Scientist with 3+ years of experience in analyzing complex data sets, developing predictive models, and providing valuable insights to optimize business strategies. Seeking an opportunity to leverage my expertise in data science and machine learning to drive data-informed decision-making and deliver measurable impact for the organization.

EXPERIENCE

Data Scientist

Edgerock Software Solution Private Limited, Hydrabad **February 2022 - Present, Hydrabad**

- Highly efficient Data Scientist/Data Analyst with 3+ years of experience in Data Analysis, Machine Learning, Data mining with large data sets of **Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data Visualization, Web Scraping.**
- Experience in statistical modelling languages with Python including Big Data technologies like Hadoop, Hive, and PySpark.
- Proficient in managing entire data science activity life cycle and actively involved in all the phases of activity life cycle including **Data Acquisition, Data Cleaning, Data Engineering, Features Scaling, Features Engineering, Statistical Modeling (Decision Trees, Regression Models, Clustering), Dimensionality Reduction using Principal Component Analysis (PCA) and Factor Analysis, Testing and Validation using ROC plot, K - fold Cross-Validation and Data Visualization.**
- Experience and deep understanding of **Statistical modeling, Multivariate Analysis, Model Testing, Problem Analysis, Model Comparison, and Validation.**
- in python-like **NLP, Pandas, NumPy, Seaborn, SciPy, Matplotlib, Sci-kit-learn, Beautiful Soup, Json, CSV**, highly skilled in using visualization tools like **PowerBI for creating dashboards.**
- Hands-on experience with big data tools like Hadoop, Spark, Hive, Pig, PySpark, Spark SQL. Hands-on experience in implementing LDA, Naive Bayes and skilled in **Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principal Component Analysis.**

PROJECT

Data-Driven Feasibility & Opportunity Assessment

Ciruss, USA

- In response to the dynamic landscape of our industry, the Company is initiating an internal project aimed at enhancing our strategic decision making processes. The primary objective is to systematically evaluate the feasibility and likelihood of securing projects based on comprehensive analysis of client information, our technological capabilities, available company resources, and a thorough understanding of our competitive landscape.
- **Roles**: Conducting comprehensive analyses of client data to identify trends, preferences, and potential opportunities.
- Leveraging statistical and machine learning techniques to extract valuable insights from large datasets. Collaborating with cross-functional teams to integrate client feedback into decision-making processes. Collaborating with teams across departments to ensure a holistic approach to project evaluation.

One Bridge Solution

AGI Global Logistics Ltd – Bristol

- Developed and optimized data pipelines to process shipment forecasts, standardize data formats, and enable predictive analytics for demand forecasting and capacity planning.
- **Key Responsibilities**:
- Designed and implemented end-to-end ETL pipelines using Azure Data Factory, Logic Apps, and Blob Storage to automate data ingestion and preprocessing.
- Standardized and transformed multi-format data (CSV, Excel, PDF) for integration with downstream systems using Python and master tables.
- Built data models and enriched datasets for capacity planning, linking hierarchical client data with key metrics like CBM (cubic meters).

• Orchestrated data storage layers (Bronze, Silver, Gold) in Couchbase NoSQL for structured and processed data.
• Collaborated with machine learning teams to integrate prediction outputs into the pipelines for real-time decision-making.
• Automated data processing reduced manual errors and accelerated forecast delivery.
• Enabled 90% ship utilization through efficient demand prediction and capacity planning.

ACA Reporting Predication of Insurance and Broker, Client, User Churn Model

Mercury Insurance Group, UK

• Ensure ACA compliance for clients by managing, validating, and analyzing large-scale data from multiple sources.
• **Key Challenges & Solutions:**
• Large Data Handling: Managed 90GB of monthly data via SharePoint API and email ingestion.
• Dynamic Data Matching: Used fuzzy matching for data accuracy.
• Cost Optimization: Leveraged Amazon Athena and AWS Glue for serverless querying and minimized infrastructure costs.
• **Responsibilities**:
• Data Ingestion: Used AWS Lambda and SharePoint API for file processing.
• Data Storage: Managed data in S3 and RDS/Redshift.
• Data Processing: Built ETL pipelines with AWS Glue and Python/PySpark.
• Querying & Reporting: Optimized Athena queries and visualized data in Quick Sight.
• Data Quality & Governance: Implemented quality checks for HIPAA/ACA compliance.
• Cost Optimization: Optimized S3 storage and used Glue/Athena for cost savings.

EDUCATION

M.Tech (PEPS)

Guru Nanak Institute of Engineering and Technology, Nagpur • September 2022 • 7.61

B.E. Electrical

Guru Nanak Institute of Engineering and Technology, Nagpur • November 2020 • 8.55

SKILLS

Machine Learning: Linear Regression, Logistic Regression, Decision Tree, Random Forest, K-means, KNN

Frameworks: Django, Flask

Programming Languages: Python, SQL

Python Libraries: Pandas, NumPy**,** Scikit Learn, TensorFlow, Seaborn, Keras

Version Control: Git, GitHub

Cloud Computing: AWS EC2, IAM, S3, AWS Glue, SES, SNS, RDS, Redshift, Sagemaker, Athena, Lambda, Step Function, ECR

Analytic Tools: PySpark, Pandas, Matplotlib, Seaborn, PowerBI

Databases: Oracle 10g, MySQL, MongoDB

IDEs/ Development Tools: PyCharm, Anaconda – Jupyter Nb, Microsoft Visual Studio

Deployment Tools: Docker, Heroku, Jenkins.

Web Service: REST –JSON, DRF (Django Rest Framework)

Programming Experience: ETL – Python | Machine learning –Python

BI Tool used: PowerBI | ETL Tool Used –Pyspark

Strong area : Python | SQL | Unix | Data Analytics –Etl , Algorithm (Machine Learning) | Artificial Intelligence | Databricks | Pyspark | Azure |Azure Data Factory

Domain Type: Banking | Payment | Telecom –Network switch/BSS | Health Care| Shipping

Gen AI : LLM Model | Gemini
