Raghuveera Narasimha

+1(680)356-8722 | raghunaras97@gmail.com | www.linkedin.com/in/raghunaras97/

Work Experience

Data Analyst, Open Financial Technology Pvt. Ltd, Bengaluru, Karnataka, India   **Mar 22 - Aug 23**
- Designed and implemented a secure, scalable data lake on ADLS Gen2, utilizing Unity Catalog for centralized governance.
- Engineered PySpark ETL pipelines on Databricks, processing 1TB+ daily data, achieving a 6-hour reduction in processing.
- Maintained a robust Snowflake data warehouse, utilizing optimized schemas and materialized views for efficient data analysis.
- Automated ETL workflows using Apache Airflow, orchestrating tasks and managing dependencies, reducing manual intervention by 40% reliably.
- Implemented data quality checks in PySpark, ensuring accuracy of critical financial data.
- Modeled complex SQL queries for reporting and analysis, achieving a 30% performance improvement in data retrieval time.
- Developed a PySpark Random Forest Classifier for payment analysis, increasing fraud detection accuracy by 25% for lending.
- Analyzed customer behavior via funnel analysis and A/B testing, increasing app engagement by 25% and reducing drop-offs.
- Built interactive dashboards in Tableau and Metabase, visualizing KPIs and providing insights, improving revenue forecasting by 15%.

Associate Data Analyst, XLDynamics Pvt. Ltd, Mysore, Karnataka, India   **Jun 21 - Mar 22**
- Streamlined loan processing and data cleansing to a U.S. mortgage company using advanced Excel functions (VLOOKUP, INDEX/MATCH, Pivot Tables), reducing report time and enhancing analytical productivity.
- Developed vibrant, actionable Power BI dashboards for loan management, improving decision-making speed and operational workflow efficiency
- Led knowledge transfer for two new hires on standardized procedures, fostering team efficiency and maintaining high accuracy in daily reporting and loan review processes.

Technical Skills & Tools

- **Big Data Tools:** AWS (Glue, Lambda, Athena, Neptune, EC2, EMR), Azure (Data Factory, Synapse, Databricks), PySpark, Hive
- **Programming Languages:** Python, R, SQL
- **Data Analysis & Manipulation:** Pandas, NumPy, Tidyverse
- **Data Visualization:** Matplotlib, Seaborn, ggplot2, Looker
- **Machine & Deep Learning:** Scikit-learn, Keras
- **Statistical & Analytical Techniques**: Descriptive, Predictive & Inferential Statistics, Time Series Analysis
- **Database Management Systems:** Mysql, Microsoft SQL Server, MongoDB, CassandraDB
- **Other Tools:** Jira, Microsoft Office, Docker, kubernetes, Terraform Git, Git Actions, Alteryx

Projects

COVID-19 Data Pipeline and Analysis with Azure Data Factory
Technologies: Azure Data Factory, Data Lake Gen2, SQL Database, Databricks, HDInsight, Azure DevOps, Power BI, CI/CD
- Developed a robust data pipeline using Azure Data Factory to process COVID-19 data, leveraging Data Flows for transformations and integrating with Databricks and HDInsight for complex processing.
- Automated data ingestion and pipeline execution using triggers, parameters, and variables, ensuring reliable data updates and implementing comprehensive monitoring and alerting.
- Implemented CI/CD with Azure DevOps and secured data access using Managed Identities and Key Vault, streamlining deployments and enhancing data security.

Airfare Prediction and Optimization with PySpark
- Utilized PySpark, MLlib, StringIndexer, OneHotEncoder, and GBTRegressor to build an airfare prediction model, achieving strong performance metrics (RMSE: 12.60, R-squared: 0.96), through rigorous data cleaning, feature engineering and selection.
- Delivered actionable insights through visualizations and reports, highlighting the key factors influencing price changes and enabling data-driven recommendations for pricing

Advance Chest Disease Detection using R-CNN | Syracuse University
- Enhanced Chest Disease detection from chest X-rays by developing a Mask R-CNN model with PyTorch, incorporating strategic data augmentation and balanced sampling for optimized dataset, thereby boosting diagnostic accuracy and model reliability.

Energy Consumption Analysis | Syracuse University
- Conducted advanced data analytics on over 4.2M records using R, improving data quality and uncovering insights with statistical tests, and built a linear regression model with a 79.8% R-squared, identifying energy efficiency drivers through Bayesian analysis.

Education

SYRACUSE UNIVERSITY, School of Information Studies (iSchool)   Syracuse, NY
Master of Science in Applied Data Science GPA:3.8   **Aug 23 - May 25**

VISVESVARAYA TECHNOLOGICAL UNIVERSITY (VTU), NIE SOUTH   Mysore, India
Bachelor of Engineering in Electrical and Electronics Engineering   **Jul 16 - Aug 20**

Certifications: Azure Data Fundamentals, Azure Data Engineer, IBM Data Science, Tableau Desktop Specialist