Gopi Chand Yenikapalli +18137341602 gopichandhemanth@gmail.com LinkedIn Portfolio GitHub

TECHNICAL SKILLS

Demonstrated Capabilities: Statistical Modeling, Time Series Analysis, Forecasting, Linear Regression Modeling, REST API
Unsupervised / Supervised Learning, Product Analytics, A/B Testing, Large Language Models (LLM's), MLOps, MLflow
Programming Languages: Python, R, Shell / Bash, Scala, JavaScript, HTML, CSS
Databases: SQL (Oracle SQL, Postgres), NoSQL (DynamoDB, Bigtable, MongoDB, Cassandra, Apache Hive, Impala),
Tools/Frameworks: PySpark (RDDs, DataFrames, Spark SQL and MLlib), Docker, Dbt, Airflow, GIT, BigQuery, AWS Redshift, AWS Glue,
Athena, Tableau, Power BI, Looker, Excel, Numpy, Pandas, Matplotlib, Plotly, Tensorflow, Pytorch, Transformers.
Additional Skills: Communication, Teamwork, Problem-solving, Critical thinking, Prompt Engineering.

WORK EXPERIENCE

Data Science and Engineering Intern | South Florida Water Management District   Jan 2024 – Present
- Actively developing a Python library for data retrieval and analysis, leveraging Python libraries, SQL connectivity, and API integration to access, retrieve, and process hydrologic, meteorological, and hydrogeologic data from the DBHYDRO database.
- Developing machine learning models to predict algal blooms using GCP Vertex AI and designing dashboards with Looker.

Graduate Research Assistant | University of South Florida (USF)   May 2023 – Present
- Designed and co-authored a course on the end-to-end Machine Learning Lifecycle, incorporating MLOps and Databricks MLflow in collaboration with Professor Varol O. Kayhan.
- Streamlined data handling by 60% using automated Python and R scripts, enhancing the efficiency of the ETL (Extraction, Transformation, and Loading) process, and improving the accuracy of environmental geospatial data analysis.
- Collaborated on the development of a statistical model for Lake Okeechobee's nutrient exports prediction.
- Developed a multi-classification machine learning model to predict the cognitive loads during specific simulated tasks.

Assistant Manager - Data Analytics | Aditya Birla Group (ABG)   Jun 2018 – Jul 2022
- Led and delivered complex data science projects involving unstructured data (text), natural language processing (NLP), and deep learning to solve ambiguous business problems.
- Employed advanced data analysis (A/B tests, quasi-experiments, causal analysis) in Oracle ML and SageMaker to optimize ABG's supply chain, boosting cost reduction, equipment lifespan, and operational efficiency via integrated app impact evaluation.
- Implemented Oracle ML for Supply Chain Optimization, reducing equipment costs by 15% through accurate cost allocation.
- Integrated Amazon SageMaker for Predictive Scheduling, reduced downtime by 20% and extending equipment life by 30%, contributing to supply chain and delivery optimization.
- Achieved budget forecasting accuracy of 95% through regression analysis techniques, ensuring financial preparedness in line with risk management approaches.
- Optimized ETL processes for big data workloads, streamlining document data extraction and transformation by 2.5 times for faster, more efficient decision-making.
- Enhanced data access by 50% through Oracle's graph capabilities and in-database machine learning, specifically for Asset Hierarchy Optimization, supporting geospatial data and logistics analysis.
- Created interactive Tableau dashboards and conducted data analysis to identify key trends, providing valuable business insights for informed decision-making.

EDUCATION

M.Sc. Information Systems (Data Science) | University of South Florida, Tampa, USA   Aug 2022 - May 2024
B.Tech. Engineering | National Institute of Technology (NIT), Raipur, India   Jun 2014 - May 2018

PROJECTS

Spatial and Temporal Variability in Concentration Discharge Relationships | R, Non-Linear Modeling, Data Visualization
- Developed an R pipeline to streamline geospatial data analysis workflows for the DBHYDRO database, integrating data cleaning, transformation, modeling, and visualization.
- Developed high-accuracy non-linear models using quantitative analysis, achieving 97% prediction accuracy for nutrient loads based on daily stream flow, outperforming existing statistical models by 10%.

An Agent-Based City Digital Twin for Pandemic Analysis and Simulation | GeoJSON, GeoPandas, Statistical Modeling, Plotly
- Collaborated in constructing a Digital Twin of Hillsborough County, comprising 52 zip codes, to examine pandemic trends.
- Utilized GeoJSON files to access spatial data and employed the GeoPandas Python package for data acquisition, data wrangling,
- Designed an interactive dashboard for the exploration and assessment of pandemic spread using Plotly and hosted it on Azure.

Predictive Analytics for Enhancing Customer Purchase Behavior in E-Commerce | Python, PySpark, SQL, Data Visualization
- Built a PySpark-based model analyzing 12K+ data points to predict customer product purchase intentions based on website visits.
- Enhanced accuracy by 27% via feature engineering, achieved an 88.62% F1 score with Linear SVC among diverse ML models.