Sai Kumar Mupparthi

+1 (302)-897-5850 | msaikumarda@gmail.com | https://www.linkedin.com/in/sai-kumar-mupp/ |Portfolio

Results-driven Data Analyst with over 2 years of experience leveraging Python, SQL, and R to transform complex datasets into actionable insights. Expertise in machine learning techniques (Scikit-learn, TensorFlow, XGBoost) to develop predictive models that optimize business processes across various industries. Proven track record of automating data pipelines using Apache Kafka, Talend, and Apache NiFi while ensuring seamless integration with cloud platforms like AWS and Google Cloud. Skilled in designing and deploying interactive dashboards using Power BI, Tableau, and Matplotlib to support data-driven decision-making. Adept at statistical analysis, including regression, hypothesis testing, and time series forecasting, delivering insights that improve operational efficiency and drive cost reductions. Committed to improving business outcomes, with notable achievements such as reducing manual work by 40% and improving supply chain forecasting by 20%.

SKILLS

- **Programming Languages:** Python, SQL, R
- **Data Visualization Tools:** Tableau, Power BI, Excel, Matplotlib, Seaborn
- **Machine Learning & Modeling:** Scikit-learn, TensorFlow, Keras, XGBoost
- **Libraries & Frameworks**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Big Data & Distributed Computing:** Apache Spark, Hadoop, Apache Kafka
- **Databases:** MySQL, PostgreSQL, MongoDB, MS SQL Server, Oracle Database
- **Data Science & Analytics Tools**: Jupyter Notebook, Anaconda, SAS Enterprise Miner
- **Statistical Analysis**: Regression Analysis, Hypothesis Testing, Time Series Forecasting, A/B Testing
- **Data Engineering & ETL Tools:** SQL-based ETL, Apache NiFi, Talend, Data Pipelines, Data Cleaning & Transformation
- **Cloud Platforms**: AWS, Google Cloud
- **Version Control**: Git, GitHub

PROFESSIONAL EXPERIENCE

Data Analyst | Johnson & Johnson Aug 2024 – Present

- Analyzed large healthcare datasets using Python (Pandas, NumPy), SQL, and Excel, delivering actionable insights to improve clinical trial outcomes and patient care decisions. Utilized advanced machine learning models in Scikit-learn and TensorFlow for predictive analysis.
- Engineered and automated interactive reporting dashboards with Power BI and Tableau, reducing manual reporting efforts by 40%, and integrated AWS Cloud storage for seamless data retrieval, enabling real-time insights for senior management.
- Applied statistical techniques, including regression analysis, hypothesis testing, and machine learning models (XGBoost), to identify key factors influencing patient outcomes, improving operational efficiency by 25%.
- Led the development and deploying predictive models to optimize supply chain forecasting, resulting in a 20% reduction in product shortages and a more efficient distribution model.
- Managed the cleaning, transforming, and merging large datasets from multiple sources using Python (Pandas, NumPy) and Apache NiFi, achieving 98% data quality accuracy. Integrated data into AWS-based pipelines for continuous analytics.

Junior Data Analyst | HCL Technologies July 2021 – Nov 2022

- Collected, cleaned, and analyzed large datasets using Python (Pandas, NumPy), SQL, and Excel, leveraging Talend for automated data extraction and improving decision-making across the business.
- Built and deployed automated data pipelines with SQL, Python, and Apache Kafka, reducing manual data handling time by 30% and enabling real-time analytics by integrating with AWS and Google Cloud platforms.
- Conducted extensive exploratory data analysis (EDA) on key business metrics, utilizing machine learning algorithms (Scikit-learn) to identify trends, resulting in a 15% improvement in operational efficiency.
- Developed interactive and automated Power BI and Tableau dashboards, cutting report generation time by 40%, and incorporated data streaming capabilities using Apache Kafka for real-time reporting and improved data visualization.
- Collaborated with cross-functional teams to perform root cause analysis, identifying issues in customer service processes and applying statistical analysis (A/B testing, regression) that resulted in a 20% reduction in customer complaints and improved customer satisfaction.

TECHNICAL PROJECTS

Exploring Weather Trends

- Developed SQL queries to extract, join, and normalize global and city-specific temperature datasets from multiple sources, enabling a robust comparative analysis of temperature trends across regions.
- Utilized Python (with Pandas and NumPy) to clean and transform raw temperature data, calculating moving averages and seasonal trends to uncover long-term temperature patterns.
- Designed dynamic, interactive visualizations in Matplotlib, Excel, and Tableau showcasing rising global and city temperatures. The visualizations highlighted key trends, such as increasing temperature anomalies, providing empirical support for the global warming hypothesis.
- Performed statistical analysis using time series forecasting techniques, including ARIMA (AutoRegressive Integrated Moving Average) to project temperature increases over the next decade. The results were communicated to stakeholders through interactive dashboards and presentations, aiding in understanding climate-related risks.

EDUCATION

Master of Science, Information Systems | Wilmington University, Delaware July 2024 | **GPA 3.89**

Bachelor of Technology, Information Technology | BIET, JNTUH Aug 2021 | **GPA 3.0**