KRISHNA CHAITANYA SARMA CHATRATHI
DATA ANALYST

(669) 600-9098 | krishnachatrathi21@gmail.com |LinkedIn| GitHub|Portfolio

SUMMARY

- **Data Analyst** with **3.5+** years of experience in **data processing, predictive analytics**, and **machine learning**, specializing in **Python, SQL,** and **R** to drive business insights and optimize decision-making.
- Expertise in **ETL**, **data warehousing, and big data technologies** such as **Snowflake**, **Apache Spark, AWS Redshift,** and **SQL Server**, improving data pipeline efficiency and processing large datasets at scale.
- Hands-on experience in **machine learning applications**, utilizing **BERT, Random Forest, and PySpark** for sentiment analysis, predictive maintenance, and fraud detection, leading to improved model accuracy and operational performance.
- Proficient in building **interactive dashboards** with **Tableau, Power BI, and Excel**, facilitating real-time **KPI** tracking and enhancing business intelligence capabilities.
- Experienced in **data quality management**, implementing **data governance, cleaning, data mining**, and **failure mode effect analysis (FMEA)** to ensure regulatory compliance and minimize inconsistencies in structured and unstructured datasets.

EXPERIENCE

CrowdDoing (Match4Action) **CA**
Data Scientist/Data Analyst **Aug 2024 - Current**

- Scraped and analyzed 100K+ herbal product reviews from various websites using **Python**, building a structured dataset for sentiment and quality analysis.
- Developed **ML models** using **BERT**, enhancing sentiment analysis accuracy by 28%, enabling stakeholders to refine marketing strategies.
- Implemented **PySpark-**based data pipelines, reducing large-scale data processing time by 37% and improving data ingestion efficiency.
- Orchestrated **data governance policies**, ensuring compliance with **GDPR** standards, reducing audit inconsistencies by 25%.
- Conducted **A/B testing** on **machine learning models**, selecting optimal algorithms for text classification, increasing model performance by **22%**.
- Designed interactive **Power BI** dashboards, visualizing customer sentiment trends, providing actionable insights for wellness strategy development.
- Integrated **AWS Glue** with **Snowflake**, streamlining **ETL** workflows and enabling real-time data updates, reducing manual intervention by 40%.
- Led stakeholder presentations, delivering insights on **customer behavior** and **product perception**, guiding data-driven decision-making.

Airbnb **Bangalore, India**
Data Analyst **Nov 2020 - Jul 2022**

- Built automated **ETL pipelines** using **SQL Server & SSIS**, processing **50K+ transactional records monthly**, reducing data discrepancies by 35%.
- Conducted **exploratory data analysis (EDA)** on 1M+ booking records, uncovering seasonal trends and demand fluctuations, leading to 12% increase in dynamic pricing efficiency.
- Optimized **Airbnb's recommendation system** using **Random Forest** and **K-Means clustering**, increasing guest booking conversions by 18%.
- Streamlined **SQL-based data reporting**, reducing **manual processing efforts by 50 hours/month** through advanced stored procedures.
- Developed **Power BI** dashboards for real-time monitoring of host performance and listing trends, improving data-driven decision-making for 50+ regional managers.
- Led **data wrangling & cleaning initiatives** using **Pandas** and **NumPy,** ensuring **88% data accuracy** across operational datasets.
- Automated data pipelines using **Apache NiFi**, integrating structured and unstructured data sources, reducing data latency by 30%.
- Collaborated with **cross-functional teams**, refining **pricing algorithms**, leading to a 7% increase in overall revenue per listing.

Tata Group **Bangalore, India**
Technical Data Analyst **Jul 2019 - Oct 2020**

- Designed **ETL** workflows in **Talend and SSIS**, optimizing **data extraction, transformation, and loading** for enterprise analytics dashboards, improving data availability by 45%.
- Conducted in-depth **data mining** using **Python** (**NumPy, Scikit-learn, PySpark**) to analyze thermal efficiency fluctuations in metallurgical processes, reducing energy waste by 10%

- Deployed **SQL-based** anomaly detection models to identify inconsistencies in supply chain logistics, enhancing procurement efficiency and reducing wastage by 18%.
- Integrated **Tableau** with **PostgreSQL**, visualizing **supply chain bottlenecks**, enabling cost-saving measures that reduced expenses by 12%.
- Automated weekly **KPI** reporting in **Excel (Pivot Tables, Macros, VBA)**, decreasing manual reporting workload by 50%.
- Conducted **Root Cause Analysis (RCA)** and **Failure Mode Effect Analysis (FMEA)**, identifying key process inefficiencies and reducing failure rates by 15%.
- Formulated **predictive maintenance models** using **Python and SciPy**, forecasting equipment failures with 85% accuracy, reducing downtime by 18%.
- Supported **mechanical engineering teams** by leveraging **Python-based automation scripts** to streamline equipment performance testing, reducing evaluation time by 30%.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R
Data Processing & Libraries: NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch, Apache Spark, Apache NiFi, Matplotlib, Seaborn, SciPy, PySpark, BERT
Machine Learning & Predictive Analytics: Linear Regression, Logistic Regression, Decision Tree, K-Means, Random Forest, Exploratory Data Analysis (EDA)
Big Data & Databases: Snowflake, MySQL, MongoDB, PostgreSQL, AWS Redshift, SQL Server, SSIS, SSAS, ETL, Talend
Data Visualization & Reporting: Tableau, Power BI, MS Excel (VLOOKUP, Pivot Tables, Macros, VBA)
Cloud Technologies: AWS (S3, EC2, Lambda, Glue), Azure
Data Management & Automation: Data Warehousing, Data Governance, Data Cleaning, Data Wrangling, Data Modeling, Data Mining, Automation Scripts, Root Cause Analysis (RCA), Failure Mode Effect Analysis (FMEA)
Methodologies: SDLC, Agile, Jira

EDUCATION

Illinois Institute of Technology **IL**
Master of Science in Data Science Aug 2022 – May 2024

Jawaharlal Nehru Technological University Kakinada **India**
Bachelor of Technology Aug 2013 – Jul 2017

PROJECTS

CTA Performance Analysis: R | Excel | Python | Pandas | AWS Lambda | AWS S3 | tensorflow | Sci-KitLearn

- Continuously retrieved real-time data from Train Tracker and Bus Tracker APIs by implementing web scraping techniques, orchestrated through an AWS Lambda function, and securely stored the data within an AWS S3 bucket.
- Performed thorough analysis of CTA bus routes and red lines, harnessing vast amounts of unstructured data gleaned from the CTA website.
- Utilized advanced predictive modeling methodologies such as Logistic Regression, Random Forest, Decision Tree Classifier, Support Vector Machines (SVM), and Neural Networks in Python and R, leveraging libraries like pandas, scikit-learn, and TensorFlow, to accurately forecast bus arrival delays.

Fraud Detection in credit card transactions: R|SparkR | PySpark | BigData

- Leveraged Apache Spark and SparkR to evaluate multiple machine learning models on large banking datasets, enhancing financial transaction security.
- Improved PySpark efficiency for handling extensive data, demonstrating the potential of big data and machine learning for proactive fraud detection in the financial industry.

Maternal Health Risk Analysis: Python | PowerBi

- Identified critical determinants of pregnant women's health, including high blood glucose levels (BS) as a significant risk factor, age over 25 as a concern, and a strong correlation between SystolicBP and DiastolicBP.
- Improved classification model accuracy from 86.7% to 90.15% through proficient data analysis, feature engineering, and machine learning model development for healthcare applications, showcasing strong data science skills and the ability to optimize model performance.

Car Manual Chatbot: Python | LangChain | RAG | NLP | LLM Integration

- Engineered context-aware car assistant chatbot, integrating vehicle manual with LLM using LangChain and RAG.
- Implemented real-time dashboard warning explanations and driving recommendations, enhancing user safety.
- Transformed static manual into interactive AI-powered knowledge base, showcasing NLP and conversational AI skills.