SAKSHAM ARORA

Troy, MI

: arorasaksham96@gmail.com | : (703)-980-3162 | :https://www.linkedin.com/in/sakshamarora96/ | : github.com/saksham-arora1796 | : arorasaksham.com

OBJECTIVE:

As a seasoned Data Scientist, I seek to leverage my extensive experience in machine learning modeling, generative AI, ETL processes, and the core aspects of data science. With a Master's degree in Data Analytics Engineering, I am proficient in Python, Machine Learning, Generative AI, LLM's, Advanced SQL, and AWS technologies. My goal is to leverage my expertise to develop and optimize cutting-edge data and analytical services, empowering executive decision-makers and frontline developers with actionable insights for strategic advantage.

WORK EXPERIENCE AND INTERNSHIPS:

Data Scientist | Daimler Truck Financial Services, Redford, MI (Duration: January 2022 – Present)

- Led the development of a sophisticated generative AI chatbot for Daimler Truck Financial Services, utilizing Azure OpenAI services for text summarization, named entity recognition, language translation, and question-and-answer functionalities.
- Leveraged advanced NLP techniques to streamline operations, collaborating closely with stakeholders to ensure seamless integration and optimize performance.
- Provided ongoing support and maintenance, refining the chatbot's capabilities to adapt to evolving business needs and technological advancements while maximizing efficiency through Azure OpenAI services.
- Led the Data Insights and Advanced Analytics team in utilizing Python to develop time series machine learning models, which were seamlessly integrated into an advanced Excel tool enabled with VBA. These models efficiently identified the organization's business staffing and acquisition needs.
- Spearheaded the formulation of a Collections Forecast model using advanced SQL and ML time series modeling techniques. This model enabled the Data Insights and Advanced Analytics team to accurately forecast delinquency numbers, measure credit losses, and analyze units sold, providing valuable insights for informed decision-making.
- Designed and implemented robust anomaly detection and drift detection systems using advanced Python libraries and statistical techniques, ensuring data quality and reliability for business-critical models. These systems effectively identified outliers and data drifts, maintaining model accuracy and preventing degradation over time.
- Demonstrated expertise in Reporting, Data Modeling, Analysis, Tableau Visualization, and Data Warehousing within a professional industry environment. The outputs of these processes were crucial for the financial institution and business stakeholders, as they catered to the credit paradigm for monthly, quarterly, and annual truck sales and repossessions.
- Successfully constructed a prediction model using SQL for data collection, blending, and cleansing, with subsequent development in RStudio and a user-friendly UI built on R Shiny. This prediction model, based on a blended version of the random forest model, accurately predicted the fair market value of trucks at the end of a lease term. The model achieved an impressive accuracy increase of over 75%, surpassing the threshold values of ±3%.

Data Scientist | Mercedes Benz Financial Services, Farmington Hills, MI (Duration: March 2021 – December 2021)

- Spearheaded the Data Science team, leveraging Data Analytics resources including Python, Tableau, and machine learning models to predict and analyze the impact of macroeconomic factors on the company's revenue generation strategies. Additionally, developed accurate forecasts for various key performance indicators (KPIs) to anticipate potential losses.
- Led the design and implementation of a robust prediction lifecycle and workflow, integrating machine learning tools, databases, and advanced SQL for iterative Model Evaluation (ME) activities, resulting in efficient and effective analysis.
- Successfully automated the company's manual legacy workflow, enhancing it with additional features that significantly improved accuracy by more than ±5 and expedited output generation by an impressive 85%.

Data Analyst | George Mason University, Fairfax, VA (Duration: September 2019 – February 2021)

- **Idioms Analytics Project -** Led the development of an analytical platform leveraging Python, Amazon Web Services, and Tableau to analyze the quality of written media based on idioms from an Oxford dictionary dataset.
- **Voting Fraud Analysis Project -** Executed a comprehensive analysis of voting trends across various states in the USA, employing Python and web scraping techniques to detect and uncover instances of voting fraud, ultimately identifying the individuals involved in the fraudulent activities.

Data Analyst | Auto Pearl, India (Duration: August 2017 - August 2018)

- Spearheaded the end-to-end ETL (Extract, Transform & Load) process, resulting in an impressive efficiency improvement of 78%.
- Managed data import and export operations utilizing Flume and Kafka, ensuring seamless data flow.
- Developed a dedicated topic to capture online purchase data, facilitating streamlined analysis.
- Leveraged Python and R to clean and structure data, achieving a notable 27% increase in processing speed.
- Implemented partitioning and bucketing techniques in Hive for optimized data storage and retrieval.
- Successfully designed and maintained the architecture of Redshift and EMR Clusters, ensuring efficient data management.
- Utilized Spark with Scala on AWS EMR to handle real-time data, enabling prompt analysis and decision-making.
- Orchestrated scheduled jobs for data loading into HDFS using Oozie, resulting in automated and expedited processes surpassing the performance of 46% of traditional in-house methods.
- Developed comprehensive dashboards and reports using Power BI, catering to the needs of sponsors and providing actionable insights.

EDUCATION:

- **MASTER OF SCIENCE (Data Analytics Engineering)** – (GPA – 3.77) (2018 – 2020)
  Volgenau School of Engineering, George Mason University, Fairfax, Virginia, USA.
- **BACHELOR OF TECHNOLOGY (Computer Science & Engineering)** – (GPA- 3.8) (2014 – 2018)
  Amity School of Engineering & Technology, Amity University, Noida, India.

SOFTWARE AND TECHNOLOGY PROFICIENCY:

Programming Languages & Concepts: Python, R, Generative AI, Large Language Models (LLM), Machine Learning, Advanced SQL, Advanced Excel.
Databases & Data Management: MySQL, MongoDB, Neo4j, PostgreSQL, Oracle 12c, HDFS (Elementary Knowledge).
Data Science & Machine Learning Tools: Stable Diffusion, Alteryx, Orange for ML, Vertex AI (for model development and deployment with Elementary Knowledge), DBeaver.
Data Visualization: Tableau, GGPlot (R for Data Visualization).
Cloud Platforms & Services: AWS (EC2, Route 53, S3, CloudFront, Data Lakes), Databricks (Elementary Knowledge), Microsoft Azure (Elementary Knowledge).
Version Control & Collaboration Tools: GitHub, SaaS (Functional Knowledge).
Data Analytics & Transformation Tools: SQL Developer Workbench, Alteryx.
Datasets & Domain-Specific Knowledge: Bloomberg Data (Financial Data Infusion), CPG (Consumer Packaged Goods) Data (Nielsen, IRI).

PROJECTS:

- **Psychiatric Chatbot:** A chatbot created using Python and NLP, which works like a psychiatrist and provides all sorts of emotional & mental support.
- **IoT based Pollution Predictor and Plotter:** Python based project that uses machine learning algorithm along with Scikit-learn and d3.js libraries. It uses MQ135 & MCP3008 interfaced with Raspberry Pi to collect pollution data (LPG, CO & Smoke) and predict the pollution levels in real time.
- **Stock Market Analysis:** This project uses both SQL and HIVE to analyze the stock market data taken from Yahoo Stocks. It then compares the efficiency with which both the languages can process a query and produce the results. It uses SQOOP to obtain the data from the dataset to HDFS.