ABHIJIT SRIRAM

+1 (703) 677-8975 | abhisriram25@gmail.com | Virginia, USA | linkedin.com/in/abhijt-s-b06701336/

EDUCATION

George Mason University **August 2022 - May 2024**
Master's, Data Science *GPA: 3.9*

PROFESSIONAL EXPERIENCE

Koch Industries **United States**
Data Scientist *November 2023 - Present*

- Designed and deployed machine learning models (regression, classification, clustering, time series) using Python, Scikit-learn, TensorFlow, and PyTorch, optimizing business processes and forecasting with ARIMA, LSTM, and Prophet.
- Built real-time data pipelines with Apache Kafka, Spark Streaming, Apache Flink, and AWS Kinesis, enabling streaming analytics and predictive decision-making at scale.
- Performed statistical analysis, A/B testing, hypothesis testing, and predictive modeling to derive actionable insights from structured and unstructured datasets using tools like Pandas, NumPy, and R.
- Developed and maintained ETL pipelines and data validation frameworks using Airflow, SQL, Snowflake, Teradata, and PostgreSQL, ensuring clean, reliable, and production-ready datasets.
- Created dashboards and visualizations with Power BI, Tableau, and Looker, automating KPI reporting and empowering business teams with data-driven insights.
- Applied NLP techniques (using spaCy, NLTK, Hugging Face Transformers) and built LLM-based prototypes for tasks like document summarization, sentiment analysis, and intelligent customer support.
- Followed MLOps best practices with MLflow, Docker, Git, Jenkins, and CI/CD pipelines to manage end-to-end model lifecycle, streamline deployment, and ensure reproducibility in cloud environments (AWS, GCP, Azure).

Markel Group **United States**
Data Scientist *December 2022 - October 2023*

- Built and deployed predictive models (Random Forest, XGBoost, Logistic Regression) using Python, R, and SAS to forecast financial outcomes and enhance investment strategies; performed time-series forecasting with ARIMA and LSTM to predict market trends.
- Engineered real-time fraud detection pipelines using Kafka, Spark Streaming, and NLP techniques (spaCy, NLTK) to analyze unstructured financial data, market reports, and sentiment, proactively identifying anomalies and mitigating risks.
- Designed and optimized financial risk models using supervised learning techniques, enhancing portfolio performance and regulatory compliance by extracting and analyzing high-volume data from SQL, NoSQL, Snowflake, and Amazon Redshift.
- Created interactive dashboards and automated reports with Tableau, Power BI, and Excel Macros, delivering timely insights to stakeholders and informing strategic financial decisions.
- Automated ETL workflows and financial data pipelines using Apache Airflow, Talend, and SQL, reducing manual intervention and streamlining financial reporting processes.
- Deployed financial models to production using CI/CD pipelines with Docker, Kubernetes, Jenkins, and Flask/FastAPI, ensuring continuous integration and model reliability in real-time applications.
- Utilized deep learning frameworks (TensorFlow, PyTorch) for complex financial forecasting and anomaly detection; applied log analytics with Elasticsearch and Splunk to monitor system performance and detect operational issues in production systems.

Fidelity Investments **India**
Data Scientist *June 2020 - July 2022*

- Extracted and analyzed large-scale financial datasets using SQL, Python, and R, uncovering trends and delivering actionable insights that optimized investment strategies and enhanced portfolio performance.
- Built and deployed machine learning models for predictive analytics, risk assessment, and trade optimization using TensorFlow, PyTorch, and Scikit-learn, improving forecasting accuracy and decision-making.
- Utilized big data tools like Hadoop, Apache Spark, and Hive for distributed processing and analysis of massive financial datasets, enabling faster and more scalable investment analytics.
- Developed interactive dashboards and visualizations using Tableau, Power BI, and Matplotlib to deliver market trend insights, performance metrics, and investment recommendations to business stakeholders.
- Optimized data storage and querying pipelines with Snowflake, Redshift, and Google BigQuery for scalable data retrieval; managed unstructured data using MongoDB, Cassandra, and Elasticsearch to support trading signals and research workflows.

- Applied advanced Natural Language Processing (NLP) techniques with spaCy, NLTK, Hugging Face Transformers, BERT, and TextBlob for sentiment analysis, text mining, and financial report summarization, aiding investment decisions and customer engagement.
- Built and managed CI/CD pipelines for model deployment using Jenkins, GitHub, Bitbucket, GitLab CI/CD, and CircleCI, ensuring robust version control, automated testing, and continuous delivery across financial data projects.

Baxter International **India**
Data Analyst *March 2018 - March 2020*

- Analyzed clinical and operational healthcare data—including patient admissions, treatment efficacy, and medication usage—using Python (Scikit-learn, Statsmodels) and R, enabling data-driven clinical decision-making and outcome optimization.
- Utilized big data frameworks such as Apache Spark and Hadoop to process large-scale electronic health records (EHRs), uncovering patterns in care delivery and identifying areas for workflow improvement.
- Designed scalable data models and data warehouses using Snowflake, Azure Synapse Analytics, and Google BigQuery, supporting robust healthcare analytics and clinical decision support systems.
- Created detailed data visualizations and trend analyses using Matplotlib, Seaborn, and Plotly to illustrate patient demographics, treatment outcomes, and hospital performance metrics for stakeholders.
- Implemented machine learning models (e.g., XGBoost, LightGBM) to predict disease progression, identify at-risk patients, and improve treatment recommendations, enhancing personalized care delivery.
- Deployed deep learning models using TensorFlow and Keras for health risk prediction and real-time monitoring of patient vitals; integrated with Apache Kafka and Apache Flink to build streaming pipelines for critical alerting in clinical settings.
- Built end-to-end CI/CD pipelines using Jenkins, GitLab CI/CD, and Azure DevOps, automating deployment of analytics applications and ML models across cloud-based healthcare environments, ensuring operational efficiency and compliance.

SKILLS

Programming Languages: Python, R, SQL, Scala, Java
Data Analysis libraries: Pandas NumPy, SciPy, Scikit-learn, Stats models NLTK Plotly, Matplotlib, scala
Big Data Technologies: HDFS. Scoop. Flume, Oozie. PySpark, Data Lake, HBase, Redshift, Kafka. YARN. Spark Streaming, ML Lib ZooKeeper
Machine Learning Libraries: Scikit-learn, TensorFlow, Karas, PyTorch, XGBoost, LightGBM
Data Modeling Tools: Toad Data Modeller SQL Server Management Studio MS Visio SAP Power designer Erwin 9.x
Databases: Teradata MYS My SQL Server,PostgreSQL,Oracle12c/11g/10g/9i MS Access 2016/2010, Hive,SQL Server 2014/20 16, Amazon Redshift, Azure SQL Database
Reporting Tools: Crystal reports Xl/2013, SSRS, Business Objects 5.x/ 6.x, Tableau, Informatica Power Center
Cloud Technologies: AWS (EC2, S3, SageMaker, Lambda, Redshift, Glue), Microsoft Azure (Azure ML, Synapse Analytics, Blob Storage), Google Cloud Platform (GCP), Databricks, Snowflake, IBM Cloud
Big Data Warehousing: Amazon Redshift, Snowflake, Google BigQuery, Azure Synapse Analytics, Teradata, Vertica, Delta Lake
Containerization: Docker, Kubernetes, Jenkins, GitLab CI/CD, CircleCI, Argo Workflows, Terraform, Helm, Ansible
Natural Language Processing (NLP): NLTK, spaCy, Gensim, BERT, GPT, Hugging Face Transformers, OpenAI APIs, TextBlob, FastText, T5, LangChain
Deep Learning & AI: TensorFlow, Keras, PyTorch, CNNs, RNNs, LSTMs, GANs, Transformers, Autoencoders, Attention Mechanisms
Automation & Scripting: Bash, PowerShell, Python scripting, VBA, Cron Jobs, Apache Airflow
Satistical Analysis & Forecasting: Hypothesis Testing, A/B Testing, Regression Analysis, Time Series Forecasting (ARIMA, Prophet), Experimental Design, ANOVA, PCA, Clustering (K-Means, DBSCAN)