MIN PRASAD GAUTAM

**Data Scientist/Machine Learning Specialist • NRH, Texas**

**254-563-1494**

**gautam.pgautam07@gmail.com**

# Objective/Summary:

Detail-oriented Data Scientist/ML Engineer with a strong background in python, data science and machine learning analytics. Skilled in gathering, interpreting, and manipulating structured and unstructured data to develop scalable, automated solutions for delivering valuable business insights. Adept at translating complex data into clear, understandable reports for decision-makers. Seeking a role where I can leverage my expertise in end-to-end project deployment, client communication, SQL, data engineering, machine learning, and collaboration to address complex data challenges and drive impactful results. Excels in fast-paced environments and thrives on solving challenging analytical problems.

* Proficiency in Hadoop,Hive and Spark for processing large datasets.
* Exceptional analytical skills for insightful decision-making.
* Proficient in ad-hoc statistical and data mining analysis using tools like R and python.
* Experienced in developing scalable data pipelines for handling large volumes of structured and unstructured data.
* Proficient in advanced statistical analysis, machine learning, and data mining techniques.
* Skilled in developing predictive, prescriptive, and inferential models to address business challenges.
* Experience leading end-to-end data analysis/data science projects from data collection and analysis to model deployment.
* Root Cause Analysis, Advanced analytical and quantitative skills.
* Researching, auditing, and documenting data analytics and underlying data sources.
* Strong presentation and communication skills for translating complex concepts into actionable insights and the ability to work in a fast-paced environment.

# EXPERIENCE:

**Data Scientist | Reece USA | Dallas, TX | 07/2021 – Present**

* Reduced data analysis processing time by 75% through advanced data processing techniques and automated workflows, resulting in 25-30% higher accuracy in data analysis.
* Led the development and optimization of distributed machine learning models in a heterogeneous domain environment, resulting in improved performance and efficiency.
* Designed, constructed, and maintained data pipelines, optimizing ETL processes for efficient data extraction, transformation, and loading.
* Conduct rigorous testing, validation, and deployment of AI models throughout their lifecycle.
* Strong problem-solving mindset, leveraging time series analysis to uncover actionable insights and drive strategic business outcomes.
* Led a team of data engineers and data scientists in optimizing ETL processes using PySpark, resulting in a 30% reduction in data processing time.
* Successfully deployed and managed machine learning applications on AWS cloud, utilizing services like SageMaker, EMR, S3, and VPC endpoint.
* Demonstrated deep expertise in data modeling and analytics using SQL and Python.
* Took the lead in driving distribution center analytics, employing techniques such as time-series forecasting, classification, regression, and clustering.
* Create compelling graphical representations of analyses and findings, effectively communicating their impact to stakeholders.
* Developed scalable data pipelines to handle large volumes of structured and unstructured data, ensuring data quality and reliability.
* Applied advanced statistical analysis, machine learning, and data mining techniques to extract actionable insights from complex datasets.
* Developed and implemented predictive, prescriptive, and inferential models to address business challenges and inform decision-making processes.
* Conducted exploratory data analysis (EDA) to gain insights from data, identify patterns, and inform data science strategies.
* Collaborated with diverse departments to architect data pipelines for ML modeling and process optimization.
* Leveraged SQL and Python to conduct ad-hoc analysis of structured and unstructured data from diverse sources, delivering valuable insights that informed strategic decisions.
* Lead and perform complex statistical analyses to identify performance trends and key drivers.
* Utilized machine learning visualization flows to enhance data analysis and interpretation.
* Achieved a 30% reduction in data processing time by optimizing PySpark data pipelines, resulting in faster insights generation and report delivery.

**Environment & Technology:** Elasticsearch, AWS, Cassandra, PostgreSQL, Kibana, Snowflake, TensorFlow, Power BI, Python, Spark, Docker, SQL, Github, Git, Redis, Tableau, Microsoft SQL Server,Streamlit

## Data Scientist | Bristol Myers Squibb | New Jersey, NY | 05/2020 – 07/2021

* Applied advanced machine learning models and statistical techniques to analyze complex clinical data, leading to 25% increase in accurate predictions and 25% enhancement in actionable insights.
* Conducted in-depth analysis of complex clinical, healthcare and supply chain datasets, resulting in the identification of cost-saving opportunities and $2 million in savings.
* Analyze healthcare data from diverse sources to identify patterns and trends, facilitating informed decision-making.
* Develop and deploy machine learning models for predictive analytics, disease diagnosis, patient risk assessment, and treatment recommendations.
* Collaborate with healthcare experts and clinicians to address healthcare challenges through data-driven solutions.
* Conduct statistical analysis and hypothesis testing to derive actionable insights from clinical and operational data.
* Apply NLP techniques to extract valuable information from medical texts and unstructured data.
* Contributed to the development and implementation of descriptive, predictive, and prescriptive statistics solutions for real-world problems using SHAP visualization hosted on AWS.
* Proficiently worked with AWS applications, including EMR, SageMaker, CloudWatch, S3 Data Lake, and more
* Applied advanced analytics techniques, such as time series models and cluster analysis, to drive data-driven decision-making.
* Built, automated, and maintained data pipelines to preprocess and transform clinical datasets for analysis, ensuring data accuracy and integrity.
* Developed and coded software programs, algorithms, and automated processes to cleanse, integrate, and evaluate large and diverse datasets.
* Deconstructed descriptive data into meaningful categories for predictive modeling and visual analysis.
* Conducted ad-hoc statistical and data mining analysis using tools like R and python

**Environment & Technology:** Python, Keras, Pytorch, Spark, AWS, MongoDB, Tableau, BioBert, R, NoSQL, Domino

## Machine Learning Engineer | Charter Communications | Maryland Heights, MO | 03/2018 – 04/2020

* Developed and deployed machine learning models that increased revenue by 20% within six months, integrating Alteryx and machine learning algorithms for order volume forecasting and improved lead predictive analytics.
* Created Model Evaluation of Time Series Analysis Tool which is able to measure how close the future predicted KPI values are with the actual KPI values when we get those KPI values.
* Applied anomaly detection methodologies z-scores and the modified z-score methodology for 4G KPIs at different granularities such as hourly level based on eNodeB and market level data.
* Successful track record of building and fine-tuning a variety of time series models, including ARIMA, SARIMA, Exponential Smoothing, and Prophet, for accurate and reliable predictions.
* Leveraged Airflow for workflow orchestration, ensuring efficient and automated data processing.
* Automated data pipeline monitoring and alerting using tools like Apache Airflow, ensuring pipeline reliability and minimizing downtime.
* Conducted data exploration and manipulation using SQL and NoSQL databases, contributing to data-driven decision-making across the organization.

**Environment & Technology:** Hbase,Azure, AWS SageMaker, Airflow, Hadoop, Spark, PySpark, Hive, Talend, Linux, R, Alteryx, Unix, Scripting

## Data Scientist | FedEx Supply Chain | Fort Worth, TX | 09/2016 – 02/2018

* Deployed predictive analytics to optimize customer service, resulting in a 30% reduction in customer complaints.
* Led the design and development of data ingestion and transformation framework, reducing data processing time by 40% and enabling seamless integration of structured and unstructured data sources.
* Utilized Spark,Hive and SQL to extract and merge data from diverse sources, creating seamless data pipelines for validation and enrichment.
* Played a key role in establishing metadata for all stages of the data pipeline, enhancing traceability and transparency throughout the process.
* Collaborated closely with multidisciplinary teams and stakeholders to define data requirements and establish interfaces through APIs, facilitating smooth data access.
* Utilized Python programming expertise to create data pipelines, validating and enriching data for machine learning models.
* Actively develop advanced Data Science skills in areas including data wrangling, statistical analysis, visualization, artificial intelligence, machine learning, and big data.

**Environment & Technology:** R, Python, AWS, Machine learning, Power BI, SAS, Pig, Cloudera, SQL, Excel, Java, C++

# Skills:

* Data Analysis: Proficient in end-to-end data analysis, including data preparation, modeling, and evaluation
* Machine Learning: Strong experience in developing machine learning systems, continuous training, and deployment
* Programming: Skilled in Python, R, and statistical programming languages for analytics and modeling
* Problem Solving: Strategy Consulting, Action-oriented Approach
* Data Insights: Ability to translate complex analysis insights into actionable strategies for non-technical audiences.
* Communication: Excellent oral and written communication skills for effective collaboration and presentation.
* Data Science: Data exploration, data mining, statistical analysis
* Cloud Environments: Familiarity with AWS/Azure
	+ Statistical and Machine Learning Techniques: Logistic Regression, Time Series Analysis, Ensemble Learning, Mixed Modeling, Multivariate Statistics, Large-Scale Predictive Modeling, XGBoost, LightGBM, Random Forests, Neural Networks
	+ Hypothesis testing, sampling, experiment design (e.g. A/B testing, MAB, etc.) and causal inference methods
	+ Databases: Experience with NoSQL databases, including MongoDB and Cassandra
	+ Healthcare Data Analysis

**Data Science Packages:** NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, PyTorch, Dask, TensorFlow, Keras, Scipy, Pydantic

**EDUCATION: Southern Methodist University** •Master of Electrical Engineering •Dallas, TX