Varshilkumar Patel
Data Scientist
varshilkumar.p@mymailwork.com | 929-613-8684 | Richardson, TX | **LinkedIn**

SUMMARY
Data Scientist with 3+ years of experience in Data Extraction, Modeling, Machine Learning, and Data Visualization. Proficient in using machine learning algorithms like Regression, Decision Trees, and NLP to drive predictive insights. Skilled in Python (NumPy, Pandas, SciPy) for data manipulation, and experienced with Tableau, Power BI, and Excel for impactful visualizations. Hands-on expertise in cloud platforms (AWS, GCP) and databases (MySQL, SQL Server, MongoDB), delivering scalable solutions and seamless data management. Adept in Agile and Waterfall methodologies, ensuring efficient project execution and delivery.

SKILLS
Programming Languages: Python, R, SQL, Scala, Julia, MATLAB
Machine Learning: Regression analysis, Bayesian Method, Decision Tree, Random Forests, Support Vector Machine, Neural Network, Sentiment Analysis, K-Means Clustering, KNN, Classification, SVM, Naive Bayes, NLP, LLM, CNN, XGBoost, Deep Learning
Deep Learning: TensorFlow, Keras, PyTorch, CNNs, RNNs, LSTMs, GANs
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly
Databases: MySQL, PostgreSQL, MongoDB, Cassandra, Redshift, Snowflake, BigQuery
Cloud Platforms: AWS (S3, EC2, Redshift, SageMaker), Google Cloud Platform (GCP), Microsoft Azure
Natural Language Processing (NLP): Text Mining, Sentiment Analysis, Topic Modeling, Named Entity Recognition (NER)
Software Development: Object-Oriented Programming (OOP), Agile, CI/CD, DevOps, Docker, Kubernetes
Business Skills: Data-Driven Decision Making, Business Strategy, Communication of Insights, Stakeholder Management
Data Analysis & Modeling: Predictive Modeling, Statistical Analysis, A/B Testing, Time Series Analysis, Clustering, Regression, Forecasting
Software/ Version Control/Other Skills: Data Cleaning, Data Wrangling, Critical Thinking, Communication Skills, Presentation Skills, Problem-solving, Decision-Making, EDA, Databricks, Data Visualization, Predictive Analytics, Pattern Recognition, JMP, Data Integrity, Quantitative Data, Data Science, Statistics, Data Analytics, Data Modeling, Data Analysis, Data Mining, SAP, Mathematics, Computer Science, Programming, Feature Engineering, Model Deployment, API Development, Experiment Design, Git, GitHub, Bitbucket
Operating System: Windows, Linux

EDUCATION & CERTIFICATIONS
Master in Computer Science | The University of Texas at Dallas, Richardson, TX
Bachelor in Computer Engineering | Charotar University of Science and Technology, Anand, Gujarat, India
Google Data Analytics.
AWS Certified Cloud Practitioner.

WORK EXPERIENCE
Intuit, USA | Data Scientist   Jan 2024 - Current
- Worked in Agile environments to deliver iterative solutions and adapt to changing project requirements.
- Conducted hypothesis tests to evaluate the significance of relationships between variables and make data-driven decisions.
- Formulated and deployed an advanced real-time AI monitoring system using XGBOOST and Random Forest Algorithm to detect and diminish fraudulent activities, achieving a 5% enhancement in fraud detection efficiency.
- Performed exploratory data analysis (EDA) to find insight into the difference in efficiency among devices using ggplot2 in R.
- Integrated Elasticsearch via Docker to enhance the analytics pipeline, achieving a 50% faster indexing and search capability, significantly improving insight extraction speed.
- Implemented Apache Spark streaming API in an active cluster environment, enabling real-time processing and analysis of insurance data streams, enhancing operational efficiency and responsiveness.
- Leveraged AWS services (EC2, VPC, RDS, EBS, S3, CloudWatch) to construct scalable and secure data science solutions, optimizing computational resources and upholding data privacy.
- Created and maintained reports to display the status and performance of the deployed model and algorithm with Tableau.
- Transformed large datasets from SQL Server and CSV files into structured data frames using PySpark.

Druva Software, India | Data Scientist   Jan 2020 - July 2022
- Involved in developing campaign waterfall reports per the business requirements to facilitate highly accurate decisions.
- Provided technical support for SAS programming in cross functional teams, specializing in data manipulation and analysis. Imputing missing values using regression methods.
- Led an in-depth comparative analysis of Neural Network and Random Forest models, achieving an 87.44% accuracy rate that significantly elevated predictive analytics performance in healthcare applications.
- Streamlined data integration by configuring AWS Data Pipeline to load data from S3 to Redshift, enhancing analytics accessibility.
- Completed advanced training in data analysis techniques, driving the adoption of best practices that enhanced team productivity by 15%.
- Designed and deployed automated data-cleaning workflows, reducing data errors by 30% & accelerating data validation efficiency by 25%.
- Maintained and optimized databases and wrote SQL queries for data extraction, transformation, and loading (ETL) processes.
- Utilized NLP techniques and libraries such as NLTK (Natural Language Toolkit) for text preprocessing, feature extraction, and text mining tasks.
- Analyzed large data sets using pandas and used regression models by using SciPy to predict future data and visualized them
- Collaborated with cross-functional teams to integrate Databricks solutions into existing data pipelines, ensuring seamless data flow.
- Implemented supervised and unsupervised learning algorithms, including LLM (Latent Dirichlet Allocation), to uncover patterns in large datasets.
- Developed complex SQL queries using stored procedures, common table expressions (CTEs), and temporary tables to support Power BI reports.
- Created custom SQL queries in MS SQL for data extraction and reporting with Crystal Reports.