

# Pujan Motiwala

Greater Dallas Area, TX

1-732-306-5127

[pujanm1991@gmail.com](mailto:pujanm1991@gmail.com)

<https://github.com/PujanMotiwala>

<https://www.linkedin.com/in/pujanmotiwala>

<https://portfolio.pujanmotiwala.com>



## Professional Summary

Innovative Senior Data Scientist with **over 9 years of experience** driving **significant business impact** through **advanced analytics** and **machine learning** across cybersecurity, finance, and marketing domains. Proven expertise in translating complex data into actionable insights, optimizing decision-making processes, and delivering cutting-edge solutions that enhance operational efficiency and profitability. Adept at leveraging emerging technologies like **LLMs** and **AutoML**, fostering collaboration, leading cross-functional teams, and effectively communicating complex insights to C-level executives.

### Key Achievements:

#### Business Impact

- Increased customer retention by **20%**, saving **\$10M annually** through predictive churn modeling.
- Boosted marketing ROI by **25%** using predictive analytics for campaign optimization.
- Reduced fraudulent activities by **30%** with advanced machine learning algorithms.
- Optimized recommendation systems, increasing user engagement by **40%**.

#### Technical Innovation

- Slashed cyber threat false positives by **74%** using ML models, significantly enhancing security accuracy.
- Improved ETL efficiency by **70%** leveraging **cloud solutions** (AWS, Azure) and **big data technologies** (HDFS).
- Implemented **federated learning** for privacy-preserving analytics across sensitive datasets.
- Leveraged **deep learning** models for image recognition, improving product categorization accuracy by **35%**.

#### Operational Excellence

- Reduced model deployment time from **weeks to days** by leading **MLOps practices** adoption.
- Improved decision-making speed by **40%** through **real-time analytics pipelines**.
- Developed and maintained scalable data pipelines, processing over **1TB of data daily** with **99.99% uptime**.
- Ensured **99.9% data accuracy** and regulatory compliance with robust data governance.

#### Leadership & Collaboration

- Led cross-functional projects, boosting success rates **35%** while aligning data science with business goals.
- Upskilled **50+ team members** through data science workshops and training sessions.
- Increased stakeholder trust and model adoption by **50%** using **explainable AI** techniques.

#### Additional Highlights

- Boosted data usability by **62%** with interactive dashboards using **D3.js, Tableau, PowerBI, and Plotly**.
- Implemented **NLP techniques** to analyze customer feedback, increasing satisfaction scores by **15%**.
- Developed automated reporting systems, saving **20+ hours per week** for the analytics team.
- Designed and conducted **A/B tests**, leading to a **20% improvement** in product feature adoption.
- Improved inventory management efficiency by **25%** using advanced **time series forecasting** models.
- Developed custom **Python libraries** for streamlining data preprocessing, reducing development time by **30%**.
- Leveraged **SAS Enterprise Miner** for customer segmentation, driving **14%** cross-sell increase and **10%** churn reduction.

Proficient in **Python, SQL, R**, and cloud platforms. Committed to driving innovation and efficiency through emerging technologies and data-driven strategies.

## Work Experience

**Senior Data Scientist** at [ABC Fitness Solutions](#) - Frisco, TX

*Feb 2023 – May 2024*

- Improved data accuracy by **15%** and reduced report generation time with **Python-driven checks**.
- Boosted user engagement by **5%** with a **machine learning prototype** for customer behavior analysis.
- Reduced ETL workflow processing time by **20%** with **Azure** and **Databricks** through collaboration with the data engineering team.
- Led **cloud migration planning**, providing detailed cost-benefit analyses and technical requirements.
- Improved team collaboration and reduced errors by **10%** with **version control** for analytical workflows.
- Developed a **predictive maintenance model** for fitness equipment, reducing downtime by **30%**.
- Implemented a **personalized workout recommendation system**, increasing member retention by **15%**.
- Created a **real-time dashboard** for monitoring gym occupancy and equipment usage, optimizing resource allocation.

**Environment:** Python, SQL, R, Azure, Tableau, PowerBI, Git, MLflow, scikit-learn, TensorFlow, Pandas, NumPy, PySpark, Airflow

**Senior Data Scientist** at [ArkoseLabs](#) - San Mateo, CA

*Sep 2021 - Feb 2023*

- Reduced false positives by **74%** and enhanced threat detection accuracy with **ML-based defense systems**.
- Increased model robustness against zero-day attacks by **30%** with a **synthetic data pipeline** using **LLMs**.
- Cut manual workload by **40%** and improved early threat detection by **25%** with a **real-time monitoring system**.
- Enhanced threat detection precision by **35%** through **pattern recognition** and **time series analysis**.
- Optimized intrusion detection with **federated learning**, preserving data privacy.
- Reduced model update cycle time from weeks to days by adopting **MLOps practices** with **AWS**.
- Developed an **anomaly detection system** for user behavior, reducing account takeover incidents by **50%**.
- Implemented a **real-time risk scoring model**, improving fraud prevention efficiency by **40%**.

**Environment:** Python, R, TensorFlow, PyTorch, scikit-learn, AWS SageMaker, Docker, Kubernetes, Git, MLflow, Pandas, NumPy, Plotly

**Data Scientist** at [USAA](#) - Plano, TX

*Feb 2019 - Sep 2021*

- Architected a scalable **Hadoop-based data lake**, reducing ETL time by **70%** and increasing data accessibility.
- Designed and deployed **20+ interactive Tableau dashboards**, improving data usability by **62%** and process visibility by **80%**.
- Developed prescriptive and predictive ML models for healthcare dashboards, reducing claim processing time by **25%** and fraudulent claims by **15%**.
- Spearheaded the adoption of **AutoML**, cutting model development time by **40%** and enabling non-technical teams.
- Leveraged **SAS** for predictive analytics, enhancing vulnerability detection and improving model accuracy by **18%**.
- Led a team in developing **customer churn prediction model**, improving retention by **20%** and saving **\$10M annually**.
- Pioneered **explainable AI** techniques (SHAP, LIME), enhancing model interpretability and increasing stakeholder trust by **35%**.
- Implemented a **customer lifetime value prediction model**, optimizing marketing spend and increasing ROI by **25%**.
- Migrated **70% of on-premise workloads (500TB)** to **AWS**, and with **Databricks** improving scalability and reducing costs.

**Environment:** Python, R, SQL, Hadoop, Tableau, SAS Enterprise Miner, AutoML, SHAP, LIME, Pandas, NumPy, scikit-learn, PySpark, Git

**Marketing Data Analyst** at [Liberty Mutual Insurance Company](#) - Portsmouth, NH

*Jan 2018 - Feb 2019*

- Engineered a **marketing analytics platform**, using advanced visualization, boosting campaign effectiveness by **20%** and customer acquisition by **12%**.
- Implemented **A/B testing frameworks**, increasing conversion rates by **15%** and optimizing budget allocation.
- Developed interactive dashboards with **D3, React, and Plotly**, reducing reporting time by **30%** and improving decision-making accuracy by **25%**.
- Utilized **SAS Enterprise Miner** for customer segmentation, increasing cross-sell by **14%** and reducing churn by **10%**.

- Created a **lead scoring model**, improving sales team efficiency and increasing conversion rates by **18%**.
- Developed a **customer journey mapping tool**, enhancing the understanding of touchpoints and improving customer experience.
- Implemented a **marketing mix modeling** approach, optimizing media spend across channels and increasing ROAS by **22%**.

**Environment:** Python, R, SQL, SAS Enterprise Miner, D3.js, React, Plotly, Tableau, Git, Pandas, NumPy, scikit-learn

**Associate Faculty** at [Columbia University](#) - New York, NY

*Mar 2017 - Dec 2017*

*Applied Analytics Department, School of Professional Studies*

- Instructed graduate-level course on **Python for Data Analysis**, covering advanced topics including GitHub workflows, market trend analysis, and machine learning algorithms, receiving an average student rating of **4.8/5**.
- Led 3 academic teams in the **BMW Innovation Challenge**, resulting in a **25% improvement** in clickstream analysis accuracy and optimized website user experience using **Hadoop clusters**.
- Developed and presented **R-Shiny dashboard reports** on social media analytics, engaging high-level technical audiences and industry professionals.
- Created and delivered workshops on advanced **data visualization techniques**, improving students' ability to communicate insights effectively.
- Mentored students on capstone projects, guiding them through real-world data science challenges and improving project success rates.

**Environment:** Python, R, R-Shiny, Hadoop, Git, Pandas, NumPy, scikit-learn, Matplotlib, Seaborn

**Data Visualization Intern** at [New York City Office of the Comptroller](#) - New York, NY

*Jun 2016 - Aug 2016*

- Enhanced NYC's asset management site using **D3.js and R**, creating interactive visualizations of key financial metrics.
- Developed responsive dashboards to improve budget transparency, contributing to an **18.47% increase** in public engagement within 3 months.
- Assisted in data cleaning and preprocessing using R, improving data accuracy for visualizations.
- Collaborated with financial analysts to identify and visualize important KPIs for city asset management.
- Implemented an **automated data update pipeline**, ensuring visualizations reflected the most current financial data.
- Created a user guide for maintaining and updating the visualization platform, ensuring long-term sustainability of the project.

**Environment:** R, D3.js, SQL, Git, ggplot2, dplyr, tidyr, Excel

**Process Analyst (Apprenticeship)** at [Stevens Institute of Technology](#) – Hoboken, NJ

*Sep 2015 – Feb 2016*

- Led process mapping for **DoD-affiliated SERC**, interviewing stakeholders and using **BPMN diagrams** and **Signavio simulations**.
- Improved resource efficiency by **24%** by optimizing reimbursement workflows for defense research projects.
- Implemented **Six Sigma DMAIC** and **Systems Architectural Diagrams**, enhancing financial processes and documentation standards.
- Developed "as-is" and "to-be" process maps with task probability simulations, driving data-based decision-making and change management.
- Created a knowledge base of best practices for process improvement, facilitating knowledge transfer across departments.
- Conducted training sessions on process mapping and analysis techniques, upskilling team members and improving overall process efficiency.

**Environment:** Signavio, Microsoft Visio, Excel, PowerPoint

**Business Analyst** at [Bajaj Allianz General Insurance](#) – Gandhinagar, GJ

*Feb 2013 – Dec 2014*

- Reduced average claim processing time by **10%** by analyzing high-cost claims data.
- Developed reporting dashboards using **SQL** and **SAP BW** contributing to an **8% reduction** in fraudulent claims.
- Implemented data cleaning protocols and standardized claim categorization, reducing misclassification errors by **15%** and improving data consistency for analysis.
- Created a **predictive model** for claim severity, allowing for more efficient resource allocation in claims processing.
- Developed and maintained a **KPI tracking system**, providing real-time insights into business performance for management.
- Conducted regular data quality audits, ensuring the integrity and reliability of insurance data used for decision-making.

**Environment:** SQL, SAP Business Warehouse, Excel, PowerPoint

## Certifications

- [IBM Machine Learning with Python](#)
- [Generative AI with Vertex AI](#)

## Skills

**Domain Expertise:** *Finance* (Risk Modeling, Customer Lifetime Value Prediction, Churn Analysis),  
*Cybersecurity* (Threat Detection, Anomaly Detection, Fraud Prevention),  
*Marketing* (Customer Segmentation, A/B Testing, Campaign Optimization),  
*Fitness & Health* (Personalized Recommendations, Engagement Analytics)

**Data Science & ML:** *Deep Learning* (TensorFlow, Keras, PyTorch),  
*ML Operations* (MLflow, Kubeflow), AutoML, LLMs

**Programming & Scripting:** *SQL, Python* (NumPy, Pandas, scikit-learn, PySpark, Flask), R, *Scala*

**Data Visualization & BI:** *PowerBI, Tableau*, D3.js, Plotly, Apache Superset

**Big Data & Cloud:** *Databricks, Apache Spark*, Hadoop, Snowflake, Alteryx,  
*AWS* (S3, EC2, SageMaker), *Azure* (Data Factory, Synapse), *GCP* (Vertex AI)

**Data Engineering:** *ETL/ELT, Data Pipelines*, Apache Airflow, ControlM, Stream Processing

**DevOps & Version Control:** Docker, Kubernetes, Git, GitHub Actions, CI/CD

**Development Environments:** Jupyter, VS Code, Databricks Notebooks

## Education

Stevens Institute of Technology – Hoboken, NJ, USA

**Master's in Information Systems** | Concentration: BPM & Service Innovation

**GPA: 3.7/4**

*Jan 2015 - Dec 2016*

Gujarat Technological University – Ahmedabad, GJ, India

**CPGA: 7.35/10**

**Bachelor's in Information Technology** | Certification: Microsoft Student Partners

*Aug 2010 - Jul 2014*