# **Portfolio content**

### **Education:**

I am currently pursuing a Master’s in Data Science at Rutgers University (September 2023 – May 2025) and am in my final semester. During my time here I have built expertise in machine learning, statistical modeling, and data mining, with a strong focus on AI, big data, time series predictions, and GIS applications. Previously, I earned my Bachelor of Engineering in Information Technology from Mumbai University (September 2016 – August 2020), where I developed a strong foundation in cybersecurity, machine learning, data structures, and algorithms, along with core IT and computer science concepts. I have also received a professional certification from IBM for data science, here is the link to the certificate: https://www.coursera.org/account/accomplishments/professional-cert/BZ4R7M2RWAHG

### **Experience:**

With a passion for data and problem-solving, I have worked across multiple domains, from predictive modeling to system architecture, always striving to enhance efficiency and drive data-driven decision-making. As a Statistics and Economics Intern at PGW (May 2024 – Dec 2024), I served as the sole Data Science intern in the customer programs department. My role involved developing predictive models and conducting statistical analyses on customer data to forecast energy burdens, monthly customer counts, and evaluate campaign effectiveness. I collaborated with cross-functional teams to gather meaningful insights, enabling data-driven communication strategies and actionable recommendations for stakeholders. Previously, as a Solutions Architect at Reliance Jio (Nov 2020 – Jul 2023), I was an integral part of the Design and Architecture team, responsible for developing scalable telecom solutions. I worked on systems for online SIM ordering, regional telecom integrations, and workflow automation. Additionally, I streamlined team operations by creating custom Python tools, significantly improving efficiency and reducing problem-solving time. In total I have 3+ years of work experience. 2 years 8 months as a solutions architect and 8 months as a statistics and economics intern.

### **Publications:**

I have contributed to research in sentiment analysis and healthcare analytics, leveraging machine learning to tackle real-world challenges. Sentiment Analysis and Classification on Twitter Spam Account Dataset (Jul 2020) – Developed a sentiment analysis model using VADER and a Random Forest classifier to accurately identify Twitter spam accounts, achieving 94% accuracy. It was published in IEEE Xplore, here is the link: <https://ieeexplore.ieee.org/document/9213206>. I have also done research focusing on ensemble classifiers and their applications in the field of diabetes detection. Type-II Diabetes Detection using Decision-Tree Based Ensemble of Classifiers (Sept 2019) – Investigated the application of Decision Tree-based ensemble classifiers for Type-II diabetes detection, achieving a maximum accuracy of 93.5% using XGBoost.It was published in IEEE Xplore, here is the link:http://ieeexplore.ieee.org/document/9129348/

### **Projects at Philadelphia Gas Works (PGW)**

During my tenure as a Statistics and Economics Intern at PGW, I led multiple data-driven projects aimed at optimizing customer engagement, forecasting energy burdens, and improving financial assistance programs. My work involved developing predictive models, conducting statistical analyses, and designing interactive dashboards to support data-driven decision-making. I collaborated with cross-functional teams to translate business challenges into actionable insights, ultimately enhancing operational efficiency and customer outreach.

As part of my internship at PGW, I developed a Monthly Customer Enrollment Forecasting Model using time-series analysis to predict customer enrollments for financial assistance programs. This project provided the company with a reliable forecasting tool to optimize budget planning and resource allocation. By leveraging the Prophet model, I incorporated seasonality and external regressors, achieving a mean absolute percentage error (MAPE) of less than 10%. The final results were delivered through an interactive Excel dashboard, allowing stakeholders to make informed financial and operational decisions.

Another key project I completed at PGW was an Analysis of Low-Income Customer Behavior, where I examined participation trends in financial assistance programs. Using Python and Tableau, I performed exploratory data analysis (EDA) on a dataset of over 50,000 customers, identifying key demographic and behavioral factors affecting program engagement. This analysis helped PGW refine outreach strategies and prompted a follow-up customer survey to address participation barriers.

To further enhance financial assistance targeting, I worked on Integrating City Government Datasets to identify potential candidates for financial aid programs. By researching and merging publicly available datasets from sources like Census data and OpenDataPhilly, I helped PGW proactively identify eligible customers, ensuring better financial aid distribution. The results enabled more effective outreach campaigns and improved customer enrollment in assistance programs.

I also led an Energy Burden Forecasting and Analysis project during my internship at PGW, aimed at identifying disparities in energy usage and energy burden across different customer segments. I integrated city-provided property assessment data with PGW’s customer records using SQL, allowing a comprehensive analysis of energy consumption patterns based on factors like construction year, living area, and building conditions. I built an interactive Tableau dashboard that provided energy burden forecasts and helped optimize program performance and resource allocation.

Another significant project I spearheaded at PGW was a Survey to Explore Financial Assistance Barriers. Based on insights from my low-income customer behavior analysis, I designed and analyzed a survey conducted among 900 customers in collaboration with Bellomy Research. This survey assessed program awareness, effectiveness, and customer engagement barriers. Using Python and Tableau, I processed and analyzed the survey data, uncovering ineffective communication channels and highlighting preferred outreach methods, leading to improved customer engagement strategies.

Lastly, I conducted an Analysis of PGW’s Recertification Strategy, which examined the annual recertification process required for continued enrollment in low-income programs. Using SQL and Tableau, I developed a dashboard to track customer retention, identify high drop-off points, and visualize future recertification volumes for better resource planning. This analysis helped PGW optimize outreach efforts, improve program efficiency, and enhance customer retention.

These projects, completed during my internship at PGW, leveraged machine learning, statistical modeling, SQL, Python, and Tableau to improve operational efficiency, enhance customer engagement, and optimize financial assistance programs. My contributions played a crucial role in helping the organization make data-driven decisions and refine strategic initiatives.

### Projects at Reliance Jio Infocomm Ltd.

### During my tenure as a Solutions Architect at Reliance Jio (Nov 2020 – Jul 2023), I was part of the Design and Architecture team, where I played a crucial role in developing scalable telecom solutions. My work ranged from designing customer-facing systems to integrating new telecom regions and automating internal processes using Python-based solutions. I also collaborated across teams to ensure seamless system implementation while optimizing workflows and improving operational efficiency.

### As part of my role at Reliance Jio, I designed and implemented the On-Demand SIM Activation and Doorstep Delivery System, which allowed users to order an activated SIM card online and have it delivered to their doorstep. This was the first such system in the telecom industry, providing a competitive advantage by enhancing accessibility for customers. I conducted market research, evaluated competitor solutions, and collaborated with development teams to integrate fraud prevention mechanisms using computer vision models. I also presented the system architecture to over 500 stakeholders, ensuring seamless implementation. The project was a major success, making Jio the first telecom provider to offer this service.Here is a link to view this work: https://www.jio.com/selfcare/selfkyc/

### Another major project I led was the Central Coupon Management System, which standardized the handling of promotional campaigns, discount vouchers, and customer onboarding incentives. Previously, campaign management was fragmented, making it difficult to track and validate historical coupons. I analyzed past campaigns, identified API inconsistencies, and designed a unified, scalable system that ensured consistency and efficiency. Using SOAP UI, Postman, and SQL, I tested and optimized API workflows for seamless data exchange. The result was a centralized system that simplified coupon tracking, validation, and campaign integration, enhancing marketing efficiency and reducing operational complexity.

### To improve operational efficiency, I developed several Python-based automation tools, addressing repetitive and error-prone tasks such as invoice validation, data extraction from PDFs, and bug tracking. These tools utilized OpenCV, tabula-py, and PyMuPDF to extract and structure data from invoices, significantly reducing manual effort and analysis time. One of these programs successfully detected 90% of defective invoices, preventing penalty costs from external audits and improving overall compliance. The automation significantly streamlined testing workflows and improved team productivity.

### In an effort to expand JioPhone adoption among younger demographics, I worked on Youth Customer Onboarding and JioPhone Sales Integration. JioPhone, a product extensively used in rural areas, needed an improved system to authenticate younger customers. I integrated Hyperverge's image recognition technology to validate identity documents and ensure smooth onboarding. I also maintained and updated the JioPhone sales system, ensuring it remained aligned with company workflows. The system continues to be a crucial revenue driver and is particularly popular among rural and elderly customers. Here is the link to view his work: https://www.jio.com/en-in/jiophone

### These projects, completed during my time at Reliance Jio, leveraged system architecture, API integrations, automation, and machine learning to enhance customer experience, streamline operations, and optimize campaign management. My contributions played a significant role in expanding Jio’s services, reducing operational inefficiencies, and improving telecom solutions at scale.

### 

### 

# 

### **Personal Projects**

In addition to my professional work, I have undertaken several personal projects that explore a wide range of machine learning, deep learning, NLP, and statistical modeling techniques. These projects showcase my ability to integrate data science methodologies with real-world applications, spanning domains such as recommendation systems, adversarial attacks, medical analytics, sentiment analysis, and career insights.

As part of my personal work, I built NextBuys.co, a recommendation engine that enhances product suggestions using Amazon transaction data. By analyzing 1.8 million transactions across 1,800 product categories, I implemented the FP-Growth algorithm to identify frequently co-purchased items and cosine similarity metrics to refine product categorization. Additionally, I designed a customer feedback loop that adjusts recommendations based on user interactions. This project combined machine learning and web development, with the recommendation engine built in Python and SQL, and the front-end developed using HTML, CSS, and JavaScript. The system successfully delivers personalized product recommendations, optimizing user engagement. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/nextbuys\_details.html

Another project, A Study on Adversarial Attacks and Defenses in Deep Learning Models, explores security vulnerabilities in AI systems. I implemented Fast Gradient Sign Method (FGSM) attacks on a PyTorch-based computer vision model for vehicle damage assessment, reducing its accuracy from 77% to 16%. To counteract this, I explored Gaussian Noise Augmentation and adversarial training, which successfully restored model robustness. This project highlights the importance of securing AI applications, demonstrating how adversarial defenses can mitigate model manipulation risks. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/inferenceattack\_details.html

In the medical analytics domain, my project Assessing Heart Failure Risk: Statistical Modeling with Logistic Regression leverages statistical modeling to predict heart failure survival. By performing exploratory data analysis (EDA) and applying stepwise feature selection, I built a logistic regression model that provided interpretable results. The model identified key risk factors such as age, ejection fraction, and creatinine levels, highlighting their impact using odds ratios. This project underscores how simple statistical models can still provide powerful and interpretable insights in healthcare analytics. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/regerssionmodel\_details.html

For natural language processing, my project Deep Learning for Sentiment Analysis: IMDb Movie Reviews with LSTMs investigates the effectiveness of bi-directional LSTM networks in sentiment classification. I preprocessed the dataset using one-hot encoding and designed a sequential neural network with multiple LSTM layers, followed by a sigmoid activation function for binary classification. The model was trained using TensorFlow and Keras, achieving strong performance on accuracy, precision, recall, and F1-score. This study highlights the power of deep learning in NLP applications, paving the way for future improvements with transformer-based models. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/lstmnlp\_details.html

Another personal project, ML-Driven Industry Analytics: LinkedIn, BERT, and GIS in Action, provides personalized career insights based on LinkedIn user data. Using the LinkedIn API, I extracted user industry information, analyzed salary trends and job openings using Bureau of Labor Statistics (BLS) data, and implemented BERT embeddings with cosine similarity to match users to relevant industries. The system also visualized geographic industry trends using GeoPandas and Folium, providing interactive insights through a Plotly Dash dashboard. This project demonstrates how NLP, machine learning, and GIS technologies can be combined to create AI-powered career advisory tools, offering actionable insights for workforce analytics. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/monty\_details.html

These personal projects reflect my ability to apply advanced data science techniques across diverse domains, from recommendation systems and AI security to medical analytics, NLP, and career forecasting. They emphasize my commitment to solving real-world problems using data-driven insights and innovative machine learning approaches.