# **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. More information about my experience can be found in my portfolio website here: https://gauravshetty98.github.io/portfolio/experience.html

### 

### **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. More information here: https://gauravshetty98.github.io/portfolio/experience.html

Monthly Customer Enrollment Forecasting Using Time-Series Analysis.  
Project type: Project done for PGW during my time as a statistics intern  
Objective: To develop a model to help with budgeting and resource allocation by forecasting the monthly number of customers enrolling in specific programs. Business Impact: The model provided a reliable way to predict monthly enrollments, enabling the organization to plan budgets, allocate resources more effectively, and optimize programs. By accounting for seasonality and external factors, it significantly improved the accuracy of quarterly forecasts and supported smarter decision-making. Methodology: I started with exploratory data analysis (EDA) in Python to identify trends, seasonal patterns, and external factors influencing enrollments. After cleaning and preprocessing the data, I used the Prophet time series model to capture seasonality, achieving a mean absolute percentage error (MAPE) of less than 10%. I fine-tuned the model iteratively by adjusting hyperparameters and incorporating external regressors for improved accuracy. Tools: I cleaned and preprocessed the data using Python and SQL, then developed the model in Python. The final results were delivered through an interactive Excel dashboard, providing easy access and usability.

Challenges: The most challenging aspect of the project was building the model. I addressed this by conducting extensive research, experimenting with various models, and fine-tuning them iteratively to optimize performance and ensure the best results. Results: The model was well-received by the team for its accuracy, accessibility, and ability to incorporate external factors into projections. These projections were reviewed during monthly meetings to evaluate customer growth and optimize budget allocation for discount programs. More information here: https://gauravshetty98.github.io/portfolio/experience.html

Analysis of Low-Income Customer Behavior  
Project type: Project done for PGW during my time as a statistics intern  
Objective: The company wanted to analyze the effectiveness and participation rates of various financial assistance programs and identify the major factors affecting these participation rates. Business Impact: The analysis helped the organization gain a deeper understanding of its low-income customer base and address critical business questions, such as identifying factors affecting participation rates, evaluating the effectiveness of existing communication strategies, uncovering barriers to program engagement, and determining correlations between demographic or behavioral patterns and program participation. Methodology: I conducted exploratory data analysis (EDA) on a dataset of over 50,000 customers, utilizing pivot tables, scatter plots, histograms, and data grouping to uncover patterns and trends. Correlation analysis was performed to identify key relationships between demographic factors and program participation rates. The data was cleaned and preprocessed in Python, with insights visualized in Tableau. Tools: I utilized Python and SQL for data cleaning and preprocessing, with analysis and visualizations created in Tableau. Challenges: Cleaning and merging data from multiple sources was a key challenge, which I effectively addressed using Python's pandas library. Results: The analysis revealed various factors, such as age and assistance type, that affected participation rates. It also raised new questions about whether current outreach methods were effective, forming the basis for a survey conducted among customers to explore similar questions.

Integrating City Government Datasets to Identify Potential Candidates for Financial Assistance Programs  
Project type: Project done for PGW during my time as a statistics intern  
Objective: As an intern, one of my tasks was to research publicly available datasets that could help identify potential customers for financial assistance programs. Business Impact: This initiative ensured early recognition of low-income customers, enabling better financial aid distribution, reducing payment delinquencies, and optimizing the utilization of government-supported program budgets Methodology: I researched publicly available datasets from sources like the Census and OpenDataPhilly. Using these sources, Performed EDA on these sources to locate low-income housing locations and identify potential customers for targeted outreach. Tools: I preprocessed the data using Python to standardize and align it with the company’s formatting before merging it with our customer data. Challenges: The main challenge was identifying publicly available datasets that addressed key organizational issues while ensuring data quality. I tackled this by prioritizing the organization's challenges, researching datasets relevant to the top problems, and focusing on government-provided or well-documented data to maintain quality standards. Results: I delivered a list of potential leads for low-income programs, enabling the team to execute targeted outreach campaigns and inform customers about financial assistance programs they might qualify for. .More information here: https://gauravshetty98.github.io/portfolio/experience.html

Energy Burden Forecasting and Analysis  
Project type: Project done for PGW during my time as a statistics intern  
Objective: The purpose of this project was to identify disparities in energy usage and energy burden between customer programs, forecast future burdens, and provide actionable insights through a user-friendly dashboard. Business Impact: This tool provided a deeper understanding of customer energy usage, helping the organization reduce operational inefficiencies, enhance program performance, and better support customers with high energy burdens. Methodology: I integrated city-provided property assessment data with customer energy usage records using SQL, enabling an analysis of energy consumption alongside housing attributes such as construction year, living area, and building conditions. Tools: I developed an interactive Tableau dashboard to visualize energy usage trends across customer programs. The dashboard provided detailed comparisons based on housing attributes like construction year and living area. Additionally, I incorporated a feature to estimate energy burden using historical averages, total customer counts, and customizable group definitions for tailored analysis. Challenges: The biggest challenge was creating an interactive Tableau dashboard that directly addressed business needs while allowing users the flexibility to explore the data further. I overcame this by studying effective dashboard designs, collaborating closely with senior team members to understand stakeholder priorities, and organizing the dashboard into clear, functional sections. Using SQL and Tableau, I incorporated essential features and ensured the tool was intuitive and impactful. Results: I created a comprehensive Tableau dashboard that offered insights into energy usage patterns, facilitated comparisons between customer programs, and provided forecasting capabilities for energy burden, significantly improving program analysis and decision-making. More information here: https://gauravshetty98.github.io/portfolio/experience.html

Survey to Explore Financial Assistance Barriers  
Project type: Project done for PGW during my time as a statistics intern  
Objective: Previous analysis revealed age as a key factor affecting participation rates in financial assistance programs. To further explore this and address broader questions surrounding low participation rates, I conducted a survey to assess program awareness, effectiveness, and the impact of various factors on engagement. Business Impact: The survey provided critical insights into customer perceptions of financial assistance programs, the effectiveness of existing outreach campaigns, and areas for improvement. By addressing key business questions, it enabled the organization to refine awareness efforts, enhance program engagement, and optimize communication strategies to better meet customer needs. Methodology: I spearheaded the survey with the help of Bellomy Research among 900 customers. The survey posed questions related to program awareness, preferred communication methods, customer satisfaction, and other relevant topics. Tools: I merged the survey responses with customer data using Excel and derived analytical insights using Tableau and Python. Challenges: Formatting the survey responses for analysis was a significant hurdle, particularly because much of the data was free text. I addressed this by using Excel and Python to preprocess and organize the responses into a structured format, ensuring they were ready for analysis. Results: The analysis identified ineffective communication channels and highlighted preferred outreach methods for different customer groups. These findings enabled the organization to refine awareness campaigns, introduce new outreach strategies, and improve overall program engagement and reach. More information here: <https://gauravshetty98.github.io/portfolio/experience.html>

Analysing the effectiveness of our recertification strategy  
Project type: Project done for PGW during my time as a statistics intern  
Objective: The objective of this project was to analyze the annual recertification process required for customers to remain enrolled in low-income programs, identify stages with the highest drop-off rates, assess its effectiveness, and provide insights to help the organization strategize and forecast future customer retention. Business Impact: The analysis provided a deeper understanding of the recertification process, helping the organization reduce customer attrition, improve the efficiency of their programs, and better anticipate recertification volumes to allocate resources effectively. Methodology: Data preprocessing was done using SQL and data visualization was done using Tableau. Developed a dashboard to track customer retention through each stage, identify drop-off points, and visualize future recertifications. Tools: I merged the survey responses with customer data using Excel and derived analytical insights using Tableau and Python. Challenges: The biggest challenge was visualizing the recertification process, as its 45-day timeline varied for each customer, making it difficult to showcase monthly drop-offs. Unclean data further complicated the task. I addressed this by performing data aggregation and grouping customers by their stage and remaining days in the process, enabling clear and interpretable visualizations that highlighted monthly drop-offs effectively. Results: The dashboard helped in identifying stages that faced the most drop-off thus improving the overall process, it also provided details about upcoming recertifications helping the organization in better planning. More information here: <https://gauravshetty98.github.io/portfolio/experience.html>

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 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. More information here: <https://gauravshetty98.github.io/portfolio/experience.html>

On-Demand SIM Activation and Doorstep Delivery System  
Project type: Project done for Reliance Jio during my time as a solutions architect   
Objective: The company wanted a system that would allow users to order an activated SIM card directly from the company website or mobile app and have it delivered to their doorstep. Business Impact: This system aimed to increase the reach of our products and make it easier for customers to place orders. As no similar system was available in the market at the time, it gave our product a competitive advantage. Methodology: Researched competitor solutions to identify shortcomings and customer-preferred workflows, guiding improvements in system design. Collaborated with development teams to assess available APIs and system integrations. After extensive brainstorming, designed a secure, end-to-end system that adhered to data privacy and security guidelines. Integrated advanced computer vision models for fraud prevention using face detection. Presented the system architecture to over 500 stakeholders and actively oversaw development and deployment, providing continuous support to resolve issues and ensure a smooth rollout. Tools: I used SOAP UI and Postman to test APIs across existing systems, ensuring functionality and reliability. SQL was utilized to query databases and validate data flow, ensuring consistency and accuracy. For issue tracking and management, I leveraged Azure DevOps, streamlining debugging and collaboration across teams. Challenges: One of the main challenges was the scale of the project and the number of stakeholders involved. Designing a solution that met everyone’s requirements was difficult. We addressed this by holding regular meetings with different teams to provide constant updates on the system and gather feedback. Additionally, we ensured that every stakeholder understood the priority and importance of various requirements, facilitating a smooth design and development cycle. Results: The system was deployed in record time, ahead of any other telecom provider, making our company the first to offer such a service at the time. Check it out for yourself: Here is a link to view this work: https://www.jio.com/selfcare/selfkyc/ and here is the link to know more about the project: https://gauravshetty98.github.io/portfolio/experience.html

### Central Coupon Management System Project type: Project done for Reliance Jio during my time as a solutions architect Objective: The organization ran multiple campaigns to onboard customers with various coupons and discount vouchers but lacked a central system to manage them, making it difficult to access historical data or use existing APIs. Business Impact: A centralized system would standardize campaign management, reducing complexity and ensuring consistency across all campaigns. It would also simplify coupon generation, management, and validation for on-field employees for all campaigns Methodology: I conducted a technical analysis of historical campaigns to identify commonalities in API structures, coupon generation mechanisms, and data handling. Leveraging existing APIs, I designed a scalable system that could manage past campaigns while seamlessly integrating new ones. Using SOAP UI and Postman, I tested and optimized API workflows to ensure smooth data exchange. SQL was employed for data validation, ensuring accuracy in coupon tracking and campaign records. The system was built with modularity in mind, enabling flexibility for future campaign integrations while maintaining consistency and security in coupon management. Tools: I used Soap and Postman to test and design APIs. SQL was used to do data validation and testing. Challenges: The biggest challenge was analyzing historical campaigns and consolidating diverse coupon generation systems into a unified framework. Each campaign had its own technical flow, APIs, and validation mechanisms, requiring extensive research to identify commonalities and standardize processes. Additionally, collaborating with multiple teams—including marketing, campaign management, and development—was complex, as each had different requirements and priorities. Balancing these inputs while ensuring system compatibility and scalability required careful negotiation, iterative refinement, and rigorous API testing. By thoroughly reviewing documentation, consulting senior members, and systematically testing integrations, I was able to develop a cohesive solution that met all business and technical needs. Results: The result of this project was the successful development of a centralized campaign management system that streamlined coupon generation, validation, and tracking across all customer acquisition campaigns. The system standardized workflows, reducing operational complexity and ensuring seamless integration with existing APIs. It enabled marketing and on-field teams to efficiently manage and retrieve historical campaign data, improving decision-making and reducing redundancy in new campaigns. The system became a foundational tool for future campaigns, significantly improving efficiency and scalability in promotional efforts. More information here: <https://gauravshetty98.github.io/portfolio/experience.html>

Python Programs for Operational Efficiency  
Project type: Project done for Reliance Jio during my time as a solutions architect  
Objective: I identified several areas within the company that could be improved through Python automation, such as reducing penalties from faulty invoices and addressing data extraction issues while fixing bugs. I developed Python-based tools to streamline these processes and effectively resolve these challenges. Business Impact: The Python-based programs significantly enhanced efficiency by automating data extraction from invoices and PDFs, reducing manual effort and analysis time. The software also identified defective invoices, minimizing penalties from external audits and improving compliance. Methodology: The approach focused on identifying repetitive tasks that consumed significant time for the team and determining whether a software solution could perform them more efficiently. After developing the tools, I presented them to the team, provided thorough documentation on their usage, and included guidelines for future modifications. Tools: The tools were developed in Python, with Azure DevOps used for version control. Since the primary focus was working with PDF - extracting data and storing it in a structured format, libraries such as OpenCV, tabula-py, and PyMuPDF were utilized. Results: The program detected 90% of defective bills, significantly reducing penalty costs for the company. It also enabled efficient extraction of text and image data from invoice PDFs, cutting analysis time for the testing team. More information here: https://gauravshetty98.github.io/portfolio/experience.html

### Youth Customer Onboarding and JioPhone Sales Integration Project type: Project done for Reliance Jio during my time as a solutions architect Objective: The organization wanted to improve the accessibility of our services and our products like JioPhone by introducing them to younger demographics. The current system didn’t have proper checks in place to cater to such needs. Business Impact: Since JioPhone was a product that was extensively used in rural areas, such a system would increase the reach of an already popular product. Methodology: I integrated the existing system with image recognition technology using Hyperverge to authenticate younger customers through valid identity documents. Additionally, I maintained and updated the JioPhone sales system to ensure it remained aligned with our other workflows. Tools: I primarily used Visio and UML to visualize the system architecture and data flow. For data validation and testing, I utilized SQL, Postman, and SOAP. Challenges: Understanding the JioPhone system was challenging due to its complexity and age. To overcome this, I thoroughly reviewed the available documentation, sought guidance from senior team members, and conducted extensive testing of the system and its API calls. This hands-on approach helped me gain a comprehensive understanding of the system's logic and functionality. Results: The system and the product are still an integral part of the company’s revenue and are crucial, especially among aged and rural customers. Check it out here: JioPhone. Here is the link to view his work: <https://www.jio.com/en-in/jiophone> and here is the link to know more about the project: <https://gauravshetty98.github.io/portfolio/experience.html>

### 

# 

### **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. More information here: <https://gauravshetty98.github.io/portfolio/projects.html>  
  
  
NextBuys.co - A Recommendation Engine  
Project type: Personal project done during my time as a masters student in rutgers university  
NextBuys.co is a recommendation engine system that enhances product recommendations using Amazon transaction data. Leveraging a dataset of over 1.8 million transactions across 1,800 product categories, the system preprocesses data to identify multi-item purchases and applies cosine similarity to refine product categorization. Algorithms like FP-Growth and metrics like cosine similarity have been used to optimize these recommendations. A customer feedback loop to change recommendations based on user input. Objective: To build a recommendation system capable of suggesting items based on purchase trends and item similarity. Key Focus: Enhancing user experience with tailored product recommendations. Methodology: Utilized FP-Growth algorithm to identify items that are frequently bought together and used metrics such as lift and confidence to refine these suggestions. Implemented cosine similarity to identify items that are similar to each other. Also designed a customer feedback loop to let the users like or dislike the recommendations based on which they are altered. A feedback loop was implemented to adjust recommendations based on user preferences, refining suggestions dynamically. Tools & Technologies: Python for data processing and algorithm implementation. SQL for data storage and retrieval. HTML, CSS, and JavaScript for the front-end interface. Outcome: Successfully developed an end-to-end recommendation system, blending machine learning with web development to provide an intuitive user experience. 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

Intelligent Portfolio Chatbot: AI-Powered Interactive Resume  
Project type: Personal project done during my time as a masters student in rutgers university  
This project is an AI-powered chatbot integrated into my portfolio website, designed to provide visitors with an interactive way to learn about my experiences, skills, and projects. Unlike a static resume, this chatbot leverages Retrieval-Augmented Generation (RAG) with an LLM model to answer questions dynamically based on structured information from my portfolio and additional documents. The goal is to enhance engagement and make my portfolio more informative and interactive. Objective: The primary objective of this project is to create a seamless and intelligent interface that allows recruiters, collaborators, and visitors to explore my professional background through natural language interactions. Instead of navigating through multiple pages, users can simply ask questions about my projects, skills, or experiences and receive accurate, well-structured responses. Key Focus: The key focus of this chatbot is to provide precise, contextually relevant answers about my work. It integrates structured and unstructured data sources, ensuring that responses are both informative and aligned with my professional experience. Additionally, it showcases my expertise in AI, NLP, and Retrieval-Augmented Generation while demonstrating my ability to deploy AI-powered applications in real-world scenarios. The chatbot was built by first creating a structured document containing comprehensive details about my projects, experiences, publications, and other professional work. This document served as the primary knowledge base. Using Python, I preprocessed the data by segmenting it into meaningful textual chunks to ensure efficient retrieval. These chunks were then converted into vector embeddings using all-MiniLM-L6-v2, allowing semantic similarity searches. To store and index these embeddings efficiently, I implemented FAISS (Facebook AI Similarity Search), enabling fast and scalable nearest-neighbor searches for retrieving relevant content based on user queries. For response generation, I integrated Gemini-2.0-Flash, a lightweight and efficient LLM model, to generate accurate answers. The core retrieval workflow involves taking a user query, using FAISS to fetch the most relevant information from the indexed database, and then passing both the query and retrieved data to the LLM for response generation. To make this accessible, I built a FastAPI backend, exposing this functionality as an API, which was then deployed on Render. The frontend, designed using HTML, CSS, and JavaScript, serves as an intuitive user interface where users can input their queries, interact with the chatbot, and view dynamically generated responses. Tools & Technologies: Used Python for preprocessing and segmenting textual data into meaningful chunks. Converted text chunks into embeddings using all-MiniLM-L6-v2 for semantic similarity search. Implemented FAISS (Facebook AI Similarity Search) for indexing and retrieving relevant information efficiently. Integrated Gemini-2.0-Flash as the LLM for generating responses based on user queries. Developed a FastAPI backend to expose the chatbot functionality as an API and deployed it in Renfer for cloud-based accessibility. Designed an interactive frontend using HTML, CSS, and JavaScript for user interaction. Outcome: The chatbot significantly enhances user experience by providing a conversational and engaging way to explore my work. It allows visitors to receive tailored responses without manually searching through different sections of the portfolio. This project not only improves accessibility but also serves as a demonstration of my expertise in AI, NLP, and chatbot development. It stands as a practical example of how AI can be leveraged to make personal branding more interactive and insightful. Here is a link to view the project in more details including website and github codebase link : https://gauravshetty98.github.io/portfolio/gauravgpt\_details.html

A Study on Adversarial Attacks and Defenses in Deep Learning Models  
Project type: Personal project done during my time as a masters student in rutgers university  
This project explores inference attacks on machine learning models, highlighting both attack strategies and defense mechanisms. As machine learning becomes increasingly prevalent, it is crucial to recognize that these systems are also vulnerable to adversarial attacks. Using a PyTorch-based computer vision model for vehicle damage assessment, we demonstrate how an attacker can exploit weaknesses using the Fast Gradient Sign Method (FGSM) to significantly degrade model accuracy. Additionally, we showcase defensive techniques such as Gaussian Noise Augmentation and adversarial training to mitigate these attacks and restore model performance. This project underscores the importance of robust AI security in real-world applications. This project aims to analyze the impact of adversarial attacks on deep learning models, focusing on inference attacks and defense mechanisms in computer vision applications. The objective is to highlight the vulnerabilities of machine learning models to adversarial manipulation and evaluate countermeasures to ensure secure and reliable AI systems in real-world applications Key Focus: Observing how vulnerable are the machine learning models against these attacks and the effectiveness of defense mechanisms. Methodology: The project follows a structured approach, beginning with the development and training of a computer vision model for vehicle damage assessment using PyTorch. After establishing a baseline accuracy of 77%, we implement FGSM attacks to craft adversarial samples and assess their impact on model performance. The model performance dropped to 16% based on these attacks. We also look into defensive techniques such as Gaussian Noise Augmentation, introducing controlled noise during training, and implement adversarial training, which retrains the model with adversarial examples to improve its resistance. The effectiveness of these defenses is measured by evaluating key performance metrics, including accuracy degradation, robustness improvements, and adversarial resistance. Visualization techniques such as confusion matrices and perturbation heatmaps further illustrate the attack and defense dynamics. Tools & Technologies: Utilized the ART library in python to program the attacks. PyTorch and Torchvision to build the convolutional neural network. LIME for model explainability. Outcome: This study demonstrates how even machine learning models are exposed to security risks like these attacks which can degrade the model performance. It also shows ways to defend against these attack emphasizing the necessity of adversarial defense mechanisms in securing AI applications. Here is a link to view the project in more details including github codebase link :https://gauravshetty98.github.io/portfolio/inferenceattack\_details.html

Deep Learning for Sentiment Analysis: IMDb Movie Reviews with LSTMs  
Project type: Personal project done during my time as a masters student in rutgers university  
This project explores sentiment analysis on a movie reviews dataset using bi-directional Long Short-Term Memory (LSTM) networks, showcasing the application of deep learning in Natural Language Processing (NLP). The model is designed to analyze the sentiment of movie reviews, leveraging one-hot encoding for vectorization to capture contextual dependencies in text data. A sequential neural network with multiple LSTM layers and a dense output layer using a sigmoid activation function was implemented for classification. The project examines the network architecture and its performance, highlighting the effectiveness of recurrent neural networks (RNNs) in sentiment analysis. Objective: The aim of this project was to explore the applications of deep learning algorithms like Long Short-Term Memory networks in the field of Natural Language Procesing. The study demonstrates the use of RNNs and their performance in classifying whether a particular moview review was positive or negative.Key Focus: Implementing a recurrent neural network and comparing its performance with other architectures for NLP tasks. Methodology: The dataset is preprocessed through one-hot encoding to transform text into numerical vectors, allowing the LSTM model to effectively capture word dependencies. A bi-directional LSTM architecture is implemented with multiple layers followed by a sigmoid activation function for binary classification. The model is trained and evaluated using standard metrics such as accuracy, precision, recall, and F1-score, while techniques like dropout regularization are applied to prevent overfitting. The model performed comparatively better than lexicon or rule based approaches like VADER. Tools & Technologies: Python for data processing and algorithm implementation. Tensorflow, Keras and Scikit-learn for model implementation and evaluation. Github for version control. Outcome: The study underscores the importance of deep learning in NLP applications, particularly in understanding sentiment and context within textual data. The findings suggest that LSTMs are well-suited for sentiment analysis, paving the way for further improvements using attention mechanisms or transformer-based models.Here is a link to view the project in more details including github codebase link <https://gauravshetty98.github.io/portfolio/lstmnlp_details.html>

ML-Driven Industry Analytics: LinkedIn, BERT, and GIS in Action  
Project type: Personal project done during my time as a masters student in rutgers university  
This project aimed to provide industry analytics based on a user's LinkedIn profile, offering insights into salary trends, job openings, and industry shifts pre- and post-COVID. A dashboard built with Plotly Dash integrated the LinkedIn API to fetch user data and analyzed Bureau of Labor Statistics datasets. Vectorization techniques and cosine similarity were used to identify similar industries, enhanced by Hugging Face’s BERT model for improved embeddings. GeoPandas helped in visualising geographic trends for those industries. The result? A data-driven, interactive tool that delivers personalized insights into industry projections, leveraging data science techniques and public datasets to inform career decisions. Objective: The project aimed to leverage machine learning and NLP techniques to deliver personalized career insights by analyzing data from the Bureau of Labor Statistics and the LinkedIn API. The goal is to offer insights into salary trends, job openings, and industry shifts before and after COVID-19. Key Focus: Combining concepts across various field to create a dashboard to show personalized industry analysis. Methodology: The project utilizes a Plotly Dash-based interactive dashboard to visualize the analysis. User details fetched from Linkedin API are preprocessed before applying NLP techniques to find user’s industry data from Bureau Of Labour statistics. A combination of word embeddings derived through BERT and cosine similarity metric is used to identify the user’s industry from the BLS database. In-depth analysis is done on BLS data to showcase different kinds of trends in industry including geographic hotspots for the industry with the help of GIS libraries like GeoPandas and Folium. Tools & Technologies: Linkedin API was used to fetch user details in real time. Libraries like BERT and Cosine similarity were used for text processing. Visualization libraries like Plotly and GIS libraries like GeoPandas, Folium were utilized for graphs and charts. The website was created using Plotly Dash. The dashboard was created using Plotly Dash. Outcome: The study helped in creating a system for providing career insights based on a users industry and displayed how different Data Science techniques can be used in real world applications. It highlights how these techniques can be used in transforming labor market insights into actionable career decisions. This project lays the foundation for AI-powered career advisory tools, demonstrating the potential of data science in workforce analytics. Here is a link to view the project in more details including 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.

### **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/