

GAURAV THORAT

United States

 682-563-0193

✉ gauravthorat1998a@gmail.com

 LinkedIn

 GitHub Portfolio

Education

University of Texas at Arlington

Masters of Science in Data Science

Jan. 2022 – Dec. 2023

Arlington, TX

Mumbai University

Bachelor of Engineering in Computer Engineering

Jun. 2017 – Jul. 2021

Mumbai, India

Experience

Bank of America

Data Scientist

Sep. 2023 – Present

Remote, USA

- Working on leading efforts in Cyber Security, focusing on "Analyzing CTI Reports" and "Testing LLM Detectors". I've applied Python, TensorFlow, and NLP to analyze open-source LLM applications, improving model accuracy and efficiency.
- In Analyzing the CTI Reports, used NLP, Python, and frameworks like TRAM and FACTOR to assess the authenticity of CTI reports and develop better detectors. Increased the TRAM model's accuracy to 55% through fine-tuning and improved decision-making processes by 15% through proxy-tuning and token synthesizing.
- I've also generated dataset from authenticated CTI Reports and developed LLM detectors and text generators, leading to a 28% improvement in training efficiency and a 20% enhancement in data quality by A/B Testing. My work has resulted in a new LLM text generator and detector with an 86% accuracy rate, outperforming market benchmarks by 12%, currently implementing Perplexity Analysis to improve model accuracy and creating a more comprehensive dataset for new models, contributing to the wider adoption of LLM technologies for text analysis applications.

UTARI (UTA Research Institute)

Data Scientist

Sep. 2022 – August 2023

Fort Worth, TX

- Developed and maintained a Java and Android Studio application integrated with Keras for a smartwatch, enhancing user engagement by offering advanced health data tracking, and facilitated efficient health monitoring through dynamic dashboards using Python, R, and Tableau, improving efficiency by 20%.
- Automated data integration between the wearable device and GCP dashboard using AutoML, reducing manual data handling by 43%, and centralized data storage and analysis by integrating Google Cloud with Azure Cloud, significantly decreasing data retrieval times.
- Enhanced data-driven decision-making by 35% through statistical analyses using Python and R, applying regression analysis, machine learning, and predictive modeling to derive actionable insights, and collaborated with a research team to design user-centric software applications, potentially increasing user satisfaction by 20%.

Zenith InfoTech

Lead Data Scientist

August 2019 – Nov. 2022

Mumbai, India

- Worked on two major projects over two years for a Denmark-based client, significantly advancing their health and finance sectors through machine learning, data analysis, and innovative technologies like blockchain.
- Key achievements include developing a secure sales application with increased efficiency, introducing generative models for personalized content, and optimizing inventory management.
- Delivered notable enhancements in fraud detection accuracy and operational cost savings by designing a new cloud data architecture, resulting in a 25% decrease in fraudulent activities, a 40% reduction in infrastructure costs, and contributing to a robust 99.9% system uptime with the implementation of an AI-bot.
- Led a diverse team of developers and Analysts on various aspects, yielding a 35% improvement in product performance, a 20% increase in company revenue, and elevating company stocks by 15%, establishing a strong foundation for future machine learning developments.

Technical Skills

Programming: Python, R, SQL, Java, PHP, SAS, JavaScript, C++, GO, Swift, Scala, Pandas, NumPy, Tableau, Power BI, Matplotlib, Seaborn, MongoDB, DynamoDB, NoSQL, Scikit-Learn, TensorFlow, PyTorch, Generative AI, Large Language Models (LLM), Keras, Neural Networks, Langchain, Hadoop, Spark, Kafka, MapReduce, Blockchain, Snowflake.

Developer Tools: Amazon Web Services (SageMaker, Lambda, EC2, Glue, RedShift), Microsoft Azure (DataBricks, ML Service, HDInsight), Google Cloud Platform (Colab, TensorFlow, BigQuery, AutoML, Vertex AI, DataFlow, DataProc), ELK Stack, MLFlow, Prometheus, Grafana, Git/GitHub, Docker, Kubernetes, Terraform, Jenkins, CI/CD Pipelines

Applications: A/B Testing, Agile Methodologies, Exploratory Data Analysis, Web Scraping, Predictive Modelling, Quantitative Analysis, Deep Learning, AI (Artificial Intelligence), Data Mining, Big Data Analysis, Project Management, Predictive Analytics, GIS

Projects

Data Analysis | *Java, SQL, Python, Android Development, Kotlin*

- Leveraged expertise in Data Analysis, Java, SQL, Python, and Android Development with Keras to create and implement a relational e-commerce database from the ground up. This included conducting requirement analysis, creating ER/EER models, performing normalization, and developing a GUI application using Android Development.
- Enhanced an application's synchronization with an SQL database, leading to a 10% increase in performance through effective indexing and schema optimization.
- Engineered the application over cloud infrastructure, resulting in a 30% improvement in data retrieval efficiency and ensuring reliable data management across concurrent transactions.

Machine Learning | *Python, Yolo, OpenCV, Java, NLP, ML (Decision Tree), GCP, API, Raspberry PI, Keras*

- Developed an Android chatbot application using Google Voice API for converting speech to text and a decision tree algorithm for predicting symptoms, achieving an accuracy rate of 75% and significant savings in time and costs for medical consultations.
- Implemented a vehicle detection and counting machine learning project utilizing Python, OpenCV, and the YOLO algorithm, reaching a 92% accuracy rate, thereby enhancing the efficiency of traffic monitoring systems and optimizing parking space management.
- Created a Python-based machine learning program at the RFID Lab that achieved an 85% accuracy rate in diagnosing facial illnesses, surpassing the previous best performance by 15%.
- Constructed a hybrid recommender system using collaborative filtering and content-based filtering, achieving 80% accuracy and providing users with personalized recommendations based on their preferences and interests.

Software Engineering | *Java, Python, Solidify, Firebase API, Google Maps API, HTML, CSS, JavaScript, PHP, XML, Docker*

- Developed an Android ridesharing application that integrates Firebase API for user authentication and Google Maps API for location tracking, enabling users to share rides and achieve an estimated 50% reduction in fuel costs.
- Engineered a scalable REST API using Python FastAPI with regex input validation, established an authentication system from the ground up, and enhanced testing and deployment processes with Docker, resulting in a 30% decrease in deployment time and an 80% reduction in input error rates.
- Created a custom detector to identify private information in smart contracts, enhancing security and data confidentiality in blockchain networks using the Slither analyzer.
- Designed and constructed a responsive website from scratch using HTML, CSS, and JavaScript, implemented form validation with JavaScript, used DOM, XPath, and XSLT for XML data querying, and developed a message board with PHP and MySQL, delivering a fully functional website.

Certifications

- | | |
|---------------------------------------|--------------------------------------|
| • AWS Certified Developer - Associate | • GCP - Certified Data Engineer |
| • AWS Certified Cloud Practitioner | • GCP - Certified Data Analyst |
| • Azure - A-Z900 Cloud Fundamentals | • Oracle Cloud Certified - Associate |