# GAURAV THORAT
United States

📞 682-563-0193  ✉ gauravthorat1998a@gmail.com  in LinkedIn  ⊙ GitHub  ⊕ Portfolio

## Education

**University of Texas at Arlington**                                   Jan. 2022 – Dec. 2023
Masters of Science in Data Science                                              Arlington, TX

**Mumbai University**                                                  Jun. 2017 – Jul. 2021
Bachelor of Engineering in Computer Engineering                               Mumbai, India

## Experience

**Bank of America**                                                    Sep. 2023 – Present
Data Scientist                                                                  Remote, USA

- Worked simultaneously on 2 projects 'Analyzing CTI Reports' and 'Detecting Deepfake Technologies'. For CTI Reports the goal is to determine their authencity and design LLM models for automated detection in future. Initiated with the TRAM by MITRE and FACTOR model for testing the reports and achieved 40% and 55% accuracy respectively.
- With added fine and proxy tuning, enhanced FACTOR's accuracy by 15% - 70% over training on 10k samples collected. Furthermore, explored TinyLlaMa and Mistral AI models with 7B parameters, achieved 68% accuracy.
- For Deepfake Technology, developed a framework for Fake Face Detection by snipping image of faces from live video feed and testing, achieved an exceptional accuracy of 95%. Working on implementing live detection on video itself.
- Worked on Testing LLM Detectors to determine whether they are sufficient to detect LLM generated text. Created an in-detail dataset, obtaining columns of Human and LLM generated text data, scraped over the Internet and generated from 5 different LLM text generators, conducted A/B testing between detectors to determine the text.
- This led to an ideal performing LLM text Detector with 98% accuracy, which is 20% higher than market standards. Also, developed a LLM text detector from scratch by reverse engineering ChatGPT's API, achieved 86% accuracy on it. Enhancing the detector by Perplexity Analysis and increasing the dataset for future iterations.

**UTARI (UTA Research Institute)**                                  Sep. 2022 – August 2023
Data Scientist                                                                  Fort Worth, TX

- Developed a Android app for wearble smartwatches, integrating cloud applications for data collection and analysis. Integrated Google's APIs for data collection, used AutoML and Databricks, to simultaneously connect GCP and Azure for seamless data transfer between them, to visualize data on Azure using Tableau and Power BI.
- Leveraged Azure as centralized data hub, reduced data transfer loss by 20%, reduced manual data handling by 43%, created dashboard for visualizing data for better health metrics increasing 38% rise in user traffic.
- Applied ML to design a model for automatic data transmission to Azure dashboard, by 99% uptime on the network, provided insights for health management, proved to 35% increase in user-satisfaction and designed an automated ticket window for feedback's, answered by Chat Bot, trained on massive dataset of FAQs.

**Zenith InfoTech**                                                  August 2019 – Nov. 2022
Lead Data Scientist                                                            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
- AWS Certified Cloud Practitioner
- Azure - A-Z900 Cloud Fundamentals
- GCP - Certified Data Engineer
- GCP - Certified Data Analyst
- Oracle Cloud Certified - Associate