SHUBHAM GUPTA

(513) 837-0799 | Austin, TX | Email: shubhamgupta2048@gmail.com | Website: http://shubhamg.me
GitHub: https://github.com/ShubhamGupta19 | LinkedIn: https://www.linkedin.com/in/shubham-gupta-664332171/

EDUCATIONAL QUALIFICATION:

Master of Engineering in Computer Science, Data Science — **April 2023**
University of Cincinnati, Ohio — **GPA: 3.75/4.0**

Bachelor of Technology in Computer Science — **July 2021**
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India — **GPA: 3.81/4.0**

TECHNICAL SKILLS:

Machine Learning:	Regression analysis, Predictive Analysis, Clustering, Feature Engineering, SkLearn, PCA, Statistical Analysis, Algorithms
Deep Learning:	CNN, RNN, NLP, LSTM, LLM PyTorch, TensorFlow 2.0, Keras, Image Processing, OpenCV, RCNN, Yolo v7,
Programming Languages:	Python, R, C, C++, Java, PHP, dart, JavaScript, SQL
Web Development:	HTML5, CSS3, JavaScript, Nodejs, Postman, Flask, Django, Flutter, Fast API, REST API
Databases	MongoDB, PostgreSQL, MySQL, Oracle SQL
Tools:	Power BI, Tableau, Git, Docker, Visual Studio, Jupyter Notebook, Spark, Apache Kafka, Confluent, Terraform, Mage, Airflow, Jira, Jenkins
Cloud Technologies:	AWS, Azure, GCP
Operating Systems:	Linux, Windows, MacOS

EXPERIENCE:

Data Science Analyst, Eversoft Technologies LLC — **September 2022 – Present**

- Analyzed extensive datasets using SQL and Python to extract trends, revealing a 10% increase in user engagement.
- Established databases, ETL pipelines, and reporting systems utilizing AWS Glue and Redshift.
- Automated data collection, processing, and reporting workflows with Python and Apache Storm, improving efficiency.
- Developed predictive model using Random Forest algorithm, reducing customer churn by 20%.
- Engineered real-time Power BI dashboard, enhancing operational performance by 15%.

Computer Vision and ML Engineer, SwiftAI — **September 2020 - March 2021**

- Deployed a real-time image search engine, offering conditional recommendations, and effectively reducing customer search time by 40%.
- Trained deep neural network models using PyTorch framework for classification purposes and constructed a user-friendly Flask-based interface.
- Orchestrated Python workflows, integrating unit testing at crucial stages to ensure goal achievement within allocated budget and time constraints.
- Implemented CI/CD pipelines with Jenkins and employed Git for version control.

Data Scientist, EkaLavya — **August 2019 – August 2020**

- Formulated and trained Naive Bayes model with 85% accuracy and 87% precision on sentiment analysis.
- Optimized model's hyperparameters, achieving additional 5% accuracy through CI/CD practices.
- Visualized results using WordCloud and confusion matrix.
- Collaborated for model deployment via Docker containers and Kubernetes.

ACADEMIC PROJECTS:

Faulty Sensor Detection

- Built a data pipeline to efficiently collect real-time sensor data, utilizing Kafka topics to feed MongoDB collections.
- Conducted exploratory data analysis on 50,000+ data points, leading to enhanced data quality and reliability.
- Trained and evaluated a ML model using XGBoost, achieving a predictive accuracy of 92% in fault detection.
- Successfully deployed the model on Amazon EC2 and S3, improving system efficiency and sensor reliability

Diabetic Retinopathy detection

- Developed an innovative medical image analysis platform using Python and OpenCV, achieving 92% diagnostic accuracy.
- Fine-tuned convolutional neural networks (CNNs) for accurate and reliable disease detection.

Solar Power Forecasting

- Orchestrated an advanced LSTM neural network project for precise solar power generation prediction, enhancing energy grid management and reducing fossil fuel reliance.
- Constructed a comprehensive data pipeline integrating weather data, solar panel specs, and environmental factors, enabling accurate forecasts hours ahead.
- Demonstrated dedication to sustainability, positioning the project as a leader in renewable energy optimization.
- Implemented a streamlined data pipeline, collecting sensor data via Kafka into MongoDB, saving time on dataset creation.
- Conducted comprehensive exploratory data analysis, covering ingestion, transformation, and validation steps.

ETL Pipeline: SQS to PostgreSQL Transformation:

- Developed ETL pipeline to process JSON data from AWS SQS to PostgreSQL, including data masking.
- Utilized Docker for local development with Localstack and PostgreSQL.
- Implemented parallel processing, error handling, and data validation for scalability and data integrity.
- Created comprehensive project documentation covering architecture, setup, and potential improvements.

PUBLICATION:

S. Gupta, K. Kavitha, P. Sharma, and R. V. S. Lalitha, "**Medicinal Plant Species Detection using Deep Learning**," *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichy, India, 2022, pp. 01-06, doi: 10.1109/ICEEICT53079.2022.9768649.

CERTIFICATIONS:

- **PowerBI Virtual Case Experience** by **PwC Switzerland**
- **Data Science Virtual Experience Programme** by **British Airways**
- **PCAP- Programming Essentials in python** by **Cisco**