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/

TECHNICAL SKILLS:

Machine Learning:	Classification, Statistics and regression analysis, Time Series Forecasting, Collaborative Filtering, Clustering, Feature Engineering, Natural Language Processing, PCA, Keras, SkLearn, Pandas
Deep Learning:	Computer Vision, Advanced Neural Networks, CNN (ResNet50, GoogLeNet, MobileNet etc.), Pytorch, Tensorflow 2.0, Keras
Programming Languages:	Python, R, C, C++, Java, PHP, dart, JavaScript, SQL
Web Development:	HTML5, CSS3, Nodejs, Postman, Flask, Django, REST API
Databases	MySQL**,** MongoDB, PostgreSQL, Oracle PL/SQL,
Tools:	MS Excel, Snowflake**,** Power BI, Tableau, Git, Docker, Spark, Apache Kafka, Confluent, Terraform, Grafana Mage, Airflow, Jira, Jenkins, Redis, Unix Shell Scripting
Cloud Technologies:	AWS, Azure, GCP

EXPERIENCE:

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

- Employed SQL and Python to analyze extensive datasets, uncovering trends that resulted in a 10% increase in user engagement.
- Established databases, ETL (Extract, Transform, Load) pipelines, and reporting systems using AWS Glue and Amazon Redshift.
- Integrated AWS Lambda functions within data processing workflows to execute real-time data transformation.
- Automated data collection, processing, and reporting workflows with Python and AWS Glue, leading to improved efficiency and scalability.
- Developed a predictive model using Python on Amazon SageMaker, successfully reducing customer churn by 20%.
- Implemented PowerBI dashboards to vividly showcase acquired insights.

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

- Deployed production search engine for conditional recommendations, reducing customer search time by 10 minutes.
- Trained deep neural network models for classification; built Flask-based user interface.
- Orchestrated Python workflows, ensuring goal achievement within budget and time constraints.
- Showcased proficiency in designing, testing, and evaluating machine learning algorithms, comparing various approaches.
- Utilized Terraform for infrastructure provisioning and management to ensure scalability, consistency, and cost-effectiveness.
- Applied SQL for querying and analyzing relational databases in the context of project requirements.

Data Scientist, *Ekalavya* **August 2019– August 2020**

- Formulated and trained Naive Bayes model with 85% accuracy and 87% precision on sentiment analysis.
- Utilized SQL for data extraction and analysis, enhancing the preprocessing phase for model training.
- Optimized model's hyperparameters, achieving additional 5% accuracy.
- Visualized results using WordCloud and confusion matrix.
- Collaborated with teams for model deployment in production environment via Docker containers and Kubernetes clusters.
- Monitored and improved model performance, reducing false positives by 10% in 6 months.

EDUCATION:

Master of Engineering in Computer Science (Specialization - 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 **CGPA: 3.81/4.0**

PROJECTS:

Realtime Air Quality Control Monitoring:

- Developed Air Quality Control App using Python, FastAPI, and ML algorithms for predictive analysis.
- Addressed binary classification challenge, achieving an accuracy of 85% in distinguishing specific factors affecting air quality (positive class) from unrelated factors (negative class).
- Implemented Docker for efficient deployment, resulting in a 30% reduction in deployment time.
- Utilized AWS infrastructure (S3, EC2, ECR) with Git Actions and Terraform for real-time monitoring, leading to a 20% improvement in system responsiveness.
- Aimed at accurate air quality assessments, minimizing false predictions for diverse locations.
- Conducted thorough analysis of model predictions, resulting in the identification of key factors influencing air quality.

ETL Pipeline: Social Media Analytics:

- Designed and implemented an ETL pipeline to gather and analyze data from social media platforms, including Twitter, Facebook, and Instagram, for a marketing analytics application.
- Developed custom Python scripts and utilized API endpoints to extract real-time data, including posts, comments, likes, and engagement metrics.
- Transformed and enriched the raw social media data by applying sentiment analysis, entity recognition, and topic modeling using natural language processing (NLP) techniques.
- Employed distributed systems and cloud services, such as Apache Kafka and Google Cloud Platform (GCP), for handling high-throughput data streams.
- Conducted data cleansing, data deduplication, and advanced SQL queries to ensure data accuracy and quality.

Faulty Sensor Detection

- Implemented 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.
- Implemented automated deployment of the model using Git and Terraform, streamlining the process by 40%.
- Successfully deployed the model on Amazon EC2 and S3, improving system efficiency and sensor reliability

Credit Card Fraud Detection Project:

- Implemented fraud detection system using Apache Spark, Pandas, and Scikit-learn.
- Designed efficient data integration with Apache Spark, handling diverse sources.
- Constructed accurate ensemble models: XGBoost, TensorFlow, achieving 95% detection accuracy.
- Enabled real-time fraud detection via Apache Kafka, reducing response time to milliseconds.
- Implemented rapid monitoring with Apache Spark, PostgreSQL, Flask, promptly identifying potential fraud incidents.

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**