Vedant Bajaj

Dallas, TX | (945)-291-0320 | vedant.bajaj99@gmail.com

www.linkedin.com/in/vedant-bajaj99 | https://github.com/VedantBajaj

PROFESSIONAL SUMMARY

Graduate student in Computer Engineering with hands-on experience in data engineering and cloud infrastructure. Proficient in building ETL pipelines, migrating data systems to GCP, and orchestrating workflows with Apache Airflow. Strong foundation in Python, SQL, and Google BigQuery for processing large-scale datasets.

EDUCATION

Master of Science in Computer Engineering Aug. 2024 – May. 2026

University of Texas at Dallas GPA – 3.7

Coursework: Machine Learning, Statistical Methods in ML/AI, Applied Data Structures and Algorithms

Bachelor of Technology in Computer Engineering Aug. 2018 – May. 2022

University of Pune GPA – 3.7

Coursework: Data Structures and Algorithms, Object-Oriented Programming, Database System, Machine Learning

TECHNICAL SKILLS

Data Engineering & Cloud: Python, SQL, Google Cloud Platform (GCP), BigQuery, Apache Airflow, Dataflow, ETL Pipelines

Database Management: MySQL, SQL Server, Data Warehousing, Query Optimization, Database Design

Programming Languages: Python, C++, Java **Data Analysis & ML:** TensorFlow, Scikit-learn, NumPy, Pandas, Data Analysis

Software Development: Object-Oriented Design, REST APIs, Version Control (Git), Agile Development

Additional Tools: OpenCV, Pytesseract, Django, Web Scraping

PROFESSIONAL EXPERIENCE

Data Engineering Intern, Conifer Health Solutions, USA Sep. 2024 – Present

- Migrating on-premises data infrastructure to GCP, implementing ETL pipelines using Apache Airflow and Dataflow for automated data processing into BigQuery data warehouse.
- Developing Python and SQL-based workflows to process large-scale healthcare datasets, optimizing for cloud performance and scalability.

Software Engineer, Rabbit and Tortoise Technology Solutions, India Nov. 2022 – Sep. 2023

- Designed modular OCR pipelines in Python using OOP principles, OpenCV, and PyTesseract, achieving 95% text detection accuracy and making the system scalable and maintainable.
- Implemented reusable classes for image preprocessing, text extraction, and validation, reducing future development effort by 40% and improving extraction efficiency.

Software Engineer Intern, Bizsol IT Services, India Aug. 2021 – Dec. 2021

- Developed SQL Server-based data integration system, reducing manual reconciliation time by 60%.
- Optimized database queries and implemented indexing strategies, improving query performance by 40%.
- Created interactive dashboards for data visualization, enhancing business intelligence capabilities.

PROJECTS

ETL Data Pipeline | University of Pune, India

- Designed end-to-end ETL pipelines using Python, NumPy, and Pandas for processing large-scale datasets.
- Applied automated data cleaning and validation workflows, reducing preparation time by 50% and improving reliability.
- Created scalable data transformation pipelines for analytics, resulting in 30% improved decision accuracy.

Real-Time Face Mask Detection System, University of Pune, India

- Engineered real-time mask detection system utilizing OpenCV and MobileNetV2, achieving 96% accuracy.
- Implemented Haar Cascade algorithms for facial recognition and deep learning models for mask classification.
- Designed detection modules using class-based architecture in Python for real-time face mask classification.

PDF OCR with ML-Based Text Extraction, University of Texas, USA | https://pdf-ocr-webapp.onrender.com

- Developed PDF OCR web application integrating Python, OpenCV, and PyTesseract for automated text extraction.
- Designed and implemented robust backend APIs for document processing with efficient data handling.
- Created responsive frontend interface ensuring seamless user experience for document submission and result retrieval.