

Anugya Chaubey

+91 9910693592 / anugyachaubey95@gmail.com / [linkedin.com/in/anugya](https://www.linkedin.com/in/anugya) / github.com/anugyachaubey

EDUCATION

BTech in CSE (AIML)	CGPA: 9.37	VIT University May 2027
Higher Secondary Education	Percentage: 91.6%	New Delhi July 2022
Secondary Education	Percentage: 93%	New Delhi July 2020

TECHNICAL SKILLS

Languages: Python, SQL, Java, C++, YAML
Frameworks: HTML, CSS, Flask
Developer Tools: Visual Studio, PyCharm
Cloud & DevOps: AWS (Glue, S3, IAM, CloudWatch), Apache Spark, ETL Pipelines
Libraries: pandas, NumPy, Matplotlib, Apache Sparks, TimescaleDB, CSV data management

EXPERIENCE

- Cloud Intern at IOTfy** [\[link\]](#) May 2025 – June 2025
- Created an automated ETL pipeline using AWS Glue and Apache Spark to transfer and partition 5GB+ of daily time-series data from TimescaleDB to S3, achieving 60% storage cost reduction via Parquet optimization and 70% faster query performance.
 - Automated data exports and cloud operations with VPC-secured connections, IAM roles, and CloudWatch monitoring, saving 10+ manual hours/week and enabling real-time analytics on historical IoT data.
- Backend Developer Intern at The Mentor** [\[link\]](#) Aug 2024 – Jan 2025
- Contributed to building and maintaining the company's educational technology (EdTech) website using Python 3 and Django, ensuring efficient backend functionality and smooth integration with the frontend.
 - Designed and managed MySQL relational database schemas with over 2,000 records, optimizing data storage and retrieval for better system performance.

PROJECTS

- Video Anomaly Detection** / Python, Computer Vision, DAE, Scikit-learn [\[link\]](#) Mar 2025 – May 2025
- Developed an explainable deep learning framework using pre-trained CNN models and SHAP interpretability. Implemented 5 motion-based features for anomaly classification with mean_motion showing highest importance score of 0.35.
 - Trained on 70 video segments (25 normal, 45 anomalous) from UCF Crime dataset achieving 81% accuracy, 88% precision for anomalies, 82% recall, F1-score of 0.85, and 0.82 AUC. Demonstrated ultra-fast training of 5 seconds on UCSD Ped1 (85.9% AUC) and 2.9 seconds on UCSD Ped2 (92.4% AUC).
 - Processed videos with frame counts ranging from 100-76,988 frames using OpenCV for preprocessing and feature extraction. Achieved 37 true positives out of 45 anomalous cases with only 5 false positives, delivering 2-5% AUC improvement potential over baseline methods through dataset expansion and model fine-tuning.
- Deepfake Analysis** / Python, TensorFlow, Keras [\[link\]](#) Dec 2024 – Jan 2025
- Developed a deepfake detection model using Python, TensorFlow, and Keras. Built an 8-layer CNN for spatial features and a 2-layer LSTM with 256 units each for temporal analysis.
 - Trained on 800 videos using Adam optimizer with a learning rate of 0.0001 and batch size of 32.
 - Utilized OpenCV for video preprocessing, including frame extraction and metadata parsing. Trained and validated the model on FaceForensics++ and Deepfake Detection Challenge datasets, achieving 81% accuracy and an F1-score of 0.90.
 - Designed a user-friendly interface for real-time video analysis and classification. Optimized model performance using GPU acceleration and data augmentation, achieving 5% higher accuracy than the baseline RNN model.
- Real-time Face Recognition Attendance System** / Python, OpenCV, NumPy [\[link\]](#) Aug 2024 – Oct 2024
- Engineered a real-time face recognition system capable of identifying individuals and recording their attendance, automating the process and reducing manual effort by 90%.
 - Executed a robust face recognition pipeline using OpenCV's Haar cascade classifiers for face detection and KNN for identification, training on a dataset of 700 facial images with an average detection time of <100ms per frame.
 - Created a data-driven attendance management solution that stores records in a structured CSV format, facilitating easy analysis and reporting. The system also provides optional text-to-speech feedback for immediate confirmation, reducing false negatives by 30%.
 - Improved the system's accuracy and efficiency through data preprocessing and model optimization, enabling the model to make correct predictions 8 out of 10 times.