AKASH PINTUKUMAR DAS

akash.pdas@outlook.com | +1 (930)-333-5031 | GitHub | LinkedIn | Portfolio

EDUCATION

Indiana University Bloomington, Bloomington, United States | *Master of Science in Data Science* **August 2023 - May 2025**
Courses: Advanced Natural Language Processing, Elements of Artificial Intelligence, Computer Vision GPA – 3.78

University of Mumbai, Mumbai, India | *Bachelor of Engineering in Electronics and Telecommunication* **August 2017 - May 2021**
Courses: Neural Network and Fuzzy Logic, Image Processing and Machine Vision, Database Management System. GPA – 3.66

SKILLS AND EXPERTISE

Programming Languages	SQL, Python, Java, Shell Scripting, C, C++
Frameworks	Apache Spark, Apache Kafka, Hadoop, Flask, Django
Libraries	Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, OpenCV, NLTK
Cloud Services	Amazon Web Services, Azure Web Services, Google Cloud Platform
Machine Learning Skills	Supervised & Unsupervised Learning, Model Development,, Hyperparameter Tuning, Model Evaluation, NLP
Data Science Skills	Data Visualization, EDA, Predictive Analytics, A/B Testing, Experiment Design, Time Series, Causal Inference

PROFESSIONAL EXPERIENCE

Indiana University, *Research Assistant* – *IUB Psychological Department* | *Python, Computer Vision* **June 2024 - Present**
- Automated annotation of 200,000+ video frames using a hybrid pipeline combining LLMs (Hugging Face Transformers) and CLIP models for multimodal scene understanding, reducing manual effort by 70% and generating high-fidelity labels to train downstream supervised learning models for cognitive development research.
- Improved cognitive milestone prediction accuracy by 30% through scalable ML pipelines integrating YOLOv5 object detection on 5M+ video frames, GIST descriptors, Clutter metrics, and synchronized Librosa audio embeddings, enabling multimodal feature fusion to model relationships between visual complexity, object presence, and auditory context.
- Conducted statistical validation using Kolmogorov-Smirnov (KS) Tests and Earth Mover's Distance (EMD) to quantify distribution shifts in visual and auditory features across age groups, while applying linear mixed-effects models (LME) in R (lme4) and Python (statsmodels) to measure the impact of environmental factors on early word acquisition in 200+ infants.

LTIMindtree Limited, *Data Scientist* – *Johnson and Johnson (Client)* | *Python, Flask, Teradata, Oracle* **July 2021 – July 2023**
- Automated 200+ hours of manual work quarterly by building a CI/CD pipeline with Apache Airflow, Python, and Shell scripting, streamlining data ingestion into Informatica pipelines, reducing manual errors by 70%, and improving production efficiency by 85% for scalable ML data workflows.
- Boosted Teradata query performance by 80% and reduced downtime by 50% by developing automated SQL optimization tools in Python, enabling faster feature extraction for downstream ML models and ensuring high availability in large-scale data processing.
- Reduced support ticket resolution time by 70% with an AI-powered chatbot using BERT-based intent detection deployed on AWS Lambda, integrated with API Gateway to provide real-time metadata retrieval and automated monitoring of data pipeline health.

Mastersoft ERP Solutions, *Research and Development Intern* | *Python, Flask, Keras, BERT models* **December 2020 – June 2021**
- Automated 70% of student inquiries and reduced response time by 50% through a scalable NLP-driven chatbot, leveraging BERT-based architectures for intent classification, hosted via containerized APIs on AWS EC2, and integrated with AWS RDS for persistent query storage.
- Improved intent recognition accuracy by 55% and reduced inference latency by 35% through hyperparameter optimization, stratified data augmentation, and efficient batch processing, with fine-tuned BERT and DistilBERT models optimized for real-time performance.
- Enabled continuous model improvement saving 150+ support hours monthly per institution, by implementing a full MLOps pipeline with CI/CD workflows, automated retraining on AWS SageMaker, and real-time deployment monitoring to support 10,000+ monthly users.

ACADEMIC PROJECTS

Mental Health Analyser | *Speech Processing, Deep Learning, Librosa, Natural Language Processing*
- Delivered 90%+ accuracy in real-time mental health monitoring by building a multimodal emotion detection system combining 1D CNNs (Librosa) for speech and 2D CNNs (OpenCV) for facial analysis.
- Engineered a multimodal emotion detection pipeline integrating BERT embeddings, spectrogram-based audio features, and facial keypoint detection, with optimized fusion layers to improve overall classification performance by 20% compared to single-modality baselines.
- Deployed as a Flask API enabling live webcam and microphone-based emotion tracking with <200ms latency.

Multi-Language Auto–Translation using Natural Language Processing | *Transformer Models, Python, Keras*
- Delivered 85%+ accurate multilingual translations and 83% script auto-detection by developing a Transformer-based NLP system using mBART, TensorFlow, and Hugging Face, supporting seamless language detection and context-aware translations across 20+ global scripts.
- Enhanced performance with self-attention encoder-decoder models, fastText language identification, and GPU-accelerated training (CUDA), deploying the solution via FastAPI and Docker for scalable, real-time translation services.

Live - Rock-Paper-Scissor Simulator | *Computer Vision, CNNs, OpenCV, Deep Learning*
- Built a real-time live Computer Vision Rock-Paper-Scissors game using OpenCV and a 2D CNN, achieving 89.6% training accuracy and 78.9% validation accuracy on 50,000+ hand gesture images, allowing players to compete through instant hand sign recognition.
- Integrated real-time hand tracking with sub-100ms latency, enabling smooth gesture detection, automated scorekeeping, and seamless gameplay entirely through live webcam-based interactions.

PUBLISHED PAPERS

- Emotion Detection Using Natural Language Processing and ConvNets (**ISBN: 978-981-19-2211-4**) (**DOI: 10.1007/978-981-19-2211-4_11**)