Kunal Shah

+1 980-899-3900 | kunaljshah03@gmail.com | linkedin.com/in/kunal-shah-86a23016b/ | GitHub | Google Scholar

I am a versatile computer science graduate with a Master's from NC State University and a Bachelor's from Pune Institute of Computer Technology. I'm proficient in Python, Java, JavaScript, and SQL, specializing in machine learning, deep learning, and computer vision. I'm skilled in cloud platforms like AWS and GCP, with experience in database management systems. I'm Seeking full-time roles in machine learning, data science, and software development to drive innovation.

Education

Master of Science: Computer Science **GPA: 3.75/4.0**
North Carolina State University *08/2022 - 05/2024*
Relevant Coursework:
- (CSC 591) Neural Networks and Deep Learning.
- (CSC 540) Database Management Systems.
- (CSC 517) Object-Oriented Design & Development.
- (CSC 505) Design & Analysis of Algorithms.
- (CSC 510) Software Engineering.

Bachelor of Engineering: Computer Engineering with **Honors: Data Science** **GPA: 9.66/10**
Pune Institute of Computer Technology (University of Pune, India) *08/2018 - 06/2022*

Skills

- **Languages:** Python, JavaScript, C++, SQL, Java.
- **Machine Learning:** Deep Learning, Computer Vision, Natural Language Processing, GenerativeAI, Prompt Engineering.
- **Frameworks & Libraries:** Spring Boot, Django, Flask, React.js, Node.js, TensorFlow, PyTorch, Keras.
- **Cloud & Databases**: Amazon Web Services(AWS), Google Cloud Platform(GCP), MySQL, MongoDB.
- **Tools:** Git, Tableau, Docker, Microsoft Office.

Experience

Research Assistant, North Carolina State University - Raleigh, NC *01/2023 - Present*
- Developed a transferable prediction model, leveraging hierarchical clustering with CARTs and transfer learning, to assess open-source(OS) project health (95% accuracy), analyzing 150+ GitHub projects with interpretable insights from LIME.
- Optimized regression models for OS health metrics using HyperOpt, reducing mean squared error by 40%.
- Contributed to open-source tools for automated data mining in OS projects and participated in NSF-funded research on adapting software engineering practices to computational science.

Data Science Intern, Fidelity Investments - Durham, NC *06/2023 - 08/2023*
- Built and deployed cutting-edge Generative AI text-to-image model using AWS SageMaker for Fidelity platform, leveraging Prompt Engineering and LoRA for error reduction and accelerated fine-tuning.
- Evaluated potential Fidelity software for automated data analysis on a 2.1M-participant dataset using EDA, feature engineering, and advanced data visualization (dashboards, statistical plots).
- Presented evaluations and recommendations to senior management, impacting decisions on software adoption.

Machine Learning Engineer Intern, Omdena - Pune, India *05/2021 - 07/2021*
- Implemented user-driven NLP solution (Beautiful Soup, Scrapy) reducing information gathering time by 70%.
- Employed LSTM/BiLSTM/Transformers for 96% accurate data extraction from diverse sources (webpages, images, databases, social media) with sentiment analysis and summarization.
- Deployed the web-scraping tool on the Google Cloud Platform with a Streamlit front-end, empowering clients to define topics, interests, and timeframes for personalized data retrieval.

Software Developer Intern, Carikture India - Pune, India *07/2020 - 09/2020*
- Launched a data-driven educational platform (Node.js, Electron.js, React.js) for 5000+ users, integrating predictive models and interactive data dashboards (Tableau) to personalize learning paths, increasing user engagement by 40%.
- Incorporated machine learning models (recommendation systems) for personalized learning paths and A/B tested UI elements (ReactJS, CSS, Bootstrap) to optimize the user experience.

Publications

EffResUNet: Encoder decoder architecture for cloud-type segmentation | *Big Data and Cognitive Computing 2022*
- Researched a cloud-type segmentation encoder-decoder UNet-architecture leveraging EfficientNet, residual blocks, and an attention mechanism with an F1 score of 0.5735, surpassing SOTA methods.

Image-Based Plant Seedling Classification Using Ensemble Learning | *ICACIE 2021 (LNNS volume 428, Springer)*
- Devised an ensemble-based model for plant seedling classification using a shallow CNN for feature extraction and K-Nearest Neighbors for classification refined through 5-fold cross-validation. Achieved a recall of 95.32%, outmatching existing techniques.

Projects

WolfMedia *Java, Spring Boot, SQL, Docker, Swagger API*
- Designed and built a scalable media streaming platform (WolfMedia) using Spring Boot and Docker, managing artists, content,

payments, and subscriptions with secure APIs and comprehensive documentation.

Real-time Video Anomaly Detection *PyTorch, Python(Pandas, NumPy, Plotly), OpenCV, YOLO*

- Created a multi-faceted deployable tool integrating real-time traffic visualization, pedestrian anomaly detection system (ROC AUC 0.932, 20% false positive reduction), and a Streamlit-powered alert system for proactive notifications to enhance traffic safety and security.

PackTravel *Python, JavaScript, Django, MongoDB, Pylint*

- Developed a ride-sharing application (PackTravel) using Django and MongoDB for university students, enabling features like multi-modal route planning, ride search/joining, and user authentication, with plans for future ML-powered route optimization and cab service integration.

Pose Detection for Sign Language Videos *Tensorflow, Mediapipe, OpenPose*

- Programmed a deep learning pipeline for sign language-to-Hamnosys translation, achieving 47% accuracy (9% above SOTA) using 3D pose landmarks from Mediapipe and transfer learning on TensorFlow CNNs, enabling multilingual communication.

Leadership &Initiatives

- Engaged as a research member at the NCSU RAISE Lab, focusing on the convergence of Artificial Intelligence (AI) and Software Engineering (SE).
- Contributed as a technical consultant for 3 startups incubated at the PICT Startup and Innovation Center as the technical head of the PICT Entrepreneurship Development Cell.
- Led seminars on Machine Learning for 500 students as a recognized PICT ACM Student Chapter (PASC) member, and contributed to 3 Best ACM Student Chapter awards.