Bhuvan Shah

213-331-1360 | bhuvansh@usc.edu | linkedin.com/in/bhuvanshah | Portfolio | github.com/Bhuvannnn

Education

University of Southern California — Los Angeles, CA
Master's of Science in Computer science — *May 2025*

Indus University — India
Bachelor's of Technology in Computer Science and Engineering — *July 2019 - May 2023*

Technical Skills

Languages & Tools: Python, SQL, Scala, R, C++, C#, Git, Excel, Mathematics, Deep Learning, Machine Learning, Transformers, LSTM, XGBoost, Pytorch, Keras, SciKit, Computer Vision
Cloud, Data & DevOps: Databricks, MySQL, MongoDB, BigQuery, Snowflake, Hadoop, Database, AWS S3, AWS SageMaker, GCP, Azure, Django, Kubernetes, Terraform, CI/CD Pipelines, Docker, Container, Lambda, Agile

Experience

Graduate Teaching Assistant — May 2024 – Present
University of Southern California — *Los Angeles, CA*
- Coordinate 360+ students in advanced game development for object-oriented programming and algorithm optimization in C#, Python, and R—resulting in 90% of teams publishing games with scalable architectures
- Design 6+ coding assignments integrating Firebase for real-time analytics —enabling students to track 10,000+ gameplay events per project and make data-driven design decisions with Agile methodologies

Data Scientist Intern — December 2022 – May 2023
Prompt Softech — *India*
- Impacted 12 traders by attaining 93% accuracy with a custom LSTM neural network with bifurcated architecture using stock prices and Twitter sentiment polarity
- Enhanced data processing efficiency by 35% reduction in retrieval time, by operating Hadoop for managing datasets over 50 GB
- Boosted trading profits to over $1200 as measured by profit increase, by optimizing trading strategies with Apache Spark for real-time processing and model validation alongside MQL4/MQL5 backtesting

Data Analyst Intern — June 2021 – July 2021
Tornado Computers — *India*
- Attained 92% accuracy in performance analysis by leveraging Apache Kafka for real-time data streaming and IoT integration, improving hardware performance and predicting system failures
- Decreased downtime and repair costs by 87% through implementing a neural network on TensorFlow for predictive maintenance, processing sensor data with Apache Airflow to automate data pipeline orchestration

Software Developer Intern — June 2020 – October. 2020
VueInternational — *India*
- Delivered a 15% increased business impact by designing and executing interactive components with React, Tailwind CSS, Node.js, and Express.js, while incorporating Redis for efficient data caching and MongoDB for data retrieval
- Strengthened transaction security by 20% through Docker-containerized Stripe API integration for secure transactions

Projects

NoteScape | *Python, FastAPI, Neo4j, NLP, React, Docker* — February 2025 – Ongoing
- Architected a semantic knowledge graph system by implementing keyword-based relationship extraction algorithms, resulting in 85% accuracy while processing 1000+ notes in under 30 seconds, enabling real-time knowledge across large datasets
- Engineered a graph database schema using Neo4j reduced relationship query time by 75% compared to traditional SQL approaches, enabling real-time visualization of 500+ interconnected notes

Healthcare Resource Optimization System | *Python, Flask, PostgreSQL, Docker* — September 2024 – November 2024
- Engineered a microservices-based healthcare management platform using Python and PostgreSQL, implementing CI/CD pipelines with Jenkins for automated testing and deployment; containerized with Docker reducing deployment time by 30%
- Built ETL pipelines processing 1000+ daily healthcare records, integrated real-time analytics dashboard using Flask and Chart.js, achieving 95% resource utilization across 4 departments and 20+ medical equipment units

Statistical Analysis of Cultural Representation in Video Games | *Python, SPSS* — August 2024 – November 2024
- Led a statistical study on the impact of cultural representation in games on player anxiety, coordinating research efforts, applying predictive models and regression analysis in Python, SPSS, and Excel, succeeding with 95% statistical power
- Developed scalable ETL pipelines for handling and transforming experimental datasets, amplifying data storage with SQL, NoSQL and Databricks, reducing processing latency by 25%

PII Detection and Masking | *PyTorch, NumPy, Pandas, Transformers, NLP, LLMs* — January 2024 – May 2024
- Led and collaborated team of 5 and achieved a 99.56% F1 Score by developing a DeBERTaV3-based NER system with advanced architecture along with embedding layers and in-context learning to boost PII detection accuracy
- Expanded detection capabilities from 15 to 94 PII tags by leveraging Elasticsearch for scalable data storage and search and integrated reinforcement learning with TensorFlow to optimize training, improving user data privacy