# VATSAL SHETH

(878) 834-9437 | vus@andrew.cmu.edu | https://www.linkedin.com/in/sheth-vatsal/ | New York | https://github.com/Vatsal-Sheth

## Education

**CARNEGIE MELLON UNIVERSITY** **August 2023 – December 2024**
Master's in Business Intelligence and Data Analytics, GPA: 3.68/4.0 Pittsburgh, USA

- Coursework: Advance Machine Learning, Math for ML, Quantum ML, Unstructured Data Analytics, Distributed Systems
- Capstone Project: Developed predictive models for credit risk using ensemble learning for Northwest Bank

**NMIMS UNIVERSITY** **July 2017 – August 2021**
Bachelor of Technology (Honors), GPA: 3.47/4.0 Mumbai, India

- Relevant Coursework: Artificial Intelligence Foundations, Neural Networks, Computer Vision, Fuzzy Logic, Statistics
- Honors Thesis: Designed CNN-based genre classification model, achieving 92% accuracy (published under IEEE)

## Work Experience

**Machine Learning Intern** **May 2024 – August 2024**
*Carnegie Mellon University* Pittsburgh, USA

- Developed machine learning models for economic research to validate statistical trends and support policy analysis
- Processed and analyzed large datasets (1TB+) using Spark and Python, improving computational efficiency by 40%
- Conducted regression analysis, chi-square tests, and data cleaning to ensure robust insights for labor economics research
- Created visualizations and replication packages with LaTeX to enhance research reproducibility and support publications

**Machine Learning Software Engineer** **June 2021 – August 2023**
*Accenture* Mumbai, India

- Developed predictive models to optimize network resource allocation, reducing downtime for telecom clients by 30%
- Built automation scripts using Python and Splunk APIs to analyze call data records, enhancing fault detection by 87%
- Analyzed time-to-churn patterns for clients using Kaplan-Meier survival analysis, identifying high-risk customer segments.
- Engineered machine learning models for churn prediction, enabling proactive retention strategies and reducing churn by 15%
- Designed and implemented ML pipelines to forecast customer churn, enabling clients to take proactive retention measures
- Created dashboards with R Shiny, PostgreSQL and Docker for visualizing model performance, improving decision-making

## Projects

**Meeting Minutes Summarization and Decision Tracking using LLMs (Research Project)**
***Tech Stack:*** Python, ReactJS, OpenAI API, LangChain, Pinecone, Pandas, Beautiful Soup, Cron Jobs

- Developed a pipeline with OpenAI API and LangChain to summarize meeting minutes, auto-generate questions, and track decision statuses (e.g., completed, on hold, canceled), improving decision-tracking efficiency by 50%.
- Designed a ReactJS interface integrated with a Pinecone vector database, enabling query-based access to meeting summaries
- Automated updates and tracking of decisions, ensuring complete visibility into discussions/changes, for absent stakeholders

**Poker AI Bot (CMU Hackathon runner-up)**
***Tech Stack*** - Python, Java, Monte Carlo Simulations, Custom Risk-Reward Algorithm

- Engineered an AI poker bot by integrating Monte Carlo simulations into a risk/reward algorithm winning 83% matches
- Adapted the bot to handle diverse scenarios, including edge cases like aggressive play styles "all-in" strategies

**Genre Classification using Neural Networks and Traditional Models (Published research under IEEE)**
***Tech Stack*** - Python, TensorFlow, Librosa, SQL, Numpy, Seaborn

- Enhanced traditional models like XGBoost with hyperparameter tuning on the GTZAN dataset, achieving 79% accuracy
- Expanded the dataset to 100,000 samples by segmenting 10,000 audio tracks, enabling a custom-developed CNN to achieve 86% accuracy, significantly outperforming the existing 66% from previous CNN approaches
- Leveraged Librosa for temporal and spectral feature extraction, including MFCCs, improving performance across all models

## Academic Experience

**Graduate Artificial Intelligence Engineer** **August 2024 – December 2024**
*Northwest Bancshares (Capstone Project)* Pittsburgh, USA

- Developed XGBoost models for loan prepayment and delinquency risk, increasing precision and recall from 66% to 79%
- Feature engineered seasonal trends(Freddie Mac) and economic factors(FRED dataset) to enhance risk prediction accuracy
- Enhanced financial document processing using LLMs to extract actionable insights, providing preliminary frameworks

**Graduate Artificial Intelligence Research Assistant** **November 2024 – May 2024**
*Block Center for Technology and Society, CMU* Pittsburgh, USA

- Built machine learning models to predict job migration ease and optimize skill development plans for workforce transitions
- Reduced application load times for 100+ GB datasets to under 50 seconds using parallelized processing with Spark
- Designed interactive dashboards with R Shiny and Docker, delivering real-time insights for worker supply chain predictions

## Technical Skills

**Programming & Software:** Python, R, Java, SQL, JavaScript, C++, Spring Boot, Angular, REST APIs
**ML/AI & Data Engineering:** TensorFlow, PyTorch, Scikit-learn, Spark, Keras, LangChain, OpenCV, Pandas, Numpy
**Cloud & Tools:** AWS (EC2, S3), GCP, Docker, Kubernetes, Hadoop, Git, Bamboo, Bitbucket, Cron Jobs, SpaCy, Hugging Face
**Data Visualization:** Tableau, R Shiny, Seaborn, Matplotlib

## Publications and Certifications

- *"Genre-Based Music Classification using Machine Learning and Convolutional Neural Networks"*, ICCCNT 2021, IEEE
- IBM Data Science Professional Certificate
- Sequence Models for Time Series and Natural Language Processing, Google Cloud