

Dhairya Jayesh Chheda

Ready to Relocate | ☎ +1 930-333-5591 | ✉ dhairyajayeshchheda@gmail.com | in [linkedin.com/chheda-dhairya](https://www.linkedin.com/in/chheda-dhairya) | 🐙 [GitHub](#) | 📁 [Portfolio](#)

EDUCATION

Master's in Data Science

Indiana University – Bloomington, Indiana, USA

August 2023 – May 2025

GPA: 3.9/4.0

Bachelor of Engineering (B.Eng.) in Information Technology

University of Mumbai – Mumbai, Maharashtra, India

August 2017 – May 2021

GPA: 9.62/10.0

TECHNICAL SKILLS

Languages: Python, R, SQL, Java, C/C++

Frameworks: Numpy, Pandas, Matplotlib, Scikit-learn, NLTK, Spacy, SciPy, TensorFlow, Keras, PyTorch, HuggingFace, FastAPI

Machine Learning: Statistics, Feature Engineering, Dimensionality Reduction, Machine Learning (Supervised & Unsupervised), Regression, Classification, Predictive Modeling, Natural Language Processing, Generative AI, Large Language Models (LLMs)

Developer Tools: Amazon Web Services (AWS), GitHub Actions, Docker, Jenkins, JIRA, CI/CD, DVC, Airflow, MLOps

PROFESSIONAL EXPERIENCE

Machine Learning Engineer Intern

September 2024 – December 2024

Jiseki: an early-stage AI startup

San Mateo, CA, USA

- Optimized **ETL pipelines** using **AWS** Glue, Amazon Redshift, and Amazon S3 to integrate and harmonize data from various sources into a centralized **data warehouse**, **enhancing data pipeline scalability and efficiency by 50%**.
- Developed a conversational agent by blending **Google's Voice Activity Detection (VAD)** to capture and convert speech to text, **OpenAI's Text-To-Speech (TTS)** and Pygame for **voice generation** and smooth audio playback.

Data Scientist Intern

May 2024 – July 2024

Indiana University

Bloomington, IN, USA

- Curated a corpus of **1M+ records**, enabling comprehensive exploratory **data analysis** and **supporting model development**.
- Optimized **multi-label** classifiers using RoBERTA and active learning, achieving **60% F-1 score with limited labeled data**.
- Enhanced a text summarization model with **BART**, allowing to extract insights from a large corpus of visually rich documents.
- Developed **information extraction** pipelines in a **few-shot incremental** setting using **Langchains**, **OpenAI GPT-4**, and **Retrieval-Augmented Generation (RAG)** pipelines for efficient meta-analysis.

Software Engineer

June 2021 – July 2023

LTIMindtree

Mumbai, Maharashtra, India

- Spearheaded** a team in achieving **75% acceleration in project completion** by utilizing the Six Sigma approach and **SCRUM project management**, and following **SDLC** and **Agile** methodologies, minimizing project duration from 40 days to 10 days.
- Streamlined **CRUD operations** with Python and SQL **automation**, saving over 4 days of manual effort.
- Constructed Python scripts for **data migration and cleaning**, particularly for **Teradata** and **PeopleSoft DB2** transfers.

RESEARCH EXPERIENCE

Artificial Intelligence Research Intern | Hawkeye MedTech

September 2019 – June 2020

Guide: Dr. Nilakshi Jain, Head of Department, University of Mumbai

Columbia, MD, USA

- Performed **named-entity recognition** to extract data from unstructured text followed by **word embedding** and **similarity matching** using ClinicalBERT, improving disease prediction accuracy by 18%.

Modeling Human Tendencies for Password Guessing 📄 | Publication 📄

August 2020 – January 2022

Guide: Dr. Dhanashree Toradmalle, Professor, University of Mumbai

Mumbai, Maharashtra, India

- Trained LSTM, GRU, and **GAN** seq2seq models totaling **4.8M+ parameters** to generate human-like passwords that matched ~55% of the **14M+ passwords** within 10^9 guesses.

PROJECTS

ConversAI: a Generative AI Chatbot 📄 LLM, Langchain, Embedding, FAISS Vector Database, Mixtral 8x7B, Gemma 7B

- Developed and managed a generative AI chatbot using **GroqAPI**, leveraging **advanced NLP techniques** to enhance user interaction by understanding intent and **generating precise responses**.

IntelliTag: a Smart Tag Recommendation System 📄

TF-IDF, OneVsRest, BERT, Llama2

- Developed a **multi-model content-based recommendation system** for Stack Overflow using **NLP techniques** and pre-trained models, optimizing tag prediction accuracy with a **43% increase in F-1 score**.

EduPredict: Student Performance Forecaster 📄

Decision Tree, Random Forest, XGBoost, Ensemble Learning

- Enhanced model performance by applying **ensemble learning** techniques, **boosting F-1 score to 0.86**.
- Implemented a **REST API** to enable seamless **integration** with other applications and **deployed** the model via Streamlit, **optimizing performance, scalability, and reliability**, and increasing user engagement.

ACHIEVEMENTS / EXTRA-CURRICULARS

2024 – President, Data Science Club: Responsible for mentoring students, inviting industry experts and organizing events.

2023 – Pat on the Back: Performance Award from the manager for my exemplary contributions to deliver the project 'MFGOG'.

2018 – Academic Merit: University Honors Award for the best academic performance among all freshmen.