

Dhairya Jayesh Chheda

+1 930-333-5591 | dhairyajayeshchheda@gmail.com | [linkedin.com/in/chheda-dhairya](https://www.linkedin.com/in/chheda-dhairya) | github.com/DhairyaC | [Portfolio](#)

EDUCATION

Master of Science (M.Sc.) in Data Science

Indiana University – Bloomington, IN, 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++, HTML/CSS, PySpark

Frameworks: Numpy, Pandas, Matplotlib, Scikit-Learn, NLTK, Spacy, Tensorflow, Keras, Pytorch, HuggingFace

Machine Learning: Statistics, EDA, Feature Engineering, Dimensionality Reduction, Supervised Learning, Unsupervised Learning, Predictive Modeling, Regression, Classification, Decision Trees, Random Forest, XGBoost, SVM, Naive Bayes, K-Means Clustering, Multi-Modal Deep Learning Techniques, Natural Language Processing, Generative AI, Large Language Models (LLM)

Developer Tools: Google Cloud Platform (GCP), Amazon Web Services (AWS), GitHub Actions, Docker, Jenkins, JIRA, DVC

PROFESSIONAL EXPERIENCE

Machine Learning Engineer Co-op Intern

Jiseki

September 2024 – Present

San Mateo, CA, USA

- 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 Summer 2024 Intern

Indiana University

May 2024 – July 2024

Bloomington, IN, USA

- Involved in curation of new text dataset of 10K+ records for comprehensive exploratory data analysis and model development.
- Optimized multi-label text classifiers using **RoBERTa** and **active learning**, achieving 60% F-1 score with limited labeled data.
- Worked on **Information Extraction** in a **Few-Shot Incremental** setting for meta-analysis using LLMs based on **OpenAI API like GPT-4, Langchains** and **Retrieval-Augmental Generated (RAG)** pipelines.
- Enhanced a text summarization model utilizing **BART** to visualize insights from a large corpus of visually rich documents.

Software Engineer

LTIMindtree

June 2021 – July 2023

Mumbai, Maharashtra, India

- Streamlined CRUD extraction 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.
- 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%.

RESEARCH EXPERIENCE

Artificial Intelligence Research Intern | Hawkeye MedTech

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

September 2019 – June 2020

Columbia, MD, USA

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

Modeling Human Tendencies for Password Guessing | Publication

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

August 2020 – January 2022

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

Image and Text Classification using Transfer Learning

EDA, DistilBERT, ResNet, Neural Networks

- Conducted Exploratory Data Analysis (EDA) on **1K+** text files using regex, pandas and stemming.
- Implemented transfer learning on transformer models like **DistilBERT** to classify images and text, achieving 95% accuracy.

Student Performance Prediction using ML

Decision Tree, Random Forest, XGBoost, Ensemble Learning

- Enhanced model performance by applying **ensemble learning** techniques such as the voting, bagging and stacking methods, boosting F-1 score to 0.86, surpassing individual prediction results.
- Created user-friendly applications using **Streamlit** for deploying ML models.

Airbnb Booking Analysis and Recommendation using Machine Learning on GCP

PySpark, BigQuery, Looker

- Utilized **PySpark** for distributed data processing and transformation, reducing data processing overhead by 40%.
- Visualized data on **GCP BigQuery** and **Looker** to uncover trends and insights, and recommend strategic business-decisions.

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'.

2022 – BU of the Year: GoMX Annual Hall of Fame Award for leading a team of 5 members to deliver a client-facing project.

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