Dhairya Jayesh Chheda

☎ +1 930-333-5591 | ✉ dhairyajayeshchheda@gmail.com | in linkedin.com/chheda-dhairya | ⌂ github.com/DhairyaC | 🌐 Portfolio

EDUCATION

Master of Science (M.Sc.) in Data Science	August 2023 – May 2025
Indiana University – Bloomington, IN, USA	*GPA: 3.9/4.0*
Bachelor of Engineering (B.Eng.) in Information Technology	August 2017 – May 2021
University of Mumbai – Mumbai, Maharashtra, India	*GPA: 9.62/10.0*

TECHNICAL SKILLS

Languages: Python, R, SQL, Java, C/C++, HTML/CSS, PySpark
Databases: Data Warehousing, Database Querying, MS SQL, MySQL, PostgreSQL, dbt, NoSQL, Salesforce, Snowflake, Databricks
Frameworks: Numpy, Pandas, Matplotlib, RShiny, ggplot2, Scikit-Learn, NLTK, Spacy, Tensorflow, Keras, Streamlit, Flask
Data Science: Statistics, Hypothesis Testing, A/B Testing, Data Processing, EDA, Data Visualization, Business Intelligence Reports and Dashboards, DAX, Key Performance Indicators (KPI), Feature Engineering, Dimensionality Reduction, Supervised Learning, Unsupervised Learning, Data Mining, Regression, Classification, Decision Trees, Random Forest, XGBoost, SVM, Predictive Modeling
Analytic Tools: Stata, Tableau, Microsoft Power BI, Alteryx, QlikView, Jenkins, JIRA, Microsoft Excel, Microsoft PowerPoint, ETL (Extract Transform Load), Google Cloud Platform (GCP), Amazon Web Services (AWS), GitHub Actions

PROFESSIONAL EXPERIENCE

Machine Learning Engineer Intern	September 2024 – December 2024
Jiseki	*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%.

Data Scientist Intern	May 2024 – July 2024
Indiana University	*Bloomington, IN, USA*

- **Extracted and gathered** 50+ historical reports and refined **index value calculations** by managing insights, streamlining data export process to **Microsoft Excel and relational databases**, achieving 50% improvement in data processing efficiency.
- Conducted Exploratory Data Analysis (**EDA**) on **1M+ data points** to create **dynamic and interactive Tableau dashboards**, enabling data-driven reporting and presenting (**KPIs**) to stakeholders, increasing funding opportunities by $200,000.
- **Communicated** findings from analyses to consultants through **visuals, data tables and models**, and assisted with the development of final reports and presentations to be shared with stakeholders, fostering effective teamwork and project success.

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 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.

PUBLICATION

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

E-Commerce Sales Analysis ⬀	*Excel, Python, Statistical Analysis, EDA, SQL, Microsoft Power BI*

- Designed a real-time **sales dashboard to track KPIs** and monitor data accuracy, increasing operational efficiency by 15%.
- Performed **financial planning and strategic analysis** and conducted time series analysis, identifying **trends, key growth factors and forecasting** sales for the next 15 days, contributing to a 30% increase in annual sales growth.

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 – **Volunteer, SankalpTaru Foundation**: Planted a seed as my contribution towards the mother nature: a DEI initiative.
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.