

# KRISHNA DESAI

Bloomington, IN | +1(930) 904-5193 | [desaikri@iu.edu](mailto:desaikri@iu.edu)

<https://github.com/krishnadesai03> | <https://www.linkedin.com/in/krishna-desai-7162b8224/>

## EDUCATION

### Indiana University Bloomington

Master of Science in Data Science (GPA: 3.7/4.0)

Semester 3 Coursework: Applied Algorithms, Information Visualization, Data Visualization

August 2024 - May 2026

Bloomington, IN, US

### University of Mumbai

Bachelor of Engineering in Information Technology (CGPI: 9.52/10.0)

January 2021 – July 2024

Mumbai, MH, India

## TECHNICAL SKILLS

**Programming Languages:** Python 3.10.9, R, C/C++, SQL

**Application Software:** Google Colab, Jupyter Notebook, Git, Github, Power BI 2, Tableau 2023.2.0, R Studio, Microsoft Excel

**Databases:** SQL (MySQL, PostgreSQL pgAdmin 4 v8), NoSQL (MongoDB)

**Certifications & Training:** DeepLearning.AI Data Engineering Professional Certificate, Linux Professional Training (IIT Bombay)

## WORK EXPERIENCE

### PROJECT 990

Bloomington, IN, USA

Data Engineering Intern

May 2025 – Present

- Designed and deployed a robust Python-based data migration pipeline that transformed unstructured IRS tax filings into structured, queryable datasets—enhancing data accessibility, reliability, and downstream analytics efficiency.
- Queried and examined 2 million+ IRS tax records using advanced SQL techniques, uncovering critical data anomalies, schema mismatches, and reporting gaps, while spearheading initiatives for continuous data-quality improvements and audit readiness.
- Defined, tracked, and enforced data-quality KPIs to keep ingested and processed data in sync, giving stakeholders the confidence to make sound, data-driven decisions.

## TECHNICAL CODING RESEARCH INNOVATION

Mumbai, India

Data Science & Machine Learning with Python Intern

January 2022 – April 2022

- Conducted analytics on an employee dataset having 35 attributes, predicting retention using Python and ML.
- Designed and implemented ETL data pipelines ensuring high-quality datasets from 2940 entries.
- Delivered 8 visualizations (pie charts, column graphs, heatmaps), aligning insights with business performance metrics.
- Developed predictive models with 98.15% training accuracy and 85.26% testing accuracy, supporting resource allocation decisions.

## PROJECTS

### IMDB EMOANALYSIS: UNDERSTANDING MOVIE SENTIMENT WITH DEEP LEARNING

May 2024

- Developed sentiment analysis models using LSTM and Bi-LSTM, achieving an accuracy of 0.86 on the final model.
- Analyzed a 50,000-record dataset, solving complex data challenges and implementing 5-6 crucial pre-processing steps.
- Applied analytical thinking to derive insights into audience sentiment trends, validating the model on live data with 0.85 accuracy.
- My model outperformed others by 7.5% in terms of accuracy and prediction quality.

### BINARY IMAGE CLASSIFICATION FOR WASTE MANAGEMENT

- Developed a deep learning model to classify waste images into 2 categories namely recyclable and non-recyclable images.
- Used a dataset of 1,800 training and 450 testing images, ensuring balanced category representation for effective model evaluation.
- Implemented data preprocessing techniques, like image augmentation and normalization, to enhance model performance by 10%.
- Designed and trained a CNN-based model using TensorFlow and Keras, achieving an accuracy of 72% on the test dataset.
- Leveraged GCP for data storage, model training using AI Platform, and deployment through Cloud Run for real-time classification.

### GLOBAL SUPERSTORE ANALYTICS – AWS END-TO-END DATA PIPELINE AND DASHBOARD

- Designed and implemented a scalable data pipeline on AWS (S3, Glue Crawler + ETL, Redshift) to process Global Superstore data, improving overall data usability by ~10% and reducing ETL runtime by ~5% through schema optimization and structured transformations.
- Queried and analyzed transformed datasets with Athena and built an interactive Power BI dashboard (10+ KPIs, sales trends, profitability insights), helping reduce manual reporting effort by ~15% and making insights more accessible to business stakeholders.
- Configured IAM roles, resolved connectivity/authentication issues, and implemented CloudWatch monitoring, improving pipeline stability and reducing troubleshooting time by ~10%.