

DEVARATA OZA

21 Courtlandt St, New Brunswick, New Jersey, USA, 08901

📞 732-522-6179 ✉️ do309@rutgers.edu 🔗 <https://www.linkedin.com/in/devarata/> 🌐 <https://github.com/devarata>

EDUCATION

Rutgers University, New Brunswick

Sep 2022 – May 2024

Master of Science in Computer Science

GPA: 3.8/4.0

Thadomal Shahani Engineering College, Mumbai

Aug 2015 – May 2019

Bachelor of Engineering in Computer Engineering

CGPA: 8.34/10

TECHNICAL SKILLS

Languages: Python, C++, JavaScript, SQL, Hive Query Language (HQL), Shell script

Frameworks & Libraries: PySpark, NLTK, spaCy, Pandas, NumPy, Matplotlib, scikit-learn, OpenCV, Tensorflow, PyTorch, Apache Spark, Apache Hive, Apache Hadoop, MapReduce, SciPy, ReactJS, Airflow, Docker, Raspberry Pi

WORK EXPERIENCE

Center for Advanced Biotechnology and Medicine, Rutgers University

Jun 2023 – Sep 2023

Research Assistant (Machine Learning Engineer)

New Brunswick, NJ

Project: Mice Monitoring System | **Skillset:** PyTorch, Deep Learning, Computer Vision, ReactJS, Raspberry Pi, SQL

- Created scalable cost-saving mice cage monitoring system with 160 cages per rack using YOLOv8 model trained on a custom dataset, comprising over **10000 annotated images** from video recordings and web scraping.
- Designed Pi-based night vision system for mice monitoring, processing video every 10s with **optimized storage**.
- Developed 3 ML models (mice counting, injury detection) and **mice disease detection** (Alopecia, Dermatitis, Tumors, Nasal discharge), while designing a **web interface using ReactJS** integrated with **Tecniplast API**.

Deutsche Bank

Jul 2019 – Jul 2022

Senior Analyst (Data Scientist)

Pune, India

Project: Rate Card and Matter-Invoice | **Skillset:** PySpark MLlib, sklearn, NLP, Spacy, NLTK, matplotlib, Hive, SQL, HQL

- Designed **machine learning-based multiclass classification system** for **automated invoice classification**, processing **approx. 4 million invoices in minutes**.
- **Optimized hyperparameters** to elevate **XGBoost Classifier** accuracy from **92.7% to 97.8%**.
- **Boosted classification by 20%** through **feature engineering** and **spaCy's Named Entity Recognition**.
- Employed **PySpark, Apache Hadoop, and Apache Hive** for Big Data processing of **5 GB daily** in Parquet format.
- Developed **Error Logging System** for Spark jobs and data pipelines, **enhancing troubleshooting by 50%**.
- **Led migration of 1 TB+ legal historical data from Oracle to Big Data** for enhanced processing.
- **Designed 200+ SQL queries** for data consolidation across data marts, adhering to business logic.

PERSONAL PROJECTS

Walk-these-ways | PyTorch, Isaac Gym Simulator, Reinforcement Learning, Actor-Critic Policy

Sep 2023 - Dec 2023

- Developed neural network for **robotic dog** in Isaac Gym, focusing on efficient **locomotion** & collision-free navigation.
- Implemented **Actor-Critic Proximal Policy (PPO)** navigation policy along with **reward function** in Isaac Gym using **reinforcement learning** for **precise, collision-free movement**.
- **Integrated obstacle avoidance** into the robot, enabling **adaptive navigation** and environmental awareness in complex scenarios.

Youtube Data Analysis | AWS (S3, Glue, Lambda, Athena), Python, Pandas, SQL

Jun 2023 - Jul 2023

- Analyzed large YouTube trendings dataset (40,949+ records) using **Python** and **Pandas**.
- Transformed **JSON and CSV to Parquet**, achieving a **40% reduction** in storage space.
- Developed **Glue crawlers & ETL jobs** using **AWS Lambda, Athena, & S3**, boosting query performance by **30%**

Question Answering System | Natural Language Processing, BERT, PyTorch, ReactJS

Mar 2023 - Apr 2023

- Devised **Question-Answering system** for low-resource languages using **mBERT** and **XLM-R** models.
- **Enhanced model performance** by incorporating **SQuAD, chaii, and FLoRes** datasets, expanding training data for **101 languages with diverse question-answer pairs**, while eliminating noise and normalizing text.
- Improved performance by **fine-tuning state-of-the-art NLU models (mBERT and XLM-R)** on preprocessed data, surpassing **TyDiQA baseline with F1 scores of 39.1 to 78.7** across multiple languages.

ACHIEVEMENTS & AWARDS

- Recognized as department's best-performing employee, receiving **Deutsche Bank Recognition Award** for H2 2020.
- Awarded 1st place at TSEC Project-Expo'19 for **Smart Waste Management**, Internet of Things (IoT) Project