DEVARATA OZA

📍 New Brunswick, New Jersey, USA, 08901 📞 732-522-6179 ✉ do309@rutgers.edu

in https://www.linkedin.com/in/devarata/ ⬤ https://github.com/devarata 🌐 https://devarata.vercel.app/

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, India   **Aug 2015 – May 2019**
Bachelor of Engineering in Computer Engineering   *CGPA: 8.34/10*

TECHNICAL SKILLS

Languages & Databases: Python, C++, JavaScript, Shell, SQL, Apache Hive, HiveQL, Oracle, MySQL, MongoDB

Big Data & Cloud: PySpark, Apache Spark, SparkSQL, Apache Hadoop, HDFS, MapReduce, AWS, Docker

ML & Analysis: Spark MLlib, TensorFlow, Pytorch, Pandas, NumPy, Matplotlib, scikit-learn, NLTK, spaCy, OpenCV

Web Development & DevOps: ReactJS, BeautifulSoup (bs4), Requests, Git, Airflow, CI/CD

WORK EXPERIENCE

Center for Advanced Biotechnology and Medicine, Rutgers University   **Jun 2023 – Oct 2023**
Research Assistant (Machine Learning Engineer)   *New Brunswick, NJ*
Project: Mice Monitoring System | **Skillset:** *Python, 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 Engineer)   *Pune, India*
Project: ELM Reporting | **Skillset:** *PySpark, MLlib, spaCy, Shell, matplotlib, Hive, SQL, HQL, Tableau, CI/CD*
- Designed **machine learning-based multiclass classification system** for **automated invoice classification**, processing **approx. 4M invoices in minutes**, replacing the previous **SQL system**.
- **Optimized hyperparameters** to elevate **XGBoost Classifier** accuracy from **95.7%** to **97.8%**.
- **Boosted classification** by **20%** using **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.
- **Accelerated reporting time** from 8 hours to 2 hours with **50+ data marts** and **10+ ETL pipelines**.

PERSONAL PROJECTS

Youtube Data Analysis | *AWS (S3, Glue, Lambda, Athena), Python, Pandas, SQL*   **June 2023 - July 2023**
- **Analyzed large YouTube trendings dataset** (40,949+ records) using **Python** and **Pandas**.
- **Designed an AWS Glue catalog** with **2 tables** for enhanced data querying and integration.
- **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%**

Stock Market Data Analysis | *Apache Kafka, Python, AWS (S3, Athena, Glue)*   **March 2023 - May 2023**
- **Built a real-time stock market engine** with **Kafka** for streaming between producers and consumers.
- **Developed Python scripts** for data generation & simulation, storing on **Amazon S3** with 100,000+ records in 5s.
- **Deployed AWS Glue Crawler and Catalog** on **Amazon S3**, achieving a **93ms data scan time** and rapid analysis via **Amazon Athena**.

Music Recommender System | *Annoy, Tensorflow, Python, Docker*   **Jul 2021 - Aug 2021**
- **Developed recommendation system** with **Annoy**, achieving a cosine similarity of **0.85** among top 10 neighbors.
- **Processed 13,000 audio examples**, generating **128D embeddings per sec** with Max Audio Embedding Generator.
- Implemented Annoy with **100 trees** for fast music recommendations, achieving an **average response time of 50 ms**.

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)