DEVARATA OZA

Apt 204, 100 Hiram Square, New Brunswick, New Jersey, USA, 08901

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

EDUCATION

Rutgers University, New Brunswick   **Aug. 2022 – May 2024**
Master of Science in Computer Science   *GPA: 3.83/4.0*

Thadomal Shahani Engineering College - Mumbai University, Mumbai, India   **Aug. 2015 – May 2019**
Bachelor of Engineering in Computer Engineering   *CGPA: 8.34/10*

TECHNICAL SKILLS

Languages & Database: Python, C++, JavaScript, Java, SQL, HTML, CSS, R, Apache Hive, Oracle, MySQL, MongoDB

Frameworks & Libraries: Pyspark, NLTK, Spacy, Pandas, numpy, matplotlib, scikit-learn, OpenCV, Tensorflow, Pytorch, Apache Spark, Apache Hadoop, MapReduce, Node.js, React.js, Express.js, BeautifulSoup, MLlib, Selenium, pytest

Tools & IDE: Git, Bitbucket, JIRA, Confluence, Raspberry Pi, Docker, Kubernetes, Tableau, CI/CD, Postman, Ansible

Technical & Soft Skills: Extract Transform Load (ETL), Data Warehousing, Big Data Processing, Data Modeling, Data Pipelines, Machine Learning, Data Science, IoT, communication, analytical skills, detail-oriented, adaptability, collaborative

WORK EXPERIENCE

Deutsche Bank   **Jul 2020 – Jul 2022**
Senior Analyst (Data Engineer)   *Pune, India*
Project: Rate Card and Matter-Invoice | **Skillset:** *PySpark, MLlib, Spacy, WordCloud, matplotlib, Apache Hive, SQL, HQL*
- Built supervised **multiclass classification** machine learning models and data pipeline to automate the classification of invoices, replacing inaccurate SQL-based system, enabling users to parse **approx. 4 million invoices in minutes**
- Improved **Random Forest Classifier** model from **95.7% to 96.8%** accuracy with hyperparameter optimization.
- Extracted entity information from invoice line items using **Name Entity Recognition**, improving **flagging by 20%**
- Developed a **33-cluster** word cloud from transactional data to visualize and validate the results of the classifier models

Deutsche Bank   **Jul 2019 – Jun 2020**
Graduate Analyst (Data Engineer)   *Pune, India*
Project: Details Individual | **Skillset:** *SQL, Apache Hive, Apache Hadoop, PySpark, MapReduce, HQL, Shell Scripting*
- Played key role in **migrating 25 GB historical data** warehouses from **Oracle to Big Data** for efficient processing.
- **Reduced** daily reporting time from **8 hours to 2 hours** by developing data marts and data pipelines using **PySpark** and **Hive Query Language (HQL)** to **process 5 GB of daily data** and shift business logic to **Big Data** platform.
- **Streamlined tendering process** by designing a data pipeline to create a **consolidated data mart from 5 tables**
- **Engineered row-level security** to restrict data row access based on region user login. This approach was **optimized** by applying "**Matrix Chain Multiplication**" concept to SQL tables enhancing querying speed by approx. **10%**.
- **Optimized** performance of **40+** data extracts and **SQL** scripts in **Tableau** resulting in accelerated report generation.

PERSONAL PROJECTS

Maze Solver Visualizer | *React.js, JavaScript, Python*   **Sep 2022 - Oct 2022**
- Generated a maze with **BFS algorithm** in **Python** and visualized paths of **3 variations of Astar using React.js.**

ZEST | *React.js, Express.js, Node.js, Firebase, WebRTC, JavaScript, Material UI*   **Aug 2020 - Sep 2020**
- Created a **4-in-1 web application** that combines features of **Zoom, Evernote, Slack,** and **Trello.**
- The **all-in-one** platform provides **video calling, note-taking, chat, project planning, and persistent sign-in/sign-up** functionalities for seamless collaboration, **boosting** team **productivity** by up to **25%**

Covid-Tracker | *React.js, JavaScript, Leaflet, Context API, React Hooks, Material UI, Chart.js*   **Jul 2020 - Aug 2020**
- A **COVID-19 tracking website** with **global** and **country-wise cases** dropdown selection using **disease.sh API**, including stats on **active cases, recoveries, & deaths**, alongside **map visualization** of affected areas using **Leaflet**

YourVocab | *Python, BeautifulSoup, Pandas, Selenium, VB script*   **Feb 2020 - Mar 2020**
- **Eliminated manual creation** of personal vocabulary lists by developing program that streamlines the entire process
- Developed a Tkinter UI to enter a word. Program leverages **web scraping, Pandas, and selenium** to **automate vocabulary list creation** and uploaded the updated excel to **OneDrive, reducing manual data entry** by **99%**

ACHIEVEMENTS & AWARDS

- Received **2000 $** equivalent as **Deutsche Bank Recognition Award** for outstanding performance in February 2021.
- Awarded **1st place** at TSEC Project-Expo'19 for **Smart Waste Management, Internet of Things (IoT) Project**