ROHAN PATEL

469-679-1892 | patelrohan0222@gmail.com | linkedin.com/in/rohanp8980 | github.com/rohan8980

SUMMARY

Gained valuable experience as a Data Engineer, executing data analysis, migration, ETL and SQL. Possess a diverse skill set including proficiency in Data Science, Python, and SQL with strong foundation in machine learning, statistical learning, and data mining

EDUCATION

Masters: Data Science - 3.9/4, University at Buffalo, The State University of New York, May 2024
- Statistical Learning, Data Mining, Numerical Mathematics, Machine Learning, Probability, Python, R Programming, MATLAB, Spark, Hadoop, SQL

Bachelors: Computer Engineering - 8.74/10, GTU, India, May 2019
- Big Data Analytics, Data Mining and Business Intelligence, Artificial Intelligence, DBMS, Data Structures, Advanced Engineering mathematics, Theory of Computation, Operating System, Advanced Java, .NET Technology, C++

SKILLS & TOOLS

Programming: Python, MS SQL, PostgreSQL, MySQL, R, Java, Hadoop, Spark, .NET, Angular, Javascript, CSS, HTML
Tools: Anaconda, SSMS, Jupyter Notebook, Tableau, Git, MATLAB, pgAdmin, PowerBI, AWS, Excel, VSCode, Bootstrap
Skills: Machine Learning, Deep Learning, Programming, Database, Data Scraping, Data Wrangling, Computer Vision, Natural Language Processing NLP, Large Language Models LLMs, Recommender Systems, Sentiment Analysis, Time Series Forecasting, Data Visualization, Tensorflow, Keras, Pytorch, Pandas, Numpy, Version Control, Communication

EXPERIENCE

Data Engineer, Meditab Software Pvt. Ltd., Ahmedabad, India: January 2019 - January 2022
- Coordinated ETL driven data migrations for healthcare organizations like IPAs, TPAs, MSOs, and ACOs from their legacy systems to MSSQL Server, seamlessly transferring over 5M patient records and 200K provider profiles, improving database speed by 35% and enhancing security by 20%.
- Managed 2.5 TB of healthcare data, ensuring HIPAA compliance. Implemented anonymization for 20M patient records and over 1M provider profiles for claims adjudication, authorizations, and case management modules.
- Conducted data analysis to develop 20-25 custom dashboards per client to provide crucial insights and saving $2 million annually.
- Implemented robust data backup, conducting weekly backups of 1.5 TB of critical data, achieving 99.9% recovery success rate.
- Engineered automated custom data importing, processing 70-80 files monthly, achieving very high accuracy of almost 95%.

PROJECT

HeroClassifier: February 2024
- Orchestrated a deep learning project for image classification of 10 superheroes, showcasing proficiency in data scraping, preprocessing, and model deployment
- Utilized CNN and transfer learning with the VGG16, achieving an impressive 88.67% accuracy on unseen data
- Implemented a user-friendly interface leveraging Streamlit for model deployment, enabling users to upload superhero images and receive predictions instantly at heroclassifier.streamlit.app

RecipeRover: April 2024
- Spearheaded the development of an end-to-end recipe recommendation system, encompassing web scraping of over 7500 recipes, meticulous data cleaning, preprocessing, modeling, and web application deployment
- Implemented a content-based recommendation algorithm leveraging TF-IDF vectorization and cosine similarity calculations to suggest similar recipes tailored to user preferences, achieving an average recommendation speed of under 100 milliseconds.
- Orchestrated the deployment as a user-friendly web application using Streamlit at reciperover.streamlit.app, successfully attracting engagement from a targeted user base of over 50 users within the initial launch phase

ReviewRadar: March 2024
- Developed a sentiment analysis model using NLP techniques for a massive dataset of 11 million reviews. Achieved 88.5% accuracy on test data using Support Vector Machines
- Implemented data cleaning and pre-processing techniques including tokenization, stop words removal, stemming, lemmatization and text normalization with spaCy.
- Generated a word cloud visualization to communicate sentiment trends within the reviews effectively

Reducing Telecom Customer Churn: October 2023
- Unveiled key drivers of telecom churn to anticipate customer defection and implement targeted retention strategies.
- Built a Support Vector Classifier on the Customer Churn data to predict customer churn with an accuracy of 81 percent
- Analyzed data to collect information used to build targeted customer retention tactics thus lowering client churn

Sunrise Transformers (sunrisetransformers.com): November 2019
- Coded and deployed a website using the latest technologies for a client to boost the reach of the company by 30%
- Employed Microsoft SQL Server for managing and storing various products, services offered, and any inquiries of any buyer
- Integrated backend coding using .NET technology with MS SQL to attain smooth data flow from DB to the front-end
- Implemented typescript for integrating with .NET API to achieve a quick page loading time of under 2 seconds for each page

CERTIFICATIONS

- Google Data Analytics Professional Certificate, June 2022 - Acquired Skills (Data cleaning, Data Analysis, & Visualization) and mastered Tools (SQL, R programming, Tableau, Spreadsheets)
- Data Science using Python Workshop, January 2018 - Skills (Data Science Workflow, Data Manipulation with Pandas, Data Visualization, Data Preprocessing, ML Python Libraries like NumPy and SciPy)

LEADERSHIP AND VOLUNTEER EXPERIENCE

Event Manager, GTU Central TechFest'18, Ahmedabad, Gujarat: March 2018
- Managed a National Level Technical Event - Programming Quiz among 500+ participants

Volunteer, Chemo Confluence'17, Ahmedabad, Gujarat: March 2017
- Volunteered at a State Level Event - Chemo-o-Live at Vishwakarma Govt. Engg. College, India