+91-9424592907
contact@shubhanshutrip.com

SHUBHANSHU TRIPATHI

shubhanshutrip.com
ShubhanshuTrip

TECHNICAL SKILLS

- Python, SQL, PySpark *(Apache Spark)*, ETL *(Extract, Transform, Load)*, Big Data, MySQL
- Microsoft Azure *(Azure Synapse Analytics, Azure Databricks, Azure Data Factory, Azur Data Lake, Azure SQL, Azure Logic Apps)*, Amazon Web Services *(AWS)*, Data Science, Machine Learning, Git / GitHub

EMPLOYMENT

Senior Software Engineer | **Capgemini** | **March 2022 - Present**

Data Engineering

Operational Data Store, Data Hubs, Data Marts, Extracts Generation

- Migrated on-prem big data ETL processes to cloud, by creating storage event and tumbling window triggered pipelines in Azure Synapse Analytics using PySpark notebooks
- Implemented pre-load, data quality, and data control checks
- Performed data cleaning and applied transformations on parquet and CSV big data feeds
- Implemented change data capture (CDC) process to store transformed data with SCD Type 2 implementation
- Loaded data to Azure SQL Database and Delta tables, & generated CSV and TXT extracts by creating dynamic pipelines to selectively fetch data by joining multiple source tables
- Implemented dynamic pipeline status email notification functionality using Azure Logic Apps and Web Activity

Software Engineering

Status and Metadata Tracking

- Dynamic real-time status and metadata tracking using Python
- Extracted metadata properties and row counts dynamically from DAT and TXT files

Miscellaneous

- Optimised Python programs to reduce average runtime by 23%
- Automated Excel macro run by creating Python scripts, to email daily consolidated status reports

EDUCATION

Master of Technology (Computer Science and Engineering) | **Birla Institute of Technology, Mesra** | **CGPA: 8.06** | **July 2018 - July 2020**

Thesis Work

Diabetes Prediction using Machine Learning (bit.ly/dbtspred)

- Achieved up to 81.6% accuracy in Diabetes Prediction on Pima Indians Diabetes Database with Random Forest classifier
- Applied and analysed accuracies of “K-Nearest Neighbors, Support Vector Machine, Decision Tree and Random Forest” classification algorithms for diabetes prediction
- Achieved up to 7.04% improvement in the accuracy of Decision Tree classification algorithm for Diabetes Prediction
- Predicted missing values present in the dataset using a set of “Linear Regression, Support Vector Regression, Decision Tree and Random Forest” regression algorithms
- Performed Dataset Balancing using SMOTE algorithm and then Feature Scaling

PROJECT WORK

CoWIN Vaccine Notifier (bit.ly/covantf)

- Developed a Python notebook to notify the user, as soon as any desired Covid Vaccine is available on CoWIN website for booking
- Implemented 4 dynamic filters on the Vaccination calendar received as a JSON response from Co-WIN API
- Helped more than 30 people to get Covid Vaccines using this notifier

ACHIEVEMENTS AND CERTIFICATIONS

- Achieved 3rd rank amongst 250 teams in i3i 2023, a Capgemini’s internal hackathon for insurance domain
- Passed “AWS Certified Cloud Practitioner” certification exam of Amazon Web Services (bit.ly/stawsccp)
- Completed “Complete Python Bootcamp” course from Udemy (bit.ly/pyud)
- Passed “Basics of Data Science and Machine Learning” course from Coding Ninjas (bit.ly/dscmlcn)
- Qualified GATE (CS) exam organised by Indian Institute of Science (IISc), Bangalore