

## EMPLOYMENT

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

Operational Data Store, Data Hubs, 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 process to store transformed data with SCD Type 2 implementation
- Loaded data to Azure SQL database and Delta tables, & generate fixed length TXT extracts, with selectively fetched data by joining multiple source tables
- Implemented dynamic pipeline status email notification functionality using Azure Logic apps and Web activity

Status and Metadata Reports Generation

- Python based automation for status and metadata report generation of data feeds
- Extracted metadata properties and row counts dynamically from DAT and TXT data feeds

Miscellaneous

- Optimized ETL python scripts average runtime by 23%
- Automated Excel macro run by creating Python scripts to send daily consolidated ETL jobs status report email at scheduled time

## EDUCATION

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

### Thesis Work

Diabetes Prediction using Machine Learning ([bit.ly/dbtspred](https://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](https://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

## CERTIFICATIONS AND ACHIEVEMENTS

- Finalist (top 10) among 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/stawssccp](https://bit.ly/stawssccp))
- Completed “Complete Python Bootcamp” course from Udemy ([bit.ly/pyud](https://bit.ly/pyud))
- Passed “Basics of Data Science and Machine Learning” course from Coding Ninjas ([bit.ly/dscmlcn](https://bit.ly/dscmlcn))
- Qualified GATE (CS) exam organised by Indian Institute of Science (IISc), Bangalore
- Earned Silver badge (3/5 stars) in Python on HackerRank

## Languages and Technologies

- Python, SQL, PySpark, Data Science, Machine Learning, Markdown
- Azure, Amazon Web Services (AWS), Google Colab, Anaconda, Jupyter Notebook, Visual Studio Code