Tanmay Bhalavi

tanmaybhlv@gmail.com | +91 79997 90443

Career Highlights

Experienced data professional proficient in handling and analyzing diverse data sets to drive organizational productivity. Skilled in utilizing a wide range of data science programming languages and tools, including R, Python, SQL, and Spark. Seeking to leverage expertise and experience as a Data Scientist to extract actionable insights and deliver impactful solutions.

Searce – 07/2022 – Present

Data Engineer

- Built optimal data pipeline architecture using Beam, Airflow, and other technologies to assemble large, complex datasets for advanced analytics and modeling purposes
- Migrated data from MongoDB to BigQuery through Fivetran and transformed & orchestrated the data from raw BQ layer to enriched & curated layer using Airflow
- Integrated SFDC reports with BigQuery to enhance data compatibility and enable efficient data processing for data science initiatives
- Conducted workshops on BigQuery, coaching participants on identifying and addressing use cases for its advanced features
- Migrated over 250 collections from MongoDB to BigTable using Python scripts, employing multi-threading and automation for an efficient and seamless migration process
- Developed a data pipeline from AWS RDS to GCP BigQuery for migrating over 10 TB of data, leveraging AWS DMS, GCP Transfer Service, and Cloud Scheduler
- Migrated 30 GB of data from AWS DynamoDB to BigQuery using Dataflow, preparing it for machine learning model development on VertexAI. Provided guidance on best practices for deploying ML frameworks using BQML.

Searce – 07/2021 – 06/2022

Data & Analytics Practice Engineer

- Leveraged Looker Studio to build and maintain interactive dashboards and reports for clients and internal teams, enabling data-driven decision-making
- Conducted data cleaning, validation, and transformation on large datasets to ensure accuracy and completeness, preparing the data for advanced analytics and modeling
- Developed rich interactive reports, dashboards, and user interface elements by integrating components from multiple sources, enhancing data visualization and insights
- Translated raw data into meaningful information and actionable insights using automated technologies, extracting data from primary and secondary sources. Communicated findings to management for improved business process productivity

Skills

- **Languages/Scripting:** Python, SQL, PySpark
- **Data Manipulation:** Pandas, NumPy
- **Data Visualization:** Matplotlib, Seaborn, Plotly
- **Cloud Platforms:** GCP, AWS
- **RDBMS:** MySQL, PostgreSQL, Cloud SQL, RDS
- **Storage:** GCS, BigQuery, BigTable, Cloud SQL, S3
- **ETL:** Dataflow, Data Fusion, Dataproc, Beam, Spark
- **BI Tools:** Looker Studio, Tableau
- **Workflow Platform:** Airflow, Cloud Composer
- **Version Control Tool:** Git

Certifications

- GCP Professional Machine Learning Engineer
- GCP Professional Data Engineer
- GCP Associate Cloud Engineer

Academic Details

Degree University Year	CGPA	Specialisation
MTech IIT Kharagpur 2019-2021	8.26/10	Industrial Engineering & Management
B.E. RGPV Bhopal 2014-2018	7.16/10	Mechanical Engineering

Internships

iNeuron	Duration: 3 Months	Period: May/2021 – July/2021
Deep Learning Engineer During the internship, I built a camera-mounted robotic arm with basic functionalities to collect real-time data and created a solution for businesses, workplaces, and stores to improve the practice of COVID guidelines of wearing masks. I also collected and labelled images dataset & developed a DL model through various algorithms to detect the person without masks.		

The Sparks Foundation	Duration: 1 Month	Period: 01/Oct/2020 – 31/Oct/2020
Business Analyst One of the tasks of this internship involved analysing the COVID-19 public dataset from the start of the pandemic and creating interactive dashboards on Tableau. In the second project I created a clustering-based ML model to identify segments of customers who are interested in specific products or product lines.		

BEPEC Solutions	Duration: 3 Months	Period: August/2020 – October/2020
Data Scientist I developed an ML solution for predicting employee attrition rate by cleaning, wrangling, and analysing HR data. I deployed the model through Flask and monitored its performance for further improvements.		

Publications

IEEE Xplore	Feb 21, 2022	ICMIAM
Network Intrusion Detection using Principal Component – Mahalanobis Taguchi System A novel approach to network intrusion detection by utilizing the MTS. A supervised learning approach based on the Mahalanobis distance is used to distinguish attack instances from normal instances and a PCA based framework is used for feature reduction. The proposed method PC-MTS is shown to achieve high accuracy on two network intrusion data sets NSL-KDD and IDS2017 surpassing existing methods.		

Smithers Information Ltd	May 1, 2017	Polymers from Renewable Resources
Natural Fibre Reinforced Polymer Composite Materials – A Review This review discusses the use of chemically treated natural fibres as a reinforcement in polymer composites. The chemically treated fibres exhibit better mechanical properties due to improved interfacial adhesion with the polymer matrix. The article also highlights the potential industrial and structural applications of natural fibre reinforced composites.		