

Hrishikesh Dhole

Data Engineer

858-308-2924 | hrishikeshdhole0@gmail.com | [LinkedIn:- Hrishikesh](#) | [GitHub:- Hrish52](#)

SUMMARY

Data Engineer who designs and tunes cloud-scale pipelines that process 500 000 + records nightly and keep key Power BI dashboards current within minutes. Hands-on with Python, SQL, PySpark, Spark, Airflow, and BigQuery, delivering 30 % faster data-processing times through smart partitioning and query optimization. Skilled at translating business goals into star-schema models and orchestrating end-to-end ELT on AWS and GCP. Bolstered by coursework and lab builds in Snowflake, Redshift, Terraform-based CI/CD, and data-quality tooling—ready to apply these practices in production to deliver fast, trustworthy analytics.

EXPERIENCE

- Data Scientist Apprentice, Costco Wholesale, Seattle, Washington, United States** **Jan 2025 - Jun 2025**
- Led a 5-member capstone project that collected and analyzed ~350 k+ social-media comments from Reddit, Instagram, X, and Costco.com, delivering data-driven insights to Costco's IT Data Science team.
 - Automated data collection with Apify, Phantombuster, Selenium, and BeautifulSoup, integrating multi-platform scraping and API extraction to cut manual data-prep time by 80 % and ensure consistent, high-quality inputs for sentiment analysis.
 - Processed and normalized unstructured comment data to create a clean, analytics-ready corpus, enabling faster topic modeling, sentiment analysis, and voice of market insights.
 - Designed a Power BI dashboard that distills social-media comments into sentiment timelines, topic clusters, and top product issues, reducing merchandiser insight-gathering time by 30 %.
 - Prototyped an Azure OpenAI summarization chatbot that answers ad-hoc questions in seconds, using fine-tuned BERT and VADER sentiment signals for accurate, context-aware responses.
- Web Developer Intern, Tantransh Solutions, Nagpur, Maharashtra, India (Remote)** **May 2020 - Aug 2020**
- Designed and developed a blogging platform using Ruby on Rails, implementing user authentication, post creation, comment management to validate core functionality and gather stakeholder feedback.
 - Optimized database schemas and query logic, reducing execution times by 30 % and improving data retrieval speed for complex joins and multi-table queries. Implemented basic caching, database indexing, and load balancing concepts to improve response times and scalability for future deployment.
 - Built responsive, mobile-first front-end interfaces using HTML, CSS, Bootstrap, and JavaScript, ensuring consistent performance across devices for early-stage user testing.
 - Collaborated with cross-functional teams (engineering, marketing, product) to refine feature requirements and present working demos, cutting strategy meeting prep time by 20 %. Integrated Google Analytics to track user interactions, uncover feature gaps, and prioritize development efforts based on real-time usage data.

EDUCATION

- Seattle University, Seattle, United States - Master, Data Science** **Sep 2022 - Jun 2025**
Focused on Statistical Machine Learning, Big Data Analytics, Data Management, Data Visualization, and Cloud Computing, gaining hands-on experience with Python, R, SQL, Hadoop, Spark, and D3.js.
- St. Vincent Pallotti College of Engineering and Technology, Nagpur, India - Bachelor, Computer Engineering** **Jul 2017 - Dec 2021**
Gained a strong foundation in Data Structures, Algorithms, Distributed Systems, Operating Systems, Computer Networks, Theory of Computation, Software Engineering, and Object-Oriented Programming (C++, Java, Python).

CERTIFICATIONS

- Google Cloud Big Data and Machine Learning Fundamentals, Coursera** **Aug 2020**
- Basic Artificial Neural Networks in Python, Coursera** **Sep 2020**
- Neural Networks and Deep Learning, Coursera** **Oct 2020**

PROJECTS

- House Tenure Prediction Analysis, [Link](#)** **Apr 2024 - Apr 2024**
- Built an SVM model on 75,000+ census records, achieving ~80% accuracy through data transformation, feature selection, and hyperparameter tuning.
 - Conducted data analysis and extraction, identifying key predictors like income, education level, and employment status, offering business insights for real estate and housing policies.
- Seattle Bird Call Classification, [Link](#)** **Apr 2024 - May 2024**
- Designed a CNN-based classifier for spectrogram-based classification, achieving 96% binary accuracy and 71% multi-class accuracy using deep learning and adaptive learning rates.
 - Preprocessed 500,000+ audio samples using Librosa (noise filtering, resampling, spectrograms) and developed a scalable data pipeline for real-time ecological monitoring, integrating big data processing & AI-driven automation.

SKILLS & INTERESTS

- Cloud & Big Data:** : AWS, Microsoft Azure, Google Cloud Platform, BigQuery, Azure Synapse, Hadoop MapReduce, PySpark, Spark Streaming, Apache Cassandra, DynamoDB, Data Pipelines, Data Warehousing, API Integration, Data Lake Management
- Programming & Scripting** : Python, SQL, R, C++, DAX
- Machine Learning & AI** : NLP (BERT, VADER), CNN, Logistic Regression, Linear Regression, Reinforcement Learning, Support Vector Machine, Clustering, Text Classification
- Data Visualization** : Tableau, Power BI, Looker, Alteryx, D3.js, Matplotlib, Seaborn, Plotly
- Frameworks and Tools** : TensorFlow, Keras, Flask, Rasa, Dialogflow, Google Workspace, Docs, NLTK (NLP Analysis), Excel, Google Sheets, Jira, Docker, VS Code
- Data Engineering and Modeling** : Data Extraction, Data Transformation, Data Modeling, ETL Pipelines, Data Quality Managemen, Data Integration, Batch and Real-Time Data Processing
- Business & Management** : Business Insights, Usability, Customer Service, Product Management, Process Optimization
- Domains** : E-Commerce, Retail, Social Media, Marketing Analytics

PUBLICATIONS

- Survey on Discrete Gesture Recognition Techniques, International Journal of Advance Research, Ideas and Innovations in Technology** **May 2021**
Published peer-reviewed paper on Gesture Recognition Techniques (IJARIIT, May 2021), comparing vision-based, data glove, and marker-based methods for Human-Computer Interaction (HCI). Evaluated algorithms like Viola-Jones, Haar Cascade, and Arc of Reference for accuracy, efficiency, and real-world applications in AI-driven interfaces and assistive technologies.