Harshal Ravindra Sonawane

Technical Data Analyst | Python | Oracle SQl | AI/ML

✉ harshals14014@gmail.com   📞 +91 9049491400   ⭘ Github   in Linkedin

PROFESSIONAL SUMMARY

Solid background in Python programming, AI/ML integration, and Oracle SQL/PLSQL, and Analytics. ability to automate ETL pipelines, and extract insights from complicated datasets. competent in assessing new AI technologies for business applications, developing dashboards (Power BI, Tableau). Experienced in designing and maintaining PL/SQL procedures and functions for ETL processes. Experienced in researching and applying emerging AI/ML tools

EXPERIENCE

Junior Data Analyst, *Balbird Industries Ltd*   04/2023 – 09/2023
                                                                                                                                                        Pune, India
- "Developed **Oracle SQL** queries and **Python scripts** for automating data extraction and reporting workflows.
- Reduced manual reporting by **20%** through **Tableau** and **Power BI** automation..
- **15%** more targeted interaction was achieved by applying **clustering algorithms** to deliver consumer segmentation insights and Presented to Stakeholders
- Automated **ETL pipelines** and created standardized reporting processes, ensuring data quality and consistency.
- Maintained data integrity across systems by enforcing data quality rules and validations.

EDUCATION

B.E. - Mechanical Engineering - 7.8 CGPA, *Sinhagad College of Engineering, Pune*   2024
                                                                                                                                                                Pune

12th - Fergusson Jr. College, Pune - 74%, *Fergusson Jr. College, Pune*   2019
                                                                                                                                                                Pune

TECHNICAL SKILLS

Languages & Tools
Python (Matplotlib, Seaborn, NumPy, Pandas), SQL, MySQL, PL/SQL, Power BI, Tableau, Qlik Sense, MongoDB, Relational Databases, MS Excel, Python Scripting, Data Quality / Data Integrity

Big Data & Cloud
Hadoop, Hive, Google Cloud Platform (BigQuery, Vertex AI), AWS (S3, SageMaker), Generative AI, LLMs

Soft Skills
Business Strategy, Collaboration Business Stakeholder, Problem Solving, Meeting Coordination, cross-functional collaboration

Machine Learning
 Linear Regression, Logistic Regression, Decision Trees, Scikit-learn, Trees, Random Forest, SVM, XGBoost, KNN, Clustering (K-Means, Hierarchical), PCA, Time SeriesForecasting, Model Evaluation (AUC, F1, ROC).

Data Engineering
ETL Pipelines, Feature Engineering, Data Cleaning, KPI Dashboards, Automation

COURSES & CERTIFICATION

SQL & MongoDB, *Learnbay*   12/2024 – 01/2025

Power BI & Tableau, *Learnbay*   10/2024 – 12/2024

Statistics & ML for Data Science, *Learnbay*   08/2024 – 12/2024

Big Data Analytics with PySpark, *Coursera - Google*   09/2024 – 11/2024

Python for Data Science, *Learnbay*   07/2024 – 08/2024

HONOURS & RECOGNITION

Silver Medalist, State Level Badminton Championship   01/09/2019

PROJECT EXPERIENCE

Breast Cancer Analysis
- Developed a Machine Learning Model Using **K-Means Clustering** to Classify Breast Cancer Types with accuracy of **95%** on a dataset containing **569 samples and 30 features**.
- Conducted exploratory **data analysis** and preprocessing to handle missing values and outliers.
- Tools Used: **Python, Scikit-learn, Matplotlib, Seaborn**

Banking Finance Project
- Used **SQL** for initial data extraction and cleaning.
- Build **classification** models **(Random Forest, XGBoost)** on **10,000+ records** to predict customer churn with **92% accuracy.**
- Applied **KNN imputation** and **hyperparameter tuning (GridSearchCV)** to improve model performance.
- Delivered actionable insights to reduce potential churn by **18%** in a simulated scenario.
- Simulated financial dataset to reflect customer behavior in banking environments.

Reddit User Persona Generator using Python & OpenAl
- Developed an automated **pipeline** using **Reddit API (PRAW)** to extract user posts/comments and generate detailed **UX**-style personas with **OpenAI GPT API**.
- Engineered **scraping scripts** to collect, clean, and structure behavioral data from Reddit, saving outputs in modular .txt files.
- Generated personalized profiles highlighting motivations, frustrations, and behaviors-boosting efficiency in UX research and persona creation.
- Automated **data pipelines** and applied **NLP** with **OpenAI** models
- Tools Used: **Python, Reddit API (PRAW), OpenAI API, Regex, Hugging Face, LangChain, Automation, Text Processing**