Yash Haval

Kolhapur, Maharashtra +91 9373101750 **yashhaval23@gmail.com** **GitHub** **LinkedIn**

SUMMARY:

Dedicated Data Science enthusiast skilled in Python, SQL, and Statistics with hands-on experience in data cleaning, EDA, feature engineering, and visualization. Proficient in MySQL, PostgreSQL, Power BI, Tableau, and Excel for data analysis and reporting. Experienced in building and deploying machine learning models (Linear & Logistic Regression, Decision Trees) and automating workflows to improve efficiency. Currently pursuing a B.Tech in Data Science with practical exposure to real-world projects.

EXPERIENCE:

Python Developer Intern, Anvistar ITS Pvt. Ltd   07/2025 – 08/2025
• Implemented Python libraries and OOP concepts to support development tasks, improving task efficiency by 15%.
• Collaborated with the team to work on real-world projects, gaining hands-on exposure to the company's tech stack and project development •
process.
• Supported debugging and code reviews, helping maintain high-quality standards and smooth project execution.

EDUCATION:

B.Tech - Computer Science (Data Science),   2026
Dr. D.Y. Patil Pratishthan's College of Engineering, Kolhapur
CGPA:8.0

HSC, Murgud Vidyalaya Jr. College   2021-2022
80%

SSC, Murgud Vidyalaya, Murgud   2019-2020
91.60%

TECHNICAL SKILLS:

Programming & Databases
Python, SQL, R, Core Java, C++, HTML, CSS, PostgreSQL, MySQL

Data Analysis & Visualization
Pandas, NumPy, Matplotlib, Seaborn, Power BI, Excel, Streamlit

ML & Statistics
EDA, Scikit-learn, Feature Engineering, Linear & Logistic Regression, Decision Trees, Ensemble Techniques (Voting, Bagging)

Tools & Platforms
Jupyter Notebook, VS Code, Git, GitHub, XAMPP, Canva

PROJECTS:

AI-Powered Data Analytics Portal | Python, Streamlit, Pandas, Scikit-learn
• Designed a Streamlit UI for real-time data analysis and visualization, improving evaluation efficiency by 20%.
• Executed data cleaning, feature selection, and ML model training (Linear Regression, Decision Trees) on 10k+ rows, reducing preprocessing time by 30%.
• Automated exploratory data analysis (EDA) and feature selection, enabling faster insights for end-users.
• Developed a performance dashboard displaying R^2, MAE, and RMSE, reducing model selection time by 25%.

YouTube Trending Video Analysis & Prediction | Python, Streamlit, XGBoost, Scikit-learn
• Analyzed Kaggle's US You Tube Trending Videos dataset to identify engagement patterns and predict views.
• Developed a Streamlit app allowing live predictions, improving prediction speed by 50% compared to manual analysis.
• Achieved $R2 = 0.82$, reducing RMSE by 18% compared to baseline models.

Movie Recommender System | Python, Pandas, Scikit-learn, Streamlit, Flask, TMDb API
• Built a web-based movie recommender using content-based filtering on 5k+ movies, improving recommendation relevance.
• Integrated the TMDb API to dynamically fetch and display posters, boosting user engagement by 30% (based on feedback).
• Implemented robust error handling for missing posters and invalid selections, decreasing Ul errors by 90%.
• Designed a Streamlit interface to present movie recommendations with images and titles, improving click-through rate by 15%.

CERTIFICATIONS:

Python for Data Science	Master Data Analysis with Python	Python for Data Science
NPTEL (2024 | Score:62%)	Udemy (2025)	Great Learning (2023)

SOFT SKILLS:

Problem-Solving | Adaptability | Time Management | Teamwork & Collaboration | Analytical Thinking