AYESHA SALEEM

+92 318 6706803 | ayeshasaleem853@gmail.com | [GitHub](#) | [Kaggle](#) | [LinkedIn](#) | [LeetCode](#)

ACHIEVEMENTS & COMPETITIONS

- *[@Kaggle](#)* Expert: Published datasets, notebooks, and competed in real-world ML challenges.
- *[@Harvard CS50x](#)* Puzzle Day Winner (2025): Solved all 9/9 puzzles; secured 1st place globally with team.
- *[@Meta](#)* Hacker Cup (2024) Qualifier: Competed in Meta's global programming competition, showcasing strong algorithmic and problem-solving skills.
- *[@UC Berkeley](#)* CALICO Informatics Competition (2024): Participated to enhance data handling and computational thinking.
- *[@LabLab.ai](#)* AI Hackathons: Participated in multiple international hackathons focused on generative and applied AI.
- *[@LeetCode](#)* 230+ DSA Problems Solved: Practiced advanced algorithmic skills and competitive coding.

TECHNICAL SKILLS

- **Programming Languages:** C++, Python
- **Data Science and Machine Learning:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Plotly, TensorFlow, PyTorch, Keras, OpenCV, Hugging face
- **Frameworks:** Streamlit, Gradio
- **Version Control:** Git, Github
- **Developer Tools:** VS Code, GoogleCollab, Anaconda, Jupyter Notebook
- **Design Tools:** Canva

EDUCATION

Emerson University Multan   Sep. 2023 – Present
Bachelor of Science in Computer Science   Multan, Punjab, Pakistan
CGPA: 3.86/4.0
Coursework: Data Structures and Algorithms, Artificial Intelligence, Programming Fundamentals, OOP

EXPERIENCE

C++ Programming Intern   May 2024 – June 2024
CodeAlpha   Remote

- Developed hands-on projects using C++ programming language.
- Gained experience in Object-Oriented Programming (OOP) and Data Structures. Contributed to real-world applications by writing clean and efficient code.

Kaggle Professional   Sep. 2024 – Present
Data Science Community Expert   Global

- Published datasets and notebooks, contributing to the Kaggle community.
- Created and shared data analysis notebooks showcasing Exploratory Data Analysis (EDA) and Machine Learning techniques.
- Actively participated in Kaggle competitions, applying Data Science skills to solve real-world problems.

PROJECTS

TasteMatch: Transformer-Based Movie Recommender System | [Live App](#)

Technologies: Python, NLP, Sentence-BERT, Pandas, Cosine Similarity

- Built a content-based recommendation engine using Sentence Transformers (MiniLM-L6-v2) to suggest similar movies from a 5K+ movies dataset.
- Engineered semantic "tags" by preprocessing metadata including plot, cast, genres, keywords.
- Used cosine similarity to compute and rank top-5 similar movies.

OncoPredict AI: Early Breast Cancer Detection | [GitHub Repository](#) **|** [Kaggle](#)

Technologies: Python, Scikit-learn, Logistic Regression, Random Forest, EDA

- Developed a machine learning pipeline for early breast cancer detection using the Wisconsin Diagnostic Dataset.
- Applied Logistic Regression and Random Forest; achieved 98% accuracy and 99% recall with ROC-AUC evaluation, minimizing false negatives.

SupplyShield 2.0 – AI-Powered Logistics Risk Detection System | [lablab.ai](#)

Technologies: Streamlit, Claude, LangChain, Plotly, ChromaDB, APIs

- Developed a real-time dashboard to detect global shipment risks using weather, news, and logistics data.
- Used Claude LLM to generate risk summaries, suggest actions, and automate stakeholder communication.
- Integrated LangChain logic, ChromaDB memory, and APIs for alerts, visualizations, and smart decisions.

EDA: Unveiling Sales Patterns | [Kaggle](#)

- Performed Exploratory Data Analysis (EDA) on sales data to uncover hidden patterns and trends.
- Applied univariate, bivariate, and multivariate analysis techniques to identify key factors influencing sales performance.
- Visualized insights using Matplotlib and Seaborn, enhancing decision-making.

LifeLens: Life Expectancy Prediction with Machine Learning | [Kaggle](#)

- Developed a machine learning model to predict life expectancy based on health, economic, and social factors using XGBoost Regressor, performing EDA to identify correlations and data patterns.
- Preprocessing data through encoding, scaling, and missing value imputation.
- Implementing GridSearchCV for hyperparameter tuning to improve model performance.