

Muhammad Haris Afridi

AI Engineer | Full-Stack Developer | ML Researcher | Automation
mharisyar.ai@gmail.com | github.com/harisyar-ai | Peshawar, Khyber Pakhtunkhwa, Pakistan

SUMMARY

Self-taught AI Engineer and Full-Stack Developer with hands-on experience building production-grade intelligent systems from the ground up. Conducted original ML research achieving 97%+ predictive accuracy on real-world Pakistani market data. Expanding into Generative AI — building N8N automation pipelines and AI-powered content generation workflows for social impact. Combines strong Python engineering with practical ML expertise and a drive to build tools that create genuine impact.

EDUCATION

Bachelor of Science in Artificial Intelligence

2023 – 2027

Agriculture University of Peshawar | GPA: 3.5 / 4.0

- Coursework spanning Machine Learning, Deep Learning, Computer Vision, NLP, and AI Systems Design.
- Active Research Assistant at the Digital Image Processing (DIP) Lab, Islamia College Peshawar.

EXPERIENCE

Sole Developer — Government MIS Platform

2026

Local Government, Elections & Rural Development Department, KPK (JICA Collaboration) | Khyber Pakhtunkhwa, Pakistan

- Sole designer and developer of IDBPM KP — a production-grade, multi-role MIS platform serving districts across Khyber Pakhtunkhwa in partnership with JICA.
- Built the entire stack independently: Flask backend, custom HTML/CSS frontend, PostgreSQL database, REST APIs, and AI-driven prioritization engine.
- System handles scheme scoring, multi-role authentication, Excel/CSV ingestion, tehsil need lists, MTDP documents, VCNC management, and district-level budget planning workflows.
- Deployed live on Railway and actively maintained based on real officer feedback.

Research Assistant — Computer Vision

2025 – Present

Digital Image Processing (DIP) Lab, Islamia College Peshawar | Peshawar, Pakistan

- Applied computer vision techniques to active academic research problems at a dedicated image processing laboratory.
- Developing practical expertise in CNNs, image segmentation, and object detection within a research context.
- Collaborating on work aimed at publication or scholarly showcase.

Machine Learning Intern

2026 – Present

CodeAlpha | Remote / Virtual

- Completing a structured virtual ML internship — building end-to-end machine learning projects including classification, regression, and predictive modelling pipelines on real-world datasets.
- Applying Scikit-Learn, Pandas, and Python to assigned tasks; sharing completed work on GitHub and LinkedIn as part of the internship programme requirements.
- Gaining practical exposure to the full ML lifecycle — data cleaning, feature engineering, model training, evaluation, and deployment-ready output — outside of academic coursework.

PROJECTS

IDBPM KP — Integrated District Budget Planning & Management Information System

2026

Stack: Python · Flask · PostgreSQL · HTML/CSS · Pandas · openpyxl · Jinja2 · Unicorn · Railway

- Full-stack governmental MIS platform built solo for the Local Government Department of KPK in collaboration with JICA. Supports multi-role authentication (Admin, Officers, Viewers), AI-driven scheme prioritization, Excel/CSV upload pipelines, and tehsil need-list management.
- Serves all target districts across Khyber Pakhtunkhwa; deployed live on Railway.
- Custom sector scoring: Drinking Water uses coverage-gap weighting; Education uses population-per-school vulnerability ratio; Health uses distance-to-facility proximity.

Pakistani Used Car Price Prediction — ML Research

2026

Stack: Python · LightGBM · Scikit-Learn · SHAP · Pandas · Streamlit

- End-to-end ML research pipeline on 75,000+ real PakWheels listings, cleaned through a rigorous 4-stage preprocessing pipeline to 60,396 usable rows.
- Best model: LightGBM — Test $R^2 = 0.9722$ (log scale) / 0.9490 (actual prices), RMSE = 13.33 Lacs, MAE = 3.62 Lacs.
- 13 engineered features including car_age, brand_origin, city_tier, trim_tier, and is_electric flag.
- Research paper drafted for IEEE submission; currently in literature review and writing phase.

Nova Books Recommender
2025

Stack: [Python](#) · [NLP](#) · [TF-IDF](#) · [Cosine Similarity](#) · [Streamlit](#)

- Content-based book recommendation engine across 10,000+ titles using TF-IDF and cosine similarity.
- Analyzes title, author, description, and genres to surface contextually meaningful recommendations.
- Includes a custom cover-image scraper and mobile-responsive Streamlit UI.

Medical Insurance Cost Prediction
2026

Stack: [Python](#) · [Random Forest](#) · [Scikit-Learn](#) · [Streamlit](#)

- Supervised regression system estimating annual health insurance charges from personal and lifestyle features.
- Applied polynomial features, interaction terms, and log transformations; benchmarked 15+ models — Random Forest achieved $R^2 = 92.46\%$.

Job Salary Prediction
2026

Stack: [Python](#) · [Gradient Boosting](#) · [Scikit-Learn](#) · [Streamlit](#)

- ML regression app estimating annual salaries based on education, experience, job title, location, and gender.
- Benchmarked 11 models with full preprocessing pipelines — Gradient Boosting achieved $R^2 = 85.49\%$.

Breast Cancer Prediction
2025

Stack: [Python](#) · [Scikit-Learn](#) · [Classification Models](#)

- ML classification project for early breast cancer diagnosis using clinical feature inputs.
- Full EDA, preprocessing pipeline, and model evaluation. Deployment in progress.

AI Content Automation Pipeline
2026 — In Progress

Stack: [N8N](#) · [OpenAI API](#) · [Prompt Engineering](#) · [Workflow Automation](#)

- N8N workflow pipeline for automated AI content generation — content brief triggers GPT-powered script generation, formats structured output, and delivers ready-to-use content.
- Built for social impact topics including polio awareness, women empowerment, and community health campaigns.

PUBLICATIONS

“LightGBM for Pakistani Used Car Price Prediction: Domain-Specific Feature Engineering and SHAP Analysis”
2026

Under Review — *Expert Systems with Applications (Elsevier, Q1)*

CERTIFICATIONS

Python Developer Sololearn	Feb 2025
AI For Everyone DeepLearning.AI	Jun 2025
Machine Learning Specialization Stanford · Coursera	Jul 2025
Deep Learning Specialization DeepLearning.AI	Nov 2025
Computer Vision Specialization Univ. of Colorado	Nov 2025
ChatGPT Prompt Engineering for Developers DeepLearning.AI	In Progress

TECHNICAL SKILLS

Machine Learning & Data Science

Entered the field of machine learning with a focus on genuinely understanding how models learn from data — not just running libraries. Worked extensively with real-world, messy, unstructured datasets: scraping, cleaning, engineering features, and training models end-to-

end. Built complete pipelines from raw data to explainable, deployable predictions — applying ensemble methods, SHAP explainability, and rigorous cross-validation. Tools: Scikit-Learn, LightGBM, Random Forest, Gradient Boosting, SHAP, Pandas, NumPy, Matplotlib.

Generative AI & Prompt Engineering

Building GenAI pipelines using OpenAI and Anthropic APIs. Experienced in structured prompt engineering for complex real-world systems — including the government MIS AI prioritization engine built entirely through prompt-driven logic. Working with AI video generation tools for social impact content creation. Tools: OpenAI API, Anthropic API, Prompt Engineering, AI Video Generation (InVideo AI, Pictory).

Workflow Automation — N8N

Designing and building N8N automation pipelines that connect AI models to real-world workflows — trigger-based content generation, structured output delivery, and multi-step task automation. Currently building a pipeline for AI-powered social impact content generation. Tools: N8N, Webhook Triggers, GPT API Integration.

Deep Learning, NLP & Computer Vision

Expanded into neural networks — learning how deep architectures perceive and interpret data at multiple levels of abstraction. Built and trained CNNs, explored NLP pipelines from TF-IDF through to Transformer-based contextual embeddings, and applied digital image processing in a live research setting. Currently deepening knowledge of advanced architectures including GANs and modern segmentation models. Tools: TensorFlow, Keras, OpenCV, Hugging Face Transformers, TF-IDF, Cosine Similarity.

Backend Development

Built complete, production-ready web backends from scratch using Python and Flask — handling routing, authentication, file processing, and REST API design. Experienced with PostgreSQL and structured data pipelines, comfortable deploying to cloud environments and managing live applications. Tools: Python, Flask, PostgreSQL, REST APIs, Werkzeug, Gunicorn, openpyxl, Pandas Excel I/O.

Frontend Development

Developed clean, functional user interfaces that prioritize usability and clarity — built on core web fundamentals without relying on heavy frameworks. Skilled in crafting responsive layouts and data-rich dashboards that non-technical users can navigate with ease. Tools: HTML5, CSS3, Flexbox, Grid, JavaScript, Jinja2 Templating, Animations, Responsive Design.

Tools & Deployment

Streamlit, GitHub, VS Code, Jupyter Notebooks, Railway (cloud deployment), openpyxl.

CURRENT FOCUS & FUTURE GOALS

Currently Working On

- Building an N8N automation pipeline for AI-powered content generation — prompt to script to structured output.
- Exploring GenAI tools for social impact video content — polio awareness, women empowerment, community campaigns.
- Pakistani Used Car Price Prediction research paper — literature review and writing phase for IEEE publication.
- Research Assistant at DIP Lab — applying computer vision techniques to real academic research problems.
- Deepening knowledge of Transformers, advanced CNNs, and GANs within AI coursework at Agriculture University.

Future Goals

- Complete and publish the Pakistani Used Car Price Prediction research paper.
- Begin research in medical computer vision — specifically image segmentation for early detection of lung cancer, kidney failure, and skin diseases.
- Develop a Transformer-based NLP project to push beyond TF-IDF into contextual understanding.
- Contribute to open-source AI tools and grow the GitHub portfolio with real-impact projects.
- Complete BSc in AI and pursue advanced research or industry roles in ML Engineering or AI Systems.