

# Mostafa Farag Abdelwahed

## Machine Learning Engineer | Data Scientist

✉ mhfarag938@gmail.com 📍 Cairo/Egypt 🌐 github.com/mostafa7hmmad ☎ 01154079827

🌐 linkedin.com/in/mostafa-hammad-154b42317

## Profile

**Third-year Computer Science student** with a strong foundation in Python, PHP, C++, and core computer science concepts like data structures, OOP, and linear algebra. Skilled in data science and AI, with hands-on experience using libraries such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, and NLTK. Built interactive data apps with Streamlit and conducted end-to-end projects involving data preprocessing, visualization, machine learning, and deep learning. Comfortable working with MySQL and applying data engineering techniques. Passionate about clean code, explainable AI, and simplifying technical concepts. Actively learning advanced topics in natural language processing and deep learning. Seeking internship opportunities in AI, data science, or machine learning.

## Education

**Bachelor student Artificial Intelligence department, faculty of information technology, Misr university for science and technology**

2022 – present

## Technical Skills

**Programming language:** Python | C++ | PHP

**Frameworks Libraries:** Pandas | Numpy | Matplotlib | Seaborn | Plotly | Bs4 | Requests | Scikit-learn | TensorFlow | Streamlit | NLTK | Laravel | Bootstrap

**Concepts:** Software Engineering | OOP | Data Structures & Algorithms | Linear Algebra | Probability & Statistics | Calculus Basics | EDA | Data Wrangling | Data Analysis | Data Visualization | Feature engineering | Time Series Analysis | Machine Learning | Deep Learning | Natural Language Processing | Model Evaluation | Model Deployment | Web Scraping | FastAPI

**Tools:** Git | Github

## Professional Experience

**Machine Learning Intern Coding Samurai | Remote | Jun 2025 – Jul 2025** 🌐

Worked on real-world machine learning projects involving data preprocessing, model building, and evaluation using Scikit-learn and TensorFlow. Gained hands-on experience in EDA, model tuning, and deploying ML models with Streamlit

## Projects

**U-Net-Powered-Brain-MRI-Segmentation-for-Medical-Diagnostics** 🌐

- Built a U-Net model with EfficientNetBo for brain tumor segmentation on 3,064 MRI images. Applied resizing, normalization, and augmentation to improve model robustness.
- Trained with BCE and Dice Loss, achieving 0.80 Dice coefficient using early stopping.
- Evaluated performance using classification metrics and mask visualizations with TensorFlow & Keras.

**Predictive-Maintenance-for-Medical-Devices** 🌐

- Developed a predictive maintenance system for medical ventilators using real-world equipment data.
- Renamed and cleaned columns (e.g., Asset\_condition, Maintenance\_complexity) and exported processed data to Excel.
- Conducted feature engineering by removing irrelevant features (e.g., Make, Model) and encoding categorical variables.
- Applied K-means clustering for unsupervised analysis and validated results using 3D t-SNE visualizations.
- Labeled time-to-failure into three classes for classification: no failure, short-term, and long-term failure.
- Trained and compared multiple models including SRN, LSTM, ANN, SVM, and Random Forest.
- Achieved up to **99.55% test accuracy** with SRN and LSTM models.

**Student-Mental-Health-Prediction-Using-Ensemble-Learning-Technique-GA** 🌐

- Built predictive models for student depression using 502 samples and 10 psychological and lifestyle features.
- Performed EDA and visualized correlations with Seaborn and Matplotlib; encoded categorical data and normalized features.
- Selected key features using correlation analysis and Random Forest importance (e.g., Academic Pressure, Suicidal Thoughts).
- Trained and tuned multiple models (SVM, RF, KNN, etc.) using GASearchCV; SVM reached 95.05% accuracy.
- Developed an ANN with dropout and early stopping in TensorFlow/Keras, achieving 97.03% accuracy and saved models for deployment.

### Breast Cancer Recognition [↗](#)

- Built a Random Forest model to classify breast cancer (benign vs. malignant) using 609 samples and 30 features.
- Conducted EDA with Pandas, Seaborn, and Plotly to explore missing values, duplicates, and feature distributions.
- Preprocessed data by handling missing values, removing outliers (IQR & Z-score), and applying MinMax normalization.
- Selected top 5 correlated features for model input using correlation analysis.
- Tuned model with GridSearchCV, achieving 94.5% test accuracy and evaluated results using classification metrics and visualizations.

### Stock Price Prediction for Apple Inc [↗](#)

- Developed a predictive model for Apple Inc.'s stock closing prices using 5 years of daily trading data (1,267 days).
- Conducted EDA with Plotly and Seaborn to analyze trends, seasonality, and key market events.
- Preprocessed data by removing irrelevant columns, handling missing values, and filtering outliers using IQR.
- Engineered features like average price, price range, and open-close difference to capture stock behavior.
- Trained a Random Forest Regressor, achieving  $R^2 = 0.9676$ , MSE = 0.000289, and MAE = 0.007067.

### Automated-Gender-Classification-Using-Facial-Recognition [↗](#)

- Developed a CNN model for automated gender classification using 47,009 training and 11,649 validation facial images.
- Applied image preprocessing and augmentation (rotation, shear, zoom, flip) to improve generalization.
- Built a CNN with Conv2D, MaxPooling, Dropout, and Dense layers, achieving **96.29% validation accuracy**.
- Used early stopping to prevent overfitting and visualized training metrics with Matplotlib.
- Saved the trained model and class mappings for deployment in real-world applications.

### yolov8-fired-detection [↗](#)

- Developed an object detection model using YOLOv8 to detect fire and smoke in images.
- Fine-tuned a pretrained YOLOv8s (11.1M parameters) on a custom dataset of 14,101 training and 3,094 validation images.
- Achieved a **mAP<sub>50</sub> of 0.66** after 30 epochs with early stopping and AMP training on a Tesla P100 GPU.
- Enhanced model robustness through data augmentation (rotation, translation, scaling, shearing).
- Visualized box, classification, and distribution focal loss to monitor training performance.
- Used Python, PyTorch, Ultralytics, and Matplotlib for modeling, optimization, and evaluation

### yolov8-drowsiness-detection-system [↗](#)

- Developed a YOLOv8-based model to detect driver drowsiness using 12,513 training and 1,191 validation images.
- Fine-tuned a YOLOv8n (3.01M parameters), achieving **mAP<sub>50</sub> of 0.983** and **mAP<sub>50-95</sub> of 0.835** after 30 epochs.
- Applied data augmentation (blur, grayscale, CLAHE) and trained using AMP on a Tesla P100 GPU.
- Used early stopping (patience=7) to prevent overfitting during model training.
- Visualized box, classification, and focal loss metrics with Pandas and Matplotlib for model evaluation.
- Implemented with Python, PyTorch, Ultralytics, and Albumentations for real-time detection in safety-critical system

## Certificates

---

Artificial intelligence and Data Science | Optical Soft [↗](#)

Foundation of machine learning and Ai |Sprints

Machine Learning for Quant Finance and Algorithmic Trading| Udemy [↗](#)

AI Development Bootcamp and Certification |Udemy [↗](#)