Pneumonia Detection Capstone Project — Workflow

Introduction

In this project, an end-to-end machine learning pipeline was developed to detect pneumonia from chest X-ray images. The goal was to achieve accurate, explainable, and practical results to support healthcare professionals in early detection.

Workflow

Dataset Preparation

- Started with chest X-ray dataset in DICOM (.dcm) format + labels CSV.
- Converted .dcm to .png for easier handling.
- Created a balanced subset (equal Normal & Pneumonia samples) to save space & time.

Exploratory Data Analysis (EDA)

- Checked class distribution (Normal vs Pneumonia).
- Visualized random X-rays.
- Confirmed dataset imbalance and solved using balanced subset.

Preprocessing

- Normalized pixel values (0–255 → 0–1).
- Resized images to 224×224.
- Converted grayscale images to 3-channel (needed for pretrained CNNs).

Model Training

- Used ResNet-18 pretrained on ImageNet.
- Replaced final layer with binary classifier (Normal / Pneumonia).
- Optimizer: Adam, Learning Rate: 1e-4.
- Trained on small balanced dataset (quick demo).

Evaluation

- Generated Classification Report (Precision, Recall, F1-score, Accuracy).
- Verified performance on validation subset.

Inference & Visualization

- Loaded saved model weights.
- Predicted on single X-ray images.
- Added Green box + label for Normal, Red box + label for Pneumonia for easy visualization.

Deployment Demo

- Created a Streamlit app → upload X-ray → get prediction & visualization.

This workflow demonstrates the full ML pipeline: data → preprocessing → training → evaluation → inference → demo.

Conclusion

This project highlighted the importance of balanced data, careful preprocessing, and interpretable results. It showcases the value of AI in healthcare and its ability to support critical decision-making.