# Karthik Premaram

🗷 premaramkarthik@gmail.com 📞 93912222493 👂 hyderbad 🕥 github 🛅 linkedin 🔗 portfolio

## **EDUCATION**

**Data Science** 08/2024 – present | Hyderabad, India

Innomatics Research Labs

**Bachelor's in Artificial Intelligence** 09/2021 – 04/2024 | Hyderabad, India

Vidya Jyothi institute of technology

**SKILLS** 

Programming LanguagesData Visualization ToolsPython, R Programming , SQLMatplotlib, Seaborn, PowerBI

T Thurston

LIbraies Exploratory Data Analysis (Web Scraping)
Numpy, Pandas, ScikitLearn, Stream Lit Beautiful Soup, Statistical Methodologies

Advanced Statistics Machine Learning
Statistics Inferential and Descriptive analysis, Supervised and Unsupervised Machine

Hypothesistesting Supervised and Unsupervised Machine learningAlgorithms

Advanced ML Concepts

Feature Engineering, Regularization Techniques,
HyperParameter Tuning

Deep Learning

Neural Networks , Natural Language Processing,
Computer Vision, Generative AI (GenAI)

**PROJECTS** 

Hybrid Recommendation System for E-commerce (Olist) ∂

**Skills:** Machine Learning, Hybrid Recommendations, Collaborative & Content-Based Filtering, Sparse Matrices, Feature Engineering, Unsupervised Learning, Hyperparameter Tuning, Model Evaluation, LLMs.

Aim:

• Develop a **hybrid recommendation system** to improve product suggestions and address the **cold-start problem.** 

#### Approach:

## 1.Data Engineering:

- Processed and integrated multiple datasets (100,000+ orders, 3,000+ sellers, 10,000+ products) to create comprehensive user-item interaction matrices.
- Created custom **visualization** to **analyze** purchasing **patterns**, product categories, and seasonal trends.
- Implemented **multilingual sentiment analysis** on Portuguese customer reviews using **Hugging Face Transformers** (tabularisai/multilingual-sentiment-analysis model).
- Engineered **sparse feature matrices** incorporating product metadata (categories, reviews, pricing) to enhance recommendation quality.

# 2. Model Implementation:

- Built a **hybrid recommendation system using LightFM** that combined collaborative filtering with content-based approaches.
- Implemented **systematic hyperparameter tuning** through randomized search to optimize model performance, testing combinations of learning rates, regularization parameters, and loss functions.

## 3. Model Evaluation:

- Measured **AUC** to assess recommendation quality, including cold-start scenarios.
- Conducted **user-case testing** for relevance and consistency.

#### Results:

• The model, built using the **LightFM algorithm**, achieved an **AUC of 0.97 (vs. 0.82 baseline)** with optimized hyperparameters.

Agricultural Pest Classification Using convolutional neural network  $\mathscr{D}$ 

#### Skills:

Deep Learning, CNNs, Image Classification, Transfer Learning, Model Evaluation, Streamlit, OpenRouter API, Hyperparameter Tuning, YOLOv8, LLMs, Deployment.

#### Aim:

Develop an pest **identification** system for detection and control recommendations using **LLM models**.

#### Approach:

# 1. Data Preprocessing & Preparation:

- Organized pest images into five categories (Caterpilla, Locusts, Slug, Gastropoda, Curculionoidea) with proper labeling and an 80-20 train-test split.
- Applied resizing, normalization, and **augmentation** (flips, rotations, zoom, contrast) for better generalization.

## 2. Model Development & Evaluation:

- Trained and evaluated ResNet50 (80%), ResNet101V4 (87.7%), MobileNetV2 (97.5%), YOLOv8 (77%), and EfficientNet (89%).
- Used early stopping, hyperparameter tuning, and monitored YOLOv8 performance (optimal at epoch 29).

## 3. Model Selection & Optimization:

- Chose **YOLOv8** for as the primary model due to its **accuracy (77%)** and excellent generalization performance.
- Optimized model architecture for efficient classification of five pest categories.

## 4. Deployment with Streamlit

- Built an **interactive web application** using **Streamlit** for pest classification.
- Integrated **OpenRouter API** with **Google's Gemini 2.0 Flash Thinking model** to provide detailed pest insights and control strategies.

#### **Results:**

• The pest classification system was deployed with **MobileNetV2** (97.5%) and **YOLOv8** (77%) as the best models. A **Streamlit app** enabled Classification, while **Gemini 2.0 via OpenRouter API** provided actionable insights including organic, chemical, and integrated pest management approaches.

#### Hate Speech Detection *⊘*

**Skills:** NLP, Text Classification, Machine Learning, Data Preprocessing, Feature Engineering, Hyperparameter Tuning, Model Evaluation, Data Visualization

#### Aim:

Developed a **hate speech detection system** for Twitter data using **machine learning** techniques to accurately classify harmful content.

#### Approach:

#### 1. Data Engineering:

- Processed and **cleaned Twitter dataset** containing **30K+ tweets** using **regular expressions** and NLTK.
- Applied text normalization techniques including tokenization, lemmatization, and stopword removal.
- Implemented **visualization** techniques to analyze hate speech distribution and word frequencies.
- Generated **TF-IDF features with n-grams** to capture contextual information.

#### 2. Model Implementation:

- Built a classification pipeline using Logistic Regression as the primary model
- Implemented systematic **hyperparameter tuning** through **GridSearchCV** to optimize model parameters (C values, solvers)
- Experimented with different n-gram ranges to capture phrase-level patterns in hate speech

#### 3. Model Evaluation:

- Utilized confusion matrices, classification reports, and accuracy scores to assess model performance
- Conducted thorough error analysis to understand **misclassification** patterns
- Visualized model predictions to identify areas for improvement

#### Results:

• The optimized **Logistic Regression** model achieved high **accuracy (94.89%)** in classifying hate speech, with **precision and recall metrics** outperforming baseline approaches. The implementation of **TF-IDF** with **n-grams (1-3)** significantly improved the model's ability to detect subtle forms of hate speech in tweets.

#### **CERTIFICATES**

- Python for Data Science, AI & Development ℰ
- Google Data Analytics Professional Certificate ∅
- Machine Learning by Stanford University & DeepLearning.AI on Coursera ℰ