Machine Learning Engineer/ Data Scientist

TA TRAN TUYEN

(+84) 344665709|tatrantuyen@gmail.com
https://www.linkedin.com/in/ta-tuyen-148479258/
https://github.com/young-the-tiny

Skills

Languages: Python, Java, C++.
Frameworks: PyTorch, Scikit-Learn.
Tools/Library: Matplotlib, Seaborn, Pandas, NumPy, Jupyter Notebook, Git, Linux, SQL, PySpark.

Technical Skills: Machine Learning, Deep Learning, Large Language Models, Statistics, Data Analysis, Natural Language Processing, Computer Vision.

Projects

SENTIMENT ANALYSIS ON GAME REVIEWS – Personal Project December 2023

- Utilized **Python** to analyze over 6 million rows of reviews.
- Implemented **Regex** techniques to comb through the job descriptions and identify the key requirements.
- Aggregated and visualized the data by using **pandas, matplotlib** and **wordcloud** to compile a professional report.
- Applied oversampling techniques such as **SMOTE (Synthetic Minority Over-sampling Technique)** to address class imbalance and enhance model performance.
- Conducted **feature extraction** using techniques like **TF-IDF** and **word embeddings** to improve model accuracy.
- Implemented many advanced **Machine Learning** models and Deep Learning models such as **RNNs** and **LSTMs** with high accuracy.
- Evaluated model performance using metrics such **as accuracy, precision, recall,** and **F1-score**, and performed hyperparameter tuning to optimize results.

TRANSLATING MANGA WITH OCR –Personal Project April 2024

- Employed **PyQT5** library to help creating an interface for the project.
- Developed and implemented **an Optical Character Recognition (OCR**) system specifically for manga text extraction.
- Using **OpenCV** library to help with image preparation.
- Integrated **Tesseract OCR engine** with custom-trained models to handle unique manga fonts and styles.
- Translating the extracted text to English using **DeepL.**
- **Documented** the OCR system architecture, workflow, and usage guidelines for future maintenance and enhancements.

FOOTBALL ANALYSIS – Personal Project July 2024

- Developed a comprehensive football analysis system using **machine learning, computer vision**, and **deep learning techniques.**
- Utilized a Kaggle dataset to train and validate the system, ensuring robust performance on diverse data.
- Employed **YOLOv8,** a state-of-the-art object detector, to detect players, referees, and footballs in images and videos.
- Fine-tuned and trained custom **YOLO models** on the Kaggle dataset to enhance detection accuracy.
- Implemented object-tracking algorithms to track detected objects across video frames.
- Applied **KMeans clustering** for pixel segmentation to accurately assign players to teams based on t-shirt colors.
- Used optical flow techniques to measure camera movement between frames, enabling precise player movement analysis.
- Employed **OpenCV's perspective transformation** to represent scene depth and perspective, allowing measurement of player movement in meters.
- Calculated player speed and distance covered using the transformed perspective data.

Education

BACHELOR OF SCIENCE IN COMPUTER SCIENCE – Sai Gon University – 273 An Duong Vuong, District 5 September 2022
I am currently a **third-year** student pursuing a Bachelor of Science in Computer Science.