THÁI BẢO AN

MY CONTACT

0902967905

thaibaoan27122003@gmail.com

786 Truong Chinh, Tan Binh, Ho Chi Minh

EDUCATION

Bachelor of Data Science

University of Economics Ho Chi Minh City

09/2021 - Present

GPA: 8.38/10

SKILLS

- Microsoft Office (Word, Excel, PowerPoint)
- Database (MySQL, MSSQL Server, PostgreSQL)
- Programming Language (Python, R, Golang)
- Visualization:
  - Python (Pandas, Matplotlib, Seaborn)
  - Tableu, Power BI
- Machine Learning and Deep Learning:
  - Python (Sklearn, Tensorflow, Pytorch)
  - Natural Language Processing (Hugging Face, NLTK, BERT)
  - Computer Vision (OpenCV, YOLO, CNNs)
  - Generative Models (GANs, Transformers, LLMs, RAG)
- Data analysis & statistical methods
- Strong Communication & Presentation
- Strong Research Abilities & Logical Thinking

CERTIFICATIONS

- IBM Data Science
- Hackerrank SQL (Advanced)
- Toeic 625

SUMMARY

I'm a fourth-year Data Science student with a passion for data analytics, machine learning, and artificial intelligence. I aim to enhance my skills and grow in these fields by applying my knowledge to create practical and impactful solutions.

EXPERIENCE

VnResource - Ho Chi Minh City *06/2024 - 10/2024*

AI Research Intern

- Developed and deployed AI models using Python to enhance HR management software solutions, with a focus on building APIs for automated CV data extraction.
- Conducted Researched and programmed algorithms, creating APIs for integration into the company's software system, including functionalities such as:
  - Automated extraction of information from applicant resumes.
  - Employee facial recognition within the HR management system.
  - Optimized the internal chatbot system using the ChatGPT API to support customer service and internal consulting by integrating a large language model with the LangChain framework.
- Gained practical experience in data processing, AI model development, and deployment while developing understanding of agile methodologies like Scrum to comprehend software development processes.

PROJECTS

Building RAG Application with Langchain

- Description: Developed a Retrieval-Augmented Generation (RAG) application leveraging LangChain and Hugging Face. The system retrieves scientific documents, processes them into a vector database, and generates context-aware answers to user queries using large language models (LLMs).
- Technologies: Python (Pytorch, Langchain, Hugging Face, FastAPI), Docker
- Results: Successfully implemented a RAG system that retrieves scientific documents, builds a vector database for efficient query retrieval, and delivers accurate, context-aware answers.
- Github: github.com/AnT2003/Langchain-RAG-Project

Sentiment Analysis System

- Description: Collected over 20,000 customer product review records from Tiki and performed analysis, developing sentiment prediction models using machine learning and deep learning techniques in natural language processing, such as AdaBoost, SGDClassifier, BiLSTM
- Technologies: Python (Request, Matplotlib, Sklearn, Tensorflow, Flask)
- Results: The model achieved 89% accuracy on test data, providing valuable insights for improving customer satisfaction and refining product offerings.
- Github: github.com/AnT2003/Sentiment-Analysis-Project

Identification System using Face Recognition

- Description: The Face Recognition System developed a CNN-based model to identify faces, generating unique facial embeddings for accurate recognition of new images. Achieving over 90% accuracy, it delivers real-time results with identity details and confidence scores.
- Technologies: Python (TensorFlow, Keras, face_recognition, OpenCV, Flask)
- Results: Achieved over 95% accuracy, providing real-time results including name, position, and confidence score for each predicted face.
- Github: github.com/AnT2003/Identification-System-using-Face-Recognition

Traffic Monitoring and Vehicle Analysis System

- Description: Developed a real-time vehicle detection and license plate recognition model that tracks vehicles, identifies types, reads license plates, and estimates speed with over 80% accuracy
- Technologies: Python (Numpy, Tensorflow, YOLO, EasyOCR, OpenCV, Flask)
- Results: identifying key trends, gaining deeper insights into customer segmentation to support targeted marketing and improved engagement
- Github: github.com/AnT2003/Traffic-Monitoring-System