Bui Thi Thanh Phuong

0948324649 | thanhphuong28022000@gmail.com | Github | Linkedin

EDUCATION

VNUHCM - University Of Science — Thu Duc City
Bachelor of Mathematics and Computer Science major — *October 2020 –Present*

OBJECTIVE

Throughout my learning journey, I have come to realize the importance of self-study and knowledge accumulation. Learning is a continuous process, and I aspire to continue learning new things in my field and master the application of acquired knowledge. Alongside accumulating experience, I set additional goals to constantly update myself with new methods, such as reading articles and researching advancements in those methodologies.

TECHNICAL SKILLS

Programming Languages: Python, C/C++, SQL (Postgres), R
Tool and Framewor: Jupyter Notebook, Git, Apache Spark, Pandas, Mathplotlib, SKlearn, Pytorch
Domain Knowledge:Language Modeling, Machine Learning, Probability and Statistics, Math concepts
Others: Excel, Word, Spreadsheets

PROJECTS

On Retrieval-Augmented Text Generation | *Python* — Oct 2023 – Dec 2023
- Utilize Retrieval-Augmented Generation - RAG to query and generate text to assist language models in producing more accurate responses

Prove the Johnson-lindenstrauss lemma and its application in dimension reduction. | *Python* `

May 2023 – June 2023

- To prove the lemma, we utilize Gaussian Random Projection to demonstrate the mapping with a probability.
- Directly apply Johnson's Lemma algorithm to perform dimensionality reduction on a dataset of 11,938 $64{\times}64$ face images, then compare the performance of the lemma with the KNeighborsClassifier model.

Classify customer sentiment analysis | *Python* — Sep 2023– Oct 2023
- Using some tools in the Natural Language Toolkit library such as: Tokenization, POS tagging... to analyze text.
- Use RoBERTa Pretrained Model to evaluate customer reviews.
- Compare the above results with star ratings using the Matplotlib library.

Hotel booking demand prediction | *Python* — Nov 2022– Dec 2022
- Constructing a supervised Machine Learning model to predict hotel room booking demand using scikit-learn and evaluating the model's performance metrics such as Decision Trees, Logistic Regression, and Random Forest. Determine the best model for solving the problem.

Program in the Windows environment | *C++* — June 2022– June 2022
- Create objects of the shape class and check if a point lies within that shape
- Find the intersection between two arbitrary objects and print a plus sign inside the intersection region
- Implement an algorithm using arrow keys to move one of the shapes, plus and minus keys to zoom in or out one of the images. In the Windows environment

ACTIVITIES

Summer volunteer program "Green summer" — July 2022 – August 2022
VNUHCM - University Of Science — *Thu Duc City*
- Create opportunities for students to participate in exchanging knowledge of Mathematics, Physics, Chemistry and Biology, while applying problem-solving skills to real-life situations

VIASM Mathematics Summer School — July 2022 – August 2022
VISAM — *DakLak Province*
- Research advanced topics in Mathematics, including optimization models and linear algebra, focusing on their applications in data science