Hoang Chu

EDUCATION

Pitzer College Claremont, CA
Bachelor's Degree, **Joint Computer Science and Mathematics (Honors)** Expected May 2024

- **GPA**: 3.68 / 4.0
- **Coursework**: Data Structures and Algorithms in C++ (**Teaching Assistant**), Intermediate Probability and Statistics, Combinatorics, Machine Learning (MIT OCW), Linear Algebra, Abstract Algebra, Mathematics of Topic Modeling
- **Awards**: 1st Prize - Vietnam Mathematics Olympiad; Winner - Citadel Invitational Datathon

WORK EXPERIENCE

Meta Menlo Park, CA
Software Engineer Intern May 2022 - August 2022

- Built Deep Learning models based on Transformer architecture, which helped improve the accuracy of the team's various Ads and Commerce classification tasks, and implemented 2 papers on multimodal models in Python.
- Wrote Dataswarm pipelines preparing the pre-training dataset and downstream tasks using internal operators.

CoHost.ai Hanoi, Vietnam
Software Engineer Intern June 2021 - August 2021

- Developed a chatbot that helps ~200+ daily users choosing suitable houses for rent.
- Constructed a statistical language model that labels Intent-Entity from users' input texts, achieving a 95% accuracy.

PROJECTS

Google Summer of Code - Python Software Foundation, Independent | Java, Python

- Reduced 4x the runtime of VOC compiler, an open-source Python Abstract Syntax Trees to Java bytecode translator.
- Refactored Python's Object abstraction by writing a method that identifies duplicated bytecode when translating e.g primitive types calling Object() multiple times, caches their values, and replaces subsequent reloads with local loads.
- Modified VOC internals in translation to use Java primitives instead of Python built-in classes to avoid duplicating instance calls in Java's CONSTANT pool, yet still able to call utility methods for primitive types that Python provides.

Risk Assessment Large Scale Data Analysis, Team of 4 | Python

- Detected from a 26 million dataset that given similar risk profiles, LendingClub charges higher interest rates than lending counterparts, and when previously similar risk profiles had multiple defaults, it fails to adjust interest rates.
- Self-designed a distance metric that helps the team's k-NN algorithm only need a small subset of features to classify groups of borrowers by interest rate and achieve a 94% accuracy in predicting interest rate of a new borrower.

ClaremontCourses - Optimized Course Searching Website, Independent | Python, JavaScript, MySQL

- Designed a non-clustered database indexing with self-designed tree data structures and an accurate course-equivalent matching algorithm that helped reduce the search engine's response time from 25 to consistently 0.8 seconds.
- Constructed algorithms scraping 4500+ courses an hour, matching equivalent courses with 98% accuracy across 5 different colleges, and self-designed data structures and regular expressions to build full-depth prerequisite trees.
- The website got featured in the college's student newspaper and raised a $3000 donation for non-profits (source).

Where-to-go Android App: Tourism Recommendation, Independent | Java

- Designed a database schema and an indexing data structure for the team's tourism recommendation system.
- Ideated and implemented a multi-threading solution achieving 33% runtime reduction in handling API requests.

Where-to-go Android App: Tourism Recommendation, Independent | Rust

- Implemented git (basic commands, branches, remotes, and plumbing) in Rust, for educational purposes only.

SKILLS

Languages: Python, C++, Java, JavaScript, SQL
Frameworks and Tools: Pandas, Numpy, Scikit-learn, Flask, MySQL, Docker, Git