Nam Vu

● namthatman@gmail.com ● namthatman.github.io
● +61 0452-614-807 ● linkedin.com/in/namthatman

OBJECTIVE

Data scientist in marketing industry, with a bachelor's background in software engineering. I am excited about obtaining data-driven insights and leveraging data science skills to solve business problems.

SKILLS

● ***Technical Skills:*** Programming, Machine Learning (supervised, unsupervised, deep learning), Databases (data warehouse, distributed database), Association Rules, Data Visualization, Recommendation, Statistics, Natural Language Processing, Big Data

● ***Tools:*** Python (Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Keras, TensorFlow, XGBoost, LightGBM, PySpark, Librosa, NLTK, BERT, Transformer, T5, Django, Selenium, BeautifulSoup), SQL, MySQL, PostgreSQL, Git/GitHub, Jupyter Notebook, VS Code, Tableau, Excel, HTML, CSS, JavaScript, RESTful API, Command Line, AWS (EC2, S3, Kinesis, DynamoDB, Lambda, Glue, Athena, EMR, SageMaker, Elasticsearch, Redshift, QuickSight)

WORK EXPERIENCE

Zalo, Data Scientist *7/2020 – 10/2020*

● Conducted comprehensive analysis of music trends, music preferences to create users's music profile for recommendation at Zingmp3.
● Optimized the performance of music recommender system by reducing the generating time of one candidate list from 6s to 2.5s.
● Deployed the spell-checking tool to detect and fix spelling and grammatical errors in Vietnamese journals for Baomoi.com.
● Adapted new approach (Transformers/T5) to the spell-checking tool, significantly reduces checking time by half from 9s to 4.5s per 1,000 text sentences.

TMA Solutions, Backend Developer Intern *7/2017 – 9/2017*

● Cleaned and integrated over 24,000 rows data of different sources from third party partner businesses.
● Created data pipeline for regular checking data quality, maintained the integrity between data warehouse and database system.
● Optimized backend code to be delivered to front-end in milliseconds, increases the effiency of API services.

SIDE PROJECTS

BOT Detection for Online Auction

● Predictive model used to detect if an online bid is placed by a machine or human, helped easily flag these bot users for removal.
● Feature engineering on history bid data, and ensembling models from Logistic Regression, Random Forest, GBTree, XGBoost, lightGBM to delivered excellent CV score of 0.9535 on train data.

GTZAN Music Audio Classification

● Classification models used to classify music audios into genres, using 1,000 audio tracks (30s long, 22kHz Mono, 16-bit, .wav format).
● Two models are built based on different approaches: CNN model (melspectrogram images), and ANN model (sound features extraction).

Ames House Prices Regression

● Performed comprehensive EDAs, applied feature engineering on missing values, outliers, encoding, and skewed features.
● Achieved RSMLE of 0.0759 by ensembling models from lasso, elastic net, SVR, GBTree, XGBoost, LightGBM.

Movies Recommender System

● RecSys is built by ensembling 4 filtering methods: Content-based, Neighborhood-based collaborative, Matrix factorization, Neural Nets.

Language Adventures App

● A scrum team project consists of a mobile application (iOS, Android), and a web-admin application.
● LA is a story-based scavenger hunting game system, with the purpose of encouraging learning languages through video game.

EDUCATION

The University of Queensland **Brisbane, QLD, Australia**

Bachelor of Science – Computer Science *2/2018 – 7/2020*

Areas of Interest: Data Science, Machine Learning. *GPA:* 5.6 / 7.0

Ho Chi Minh University of Technology **Ho Chi Minh City, Viet Nam**

Bachelor of Computer Science *9/2015 – 6/2017*

Awards: Academic Dean's List (2015 - 2017) awarded for students with excellent academic performances.