**Nam Vu** ● [namthatman@gmail.com](mailto:namthatman@gmail.com) ● [namthatman.github.io](https://namthatman.github.io/)

● 0348-338-392 ● [linkedin.com/in/namthatman](https://www.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, Data Analytics, Machine Learning (supervised, unsupervised, deep learning), Data Ming, 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, NoSQL, 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*** *06/2020 – 10/2020*

● Conducted comprehensive analysis of music trends, music preferences to create users’s music Z-profile for recommendation at Zingmp3.

● Attempted to optimize the performance of music recommender system by applying new approach using Z-profiles.

● Deployed the spell-checking tool to detect and fix spelling and grammatical errors in Vietnamese journals for Baomoi.com.

● Adapted new approach (T5) for spell-checking tool, significantly reduces checking time by half from 14s to 6.5s per 1,000 text sentences.

***TMA Solutions, Data Engineer Intern*** *07/2017 – 09/2017*

● 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**

**[Online Auction: BOT Detection](https://namthatman.github.io/" \l "portfolio)**

● 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.

**[Music Audio Classification](https://namthatman.github.io/" \l "portfolio)**

● 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).

**[AWS Analytics Workflow for eCommerce](https://namthatman.github.io/" \l "portfolio)**

● Automated the workflow from serverlogs collection, product recommendations, rate alarms, log analysis, data warehousing, visualization.

● AWS services used including Kinesis, Lambda, DynamoDB, S3, EMR, SNS, ElasticSearch, Glue, Athena, Redshift, QuickSight.

**[Ames House Prices Regression](https://namthatman.github.io/" \l "portfolio)**

● 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](https://namthatman.github.io/" \l "portfolio)**

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

**[eCommerce: Customer Segmentation and Analytics](https://namthatman.github.io/" \l "portfolio)**

● Performed cohort analysis, conducted market-basket-analysis by applying Association Rule to extract frequent itemsets.

● Applied RFM-Recency Frequency Monetary model to identify the company’s best customers based on their behaviors.

**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*