Thanh Vuong

Data Scientist / Machine Learning Engineer

Hochiminh, Vietnam | (+84) 966097061 | phucthanhvuong@gmail.com
GitHub: thanhvuong1605 | LinkedIn: thanh-vuong-182449b0

Education

Master of Artificial Intelligence, RMIT University   2024 - Present
Scholarship Awarded
Bachelor of Engineering, Vaasa University of Applied Sciences   2014 - 2018
Software Engineering

Work Experience

Be Group JSC - Ride-hailing Application   08/2022 – Present
Senior Data Scientist/Machine Learning Engineer Lead

Led a team of 4 ML engineers to design and maintain real-time ML models for pricing, dispatch, and recommendations, car pool and delivery systems.
Developed and optimized production ML pipelines, improving scalability and efficiency.

Dynamic Pricing Engine

- Designed and implemented a surge pricing model, increasing revenue by **70-100%** during peak hours and rainy conditions.
- Developed a demand-supply forecasting model to predict market conditions for the next 30 minutes (6 time steps).
- Built a pricing engine that dynamically adjusts surge factors based on predicted demand and supply, optimizing balance in future time steps.

Multi-Dispatch Engine

- Built a predictive model to score rider-driver pairings for optimal matching.
- Implemented a linear sum assignment algorithm to maximize overall rider-driver match scores.
- Increased **completion rate (CR) by 10%** and reduced **average pickup distance by 30%**.

Food Recommendation Systems

- Developed a hybrid recommendation model combining content-based filtering and collaborative filtering.
- Implemented a DeepFM model, improving ranking **NDCG by 20%** and achieving a **15%** increase in CTR through A/B testing.
- Recommendation engine contributes **20%** of food product traffic revenue.
- Applied byte-pair encoding for Vietnamese tokenization in the food domain.
- Designed a keyword extraction model that analyzes user interactions to recommend relevant dishes.

Be Search Engine - Food and Map Search

- Architected the search systems for **beMap** and **beFood**.
- Designed search logic optimized for business objectives, ranking items based on rating, partnerships, distance, and relevance.
- Developed a keyword suggestion feature using Elasticsearch, ranking terms dynamically via exponential moving average (EMA).
- Search system drives **60%** of total food product traffic revenue.
- Enabled real-time data synchronization.

Waiting Time Estimation Model

- Built a predictive model to estimate driver waiting times within geofenced areas.
- Developed a demand forecasting model to generate real-time heatmaps for dynamic driver allocation.

Food Delivery ETA Model

- Designed a predictive model for estimating delivery time ranges, enhancing customer experience and reducing driver idle time.
- Improved **on-time delivery rate by 50%** compared to baseline statistical methods.

Dropoff Recommendation System

- Developed a model predicting user drop-off locations based on historical interactions, including location, time of day, and day of the week.
- Enhanced user experience, with **80%** of users adopting the feature.

_____

Tiki - E-commerce Platform   05/2022 – 08/2022
Senior Data Scientist - Financial Product

Customer Segmentation

- Expanded customer segmentation in the CDP to enhance targeting for insurance marketing campaigns.
- Developed segmentation models and analyzed cluster characteristics to derive actionable insights.
- Collaborated with the business team to refine customer targeting strategies and optimize marketing effectiveness.

Propensity Model

- Designed propensity models to predict customer likelihood of clicking or purchasing.
- Achieved **90% precision** and **70% recall**, generating a high-confidence list of potential customers for Tiki's insurance products.

————————————————————

Sen Do Technology JSC - E-commerce Platform   07/2019 – 05/2022
Data Scientist

Brand Detection

- Developed **Computer Vision AI models** to extract brands from product images.
- Implemented **Faster R-CNN, ResNet, and OCR models** to detect and classify brands into 500 categories.
- Achieved **over 90% accuracy**, improving market management efficiency by **300%**.

Sendo Search Engine Optimization *Led efforts to enhance the Sendo search engine, designing ranking logic and optimizing product visibility.*
*Result: CTR improved from **38% to 47%**.*

Search Query Category Mapping

- Developed a **deep learning model** to categorize search queries, improving product relevance.
- Achieved **97% precision and 80% recall**, increasing overall search CTR by **1%**.

Search Query Auto-Correction

- Implemented a **misspelling correction model** that refines user queries before processing in Elasticsearch.
- Based on research paper: *Using the Web for Language-Independent Spellchecking and Autocorrection (Google)*.
- Improved query accuracy by **60%** from test datasets.

Search Query Auto-Filtering

- Developed a **sequence labeling model (CRF)** to extract product attributes (price, size, color, location) from queries.
- Automatically applies relevant filters to refine search results.

Synonym Dictionary for Elasticsearch

- Built a **dictionary of 2,000+ synonym pairs** to improve query interpretation.
- Trained a **Word2Vec skip-gram model** on 19 million product descriptions.
- Extracted **highly similar words** to enhance Elasticsearch search accuracy.

Product Title Scoring System

- Developed a **scoring model** to rank product titles based on relevance and quality.
- Implemented **two distinct models**:
    - **Title Spamming Score**: Detects and penalizes spam-like product titles.
    - **Title Completeness Score**: Rewards well-structured, informative titles.
- Titles flagged as spam are **excluded from top search results**, improving user experience.

————————————————————

FPT Software   07/2018 – 07/2019
Machine Learning Engineer

OCR-Based Vehicle Registration Recognition

- Developed an OCR-based model for vehicle registration recognition, achieving **90% accuracy**.
- Improved document processing efficiency through automated extraction of license plate details.

Anomaly Detection in Automotive Sensor Data

- Implemented an anomaly detection system to identify irregularities in automotive sensor data.
- Enhanced predictive maintenance by detecting potential failures before they occur.

Certifications and Awards

The Contributor Award - Be Group	2024
Best Employee Recognition	
Top Performer of The Year - Data Scientist at Be Group	2023
Top Performer of The Year - Data Scientist at Be Group	2022
Best Employee Of Organic Growth Tribe - Data Scientist at Sendo	2021
Best Employee Of The Year Award - Data Scientist at Sendo	2020
Data Science Mentor at MASSP - Math and Summer Science Program	2021

Technical Skills

Programming: Python, SQL
Machine Learning: Deep Learning, Time Series, NLP, Recommender Systems
Big Data Technologies: Apache Spark, Kafka, Hadoop, Airflow, GCP, BigQuery, PubSub
DevOps & Deployment: Docker, Kubernetes, CI/CD Pipelines, FastAPI, Grafana
Databases: Elasticsearch, PostgreSQL, MongoDB, Redis

Languages

Vietnamese: Native
English: Professional working proficiency

References

Assoc. Prof. Xuan-Hieu Phan, Head of Research and Development Center - Sendo Technology JSC
Email: pxhieu@gmail.com
PhD. Hoang-Huy Nguyen - AI Director at Techcombank
Email: nhhuy2810@gmail.com
Hieu Nguyen - Head of Data Science at Be Group
Email: vuondenthanhcong11@gmail.com