To Duc Anh

toducanh2001@gmail.com
My Phone: +84 983691898
My LinkedIn: To Duc Anh

 Hyprnx
</> Python, PySpark, AWS
See light theme

Introduction

Results-driven Management Associate with a High Distinction degree in Data Science from National Economics University. Skilled in Mathematics, Artificial Intelligence, and Engineering. Proficient in self-study, communication, and work in both on-premise and on-cloud environments. Dedicated to deliver high-quality work.

Experiences

- **Techcombank Vietnam (TCB) - Data and Analytics Division** — Hanoi, Vietnam
  Senior Data Engineer - Key Talent — *January 2023 - July 2024*
  Data Engineering and Delivery - Risk Domain Delivery Squad — *Aug 2024 - now*
  - Develop new ETLs that serve Risk related domain, comply with circulars from SBV and decrees from the government.
  - Develop serving data pipeline via GraphQL/Go Anywhere MFT to match with the requirements from Vietnam Government based on TCB's Databricks data platform.
  - Participated in "Save Cost" campaign, optimized existing ETL pipeline, saved 16.000+ USD annually.
  - Contributed 15+ features in Enterprise Feature Store.
  - Contributed serving writer to TCB's Datalake orchestration framework.

- **Techcombank Vietnam (TCB) - Data and Analytics Division** — Hanoi, Vietnam
  Management Associate - Key Talent — *January 2023 - July 2024*
  Data Engineering and Delivery - Landing Position
  - Maintain a went-live strategic project. Fabricated 10+ ETLs jobs and contributed 200+ features to Risk Datamart. Take part in building Credit Application Fraud system.

  Data Governance
  - Designed, implemented and automated the process of updating Hierarchy Address to align with data published by Vietnamese General Statistics Office, resulting in improved onboarding for new-to-bank (NTB) customers and standardized exist-to-bank (ETB) customers.

  Advanced Analytics and Innovation
  - Built CI/CD pipeline that tested and validated Airflow DAGs, facilitating other divisions in orchestrating Features Store. Developed ETL pipeline to transfer Machine Learning models' inference results to TCB's DataLake. Analyzed and developed Machine Learning/Deep Learning models for analyzing and predicting customer transactions. Assessed business requirements and created AWS ad-group for business users to facilitate building and migrating Machine Learning pipelines to the cloud environment.
  - Implemented OpenAI's Whisper based inference engine as a core module of a innovation (HienOi) project to transcribe customers's scheech to text for sentiments analysis.

  Business Intelligence
  - Played a deciding role in developing a strategic initiative project for Techcombank as Technology Advisor for Accenture Outsourcing Team. Participated and delivered User's Workbench, a service that enabled business users to build and analyze isolated data specifically tailored.

  Data Engineering and Delivery
  - Analyzed data updating pattern for CDC and SCD types resulting in improved data accuracy. Successfully delivered 30 data ETLs based on PySpark that run daily. Developed Loyalty ID generation function. Proficient in Techcombank's tech stack and cloud environments.

- **AI for Everyone** — Online
  Teaching Assistant — *March 2023 - Now*

- ○ Assisted in constructing and evaluating students' homework, resulting in a 20% increase in completion rates. Supported students with questions on Machine Learning, Deep Learning, and Model Deployment.

- **KiotViet (Former Citigo Software) - AI team, Data Department**   Hanoi, Vietnam
  Intern Data Scientist   *August 2022 - December 2022*
  - ○ Automate product classification process by building text classification machine learning model to predict categories using product names. Designed and implemented a package that extracted brand names from product names through fuzzy search, regex, and Trie Data Structure, which later have been published to internal Gitlab. Upgraded the Address Suggestion module to enhance suggestion accuracy utilizing existing module and Elasticsearch, resulting in a 20% increase in accuracy.

- **DSLAB NEU**   Hanoi, Vietnam
  Research Assistant   *November 2021 - Present*
  - ○ Responsible for building a face attendance system on nVidia Jetson Nano. The model recorded a 95% accuracy on testset

Following sections items are clickable on the PDF version.

Education

- **National Economics University**   Hanoi, Vietnam
  Major: Data Science in Economics and Business   *September 2019 – May 2023*
  - ○ Teaching Assistant for Data Structures & Algorithms, and Machine Learning classes.
  - ○ GPA: 3.66/4.0
  - ○ Third prize in student scientific research contest
  - ○ Forth consecutive semester nominated for outstanding results and performance
  - ○ Achieved A+ in Data Structure and Algorithm, Machine Learning, Data Preparation and Visualization course

Courses

- ○ Advanced Architecting on AWS
- ○ AWS Security Governance at Scale
- ○ Architecting on AWS
- ○ AWS Cloud Practitioner Essentials.

Certifications

- **Certificate**   Received Date
  (Name and score(s) are given, if possible below)
  - Third-prize AWS x Techcombank Deepracer   *Nov 2023*
  - IELTS 7.0   *April 2019*
  - The Internet and Computing Core Certification: 2520   *December 2019*
  - BLOCKCHAIN MATHEMATICS AND COMPUTING   *July 2021*
  - Qiskit Localization Contributor   *May 2021*

Personal Projects

- Car Specifications: Designed and implemented an open-source project that was created with the primary purpose of saving time for the cars-related research community. The project offers the community the dataset that includes 44934 car models and variations that were mass produced from 1985 to early 2022. The project also comprised a Scrapy-based crawler that have clear instruction to re-crawl if needed.

- Text Classification: Directed and collaborated with a team of four to design and implement a classification module to classify product names on e-commerce platforms into four distinct categories. The module was later deployed on Streamlit Cloud for demonstration. Models that are based on transformer architecture (sBERT and phoBERT) were used to calculate text embedding before passing to custom-made two layers Neural-Net for the classification task. The ONNX version of the models was used to accelerate the inference time.

- Personal Website: A website were made using HTML and TailwindCSS framework, with the main purpose of making a portfolio/introduction page for myself.

Publications

- **Used Cars Prices Prediction**   Published on National Scientific Conference
  Hanoi, April, 2022

  Archived Third Prize in NEU Scientific Research Competition

  ISBN: 978-604-358-602-2
  -Investigated and extracted used cars information that are selling on Vietnamese E-commerce site.
  -Constructed a Machine Learning model to help Vietnamese choosing the right used cars for their demands.
  -Conducted a dataset consists of all cars specifications (with variants) while doing Data Prepossessing.

Techstack

- Platform: AWS, Databricks

- Operating System: Windows, MacOS, Linux (Ubuntu)

- Programming Languagues: Python, C.

- Databases/Table Format: Apache Iceberg, Databricks DeltaLake, AWS Athena, AWS Dynamo. MSSQL, Postgresql

- Frameworks: Apache Spark, Pytorch, Onnx, Sklearn, Flask.

- ETL tools: AWS Glue, AWS S3, AWS Lambda, AWS Step Functions.

- IaC: Terraform, Terragrunt.

- Version Control System: git, Github, Gitlab.

Languages

- Vietnamese: Native.

- English: Full working proficiency.