Duy Tran (Jude)

(+84) 949910412 | anhduy412works@gmail.com | linkedin.com/in/anhduy12 | Portfolio

SUMMARY

I have over 3 years of experiences working multiple roles in the data and analytics field such as Data Scientist, Data Engineer, Data Analyst; proven records in delivering resultful, data-driven machine learning, data analysis, and Python-based solutions. Now, I am eager to apply my analytical mindset, technical skills, and curiosity to contribute and innovative solutions and continuing to grow as a data professional.

EXPERIENCES

Data Engineer | FPT Software

December 2025 – Present

- Join projects as Data Engineer

Researcher & Analyst – Data & AI | FPT Digital Co. Ltd

June 2025 – September 2025

- Researched and made reports on current status and upcoming trends of applying data analysis and advanced data analysis in multiple domains including retail and production.
- Provided consultancy on data strategy, data governance, and data architecture to help clients leverage their data assets effectively.
- Assisted clients in identifying data-related challenges and opportunities, and recommended solutions to improve data quality, accessibility, and usability.
- Collaborated in FRT Long Chau project from July 2025 to September 2025 as a Data Analyst with main responsibilities are getting weather, holiday and disease data for feature engineering and normalise SKUs data.

Data Scientist | BlueInfo Joint Stock Company

November 2024 – May 2025

- Developed and deployed a Lead Generation and Credit Scoring model for VNPT Vinaphone subscribers, incorporating telecom and financial behaviour to generate propensity scores.
- Integrated real-time scoring APIs into production systems, improving customer targeting and segmentation for 6+ partners including MCredit, FE Credit, and VNPT Money.
- Collaborated cross-functionally to ensure smooth integration of the model into existing systems.
- Contributed to research and ideation for 3+ upcoming projects, including an SME credit scoring model for Viettel Digital Solutions.

Contract Data Staff | Asia Trade Fair & Business Promotion Joint Stock Company

September 2024 – October 2024 & March 2025 – April 2025

- Built a web scraper using Python with BeautifulSoup4, requests to get business data collection from public sources.
- Compiled a cleaned dataset of 6,700+ businesses across multiple sectors for event and client outreach.
- Enabled data-driven targeting for trade events in lifestyle, healthcare, construction, and transportation.

Data Scientist | KPIM Joint Stock Company

April 2023 – July 2024

- Led and contributed to various data science initiatives including a content-based recommendation system and a predictive model for agricultural water quality.
- Developed and integrated machine learning modules and APIs into COMPACLASS, an internal education platform.
- Built and maintained data infrastructure using Microsoft SQL Server and SQL Server Integration Services (SSIS) for 7+ enterprise clients, including Sonion Vietnam, Jeisys Medical Vietnam, Aeon Mall Vietnam, and Consortio Vietnam Technologies.
- Designed a scalable medallion architecture using Microsoft Fabric, enabling structured data progression and real-time reporting.
- Nominated "Best New Employee of 2023" for outstanding contributions.

Data Scientist | XTEK Technology Co., LTD

April 2022 – July 2023

- Took part in developing a face recognition system using TensorFlow and OpenCV as a part of the collaboration between the company and the Korea International Cooperation Agency (KOICA), supporting accurate, real-time identification.
- Recognized by the CTO for technical excellence and strong work ethic.

Software Engineer | INTER ITS., JSC

November 2021 – May 2022

- Built and maintained features for web apps across 8+ clients using Flask, FastAPI, MongoDB.
- Named "The Best Employee" in November 2021, March 2022, April 2022 praised by C-level executives for reliability and performance.

EDUCATION

Electric Power University

Software Engineering | 2019 – 2024

- Achieved multiple awards in hackathons, scientific research and start-up competitions throughout the studying process.

SKILLS

Technical Expertise

- **Machine Learning:** Familiar with classification, regression, clustering; with experiences in scoring models, recommendation systems, computer vision.
- **Data Analysis:** Data storytelling, visualization with Python libraries or Microsoft Power BI.
- **Data Engineering:** ETL pipelines, data warehousing in medallion architecture, database management.

Tools & Technology

- **Machine Learning and Deep Learning frameworks:** PyTorch, TensorFlow, scikit-learn.
- **Data Science libraries:** PySpark, NumPy, Pandas.
- **Databases:** SQL (Microsoft SQL Server, PostgreSQL, etc.), NoSQL (MongoDB).
- **Programming:** proficient in Python, basic in R.

- **Software Development:** system design, design patterns, clean code, Agile/Scrum.
- Familiar with **Docker, Ubuntu, VMWare.**

Soft Skills

- Using English as the second language (IELTS 7.0 OVR in 2014).
- Strong research, analytical and critical thinking skills.
- Adaptive learner with adaptability to recent technologies.
- Ability to work independently.

FEATURED PROJECTS

PROJECT	PERIOD	DESCRIPTION
Lead generation score and credit score project	11/2024 – 05/2025	**Overview:** Built a predictive model to assess customer response to loan offers and credit scores using multi-source telecom data. **Role:** Data Scientist, Data Engineer | **Team Size:** 3 **Key Contributions:** • Defined target variable, cleaned data, managed missing values, and removed low-variance features. • Engineered indicators from calling behaviour, data usage, and financial metrics; normalized for modelling. • Developed sub-models with XGBoost and combined them into an ensemble model. • Automated the data pipeline using Airflow with configurable DAGs for flexible scheduling. **Result:** • Delivered a scoring model (0–100 for lead generation, 300–850 for credit) to identify high-potential loan prospects. • Enabled real-time scoring via API integration to enhance marketing precision. **Tools:** PySpark, Apache Airflow, XGBoost, pandas, numpy.
Business info crawling tool	09/2024 – 10/2024; 03/2025 – 04/2025	**Overview:** Developed a web crawler to automate extraction of business information from trangvangvietnam to create a comprehensive vendor dataset. **Role:** Data Staff | **Team Size:** 1 **Key Contributions:** • Collaborated with the ATFA team to define scraping and data requirements. • Built and deployed a custom scraper using BeautifulSoup4 and requests.

PROJECT	PERIOD	DESCRIPTION
		• Cleaned and validated extracted data to manage parsing inconsistencies. **Result:** • Generated a clean dataset of 6,000+ businesses with full contact details (name, address, phone, etc.). **Tools:** Python, BeautifulSoup4, requests, pandas, Excel
Jeisys Medical Vietnam data infrastructure	06/2023 – 07/2024	**Overview:** Designed and implemented a data infrastructure using Microsoft Fabric to improve data storage, integration, and analytics capabilities. **Role:** Data Scientist, Business Analyst | **Team Size:** 10 (5 from KPIM., JSC; 5 from Jeisys Medical Vietnam). **Key Contributions:** • Collaborated with stakeholders to assess data challenges and requirements. • Built a 3-layer medallion architecture (Bronze, Silver, Gold) using Microsoft Fabric. • Created automated ELT pipelines via Dataflow Gen2 to ingest data from 5+ sources (Odoo CRM, SAP, etc.). • Applied PySpark to clean and transform data across layers. • Produced data visualizations for validation and documented data lineage and governance processes. **Result:** • Reduced manual data extraction by 70% and improved data quality for analytics by 90%. • Established scalable, AI/ML-ready infrastructure for future business intelligence initiatives. **Tools:** Microsoft Fabric, PySpark, PostgreSQL, Power Query.
Sonion Vietnam factory data infrastructure	07/2024 – 12/2023, 05/2024 – 07/2024	**Overview:** Built and maintained a data infrastructure as a part of the reporting system to support real-time production line operation management across two factories. **Role:** Data Scientist **| Team Size:** 2 (07/2024 – 12/2023), 5 (05/2024–07/2024). **Key Contributions:** • Maintained and optimised two factory data warehouses for performance and reliability. • Developed and deployed ETL pipelines using SSIS, reducing manual processing time by 50%.

PROJECT	PERIOD	DESCRIPTION
		• Designed new tables and features to support custom dashboards and reports based on stakeholder needs. **Result:** • Improved data accessibility and reporting accuracy for production managers, enhancing decision-making. **Tools:** T-SQL, Microsoft SQL Server, SSIS.
COMPACLASS	**05/2023 – 07/2024**	**Overview:** Contributed to the development of an online learning platform ecosystem aimed at enhancing teaching and learning experiences. **Role:** Data Scientist, Software Engineer | **Team Size:** 11 **Key Contributions:** • Designed and implemented a **Practice Module** using Python and ML/DL libraries to support post-course knowledge reinforcement. • Worked cross-functionally with business analysts and developers to ensure seamless integration. **Result:** • Improved course completion rates by 25% through data-driven learning features and interactive practice tools. **Tools:** Python, including Pandas, NumPy, Scikit-learn, PyTorch, FastAPI, Typer.
Agriculture water quality predictive model	**10/2023 – 11/2023**	**Overview:** Developed an ML-powered solution to predict water quality for agricultural use, supporting efficient irrigation practices. **Role:** Project Manager, Data Scientist, Software Engineer. **Team Size:** 3 **Key Contributions:** • Researched water quality challenges and co-authored a business proposal to align model design with stakeholder goals. • Built a predictive model using XGBoost and deployed it via a production-ready API. • Led the project from concept through deployment, ensuring technical and business alignment. **Result:** • Improved irrigation decision-making with predictive insights, reducing resource waste and enhancing sustainability.

PROJECT	PERIOD	DESCRIPTION
		Tools: Python, XGBoost, Flask, Streamlit.
KOICA face recognition application	04/2022 – 07/2023	**Overview:** Built and deployed a face recognition system for KOICA, using deep learning for high-accuracy identification. **Role:** Data Scientist | **Team Size:** 3 **Key Contributions:** • Collaborated with KOICA to gather requirements and define application scope. • Developed the system using TensorFlow and OpenCV for real-time facial recognition. **Result:** • Deployed the solution in KOICA's operational environment, enabling reliable daily identifications. **Tools:** Python, TensorFlow, OpenCV

CERTIFICATIONS

Google Data Analytics – Coursera
Google AI Essentials – Coursera
Career Essentials in Data Analysis – *Microsoft and LinkedIn*

LIST OF ACHIVEMENTS

Semi-finalist at Global E-Commerce Challenge 2024 by Alibaba International
Topic: Flashback Altems to turn abandoned cart into sales in a more fun and engaging way.
Top 12 of university students' group at 2022 Data Science Talent Competition
Topic: Unemployment predictor using Machine Learning algorithms.
Top 12 at G'Contest 2022: Digital Economy
Topic: Debt predicting of credit card users using Machine Learning algorithms.
Third prize at 2021 Start-up Competition by Electric Power University
Topic: Warning and preventing of Malware, APT, BOTNET threats product for SMEs.
Second prize at 2021 Scientific Research Competition, Electric Power University
Topic: SCADA-PDS project to protect the national electric system against cybersecurity threats.
First prize at 2020 Scientific Research Competition, Electric Power University
Topic: Face recognition application for anti-cheating in examinations.