Khanh Duong TRAN

Data Scientist Intern | Île-de-France

raydiwill | Khanh Duong TRAN | duong.tranhn1102@gmail.com | +33 6 62 48 81 52

Experience

VIETTEL GROUP | Associate Data Analyst

April 2022 - June 2022 | **Hanoi, Vietnam**

- Utilised SQL for the analysis of telecommunications customer usage data.
- Assisted in compiling reports evaluating the effectiveness of Data Science applications compared to traditional telecommunication campaigns.
- Researched tree-based Machine Learning algorithms for customer churn predictions.

Projects

CHURN PREDICTION ANALYTICS PLATFORM | Tensorflow, FastAPI, Airflow

https://github.com/raydiwill/dsa-4-action-learning | **Machine Learning Engineer**

Objective: Researched state-of-the-art deep learning models to develop a Minimum Viable Product for telecommunication churn prediction.

Details:

- Designed and implemented a **hybrid ensemble deep learning model** (DNN and BiLSTM) for telecom churn prediction, achieved an **88% accuracy**.
- Logged models metadata via **MLflow**.
- Deployed the model into a **web application** using **FastAPI** for real-time predictions and facilitating data retrieval from the database.
- Automated scheduled prediction tasks using **Apache Airflow** via **Docker**, enhancing operational efficiency and model applicability.
- Integrated Great Expectations for **data quality validation**.
- Developed **interactive dashboards** monitoring model performance with **Grafana** and integrated into a **Streamlit UI**.

Results: Gained skills in building Deep Learning models and API development.

BANKING CUSTOMERS CHURN PREDICTION | Imblearn, Scikit-learn, XGBoost

https://github.com/raydiwill/customers-churn-ml-app | **Data Scientist**

Objective: Applied knowledge from previous courses to deploy churn prediction machine learning models into production through a 5-member group project.

Details:

- Conducted Exploratory Data Analysis on customer data, **identifying key features, class imbalances and correlations influencing target variable**.
- Implemented **scikit-learn pipelines** for efficient feature processing.
- Applied **SMOTE** and **class weight adjustments** to tackle class imbalance, aiming to enhance model equity and fairness across classes.
- **Implemented and evaluated** Logistic Regression, Random Forest, Gradient Boosting, and XGBoost, using **confusion matrices, ROC curves and AUC** against baseline models.
- **Employed RandomSearchCV, Optuna** for hyperparameter tuning across models; **gained proficiency in hyperparameter tuning** for model optimization.
- **Implemented feature selection** through features importance to reduce training time and boost model performance, aiming to optimize the predictive pipeline.

Results: Enhanced model building and hyperparameter tuning capabilities.

Skills

PROGRAMMING

Proficient:
Python • SQL

Familiar:
Java • Shell Script • Javascript • PHP • C/C++ • R • HTML & CSS

LIBRARIES/FRAMEWORKS

Pandas • Scikit-learn • Tensorflow • Optuna • Matplotlib • XGBoost • Imblearn • Numpy • FastAPI • Streamlit

TOOLS/PLATFORMS

Git • GitHub • Dataiku • AWS • Tableau • Docker • MLflow • MongoDB • MySQL • DBeaver • Apache Airflow • VirtualBox

OPERATING SYSTEM

Unix

LANGUAGES

- **English (C1)**
- **French (B1)**
- Vietnamese

SOFT

- Teamwork
- Inquisitive
- Adaptability
- Autonomous
- Problem solving

Education

EPITA - SCHOOL OF ENGINEERING AND COMPUTER SCIENCE

M.S.c in Computer Science

March 2023 - Present | Paris, France

- Major in Data Science and Analytics.
- Rank: 2/45 in cohort.
- Rank: 1/18 in specialization.