Linh Nguyen

npl02042000@gmail.com | (484) 343-5733 | Philadelphia, PA

https://npl0204.github.io | https://linkedin.com/in/linhpn0204 | https://github.com/npl0204

PROFESSIONAL EXPERIENCE

Data Scientist Co-Op @ Oracle | Remote, PA September 2023 – Present

- Led research initiatives in Federated Learning for Neural Network and Transfer Learning using TensorFlow and PyTorch, focusing on innovative strategies to enhance product optimization and performance, while facilitating robust, scalable, and privacy-preserving machine learning models.
- Conducted independent research on K-means and Gaussian Mixture Models for clustering, aimed at optimizing existing models. This involved developing novel approaches, analyzing their effectiveness, and integrating these methods into our workflow.
- Spearheaded the adoption of Oracle Cloud Infrastructure (OCI) for the team, integrating it with transfer learning and clustering methodologies.

Software Development Engineer Intern @ Amazon | Sunnyvale, CA June – September 2023

- Created a JIRA-enhancing tool to optimize task automation and workflow efficiency. The product is currently in use, contributing to improved efficiency in project management and communication within the organization, reduced time required for management tasks from weeks to 5 minutes per session.
- Utilized AWS DynamoDB for data transformation, harnessed AWS EventBridge, Lambda, and SES for automation. Employed Java for backend development, MVC with JSP for frontend, and Coral Services for API integration.
- Developed new functions for existing packages and seamlessly integrated them into Native AWS products using Java and AWS external/internal tools.

Data Scientist Intern @ Intermountain Health | Remote, UT January – March 2023

- Developed AI/ML recommendation system for clinical/supply chain staff for Doctor Preference Cards (DPC) decision-making process based on analysis of complex data sets of over 55,000 DPC using statistical functions, PySpark, and SQL, preventing a potential $13.65M yearly loss of surgical savings.
- Provided executive-level, data-driven insights and visualizations that replace Tableau by implementing a more effective analytic application.
- Designed baseline ML models to predict optimal supplies purchase for perioperative procedures, expected to result in savings of $7M/year by reducing surgical supply spending, optimizing working capital, and reducing inventory spoilage.

Research Assistant @ Rezapour Lab Drexel University | Philadelphia, PA June 2022 – September 2023

- Conducted data collection and analysis on social media to extract and investigate unstructured user-generated content, utilizing natural language processing ML models to classify characteristics of drug-related discussions for the baseline of an AI chatbot.
- Conducted in-depth research and analysis on computational social science, focusing on evaluating moral values in texts, optimism/pessimism detection, and methods for maintaining a positive sentiment in online communities.

Data Analyst Intern @ Vietnam Ministry of Health – Nanocovax Vaccine (Phase II and III) | Hanoi, Vietnam June – December 2021

- Imported and pre-processed biological data of 13,000 participants into a SQL database and performed ETL process and exploratory data.
- Designed machine learning models to predict the hospital length of stay (LOS) during COVID-19 based on pre-admission features in different hospitals and to predict the total number of cases by day in Vietnam based on previous data in multiple regions.
- The LOS prediction model achieved 85% accuracy during pandemic's peak period, assisting in changing hospital structure and management system, resulting in a decrease in healthcare-associated infection rate by 63% and an increase in hospital profit by 78% over 3 months.

Intern @ Vinmec International Hospital | Hanoi, Vietnam June – September 2020

- Devised deployment plan of Vinmec online pharmacy service. Worked as project manager and collaborated with 3+ application/website developers.
- Analyzed data on customers' behavior and the sales between the traditional and online pharmacies in two months using NumPy and Pandas and visualized the result using Tableau, resulting in an increase in sales of 125% compared to before the campaign.

Math Tutor @ Tutorful | Birmingham, UK December 2018 – December 2021

- Delivered 3000+ hours of tailored mathematics instruction to multiple levels, including KS3, GCSE, A-Level, and other qualifications.

LEADERSHIP AND ACTIVITIES

Ambassador @ Women in Data Science | Philadelphia, PA November 2023 – Present

Founder/President @ Drexel Association for Data Science | Philadelphia, PA February 2023 – Present

Senior CCI Dean's Ambassador @ Drexel University | Philadelphia, PA September 2022 – Present

President @ Vietnamese Society of University of Birmingham | Birmingham, UK 2019 – 2021

Competitions Sub-committee and International Representative @ British Pharmaceutical Students' Association | UK 2019 – 2021

EDUCATION

Drexel University, GPA: 4.0/4.0, Dean's List January 2022 – September 2024

- B.S. in Data Science; minor in Computer Science — GHC 23 Scholar, Anthony J. Drexel Merit Scholarship, Day Alumni Fund
- Teaching Assistant (20 hrs/week) for Data Science Programming, Database Management Systems, Introduction to Computing and Security Technology

University of Birmingham, GPA: 4.25/4.0, First-Class Honors September 2018 – June 2021

- Bachelor of Pharmacy (finished 3 out of 4 years) — International Scholarship Award (Highest Scholarship)

TECHNICAL PROJECTS

Fake News Detection – Text-Processing and Natural Language Processing

- Led a team of four to develop an efficient NLP model to predict the authenticity of 72,134 online news articles.
- Performed text preprocessing such as removing noise, tokenization, lemmatization, performing TF-IDF, etc.
- Examined seven Neural Network models (with Long Short-Term Memory) achieved the highest accuracy (94.51%) and F1 (94.71%) scores.

Brain Tumor Detection from MRI Images – Image-processing and Deep Learning

- Developed deep learning models for brain tumor detection using MRI images by employing MLP, CNN, VGG16, EfficientNet B2, and InceptionV3 models. Achieved the best performance using InceptionV3 with an F1 score of 93.55% and an accuracy score of 94.74%.
- Performed pre-processing using data generators for image processing and augmentation with the ImageDataGenerator.

Credit Card Default Classification – Data Mining and Machine Learning

- Performed EDA to visualize/investigate the dataset, and built classification models using Scikit-learn, PyTorch, and Matplotlib.
- Improved the F1 score of a simple Logistic Regression from 0% to 54.62% with an accuracy of 80.33% by tuning the decision threshold and other hyperparameters of the Extreme Gradient Boosting (XGBoost) model.

RELEVANT COURSEWORK AND SKILLS

Python, SQL, R, Java, C, Dart, MVC, Microsoft Excel, Tableau, Algorithms and Data Structures, Machine Learning, Artificial Intelligence, Big Data, Statistical Data Analysis, Hypothesis Testing, Database Management Systems, Data Visualization, System Design, Linear Algebra, Web Development