

Minnie Nguyen

📞 Phone: 438-764-0121 | ✉ Email: minniexnguyen@gmail.com

🌐 Website: <https://nminnie.github.io> | 🐙 Github: nminnie | 🔗 LinkedIn: nminnie

Skills

Machine Learning Classification, Regression, Time series analysis, Feature engineering, Visualization, Embeddings

Python Modules Scikit-Learn, Keras, PyTorch, Pandas, NumPy, Matplotlib, PyLab, StatsModel, Gensim, NLTK

Coding Python, SQL, JavaScript, Java, R, Matlab

Tools & Frameworks Git, Jupyter, Node.js, Django, HTML5, Android

Work Experience

Dynamicly Inc. (Startup)

Montreal, Canada

DATA SCIENTIST INTERN

Jun - Oct 2019

- Spearheaded an R&D project in the Data Intelligence team to build a multi-class text classification model for virtual assistants' NLU, using TF-IDF, FastText, and Word2Vec embeddings, which improved prediction accuracy by 7%.
- Conducted research and implemented state-of-the-art classification models such as LSTM, Doc2Vec, SVM, and various ensemble methods using Scikit-Learn and Keras.
- Implemented a flexible, intelligent pattern matching algorithm in Python regex that allows phrases of variable length and synonym matches using GloVe word embeddings.
- Extracted text data from user interactions, cleaned the data with stop-words removal, tokenization and lemmatization, and resolved data imbalance.

McGill University, Schulich School of Music

Montreal, Canada

STUDENT SOFTWARE DEVELOPER

May - Aug 2018

- Implemented 8 new functionalities for the Interactive Classifier, a web-based symbol classification software used by musicologists at McGill University and Dalhousie University, using Git, JavaScript, Node.js in a Django framework.
- Analyzed and optimized the runtime performance of zooming and drag-select actions by 90% and delete actions by 75%, using Chrome Developer Tools.
- Presented about the software and its advancements at multiple music seminars workshops at McGill University.

Projects

Time series Analysis & Price Forecasting (PyLab, StatsModel, Pandas, NumPy, Matplotlib)

- ARIMA time series model that analyzes the trends and seasonality of avocado prices in the last 5 years in the US and accurately predicts future prices with a low RMSE of \$0.127.

Book Recommendation System (Scikit-Learn, SciPy, Pandas, NumPy)

- kNN model that uses collaborative filtering to recommend the top-n most similar books based on 6 million ratings of 10,000 books on Goodreads.

Reddit Popularity Predictor (Scikit-Learn, Pandas, NumPy, Matplotlib)

- Constructed word features from a dataset of 12,000 Reddit comments and employed linear regression using gradient descent to predict the popularity score of Reddit comments.

Handwritten Digit Recognition (PyTorch, Keras, OpenCV, Scikit-Learn, Pandas, NumPy, Matplotlib)

- 18-layer ResNet (CNN) for the MNIST handwritten digit dataset with accuracy of 97.5%, implemented with PyTorch and uses OpenCV for noise removal and image transformations for data augmentation.

Education

McGill University

Montreal, Canada

B.Sc. IN MATHEMATICS & COMPUTER SCIENCE, GPA: 3.73

Sep 2015 - May 2019

- Awards: Major Entrance Scholarship, Tomlinson Engagement Award for Mentoring (4-time recipient)
- Courses: Applied Machine Learning, Statistics, Probability, Stochastic Processes, Software Design, Database Systems, Algorithms & Data Structures