

Employment

Springboard Data Science Career Track

Data Science Fellow - NLP

Toronto, CA
May '19 - Current

- Worked on over 550 hours of hands-on curriculum mentored 1:1 by Data Science Mentors.
- Mastered skills in Python, SQL, data analysis, data visualization, hypothesis testing and machine learning, culminating in 2 in-depth capstone projects.

Independent Instructor

Academic Coach

Toronto, CA
Oct. '17 - Feb. '20

- Designed lessons to fit the needs of client through digital and face-to-face training, with 90% satisfaction rate.
- Directly coached +30 students in Data Management and Mathematics, Business, Python and R.

Wipro Technology

Business Analyst

Hanoi, VN
Jun. '14 - Jul. '15

- Conducted daily spreadsheet analysis with JIRA issue tracking to aggregate, summarize and highlight functional and technical defects through pivot table and VBA automation, saving supervisor 2 hours/day of manual data manipulation.
- Built weekly defect report dashboard tool with insights, prioritization and recommendations, allowing management to make important resource allocation and risk mitigation decisions, contributing significantly to the timely completion the project.

Projects

Emotionally Aware Movie Recommender WebApp - Fast Film

- Tools: ScikitLearn-Surprise, Gensim, Universal Sentence Encoder, Flask, HTML/CSS, Docker, AWS
- Won First-prize in Data Product Competition hosted by Aggregate Intellect by deploying an Emotionally Aware Movie Recommender WebApp with the aim to save 20% of Netflix user's film watching time.
- Conducted AI product development including Ideation, Business Strategy, Product Strategy, User Testing and Alpha & Beta Deployment and Demo.
- Conducted full-stack data modelling and engineering, including movie data scraping, NLP feature modeling & development and UI development.

Deployment of Spam Filtering WebApp

- Tools: Pandas, Json, Nltk, Matplotlib, Scikit-Learn, Spacy, Statsmodel, Docker, Flask, AWS
- Deployed Spam Filter WebApp onto AWS instance with Dockerization so that anyone could filter Spams.
- Achieved a model with 99.4% accuracy score and 0.997 AUC (Area Under the Curve) Score.
- Created end-to-end NLP pipeline for spam classification, filtering spam at throughput of +100 sms/s.
- Converted unstructured data to Bag of Word features and modeled the training data to several algorithms Logistic regression, Naive Bayes, SVC and XGBoost.
- Regularized best algorithm architecture with F-Beta score to ensure alignment with business objectives.

Medical Review Sentiment Analysis using Neural Network Techniques

- Python tools: Pandas, Seaborn, Scikit-Learn, Spacy, Gensim, Sci-Py, Tensorflow/Keras.
- Performed sentiment analysis on a highly realistic dataset provided by Analytics Vidha with over 100 drugs for a hypothetical Canadian pharmaceutical, achieving a model with 0.5 F1 macro score with Grid-Search for Hyperparameter Tuning and Cross Validation.
- Aligned the best algorithm with business objectives by creating trade-off gradient to identify scenarios in which automation of the Sentiment Analysis task is profitable, potentially with a saving of +\$400 per 1000 reviews.
- Ran dimensionality reduction algorithms such as TSNE and LDA (Latent Dirichlet Allocation) to explore language dataset structure.
- Modeled the training data to several algorithms, including Word embedding algorithms (with Skip-Gram, Matrix Factorization and Transfer Learning) and Recurrent Neural Network with LSTM (Long Short Term Memory) component.

Powerful Visualizations Demo to learn Exoplanet properties

- Python tools: Matplotlib, Seaborn, Astropy, Bokeh, Plotly, Ipywidgets, SciPy.
- Extrapolated data from API of NASA Exoplanet Archive displaying properties of 4031 exoplanets.
- Created interactive visualizations such as 3D scatterplot and Bokeh tabular scatterplot to derive astrophysics theories through empirical data and first principles.

Volunteering

Aggregate Intellect - NLP Stream Owner

Current

- Coordinating with leading researchers and authors to present their work relating to NLP.
- Facilitating discussions on YouTube channel with more than 10,000 subscribers.
- Presenting a spotlight talk on practical business applications of NLP

Contact

✉ zachnguyen0111@gmail.com

☎ (647)-606-5628

📍 Toronto, ON (Open to relocate)
in haingocnguyen

🌐 zachnguyen

Education

Schulich School of Business

May '15 - Oct. '17

Masters Business Administration

- Underwent Deloitte Distinction Program for consulting training.
- Worked as Vice President of Business and Technology Management Association.

Colorado State University

Sep. '09 - May '13

Bachelors Economics

- Used economic regression model on Capstone Project to conduct economic consultation for local government of Larimer County

Skills

PROGRAMMING SKILLS

Python

SQL

R

SAS

Visual Basics for Applications (VBA)

MACHINE LEARNING & ENGINEERING

Supervised Learning (Regression, Classification)

Unsupervised Learning (Clustering, Dimensionality Reduction)

ML Frameworks (Scikit-Learn, Pyspark, Keras/Tensorflow)

Version Control (Git)

MLOps (Flask, Docker, AWS)

NLP (Spacy, Gensim, Nltk)

DATA ANALYSIS

Data Wrangling (Pandas, Numpy)

Data Visualization (Seaborn, Bokeh, Plotly)

Data Collection (csv, json, api, scrapy, bs4)

STATISTICS AND PROBABILITY

Inferential Statistics

Descriptive Statistics

Probability Distributions

Bootstrap Statistics

Bayesian Inference (pymc3)

A/B Testing

Hypothesis Testing

SOFT-SKILLS

Storytelling

Business Acumen

Communication

Product

Coaching

Negotiation

Project Management