

MERVIN S. McDOUGALL | DATA SCIENTIST

Baldwinsville, NY 13207 / mcdmerv@yahoo.com / (315) 749-5129 / <https://www.linkedin.com/in/mervin-mcdougall/>

GitHub URL: <https://github.com/mervinmcdougall>

Data scientist with a history in web development and a strong foundation in machine learning, natural language processing, and data analytics, seeking to leverage skills in a dynamic data scientist role to contribute to data-driven decision-making processes.

- Programming Languages: SQL, PL/SQL, XML, JavaScript, R, Python, C, Ruby
- Operating Systems: Windows, CentOS, FreeBSD, Debian Linux
- Desktop / Server Software: Oracle, MSSQL, MySQL, MongoDB, Pandas, TensorFlow Keras, Git, Mercurial, Jupyter Notebooks, Power BI, Tableau, MS Excel, Azure Data Studio, SciKit Learn, NLTK, PySpark, Pandas

RELEVANT EXPERIENCE

Senior Web Application Developer / *Millennium Communications, Inc.*

Dec 2012 – July 2024

- Developed SQL pivots for extracting data in MySQL, MSSQL, and Oracle databases for extracting data for further cleansing and data analysis.
- Developed interfaces that communicated live statistical reports on highly trafficked web applications to stakeholders.
- Polled APIs for social networks to filter and import data into local databases for sentiment analysis utilizing Python. The networks included – Facebook, Twitter, Instagram, and YouTube.
- Configured websites for collecting data for A-B testing.
- Reviewed PHP, Python, and JavaScript code for quality assurance and security concerns with a focus on OWASP vulnerabilities.
- Merged Google Analytics data with custom user activity collected via websites to provide a more comprehensive view of the performance of marketing campaigns.

EDUCATION

Syracuse University, Syracuse, NY

Aug 2023 – Present

Master of Science, Applied Data Science

Anticipated graduation: May 2025

GPA: 3.77

Relevant Course Work: Machine Learning, Natural Language Processing, Quantitative Reasoning, Business Analytics, Database Management, Deep Learning Fundamentals, Big Data Analytics, Text Mining, Responsible AI, and Introduction to Data Science.

SUNY Institute of Technology, Utica, NY

Jan 2007 – Aug 2008

Bachelor of Science, Computer Information Systems

University of the Virgin Islands, St. Thomas, USVI

Aug 2004 – Dec 2006

Associate of Arts, Computer Information Systems

MERVIN S. MCDUGALL | DATA SCIENTIST

Baldwinsville, NY 13207 / mcdmerv@yahoo.com / (315) 749-5129 / <https://www.linkedin.com/in/mervin-mcdougall/>

PROJECTS

ASL Translator / <https://github.com/mervinmcdougall/asl-ai-translator/>

Project Date: 2024

An AI model built with Python, Tensorflow, Keras, MediaPipe, and Pandas that translates the American Sign Language Alphabet into text and speech

- Performed Exploratory Data Analysis of hand gesture images to determine what was usable for training data.
- Performed data transformation into hand landmark vectors through MediaPipe.
- Normalized and resampled data to reduce model bias during training
- Experimented and Trained ANN to identify hand gestures with up to 98% F1 accuracy
- Reviewed data to identify the cause of hand gestures that could not be translated.

Toxicity Detector / <https://github.com/mervinmcdougall/social-media-toxicity-detection>

Project Date: 2024

A novel approach to toxicity detection in social media posts that emphasizes the use of topic modeling and Toxic BERT for identifying toxic posts and the relationship between the level of toxicity and the topics in the post

- Imported data into Apache Spark for managing cleaning and preparing the data for further analysis
- Performed Exploratory Data Analysis on Reddit posts within dataset using NLTK, Spacy and Keras
- Performed topic modeling on various posts to gather insights within the posts and among posts
- Performed Toxic BERT analysis of the posts to identify the level of toxicity of posts and correlations between the metrics of the post and the toxicity score.

NBA Predictor / <https://github.com/mervinmcdougall/nba-game-predictor>

Project Date: 2024

An AI model built with Python, Tensorflow, Keras, and Pandas that utilizes data sets from multiple sources to predict the winner and spread for future basketball games

- Performed Exploratory Data Analysis of various Data Sets, identifying any correlations and possible sources of skewed data
- Performed data cleaning, and munging for import into the AI model
- Developed wide and deep ANN model for processing data inputs
- Tested different ANN activations to improve the performance and predictability of the model.

RSS Feed Tokenizer / <https://github.com/mervinmcdougall/rss-feed-tokenizer>

Project Date: 2024

A Python tool for amalgamating the content from various websites and tabulating the frequency distribution to discover the most frequently used words/terms across different sites and identify trending terms.

- Developed Python script for importing content from RSS feed.
- Built custom filters using stop words and regular expressions for filtering content
- Built code for developing a frequency distribution using NLTK
- Created code for presenting the tabulated results of tokenized data across all source feeds to identify patterns