

SKILLS

MACHINE LEARNING: Linear regression/classification, logistic regression/classification, SVM regression/classification,

Decision Tree regression/classification, KNN, Ensemble Models (boosting and bagging), K-Means Clustering, DBSCAN, PCA, topic modeling, TFIDF, NMF, LSA, LDA, cross validation, hyperparameter tuning, model evaluation, regularization, exploratory data analysis

DEEP LEARNING: Dense Neural Networks, Convolutional Neural Networks, Sequence Models (RNN, GRU, LSTM), data augmentation

VISUALIZATION: t-SNE plots, pair plots, 2D/3D scatter plots, histograms, bar plots, line plots, word clouds, decision tree graphs, neural network graphs

FRAMEWORKS AND PROGRAMMING LANGUAGES: python, numpy, pandas, matplotlib, seaborn, scikit learn, sklearn, NLTK, spaCy, tensorflow, keras, RAPIDS, C, C++, OpenCL, HTML, Flask, SQL, mongo DB, git, GitHub, BitBucket, BeautifulSoup, Selenium, PERL, bash, tcsh, ksh, performe, ClearCase, object-oriented programming

MISCELLANEOUS: Can build GPU accelerated machine learning workstation,

Linux: Can setup small scale services such as web services and database services, Markdown, Microsoft Office, Doxygen, Natural Docs, MathJax, emacs, Visual Studio, Xcode, Eclipse

EXPERIENCE

Metis, Data Scientist, New York, NY

Sept. 2020 - Current

Completed Metis's 12-week accredited data science bootcamp focused on Python programming, machine learning, statistical modeling, data visualization, project design, and communication. Designed, implemented, and presented the following projects (details provided in the projects section):

- Predicting Star Ratings for Movie Reviews Using a Multi-Category Classifier
- Customer Feedback Analysis with Topic Modeling
- Automatic Fraudulent Transaction Detector for Mobile Payment System
- IMDB Star Rating Predictor
- MTA Turnstiles Data Analysis

Octopus Design Automation, Founder, San Jose, CA

May 2019 - June 2020

- Devised, wrote the specification and initial revision of the claims for patent 10,699,046 B2, A Novel System and Method for Achieving Functional Coverage Closure for Electronic System Verification.
- Developed a data visualization web service using C++, mongo DB, and flask. The service enables chip design engineers to achieve functional coverage closure with few resources and greater velocity.

Advanced Micro Devices, Senior Member of Technical Staff, Santa Clara, CA

Mar. 2016 - May 2019

- Developed automated system to verify performance of memory controller IP for discrete GPUs and SoC's with python, csh, cron, HTML, and UVM SystemVerilog. This enabled automatic tracking of configurable memory controller IP performance for multiple projects running simultaneously across multiple sites using minimal resources.
- Lead a cross-site team to perform performance verification of the GDDR and HBM DRAM interfaces of high-performance GPUs and SoCs.

Samsung Research America, Senior Staff Verification Engineer 1, Mountain View, CA

Apr. 2014 - Mar. 2016

- Developed from scratch the UVM SystemVerilog testbench, wrote the test plan, and executed verification of the GPU shader core and warp scheduler, and shader constructors.
- Championed and sold the team on using industry standard UVM testbench architecture to verify the various blocks within the GPU and enabled reuse by development of in-house verification IP (VIP).
- Fixed and enhanced Python programs that translate high-level architectural descriptions of the GPU to SystemVerilog.
- Served as technical lead for GPU pipeline control verification team: Trained teammates and management on the verification environment, UVM SystemVerilog and how it's applied to the project, and GPU design information (microarchitecture, architecture), held weekly cross functional meetings with design and architecture teams to resolve issues and engaged in process improvement for the project (verification methodology improvements, CAD environment improvements).

PROJECTS

Predicting Star Ratings for Movie Reviews Using a Multi-Category Classifier

Developed a bidirectional LSTM with pretrained GloVe word embedding using keras that classified a movie review using the Net Promoter Score (NPS). The model was trained on movie reviews scraped from imdb.com using request, Selenium, and BeautifulSoup. The model was deployed as a Flask web application that accepts the movie title and the free-form text review of the movie and then sends back to the user a predicted NPS. It is able to capture these reviews and the users desired score into a database that could then be repurposed to further train the model.

Customer Feedback Analysis with Topic Modeling

Developed a natural language processing (NLP) pipeline that processed over 500,000 free-form text product reviews from Amazon. Enabled product developers to quickly acquire actionable information about their products. It identified specific themes within the product reviews with little human intervention. The pipeline performed topic modeling via TFIDF document vectorization, text processing using regex, lemmatization using NLTK, and Latent Dirichlet Allocation using sklearn. Visualizations were developed using seaborn, matplotlib, and pyLDAvis.

Automatic Fraudulent Transaction Detector for Mobile Payment System

Built a mobile fraudulent transaction detection web application using Flask, RAPIDS, sklearn, XGBoost, pandas, numpy, matplotlib, and MariaDB SQL.

Via EDA engineered a new feature from a highly imbalanced dataset of over 6 million transactions (8000 of which were fraudulent), which enabled the XGBoost classifier to obtain an F1 score of 0.97 with a precision of 0.99 and a recall of 0.97 for the fraudulent class, and 1.00 precision and recall, and F1 score for the non-fraudulent class.

IMDB Star Rating Predictor

Developed a linear regression model using sklearn to predict the star rating of a movie based only on data that would be available about said movie on opening day. The project involved acquiring the data set via web scraping 4000 movies released over a 10-year period from imdb.com using BeautifulSoup. Data cleaning and exploratory data analysis (EDA) were performed using numpy, pandas, seaborn, and matplotlib. Engineered a feature from the data set to improve model performance using Academy Awards history for movies and actors appearing within the data set.

MTA Turnstiles Data Analysis

Working on a team of 4 data scientists, analyzed 3 months of turnstile traffic data within the city of New York using numpy, pandas, and matplotlib. We identified optimal turnstile locations and specific points in time at which a non-profit organization should place personnel to maximize outreach for a future event.

For each of the projects above, presented results including a detailed discussion of the model selection and evaluation process to a team of data scientists. Documented the project using markdown and managed development of project collateral using GitHub.

Automatic Maze Traversal Using Q-Learning

Aug. 2017 - Aug. 2017

Developed using Python a virtual robot under the constraints of the Micromouse challenge that automatically traverses various mazes based on the Q-learning algorithm. The trained Q-learning agent was able to reliably traverse several mazes including mazes upon which it was not trained in less time than the baseline maze traversal agent.

Visualization of the virtual robot was developed using turtle graphics.

EDUCATION

University of Michigan, Ann Arbor

MSE Electrical Engineering

Kettering University

BS Electrical Engineering

MICHAEL GREEN

DATA SCIENTIST / MACHINE LEARNING ENGINEER

✉ green.michael.alexander@gmail.com ☎ 408-332-8248 📍 San Jose, CA 95123 **in** <https://www.linkedin.com/in/magreen> 🌐 michael-a-green