Elizabeth Campau

West Olive, MI | +1 (616) 990-8149 | campauel@mail.gvsu.edu

AI and Machine Learning Projects

DOCUMENT CLASSIFICATION | NAÏVE BAYES | PYTHON

- The purpose of this project was to create a classifier that labels documents with their main topic. I developed a Naïve Bayes classifier based on main topic probabilities and within-topic word probabilities. In order to further increase the accuracy of the classifier, I created a word utility measure, based on the ratio of a word's maximum within-topic probability to its minimum within-topic probability. Words with insufficient max/min within topic probabilities were removed from analysis, resulting in ~2% increase in accuracy above the basic Naïve Bayes algorithm.

HAND-WRITTEN DIGIT IDENTIFICATION | NEURAL NETWORK | PYTHON

- The purpose of this project was to train a neural network to recognize handwritten digits in compressed images. The function I wrote created a single hidden layer neural network, using a sigmoid function for hidden layer activation and softmax for output nodes. Training continued until validation error increased, calculates as Sum of Squared Error. A neural network created by this function identified digits in test data with ~94% accuracy.

TRAINING 'BOT' PLAYER | Q-LEARNING | PYTHON

- The goal of this project was to train a bot representing the second tic-tac-toe player. The bot was trained against a randomly moving first player using the Q-learning algorithm. After training, the bot won 92% of the time and tied 8% against a randomly moving first player.

CLASSIFICATION | DEEP LEARNING | R

- The purpose of this project was to predict obesity from demographic data and walking data, which consisted of position and rotation values in the x, y, and z directions over a very short period of walking (~40 sec) on a treadmill. I used the R package `keras` to create a multi-layered neural network model.

CLASSIFICATION | DECISION TREE | PYTHON

- The purpose of this project was to create a function which could create a decision tree to classify any dataset. Data had to be in the correct form, with columns representing attributes and rows instances, and the last column representing the categorical target variable. After training, the tree was returned and could be used for classification.

FINDING BEST CLASSIFIER | VARIOUS | R

- The purpose of this project was to predict survey response from demographic information, and to find the best classifier with which to do so. Of decision tree, Naïve Bayes, K-nearest neighbor, and support vector machines, implemented using already built functions from common R packages, a decision tree performed the best.

FINDING BEST CLASSIFIER | VARIOUS | R

- The purpose of this project was to predict the presence of disease in patients and to find the best classifier with which to do so. Data attributes consisted of five numeric medical tests and demographic

information. Of decision tree, neural network, support vector machines, and Naïve Bayes, implemented using already built function from common R packages, a decision tree performed the best.

FINDING BEST CLASSIFIER | VARIOUS | R

- The purpose of this project was to classify individuals into 3 or 7 obesity levels based on survey data collected from Mexico, Peru, and Columbia about eating and lifestyle habits, and to find the best classifier with which to do so. Of neural networks, Naïve Bayes, and decision trees, artificial neural networks performed the best for both 7 and 3 levels of obesity, achieving 73.4% and 80.8%, respectively, compared to 59.5% and 68.8% of Naïve Bayes and 63% and 72.7% of a decision tree.

Education

P.S.M | DEGREE ANTICIPATED AUGUST 2022 | GRAND VALLEY STATE UNIVERSITY

- Data Science and Analytics
- Current GPA 3.95

B.S | 2020 | GRAND VALLEY STATE UNIVERSITY

- Major: Statistics
- Minor: Philosophy
- GPA 3.8

Skills & Abilities

LANGUAGES

- Python
- C++
- R
- MySQL
- SAS

OTHER PROGRAMMING ABILITIES

- High-Performance Computing
- Machine Learning
    - Deep Learning

STATISTICAL/MATHEMATICAL METHODS

- Multiple Linear Regression
- Logistic Regression
- Ridge Regression
- LASSO
- ANOVA/MANOVA/ANCOVA
- Strong Mathematical background, including:

- Mathematical foundations of Machine Learning algorithms
- Linear algebra
- Calculus through Calc 3

Interests

- Biking, hiking, reading, writing poetry, and learning languages (human and computer)