Title: **Medical Malpractice Lawsuit Predictor**

Description: Using National Practitioner Data Bank data to train linear, tree-based, and neural net models which predict the outcome of a medical malpractice lawsuit (Scikit-Learn, Apache Spark, Seaborn).

Title: **Deep Learning Image Classifier**

Description: Building and tuning two neural net architectures (feed forward and convolutional) for image classification (Pytorch).

Title: **Fake News Detector**

Description: Building an NLP binary classifier that predicts whether a news article is fake (TensorFlow, R).

Title: **Unsupervised Classification of Cancer Subtypes**

Description: Using principal component analysis and K-means clustering to classify breast cancer subtype based on genomic sequencing data from The Cancer Genome Atlas (R).

Title: **Social Media Database**

Description: Designing a Boyce-Codd normal form relational database and backend architecture of a social media application (MySQL, AWS, JavaScript, Node.Js).

Title: **PGE Stock Forecast**

Description: Fitting an ARMA-GARCH time series model to forecast return on Pacific Gas and Electric (PGE) common stock based on stock prices data between 1998 to 2021 (R).

Title: **Hard Margin SVM**

Description: Building a hard margin support vector machine (version h2) that optimizes the distance between SVM hyperplane and binary classified points (MATLAB).

Title: **U.S. Fatal Crash Locations Clustering**

Description: Using the HDBSCAN algorithm on NHTSA's location data of fatal heavy duty truck crashes to find clusters of geographical areas especially prone to fatalities between 2013 and 2023 (ArcGIS).

Title: **Sudoku Solver AI**

Description: Using AC-3 constraint satisfaction algorithm to create an efficient solver for high difficulty sudoku puzzles (Python).

Title: **Educational Factors and Economic Growth**

Description: Using World Bank Educational Data to train linear, random forest, Ada boost, and KNN models and perform causal inference on impact of a country's educational variables on economic growth (Scikit-Learn, Seaborn).

Title: **COVID-19 Mortality Factors Study**

Description: Using LASSO regression to identify risk factors contributing to mortality in COVID-19 deaths (R).

Title: **Bezier Curves**

Description: Drawing Bezier curves with deCasteljau subdivision, de-Boors control points, and curve interpolation (MATLAB).

Title: **Gridworld AI**

Description: Implementing a policy and value iteration agent that uses Markov decision processes to efficiently solve Gridworld path searches (Python).

Title: **Time Series Modeling of U.S.-China Imports**

Description: Using seasonal ARIMA approach to model the dollar value of Chinese imports to the U.S. from 1989 to 2019 (R).

Title: **Yelp Review Sentiment Analysis**

Description: Applying random forest, logistic regression, LASSO, and an ensemble method to perform sentiment analysis on Yelp review data and build classifiers predicting rating based on words (Keras, R).

Title: **Lady Pinkham Advertising Impact Analysis**

Description: Constructing a linear model comparing sales and advertisement with sales as the target variable, using Lady Pinkham sales data between 1938 and 1952 (R).

Title: **Youtube Views Regressor**

Description: Using dimensionality reduction and random forest models to predict the success of youtube videos (Scikit-Learn, Apache Spark, Seaborn).

Title: **MATLAB Ridge Regression**

Description: Implementing multiple approaches to fit ridge regression models to datasets (MATLAB).

Title: **U.S. Grocery Store Sales Forecast**

Description: A linear, multiplicative decomposition model forecasting U.S. grocery store sales based on time series sales data between 1992-2021 (R).

Title: **Informed Search Solvers**

Description: Implementing iterative deepening depth-first search (ID-DFS) and A* algorithms to solve for several puzzle cases (Python).

Title: **Uninformed Search Solvers**

Description: Implementing breadth-first-search and depth-first-search algorithms to solve for several puzzle cases (Python).

Title: **Dominoes Adversarial Agent**

Description: Implementing adversarial search algorithm with alpha-beta pruning to create an AI which plays dominoes (Python).

Title: **Diabetic Patients Readmission Predictor**

Description: Using LASSO regression to predict the 30-day readmission of diabetic patients based on patient characteristics (R).

Title: **U.S. Used Car Sales Market Analysis**

Description: A linear, multiplicative decomposition modeling of U.S. used car sales based on 1992-2021 data with auto-correlation and dynamic seasonality presence (R).

Title: **Pacman Reinforcement Learning Agent**

Description: Implementing a Q-learning agent that uses epsilon-greedy action selection to play Pacman (Python).

Title: **PennFlix MovieApp**

Description: Implementing a BCNF relational database, entity relationship diagram, and backend architecture for a streaming app (MySQL, AWS, Node.Js).

Title: **Perceptrons**

Description: Building binary and multiclass perceptrons from the ground up to create classification systems (Python).

Title: **Women in Data Science**

Description: Analyzing National Science Foundation data of science and engineering degrees awarded between 2006-2016 to estimate the representation of women in the field of data science (R).