Portfolio Project Choices

David D Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

September 22, 2024

Portfolio Project Choices

For my portfolio, I have chosen ten of my previous projects from the Master’s curriculum courses. The selected projects come from four previous courses: Data Mining, Data Preparation, Predictive Analytics, and Data Visualization and Presentation. I will list each project with a brief description below.

1. Restaurant Chains EDA, Data Mining, and Model Performance – this DSC 550 term project showcases the top 100 restaurant chains in the United States in 2020 and the insights that can be gained from inquiring about certain valuable tidbits of information such as what restaurant sector was the most lucrative, the relationship between menu category and total sales, and more. A thorough exploratory data analysis was performed along with several data transformations and the creation of a linear regression model to understand the significance of certain factors like change in units and change in sales as they relate to total revenue.
2. Web Scraping and Transformation of Pokemon Data – this DSC 540 project milestone highlights my ability to scrape data from a web page via Python’s bs4 library function known as BeautifulSoup, perform various transformations on the acquired data, and output a human-readable dataset after parsing the typically hard-to-read underlying HTML structure.
3. API Calling and Transformation of Pokemon Data – this DSC 540 project milestone displays my capability to establish a secure connection to a REST API using Python’s request library, parse data from the underlying JSON data structure, perform various transformations on the acquired data, and output a human-readable dataset after using the most Python data manipulation libraries, such as Pandas and NumPy.
4. Best Model Selection Using Hyperparameters – this DSC 550 assignment allows a potential employer to see how I address a classification problem using many different classification models such as K-Nearest Neighbors, logistic regression, and random forest algorithms. The best-performing model for the data is chosen using hyperparameter tuning techniques such as scaling the data using a min-max scaler and creating a grid search with a pipeline, a search space, and cross-validation folds.
5. Neural Network Image Classifier – this DSC 550 assignment depicts the creation of a convolutional neural network to recognize and classify pictures of written numbers as the correct number the picture is representing. Neural networks are some of the most involved concepts of data science, so this experience in crafting them and verifying their performance can have an employer see potential in me with my work.
6. K-Means Clustering and PCA Analysis – this DSC 630 assignment shares my experience in crafting an unsupervised machine learning algorithm to classify patients as having or not having ALS based on the variables within the data. Dimensionality reduction has also been taken into account within the assignment to omit highly correlated features while retaining the most amount of variance within the data as possible.
7. Movie Recommendation System – this DSC 630 assignment puts forth my ability to create a movie recommendation system using cosine similarity scores and combines that with the creativity to implement a user-friendly program in which they choose which movie they would like recommendations based on similarity. The knowledge and expertise in these projects are paramount to interest potential employers in your skillset to eventually get hired.
8. Airline Safety Blog Post – this DSC 640 project milestone provides the opportunity to see my ability to present a real-world topic to an audience and tell the data’s story relating to the issue of airline safety. Writing a blog post that uses a combination of eye-catching and effective visual aids and swift yet powerful wordplay to support the position that the airline industry is being sabotaged by the media is a unique avenue for a potential employer may be interested in. The power of data visualization and communication as hard and soft skills respectively is relayed well with this project.
9. Airline Safety Infographic Presentation – this DSC 640 final project milestone solidifies my ability to craft informative and attention-grabbing visualizations, communicate effectively by telling the true story of the data relating to airline safety, and put together a presentation that can be understood and appreciated at almost all professional or personal levels. My experience with visualization tools like PowerBI, PowerPoint, and Canva are also at work in this milestone, prompting a potential employer to inquire more about my skillset.
10. Predicting Monthly Revenue in the Restaurant Industry – this DSC 630 term project is the best example I have of executing a full-scale data science project, from acquiring the data and mining it to crafting multiple models of varying types and outputting their statistics. My presentation on predicting monthly restaurant revenue is the culmination of everything taught so far here at Bellevue University.