Jarrale Butts

Southern New Hampshire University

Enhancement Two Narrative: Databases

What is the Artifact:

The project I created is a data mining enhancement that analyzes animal outcome data. This enhancement was developed to add functionality to my MERN (MongoDB, Express.js, React, Node.js) full-stack web application. The primary focus of the project is to implement a Decision Tree Classifier to predict the outcome types of animals based on various features, such as breed, age, and sex upon outcome. I utilized Pandas, Scikit-learn, and PyMongo libraries to facilitate data manipulation, model training, and evaluation.

Why Include in ePortfolio:

I selected this for inclusion in my ePortfolio because it effectively demonstrates my software development and problem-solving skills. The enhancement showcases my ability to connect to a MongoDB database, perform data preprocessing, and build a predictive model using a Decision Tree Classifier. Specifically, the implementation of Label Encoding for categorical variables, data cleaning by handling missing values, and the evaluation of the model using confusion matrices and classification reports highlight my technical proficiency. Throughout the project, I also made improvements by enhancing the data preprocessing steps and fine-tuning the model, which helped achieve better accuracy and reliability. The enhancement is connected to the same MongoDB Atlas database as the MERN web app.

Course Outcomes:

The Data Mining enhancement in the Databases category has successfully achieved two course outcomes. I have successfully included a data mining feature allowing users to gain

deeper insights and identify patterns in the data, thus demonstrating an ability to use well-founded and innovative techniques, skills, and tools in computing practices for the purpose of implementing computer solutions that deliver value and accomplish industry-specific goals. While also requiring a password to access the cloud-hosted database, thus developing a security mindset that anticipates adversarial exploits in software architecture and designs to expose potential vulnerabilities, mitigate design flaws, and ensure privacy and enhanced security of data and resources. My outcome coverage plans remain the same and are on track.

Reflect:

As I worked on the project, I learned how to effectively handle missing data and the impact of feature selection on model performance. One significant challenge I faced was achieving a balanced model, as certain outcome types were underrepresented in the dataset, which led to poor classification results. Moving forward I plan on refining the feature selection to improve model performance.