

## Data Science Capstone Topic Approval Form

**Student Name:** Eric Williams

**Student ID:** 001750584

**Capstone Project Name:** Using Machine Learning to Predict Car Value

**Project Topic:** Machine Learning (Random Forests) to predict car value

☒ **This project does not involve human subjects research and is exempt from WGU IRB review.**

**Research Question:** What factors influence the value of used cars, and can our Random Forests model accurately predict car prices based on their features?

**Hypothesis:** There is a significant relationship between the features of a car and its market value.

**Null Hypothesis:** There is no significant relationship between the features of a car (brand, model year, mileage, fuel type, and engine type) and its market value.

**Alternate Hypothesis:** There is a significant relationship between the features of a car (brand, model year, mileage, fuel type, and engine type) and its market value.

**Context:** If we can create an automated model to predict car values based on the input variables, we can take a lot of the guess work and human error when determining car values. This can then save the used car company a lot of money by making sure cars are accurately priced while simultaneously ensuring the prices are competitive in the market.

**Data:** I will not need to collect data for this project as I am using the supplied data from WGU.

The dataset provides 4,009 cars with a specific brand, model, year, milage, fuel type, engine type, interior and exterior colors, title type, and price.

**Data Gathering:** As the data was provided by the school, I will not need to gather any data.

**Data Analytics Tools and Techniques:** To predict car prices, I will use the Random Forest classification tool, utilizing decision trees.

**Justification of Tools/Techniques:** To predict price, this will essentially be a regression task which Random Forest is great at analyzing. I will also use a test/train split to partition the data to create a trained model to predict the prices. Between those two tools, I will be able to create a prediction model for the prices based on the many variables that might contribute to the overall value of the car.

**Project Outcomes:** I predict that the model I create will use the random forest classification tool to accurately predict the value of a car. The value of a used car can be determined by model, brand, year, milage, and other specifications and my model should be able to predict that value by analyzing those variables.

**Projected Project End Date:** 01/20/2025

**Sources:** No sources were used besides official WGU materials, including the dataset.

**Instructor Signature/Date:**

☒ The research is exempt from an IRB Review.

☐ An IRB approval is in place (provide proof in appendix B).

Instructor's Approval Status: **Approved**

Date: **January 7, 2025**

Reviewed by: **Daniel J. Smith, PhD, MBA**

Comments: [Click here to enter text.](#)

