|  |
| --- |
| Western Governors University |
| Advanced Data Analytics |
| D214: CAPSTONE EXECUTIVE SUMMARY |

|  |
| --- |
| Allison Casey  3-20-2025 |

Since cars depreciate over time, it is common for consumers to buy used cars and dealers to sell used vehicles. As a result, it is beneficial to buyers and sellers of used cars to accurately determine the value of a used car. The model's accuracy was determined based on the R² value being 0.8 or the model explaining 80% of the variance in the test data. This study utilized a predictive neural network to see if this model can be used to predict this value accurately. The null hypothesis was that a model with greater than 80% accuracy could not be created, and the alternate hypothesis was that it could. The data was a public dataset from Kaggle. It was extracted and explored to check for null values, duplicates, and other issues before being used for the model using the Python Pandas library and its functions. One hot encoding and scaling techniques were also used to prepare the data for analysis. Finally, the data was split into training and test sets using the train\_test\_split method to prevent overfitting and test the model against data it was not trained against. A Keras Functional API model was created to analyze the data. This neural network model can handle models with non-linear topology, shared layers, and even multiple inputs or outputs (Keras, 2025). This model is ideal for this scenario because of its flexibility, handling of various inputs, and complex relationships likely present within this data set. The model was then fit to the data using an early stopping monitor. The benefit of this is that the model will stop training when the performance of the model stops improving, which will prevent overfitting and save computational resources. Once the model was trained using the prepared data, it was evaluated using the test data. Since regression models don't have the accuracy, in the same way, a classification model does the accuracy, in this case, is defined by how well the model explains the variance in the target variable or the R² value. The R² score was calculated as .9998 using the test data. So, the model is approximately 99% accurate. The results show that the null hypothesis can be rejected, and a model with greater than 80% accuracy can be created. The model appears to have performed well, although some limitations should be noted for the tools and techniques used. One disadvantage of one hot encoding is that it increases the dimensionality of the data, which, in this case, could result in overfitting and greater computational resource consumption. A disadvantage of scaling would be that it can make the feature values and model output less interpretable than their original values. Splitting the data can be a disadvantage because it limits the testing and training data. The disadvantage of using a Keras Functional API model is that it is more complicated to interpret than other methods, such as linear regression. The disadvantage of using an early stopping monitor is that the training could be stopped prematurely. The biggest concern with the model is that it is too overfit to this particular data set due to the very high R² value, but this would need to be confirmed with further data and testing. The recommended course of action would be for the model to start to be used while also using whatever previous methods were being used for evaluation to continue to evaluate the model and determine if it is genuinely accurate or overfit to the data set it was trained with. With further assessments like this and further evaluations with more data gathered, the model can be used even more reliably. Along these lines, one direction for future study of the data could be to continue to gather more data to test and potentially re-train the model. Another approach for future data research could be to create other models, such as a random forest or a linear regression model, to predict the price and compare it to this model. The expected benefits of the study would be the ability to accurately predict car values for buyers and sellers to maximize profit for sellers and help buyers not spend too much on a used car.

**Sources**

Elleh, Festus. “Advanced Data Analytics - Task 1.” D213 Task 2 Cohort Webinar. 2025.

“ReLU Activation Function in Deep Learning.” *GeeksforGeeks*, GeeksforGeeks, 29 Jan. 2025, www.geeksforgeeks.org/relu-activation-function-in-deep-learning/.

Team, Keras. “Keras Documentation: Earlystopping.” *Keras*, keras.io/api/callbacks/early\_stopping/. Accessed 11 Mar. 2025.

Team, Keras. “Keras Documentation: The Functional Api.” *Keras*, keras.io/guides/functional\_api/#training-evaluation-and-inference. Accessed 11 Mar. 2025.