SHAP Interpretation for Spaceship Titanic Prediction Model

Model Context

A CatBoostClassifier was trained to predict whether passengers aboard the Spaceship Titanic of this Kaggle Dataset would be transported to an alternate dimension. It uses minor manually engineered features, but highly relies on CatBoost's ability to capture accurately non-linear relationships between features. An API (check the repository) was also deployed to allow other users to test the model.

Global Importance

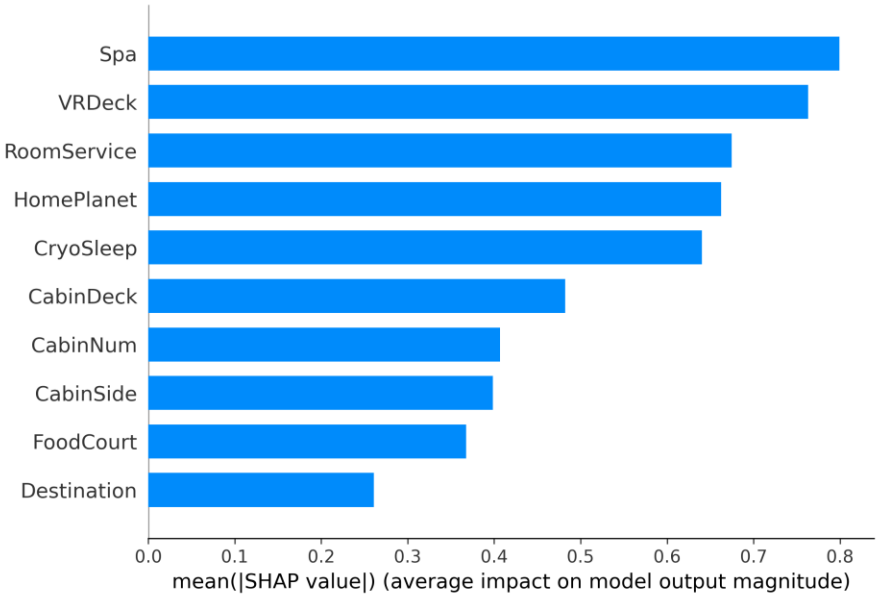

Figure 1: SHAP Bar Plot comparing the Average Contribution by the Top 10 Features

The above plot shows that features related to a passenger's spending (Spa, VRDeck and RoomService) within the spaceship have the greatest contribution when predicting who were transported to an alternate dimension. Two categorical features, HomePlanet (the planet the passenger originates from) and CryoSleep (whether the passenger is in an unconscious sleep) closely follow, but are not as powerful. These features are likely to be the major driving factors because they give insight to the passenger's lifestyle, and financial status.

Feature Effect Analysis

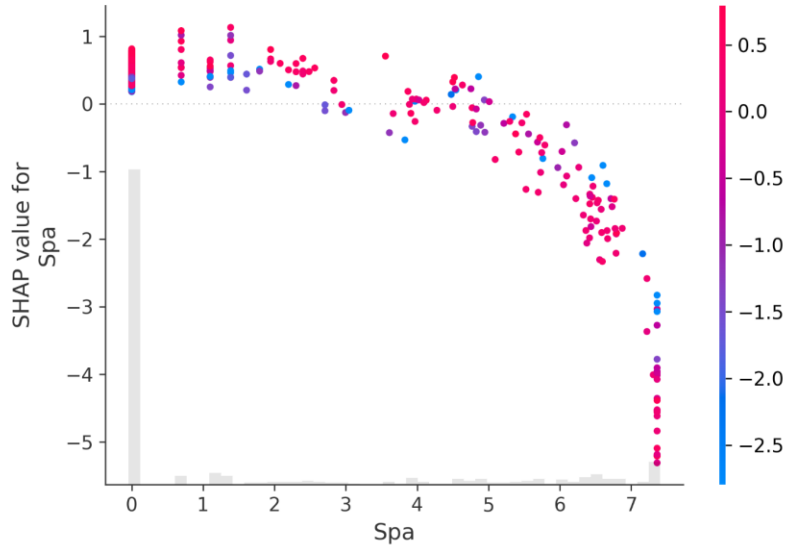

Figure 2: SHAP Dependence comparing the SHAP and Raw Values of the Spa Feature

Figure 2 shows that the SHAP and raw values of Spa are not positively correlated. This means that as the spending on Spa increases, the SHAP value (average contribution to prediction) decreases at a non-linear rate. The reason behind this is that less people spend more on the spa, so there are less data points in high values of Spa, because most passengers see no value in it.

Local Explanation

The SHAP force plot shows a single prediction from the model, and how each feature influences the model's prediction.

It shows that this passenger has characteristics that push the model to predict that they were transported, with the opposite also being true. The blue side features, pushing the prediction to "Not Transported", dominate overall, so the final prediction is below the baseline. For example, this passenger has an Age of 6, is in a group of 3, and spent no money on Spa, VRDeck or RoomService. The model has determined that these features push the prediction towards "Transported". However, this passenger's HomePlanet of Earth, Destination of TRAPPIST-1e, cabin location, and not being in cryo sleep all push the prediction downward to "Not Transported". This combination results in the model predicting a lower likelihood of being transported for this passenger.

Conclusion

In conclusion, the model's SHAP validation shows that the model places a high weight on spending features. While some SHAP values of features are difficult to explain at first, they can be deciphered with a little thought. The model's lack of reliance for manually feature engineered interactions while making accurate predictions, demonstrate the model's ability to capture non-linear feature interactions on its own.