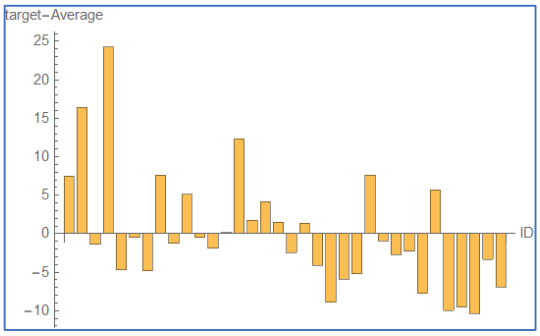
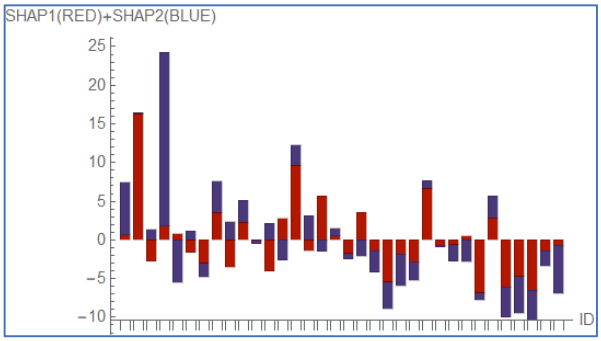
Theoretical Explanation and Case Studies of Shapley Values in Machine Learning Regression

Prof. Dr. Yukari Shirota, Gakushuin University, Japan
Prof. Dr. Basabi Chakraborty, Dean and Distinguished Professor, School of Computing, Madanapalle Institute of Technology and Science (India), and Professor Emeritus of Faculty of Software and Information Science, Iwate Prefectural University (Japan)

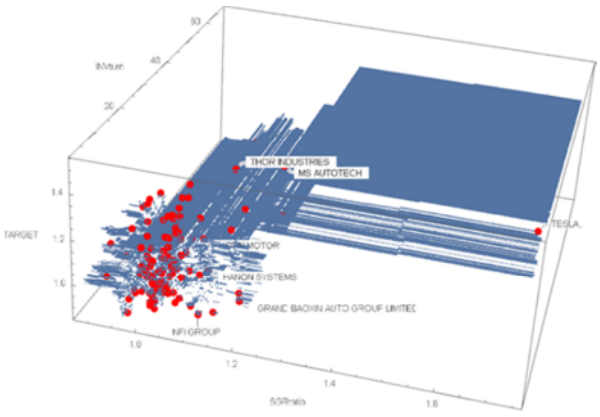
Lundberg's SHAP and its libraries have been widely used in machine learning regression analysis and have contributed significantly to the semantic interpretation of regression results. In this tutorial, Shapley's formula, the original theory behind SHAP, will be visually explained using graphical materials. In the regression analysis, Shapley's formula gives the unique solution how each explanatory variable contributes to the target value of the sample data among N explanatory variables, considering each data's characteristics. For example, in the following sample of 20 data with two explanatory variables, the right figure shows the deviation of target values. The right figure illustrates the individual data's two SHAP values; a red part is the first variable's SHAP value and a blue part is the second one. The advantage is that each variable's contribution can be evaluated considering each data's characteristics.





A SHAP value means a contribution of each predictor to the target.

Shapley's formula is such a breakthrough that it is a useful tool in the field of data engineering.



The presenters will explain the SHAP-based approach visually using the regression models and its heatmaps.

From the medical field to the economics field, in every field, SHAP-based approach can be applied. In Operational Management field, we obtained Decisions Sciences Institute (DSI) Best Paper Award in P&OM Nara 2020, by the SHAP-based analysis. Using the resultant data, the way of finding an invisible relationship will be illustrated.

The lecture will take 60 minutes and the main target audience is data analysts using regressions.