



FLORIDA ATLANTIC UNIVERSITY

DATA SCIENCE, ANALYTICS, AND ARTIFICIAL INTELLIGENCE CONFERENCE

SATURDAY
NOVEMBER 14, 2020

FAU.EDU/DATA

BREAKOUT SESSION PRESENTATION DESCRIPTION

CHARLES E. SCHMIDT COLLEGE OF SCIENCE

fau.edu/data

Opening the Black Box

Fowler, Ben; Chelsey Meise

Providing a method to interpret and explain how an advanced machine learning model is performing is essential for both modelers and business users. Modelers need to have confidence that a model is going to generalize to unseen data making sensible predictions. Business users need to be able to have insight into how the model will perform when feature values change.

In a regulated industry such as finance, it's imperative to be able to have reproducible and explainable models. Not only do we have to tell consumers why loans were rejected, we then have to explain to auditors why this decision was not based on a protected class. Then, this model must be explained to a CFO, in which case explainability is of great concern.

This presentation is an applied presentation intended for data scientists building models. While this presentation focuses on an application to finance, the methodologies can be applied to any situation in which a machine learning model is being placed into production.

SHapley Additive exPlanations are the foundation of the SHAP library, which provides a method to easily interpret machine learning models using a model-agnostic approach. We'll dig into the math behind SHAP and plots that you'll want to use to interpret your model on both a global and local scale. Additionally, we'll discuss the ethical responsibility data scientists have in not building machine learning models that have a bias that impact certain demographic groups. This discussion of bias will show SHAP plots demonstrating what can happen to your model if you include a protected feature and how you can use interpretability techniques to ensure that your model is free from bias.

As we compare the behavior of different models, we'll do so using multiple feature selection techniques on an open-source financial dataset. The experimental approaches for feature selection will use the CatBoost machine learning algorithm contrasting model metrics from the following approaches:

- Prediction values change
- Loss function change
- Permutation importance
- SHAP values

Finally, we'll conclude the end-to-end example presenting how business users can engage with the model using Streamlit. We'll provide an example UI showing the predictions from a model on new data with local reason codes from SHAP interpreting how the model arrived at the prediction it did. Furthermore, our UI will provide insights on a global scale showing the most important features in the model along with a geographic map highlighting where defaults are most prevalent.