Model Explainability with SHAP in DataRobot

Mark Romanowsky, Data Scientist - Explainable AI
Rajiv Shah, Data Scientist - Customer Success
Agenda

- Overview of Shapley values and SHAP
- An example use case: housing prices
- SHAP in the DataRobot UI
- Example use case: what did we learn?
- SHAP in the DataRobot API client
- Advanced topics
Overview of Shapley Values and SHAP
Background: Shapley values and machine learning

Lloyd Shapley: “how should we divide a payout among a cooperating team whose members made different contributions?”

Key insight:
- Shapley value for member X is the amount of credit they get.
- Total payout is the sum of Shapley values over members.
- To compute: For every possible subteam, how much marginal value does member X add when they join the subteam? Shapley value is the weighted mean of this marginal value.
Background: Shapley values and machine learning

Lloyd Shapley: "how should we divide a payout among a cooperating team whose members made different contributions?"

Key insight:
- Shapley value for member X is the amount of credit they get.
- Total payout is the sum of Shapley values over members.
- To compute: For every possible subteam, how much marginal value does member X add when they join the subteam? Shapley value is the weighted mean of this marginal value.

Machine learning researchers: "How should we divide credit for a prediction from a model whose features made different contributions?"

- Shapley value for feature X is the amount of credit it gets.
- Total prediction is the sum of Shapley values over features (plus the model baseline).
- Linear case is intuitive and simple:
  \[
  \text{shapley_value}(X_i=x) = \text{coef}[i] \times (x - \text{mean}(X_i))
  \]
- General computation is lengthy...
Background: Shapley values and SHAP

SHAP is an open-source library implementing Shapley values for ML.

Key contributions:
● An efficient algorithm to calculate exact Shapley values for tree ensembles (TreeSHAP).
● Fast and good approximate algorithm for deep learning (DeepSHAP).
● Model-agnostic, but slower and approximate, algorithm for any model (KernelSHAP).

Now used within DataRobot!
● Works with linear, tree, and deep learning models, and multi-stage combinations of these
● Whole blueprints, not just components

https://github.com/slundberg/shap
Why use SHAP?

Benefits:

- It's fast for many blueprints, including complex tree ensembles and deep learning networks.
- Additive explanations are concrete and intuitive.
- It's open source, so auditors can go deep.
Why use SHAP?

Benefits:
- It's fast for many blueprints, including complex tree ensembles and deep learning networks.
- Additive explanations are concrete and intuitive.
- It's open source, so auditors can go deep.

Limitations:
- Some AutoML blueprints don't support it (but most do).
- Times series and multiclass projects don't support it.
- Additivity is a little complicated in binary classification problems.
Example use case: housing prices

Dataset
Beijing_housing.csv (from Kaggle open datasets)

Target
TotalPrice (x 10,000 yuan)

Model
Light Gradient Boosting on ElasticNet Predictions

Questions to consider
● What will a property sell for?
● Which features drove that prediction the most?
● Which features, if changed, could change the outcome for a given case?
● Which features matter the most in general?
SHAP in the DataRobot UI

Tour based on our use case
SHAP prediction explanations in the UI

High predicted sale price -- why?
- Sale occurred in 2017 (prices rise over time through inflation)
- 110 square meters (big for this dataset)
- Average prices in the surrounding community (high)
SHAP explanations in the UI

- Sale year, size, surrounding prices
- Number of "living room"
- Stairs ratio

- Sale year, size, surrounding prices
- Sale month
- Floor

- Sale year, size, surrounding prices
- Renovation condition
- Latitude (location)

- Sale year, size, surrounding prices
- Stairs ratio
- Latitude (location)
Calculate SHAP for all validation rows
How is SHAP different from XEMP?

**XEMP prediction explanations**

- The existing standard in DataRobot.
- Model agnostic, works for all types.
- Based on a local partial dependence plot, focusing on "exemplar" values in each feature.
- Scores indicate the relative effect of each features on the prediction.

**SHAP prediction explanations**

- Now available in version 6.1
- Works with most top-scoring models.
- Computes Shapley values based on the corresponding algorithm, e.g. TreeSHAP, DeepSHAP, LinearSHAP.
- Scores sum to difference between prediction and the baseline.
SHAP Impact: aggregated SHAP explanations

\[ \text{shap\_impact}[j] = \text{mean}(\text{abs}(\text{shap\_values}[:, j])) \]
How is SHAP different from Permutation?

**Permutation Feature Impact:**
- The existing standard for DataRobot.
- Model agnostic, supported for all models.
- Measures how model loss is affected by “scrambling” each feature.
- This metric involves knowing “ground truth” of your data.

**SHAP Feature Impact:**
- Now available in version 6.1.
- Supported for trees, linear, deep learning models.
- Measures how much a feature tends to move predictions away from baseline.
- No reference to “ground truth” of training data.
Prediction explanations in deployments

Integration Scoring Code

Use this Python script to integrate DataRobot predictions into your current python project.

```python
...
Usage:
    python datarobot-predict.py <input-file.csv>

This example uses the requests library which you can install with:
    pip install requests
We highly recommend that you update SSL certificates with:
    pip install -U urllib3[secure] certifi
...
import sys
import json
import requests

AP2_URL = 'https://mlops-dev.dynamic.orm.datarobot.com/predApi/v1.0/deployments/(deployment_id)/predictions'
AP2_KEY =
DATAROBOT_KEY =
DEPLOYMENT_ID =
MAX_PREDICTIONS_FILE_SIZE_BYTES = 52428800 # 50 MB
```
Prediction explanations in deployments

```python
def make_datarobot_deployment_predictions(data, deployment_id):
    """
    Make predictions on data provided using DataRobot deployment_id provided.
    See docs for details:
    
    Parameters
    -----------
    data : str
        Feature1,Feature2
    numeric_value : str
    deployment_id : str
        The ID of the deployment to make predictions with.
    
    Returns
    -------
    Response schema:
    https://app.datarobot.com/docs/users-guide/predictions/api/new-prediction-api.html#response-schema
    
    Raises
    ------
    DataRobotPredictionError if there are issues getting predictions from DataRobot
    """
    # Set HTTP headers. The charset should match the contents of the file.
    headers = {
        'Content-Type': 'text/plain; charset=UTF-8',
        'Authorization': 'Bearer ()'.format(APY_KEY),
        'DataRobot-Key': DATAROBOT_KEY,
    }
    url = API_URL.format(deployment_id=deployment_id)
    # Parameterize Prediction Explanations with query parameters listed in the docs:
    # https://app.datarobot.com/docs/users-guide/predictions/api/new-prediction-api.html#request-pred-explanations
    params = {
        'maxExplanations': 'all',
    }
    # Make API request for predictions
```
Use case: what did we learn?

Overall top features
- squareMeters
- tradeTime (Year)
- communityAverage
- tradeTime (Month)
- Lat / lng
- ....

Explanations for individual cases
- If I'm buying a property like #28472: think about whether I could renovate for less than 100,000 yuan (SHAP value for renovationCondition=3 was -10.78, units are 10k yuan).
- ....
SHAP in the DR python client

- **SHAP feature impact**
  - Just like the UI "Export" button

- **SHAP formatted & ranked explanations**
  - Just like the UI "Calculate / Download" button
  - Or choose your own top N features, up to 100

- **SHAP values raw matrix**
  - Minimal formatting, no rank-ordering
  - No limit to number of features returned
Available on the DR Community GitHub!

Advanced topics

- **Explanation clustering**
  - The raw SHAP matrix provides the "flattened" explanations input to the clustering script.

- **Customize your visualizations**
  - The raw SHAP matrix can be fed into the SHAP library charts -- or make your own.
  - More info: [https://github.com/slundberg/shap](https://github.com/slundberg/shap)

- **Model monitoring**
  - Measure SHAP values over time and see if there is a shift in the most important explanations.
Questions & Answers
DataRobot Community

- **Engage, learn, and accelerate** your AI/ML journey
- **Connect** with peers to find solutions to AI challenges
- **Explore** helpful content to take your AI to the next level
- **Build** your brand as an AI expert & thought leader
- **Join** your peers today at community.datarobot.com
- **Questions:** aisuccess-webinars@datarobot.com

**Latest Topics**

- Challenge of Predicting Time
  - by stackblitz in AI & ML General Discussions 3 hours ago
- Unsupervised Learning with no Target
  - by spam in Automated Machine Learning Discussions 4 hours ago
- Multivariable dataset in Regression
  - by raj in AI & ML General Discussions 6 hours ago
- US Cloud—Planned DB Maintenance (March 14th—March 15th)
  - by stackblitz in AI & ML General Discussions 3 hours ago
- Can DR Indicate Trends in my data
  - by stackblitz in Automated Machine Learning Discussions
- Academic License
  - by spam in AI & ML General Discussions
- Predictive Maintenance (of NASA turbines) using DataRobot
  - by stackblitz in AI & ML Knowledge Base
- Europe User’s Guide
  - by spam in Europe Tuesday
- Student/Academic License For Europe
  - by stackblitz in Automated Machine Learning Discussions
- Webinar March 10: Model Building—AutoPilot Done, Now What?
  - by stackblitz in Europe Tuesday

**DataRobot Platform**
Unlock the full power and potential of AI in your company with the DataRobot Platform. Connect, share, and conquer.

**Learning Center**
Get started on your AI journey using DataRobot Platform. Learn about AI and ML, and how to use the DataRobot platform to achieve success.

**AI & ML General**
Get beyond the buzz words. Set in tune with DataRobot people and processes as well as technologies of the AI and ML industry.

**About Community**
Get the latest DataRobot Community news, updates, and resources to make the most out of your time here.
Backup material

In case of questions
SHAP additivity, margins, and link functions

"This doesn't add up to 0.603..."

- Most classification (and some regression) models apply a nonlinear "link function", usually the logistic function.
  - Logistic regression models use it to convert an unbounded linear function to a probability bounded by [0, 1].
  - Even in bounded models like tree ensembles, using it often improves accuracy.
- SHAP values are changes in log-odds -- not directly changes in probabilities.
- SHAP values are additive before the link function.

\[
f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}
\]
SHAP additivity, margins, and link functions

- Predicted probability: 0.603
- Probability base value: 0.101
- Log-odds base value: 0
- Log-odds: -6 to 6

Graph showing the relationship between log-odds and probability with negative and positive effects indicated.