SHAP: Bridging the Gap Between Machine Predictions and Actionable Recommendations

Navigating from Predictive to Prescriptive

Igor Alcantara
Director of Data Science & Education

igor.alcantara@ipc-global.com

Introduction

In today's data-centric landscape, much fanfare surrounds the predictive prowess of machine learning, often casting it as the beacon of the digital age. Yet, while prediction is undeniably a potent facet, the true crown jewel of machine learning lies in its prescriptive capabilities. Moving beyond simply forecasting future events, the power of prescriptive analysis provides actionable insights and recommendations, enabling organizations to not just foresee but actively shape their desired outcomes. At the heart of this prescriptive potential is a tool that ensures clarity and understanding – the SHAP statistics.

The Importance of SHAP Statistics

In the modern era of data-driven decision-making, the ability to not only predict but also understand and interpret model outcomes is vital. As businesses increasingly depend on complex machine learning models, the demand for model interpretability tools has grown. This is where SHAP (**SH**apley **A**dditive ex**P**lanations) values come into play. Particularly in prescriptive analysis, understanding these values can be a game-changer.

What is Prescriptive Analysis?

Prescriptive analysis is the final frontier in the data analytics evolution. While descriptive analysis looks at past data to identify what happened, and predictive analysis forecasts future possibilities, prescriptive analysis recommends actions to address those predictions, ensuring optimal outcomes. It's about making actionable recommendations based on the findings of a predictive model.

Why Model Interpretability Matters

Model interpretability is vital for several reasons:

- **Trust**: If stakeholders do not understand or trust the model, they're less likely to act upon its recommendations.

 - **Regulatory Compliance**: In some industries, understanding the reasoning behind an algorithm's decision is not just a best practice, but a regulatory necessity.

- **Debugging**: Interpreting models helps in diagnosing potential issues, allowing for more accurate and reliable models.

SHAP Values

SHAP values are rooted in cooperative game theory and attribute the change in a model's prediction to each feature value. In other words, SHAP values explain the output of machine learning models by averaging all the possible marginal contributions of each feature across all possible sets.

How does SHAP Values work?

Let's take a dive into a concept using Shap Values. Imagine three individuals walk into a room and take part in a **cooperative** (this is the key word for this) game, with the stake being $100. We'll name them Player A, Player B, and Player C. Given that each brings a unique skillset to the table and contributes distinctively, how can we fairly split the $100 among them? In this discussion, I'm particularly interested in figuring out Player A's deserving cut.

Consider this step-by-step process:

1. Let A come into the room solo, engage in the game, and note the resulting payout.

2. Then, invite B to join A in the room, play again, and write down the outcome.

3. Finally, bring in C to team up with A and B, participate, and document the outcome.

The increased amount each participant brings to the winnings as they enter can guide the division. However, a snag arises. What if Player A and Player B share similar expertise? The one who steps in first would naturally make a more substantial impact since the following player might not introduce much novelty. So, it's worthwhile to also explore scenarios where B comes in before A.

The key and burden is to experiment with every conceivable player arrangement and average out their contributions.

Diving deeper, the illustration that follows presents all potential sequences of their entries, focusing primarily on A's sequence. These sequences are categorized (evidenced by horizontal lines) based on who's already present in the room before and after A's entry. It's crucial to understand that for A, the sequence in which others arrive before or after him doesn't impact his game. What truly matters are the group dynamics. For instance, there are scenarios where both B and C join after A and some where they precede him. When estimating A's role, the sequential order of subsequent players post his entry is irrelevant since A's contribution is already established. Similarly, the sequence of arrivals before A doesn't influence his game; in his eyes, he's partnering with the same players regardless of who entered first.

A B C
A C B
B A C
C A B
B C A
C B A

SHAP STATISTICS
ipc-global.com

Below, you'll find the formula to determine Player A's Shap Value. At first, the set symbols and factorials might give it a complex look. However, I found that reading the formula with some annotations and color highlights will make it a bit more digestible. I could keep going this path and break down each piece of the equation but I believe for the purpose of this eBook, this is deep enough to give you an idea of how this important metric is calculated.

$$\phi_A = \sum_{S \subseteq F \setminus \{A\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{A\}) - f(S)]$$

Number of combinations before A

Number of combinations after A

A's contribution to this coalition

A B C
A C B
B A C
C A B
B C A
C B A

Summation is over these 4 coalition blocks

Total number of combinations/ coalitions

Coalition's value after A is added

Coalition's value before A is added

Interpretability vs. Complexity: Finding the Balance

The trade-off between interpretability and complexity is at the core of many data science discussions. While complex black box models may offer superior predictive performance, their lack of transparency can hinder their real-world application, especially in domains where understanding the decision-making process is crucial.

This is where tools like SHAP come in, bridging the gap by offering a lens into the black box, allowing for a better understanding of even the most complex models. In essence, as businesses and industries continue to evolve, finding the right balance between complexity for accuracy and interpretability for trust becomes a pivotal consideration in model selection and deployment.

Following the exploration into SHAP values and prescriptive analysis, it's imperative to understand the broader context of model interpretability, which brings us to the dichotomy of white box and black box models.

White Box Models (Interpretable Models)

White box models, often termed as "interpretable" or "transparent" models, are those whose internal workings and decision-making processes can be easily understood by humans. Linear regression, logistic regression, and decision trees are common examples. Using simple math equations, one can easily understand why a, b, and c are used to predict x.

White box models come with their distinct set of advantages and limitations. Their primary strength lies in transparency, providing a decision-making process that is both clear and straightforward. This clarity doesn't only illuminate how decisions are made but also streamlines the debugging process, making it easier to identify and

IPC Global, Inc. All rights reserved.

rectify issues. As a result, stakeholders often find it easier to place their trust in these models, owing to their grasp of the underlying mechanics. However, this transparency comes at a cost. The simplicity inherent to white box models can be a double-edged sword. While they offer clarity, they may struggle to capture the more intricate nuances, complex patterns, or non-linear relationships that are sometimes present in the data.

*Black Box Models (Complex Models)*

Black box models are those where the internal workings are not easily interpretable by humans. These are typically complex models that can capture intricate patterns in data but do so at the expense of transparency. Neural networks, random forests, and some ensemble methods are common black-box models.

On the upside, they frequently outshine white-box models in terms of performance, often delivering superior accuracy and more precise predictions, especially when grappling with multifaceted data. Their inherent flexibility allows them to discern and capture non-linearities and intricate patterns that might elude simpler models. However, this capability doesn't come without its pitfalls. The very essence of a black box model means there's an inherent lack of transparency, making the reasoning behind any given prediction more elusive. This opacity can sometimes lead to skepticism and resistance from stakeholders, who might find it challenging to trust what they cannot readily understand or interpret.

Why SHAP for Prescriptive Analysis?

Although there are other techniques for interpretability and prescriptive analysis, SHAP is by far one of the best options to achieve this objective. These are some of the reasons.

- **Consistent Interpretation**: Unlike other methods which may provide inconsistent explanations for a given prediction, SHAP values guarantee consistency.

- **Fair Attribution**: As they derive from Shapley values in game theory, they ensure a fair distribution of contributions among the features.

- **Granularity**: SHAP values can be calculated for each individual prediction, offering fine-grained insights.
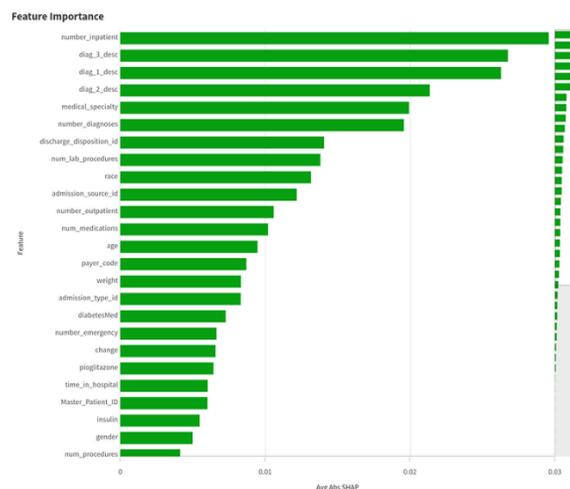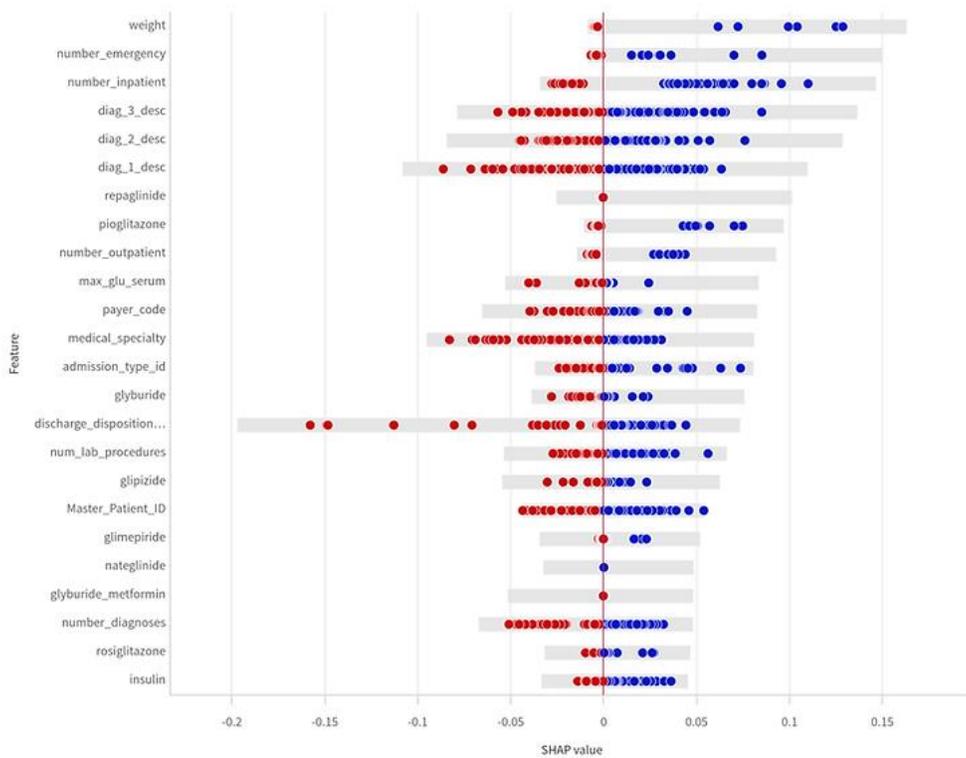

Summary SHAP chart produced by Qlik

Making Recommendations More Actionable

With SHAP values, businesses can understand which features (or factors) most influence a particular prediction. By adjusting these influential factors, businesses can strategically change the predicted outcome.

For instance, in a customer churn prediction model, if a specific feature like 'customer support interactions' heavily influences a customer's likelihood to churn, SHAP values can help quantify this influence. By understanding this, a business can enhance its customer support experience or streamline its interaction processes, thereby reducing the chance of customers leaving in the future.

Detailed SHAP chart produced by Qlik Analytics

Enhancing Stakeholder Trust

When stakeholders can visualize and understand how a machine learning model arrives at a particular recommendation, they are more likely to trust its advice. SHAP values, represented visually, can offer stakeholders an intuitive grasp of the model's inner workings.

Conclusion

In the quest for more transparent and understandable machine learning models, Qlik AutoML emerges as a forerunner by integrating SHAP statistics for every row in the prediction dataset. This integration means that decision-makers are equipped not just with cutting-edge predictions but also with deep insights into the factors driving those predictions. By understanding the 'why' behind the model's output, stakeholders can develop a more nuanced understanding of their business operations. This clarity paves the way for data-informed strategies, empowering businesses to navigate complex landscapes with precision and confidence.

In a world that's constantly evolving with technology, the capacity to make informed and intelligent decisions quickly is invaluable. SHAP statistics, by enhancing model interpretability, take prescriptive analytics from a black-box recommendation system to a transparent guide. By integrating SHAP into prescriptive analysis workflows, businesses can make more informed decisions, fostering trust and facilitating actionable insights. To conclude, the main three keywords related to SHAP and its power are: transparency, prescriptive, and actionable.

About IPC Global

IPC Global is a diverse, nationwide team of experienced, engaging, and effective data analytics experts. We provide advisory, consulting and managed services to solve your organization's Enterprise Intelligence challenges.

For more than 20 years we have been the go-to source for Artificial intelligence, Data Analytics, Data Science & Research, Business intelligence, Cloud Solutions, Data Integration, Education, Advisory services and more.