Appsilon
DATA SCIENCE

Explaining black-box models with xspliner to make deliberate business decisions

eRum 2020 | Krystian Igras | 06/2020

krystian@appsilon.com

Why should we explain our models?

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

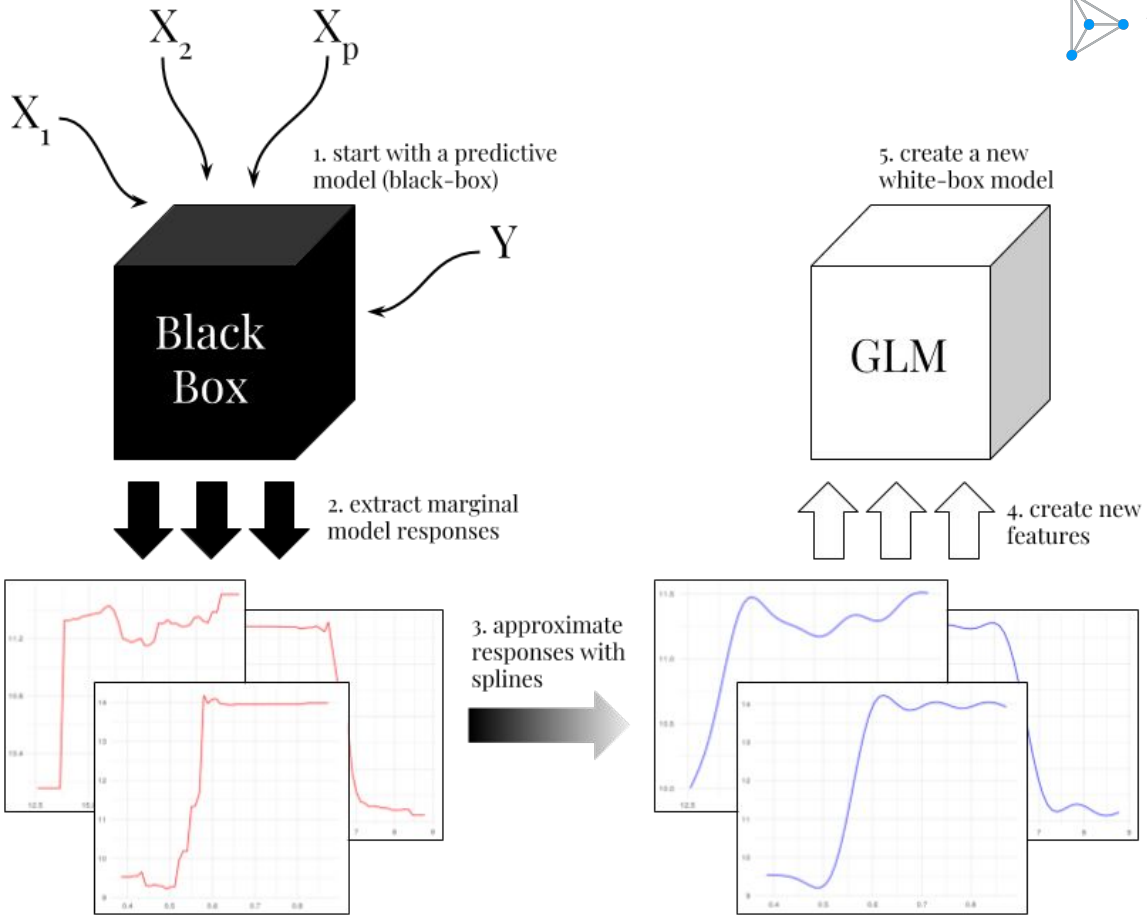
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Performance vs interpretability

https://dzone.com/articles/prediction-explanation-adding-transparency-to-mach

Appsilon
DATA SCIENCE

The way it works



Marginal response - PDP

Construction of the new, GLM, model

GLM with transformed variables

$$Y \sim \beta_0 + \beta_1 \cdot g_1(X_1) + \ldots + \beta_p \cdot g_p(X_p)$$

i-th PDP of GLM model

$$f_{GLM_i} = \beta_0 + \beta_i \cdot g_i$$

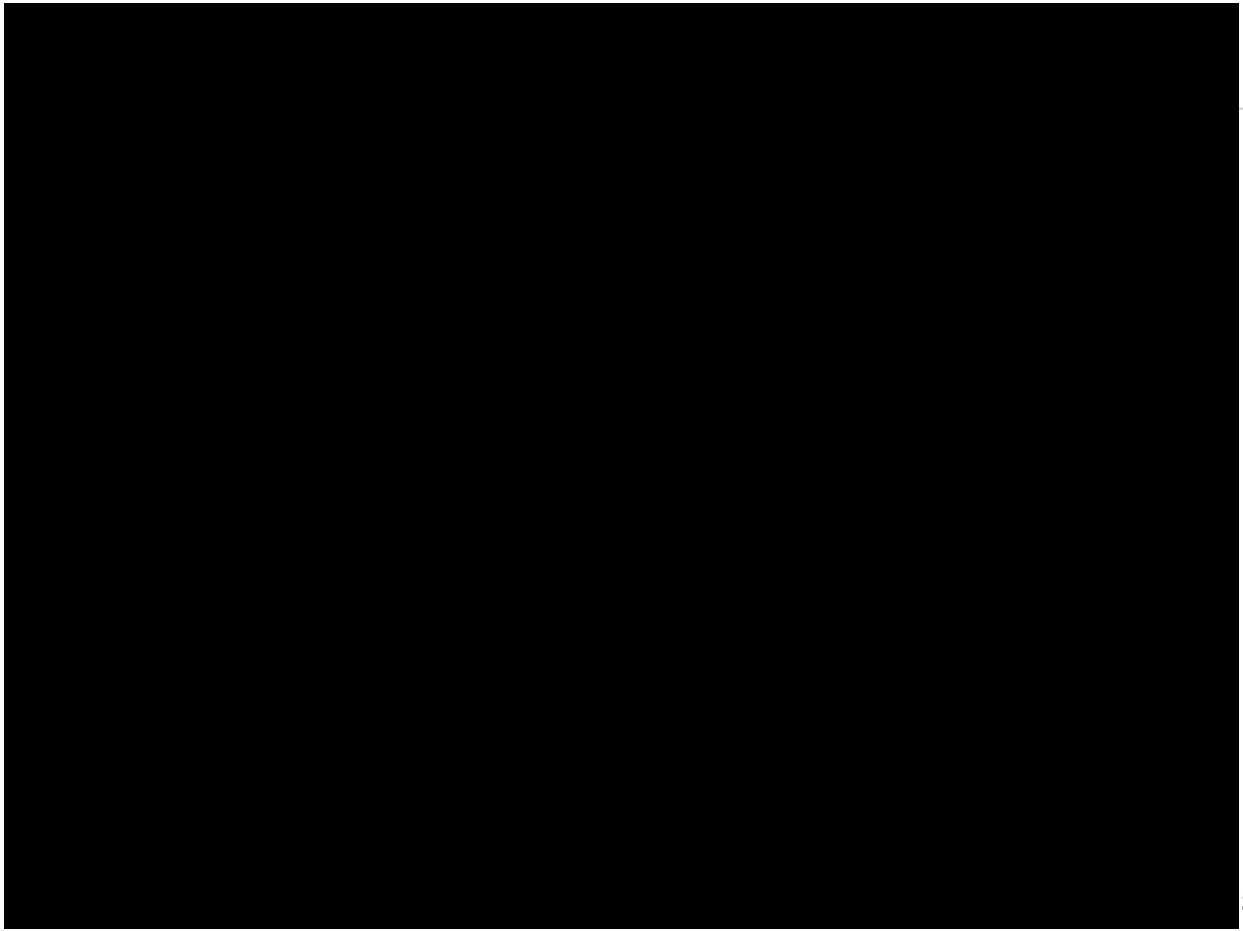
GLM with spline-based approximated PDP transformations

$$Y \sim \beta_0 + \beta_1 \cdot \tilde{f}_{BB_1}(X_1) + \ldots + \beta_p \cdot \tilde{f}_{BB_p}(X_p)$$

- https://modeloriented.github.io/xspliner/
- https://github.com/ModelOriented/xspliner

Credit scoring - use case

Data:
https://www.kaggle.com/c/GiveMeSomeCredit
Base model:
https://www.kaggle.com/fastwalker/give-me-credit



Thanks!

Feel free to reach out:

- krystian@appsilon.com
- github.com/krystian8207
- twitter.com/krystian8207
- linkedin.com/in/krystian-igras-a3068b152/

Presentation theme source: slidescarnival.com