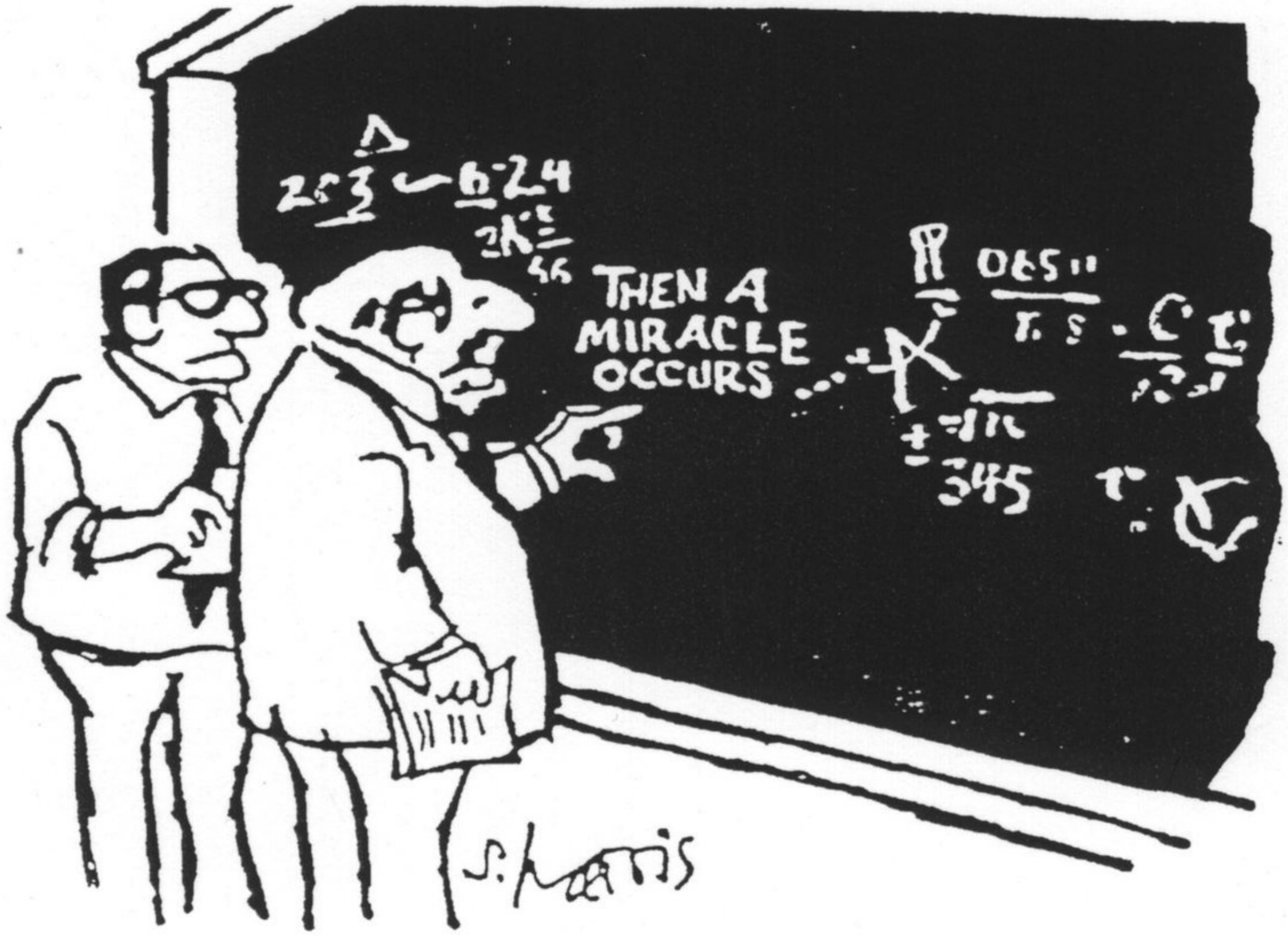


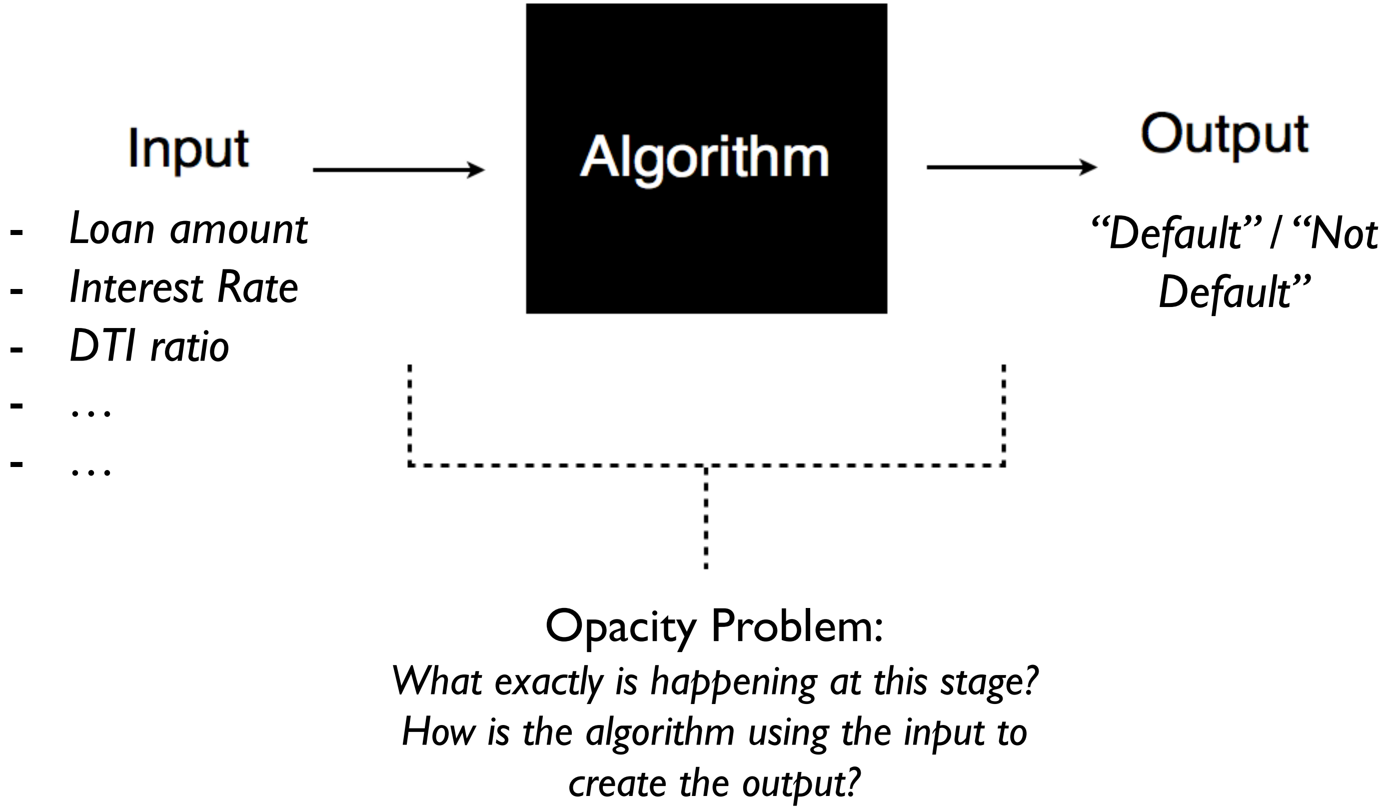
"I think we need to be a little more specific during step 2"

EXPLAINING BLACK BOX CREDIT RISK MODELS

JOSEPH MCHUGH

THE PROBLEM OF EXPLAINABILITY

- Complex machine learning algorithms are being increasingly used by lending institutions
- The lack of explainability that these black box models present has become problematic.



FICO®

"it is often the case that neither the trained model nor its individual predictions are readily explainable, but regulators and consumers demand explanations"

–FICO, 2018

WHY DOES IT MATTER?

1. Regulatory Compliance
2. Customer Understanding
3. Identifying Biases

DATA USED

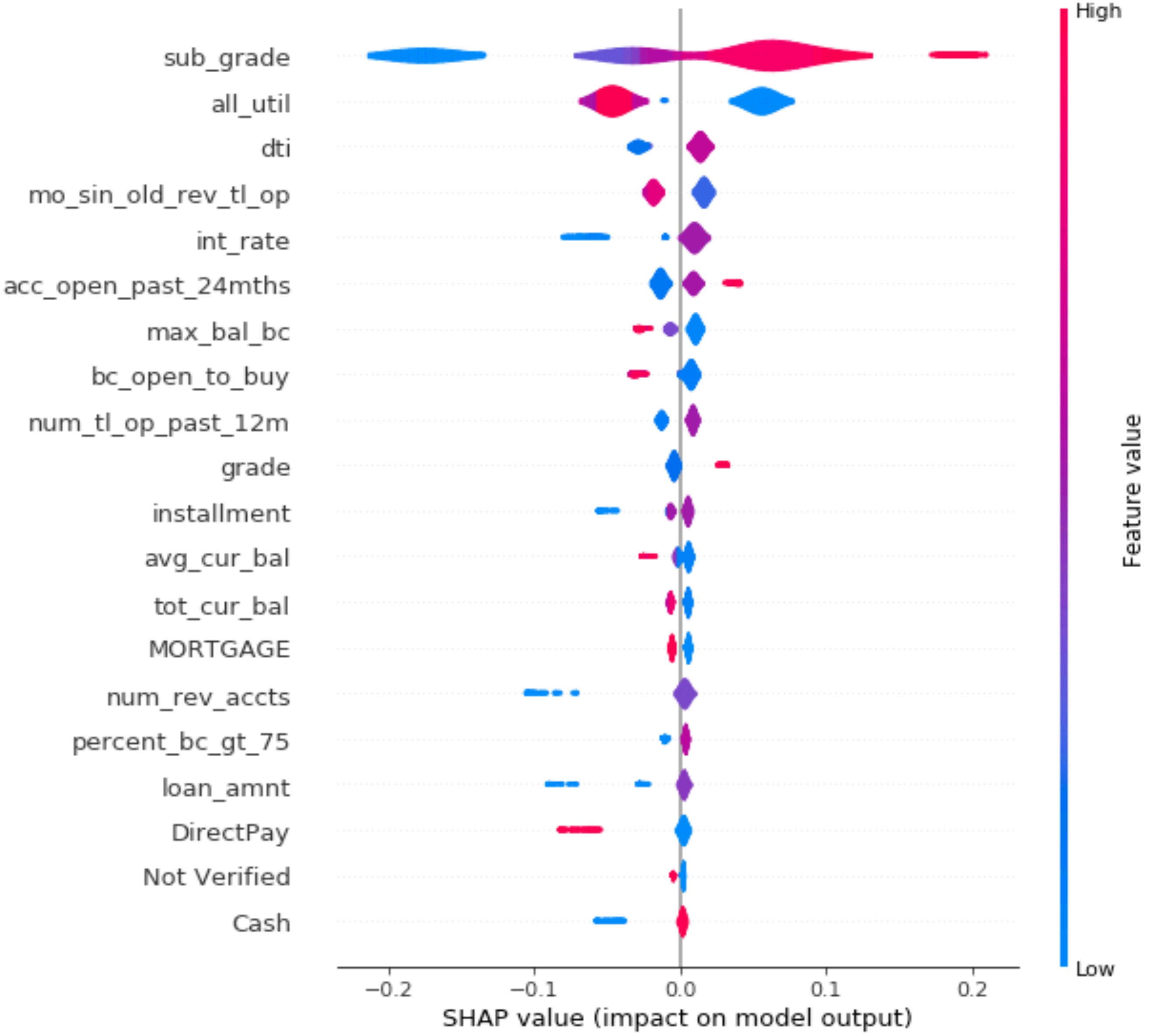
- Lending Club Data
- Period is from 2007-2018
- 2.6 million observations
- 148 feature columns

THE MODEL: SEQUENTIAL NEURAL NETWORK

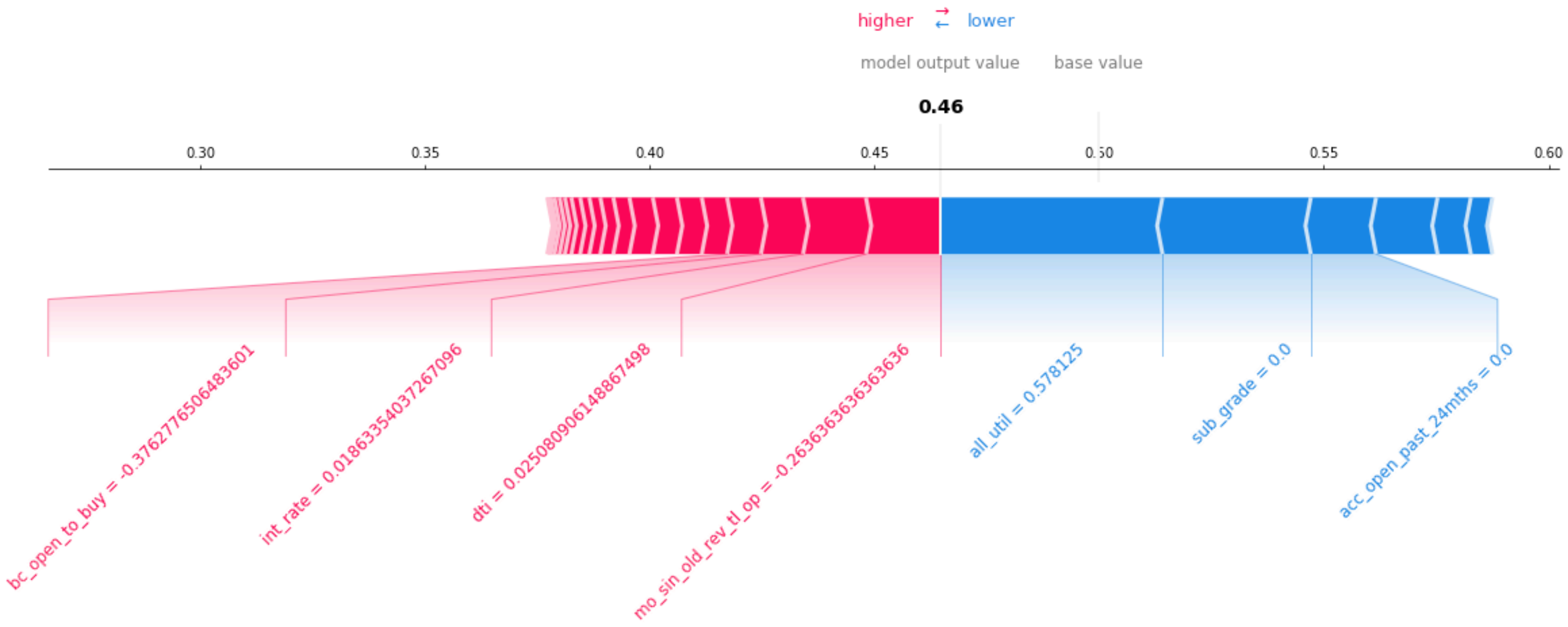
Recall: 0.806

AUC: 0.762

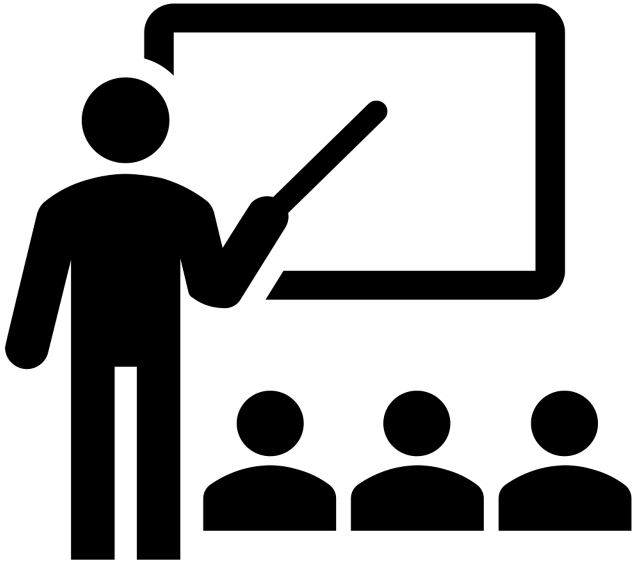
USING SHAP VALUES TO OPEN THE BLACK BOX:



MORE IMPORTANTLY: LOCAL EXPLANATIONS



APPLICATIONS



- Lenders can now highlight at a global level the interior workings of an algorithm to regulators
- Customers can have explanations as to why they were denied a loan
- Customers can receive "feedback" on how to improve loan applications in the future

FUTURE WORK

- Build software that can highlight information for bank applicants in a succinct way
- Optimizing models with different probability thresholds to achieve optimal tradeoffs between risk mitigation and loan revenue