Deep Learning Indaba 2025, Kigali –-August 17--22

pastel

White Box Finance: Interpreting AI Decisions in Finance through Rules and Language Models

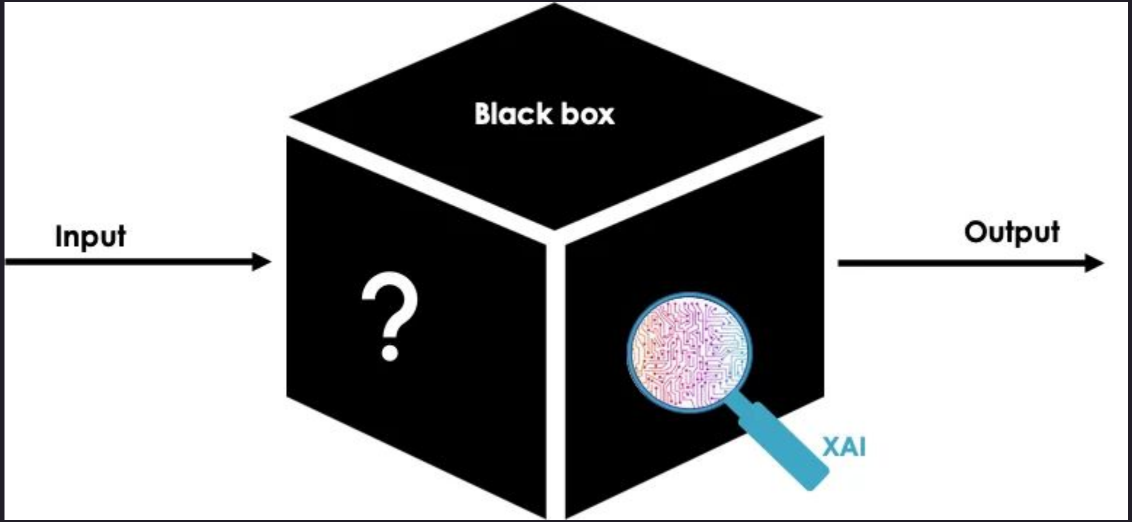
Authors: Oluwafemi Azeez, Samson Tontoye, Olorunleke White, Abuzar Royesh

Motivation

Loan defaults → major financial losses.

ML models (e.g., XGBoost) improve prediction, but are black-boxes.

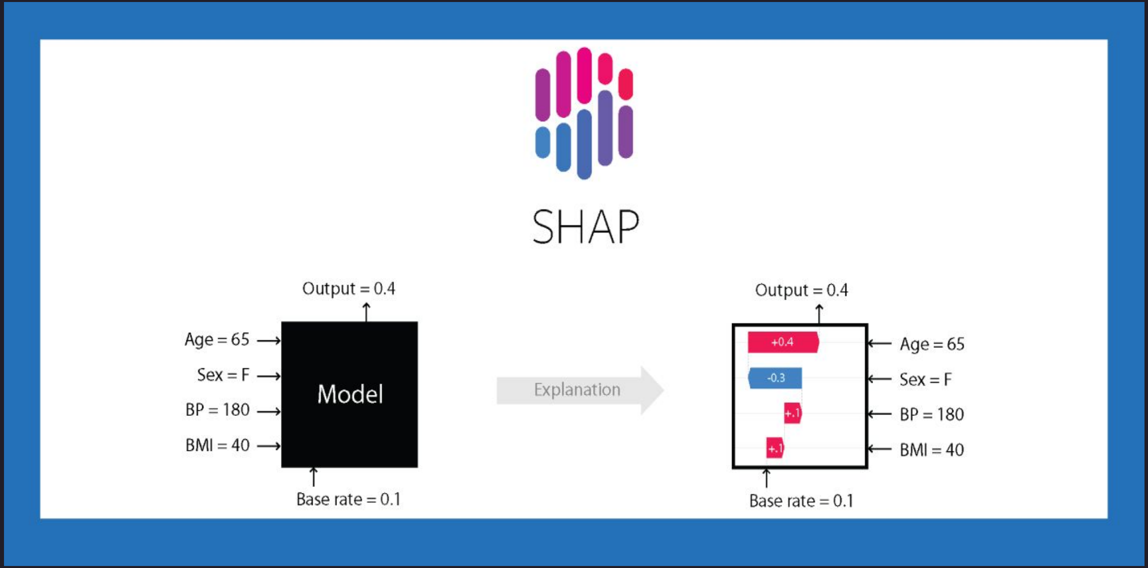
Finance requires transparent, auditable explanations for regulators, loan officers, and customers.



Methodology

Enhance interpretability and trust in AI credit risk models by creating and comparing:

SHAP + GPT-4 → feature-based + natural language explanations.



＋ GPT-4

Methodology

Enhance interpretability and trust in AI credit risk models by creating and comparing:

Rule-based logic → transparent, business-aligned decision rules.

Experimental Setup

- Dataset: Anonymized loan applicant records containing demographics, employment, credit history, and repayment behavior.
- Preprocessing: Missing values removed (<1%), categorical variables frequency-encoded, numerical features preserved.
- Model: XGBoost classifier trained with 5-fold stratified cross-validation. Class imbalance addressed using scale_pos_weight.
- Evaluation Metrics: Area Under the Curve (AUC), Precision, Recall, F1-score, and Confusion Matrix analysis

Experimental Setup

Explanation Modules

Two complementary explanation pipelines were applied to model predictions:

- SHAP + GPT-4: Local feature attributions → top 3–5 contributors → converted into business-friendly textual narratives.
- Rule-Based Logic: Categorical histograms and KDE plots used to derive interpretable decision rules aligned with institutional underwriting heuristics.

Results

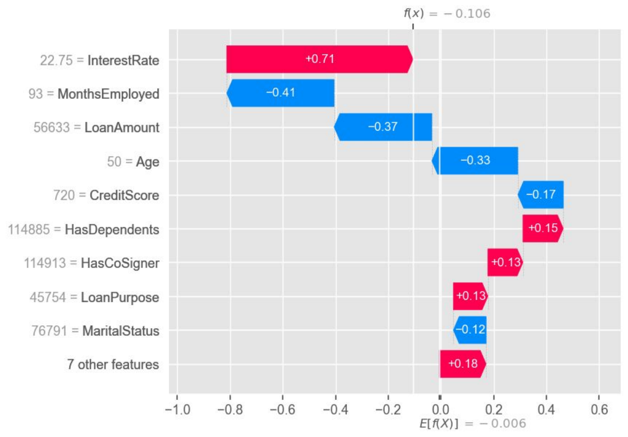
Model Prediction

Age	Income	Loan Amount	Credit Score	Months Employed	Interest Rate	DTI Ratio	Education
36	80846	179949	347	20	23.96	0.9	PhD

Results

SHAP + GPT-4

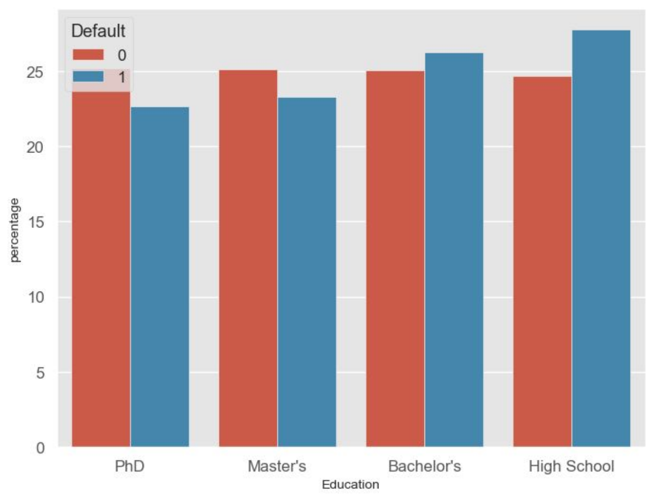
The interest rate on the loan is quite high at 22.75%. This significantly increases the cost of borrowing, making it more challenging for the customer to manage their monthly payments. The high SHAP impact of 0.71 indicates that this factor is a strong contributor to the default risk

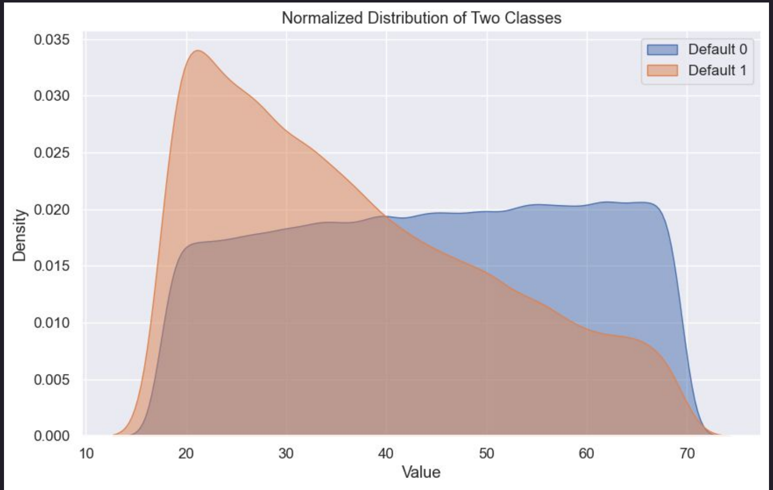


Results

Rule-Based Logic

```
if row["Age"] < 40:
    explanations.append("Young age may indicate lack of financial experience.")
```





Conclusion and QA

- GPT Explanations → Rich, nuanced, human-friendly
- Rule-Based Explanations → Transparent, audit-ready, regulatory aligned
- Hybrid Approach = Best of both worlds: trust + compliance