Model Interpretation and Identification of Key Features Behind Company Failure

Anecka Estocado | Xylgei Fuentespina

Project Goal

- Identify & Separate
- Interpret Financial Signals

Model Results

1874 companies, 80/20 split, 5-fold Cross Validation

Model	ROC-AUC (Mean + Std	F1 (Mean + Std)	Precision	Recall
LightGBM	0.9330 ± 0.0471	0.4338 ± 0.0495	0.6546	0.3271
XGBoost	0.9247 ± 0.0353	0.5029 ± 0.0904	0.6494	0.4162
Random Forest	0.9200 ± 0.0388	0.3621 ± 0.1156	0.6483	0.2576
Logistic Regression	0.6102 ± 0.0239	0.0182 ± 0.0364	0.1000	0.1000

LightGBM Model Results

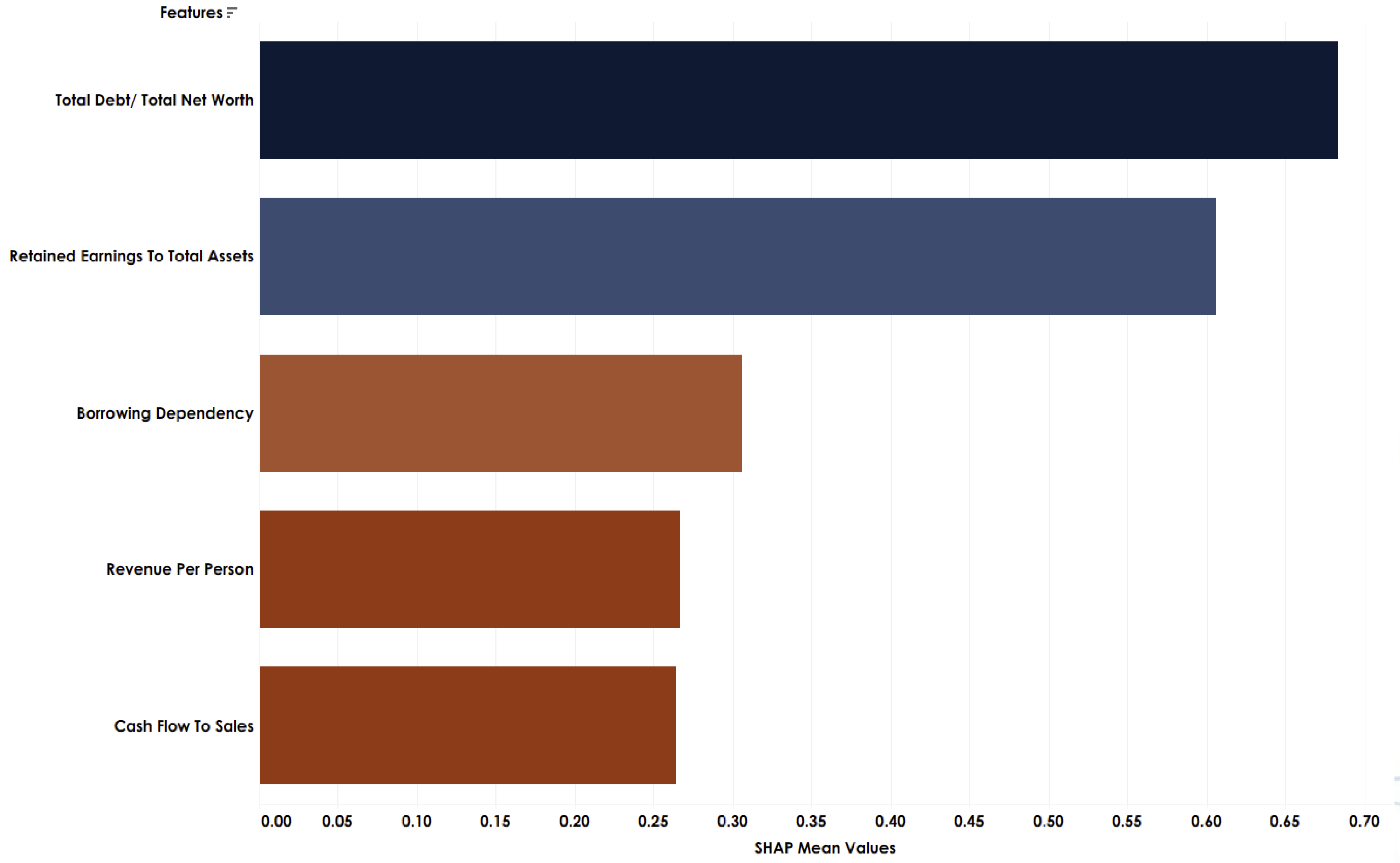
	Predicted Label	
Actual Label	**Non-Bankrupt**	**Bankrupt**
Non-Bankrupt	353	2
Bankrupt	13	7

93.2% ROC-AUC

65% Precision

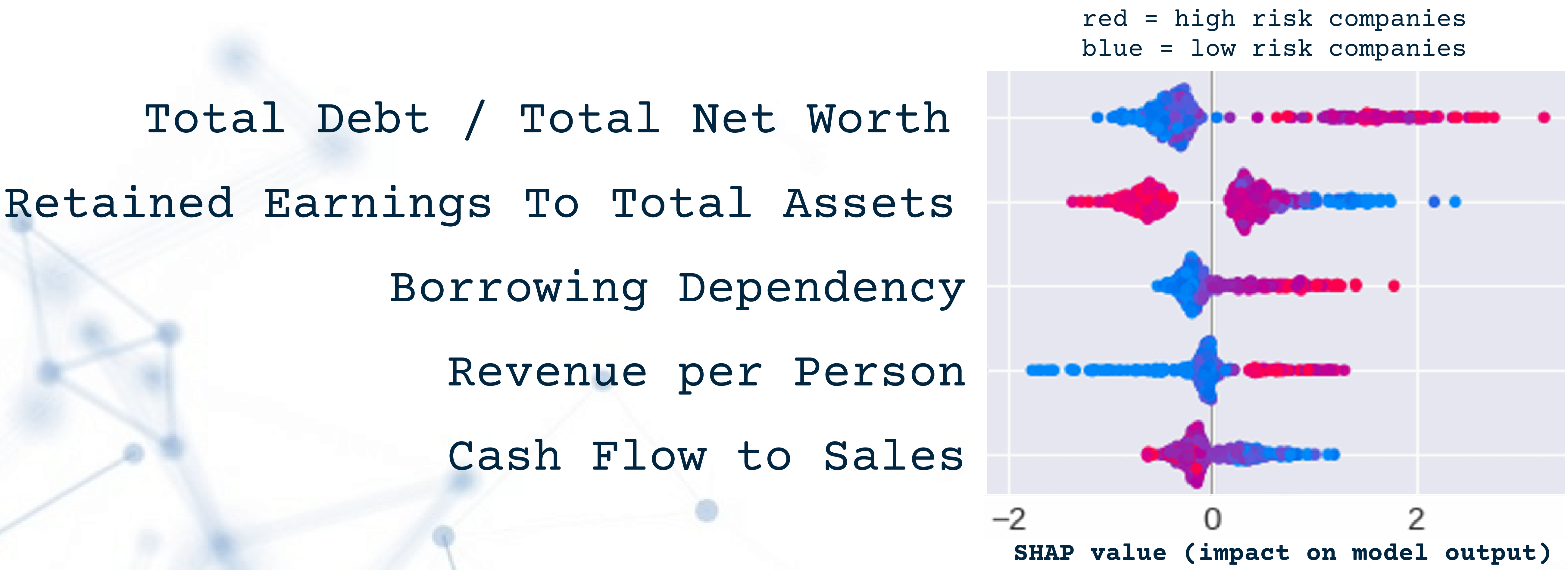
33% Recall

Top Features



SHAP Beeswarm Plot

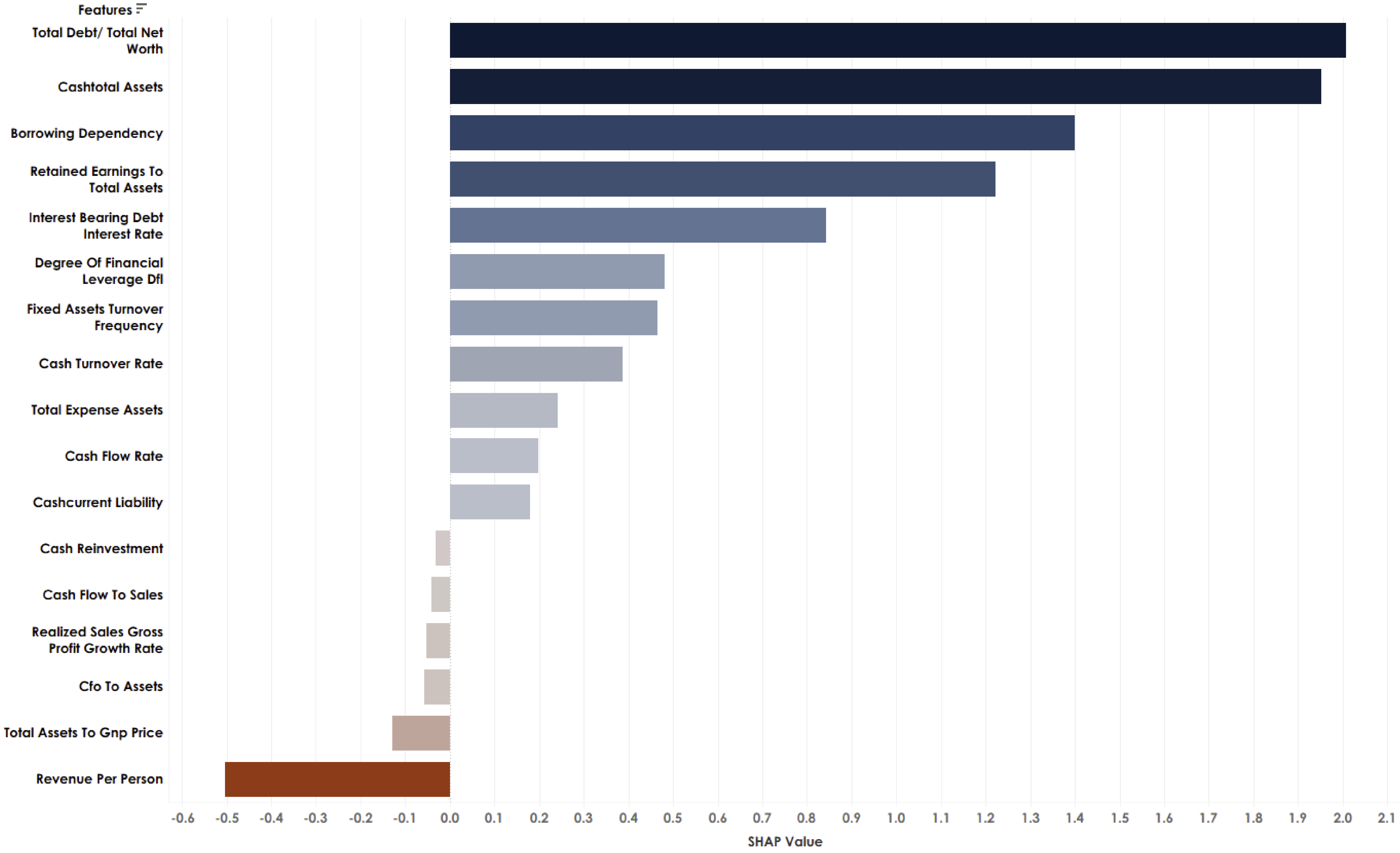
Each dot is a company; position shows impact
on bankruptcy prediction



Why is the model wrong?

Case A: Non-Bankrupt Company but Predicted Bankrupt

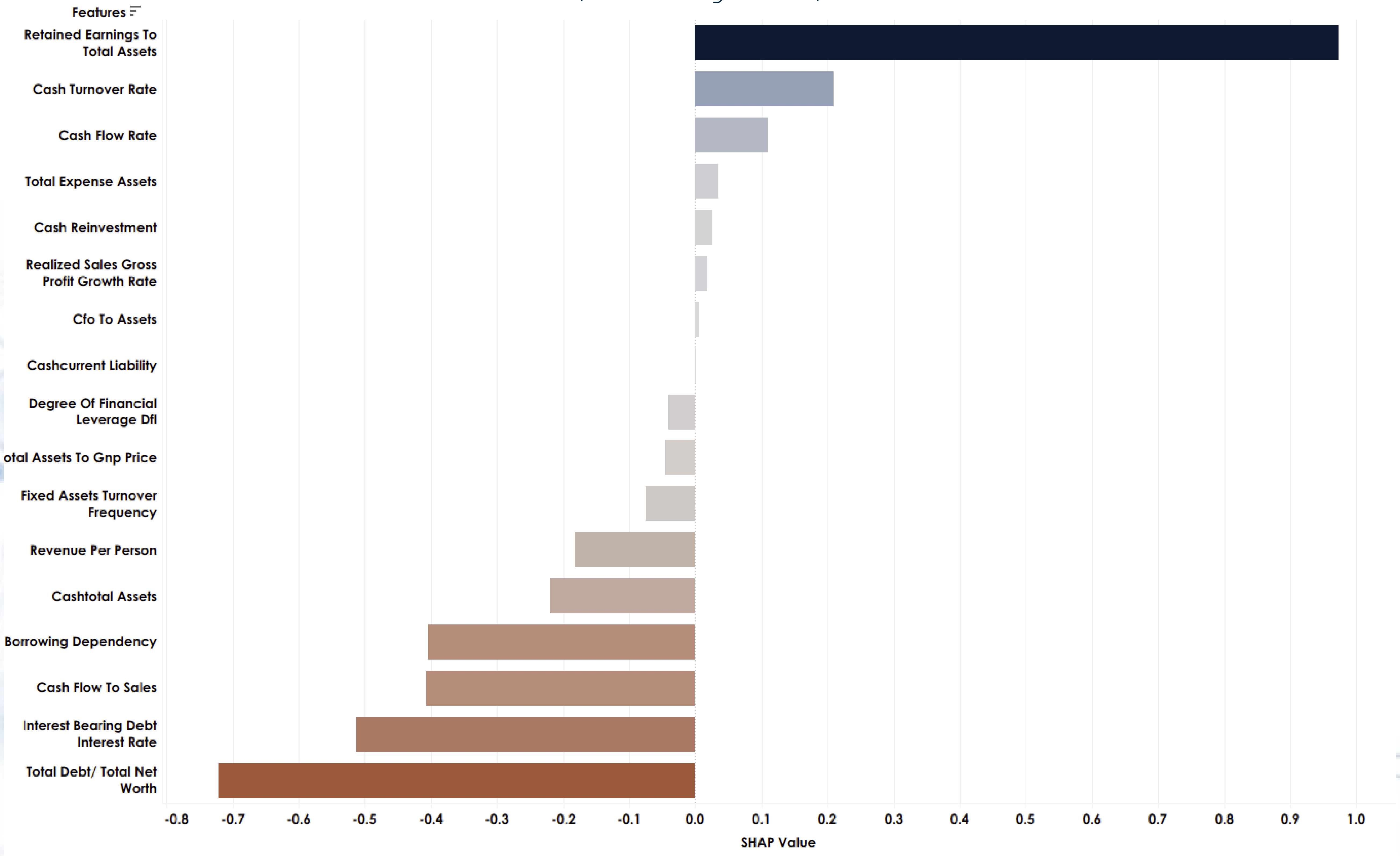
(False Positive)



Why is the model right?

Case B: Non-Bankrupt

(True Negative)



Summary

- High **Total Debt vs. Net Worth** is the #1 predictor of failure.
- The Safety Net: **Retained Earnings**
- Top Predictors: Beyond debt and savings, **Borrowing Dependency** and **Revenue Efficiency**
- The **LightGBM** model achieved a **93.2% CV-AUC**

Question and Answer