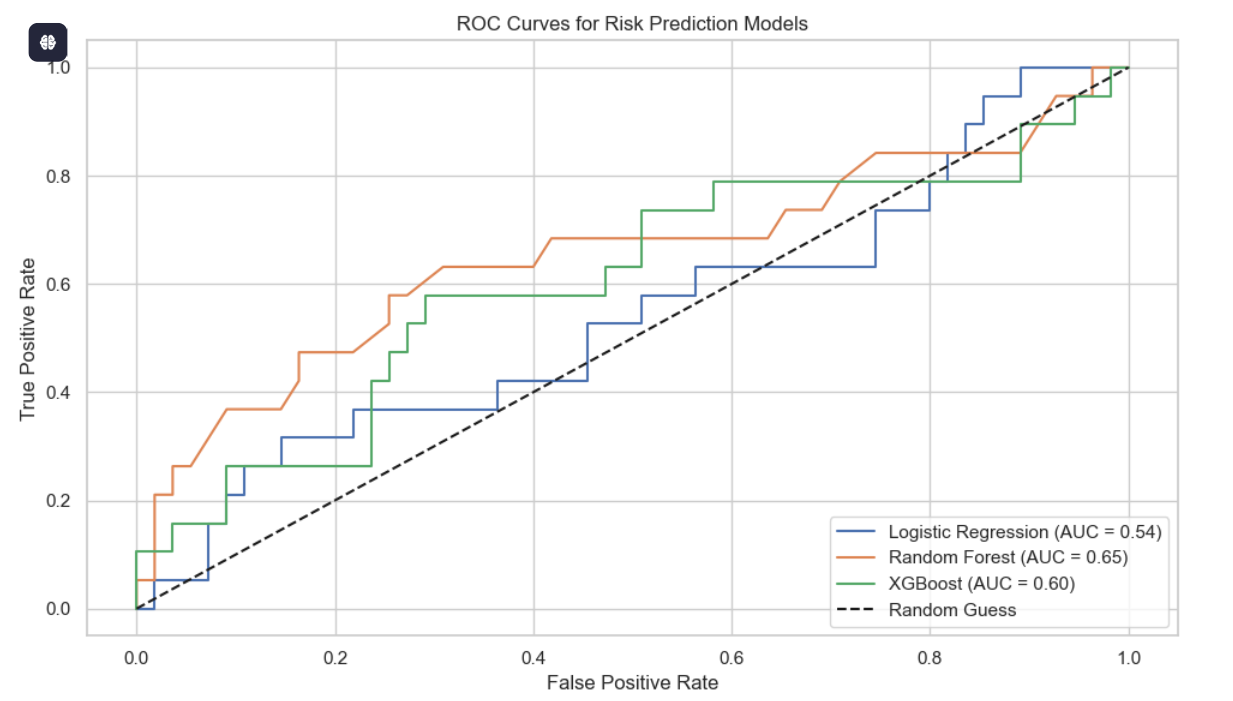
**Recommendations & Executive Summary**

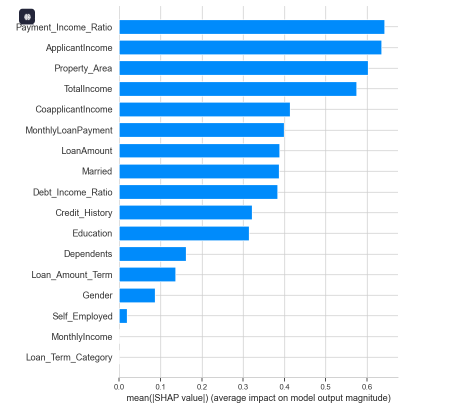
**Business Insights**

Based on the model predictions and performance:

* **Flag high-risk customers**: With AUC scores only moderately above random guess, models can still help flag *potentially risky* applicants for closer scrutiny—especially when combined with domain knowledge.



* **Prioritize Payment-Income Ratio & Applicant Income**: These are the **top drivers** of default prediction based on SHAP analysis. Customers with poor ratios or low income should trigger automatic flags or stricter loan terms.



* **Adjust loan terms accordingly**:
  + Offer **shorter terms or smaller amounts** to higher-risk customers.
  + Allow **better interest rates or larger loans** to applicants with strong repayment indicators (e.g., high income, favorable ratios, married status).
* **Improve data quality and feature engineering**: Since even the best model only achieved ~0.53–0.54 AUC, better data and feature creation could improve predictive power.

**Executive Summary**

**What We Did:**

* Developed and evaluated 3 ML classification models: **Logistic Regression, Random Forest, and XGBoost**.
* Used SMOTE for class balancing and cross-validation to assess robustness.
* Applied SHAP for model explainability to identify the most influential features.

**What Worked:**

* **Random Forest performed best** (AUC ≈ 0.53), followed by XGBoost (AUC ≈ 0.52).
* **Logistic Regression underperformed** with an AUC ≈ 0.50, barely above random chance.
* SHAP identified **Payment\_Income\_Ratio**, **ApplicantIncome**, and **Property\_Area** as top features influencing loan default prediction.

**What the Business Should Do:**

1. **Use model predictions to support decision-making**, not replace it—models offer moderate insight but require human review.
2. **Incorporate key features like Payment\_Income\_Ratio** into internal risk scoring systems.
3. **Enhance the dataset** with additional behavioral, transactional, and credit bureau data to improve future model performance.
4. **Iteratively retrain models** as more labeled data becomes available and model feedback loops are integrated into decision systems.