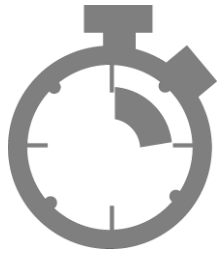
DIGICA BOOTCAMP - DS FINAL PROJECT

Machine Learning Assisted Loan Approval Prediction

Erick Khosasi

Implementing Machine Learning (ML) model can significantly **boost revenue growth**



2 Hours → **30 Minutes**

4x Processing Capacity

Up to **billions** of Revenue Growth

Current loan approval is inefficient

- Manual → time-consuming and prone to **bias/error**
- Processing limit → **bottleneck** in customer acquisition
- Approval speed ~ customer satisfaction, market share, and

  Revenue growth

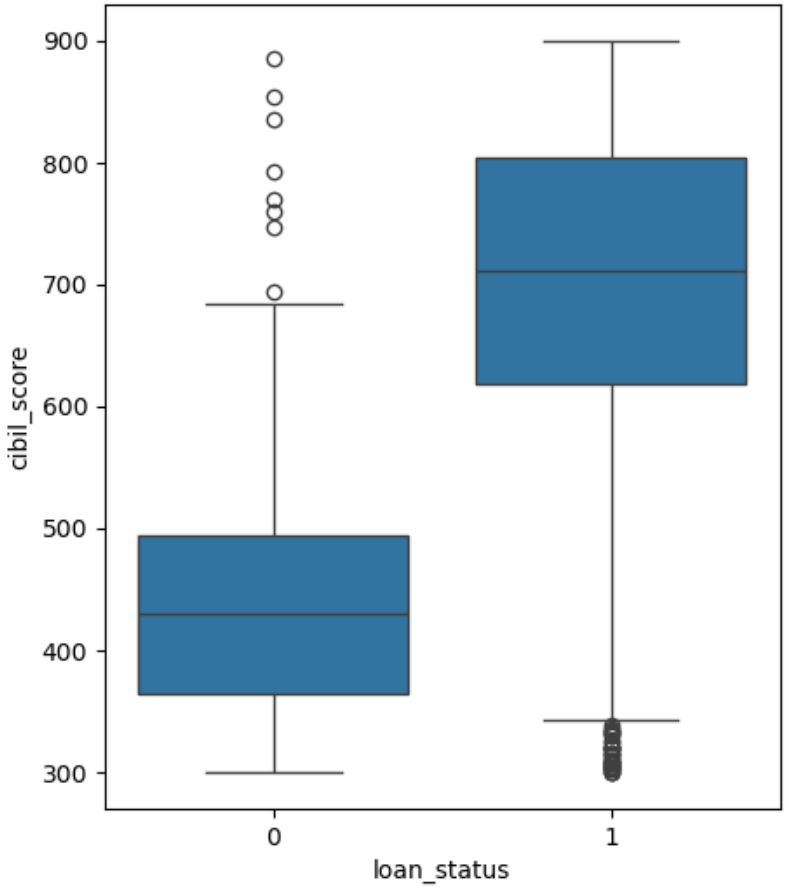
Proposed ML-Assisted Solution

- ML model with **Random Forest** algorithm
- Model performance:
  - Precision ≈ **1.0** → Avoid **false positive**
  - Recall ≈ **0.95** → Avoid **false negative**
- Applications screening within seconds
- Flag high-risk/edge cases

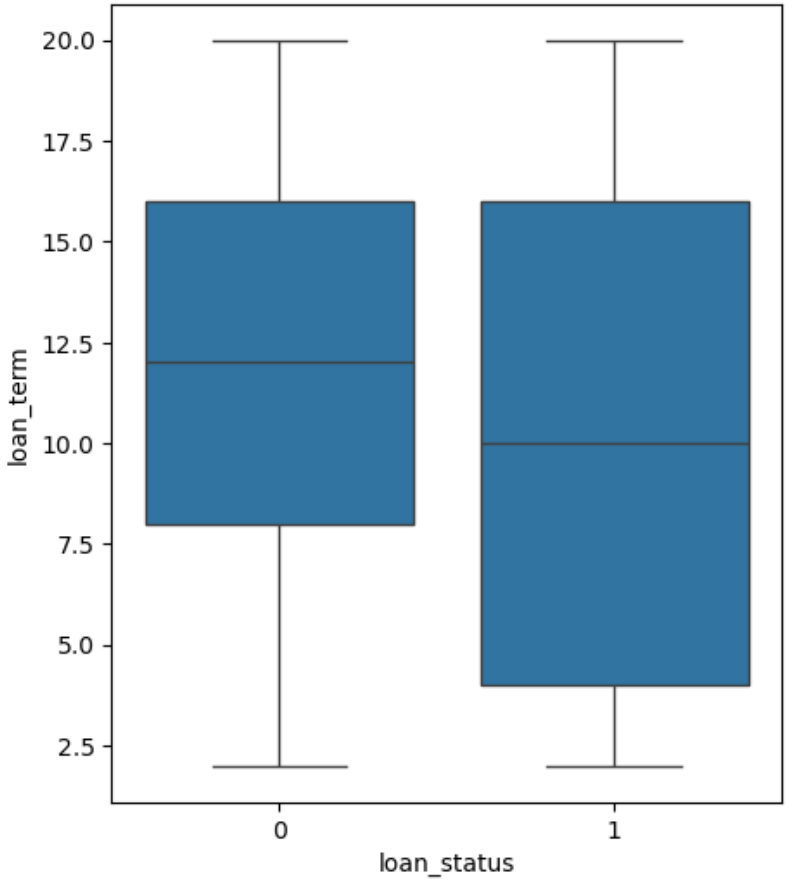
Training Data

- ML model was trained using existing historical data
  - **Cibil Score**
  - **Loan Term**
  - Education
  - Employment
  - Loan Amount
  - Annual Income
  - No. of Dependents
  - Residential Asset
  - Commercial Asset
  - Luxury Asset
  - Bank Asset
  - Loan Status

Feature Relationship with Approval Status



Point-Biserial : 0.7705



Point-Biserial : 0.1130

Revenue Growth Potential

Processing Capacity Increase:

- Current: ~ 4,269 applications per month
- ML assisted: ~ 16k applications per month (**4x increase**)

Extra capacity → onboard more customers → more loan

Using our historical data distribution

Approved loan amount (median)	=	**₹14,600,000**
Loan Interest (simple interest)	=	**10% p.a.**
Annual Interest Revenue	=	**₹1,460,000**

Let’s calculate the potential revenue

Application/Month (historical) = **4,269** applications

- **2,656** (**62.2%**) Approved
- **1,613** (**37.8%**) Rejected

Lost Approvals (95% recall)	=	5% × 2,656	=	**133** loss
Revenue Loss	=	133 × 1,460,000	=	**₹194M** loss
New Approval Rate	=	62.2% × 95%	=	**59.1**%
Extra Applications/Month	=	133 / 0.591	=	**225** applications

Revenue Scenarios

Incremental applications/month		Expected incremental approvals (59.1%)	Annual Interest Revenue	Net Annual Interest Revenue
+10%	500	296	₹432M	₹238M
+25%	1,000	591	₹863M	₹669M
+50%	2,000	1,182	₹1.73B	₹1.53B
+100%	4,000	2,364	₹3.45B	₹3.26B

Limitations and Opportunities

Limitations

- No repayment risk prediction
- Recall trade-off
- Performance degradation
- Bias from historical data

Improvement Opportunities

- Integrate repayment risk model
- Hybrid decision-making
- Continuous learning

Executive Summary

- **Efficiency** : Increase processing capacity by **4x**
- **Risk** : ~0% bad loans approved
- **Loss** : ~5% (recall) good loans rejected
- **Breakeven** : ~225 extra applications/months
- **Flexibility** : Decision threshold can be tuned

Recommendation

Implement **hybrid decision approach**:

- Model prediction + manual review for first few months
- Validate results and fine-tune performance

DIGICA BOOTCAMP - DS

FINAL PROJECT

by **Erick Khosasi**

Thank You
For Your Attention