



Research seminar II

MACHINE LEARNING MODELS FOR PREDICTING PERSONAL LOAN ACQUISITION

UNIVERSITY OF PRIMORSKA
FACULTY OF MATHEMATICS, NATURAL SCIENCES
AND INFORMATION TECHNOLOGIES

PIKA POVH MAVRIČ

MENTOR: BALÁZS DÁVID

MAY 2026

MOTIVATION

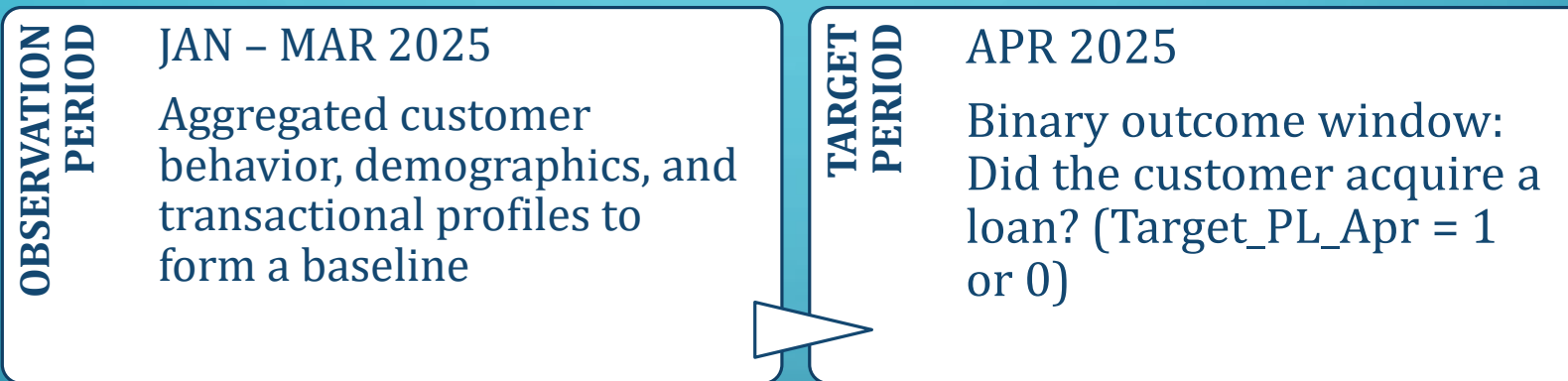
- **Strategic Importance:** Consumer loans are one of the most profitable products for retail banks.
- **Digital Transformation:** Banks now collect vast amounts of data via mobile apps and online transactions, creating a major opportunity for analytics.
- **Customer-Centric Banking:** Modern banking is shifting from selling products to everyone to understanding the specific life stages and needs of individuals.
- **Competitive Advantage:** Using data to stay ahead of customer needs ensures long term loyalty and higher conversion rates.

PROBLEM STATEMENT

- **Inefficiency of Mass Marketing:** Sending the same loan offers to the entire customer base is expensive and leads to very low response rates.
- **Marketing Fatigue:** Customers are overwhelmed with irrelevant notifications, which they often perceive as spam.
- **The Data Gap:** Having a lot of data is not enough; the real challenge is identifying the hidden patterns that signal a customer's intent to borrow.
- **The Research Question:** How can we use machine learning to filter through thousands of customers and find the few who actually need a loan?

METHODOLOGY

- **Temporal Framework:** The study is built on a two-stage temporal structure to capture cause and effect.



- **Target Variable:** A binary indicator (Target_PL_Apr) where 1 represents a successful purchase and 0 represents no purchase.
- **Outcome:** A total of 566 actual loan purchases were recorded in the target month.

DATA PREPARATION & ALGORITHMS

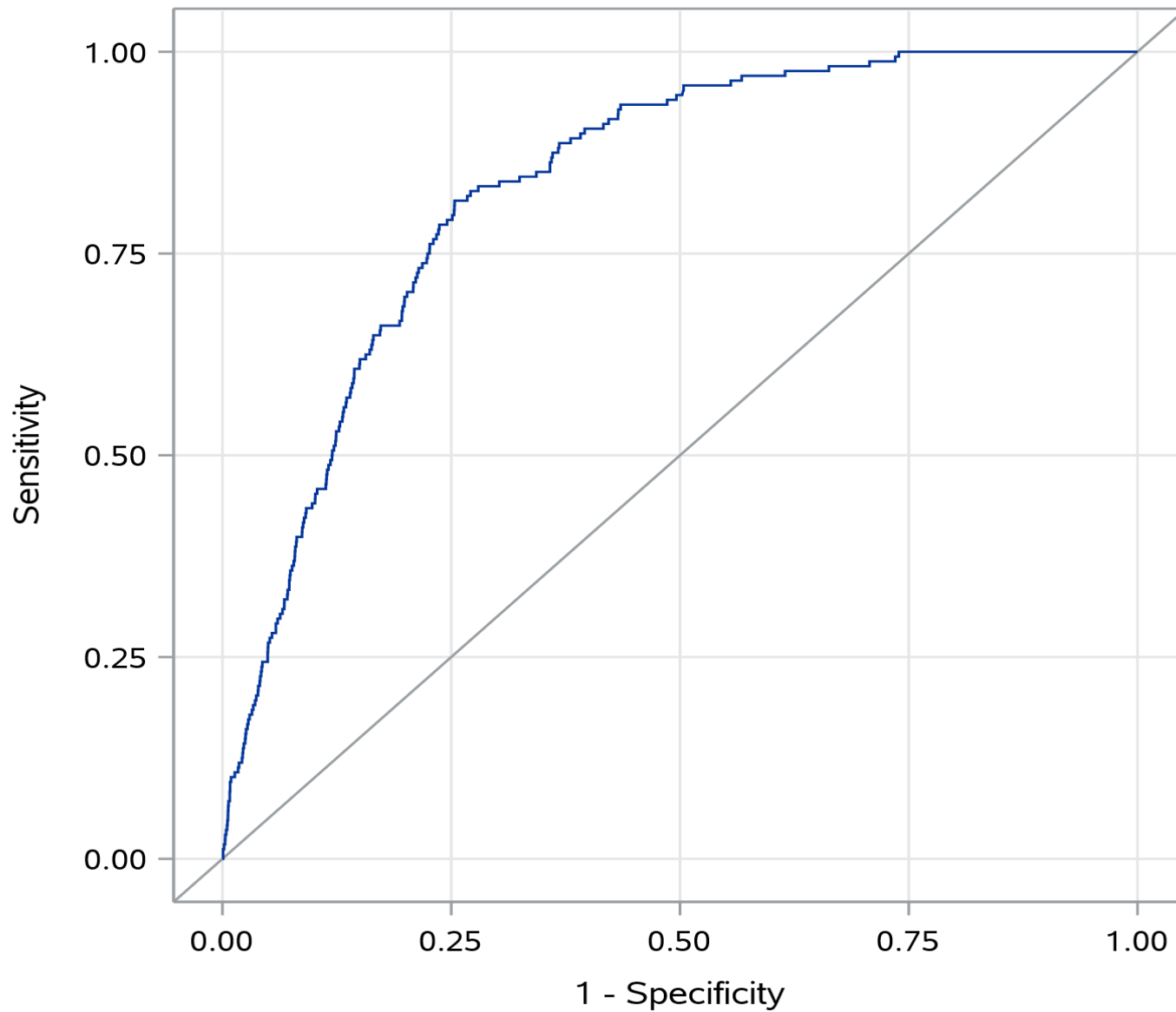
- **Analytical Base Table (ABT):** A dataset merging three domains: Product Ownership, Customer Demographics, and Marketing History.
- **Data Integrity & "Anti-Leakage":** Customers who took a loan in Q1 were excluded to ensure the model doesn't learn from future data patterns.
- **Outcome:** A total of 566 actual loan purchases were recorded in the target month.
- **Modeling Approach:** Algorithm - Binary Logistic Regression.
- *Environment:* SAS Programming Environment.
- *Validation:* Data split into 70% Training and 30% Validation sets to ensure unbiased performance testing.

MODEL PERFORMANCE: ROC CURVE

- **Analytical Base Table (ABT):** A dataset merging three domains: Product Ownership, Customer Demographics, and Marketing History.
- **Primary Metric:** C-statistic (Area Under the Curve - AUC).
- **Result:** 0.835
- **Significance:** An AUC over 0.80 is considered a "very strong" model in the banking industry.

ROC Curve for Validacija (30% podatkov)

Area Under the Curve = 0.8349

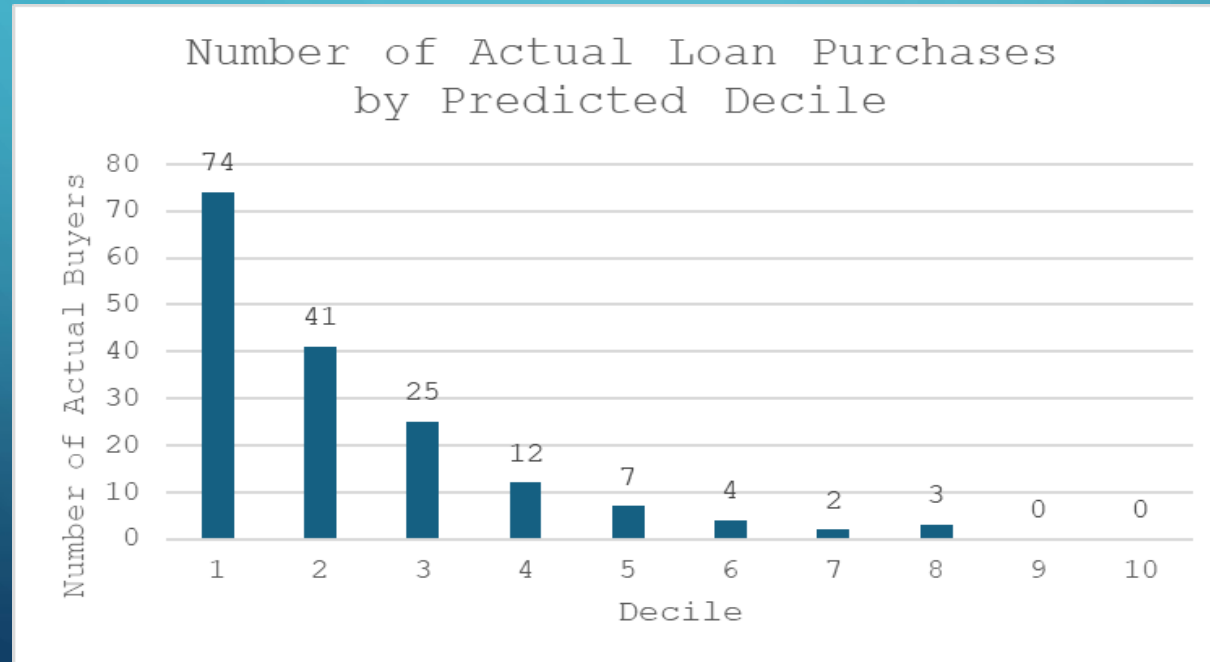


KEY PREDICTIVE VARIABLES

- **Strongest Predictor:** Mobile App Usage (Flg_App)
- **Positive Drivers:** Credit Card ownership and high ATM transaction volume
- **Negative Driver:** High Managed Deposits/Savings (Self-sufficient customers don't need loans).

DECILE ANALYSIS

- **Goal:** Rank customers by probability of purchase
- **Result:** The vast majority of actual buyers were concentrated in the top 2 deciles
- **Efficiency:** Targeting the top 10% (Decile 1) finds 74 actual buyers, while Decile 10 finds zero.



PROFILE OF THE "IDEAL CUSTOMER"

Metric (Variable)	Decile 1 (High Probability)	Decile 10 (Low Probability)
Avg. Monthly Income	2,172.20 EUR	282.60 EUR
Mobile App Usage	36.6 transactions (in 3 months)	3.5 transactions (in 3 months)
ATM Usage	18.9 transactions (in 3 months)	1.7 transactions (in 3 months)
Primary Age Group	35 – 50 years (over 57% of group)	Under 18 and Over 65
App User Ratio	98.4% (3,850 out of 3,914)	25.8% (1,010 out of 3,914)

DISCUSSION AND CONCLUSION

- **Behavior > Demographics:** Digital activity is a much stronger predictor of intent than age or gender
- **Avoiding Annoyance:** By identifying low-probability clients, the bank avoids sending spam offers to high-deposit savers.
- **Success:** Established a clear link between past digital patterns and future borrowing
- **Business Value:** Lower marketing costs and higher ROI through precise targeting
- **Future Work:** Exploring nonlinear models like Random Forests or real time triggers.



THANK YOU!