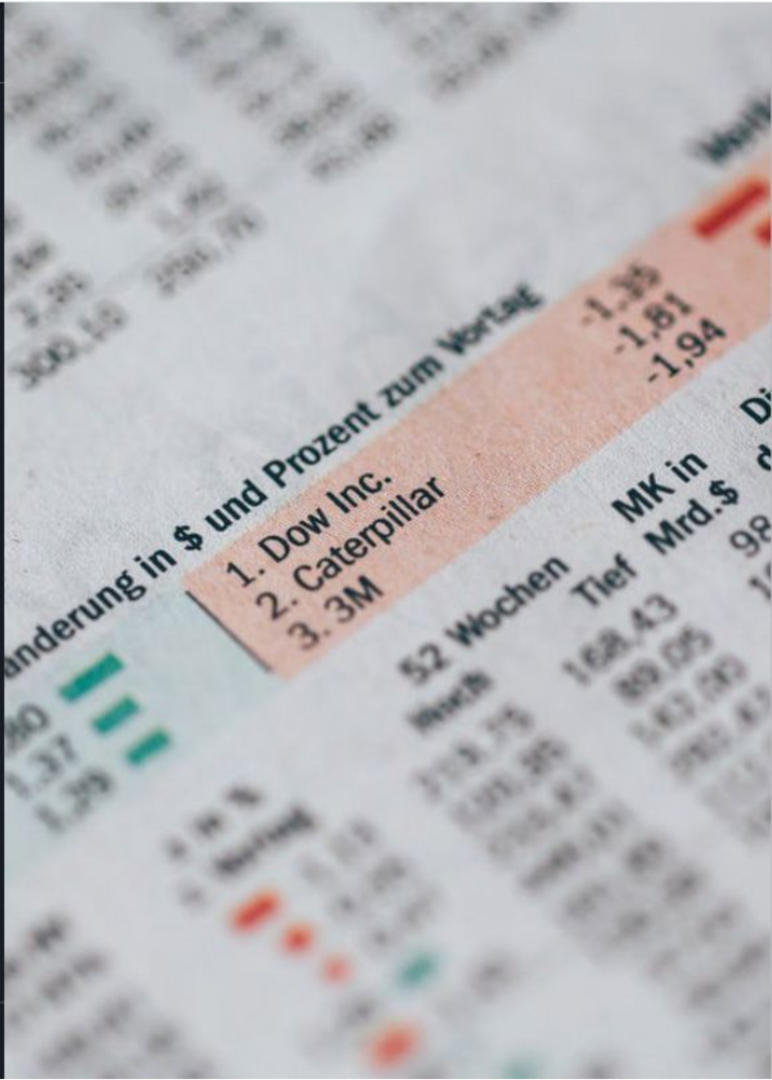
Loan Default Prediction Model

This presentation explores the application of machine learning algorithms to accurately predict loan defaults, enhancing financial decision-making and risk management.

Presenter: Lucas Tourinho Mamede



Enhancing Loan Approval Processes

Insights on Predicting Loan Defaults Effectively

- **Improved Approval Quality**

  Predicting loan defaults significantly enhances the quality of loan approvals.

- **Final Model Performance**

  The tuned Random Forest model achieved a Recall of 0.83 and AUC of 0.88, indicating high predictive accuracy.

- **Business Impact Potential**

  The implementation of this model could lead to a reduction in defaults by up to 25%, improving financial stability.

- **Next Steps Forward**

  The immediate next step involves integrating the predictive model into the current underwriting process.

Understanding Loan Default Risks

Exploring the challenges in predicting loan defaults

High Default Rate Impacts Profit

A high default rate leads to significant profit loss for financial institutions.

Predicting Default Risk

The primary goal is to predict default risk at the time of loan application.

Binary Classification Problem

This scenario is a binary classification problem, categorizing loans as BAD (0 or 1).

Key Insights on Loan Default Predictors

High DEROG, DELINQ, DEBTINC in defaulters

Defaulters show significantly high levels of derogatory marks, delinquencies, and debt-to-income ratios, indicating financial distress.

Lower YOJ and CLAGE in defaulters

Defaulters tend to have shorter years on the job (YOJ) and lower credit age (CLAGE), suggesting instability in credit history.

More inquiries (NINQ) → higher risk

An increased number of inquiries (NINQ) correlates with a higher risk of loan defaults, reflecting greater credit-seeking behavior.

Loan Default Prediction Dataset Overview

Analyzing loan defaults with key data insights

Dataset Size	Features	Target	Default Rate	Data Treatments
5960 loans	13 features	BAD	20%	Missing values, outliers, skewness

Effective Modeling Approaches for Loan Defaults

Logistic Regression

A statistical model used for binary classification, ideal for predicting loan defaults.

Decision Tree + Tuned

A model that splits data into branches to make predictions; tuning enhances its accuracy.

Random Forest + Tuned

An ensemble of decision trees that improves prediction accuracy through averaging.

Focus on Recall Metric

Recall is prioritized to effectively catch loan defaulters, ensuring fewer missed predictions.

Summary of Model Performance Metrics

Evaluating predictive models for loan defaults

Model	Recall	AUC
Logistic Regression	0.65	0.69
Decision Tree (Tuned)	0.84	0.87
Random Forest (Tuned)	0.83	0.88 ✅

Final Decision on Model Selection

Choosing the Best Model for Loan Default Prediction

Model Selection

We selected the Tuned Random Forest for its optimal balance of recall, generalization, and interpretability.

Low Overfitting

The Tuned Random Forest maintains low overfitting, making it reliable across different datasets.

Deployment Ease

One of the advantages of this model is its ease of deployment in practical applications.

High Recall

This model demonstrates high recall, ensuring that most relevant cases are identified effectively.

Feature Importance

Key features influencing predictions include DEROG, DEBTINC, and NINQ, highlighting their relevance.

Impact of Machine Learning on Loan Defaults

- **Potential reduction in defaults**

  Implementing ML could lead to a 25% drop in loan defaults, translating to substantial financial savings.

- **Financial savings estimate**

  This reduction could result in approximately $300,000 in monthly savings for the business.

- **Enhanced credit policy**

  Utilizing scores from ML models allows for better tiering of applicants, improving approval processes.

- **Flexible deployment options**

  The solution can be deployed as an API or batch system, offering flexibility in integration.

Implementation Plan and Risk Management

Strategies for Effective Model Implementation

Model Deployment

Deploy model into the approval pipeline to streamline processes.

Performance Monitoring

Monitor performance monthly to detect model drift and ensure accuracy.

Identifying Risks

Key risks include model bias and threshold tuning that could affect outcomes.

Bias Mitigation Strategy

Implement a human-in-the-loop review early to mitigate risks associated with model bias.

Thank You