

# Mentoria em Risco de Crédito e Ciência de Dados

## Resumo Técnico das Aulas 1 & 2

### Índice

- 1. [Objetivo do Material](#)
- 2. [Panorama Geral do Risco de Crédito](#)
- 3. [Principais Conceitos e Métricas](#)
- 4. [Preparação & Elegibilidade dos Dados](#)
- 5. [Construção de Targets](#)
- 6. [Amostragem & Balanceamento](#)
- 7. [LGD – Workout & Cure](#)
- 8. [Validação & Monitoramento](#)
- 9. [Boas Práticas & Armadilhas Comuns](#)
- 10. [Exercícios Propostos](#)
- 11. [Referências Essenciais](#)

## 1. Objetivo do Material

Este documento resume, de forma **estruturada e técnica**, os conteúdos sobre *Risco de Crédito* e *Ciência de Dados* apresentados nas primeiras aulas da mentoria. Ele foi depurado para:

- **Corrigir imprecisões técnicas** detectadas em aula.
- **Remover nomes próprios, empresas e comentários irrelevantes** ou sensíveis.
- Servir como **base de estudo** para alunos e como **guia de implementação** em projetos reais.

## 2. Panorama Geral do Risco de Crédito

“*Todo modelo responde a uma única pergunta.*” – Princípio adotado na mentoria.

Os modelos de risco de crédito são classificados, de forma simplificada, em:

Abreviação	Pergunta que responde	Horizonte típico
<b>PD</b> (Probability of Default)	Qual a probabilidade de o contrato entrar em <i>default</i> ?	12 m → PD12
<b>LGD</b> (Loss Given Default)	Dado o <i>default</i> , quanto será perdido?	Vida da recuperação
<b>EAD</b> (Exposure at Default)	Qual a exposição financeira no instante do <i>default</i> ?	Instantâneo

Além disso, o **IFRS 9** e a **Res. Bacen 4.966/2021** exigem estimativas de **Perda Esperada (EL)** em diferentes estágios:  $EL = PD \times LGD \times EAD$

## 3. Principais Conceitos e Métricas

### 3.1 Default & Buckets de Atraso

- **Default regulatório** (varejo): atraso  $\geq 90$  dias.
- Buckets usuais: 0-15, 16-30, 31-60, 61-90, >90 dias.

### 3.2 Flags Auxiliares

Flag	Significado	Uso principal
<b>Flag_Acordo</b>	Renegociação devedor <i>problemática</i> (concessão de benefício relevante)	Amplia o target "Mal" mesmo sem 90 dias de atraso
<b>Flag_Óbito</b>	Óbito sem seguro prestamista	Pode antecipar <i>default</i>
<b>Spell_ID</b>	"Vida" do contrato após curing	Permite múltiplas vidas no painel

### 3.3 Targets de PD

Notação	Definição (janelada)	Sensibilidade
<b>Over90 M12</b>	Atraso $\geq 90$ d <b>no último</b> mês da janela de 12 m	Baixa (captura inadimplências persistentes)
<b>Ever90 M12</b>	Atraso $\geq 90$ d em <b>qualquer</b> mês da janela	Alta
<b>MAL</b>	<b>Ever 90 M12 OR Flag_Acordo</b>	Abrange acordos ruins

### 3.4 Workout & Cure (para LGD)

1. **Default**: atraso  $\geq 90$  d (ou ativo problemático).
2. **Workout**: fase em que há **esforço de cobrança** e/ou pagamentos parciais.
3. **Cure**: período *sem atraso* após o workout.
  - Prática de mercado:  $\geq 3$  meses (varejo); pode ser maior em atacado.

## 4. Preparação & Elegibilidade dos Dados

### 4.1 Grupo Perform por Modelo

Modelo	Registros elegíveis
<b>PD</b>	Contratos <b>sem default</b> atual nem <b>Flag_Acordo</b> .
<b>LGD</b>	Registros <b>em default</b> + <b>Workout</b> (exclui <i>Cure</i> ).
<b>EAD</b>	Mesmos de PD + primeiro registro em <i>default</i> .

### 4.2 Teste de Continuidade

1. **Detectar quebras** de histórico (*gaps*) por contrato.
2. Classificar impacto: **imaterial** (<1 %) vs **material**.
3. Ações: *fill-forward*, exclusão pontual ou solicitar correção na base-fonte.

#### 4.3 Tratamento de "Re-aging" (zeragem de atraso após acordo)

- **Proibido** zerar *days past due* sem período de *Cure*.
- Manter vínculo via **ID\_Contrato\_Original**.

## 5. Construção de Targets

1. **Criar coluna** \*\*\*\*\***Over\_90**: 1 se **DPD** ≥ 90, caso contrário 0.
2. \*\*Derivar \*\***Ever\_90\_M12** com *rolling window* de 12 meses.
3. **Montar** \*\*\*\*\***MAL**: **MAL** = (**Ever\_90\_M12** == 1) or (**Flag\_Acordo** == 1).
4. \*\*Gerar \*\***Target\_PD** (Ever ou Over) conforme requisito regulatório.

Exemplo (Python pandas)

```
# suposição: df ordenado por ID_Contrato & Data_Referencia
df['Over_90'] = (df['dpd'] >= 90).astype(int)
# janela móvel de 12 linhas (mensal)
df['Ever_90_M12'] = (
    df.groupby('id_contrato')['Over_90']
    .transform(lambda x: x.rolling(12, min_periods=1).max())
)
# MAL
df['MAL'] = np.where((df['Ever_90_M12'] == 1) | (df['Flag_Acordo'] == 1), 1, 0)
```

## 6. Amostragem & Balanceamento

### 6.1 Painel × Snapshot

<b>Critério</b>	<b>Painel (filme)</b>	<b>Snapshot (foto)</b>
Modelo indicado	Behavioral (PD, LGD)	Concessão (Application Score)
Vantagens	Mantém variáveis históricas	Simples & leve
Riscos	<i>Overfitting</i> em contratos longos	<b>Viés de recência</b> e perda de história

**Recomendação** da mentoria: usar **Painel** e só recorrer a Snapshot para datasets enormes ou uso exploratório.

### 6.2 Técnicas de Balanceamento

- **Sample Weights** (preferencial)
  - Ponderar a classe minoritária (**MAL** = 1).
  - Permite granularidade (e.g., peso por *recência*).

- **Class Weight** = **"balanced"** (rápido, porém menos flexível).
- **Reamostragem** (undersampling/oversampling)
  - Usar com moderação; evita-se *SMOTE* para risco de crédito.

**Dica:** Otimize hiper-parâmetros e pesos de classe conjuntamente (e.g., *Optuna*).

## 7. LGD – Workout & Cure

### 7.1 Definindo o Período de Workout

1. Traçar a **curva cumulativa de recuperação** pós-default.
2. Identificar o **plateau** ( $\Delta < 1\%$  por  $\geq 3$  meses).
3. Fixar **horizonte LGD** até esse ponto.

### 7.2 Estimativa de LGD

$$LGD_t = 1 - \frac{\text{Recuperação Acumulada}_t}{\text{EAD}_0}$$

- **LGD final:** valor no término do workout.
- **Modelagem:** regressão Beta ou Gradient Boosting com logit.

## 8. Validação & Monitoramento

Métrica	Uso	Referência
<b>KS (Kolmogorov–Smirnov)</b>	Discriminante entre <i>good vs bad</i>	$\geq 0.30$ (varejo)
<b>AUC / Gini</b>	Rankeamento	$Gini = 2 \cdot AUC - 1$
<b>PSI</b> (Population Stability)	Drift de entrada	$\leq 0.10$ (aceitável)
<b>Brier Score</b>	Calibração PD	Quanto menor melhor

Procedimento de *Backtesting* (ex.: Res. 4.966)

1. Congelar parâmetros.
2. Projetar 12 m.
3. Comparar PD-real vs PD-estimado.

## 9. Boas Práticas & Armadilhas Comuns

- **Documentação** minuciosa do *data lineage*.
- **Não “zerar” DPD** após renegociação sem *Cure*.
- **Evitar SMOTE** e geradores sintéticos em risco de crédito.
- **Sempre** separar *train / validation / test* em corte temporal (evitar *data leakage*).
- **Monitorar** entradas & saídas do modelo mensalmente.

## 10. Exercícios Propostos

1. **Teste de Continuidade:** identifique contratos com *gaps* > 1 m e proponha correção.
  2. **\*\*Construção do \*\*MAU em Python usando a base sintética fornecida.**
  3. **Painel vs Snapshot:** compare AUC e PSI de ambos os métodos em 3 horizontes.
  4. **Workout Analyzer:** plote curva de recuperação e determine *t\_cure*.
- 

## 11. Referências Essenciais

- Basel Committee on Banking Supervision – *IRB Approach: Guidance on PD, LGD & EAD*.
  - European Banking Authority – *Guidelines on PD Estimation, LGD Estimation & Treatment of Defaulted Exposures* (EBA/GL/2017/16).
  - Resolução CMN 4.966/2021 – *Modelos Internos de Risco de Crédito*.
  - *Credit Risk Modelling* – Körner & Zimmermann (2022).
  - *Elements of Statistical Learning* – Hastie, Tibshirani & Friedman (chap. 12).
- 

© 2025 RiskPilot Mentoria em Risco de Crédito – Conteúdo de uso exclusivo dos alunos.