



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Mineração de dados do portal IF.data do Bacen  
relativos às Instituições Financeiras que operam no  
Brasil**

Hudson Silva Alves

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Orientador  
Prof. Dr. Jan Mendonça Corrêa

Brasília  
2021



# Dedicatória

Dedico este trabalho à universidade pública, pois sem a existência dela eu nunca teria acesso ao conhecimento que obtive ao longo de minha graduação e nem aos recursos de aprendizado que pude aproveitar. Que a Universidade de Brasília e as demais universidades públicas e institutos federais do Brasil continuem oferecendo oportunidades de formação à população.

# Agradecimentos

Agradeço primeiramente à minha companheira Emanuelle, que foi minha maior incentivadora e parceira, não apenas ao longo da realização desse trabalho, mas de toda a minha jornada acadêmica e profissional. Agradeço aos meus pais Alessandra e Vilmon que sempre se esforçaram para que eu pudesse ter uma educação de qualidade e me deram todo o suporte que eu precisava para conseguir ingressar na universidade. Agradeço ao meu orientador Jan Mendonça Corrêa, que me auxiliou na realização deste trabalho e, por fim, agradeço aos demais professores do CIC e de outros departamentos que contribuíram para a minha formação.

# Resumo

Trimestralmente, o Banco Central do Brasil divulga alguns relatórios contendo diversos dados de instituições financeiras através do Portal IF.data. Esses dados, apesar de sua riqueza, precisam ser tratados e explorados para a geração de informações não triviais. Deste modo, o presente trabalho tem o propósito de aplicar técnicas de mineração de dados aos dados do Portal IF.data do Banco de Central do Brasil de modo a obter a previsão dos valores de inadimplência total das instituições financeiras que operam no Brasil.

**Palavras-chave:** Mineração de Dados, Sistema Financeiro Nacional, Banco Central do Brasil, Instituições Financeiras, Inadimplência

# Abstract

On a quarterly basis, the Central Bank of Brazil publishes some reports containing various information from financial institutions through the IF.data Portal. These data, despite their wealth, must be treated and explored to generate non-trivial information. Thus, the present work has the purpose of applying data mining techniques to the data of the IF.data Portal of the Central Bank of Brazil in order to obtain the forecast of the total default values of financial institutions operating in Brazil.

**Keywords:** Data Mining, Brazilian National Financial System, Central Bank of Brazil, Financial Institutions, Default

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivo Geral . . . . .	2
1.3	Objetivos Específicos . . . . .	2
1.4	Metodologia . . . . .	3
1.5	Organização do Trabalho . . . . .	3
<b>2</b>	<b>O Sistema Financeiro Nacional</b>	<b>4</b>
2.1	Organização do SFN . . . . .	4
2.2	O Banco Central do Brasil . . . . .	5
2.3	Portal IF.Data . . . . .	6
<b>3</b>	<b>Mineração e Análise de Dados</b>	<b>8</b>
3.1	Dado, informação e conhecimento . . . . .	8
3.2	Mineração de Dados . . . . .	9
3.2.1	CRISP-DM . . . . .	9
3.2.2	Técnicas para Mineração de Dados . . . . .	11
3.2.2.1	Regressão Linear . . . . .	12
3.2.2.2	Clusterização . . . . .	13
3.3	Python . . . . .	13
3.4	PostgreSQL . . . . .	14
3.5	Weka . . . . .	14
3.6	Orange . . . . .	15
<b>4</b>	<b>Estudo de Caso: Dados de Instituições Financeiras que operam no Brasil</b>	<b>16</b>
4.1	Entendimento do Negócio . . . . .	16
4.2	Coleta e Entendimento dos Dados . . . . .	17
4.3	Preparação dos Dados . . . . .	19
4.3.1	Tratamentos realizados . . . . .	19

4.3.2	Carga no Banco de Dados . . . . .	21
4.4	Previsão da Inadimplência Total . . . . .	22
4.4.1	Importação dos dados no Weka . . . . .	22
4.4.2	Geração do Modelo Preditivo . . . . .	22
4.4.2.1	Banco do Brasil . . . . .	25
4.4.2.2	Banco Nacional de Desenvolvimento Econômico e Social . . . . .	29
4.4.2.3	Bradesco . . . . .	32
4.4.2.4	Caixa Econômica Federal . . . . .	34
4.4.2.5	Banco CSF . . . . .	37
4.4.2.6	Itaú . . . . .	40
4.4.2.7	Nubank . . . . .	43
4.4.2.8	Banco Pan . . . . .	45
4.4.2.9	Santander . . . . .	48
4.4.2.10	Banco Votorantim . . . . .	51
4.4.3	Comparação entre as Instituições Financeiras . . . . .	54
<b>5</b>	<b>Discussão dos Resultados</b>	<b>56</b>
5.1	Previsão da Inadimplência . . . . .	56
5.1.1	Comparação com os dados reais . . . . .	56
5.1.1.1	Trimestre encerrado em Junho de 2020 . . . . .	57
5.1.1.2	Trimestres encerrados em Junho de 2017, Junho de 2018 e Junho de 2019 . . . . .	58
5.1.2	Comparação com a média simples . . . . .	60
5.2	Clusterização . . . . .	61
<b>6</b>	<b>Considerações Finais</b>	<b>64</b>
	<b>Referências</b>	<b>66</b>
	<b>Anexo</b>	<b>67</b>
<b>I</b>	<b>Script de importação e tratamento dos arquivos CSV</b>	<b>68</b>
<b>II</b>	<b>Relação de abreviaturas utilizadas para as colunas da tabela</b>	<b>75</b>
<b>III</b>	<b>Modelos preditivos gerados com o Weka</b>	<b>81</b>
<b>IV</b>	<b>Script de geração dos gráficos de inadimplência total</b>	<b>94</b>



# Lista de Figuras

2.1	Composição e segmentos do Sistema Financeiro Nacional. . . . .	5
3.1	Hierarquia entre Dado, Informação e Conhecimento. . . . .	9
3.2	O ciclo de vida da mineração de dados. . . . .	10
3.3	Exemplo de Dendograma. . . . .	14
4.1	Tela Inicial do Portal IF.data. . . . .	17
4.2	Download do arquivo em formato .csv no Portal IF.data. . . . .	18
4.3	Exemplo de arquivo em formato .csv baixado do Portal IF.data. . . . .	18
4.4	Dataframe gerado após a importação dos arquivos CSV. . . . .	20
4.5	Criação do Banco de Dados na ferramenta pgAdmin. . . . .	21
4.6	Modificação realizada no arquivo Weka.ini. . . . .	22
4.7	Modificação realizada no arquivo DatabaseUtils.props.postgresql. . . . .	22
4.8	Opção Open DB do Weka. . . . .	23
4.9	Janela do Weka após internalização dos dados. . . . .	23
4.10	Configurações básicas - Weka Forecast. . . . .	25
4.11	Configurações avançadas - Weka Forecast. . . . .	26
4.12	Métricas de erro do modelo preditivo para o BB. . . . .	27
4.13	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - BB. . . . .	27
4.14	Valores previstos para os trimestre de Junho de 2017, Junho de 2018 e Junho de 2019 - BB. . . . .	28
4.15	BB - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	29
4.16	Métricas de erro do modelo preditivo para o BNDES. . . . .	30
4.17	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - BNDES. . . . .	30
4.18	Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - BNDES. . . . .	31

4.19 BNDES - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	31
4.20 Métricas de erro do modelo preditivo para o Bradesco. . . . .	32
4.21 Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Bradesco. . . . .	33
4.22 Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Bradesco. . . . .	34
4.23 Bradesco - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	34
4.24 Métricas de erro do modelo preditivo para a CEF. . . . .	35
4.25 Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - CEF. . . . .	36
4.26 Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - CEF. . . . .	37
4.27 CEF - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	37
4.28 Métricas de erro do modelo preditivo para o Banco CSF. . . . .	38
4.29 Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco CSF. . . . .	39
4.30 Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco CSF. . . . .	40
4.31 Banco CSF - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	40
4.32 Métricas de erro do modelo preditivo para o Itaú. . . . .	41
4.33 Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Itaú. . . . .	42
4.34 Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Itaú. . . . .	43
4.35 Itaú - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	43
4.36 Métricas de erro do modelo preditivo para o Nubank. . . . .	44

4.37	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Nubank. . . . .	45
4.38	Valores previstos para os trimestres Junho de 2018 e Junho de 2019 - Nubank.	45
4.39	Nubank - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	46
4.40	Métricas de erro do modelo preditivo para o Banco Pan. . . . .	46
4.41	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco Pan. . . . .	47
4.42	Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco Pan. . . . .	48
4.43	Banco Pan - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	48
4.44	Métricas de erro do modelo preditivo para o Santander. . . . .	49
4.45	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Santander. . . . .	50
4.46	Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Santander. . . . .	51
4.47	Santander - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	51
4.48	Métricas de erro do modelo preditivo para o Banco Votorantim. . . . .	52
4.49	Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco Votorantim. . . . .	53
4.50	Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco Botorantim. . . . .	54
4.51	Banco Votorantim - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo. . . . .	54
4.52	Evolução da inadimplência total das 10 IFs entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com os modelos. . . . .	55
5.1	Clusterização - Processo de geração dos dendogramas usando o Orange. . .	61
5.2	Dendograma 1 - 10 instituições financeiras selecionadas. . . . .	62
5.3	Dendograma 2 - Todas as instituições financeiras selecionadas. . . . .	63

# Lista de Tabelas

5.1	Comparação dos valores previstos com os valores reais em milhões de R\$ - Trimestre Junho de 2020. . . . .	57
5.2	Diferenças percentuais entre os valores previstos e os valores reais - Trimestres Junho de 2017, Junho de 2018, Junho de 2019 e Junho de 2020. . . . .	59
5.3	Diferenças percentuais entre os valores previstos para o trimestre encerrado em junho de 2020 e a média dos valores dos últimos 3 anos (2017, 2018 e 2019). . . . .	60

# Lista de Abreviaturas e Siglas

**Bacen** Banco Central do Brasil.

**BB** Banco do Brasil.

**BNDES** Banco Nacional de Desenvolvimento Econômico e Social.

**CEF** Caixa Econômica Federal.

**Cosif** Plano Contábil de Instituições Financeiras.

**CRISP-DM** Cross Industry Standard Process for Data Mining.

**DLO** Demonstrativo de Limites Operacionais.

**SCR** Sistema de Informações de Crédito do Banco Central do Brasil.

**SFN** Sistema Financeiro Nacional.

**SGBD** Sistema Gerenciador de Bancos de Dados.

**SQL** Structured Query Language.

# Capítulo 1

## Introdução

Atualmente, o uso de dados eletrônicos é amplamente difundido no mundo, considerada a alta conectividade na qual muitas pessoas estão inseridas. Segundo Davenport [1], no ano de 2012 foram consumidos globalmente mais de 2,8 zettabytes em dados, isso representa 2,8 trilhões de gigabytes. Mesmo assim, apenas 0,5% desses dados são devidamente analisados. Isso ocorre já que muitas vezes os dados não estão disponíveis em formatos compreensíveis, e se não forem utilizadas técnicas e ferramentas adequadas para seu manuseio não é possível extrair o seu real valor.

No sistema financeiro brasileiro não é diferente. Um estudo da Febraban - Federação Brasileira de Bancos [2] indica que em 2019 foram realizadas 89,9 bilhões de transações bancárias, e desse total, 19% foram realizadas por computador e 44% por celular. Todas essas transações feitas através de canais digitais acabam gerando vários dados que são utilizados pelos bancos. Além disso, com a finalidade de atender a diversos requisitos legais, os bancos enviam alguns documentos ao Banco Central do Brasil, contendo dados variados sobre os seus resultados, sua carteira de crédito, dentre outros.

De forma a disponibilizar parte desses dados ao público em geral, o Bacen mantém duas iniciativas: o Portal de Dados Abertos e o Portal IF.data. O Portal de Dados Abertos é um repositório de dados do sistema financeiro, que concentra diversos tipos de dados, desde estatísticas monetárias, até dados de correspondentes bancários no país. Já o Portal IF.data detém alguns dos dados enviados pelos bancos e instituições financeiras do país com fins regulatórios, como os dados do Sistema de Informações de Crédito do Banco Central do Brasil e do Demonstrativo de Limites Operacionais.

### 1.1 Motivação

Apesar de esses dados serem disponibilizados pelo Bacen para utilização do público geral, na forma em que se encontram é preciso que haja a realização de uma série de tratamentos

e utilização de técnicas para que se possa extrair alguma informação mais complexa deles. Desse modo, a aplicação de algoritmos e técnicas de tratamento e mineração de dados pode ser útil para extração de informações mais complexas dos dados disponibilizados, e essas informações podem permitir a realização de análises mais aprofundadas sobre o sistema financeiro brasileiro.

Desta maneira, optou-se pela realização da previsão da inadimplência total das instituições financeiras, tendo em vista sua relevância no cenário econômico. A inadimplência pode ser utilizada para avaliar a saúde financeira dos tomadores de empréstimos e das instituições financeiras. Índices elevados de inadimplência servem como alerta para melhorar o gerenciamento de risco de empréstimos e financiamentos e também podem ser indicadores de uma crise econômica, por exemplo.

A previsão desses valores pode auxiliar na avaliação de cenários futuros e contribuir para tomada de decisões nas instituições financeiras, nos reguladores do Sistema Financeiro Nacional e demais agentes públicos e nas empresas de maneira geral. Sendo assim, a previsão da inadimplência total das instituições financeiras se mostra como uma informação de grande valor que pode ser extraída dos dados disponibilizados pelo Bacen no portal IF.data.

## 1.2 Objetivo Geral

O trabalho tem como objetivo realizar a previsão da inadimplência das instituições financeiras que operam no Brasil, para o trimestre encerrado em junho/2020, através da mineração e análise dos dados disponibilizados no portal IF.data do Banco Central do Brasil.

## 1.3 Objetivos Específicos

Para alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- Realizar a importação e tratamento dos dados de instituições financeiras no portal IF.data do Banco Central do Brasil;
- Construir uma previsão dos dados de inadimplência da carteira de crédito das principais Instituições Financeiras para o trimestre encerrado em Junho/2020 através da utilização de algoritmo de predição e tendo a série histórica como base;
- Realizar comparação da previsão construída com os dados reais divulgados para o trimestre de Junho/2020;

- Identificar outras relações relevantes através da aplicação de técnicas de mineração de dados.

## 1.4 Metodologia

Para alcançar os objetivos propostos, a realização do trabalho consistiu nas seguintes etapas:

- Elaboração de referencial teórico dividido em dois grupos, sendo o primeiro grupo referente à organização do Sistema Financeiro Nacional e ao papel do Banco Central do Brasil, principalmente quanto à disponibilização das informações das instituições financeiras que atuam no Brasil. Já o segundo grupo trata de conceitos relacionados à análise e mineração dados e às ferramentas que serão utilizadas para realização do trabalho;
- Obtenção dos dados através do portal IF.data do Bacen, importação dos dados em banco de dados utilizando o SGBD *PostgreSQL* [3] e tratamento dos dados de modo a prepará-los para análise e mineração utilizando a linguagem *Python* [4];
- Realização da mineração e análise dos dados, utilizando algoritmos de predição com a ferramenta *Weka* [5] e de clusterização com a ferramenta *Orange* [6].

## 1.5 Organização do Trabalho

O presente texto foi dividido em quatro capítulos, da seguinte forma:

No Capítulo 2 será apresentado um referencial teórico sobre o SFN, incluindo informações sobre o papel do Banco Central do Brasil e sua atuação e sobre o Sistema de Informações de Crédito do Banco Central do Brasil, além de detalhar o portal IF.data, de onde os dados a serem analisados serão obtidos.

O Capítulo 3 irá expor um referencial teórico sobre análise de dados, incluindo informações gerais sobre mineração de dados e sobre as técnicas de mineração que foram adotadas. Também será incluído neste capítulo informações sobre as ferramentas que serão utilizadas no trabalho: *Python*, *PostgreSQL*, *Weka* e *Orange*.

No Capítulo 4, haverá um estudo de caso contendo a descrição dos dados analisados e os resultados da análise descritiva, da aplicação do modelo preditivo e da aplicação das técnicas de mineração. Além dos comparativos entre os dados previstos e os dados reais divulgados.

Por fim, o capítulo 5 será uma conclusão do trabalho e discussão sobre os resultados obtidos.



# Capítulo 2

## O Sistema Financeiro Nacional

Este capítulo tem por objetivo descrever o Sistema Financeiro Nacional (SFN), abordando principalmente sua estrutura e organização, com foco na atuação do Banco Central do Brasil (Bacen) [7]. Esse levantamento é importante uma vez que as instituições financeiras que atuam no Brasil, objeto de análise do trabalho, estão inseridas no ambiente do SFN. Dessa forma, é essencial a compreensão desse sistema como um todo.

A primeira parte abordará a organização do SFN, citando seus principais componentes e falando um pouco sobre o seu papel. A segunda parte trará um foco sobre um dos componentes do SFN, que é o Bacen, responsável pela fiscalização das instituições financeiras do Brasil. Já as duas últimas partes, falarão um pouco mais sobre dois instrumentos do Bacen que são de suma importância ao presente trabalho, o Sistema de Informações de Crédito do Banco Central do Brasil (SCR) e o Portal IF.Data.

Essas descrições irão contribuir para que o leitor fique mais familiarizado com o objeto de análise e com a fonte dos dados utilizada para a realização do trabalho, facilitando a compreensão dos passos executados e dos resultados obtidos.

### 2.1 Organização do SFN

O Sistema Financeiro Nacional (SFN) é um dos pilares da economia brasileira, estando previsto no artigo 192 da Constituição da República Federativa do Brasil de 1988. Ele foi "estruturado de forma a promover o desenvolvimento equilibrado do País e a servir aos interesses da coletividade" [8]. Atualmente, ele está organizado em três grandes blocos: órgãos normativos, supervisores e operadores, conforme ilustrado na Figura 2.1.

Os órgãos normativos são aqueles que exercem a política monetária nacional com o comando do governo federal. Isso é realizado através da publicação de normas, resoluções e outros dispositivos legais. Entre eles está o Conselho Monetário Nacional (CMN), que tem função normativa em relação à moeda, crédito, capitais e câmbio, o Conselho Nacional

	Moeda, crédito, capitais e câmbio		Seguros Privados	Previdência Fechada
<b>Órgãos normativos</b>	CMN (Conselho Monetário Nacional)		CNSP (Conselho Nacional de Seguros Privados)	CNPC (Conselho Nacional de Previdência Complementar)
<b>Supervisores</b>	Bacen (Banco Central do Brasil)	CVM (Comissão de Valores Mobiliários)	Susep (Superintendência de Seguros Privados)	Previc (Superintendência Nacional de Previdência Complementar)
<b>Operadores</b>	Bancos e Caixas Econômicas	Administradoras de Consórcios	Bolsa de Valores	Entidades fechadas de Previdência Complementar (Fundos de Pensão)
	Cooperativas de Crédito	Corretoras e Distribuidoras	Bolsa de mercadorias e futuros	
	Instituições de Pagamento	Demais Instituições não Bancárias		

Figura 2.1: Composição e segmentos do Sistema Financeiro Nacional (Baseado em: [9]).

de Seguros Privados (CNSP), responsável pelos seguros privados e o Conselho Nacional de Previdência Complementar (CNPC), que dita as regras sobre previdência fechada.

Os supervisores são aqueles que tem o papel de fiscalizar a execução das normas que são criadas pelos órgãos normativos. Eles acompanham o funcionamento dos operadores para garantir que os dispositivos legais estão sendo seguidos corretamente. A Superintendência Nacional de Previdência Complementar (Previc) exerce esse papel em relação às entidades fechadas de previdência complementar, a Superintendência de Seguros Privados (Susep) quanto aos seguros privados, a Comissão de Valores Mobiliários (CVM) no que diz respeito ao mercado de capitais e o Bacen supervisiona as normas sobre moeda, crédito e câmbio [10].

Já os operadores são as instituições financeiras ou não, que atuam dentro do escopo do SFN, como os bancos e caixas econômicas, administradoras de consórcio, seguradoras etc. Essa é a parte do SFN que está mais próxima dos clientes finais, que são os consumidores dos produtos financeiros.

## 2.2 O Banco Central do Brasil

O Banco Central do Brasil (Bacen) foi criado em 1964 pela Lei 4595 [11] e, como levantado na seção anterior, é um dos componentes do SFN que atua na fiscalização da execução

das normas. Além disso, desempenha algumas outras funções no sentido de zelar pelo poder de compra da moeda brasileira e por manter a solidez do SFN [10].

Dentre as suas atividades mais importantes está a fiscalização e acompanhamento das instituições financeiras. Essa tarefa é feita em grande parte através do recebimento e verificação de documentos disponibilizados pelas instituições. Alguns exemplos desses documentos são: o Demonstrativo do Plano Contábil de Instituições Financeiras (Cosif), o Demonstrativo de Limites Operacionais (DLO) e o Detalhamento do Sistema de Informações de Crédito do Banco Central do Brasil (SCR).

O Cosif contém dados da divulgação de resultados das instituições e os valores que são alocados em suas contas principais, através dele é possível a identificação do lucro, das despesas e do patrimônio das instituições [12]. O DLO concentra dados sobre alocação de capital, trazendo um levantamento sobre os ativos das instituições e a utilização de seus recursos.

Já o SCR é um sistema que registra diversas informações sobre empréstimos, financiamentos e outras operações de crédito, de maneira que o Bacen consegue acompanhar, por exemplo, o volume de operações contratadas pelas instituições financeiras, a quantidade de operações em dia e inadimplentes, a região geográfica onde as operações são contratadas e diversos outros dados. Ele também pode ser utilizado pelos consumidores das operações, de forma a acompanharem seu endividamento junto às instituições que fazem parte do SFN [13].

Tendo em vista a riqueza de dados que esses documentos possuem, o Bacen faz a disponibilização de alguns relatórios baseados neles, de modo que usuários externos consigam utilizá-los de maneira agregada, mantendo-se o sigilo bancário dos consumidores. Uma das formas em que esses relatórios são disponibilizados é através do portal IF.data.

## 2.3 Portal IF.Data

O Portal IF.Data [14] é um portal que foi criado pelo Bacen com o objetivo de atender ao disposto na Lei 12.527 de 2011, popularmente conhecida como Lei de Acesso à Informação. Com a promulgação dessa lei "o Poder Público passou a ter o dever de não apenas publicar as informações na internet, mas de disponibilizá-las em formatos abertos e processáveis por máquina" [15].

Neste contexto, ele foi criado como um portal interativo que armazena um compilado de relatórios com dados abertos sobre as instituições financeiras. Esses relatórios são divulgados trimestralmente e podem ser visualizados em tela ou baixados em formato *.csv*.

Dentre os disponíveis no portal, estão os dados de Informações Contábeis e de Capital, Informações de Crédito, Informações de Câmbio e Segmentação de instituições financeiras. Embora esses dados tenham passado a ser disponibilizados nos últimos anos, o portal contém dados mais antigos. O formato atualmente utilizado, abrange dados do ano 2000 em diante, mas também, é possível obter dados de anos anteriores, a partir de 1994 até 2000, em formatos específicos. Para a visualização ou *download* dos dados, o usuário pode selecionar na tela inicial do portal o trimestre referência, o tipo de instituição e o tipo de relatório que deseja obter.

A riqueza dos dados presentes nesses relatórios, a clareza da documentação existente e a disponibilidade de uma série histórica contendo vários períodos, foram os fatores que levaram à escolha do Portal IF.data como fonte de dados para realização do presente trabalho.

# Capítulo 3

## Mineração e Análise de Dados

Neste capítulo será apresentado um breve levantamento sobre conceitos utilizados ao longo do trabalho, como dado, informação e conhecimento. Esse enfoque visa elucidar os termos ao leitor, para evitar interpretações equivocadas.

Além disso, o capítulo também abordará de maneira descritiva a metodologia e as ferramentas utilizadas para a realização do trabalho. Desta forma, os próximos itens do capítulo tratarão das técnicas e algoritmos de mineração de dados aplicados na análise, bem como sobre as ferramentas *Python*, *PostgreSQL*, *Weka* e *Orange*, que foram utilizadas para obtenção dos resultados.

### 3.1 Dado, informação e conhecimento

A compreensão dos conceitos de dado, informação e conhecimento antecede a realização de atividades de mineração e análise de dados. Esses são conceitos fundamentais para o entendimento das etapas de mineração e para possibilitar que os resultados pretendidos possam ser alcançados. Amaral [16] explica que "dados são fatos coletados e normalmente armazenados. Informação é o dado analisado e com algum significado. O conhecimento é a informação interpretada, entendida e aplicada para um fim".

Já Goldschmidt e Passos [17], entendem que dado, informação e conhecimento fazem parte de uma hierarquia que pode ser representada por uma pirâmide, visualizada na Figura 3.1, onde é considerado o exemplo de uma base de dados de uma financeira que armazena dados de seus clientes.

O dado, na base da pirâmide, é o elemento mais estrutural, representado no exemplo pela renda mensal e despesas mensais. A informação, que ocupa o meio da pirâmide, é o resultado do processamento dos dados, e está localizada em um cenário bem determinado. No exemplo em pauta ela é representada pela capacidade mensal de endividamento. Por fim, o conhecimento, que ocupa o topo da pirâmide, se trata de uma interpretação das

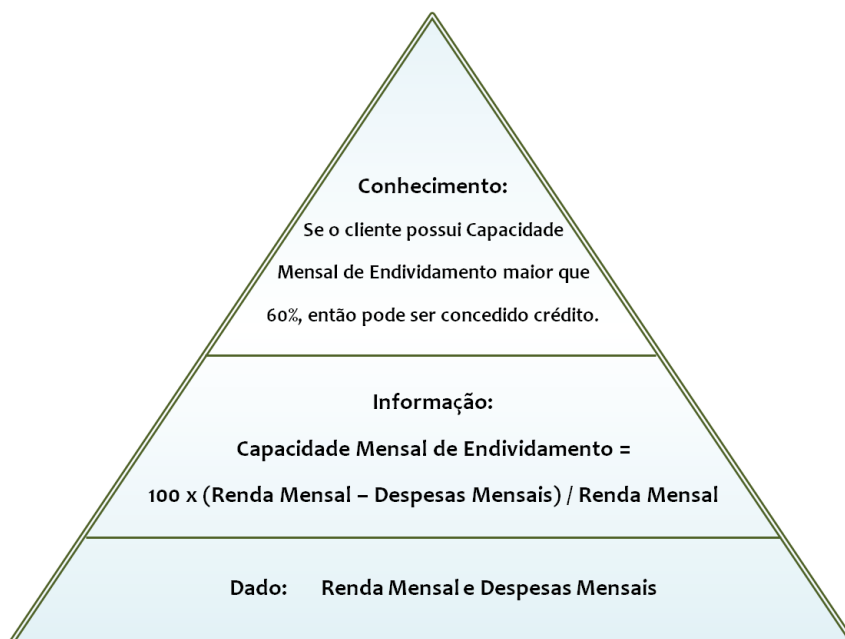


Figura 3.1: Hierarquia entre Dado, Informação e Conhecimento (Baseado em: [17]).

informações em conjunto com os dados, retratada na Figura 3.1 como uma regra para decidir se um cliente pode ou não fazer uso de crédito.

## 3.2 Mineração de Dados

Alguns autores defendem que mineração de dados é um sinônimo do termo *Knowledge Discovery in Databases* (KDD) ou descoberta de conhecimento em bancos de dados, já outros autores afirmam que a mineração de dados é apenas uma etapa da descoberta de conhecimento [18]. Neste trabalho, o conceito de mineração de dados adotado é o seguinte: mineração de dados é o processo de análise de conjuntos de dados com a finalidade de se obter informações e conhecimento, e encontrar relacionamentos não triviais [19].

Para a realização desse processo é preciso conduzir diversas etapas, que vão desde o entendimento do contexto em que os dados estão inseridos até a realização de testes para validação dos resultados obtidos. Existem algumas metodologias que facilitam a realização desse ciclo, separando as etapas a serem efetuadas e permitindo melhor organização da análise.

### 3.2.1 CRISP-DM

No presente trabalho, a organização das etapas do processo de mineração de dados foi realizada através da metodologia CRISP-DM. *Cross Industry Standard Process for Data*

*Mining* (CRISP-DM) [20] é uma metodologia utilizada para orientação de processos de mineração de dados que "inclui descrições das fases típicas de um projeto, as tarefas envolvidas em cada fase e uma explicação dos relacionamentos entre essas tarefas" [20]. Essa metodologia divide o processo de mineração de dados em 6 etapas, de acordo com o observado na Figura 3.2.

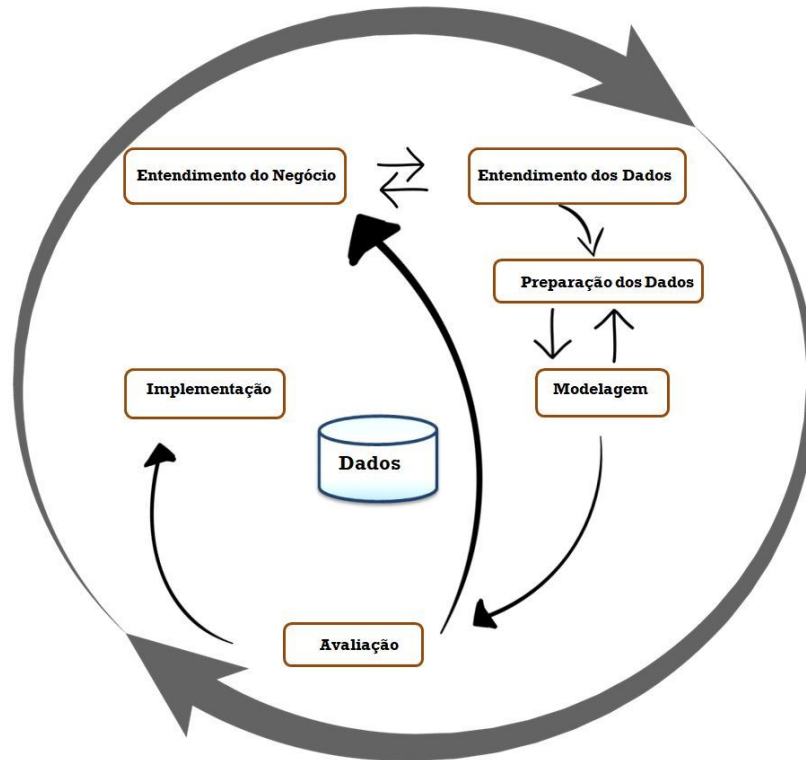


Figura 3.2: O ciclo de vida da mineração de dados (Baseado em: [20]).

**Entendimento do Negócio:** A primeira etapa é o entendimento do negócio, que consiste no aprofundamento acerca do tema que será tratado e na definição dos objetivos a serem atingidos com o processo de mineração de dados. Nessa etapa são reunidas informações sobre o contexto do projeto e então são definidas metas a serem alcançadas com a atividade de análise.

**Entendimento dos Dados:** A segunda etapa compreende o entendimento dos dados, nela os dados disponíveis para utilização no processo são explorados, elencados e organizados, e também, avalia-se a necessidade de obtenção de dados adicionais. Nesse momento, são definidas as colunas que podem ser utilizadas ou descartadas. Essa etapa é muito importante para mitigar a ocorrência de problemas na etapa seguinte, que é a preparação dos dados.

**Preparação dos Dados:** A terceira etapa, preparação dos dados, é a etapa que geralmente consome maior tempo no processo de mineração de dados. Nela são realizadas diversas atividades, como a junção de diferentes conjuntos de dados, a seleção de subconjuntos para utilização no processo, a criação de novas colunas utilizando as existentes, a classificação dos dados que mais se adequam à modelagem e a remoção de valores nulos ou vazios. Essa etapa é essencial para que bons resultados possam ser alcançados, pois se o conjunto de dados de entrada não tiver uma boa qualidade, os dados de saída não são confiáveis.

**Modelagem:** A quarta etapa do processo é a modelagem, é nessa etapa em que os primeiros resultados do projeto começam a surgir. Nesse estágio são selecionadas as ferramentas e técnicas adequadas para alcançar as metas que foram definidas na primeira fase. Normalmente, são testados diversos modelos com parâmetros variados, selecionando-se os que tiverem melhor performance. Com os modelos construídos e executados, têm-se os resultados disponíveis para avaliação do atingimento dos objetivos propostos.

**Avaliação:** A quinta etapa é a avaliação, ponto em que a maior parte do processo já estará realizado. Nesse passo, os resultados são avaliados com base nos critérios de aceitação propostos para o processo de mineração de dados. Além disso, é feito o levantamento sobre as informações que foram adquiridas e as descobertas efetuadas. Caso entenda-se que os resultados não foram satisfatórios, as etapas de preparação dos dados ou modelagem devem ser refeitas para adequação ao proposto inicialmente.

**Implementação:** Por último, a sexta etapa é a implementação. É nela em que os resultados obtidos no processo de mineração de dados são utilizados, seja para melhoria de rotinas com base nas percepções obtidas ou na criação de novas rotinas com base nos resultados da análise. Nessa fase também são realizadas eventuais revisões do projeto e produzidos os relatórios necessários para documentação do processo de mineração de dados.

### 3.2.2 Técnicas para Mineração de Dados

Como descrito na seção anterior, na etapa de modelagem do processo de mineração de dados é necessário que se faça a escolha das ferramentas e técnicas que serão utilizadas para obtenção dos resultados. Essas técnicas são escolhidas através do conhecimento dos objetivos a serem atingidos e do entendimento dos dados que serão utilizados. Além disso, também podem ser feitos alguns testes para avaliar a técnica mais adequada. No presente



trabalho, foram utilizadas duas técnicas de mineração de dados: a regressão linear e a clusterização.

### 3.2.2.1 Regressão Linear

A regressão linear teve o papel de executar a previsão dos dados de inadimplência total. Essa técnica foi escolhida após a realização de testes com os dados do Banco do Brasil e do Bradesco até o trimestre de Junho de 2019. Além da regressão linear, foram testados os algoritmos *Gaussian Process* e *Multilayer Perceptron*. Como as previsões obtidas com a regressão linear apresentaram maior proximidade com os dados reais, ela foi escolhida para a realização das previsões dos demais trimestres e instituições financeiras.

Para Goldschmidt e Passos [17], a regressão "compreende, fundamentalmente, a busca por funções, lineares ou não, que mapeiem os registros de um banco de dados em valores reais", sendo que a regressão linear é uma das formas de regressão no qual se obtém uma função linear.

De forma prática, se trata da obtenção do valor de uma variável através do uso dos valores de outras variáveis, onde a variável a ser obtida é chamada de alvo e as demais variáveis são chamadas de independentes.

Segundo Hoffmann [21] existem duas formas da regressão linear: a simples e a múltipla. Na regressão linear simples o valor da variável alvo é obtido a partir da única variável independente existente, como pode ser observado na Equação 3.1, onde  $Y$  é a variável alvo,  $\alpha$  e  $\beta$  são os coeficientes de regressão linear,  $X$  é a variável independente e o  $u$  representa possíveis erros de mensuração e efeitos de outras variáveis não previstas. O índice  $i$  representa a numeração do par: variável alvo mais variável independente, considerando a existência de  $n$  pares.

$$Y_i = \alpha + \beta X_i + u_i \quad (3.1)$$

Já a regressão linear múltipla considera  $k$  variáveis como independentes, ou seja, o valor da variável alvo é previsto com base nos valores de diversas outras variáveis, conforme visualizado na Equação 3.2. De mesmo modo,  $Y$  é a variável alvo,  $\alpha$  e  $\beta$  são coeficientes de regressão linear e  $u$  representa possíveis erros de mensuração e efeitos de outras variáveis não previstas. Entretanto, podem existir  $k$  variáveis independentes, representadas por  $X$  e há um coeficiente  $\beta$  para cada uma delas.

Desta forma, são somados os produtos de cada coeficiente  $\beta$  pela variável independente correspondente, onde a numeração de cada um dos pares é representada pelo índice  $i$ . O índice  $j$  é utilizado para especificar a numeração do conjunto de variável alvo com as respectivas variáveis independentes e o erro existente, assumindo a existência de  $n$  conjuntos

compostos pela variável alvo mais as respectivas variáveis independentes correspondentes. Neste trabalho, essa situação é exemplificada pelo trimestre referência, pois existe um conjunto de valores de inadimplência total e outras variáveis das instituições financeiras para cada trimestre.

$$Y_j = \alpha + \sum_{i=1}^k \beta_i X_{ij} + u_j \quad (3.2)$$

Considerando que para a realização da presente análise havia a disponibilidade de diversas variáveis relacionadas à variável alvo, optou-se pela utilização da regressão linear múltipla, onde a inadimplência total foi a variável alvo e os demais atributos as variáveis independentes.

### 3.2.2.2 Clusterização

A clusterização foi responsável pelo agrupamento das instituições financeiras considerando suas características em comum. A técnica de clusterização, também conhecida como agrupamento, consiste na identificação de registros similares e junção desses registros em vários grupos ou *clusters* [17]. Esse processo é realizado para agrupar os dados com características semelhantes e identificar relações entre eles.

O agrupamento pode ser realizado de diversas maneiras. No caso do presente trabalho foi empregada a clusterização hierárquica utilizando distância euclidiana. A clusterização hierárquica é efetuada com a criação de um *cluster* inicial para cada registro de dados, em seguida, os *clusters* vão se juntando ao *cluster* mais próximo, até formar um único *cluster* no final, que é o agrupamento de todos os dados. Esse agrupamento pode ser visualizado através de um dendograma, que nada mais é do que uma árvore em que a raiz é a união de todos os dados e as folhas são os *clusters* iniciais, de acordo com o visualizado na Figura 3.3.

Nessa análise, a clusterização foi realizada com base nas instituições financeiras, adotando todas as outras variáveis para avaliar a distância entre cada instituição. Os resultados foram representados em dendogramas contendo a similaridade entre as instituições financeiras avaliadas.

## 3.3 Python

*Python* é uma linguagem de programação interpretada e orientada à objetos, que foi criada em 1991 por Guido van Rossum [4]. Ela é amplamente utilizada por ser uma ferramenta de código aberto, ter alta portabilidade, atender à diversos propósitos e possuir grande comunidade ativa, além de diversas bibliotecas que incluem recursos adicionais.

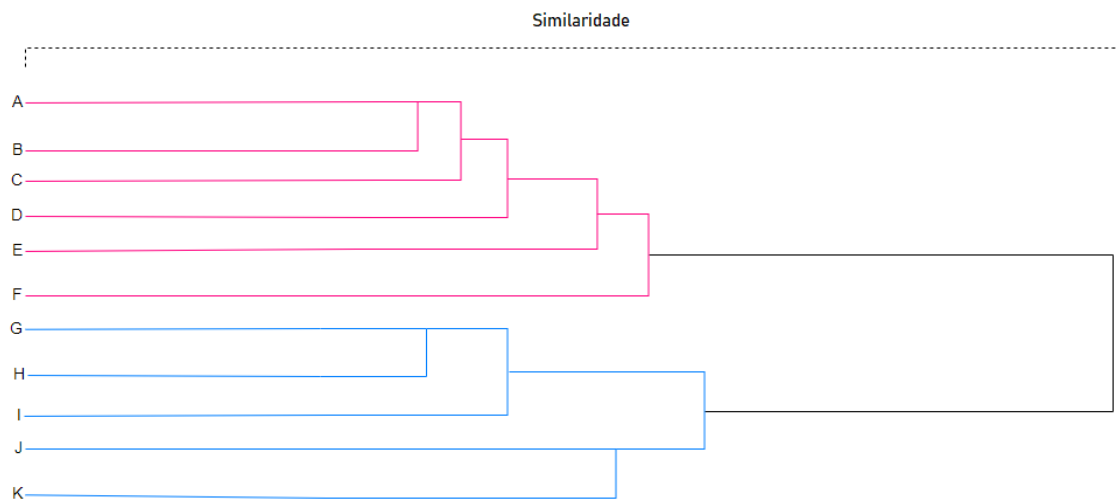


Figura 3.3: Exemplo de Dendograma (Baseado em: [22]).

Neste trabalho, optou-se pela utilização da linguagem *Python*, principalmente pela possibilidade de adição da biblioteca *Pandas* [23], que é uma biblioteca de código aberto utilizada para análise de dados. A biblioteca *Pandas* foi publicada em 2012 e através dela pode-se importar, tratar e exportar dados, com o auxílio de tabelas em formato *Data frame*, que possuem fácil visualização e manipulação.

### 3.4 PostgreSQL

*PostgreSQL* [3] é um Sistema Gerenciador de Bancos de Dados (SGBD) de código aberto, que utiliza a *Structured Query Language* (SQL) acrescida de alguns recursos adicionais. Sua origem é resultado de um projeto iniciado em 1986 na *University of California at Berkeley* e atualmente é mantido pelo *The PostgreSQL Global Development Group*.

A sua escolha para utilização no presente trabalho se deu pela facilidade de criação e manipulação dos bancos de dados na ferramenta, além da disponibilidade de extensa documentação e suporte da comunidade.

### 3.5 Weka

O *Weka* [5] é um *software* de código aberto, desenvolvido em linguagem *Java* pela universidade de Waikato na Nova Zelândia, que agrupa uma série de algoritmos de mineração de dados e aprendizado de máquina. Ele possui 5 diferentes implementações de interface: *Explorer*, *Experimenter*, *KnowledgeFlow*, *Workbench* e *SimpleCLI*. No presente trabalho

utilizou-se a interface *Explorer* que permite a importação, pré-processamento e aplicação de algoritmos aos dados carregados com o uso de uma interface amigável [17].

A ferramenta possui um tipo de arquivo específico para utilização, que é o formato *.arff*, porém, também é possível a importação de dados através de arquivos no formato *.csv* e por conexão com bancos de dados. Nessa análise, os dados foram importados para o *Weka* a partir da conexão com o *PostgreSQL* e a execução de consultas SQL.

Uma grande vantagem do *Weka* é a possibilidade de customização e expansão através da instalação de novos algoritmos e pacotes. Para a realização da análise em pauta, foi incluído no *Weka* o pacote *timeseriesForecast* [24], que possui algoritmos especializados na análise de séries temporais, o que facilitou a execução de algumas etapas essenciais do processo.

Deste modo, a ferramenta *Weka* foi escolhida para a realização do trabalho tendo em vista a usabilidade da interface gráfica da ferramenta, a facilidade de conexão com o banco de dados junto ao *PostgreSQL* e a possibilidade de customização através da instalação do pacote *timeseriesForecast*.

## 3.6 Orange

O *Orange* [6] é uma ferramenta de código aberto, que inclui diversos módulos de mineração e visualização de dados. Ela foi desenvolvida pela Universidade de Liubliana na Eslovênia. A ferramenta apresenta uma interface gráfica de fácil utilização, onde as tarefas são executadas a partir da inclusão de módulos em um fluxo de processamento.

Optou-se por seu uso no presente trabalho tendo em vista os recursos disponíveis para geração de dendogramas, a facilidade de utilização de sua interface e a possibilidade de conexão com banco de dados através do *PostgreSQL*.

# Capítulo 4

## Estudo de Caso: Dados de Instituições Financeiras que operam no Brasil

No presente trabalho foi feito um estudo de caso das instituições financeiras que operam no Brasil com o objetivo de realizar a previsão da inadimplência total de cada instituição financeira para os trimestres de Junho de 2020 e Setembro de 2020 e também de buscar relações entre as instituições através de um algoritmo de clusterização. De forma a orientar a condução das tarefas, as etapas foram divididas de acordo com o processo CRISP-DM.

### 4.1 Entendimento do Negócio

A primeira etapa consistiu no entendimento do negócio e na definição dos objetivos a serem alcançados com o processo de mineração de dados. Para iniciar essa etapa foi feita a avaliação do portal IF.data buscando a identificação dos dados disponíveis para obtenção e utilização no trabalho.

Na página inicial do portal IF.data [14] é possível selecionar a data-base do relatório desejado, o tipo de instituição e o tipo do relatório, conforme observado na Figura 4.1. Os relatórios são divulgados trimestralmente, 3 meses após o fechamento do trimestre referência, ou seja, o relatório referente ao trimestre encerrado em Março de 2020 é divulgado no último dia de Junho de 2020, por exemplo. Ao longo da realização desse trabalho, estiveram disponíveis os dados do período de Março de 2000 até Junho de 2020.

Quanto ao tipo de instituição é possível selecionar entre: Conglomerados Financeiros e Instituições Independentes, Conglomerados Prudenciais e Instituições Independentes, Instituições Individuais ou Instituições com Operações de Câmbio. A tela inicial do portal apresenta um breve resumo sobre cada um dos tipos de instituição disponíveis para seleção.

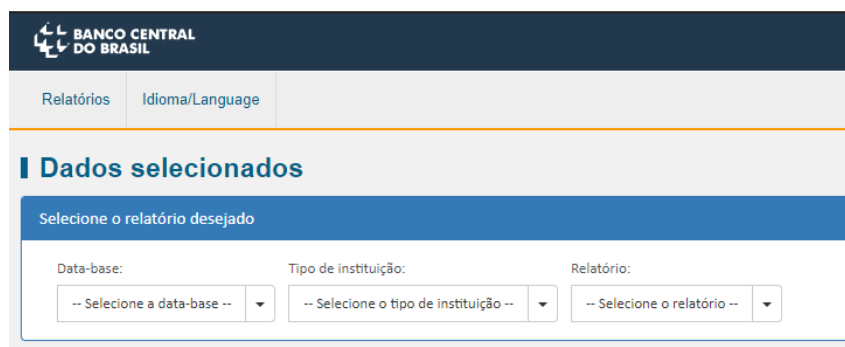


Figura 4.1: Tela Inicial do Portal IF.data.

Já na aba de seleção do tipo de relatório se pode optar pelo relatório a ser obtido. As opções de relatórios alternam de acordo com o tipo de instituição e a data-base escolhida. Estão disponíveis 23 relatórios diferentes a depender da combinação de filtros aplicada, os relatórios são gerados a partir dos dados enviados pelas instituições financeiras ao Bacen nos arquivos Plano Contábil de Instituições Financeiras, Demonstrativo de Limites Operacionais, Sistema de Informações de Crédito do Banco Central do Brasil e Demonstrativo do Sistema de Câmbio.

Após a identificação dos dados que poderiam ser obtidos e de quais informações teriam potencial de serem extraídas, o escopo do trabalho ficou restringido à realização de duas principais tarefas: a previsão da inadimplência total das instituições financeiras e a classificação das instituições financeiras através de um algoritmo de clusterização.

Com essas tarefas definidas, foi possível delimitar os dados que seriam coletados no portal. Dessa forma, foram selecionados todos os relatórios dos Conglomerados Financeiros e Instituições Independentes do período de Junho de 2014 até Junho de 2020. A escolha desse período se deu em função de os dados da carteira de crédito das instituições financeiras serem fundamentais para a realização da previsão da inadimplência total, e estes estavam disponíveis apenas a partir do trimestre Junho de 2014 com o início da publicação dos dados do SCR.

## 4.2 Coleta e Entendimento dos Dados

Com a definição dos objetivos e a delimitação escopo já realizada, começa a etapa de coleta e entendimento dos dados, onde os dados que serão utilizados são obtidos e analisados de forma exploratória para sua melhor compreensão.

Os dados do portal IF.data podem ser obtidos no formato *.csv*, bastando selecionar o relatório desejado e clicar no *link* CSV, conforme demonstrado na Figura 4.2. Previamente ao *download* do arquivo, é apresentada uma visualização na página com o conteúdo do

**Dados selecionados**

Selecione o relatório desejado

Data-base: 03/2020 Tipo de instituição: Conglomerados Financeiros e Instituições Independentes Relatório: Resumo

Resumo  
Valores monetários em R\$ mil  
Informações com base nos documentos entregues até: 23/06/2020  
CSV Composição de Colunas Composição de Colunas em PDF

Instituição financeira	TCB	SR	TD	TC	Cidade	UF	Data	Ativo Total
ITAU	b1	S1	C	2	SAO PAULO	SP	03/2020	1.861.576.247
BB	b1	S1	C	1	BRASILIA	DF	03/2020	1.581.652.548
CAIXA ECONOMICA FEDERAL	b1	S1	I	1	BRASILIA	DF	03/2020	1.314.428.681

Figura 4.2: Download do arquivo em formato .csv no Portal IF.data.

relatório que foi selecionado, de forma que já se pode ter uma prévia do arquivo que será baixado.

Para o tipo de instituição que foi escolhido (Conglomerados Financeiros e Instituições Independentes), estão disponíveis 12 relatórios diferentes: Resumo, Ativo, Passivo, Demonstração de Resultado, e as visões da Carteira de Crédito: Pessoa Física por modalidade e prazo, Pessoa Jurídica por modalidade e prazo, Pessoa Jurídica por atividade econômica (CNAE), Pessoa Jurídica por porte do tomador, quantidade de clientes e operações, Carteira de crédito por nível de risco da operação, Carteira de crédito por indexador e Carteira de crédito por região geográfica. Cada relatório gera um arquivo .csv para cada trimestre referência, no trabalho, foram utilizados 25 trimestres, o que totalizou 300 arquivos .csv.

Dado o elevado número de arquivos, a etapa de coleta dos dados foi uma das que mais demandou tempo para a realização do trabalho, pois os arquivos precisaram ser baixados um a um de forma manual no portal If.data, selecionando os filtros correspondentes para então fazer o *download* do arquivo .csv.

```

1 |Instituição financeira;Código;TCB;SR;TD;TC;Cidade;UF;Data;A
2 |BB;49906;b1;;C;1;BRASILIA;DF;06/2014;1.267.430.066;628.565.
3 |ITAU;10069;b1;;C;2;SAO PAULO;SP;06/2014;1.029.025.053;414.8
4 |CAIXA ECONOMICA FEDERAL;360305;b1;;I;1;BRASILIA;DF;06/2014;

```

Figura 4.3: Exemplo de arquivo em formato .csv baixado do Portal IF.data.

Após a realização do *download* de todos os arquivos, eles foram agrupados em pastas

organizadas por trimestre referência de modo a facilitar sua exploração e posteriormente sua importação. Com isso, foi feita uma checagem nos arquivos de modo a identificar eventuais problemas no momento da coleta e levantar os tratamentos que seriam necessários na próxima etapa, de preparação dos dados. Os arquivos obtidos apresentam uma estrutura de texto separada por ponto e vírgula, onde a primeira linha contém o nome de cada coluna, conforme observado na Figura 4.3.

## 4.3 Preparação dos Dados

### 4.3.1 Tratamentos realizados

Com a coleta dos dados realizada, foram obtidos todos os arquivos *.csv* necessários para a realização da análise, porém, antes de iniciar a análise propriamente dita foi necessária a realização de alguns processos de maneira a deixar os dados no formato adequado para serem utilizados nas ferramentas *Weka* e *Orange*.

O primeiro passo para a preparação dos dados foi realizado utilizando o editor de textos *Notepad++*. Inicialmente foi feito um pré-processamento em cada arquivo *.csv* para remoção das últimas linhas dos arquivos, que representavam algumas sumarizações dos relatórios e não tinham o leiaute correspondente às demais linhas do arquivo. Além disso, os valores NI (não informado) foram substituídos por espaços, de forma a serem tratados como valores vazios quando os arquivos fossem importados.

Com o pré-processamento realizado, para a etapa seguinte foi utilizada a linguagem *Python*, com o auxílio das bibliotecas *Pandas* e *Numpy*. Para facilitar a visualização da execução de cada processo, os *scripts* foram construídos na ferramenta *Jupyter Lab*. Para importação e tratamento dos arquivos foi utilizado o *script* do Anexo I. Primeiro foram construídos 12 laços de repetição, de modo que em cada laço foi feita a importação dos arquivos de todos os trimestres do mesmo relatório em um *dataframe*, para isso foi utilizado o método *read\_csv* da biblioteca *Pandas*.

Em cada um desses laços de repetição foram definidos nomes abreviados para as colunas nos parâmetros do método de importação, de maneira a facilitar a manipulação da tabela. A relação das abreviaturas utilizadas está listada no Anexo II. Com isso, os laços foram executados e geraram 12 tabelas diferentes, uma para cada relatório contendo todos os trimestres disponíveis, como relatado na seção 4.2.

Depois da formação das 12 tabelas, foram removidas todas as linhas que possuíam apenas valores nulos, pois essas linhas não poderiam ser utilizadas na análise já que não continham nenhum valor. Feito isso, foi definida uma chave para identificação única de cada registro em cada tabela e através da utilização dessa chave foi feito um relaciona-



mento, do tipo *left join*, unindo as tabelas uma a uma para formar uma única tabela contendo os dados de todos os relatórios.

Por fim, foi adicionada uma nova coluna à tabela com a soma dos valores de inadimplência de cada instituição, para cada trimestre. Para isso, foram utilizadas as visões por modalidade e prazo de vencimento das carteiras de crédito de pessoa física e pessoa jurídica. As demais visões não foram utilizadas para não haver duplicidade de valores, pois tratam-se apenas dos mesmos valores da carteira de crédito segmentados de uma maneira diferente. Essa nova coluna gerada será a nossa variável alvo no momento de realizar a previsão da inadimplência total com base nos dados do histórico.

Praticadas todas essas ações para tratamento dos dados, o produto final foi uma tabela contendo 333 colunas e 34.283 linhas, de acordo com o observado na figura Figura 4.4. Cada coluna tem origem em uma das colunas de um dos relatórios utilizados com base e cada linha representa os dados de um trimestre de uma instituição financeira.

	IF	COD	TCB	SR	TD	TC	CID	UF	DATA	ATIV_TOT	OUT_IND_PREC	OUT_INDEX	TOT_N_INDIV	SUDESTE	CENTRO_OESTE	NORDESTE	NORTE	SUL	REG
0	BB	49906	b1	NaN	C	1	BRASILIA	DF	2014-09-01	1.291560e+09	...	2564633.0	2117776.0	52023054.0	295074988.0	70541132.0	67245462.0	26330102.0	113124671.0
1	ITAU	10069	b1	NaN	C	2	SAO PAULO	SP	2014-09-01	1.069552e+09	...	14381.0	6995960.0	181239129.0	203958157.0	20907528.0	39776215.0	11150295.0	49562526.0
2	CAIXA ECONOMICA FEDERAL	360305	b1	NaN	I	1	BRASILIA	DF	2014-09-01	1.018865e+09	...	10764510.0	1371617.0	0.0	284074574.0	67085038.0	89239888.0	23685708.0	104725160.0
3	BRADESCO	10045	b1	NaN	C	2	OSASCO	SP	2014-09-01	8.488383e+08	...	0.0	6814753.0	83262720.0	161880080.0	24333486.0	40455172.0	14046860.0	45909582.0
4	BANCO NACIONAL DE DESENVOLVIMENTO ECONOMICO E ...	33657248	b4	NaN	I	1	RIO DE JANEIRO	RJ	2014-09-01	8.235018e+08	...	0.0	0.0	0.0	187841757.0	29133130.0	38830887.0	5280923.0	23863523.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
34281	COOPERATIVA DE ECONOMIA E CREDITO MUTUO DOS TR...	7494300	b3S	S5	I	2	ERECHIM	RS	2020-06-01	NaN	...	0.0	0.0	4.0	13.0	0.0	1.0	0.0	14750.0
34282	COOPERATIVA DE CREDITO RURAL DO AGRESTE CENTRAL...	8202793	b3S	S5	I	2	ARAPIRACA	AL	2020-06-01	NaN	...	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
34283	F.D'GOLD - DISTRIBUIDORA DE TITULOS E VALORES ...	8673569	n2	S4	I	2	SAO PAULO	SP	2020-06-01	NaN	...	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
34284	MUTUAL APETRIM CREDITO IMOBILIARIO S/A	18941229	n1	S4	I	2	BELO HORIZONTE	MG	2020-06-01	NaN	...	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
34285	COOPERATIVA DE CREDITO - SICOOB CREDICOONAI	53923116	b3S	S4	I	2	RIBEIRAO PRETO	SP	2020-06-01	NaN	...	0.0	0.0	5.0	365774.0	72.0	326.0	36.0	231.0

34286 rows x 333 columns

Figura 4.4: Dataframe gerado após a importação dos arquivos CSV.

Subsequentemente à obtenção dessa tabela única, foi feito um último tratamento para possibilitar a carga no banco de dados, que foi a conversão do formato do campo 'DATA' para o formato *datetime*, esse tratamento foi realizado com o objetivo de facilitar a identificação da tabela como uma série temporal no momento de sua importação para o aplicativo *Weka*.

### 4.3.2 Carga no Banco de Dados

Optou-se por carregar a tabela resultado dos tratamentos em um banco de dados, para facilitar o acesso pelas ferramentas *Weka* e *Orange*. A realização da análise depende de algumas subtabelas, dessa forma, a execução de consultas do tipo *select* diretamente do banco de dados é uma opção mais prática e segura do que a geração de diversos arquivos *.csv*. Isso ocorre, a medida em que a possibilidade de erros de importação e divergências de dados é reduzida.

Para a realização da carga da tabela no banco de dados, o primeiro passo foi a criação do banco através da ferramenta do SGBD, que no caso do *PostgreSQL* é o *pgAdmin*. Exercida a abertura do *pgAdmin*, selecionamos a opção *Create > Database*, de acordo com a Figura 4.5. Com isso, foi preciso definir apenas o nome do banco de dados a ser criado, já que as demais definições estavam preenchidas por padrão.

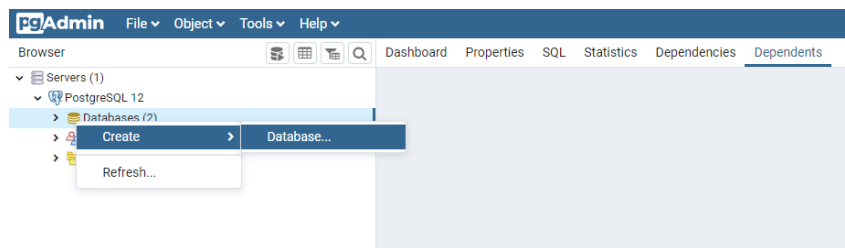


Figura 4.5: Criação do Banco de Dados na ferramenta pgAdmin.

Desempenhada criação do banco de dados, o *script* utilizado para tratamento dos dados foi reaberto e através dele a carga dos dados ocorreu com a utilização do pacote *psycopg2*. Para sua execução, foram definidos como parâmetros o endereço e o nome do banco de dados criado, além do nome do *schema* da tabela a ser carregada.

Uma vez que a carga da tabela foi concluída, apresentou-se necessária a instalação do *driver* do *PostgreSQL* para que o banco de dados pudesse ser acessado pela ferramenta *Weka*. De modo a realizar a instalação, o *driver* mais recente foi obtido no sítio [postgresql.org](http://postgresql.org) resultando no *download* do arquivo *postgresql-42.2.16.jar*. Esse arquivo foi copiado para a pasta onde o *Weka* estava instalado.

Em seguida, foi preciso realizar a edição de dois arquivos de configurações do *Weka*: o arquivo *RunWeka.ini* que está presente na pasta onde o *Weka* foi instalado e o arquivo *DatabaseUtils.props.postgresql* que pode ser localizado no diretório *weka/experiment* do arquivo *weka.jar*. De modo a exercer a edição do arquivo *DatabaseUtils.props.postgresql* é necessário descompactá-lo para a pasta de instalação do *Weka* e remover o sufixo *.postgresql*. A edição realizada no arquivo *RunWeka.ini* pode ser observada na Figura 4.6 e a efetuada no arquivo *DatabaseUtils.props.postgresql* pode ser verificada na Figura 4.7.

```

38 # The classpath placeholder. Add any environment variables or jars to it that
39 # you need for your Weka environment.
40 # Example with an enviroment variable (e.g., THIRD_PARTY_LIBS):
41 # cp=%CLASSPATH%;%THIRD_PARTY_LIBS%
42 # Example with an extra jar (located at D:\libraries\libsvm.jar):
43 # cp=%CLASSPATH%;D:\\\\libraries\\\\libsvm.jar
44 # Or in order to avoid quadrupled backslashes, you can also use slashes "/":
45 # cp=%CLASSPATH%;D:/libraries/libsvm.jar
46 cp=%CLASSPATH%;postgresql-42.2.16.jar

```

Figura 4.6: Modificação realizada no arquivo Weka.ini.

```

11 # JDBC driver (comma-separated list)
12 jdbcDriver=org.postgresql.Driver
13
14 # database URL
15 jdbcURL=jdbc:postgresql://localhost:5432/base_ifdata

```

Figura 4.7: Modificação realizada no arquivo DatabaseUtils.props.postgresql.

## 4.4 Previsão da Inadimplência Total

Após finalizada a etapa de preparação dos dados, as etapas seguintes foram a modelagem e a avaliação. De modo a facilitar a explanação do trabalho de análise, essas duas etapas serão abordadas em conjunto no texto.

### 4.4.1 Importação dos dados no Weka

Para a montagem do modelo preditivo, foi utilizada a aplicação *Explorer* do *Weka*. A importação dos dados pode ser feita na aba *Preprocess*, onde seleciona-se a opção *Open DB* para carregar o editor de consultas. Com o editor de consultas aberto deve-se preencher as credenciais para conexão com o banco de dados: a *url* do banco, o usuário e a senha. Concretizada a conexão, pode-se utilizar o espaço em branco para a digitação da consulta a ser efetuada, de acordo com a Figura 4.8.

Realizada a execução da consulta, a tela apresenta uma prévia do resultado com 100 linhas e na caixa *info* que fica na parte inferior da janela, é mostrado o número total de linhas selecionadas na consulta que foi executada. Ao confirmar o resultado com o botão *Ok* os dados são internalizados para o *Weka* e as demais abas ficam disponíveis, conforme Figura 4.9.

### 4.4.2 Geração do Modelo Preditivo

Visto que os dados que serão utilizados na análise estão organizados em uma série temporal, optou-se pela utilização do pacote *timeseries Forecast* [24] disponível para instalação no gerenciador de pacotes do *Weka*. O pacote é dedicado especificamente à análise de

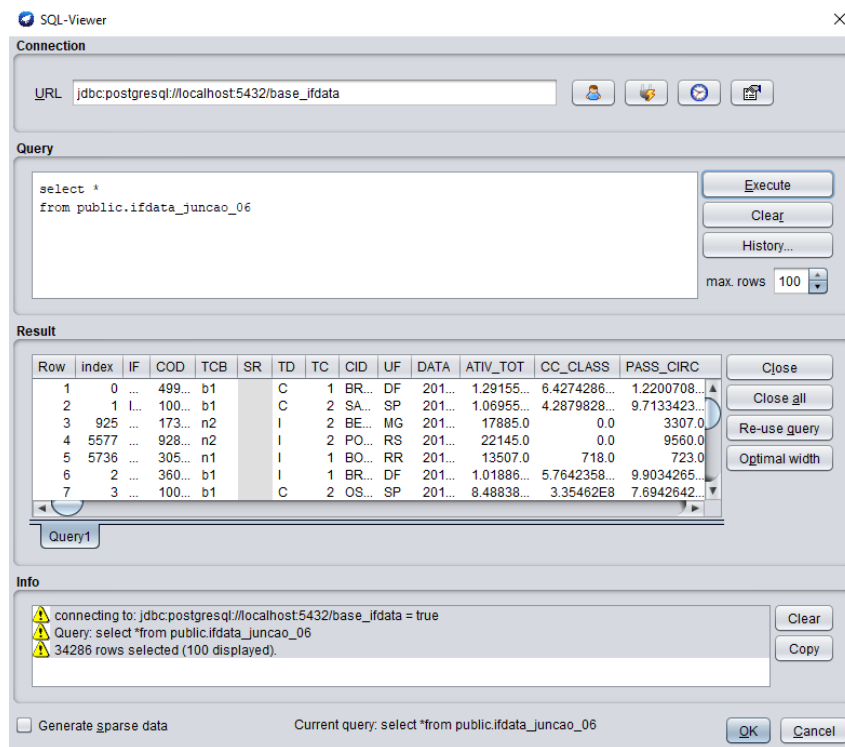


Figura 4.8: Opção Open DB do Weka.

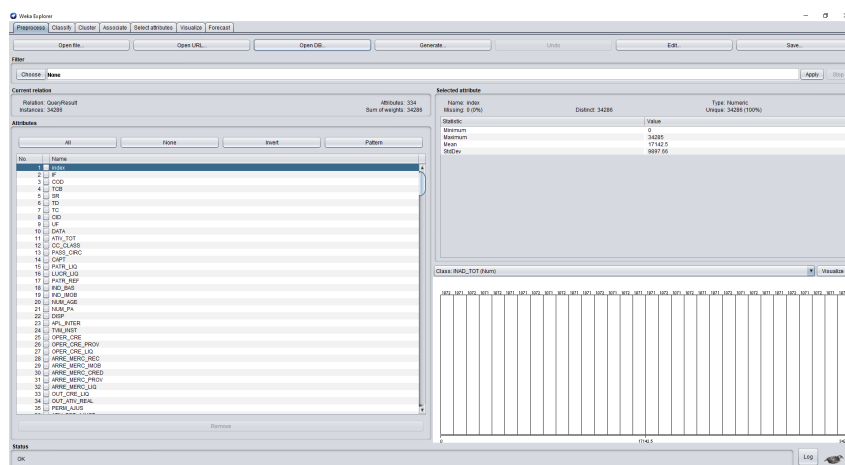


Figura 4.9: Janela do Weka após internalização dos dados.

séries temporais e tem mecanismos para identificação das séries temporais e criação automatizada de variáveis com *lag*.

A criação das variáveis com *lag* é feita com o deslocamento da variável que indica a passagem de tempo, causando um atraso nos dados e permitindo a comparação dos dados atuais com os dados mais antigos. No caso da tabela do presente trabalho, a variável que indica a passagem do tempo é a coluna 'DATA', onde o trimestre referência dos dados é

apontado.

A quantidade de unidades utilizada para aplicar o *lag* pode ser definida manualmente ou automaticamente no pacote *timeseries Forecast*. Para construção dos modelos de predição da inadimplência, utilizou-se a definição automática que aplicou 4 unidades de tempo de *lag*, ou seja, um trimestre referência é comparado ao mesmo trimestre de 1 ano atrás.

Com os dados já internalizados no *Weka* foi iniciada a geração dos modelos preditivos para a previsão da variável *INAD\_TOT*, que foi definida como variável alvo. Na primeira tentativa foram utilizadas todas as colunas disponíveis na tabela como variáveis independentes com o objetivo de construir um modelo que levasse em conta os dados de todas as colunas e de todas as instituições. Todavia, observou-se que o tempo de processamento fora muito grande, superando 12 horas.

De maneira a otimizar a construção dos modelos, definiu-se pela priorização das instituições financeiras com maiores valores de inadimplência e também pela limitação das variáveis a serem utilizadas para construção do modelo com base em sua preditibilidade.

Com isso, o escopo do trabalho ficou restrito a 10 instituições financeiras, onde a criação do modelo e a análise para cada uma delas foi feita separadamente. As instituições escolhidas foram: Banco do Brasil, Banco Nacional de Desenvolvimento Econômico e Social, Bradesco, Caixa Econômica Federal, Banco CSF, Itaú, Nubank, Banco Pan, Santander e Banco Votorantim. Os critérios para escolha foram: maior valor apresentado na coluna referente à inadimplência total no trimestre Março de 2020 e relevância da instituição no SFN em termos de ativos e carteira de crédito.

Além disso, também foram definidas 23 colunas como variáveis independentes. As colunas foram selecionados gerando alguns modelos de regressão no *Weka* com todas as variáveis disponíveis. Com os modelos gerados, observou-se quais as variáveis que possuíam maior coeficiente dentro da equação de regressão que foi obtida, escolhendo as variáveis associadas a coeficientes mais elevados.

Desta forma, a lista de colunas escolhidas foi: Ativo Total, Carteira de Crédito Classificada, Captações, Patrimônio Líquido, Provisão sobre Operações de Crédito, Passivo Total, Despesas de Pessoal, Despesas Administrativas, Despesas Tributárias, Total da Carteira Pessoa Física, Total da Carteira Pessoa Jurídica, Total Geral, Risco AA, Risco A, Risco B, Risco C, Risco D, Risco E, Risco F, Risco G, Risco H e Total Exterior.

Cumprida a definição das instituições a serem analisadas e as variáveis utilizadas como independentes, o processo de geração dos modelos foi realizado para cada uma das instituições, restringindo-se inicialmente aos dados disponíveis até Março de 2020.

#### 4.4.2.1 Banco do Brasil

Para a geração do modelo preditivo do BB, o primeiro passo foi a importação dos dados dessa instituição para o *Weka*, já com os filtros das colunas e dos trimestres a serem considerados. Para isso, foi utilizada a consulta SQL abaixo, onde os filtros aplicados no argumento *where* retiram os dados de Junho de 2020, que é o trimestre a ser previsto e selecionam apenas os dados da instituição financeira que será analisada, através de seu código identificador.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",  
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",  
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",  
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",  
"TOT_EXT", "INAD_TOT"  
from public.ifdata_juncao_06  
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 49906
```

Com a consulta executada, seleciona-se a aba *Forecast* do *Weka* para definição dos parâmetros a serem utilizados para geração do modelo, conforme observado na Figura 4.10. Na aba *Basic Configuration* é possível selecionar a variável ou variáveis que serão alvo do modelo, o número de unidades de tempo para realização da previsão, a variável utilizada com referência temporal e a periodicidade da série temporal utilizada. Além disso, caso a caixa *Perform evaluation* seja marcada, o modelo é gerado em modo de teste, retornando as métricas de erro.

Neste caso, foram definidas como variáveis alvo todas as colunas selecionadas, mesmo que o objetivo em pauta seja a previsão apenas da variável *INAD\_TOT*, pois dessa forma é gerado um modelo para cada variável considerando-as como dependentes. Utilizou-se a coluna '*DATA*' como referência temporal e 2 trimestres como número de períodos para previsão. Na periodicidade foi mantida a opção *Detect automatically*, com isso o *Weka* detectou automaticamente que a periodicidade é trimestral.

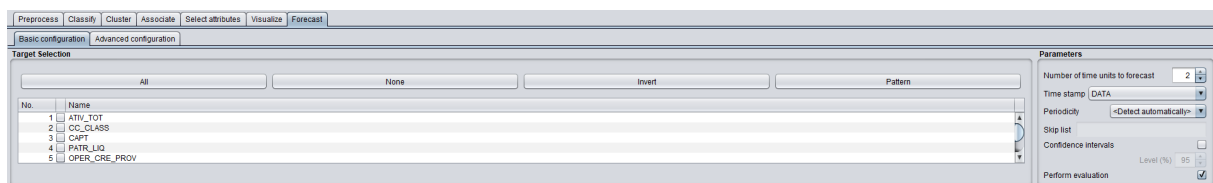


Figura 4.10: Configurações básicas - Weka Forecast.

Já na aba *Advanced configuration* existe a possibilidade de alterar as configurações seguintes: *Base learner*, *Lag creation*, *Periodic attributes*, *Overlay data*, *Evaluation* e

*Output*, de acordo com o mostrado na Figura 4.11. Neste contexto, limitou-se à utilização das configurações de *Base learner*, *Lag creation* e *Evaluation*.

Na aba *Base learner* é possível selecionar o algoritmo de predição que será utilizado, optou-se pela utilização do algoritmo *LinearRegression*.

As definições das configurações de *lag* podem ser realizadas na aba *Lag creation*, manteve-se as configurações originais de forma que o *Weka* defina automaticamente a quantidade de unidades de tempo a ser usada, que nesse caso foi de 4 trimestres.

A seleção das métricas de erro que serão exibidas quando feita uma avaliação do modelo é efetuada na aba *Evaluation*. Foram selecionadas como métricas a serem exibidas: *Mean absolute error (MAE)* que considera a média do erro absoluto na unidade numérica da amostra, que no caso dos dados do trabalho é em R\$, *Root mean squared error (RMSE)* que é o erro quadrático médio, *Mean absolute percentage error (MAPE)* que é o erro absoluto médio percentual e *Relative absolute error (RAE)* que é o erro absoluto em relação à média do atributo.

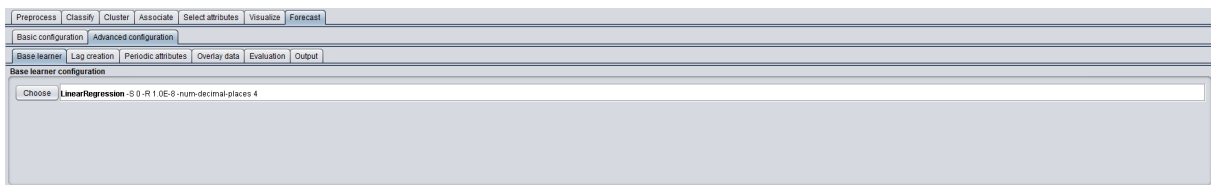


Figura 4.11: Configurações avançadas - Weka Forecast.

Com os parâmetros definidos, executou-se a geração do modelo com a opção *Perform evaluation* marcada, de forma a realizar um teste do modelo usando parte dos dados como treinamento e gerar as métricas de erro. O resultado pode ser observado na Figura 4.12, que mostra que as métricas de erro com dados de treino são bem próximas de 0 quando se tenta prever um trimestre à frente. Com dois trimestres à frente o erro médio absoluto e o erro médio quadrático são um pouco maiores, entretanto, não representam diferenças significativas em termos percentuais.

Após executada avaliação, a marcação da opção *Perform evaluation* foi desfeita e nova execução para geração do modelo com todos os dados disponíveis foi performada. Como resultado, foram gerados os valores previstos para os trimestres de Junho de 2020 e Setembro de 2020, conforme destacado na Figura 4.13, onde podem ser visualizados todos os valores da inadimplência total do trimestre Setembro de 2014 até o trimestre Setembro de 2020, sendo que os dois últimos trimestres, destacados com '\*', representam as previsões realizadas para Junho e Setembro de 2020. Além disso, essa execução também gerou a equação da regressão linear, que pode ser verificada no Anexo III.

```

=== Evaluation on training data ===
Target                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                    19            18
Mean absolute error   0.0008        13.3247
Relative absolute error 0            0.0016
Mean absolute percentage error 0            0.0001
Root mean squared error 0.001         13.968

```

Figura 4.12: Métricas de erro do modelo preditivo para o BB.

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00 6573393
2014-12-01T00:00:00 6666321
2015-03-01T00:00:00 7335013
2015-06-01T00:00:00 6985068
2015-09-01T00:00:00 8267989
2015-12-01T00:00:00 8257849
2016-03-01T00:00:00 9767507
2016-06-01T00:00:00 9030321
2016-09-01T00:00:00 9790665
2016-12-01T00:00:00 10758039
2017-03-01T00:00:00 13004504
2017-06-01T00:00:00 12574725
2017-09-01T00:00:00 13258975
2017-12-01T00:00:00 13496944
2018-03-01T00:00:00 12336617
2018-06-01T00:00:00 12348754
2018-09-01T00:00:00 9507843
2018-12-01T00:00:00 8600510
2019-03-01T00:00:00 9216026
2019-06-01T00:00:00 10000188
2019-09-01T00:00:00 9756100
2019-12-01T00:00:00 9361397
2020-03-01T00:00:00 9749410
2020-06-01T00:00:00* 11629306.1561
2020-09-01T00:00:00* 10085059.9117

```

Figura 4.13: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - BB.

Deste modo, observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 11,63 milhões e 10,09 milhões, respectivamente.

Em seguida, os passos descritos anteriormente foram replicados para geração dos modelos considerando os trimestres encerrados em Junho de 2017, Junho de 2018 e Junho de 2019, alterando-se apenas o filtro aplicado no parâmetro *where* da consulta SQL para restringir os dados até o trimestre imediatamente anterior ao que seria previsto, ou seja, Março de 2017, Março de 2018 e Março de 2019.



A realização desse processo ocorreu para que houvesse mais uma avaliação do uso da regressão linear para se obter as previsões, visto que é possível comparar os dados previstos com os dados reais dos anos anteriores e nesses trimestres não havia efeitos ligados à pandemia da COVID-19. Isso possibilita com que seja verificado o erro do processo de regressão, de forma que seja comparado ao erro obtido para a previsão de Junho de 2020.

A Figura 4.14 apresenta os resultados obtidos no *Weka*. Análoga a previsão de Junho de 2020, estão expressos todos os valores a partir do trimestre Setembro de 2014. O último valor da lista, destacado com '\*', representa a previsão que foi realizada. Observa-se na figura que os valores obtidos para inadimplência total dos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram de 13,68 milhões, 12,9 milhões e 8,84 milhões, respectivamente.

### Banco do Brasil

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	6573393	2014-09-01T00:00:00	6573393	2014-09-01T00:00:00	6573393
2014-12-01T00:00:00	6666321	2014-12-01T00:00:00	6666321	2014-12-01T00:00:00	6666321
2015-03-01T00:00:00	7335013	2015-03-01T00:00:00	7335013	2015-03-01T00:00:00	7335013
2015-06-01T00:00:00	6985068	2015-06-01T00:00:00	6985068	2015-06-01T00:00:00	6985068
2015-09-01T00:00:00	8267989	2015-09-01T00:00:00	8267989	2015-09-01T00:00:00	8267989
2015-12-01T00:00:00	8257849	2015-12-01T00:00:00	8257849	2015-12-01T00:00:00	8257849
2016-03-01T00:00:00	9767507	2016-03-01T00:00:00	9767507	2016-03-01T00:00:00	9767507
2016-06-01T00:00:00	9030321	2016-06-01T00:00:00	9030321	2016-06-01T00:00:00	9030321
2016-09-01T00:00:00	9790665	2016-09-01T00:00:00	9790665	2016-09-01T00:00:00	9790665
2016-12-01T00:00:00	10758039	2016-12-01T00:00:00	10758039	2016-12-01T00:00:00	10758039
2017-03-01T00:00:00	13004504	2017-03-01T00:00:00	13004504	2017-03-01T00:00:00	13004504
2017-06-01T00:00:00*	13684867.6634	2017-06-01T00:00:00	12574725	2017-06-01T00:00:00	12574725
		2017-09-01T00:00:00	13258975	2017-09-01T00:00:00	13258975
		2017-12-01T00:00:00	13496944	2017-12-01T00:00:00	13496944
		2018-03-01T00:00:00	12336617	2018-03-01T00:00:00	12336617
		2018-06-01T00:00:00*	12902030.2247	2018-06-01T00:00:00	12348754
				2018-09-01T00:00:00	9507843
				2018-12-01T00:00:00	8600510
				2019-03-01T00:00:00	9216026
				2019-06-01T00:00:00*	8844259.2284

Figura 4.14: Valores previstos para os trimestre de Junho de 2017, Junho de 2018 e Junho de 2019 - BB.

Por fim, de modo a observar os valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 em comparação aos demais valores da série histórica, foi gerado um gráfico com a evolução dos valores, que pode ser observado na Figura 4.15. O *script* utilizado para geração dos gráficos está disponível no Anexo IV.

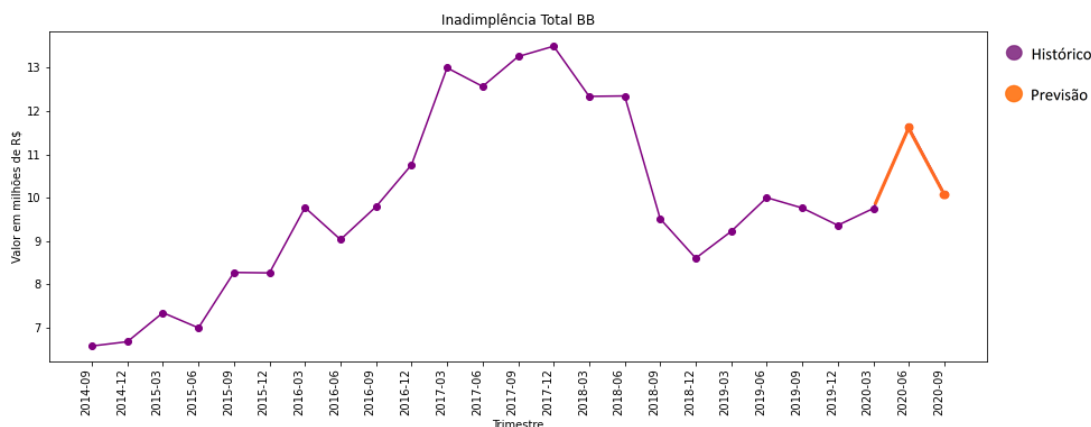


Figura 4.15: BB - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

#### 4.4.2.2 Banco Nacional de Desenvolvimento Econômico e Social

Para a geração dos modelos do BNDES foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o BNDES foi efetuada com a execução da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 33657248
```

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração

```

=== Evaluation on training data ===
Target                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                    19             18
Mean absolute error   0.0003         0.0003
Relative absolute error    0             0
Mean absolute percentage error  0             0
Root mean squared error  0.0003         0.0004

```

Figura 4.16: Métricas de erro do modelo preditivo para o BNDES.

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00    168692
2014-12-01T00:00:00    149426
2015-03-01T00:00:00    1641712
2015-06-01T00:00:00    259661
2015-09-01T00:00:00    142067
2015-12-01T00:00:00    776715
2016-03-01T00:00:00    1734731
2016-06-01T00:00:00    2804220
2016-09-01T00:00:00    937490
2016-12-01T00:00:00    2444602
2017-03-01T00:00:00    2160301
2017-06-01T00:00:00    2277219
2017-09-01T00:00:00    2354246
2017-12-01T00:00:00    527888
2018-03-01T00:00:00    942202
2018-06-01T00:00:00    730499
2018-09-01T00:00:00    1176362
2018-12-01T00:00:00    730726
2019-03-01T00:00:00    605847
2019-06-01T00:00:00    211565
2019-09-01T00:00:00    150816
2019-12-01T00:00:00    606093
2020-03-01T00:00:00    1446834
2020-06-01T00:00:00*   1731195.1218
2020-09-01T00:00:00*   1097768.156

```

Figura 4.17: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - BNDES.

do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.16 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.17 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 1,73 milhão e 1,09 milhão, respectivamente. Já na Figura 4.18 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram

## BNDES

2017		2018		2019	
Time	INAD_TOI	Time	INAD_TOI	Time	INAD_TOI
2014-09-01T00:00:00	168692	2014-09-01T00:00:00	168692	2014-09-01T00:00:00	168692
2014-12-01T00:00:00	149426	2014-12-01T00:00:00	149426	2014-12-01T00:00:00	149426
2015-03-01T00:00:00	1641712	2015-03-01T00:00:00	1641712	2015-03-01T00:00:00	1641712
2015-06-01T00:00:00	259661	2015-06-01T00:00:00	259661	2015-06-01T00:00:00	259661
2015-09-01T00:00:00	142067	2015-09-01T00:00:00	142067	2015-09-01T00:00:00	142067
2015-12-01T00:00:00	776715	2015-12-01T00:00:00	776715	2015-12-01T00:00:00	776715
2016-03-01T00:00:00	1734731	2016-03-01T00:00:00	1734731	2016-03-01T00:00:00	1734731
2016-06-01T00:00:00	2804220	2016-06-01T00:00:00	2804220	2016-06-01T00:00:00	2804220
2016-09-01T00:00:00	937490	2016-09-01T00:00:00	937490	2016-09-01T00:00:00	937490
2016-12-01T00:00:00	2444602	2016-12-01T00:00:00	2444602	2016-12-01T00:00:00	2444602
2017-03-01T00:00:00	2160301	2017-03-01T00:00:00	2160301	2017-03-01T00:00:00	2160301
2017-06-01T00:00:00*	5989235.049	2017-06-01T00:00:00	2277219	2017-06-01T00:00:00	2277219
		2017-09-01T00:00:00	2354246	2017-09-01T00:00:00	2354246
		2017-12-01T00:00:00	527888	2017-12-01T00:00:00	527888
		2018-03-01T00:00:00	942202	2018-03-01T00:00:00	942202
		2018-06-01T00:00:00*	870563.9571	2018-06-01T00:00:00	730499
				2018-09-01T00:00:00	1176362
				2018-12-01T00:00:00	730726
				2019-03-01T00:00:00	605847
				2019-06-01T00:00:00*	492201.9902

Figura 4.18: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - BNDES.

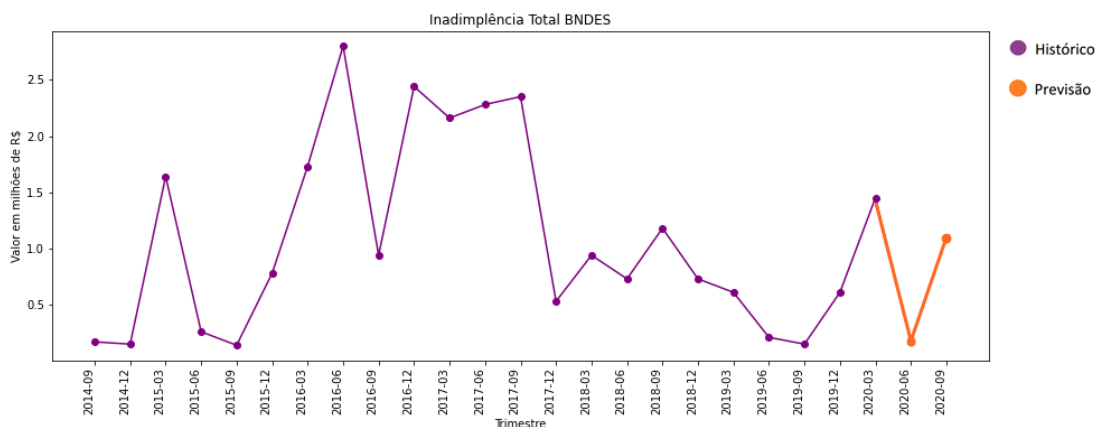


Figura 4.19: BNDES - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

de 5,99 milhões, 0,87 milhão e 0,49 milhão, respectivamente. Por fim, a Figura 4.19 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

### 4.4.2.3 Bradesco

Para a geração dos modelos do Bradesco foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Bradesco foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 10045
```

```
=== Evaluation on training data ===
Target                               1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     19            18
Mean absolute error                   0.0005        0.0007
Relative absolute error                 0             0
Mean absolute percentage error         0             0
Root mean squared error                0.0006        0.0008
```

Figura 4.20: Métricas de erro do modelo preditivo para o Bradesco.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.20 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00 8635124
2014-12-01T00:00:00 8622981
2015-03-01T00:00:00 9385702
2015-06-01T00:00:00 9740514
2015-09-01T00:00:00 10529218
2015-12-01T00:00:00 10732883
2016-03-01T00:00:00 11420797
2016-06-01T00:00:00 11915942
2016-09-01T00:00:00 14955758
2016-12-01T00:00:00 16919839
2017-03-01T00:00:00 15729634
2017-06-01T00:00:00 13322090
2017-09-01T00:00:00 12148023
2017-12-01T00:00:00 11051836
2018-03-01T00:00:00 11520704
2018-06-01T00:00:00 11026519
2018-09-01T00:00:00 10451848
2018-12-01T00:00:00 9819225
2019-03-01T00:00:00 9846433
2019-06-01T00:00:00 10175935
2019-09-01T00:00:00 11545066
2019-12-01T00:00:00 10942537
2020-03-01T00:00:00 12257988
2020-06-01T00:00:00* 11218828.0962
2020-09-01T00:00:00* 11781500.7058

```

Figura 4.21: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Bradesco.

Na Figura 4.21 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 11,22 milhões e 11,78 milhões, respectivamente. Já na Figura 4.22 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram de 17,6 milhões, 11,42 milhões e 10,28 milhões, respectivamente. Por fim, a Figura 4.23 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Bradesco

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	8635124	2014-09-01T00:00:00	8635124	2014-09-01T00:00:00	8635124
2014-12-01T00:00:00	8622981	2014-12-01T00:00:00	8622981	2014-12-01T00:00:00	8622981
2015-03-01T00:00:00	9385702	2015-03-01T00:00:00	9385702	2015-03-01T00:00:00	9385702
2015-06-01T00:00:00	9740514	2015-06-01T00:00:00	9740514	2015-06-01T00:00:00	9740514
2015-09-01T00:00:00	10529218	2015-09-01T00:00:00	10529218	2015-09-01T00:00:00	10529218
2015-12-01T00:00:00	10732883	2015-12-01T00:00:00	10732883	2015-12-01T00:00:00	10732883
2016-03-01T00:00:00	11420797	2016-03-01T00:00:00	11420797	2016-03-01T00:00:00	11420797
2016-06-01T00:00:00	11915942	2016-06-01T00:00:00	11915942	2016-06-01T00:00:00	11915942
2016-09-01T00:00:00	14955758	2016-09-01T00:00:00	14955758	2016-09-01T00:00:00	14955758
2016-12-01T00:00:00	16919839	2016-12-01T00:00:00	16919839	2016-12-01T00:00:00	16919839
2017-03-01T00:00:00	15729634	2017-03-01T00:00:00	15729634	2017-03-01T00:00:00	15729634
2017-06-01T00:00:00*	17604918.6027	2017-06-01T00:00:00	13322090	2017-06-01T00:00:00	13322090
		2017-09-01T00:00:00	12148023	2017-09-01T00:00:00	12148023
		2017-12-01T00:00:00	11051836	2017-12-01T00:00:00	11051836
		2018-03-01T00:00:00	11520704	2018-03-01T00:00:00	11520704
		2018-06-01T00:00:00*	11418085.4911	2018-06-01T00:00:00	11026519
		2018-09-01T00:00:00		2018-09-01T00:00:00	10451848
		2018-12-01T00:00:00		2018-12-01T00:00:00	9819225
		2019-03-01T00:00:00		2019-03-01T00:00:00	9846433
		2019-06-01T00:00:00*		2019-06-01T00:00:00*	10280191.2858

Figura 4.22: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Bradesco.

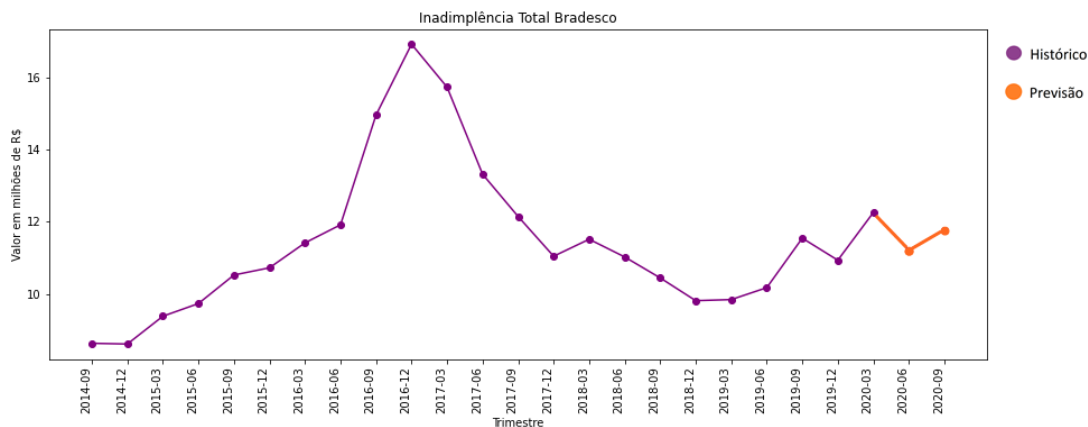


Figura 4.23: Bradesco - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.4 Caixa Econômica Federal

Para a geração dos modelos da CEF foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para a CEF foi efetuada através da execu-

ção da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 360305
```

```
==== Evaluation on training data ====
Target                                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     19            18
Mean absolute error                   0.0006        0.0009
Relative absolute error                0             0
Mean absolute percentage error         0             0
Root mean squared error                0.0008        0.0012
```

Figura 4.24: Métricas de erro do modelo preditivo para a CEF.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.24 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.25 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 13 milhões e 17,08 milhões, respectivamente. Já na Figura 4.26 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram



```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00 14368789
2014-12-01T00:00:00 14463964
2015-03-01T00:00:00 15768524
2015-06-01T00:00:00 17736223
2015-09-01T00:00:00 19213744
2015-12-01T00:00:00 20319840
2016-03-01T00:00:00 21131507
2016-06-01T00:00:00 21668072
2016-09-01T00:00:00 21671541
2016-12-01T00:00:00 17116310
2017-03-01T00:00:00 16489107
2017-06-01T00:00:00 15340214
2017-09-01T00:00:00 15266852
2017-12-01T00:00:00 14364166
2018-03-01T00:00:00 15848416
2018-06-01T00:00:00 14079021
2018-09-01T00:00:00 12958022
2018-12-01T00:00:00 12699693
2019-03-01T00:00:00 12733190
2019-06-01T00:00:00 11101761
2019-09-01T00:00:00 11388696
2019-12-01T00:00:00 10636286
2020-03-01T00:00:00 12193558
2020-06-01T00:00:00* 13001775.7372
2020-09-01T00:00:00* 17082543.2266

```

Figura 4.25: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - CEF.

de 16,85 milhões, 15,7 milhões e 12,76 milhões, respectivamente. Por fim, a Figura 4.27 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Caixa Econômica Federal

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	14368789	2014-09-01T00:00:00	14368789	2014-09-01T00:00:00	14368789
2014-12-01T00:00:00	14463964	2014-12-01T00:00:00	14463964	2014-12-01T00:00:00	14463964
2015-03-01T00:00:00	15768524	2015-03-01T00:00:00	15768524	2015-03-01T00:00:00	15768524
2015-06-01T00:00:00	17736223	2015-06-01T00:00:00	17736223	2015-06-01T00:00:00	17736223
2015-09-01T00:00:00	19213744	2015-09-01T00:00:00	19213744	2015-09-01T00:00:00	19213744
2015-12-01T00:00:00	20319840	2015-12-01T00:00:00	20319840	2015-12-01T00:00:00	20319840
2016-03-01T00:00:00	21131507	2016-03-01T00:00:00	21131507	2016-03-01T00:00:00	21131507
2016-06-01T00:00:00	21668072	2016-06-01T00:00:00	21668072	2016-06-01T00:00:00	21668072
2016-09-01T00:00:00	21671541	2016-09-01T00:00:00	21671541	2016-09-01T00:00:00	21671541
2016-12-01T00:00:00	17116310	2016-12-01T00:00:00	17116310	2016-12-01T00:00:00	17116310
2017-03-01T00:00:00	16489107	2017-03-01T00:00:00	16489107	2017-03-01T00:00:00	16489107
2017-06-01T00:00:00*	16847479.7267	2017-06-01T00:00:00	15340214	2017-06-01T00:00:00	15340214
		2017-09-01T00:00:00	15266852	2017-09-01T00:00:00	15266852
		2017-12-01T00:00:00	14364166	2017-12-01T00:00:00	14364166
		2018-03-01T00:00:00	15848416	2018-03-01T00:00:00	15848416
		2018-06-01T00:00:00*	15704779.3655	2018-06-01T00:00:00	14079021
				2018-09-01T00:00:00	12958022
				2018-12-01T00:00:00	12699693
				2019-03-01T00:00:00	12733190
				2019-06-01T00:00:00*	12756499.7313

Figura 4.26: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - CEF.

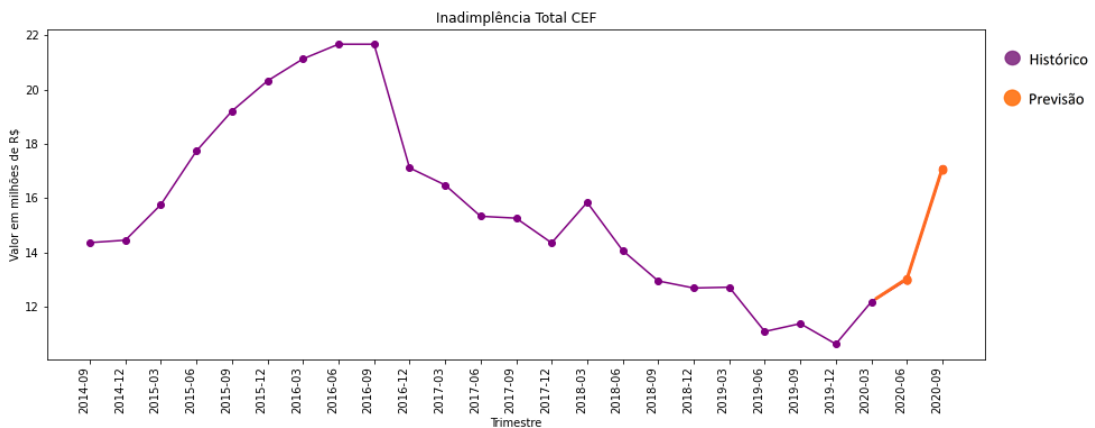


Figura 4.27: CEF - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.5 Banco CSF

Para a geração dos modelos do Banco CSF foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Banco CSF foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram man-

tidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 8357240
```

```
=== Evaluation on training data ===
Target                                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     19            18
Mean absolute error                   0.0001        0.0001
Relative absolute error                0             0
Mean absolute percentage error         0             0
Root mean squared error                0.0001        0.0001
```

Figura 4.28: Métricas de erro do modelo preditivo para o Banco CSF.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.28 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.29 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 1,59 milhão e 1,64 milhão, respectivamente. Já na Figura 4.30 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00    603170
2014-12-01T00:00:00    612001
2015-03-01T00:00:00    715829
2015-06-01T00:00:00    799273
2015-09-01T00:00:00    904174
2015-12-01T00:00:00    925519
2016-03-01T00:00:00    1008797
2016-06-01T00:00:00    1037785
2016-09-01T00:00:00    1007143
2016-12-01T00:00:00    952705
2017-03-01T00:00:00    954923
2017-06-01T00:00:00    880820
2017-09-01T00:00:00    807387
2017-12-01T00:00:00    810279
2018-03-01T00:00:00    870636
2018-06-01T00:00:00    926965
2018-09-01T00:00:00    950427
2018-12-01T00:00:00    984467
2019-03-01T00:00:00    1056962
2019-06-01T00:00:00    1130789
2019-09-01T00:00:00    1225642
2019-12-01T00:00:00    1309703
2020-03-01T00:00:00    1521144
2020-06-01T00:00:00*   1591559.8355
2020-09-01T00:00:00*   1641505.8295

```

Figura 4.29: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco CSF.

de 1 milhão, 0,88 milhão e 1,05 milhão, respectivamente. Por fim, a Figura 4.31 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Banco CSF

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	603170	2014-09-01T00:00:00	603170	2014-09-01T00:00:00	603170
2014-12-01T00:00:00	612001	2014-12-01T00:00:00	612001	2014-12-01T00:00:00	612001
2015-03-01T00:00:00	715829	2015-03-01T00:00:00	715829	2015-03-01T00:00:00	715829
2015-06-01T00:00:00	799273	2015-06-01T00:00:00	799273	2015-06-01T00:00:00	799273
2015-09-01T00:00:00	904174	2015-09-01T00:00:00	904174	2015-09-01T00:00:00	904174
2015-12-01T00:00:00	925519	2015-12-01T00:00:00	925519	2015-12-01T00:00:00	925519
2016-03-01T00:00:00	1008797	2016-03-01T00:00:00	1008797	2016-03-01T00:00:00	1008797
2016-06-01T00:00:00	1037785	2016-06-01T00:00:00	1037785	2016-06-01T00:00:00	1037785
2016-09-01T00:00:00	1007143	2016-09-01T00:00:00	1007143	2016-09-01T00:00:00	1007143
2016-12-01T00:00:00	952705	2016-12-01T00:00:00	952705	2016-12-01T00:00:00	952705
2017-03-01T00:00:00	954923	2017-03-01T00:00:00	954923	2017-03-01T00:00:00	954923
2017-06-01T00:00:00*	996406.4179	2017-06-01T00:00:00	880820	2017-06-01T00:00:00	880820
		2017-09-01T00:00:00	807387	2017-09-01T00:00:00	807387
		2017-12-01T00:00:00	810279	2017-12-01T00:00:00	810279
		2018-03-01T00:00:00	870636	2018-03-01T00:00:00	870636
		2018-06-01T00:00:00*	878865.2131	2018-06-01T00:00:00	926965
				2018-09-01T00:00:00	950427
				2018-12-01T00:00:00	984467
				2019-03-01T00:00:00	1056962
				2019-06-01T00:00:00*	1054009.6585

Figura 4.30: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco CSF.

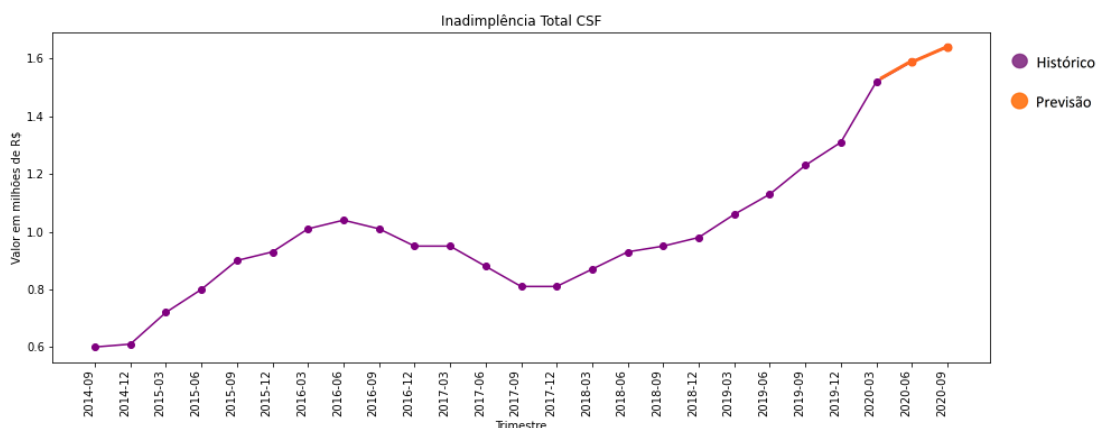


Figura 4.31: Banco CSF - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.6 Itaú

Para a geração dos modelos do Itaú foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Itaú foi efetuada através da execu-

ção da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 10069
```

```
=== Evaluation on training data ===
Target                               1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     19            18
Mean absolute error                   0.0005        0.0007
Relative absolute error                 0             0
Mean absolute percentage error          0             0
Root mean squared error                 0.0006        0.0009
```

Figura 4.32: Métricas de erro do modelo preditivo para o Itaú.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.32 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.33 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 13,74 milhões e 15,26 milhões, respectivamente. Já na Figura 4.34 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00  10188178
2014-12-01T00:00:00  10354929
2015-03-01T00:00:00  10928712
2015-06-01T00:00:00  12227485
2015-09-01T00:00:00  12674111
2015-12-01T00:00:00  12304525
2016-03-01T00:00:00  13016887
2016-06-01T00:00:00  14043276
2016-09-01T00:00:00  13401028
2016-12-01T00:00:00  11214287
2017-03-01T00:00:00  11798939
2017-06-01T00:00:00  10244683
2017-09-01T00:00:00   9711566
2017-12-01T00:00:00  10728353
2018-03-01T00:00:00  10226696
2018-06-01T00:00:00   9625295
2018-09-01T00:00:00   9800520
2018-12-01T00:00:00  10686921
2019-03-01T00:00:00  11485517
2019-06-01T00:00:00  11359498
2019-09-01T00:00:00  12079010
2019-12-01T00:00:00  12614323
2020-03-01T00:00:00  13853308
2020-06-01T00:00:00* 13742163.95
2020-09-01T00:00:00* 15261225.288

```

Figura 4.33: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Itaú.

2019 foram de 12,28 milhões, 10,06 milhões e 11,85 milhões, respectivamente. Por fim, a Figura 4.35 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Itaú

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	10188178	2014-09-01T00:00:00	10188178	2014-09-01T00:00:00	10188178
2014-12-01T00:00:00	10354929	2014-12-01T00:00:00	10354929	2014-12-01T00:00:00	10354929
2015-03-01T00:00:00	10928712	2015-03-01T00:00:00	10928712	2015-03-01T00:00:00	10928712
2015-06-01T00:00:00	12227485	2015-06-01T00:00:00	12227485	2015-06-01T00:00:00	12227485
2015-09-01T00:00:00	12674111	2015-09-01T00:00:00	12674111	2015-09-01T00:00:00	12674111
2015-12-01T00:00:00	12304525	2015-12-01T00:00:00	12304525	2015-12-01T00:00:00	12304525
2016-03-01T00:00:00	13016887	2016-03-01T00:00:00	13016887	2016-03-01T00:00:00	13016887
2016-06-01T00:00:00	14043276	2016-06-01T00:00:00	14043276	2016-06-01T00:00:00	14043276
2016-09-01T00:00:00	13401028	2016-09-01T00:00:00	13401028	2016-09-01T00:00:00	13401028
2016-12-01T00:00:00	11214287	2016-12-01T00:00:00	11214287	2016-12-01T00:00:00	11214287
2017-03-01T00:00:00	11798939	2017-03-01T00:00:00	11798939	2017-03-01T00:00:00	11798939
2017-06-01T00:00:00*	12281195.2645	2017-06-01T00:00:00	10244683	2017-06-01T00:00:00	10244683
		2017-09-01T00:00:00	9711566	2017-09-01T00:00:00	9711566
		2017-12-01T00:00:00	10728353	2017-12-01T00:00:00	10728353
		2018-03-01T00:00:00	10226696	2018-03-01T00:00:00	10226696
		2018-06-01T00:00:00*	10059642.3647	2018-06-01T00:00:00	9625295
				2018-09-01T00:00:00	9800520
				2018-12-01T00:00:00	10686921
				2019-03-01T00:00:00	11485517
				2019-06-01T00:00:00*	11851233.0388

Figura 4.34: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Itaú.

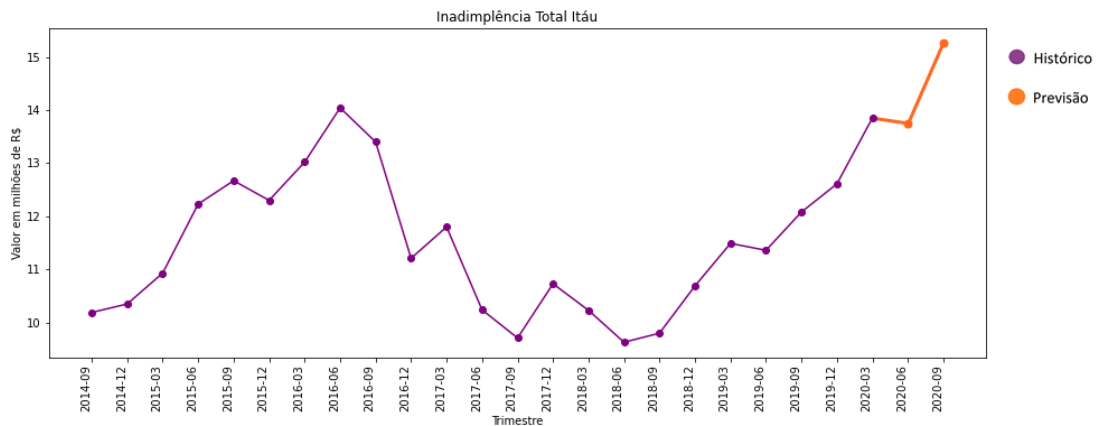


Figura 4.35: Itaú - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.7 Nubank

Para a geração dos modelos do Nubank foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Nubank foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram man-



tidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 18236120
```

```
=== Evaluation on training data ===
Target                               1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     6             5
Mean absolute error                   0.0001        0.0001
Relative absolute error                 0             0
Mean absolute percentage error          0             0
Root mean squared error                 0.0001        0.0001
```

Figura 4.36: Métricas de erro do modelo preditivo para o Nubank.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto. No caso do Nubank não foi realizada previsão para o trimestre de Junho de 2017, pois nesse trimestre o banco ainda não tinha sido criado.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida duas execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.36 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.37 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 0,75 milhão e 0,84 milhão, respectivamente. Já na Figura 4.38 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2018 e Junho de 2019 foram de 0,11 milhão

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2017-12-01T00:00:00      0
2018-03-01T00:00:00    210366
2018-06-01T00:00:00    224714
2018-09-01T00:00:00    266003
2018-12-01T00:00:00    382932
2019-03-01T00:00:00    432260
2019-06-01T00:00:00    481699
2019-09-01T00:00:00    524213
2019-12-01T00:00:00    566237
2020-03-01T00:00:00    720512
2020-06-01T00:00:00*   750221.6872
2020-09-01T00:00:00*   840003.4849

```

Figura 4.37: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Nubank.

e 0,59 milhão, respectivamente. Por fim, a Figura 4.39 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Nubank

<b>2017</b>		<b>2018</b>	
Time	INAD_TOT	Time	INAD_TOT
2017-12-01T00:00:00	0	2017-12-01T00:00:00	0
2018-03-01T00:00:00	210366	2018-03-01T00:00:00	210366
2018-06-01T00:00:00*	105183	2018-06-01T00:00:00	224714
		2018-09-01T00:00:00	266003
		2018-12-01T00:00:00	382932
		2019-03-01T00:00:00	432260
		2019-06-01T00:00:00*	594252.813

Figura 4.38: Valores previstos para os trimestres Junho de 2018 e Junho de 2019 - Nubank.

### 4.4.2.8 Banco Pan

Para a geração dos modelos do Banco Pan foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Banco Pan foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram man-

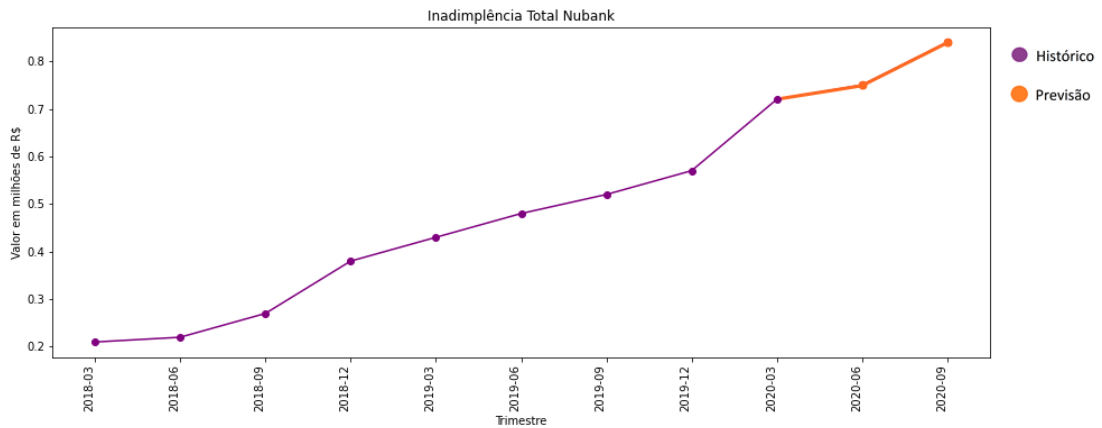


Figura 4.39: Nubank - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

tidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 31323
```

```
=== Evaluation on training data ===
Target          1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                19           18
Mean absolute error      0.0001      0.0001
Relative absolute error      0           0
Mean absolute percentage error  0           0
Root mean squared error    0.0001      0.0002
```

Figura 4.40: Métricas de erro do modelo preditivo para o Banco Pan.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.40 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00    568471
2014-12-01T00:00:00    582818
2015-03-01T00:00:00    656880
2015-06-01T00:00:00    637586
2015-09-01T00:00:00    675820
2015-12-01T00:00:00    806851
2016-03-01T00:00:00    904464
2016-06-01T00:00:00    951190
2016-09-01T00:00:00    1035361
2016-12-01T00:00:00    1113019
2017-03-01T00:00:00    1149372
2017-06-01T00:00:00    1195503
2017-09-01T00:00:00    1188456
2017-12-01T00:00:00    1036978
2018-03-01T00:00:00    1101824
2018-06-01T00:00:00    1311021
2018-09-01T00:00:00    1034179
2018-12-01T00:00:00    1151436
2019-03-01T00:00:00    1076643
2019-06-01T00:00:00    1051471
2019-09-01T00:00:00    1038577
2019-12-01T00:00:00    1025658
2020-03-01T00:00:00    1112769
2020-06-01T00:00:00*   1046204.9123
2020-09-01T00:00:00*   1003346.9437

```

Figura 4.41: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco Pan.

Na Figura 4.41 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 1,05 milhão e 1 milhão, respectivamente. Já na Figura 4.42 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram de 1,27 milhão, 1,19 milhão e 1,22 milhão, respectivamente. Por fim, a Figura 4.43 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Banco Pan

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	568471	2014-09-01T00:00:00	568471	2014-09-01T00:00:00	568471
2014-12-01T00:00:00	582818	2014-12-01T00:00:00	582818	2014-12-01T00:00:00	582818
2015-03-01T00:00:00	656880	2015-03-01T00:00:00	656880	2015-03-01T00:00:00	656880
2015-06-01T00:00:00	637586	2015-06-01T00:00:00	637586	2015-06-01T00:00:00	637586
2015-09-01T00:00:00	675820	2015-09-01T00:00:00	675820	2015-09-01T00:00:00	675820
2015-12-01T00:00:00	806851	2015-12-01T00:00:00	806851	2015-12-01T00:00:00	806851
2016-03-01T00:00:00	904464	2016-03-01T00:00:00	904464	2016-03-01T00:00:00	904464
2016-06-01T00:00:00	951190	2016-06-01T00:00:00	951190	2016-06-01T00:00:00	951190
2016-09-01T00:00:00	1035361	2016-09-01T00:00:00	1035361	2016-09-01T00:00:00	1035361
2016-12-01T00:00:00	1113019	2016-12-01T00:00:00	1113019	2016-12-01T00:00:00	1113019
2017-03-01T00:00:00	1149372	2017-03-01T00:00:00	1149372	2017-03-01T00:00:00	1149372
2017-06-01T00:00:00*	1270769.3723	2017-06-01T00:00:00	1195503	2017-06-01T00:00:00	1195503
		2017-09-01T00:00:00	1188456	2017-09-01T00:00:00	1188456
		2017-12-01T00:00:00	1036978	2017-12-01T00:00:00	1036978
		2018-03-01T00:00:00	1101824	2018-03-01T00:00:00	1101824
		2018-06-01T00:00:00*	1187396.7553	2018-06-01T00:00:00	1311021
				2018-09-01T00:00:00	1034179
				2018-12-01T00:00:00	1151436
				2019-03-01T00:00:00	1076643
				2019-06-01T00:00:00*	1215045.1793

Figura 4.42: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco Pan.

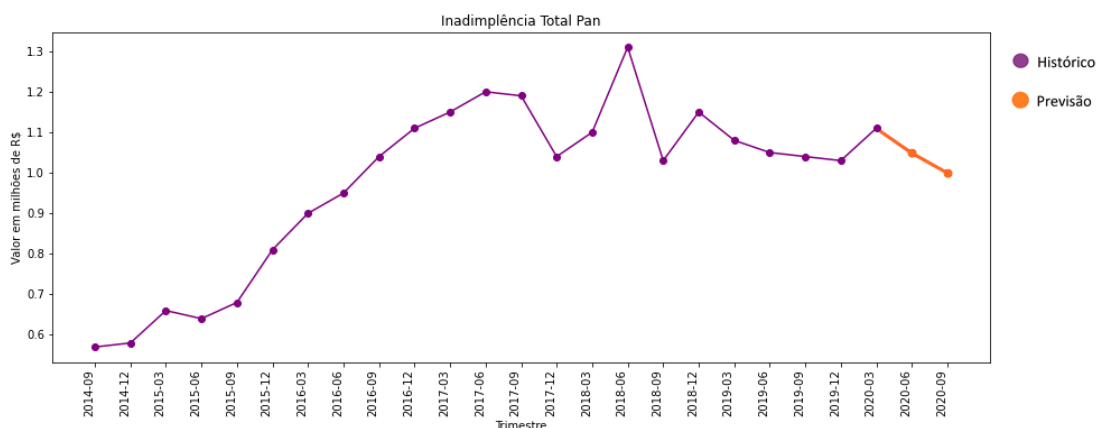


Figura 4.43: Banco Pan - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.9 Santander

Para a geração dos modelos do Santander foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados

obtidos. A obtenção dos dados para criação do modelo para o Santander foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 30379
```

```
=== Evaluation on training data ===
Target                                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                                     19            18
Mean absolute error                   0.0002        0.0004
Relative absolute error                 0             0
Mean absolute percentage error          0             0
Root mean squared error                 0.0003        0.0004
```

Figura 4.44: Métricas de erro do modelo preditivo para o Santander.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.44 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.45 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 10,04 milhões e 11,03

```

=== Future predictions from end of training data ===
Time                               INAD_TOT
2014-09-01T00:00:00                7057964
2014-12-01T00:00:00                6526922
2015-03-01T00:00:00                6274248
2015-06-01T00:00:00                6658404
2015-09-01T00:00:00                6867767
2015-12-01T00:00:00                7192608
2016-03-01T00:00:00                7337545
2016-06-01T00:00:00                7028751
2016-09-01T00:00:00                6594212
2016-12-01T00:00:00                6596539
2017-03-01T00:00:00                7444209
2017-06-01T00:00:00                7179343
2017-09-01T00:00:00                7452049
2017-12-01T00:00:00                7995275
2018-03-01T00:00:00                7237281
2018-06-01T00:00:00                7752220
2018-09-01T00:00:00                7914844
2018-12-01T00:00:00                8547391
2019-03-01T00:00:00                8551975
2019-06-01T00:00:00                8781584
2019-09-01T00:00:00                8634971
2019-12-01T00:00:00                8860597
2020-03-01T00:00:00                9038439
2020-06-01T00:00:00*              10038114.5947
2020-09-01T00:00:00*              11030832.9553

```

Figura 4.45: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Santander.

milhões, respectivamente. Já na Figura 4.46 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram de 7,61 milhões, 7,66 milhões e 8,47 milhões, respectivamente. Por fim, a Figura 4.47 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Santander

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	7057964	2014-09-01T00:00:00	7057964	2014-09-01T00:00:00	7057964
2014-12-01T00:00:00	6526922	2014-12-01T00:00:00	6526922	2014-12-01T00:00:00	6526922
2015-03-01T00:00:00	6274248	2015-03-01T00:00:00	6274248	2015-03-01T00:00:00	6274248
2015-06-01T00:00:00	6658404	2015-06-01T00:00:00	6658404	2015-06-01T00:00:00	6658404
2015-09-01T00:00:00	6867767	2015-09-01T00:00:00	6867767	2015-09-01T00:00:00	6867767
2015-12-01T00:00:00	7192608	2015-12-01T00:00:00	7192608	2015-12-01T00:00:00	7192608
2016-03-01T00:00:00	7337545	2016-03-01T00:00:00	7337545	2016-03-01T00:00:00	7337545
2016-06-01T00:00:00	7028751	2016-06-01T00:00:00	7028751	2016-06-01T00:00:00	7028751
2016-09-01T00:00:00	6594212	2016-09-01T00:00:00	6594212	2016-09-01T00:00:00	6594212
2016-12-01T00:00:00	6596539	2016-12-01T00:00:00	6596539	2016-12-01T00:00:00	6596539
2017-03-01T00:00:00	7444209	2017-03-01T00:00:00	7444209	2017-03-01T00:00:00	7444209
2017-06-01T00:00:00*	7606307.2014	2017-06-01T00:00:00	7179343	2017-06-01T00:00:00	7179343
		2017-09-01T00:00:00	7452049	2017-09-01T00:00:00	7452049
		2017-12-01T00:00:00	7995275	2017-12-01T00:00:00	7995275
		2018-03-01T00:00:00	7237281	2018-03-01T00:00:00	7237281
		2018-06-01T00:00:00*	7662310.1003	2018-06-01T00:00:00	7752220
				2018-09-01T00:00:00	7914844
				2018-12-01T00:00:00	8547391
				2019-03-01T00:00:00	8551975
				2019-06-01T00:00:00*	8465941.2013

Figura 4.46: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Santander.

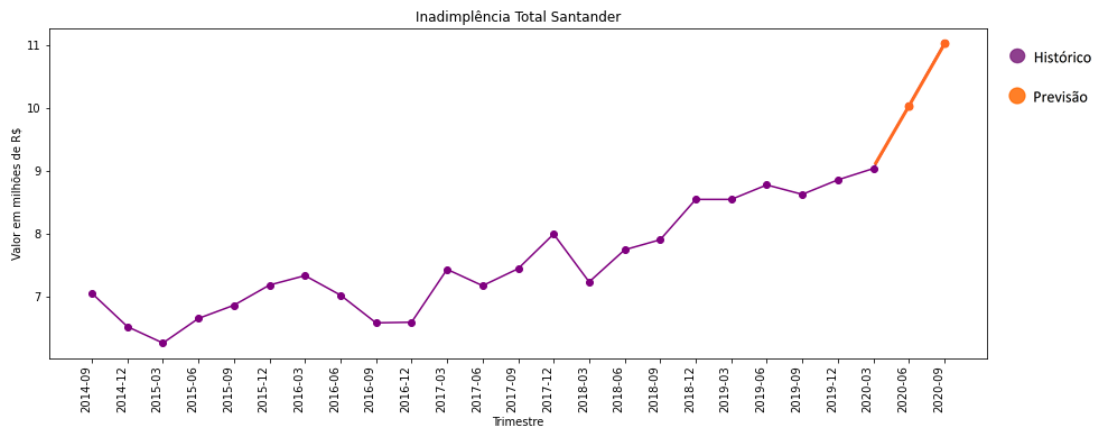


Figura 4.47: Santander - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.2.10 Banco Votorantim

Para a geração dos modelos do Banco Votorantim foram seguidos os mesmos passos descritos na seção 4.4.2.1, alterando-se as consultas para obtenção dos dados no *Weka* e os resultados obtidos. A obtenção dos dados para criação do modelo para o Banco Votoran-



tim foi efetuada através da execução da consulta SQL abaixo, onde os demais parâmetros foram mantidos modificando-se apenas o atributo 'COD', que representa o código identificador da instituição financeira.

```
select "DATA", "ATIV_TOT", "CC_CLASS", "CAPT", "PATR_LIQ",
"OPER_CRE_PROV", "PASS_TOT", "DESP_PESS", "DESP_ADM", "DESP_TRIB",
"TOT_CAR_PF", "TOT_CAR_PJ", "TOT_GER", "RISCO_AA", "RISCO_A", "RISCO_B",
"RISCO_C", "RISCO_D", "RISCO_E", "RISCO_F", "RISCO_G", "RISCO_H",
"TOT_EXT", "INAD_TOT"
from public.ifdata_juncao_06
where "DATA" <> '2020-06-01 00:00:00.0' and "COD" = 51011
```

```
=== Evaluation on training data ===
Target                1-step-ahead  2-steps-ahead
=====
INAD_TOT
N                    19             18
Mean absolute error   0.0001         0.0001
Relative absolute error  0              0
Mean absolute percentage error  0              0
Root mean squared error 0.0001         0.0001
```

Figura 4.48: Métricas de erro do modelo preditivo para o Banco Votorantim.

Do mesmo modo, o processo foi replicado de forma a se obter as previsões para os trimestres de Junho de 2017, Junho de 2018 e Junho de 2019, substituindo os parâmetros da cláusula *where* para limitação dos dados ao trimestre imediatamente anterior ao trimestre a ser previsto.

Sendo assim, foi realizada uma primeira execução para avaliação do modelo e obtenção das métricas de erro, uma nova execução para a previsão dos valores de Junho e Setembro de 2020 e em seguida três execuções com dados anteriores para realizar a previsão dos trimestres Junho de 2017, Junho de 2018 e Junho de 2019. Por último foi feita a geração do gráfico com a evolução da série temporal e os valores previstos para os dois últimos trimestres.

A Figura 4.48 apresenta as métricas de erro, onde se pode observar que os valores foram bem próximos de 0 com a utilização de dados de treino, tanto para previsão de um trimestre à frente quanto para previsão de dois trimestres à frente.

Na Figura 4.49 observa-se que os valores alcançados para inadimplência total referente aos trimestres de Junho de 2020 e Setembro de 2020 foram de 1,52 milhão e 1,83 milhão, respectivamente. Já na Figura 4.50 pode-se verificar que os valores previstos para inadimplência total nos trimestres de Junho de 2017, Junho de 2018 e Junho de 2019 foram

```

=== Future predictions from end of training data ===
Time                INAD_TOT
2014-09-01T00:00:00    2115855
2014-12-01T00:00:00    2282550
2015-03-01T00:00:00    2362643
2015-06-01T00:00:00    1715719
2015-09-01T00:00:00    1714147
2015-12-01T00:00:00    1551588
2016-03-01T00:00:00    1303599
2016-06-01T00:00:00    1826052
2016-09-01T00:00:00    1716947
2016-12-01T00:00:00    1675712
2017-03-01T00:00:00    1316912
2017-06-01T00:00:00    1197535
2017-09-01T00:00:00    1165424
2017-12-01T00:00:00    1108139
2018-03-01T00:00:00    1114250
2018-06-01T00:00:00    1107380
2018-09-01T00:00:00    1110675
2018-12-01T00:00:00    1123810
2019-03-01T00:00:00    1216377
2019-06-01T00:00:00    1302620
2019-09-01T00:00:00    1343062
2019-12-01T00:00:00    1405553
2020-03-01T00:00:00    1530052
2020-06-01T00:00:00*   1521774.3806
2020-09-01T00:00:00*   1834879.9901

```

Figura 4.49: Valores previstos para os trimestres de Junho de 2020 e Setembro de 2020 - Banco Votorantim.

de 1,53 milhão, 1,18 milhão e 0,97 milhão, respectivamente. Por fim, a Figura 4.51 apresenta o gráfico da evolução da série temporal, acrescentando a previsão realizada para os trimestres de Junho de 2020 e Setembro de 2020.

## Banco Votorantim

2017		2018		2019	
Time	INAD_TOT	Time	INAD_TOT	Time	INAD_TOT
2014-09-01T00:00:00	2115855	2014-09-01T00:00:00	2115855	2014-09-01T00:00:00	2115855
2014-12-01T00:00:00	2282550	2014-12-01T00:00:00	2282550	2014-12-01T00:00:00	2282550
2015-03-01T00:00:00	2362643	2015-03-01T00:00:00	2362643	2015-03-01T00:00:00	2362643
2015-06-01T00:00:00	1715719	2015-06-01T00:00:00	1715719	2015-06-01T00:00:00	1715719
2015-09-01T00:00:00	1714147	2015-09-01T00:00:00	1714147	2015-09-01T00:00:00	1714147
2015-12-01T00:00:00	1551588	2015-12-01T00:00:00	1551588	2015-12-01T00:00:00	1551588
2016-03-01T00:00:00	1303599	2016-03-01T00:00:00	1303599	2016-03-01T00:00:00	1303599
2016-06-01T00:00:00	1826052	2016-06-01T00:00:00	1826052	2016-06-01T00:00:00	1826052
2016-09-01T00:00:00	1716947	2016-09-01T00:00:00	1716947	2016-09-01T00:00:00	1716947
2016-12-01T00:00:00	1675712	2016-12-01T00:00:00	1675712	2016-12-01T00:00:00	1675712
2017-03-01T00:00:00	1316912	2017-03-01T00:00:00	1316912	2017-03-01T00:00:00	1316912
2017-06-01T00:00:00*	1529152.3598	2017-06-01T00:00:00	1197535	2017-06-01T00:00:00	1197535
		2017-09-01T00:00:00	1165424	2017-09-01T00:00:00	1165424
		2017-12-01T00:00:00	1108139	2017-12-01T00:00:00	1108139
		2018-03-01T00:00:00	1114250	2018-03-01T00:00:00	1114250
		2018-06-01T00:00:00*	1178490.8194	2018-06-01T00:00:00	1107380
				2018-09-01T00:00:00	1110675
				2018-12-01T00:00:00	1123810
				2019-03-01T00:00:00	1216377
				2019-06-01T00:00:00*	965506.0749

Figura 4.50: Valores previstos para os trimestres Junho de 2017, Junho de 2018 e Junho de 2019 - Banco Botorantim.

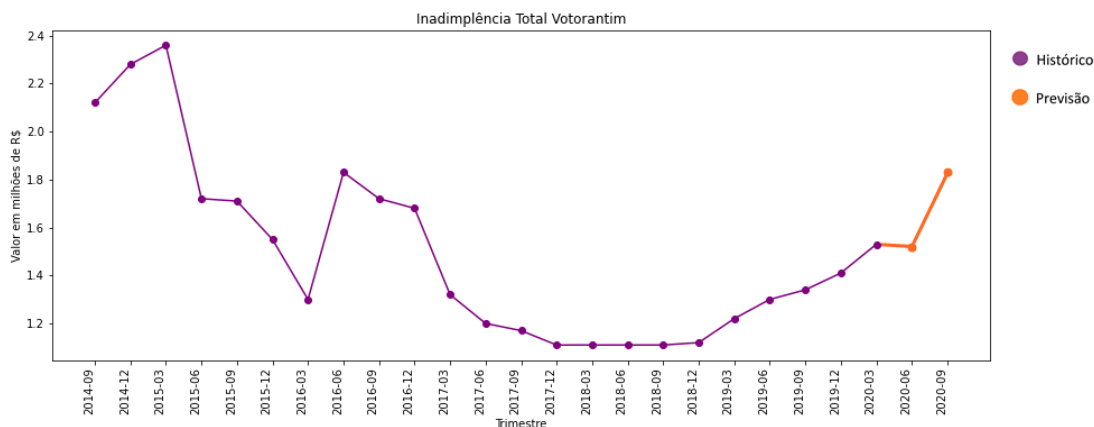


Figura 4.51: Banco Votorantim - Evolução da inadimplência total entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com o modelo.

### 4.4.3 Comparação entre as Instituições Financeiras

Ao final do processo de geração dos modelos e obtenção dos resultados, foi gerado um gráfico com a série da inadimplência de todas as Instituições Financeiras, de forma que os valores observados e previstos de cada uma delas pudessem ser comparados mais facil-

mente. O gráfico pode ser visualizado na Figura 4.52, sendo os valores de inadimplência de abril/2014 até março/2020 observados nos dados do histórico da instituição e os valores de junho/2020 e setembro/2020 previsões geradas pelos modelos.

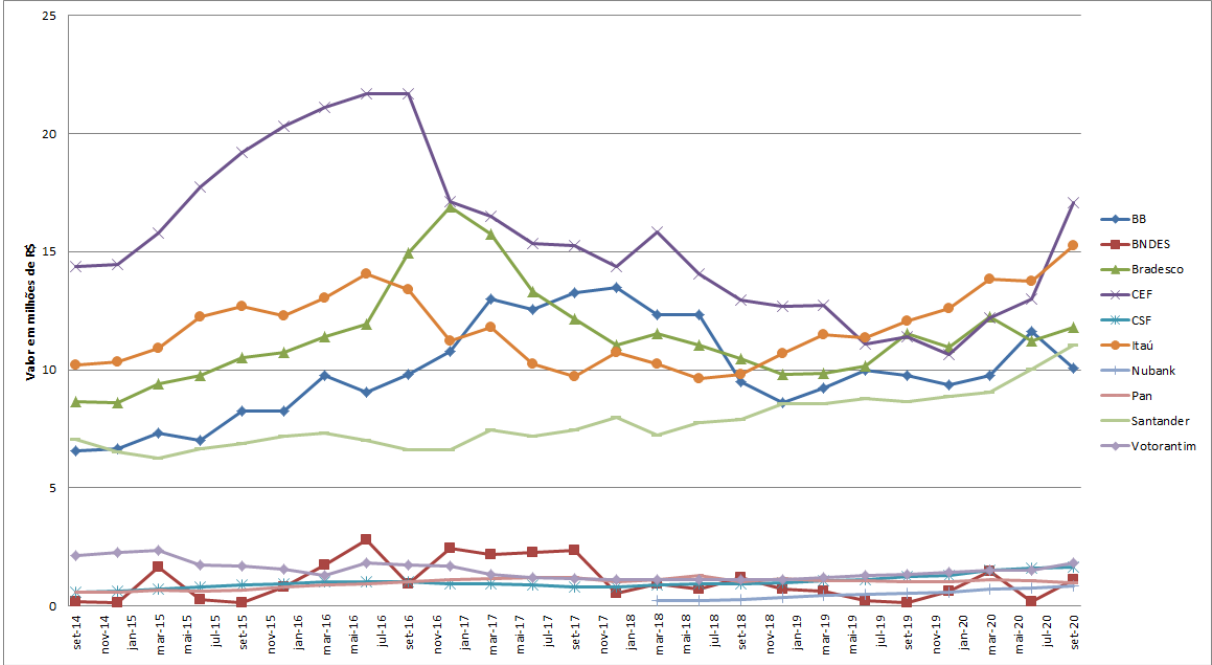


Figura 4.52: Evolução da inadimplência total das 10 IFs entre Setembro de 2014 e Setembro de 2020 - os valores dos dois últimos trimestres são previsões realizadas com os modelos.

# Capítulo 5

## Discussão dos Resultados

Neste capítulo, será feita uma discussão acerca dos resultados obtidos com a aplicação dos processos de mineração de dados relatados no Capítulo 4. Além disso, serão abordadas algumas estratégias que foram utilizadas como forma de validação desses resultados.

### 5.1 Previsão da Inadimplência

#### 5.1.1 Comparação com os dados reais

Após a obtenção das previsões para as instituições financeiras escolhidas como objeto de análise, os resultados foram agrupados de modo a serem comparados com os dados reais divulgados pelo Bacen para os trimestres escolhidos como referência. Com essa comparação, é possível verificar a proximidade dos valores previstos pelo modelo preditivo com os valores reais. Essa possibilidade existe nos casos em que os dados oficiais já estavam disponíveis.

As comparações dos trimestres de anos anteriores contribuem para a avaliação da confiabilidade da aplicação da regressão sobre os dados. O percentual de erro apresentado nas previsões para esses anos pode ser comparado ao percentual de erro obtido para as previsões do ano de 2020.

Dentre os resultados obtidos, as previsões geradas para a instituição BNDES apresentaram um desempenho fora da curva em relação às demais instituições analisadas. A diferença observada entre o valor previsto e o valor real para o BNDES foi de 90,83% no trimestre encerrado em junho de 2020. Já no trimestre encerrado em junho de 2019, essa diferença foi de -133,33%.

Uma hipótese levantada para essa grande diferença são os picos e quedas observados na série histórica da inadimplência total dessa instituição, que podem ser visualizados no gráfico da Figura 4.19.

Uma possível explicação para esse comportamento é a composição da carteira de crédito da instituição, que está concentrada em operações com pessoas jurídicas de grande porte. Nesse sentido, crises econômicas e eventos adversos que afetam macrossetores específicos podem alterar de maneira substancial a inadimplência observada em determinado período.

Isso faria com que a previsibilidade da inadimplência do BNDES fosse menor do que a das instituições que possuem uma carteira de crédito mais pulverizada, já que tem maior exposição a eventos externos aos dados utilizados para criação dos modelos preditivos do presente trabalho.

Portanto, optou-se por suprimir os resultados obtidos para a instituição BNDES nas demais comparações realizadas a seguir, desconsiderando o efeito das previsões obtidas para essa instituição nas médias de erro apresentadas.

#### 5.1.1.1 Trimestre encerrado em Junho de 2020

A comparação foi iniciada pelos dados do trimestre encerrado em Junho de 2020, que foram os primeiros a serem produzidos pelo modelo construído. A relação dos valores previstos em confronto com os valores reais observados está exposta na Tabela 5.1.

Através da análise da tabela, pode-se inferir que a média de erro percentual absoluta, que considera os modelos das 9 instituições financeiras, foi de 16,05%. Assim como houveram apenas 3 instituições em que o percentual de erro foi maior que 20%. Também é possível verificar nos resultados obtidos que entre as 9 instituições financeiras analisadas, 7 tiveram um valor previsto maior do que o valor real. Somente as instituições Banco CSF e Banco Pan tiveram o valor real maior do que o valor previsto, sendo que no caso do Banco CSF o percentual de erro foi de apenas 1,13%.

Tabela 5.1: Comparação dos valores previstos com os valores reais em milhões de R\$ - Trimestre Junho de 2020.

Instituição Financeira	Valor Previsto	Valor Real	Diferença
BB	11,63	9,02	-28,92%
Bradesco	11,22	9,82	-14,24%
CEF	13	10,71	-21,40%
Banco CSF	1,59	1,61	1,13%
Itaú	13,74	12,82	-7,14%
Nubank	0,75	0,73	-3,06%
Banco Pan	1,05	1,25	16,10%
Santander	10,04	6,97	-44,09%
Banco Votorantim	1,52	1,40	-8,35%

Dentre os fatores que podem ter sido consideravelmente impactantes para que a maioria das previsões tenham gerado valores maiores do que os valores reais, estão as medidas adotadas pelas instituições financeiras por conta da pandemia da COVID-19. A pandemia gerou uma redução na renda e capacidade de pagamento de muitos brasileiros [25], e com isso, as instituições financeiras adotaram algumas medidas para evitar a inadimplência de seus clientes. Uma das medidas tomadas foi a prorrogação do vencimento das parcelas das operações por 60 dias [26].

A aplicação dessas medidas provavelmente foi responsável pela estabilização da inadimplência nesse período, o que fez com que os valores aumentassem abaixo do esperado no período em referência, resultando em previsões de valores maiores do que o observado.

Outro ponto de destaque é o resultado da comparação do Banco Pan, em que a diferença foi de 16,10%. Isso pode ser justificado pela postura adotada pelo banco em relação à prorrogação de parcelas das dívidas. Essa prorrogação ficou limitada à algumas operações com garantia, o que afetou apenas 1% de sua carteira de crédito [27]. Deste modo, ao contrário das demais instituições em que as ações tomadas em relação à pandemia tiveram um efeito de redução da inadimplência total, no Banco Pan houve um acréscimo da inadimplência comparado ao esperado por conta da limitação das medidas adotadas.

Dado o exposto, mesmo verificada a média de erro apresentada pelos modelos e a existência de excepcionalidades, a previsão da inadimplência total das instituições financeiras pôde ser realizada com algum nível de precisão, pois, observou-se que exceto em casos pontuais o valor previsto aproximou-se do valor real. Mesmo que o valor não seja totalmente preciso, ele pode servir como uma indicação de possíveis cenários, de maneira a auxiliar na tomada de decisões.

#### **5.1.1.2 Trimestres encerrados em Junho de 2017, Junho de 2018 e Junho de 2019**

De modo a performar uma validação dos modelos preditivos gerados, foram realizadas previsões dos valores para os trimestres encerrados em Junho de 2017, Junho de 2018 e Junho de 2019, utilizando os dados até o trimestre imediatamente anterior como base. Tal validação foi realizada para que os modelos pudessem ser testados em um cenário onde não há efeitos da pandemia da COVID-19, conforme detalhado na seção 4.4.2. Nesse sentido, os resultados obtidos também foram comparados aos dados reais para verificação dos percentuais de erro. O resultado pode ser visualizado na Tabela 5.2.

O exame da tabela mostra que a média de erro percentual absoluta foi de 15,46% em 2017, 10,66% em 2018 e 11,88% em 2019. Das 9 instituições financeiras analisadas apenas o Nubank e o Banco Votorantim tiveram um percentual de erro maior que 20% em mais de um trimestre referência. Comparando os resultados do trimestre Junho de

Tabela 5.2: Diferenças percentuais entre os valores previstos e os valores reais - Trimestres Junho de 2017, Junho de 2018, Junho de 2019 e Junho de 2020.

<b>Instituição Financeira</b>	<b>Erro 2017</b>	<b>Erro 2018</b>	<b>Erro 2019</b>	<b>Erro 2020</b>
BB	-8,83%	-4,45%	11,60%	-28,92%
Bradesco	-32,13%	-3,54%	-0,98%	-14,24%
CEF	-9,84%	-11,51%	-14,95%	-21,40%
Banco CSF	-13,64%	5,38%	7,08%	1,13%
Itaú	-19,92%	-4,47%	-4,31%	-7,14%
Nubank		50%	-22,92%	-3,06%
Banco Pan	-5,83%	9,16%	-16,19%	16,10%
Santander	-5,99%	1,16%	3,53%	-44,09%
Banco Votorantim	-27,5%	-6,31%	25,38%	-8,35%

2019 aos resultados do trimestre de Junho de 2020, as instituições Banco CSF, Nubank e Banco Votorantim tiveram um maior erro percentual nas previsões realizadas para junho de 2019.

Em relação ao Nubank, que foi criado em 2013, por ser um banco relativamente novo a quantidade de trimestres de histórico disponíveis é bem menor do que das demais instituições financeiras. Estão disponíveis apenas os dados de 6 trimestres até Junho de 2019 em contraposição aos 19 trimestres disponíveis para as demais instituições analisadas. Já até Junho de 2018 a disponibilidade é de apenas 2 trimestres, sendo que o primeiro deles registra 0 no valor informado de inadimplência total.

No trimestre de Junho de 2017 o Nubank ainda não possuía dados disponíveis no portal IF.Data, desse modo, a Tabela 5.2 apresenta um valor vazio para esse trimestre. Esse é um fator que pode ter afetado consideravelmente a eficácia do modelo preditivo que foi gerado. A média de erro percentual absoluta cai para 5,75% em 2018 se desconsiderado o Nubank.

Quanto aos Bancos CSF e Votorantim, não foi possível identificar os motivos pelos quais as diferenças dos trimestres de anos anteriores se apresentaram maiores do que as diferenças do trimestre Junho de 2020.

Por fim, esclarecidos os casos pontuais acima, para a maioria das instituições financeiras o modelo preditivo teve uma performance melhor nos trimestres dos anos anteriores do que no trimestre de Junho de 2020. Isso mostra que o erro apresentado nas previsões do trimestre Junho de 2020 foi maior que o esperado de acordo com as previsões realizadas para os trimestres dos anos anteriores. Tendo em vista que nos anos anteriores a pandemia da COVID-19 não afetou os resultados da inadimplência nas instituições financeiras, é fortalecida a hipótese levantada na seção 5.1.1.1 de que seus efeitos tiveram impacto nas previsões realizadas para o trimestre de Junho de 2020.



Tabela 5.3: Diferenças percentuais entre os valores previstos para o trimestre encerrado em junho de 2020 e a média dos valores dos últimos 3 anos (2017, 2018 e 2019).

<b>IF</b>	<b>Previsão</b>	<b>Média</b>	<b>Valor Real</b>	<b>Erro Previsão</b>	<b>Erro Média</b>
BB	11,63	11,64	9,02	-28,92%	-29,03%
Bradesco	11,22	11,51	9,82	-14,24%	-17,17%
CEF	13,00	13,51	10,71	-21,40%	-26,14%
Banco CSF	1,59	0,98	1,61	1,13%	39,04%
Itaú	13,74	10,41	12,82	-7,14%	18,83%
Nubank	0,75	0,23	0,73	-3,06%	67,86%
Banco Pan	1,05	1,19	1,25	16,10%	5,14%
Santander	10,04	7,90	6,97	-44,09%	-13,44%
Banco Votorantim	1,52	1,20	1,40	-8,35%	14,16%
<b>Erro Médio</b>				16,05%	25,65%

### 5.1.2 Comparação com a média simples

Outra estratégia utilizada para validação das previsões realizadas foi a comparação dos valores obtidos com a média dos valores dos últimos 3 anos. Nesse caso, utilizou-se a média entre os valores observados nos trimestres encerrados em junho de 2017, junho de 2018 e junho de 2019.

A comparação pode ser visualizada na Tabela 5.3, onde observa-se que em 7 dos 9 casos (77,8%) as previsões realizadas com a utilização do modelo foram melhores do que a utilização da média simples. Além disso, o erro percentual médio foi 9,6% menor nas previsões realizadas com o modelo, em comparação ao percentual de erro obtido com a utilização da média simples.

Entre os casos que apresentaram melhor performance com a utilização da média, estão o Banco Pan e o Banco Santander. Quanto ao Banco Pan, os prováveis motivos para que a média tenha obtido melhor desempenho podem estar relacionados aos fatores que já foram levantados anteriormente na 5.1.1.1. Já em relação ao Banco Santander, não foram encontrados indicativos que possam explicar o motivo pelo qual a média se apresentou uma melhor opção do que a previsão realizada com a utilização do modelo.

Dessa forma, ressalvando-se os casos acima citados, nota-se que a utilização do modelo gerado como forma de se obter uma previsão da inadimplência das instituições financeiras obteve melhor desempenho do que a utilização da média simples dos últimos 3 anos.

## 5.2 Clusterização

A última etapa da análise dos dados foi a clusterização, esse processo consiste em um agrupamento dos dados com base em suas características. Utilizou-se a clusterização hierárquica de modo a classificar as instituições financeiras quanto à proximidade entre elas. Tal análise pode indicar possíveis semelhanças entre as instituições financeiras que facilitam a compreensão de seu comportamento.

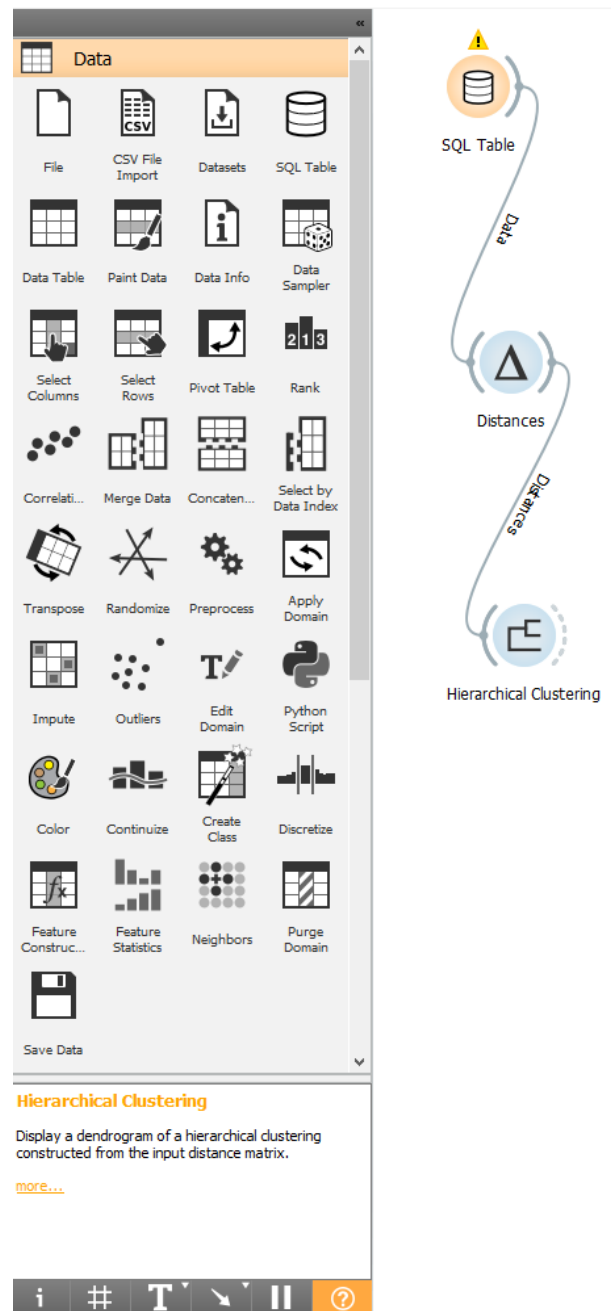


Figura 5.1: Clusterização - Processo de geração dos dendogramas usando o Orange.

Nesta análise, foram utilizadas todas as colunas disponíveis de cada instituição financeira relacionadas no Anexo II, inclusive as colunas que não foram utilizadas para a realização da previsão da inadimplência total. Os valores obtidos nas previsões de inadimplência não foram utilizados, pois não possuem a mesma confiabilidade dos demais dados. A condução dessa etapa se deu com o uso da ferramenta *Orange*. Procedeu-se com a obtenção dos dados através da conexão com o banco de dados no *PostgreSQL*, que já havia sido criado para a realização da previsão da inadimplência total. A consulta limitou-se ao trimestre de Março de 2020 para que a comparação fosse realizada sob a ótica do mesmo período de referência.

Com os dados já importados no *Orange*, os módulos *Distances* e *Hierarchical Clustering* foram acionados. O primeiro tem o papel de calcular as distâncias entre as instituições, tomando a distância euclidiana como métrica, já o segundo consome os resultados do primeiro para gerar os dendogramas, que consistem em gráficos com a finalidade de observar os resultados da clusterização. O processo de geração dos dendogramas no *Orange* pode ser visualizado na figura Figura 5.1.

Deste modo, foram gerados dois dendogramas, que podem ser visualizados nas Figuras 5.2 a 5.3. A construção do primeiro dendograma utilizou apenas as 10 instituições financeiras que foram selecionadas na seção anterior para a geração das previsões da inadimplência total, já para a construção do segundo, todas as instituições financeiras da base de dados foram aproveitadas, limitando a visualização à 32 linhas no gráfico.

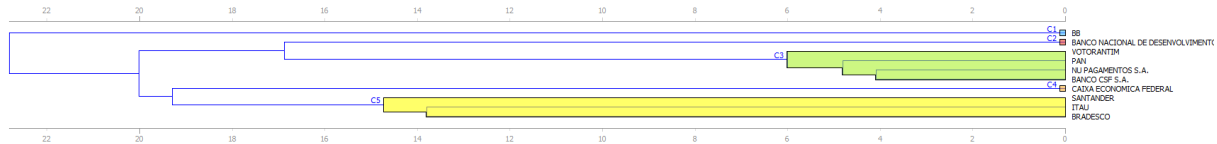


Figura 5.2: Dendograma 1 - 10 instituições financeiras selecionadas.

Em ambos dendogramas, definiu-se como ponto de corte o agrupamento das instituições financeiras em 5 *clusters* de acordo com a proximidade entre elas, nomeados de C1 à C5. Um ponto interessante é que 4 dos 5 *clusters* são iguais nos dois dendogramas. Apenas o C3, que é o *cluster* com maior número de instituições financeiras, diverge entre os dendogramas. Os *clusters* C1, C2 e C4 incluem apenas uma instituição financeira cada, sendo o BB, BNDES e CEF, respectivamente.

O provável motivo do isolamento dessas instituições são as composições de suas carteiras de crédito. O BB concentra boa parte em agronegócios, enquanto o BNDES em serviços industriais e a CEF em financiamento habitacional. Os bancos Santander, Itaú e Bradesco que possuem carteiras de crédito mais diversificadas, mas que ainda concentram grandes valores dentro do SFN, foram agrupados no *cluster* C5.

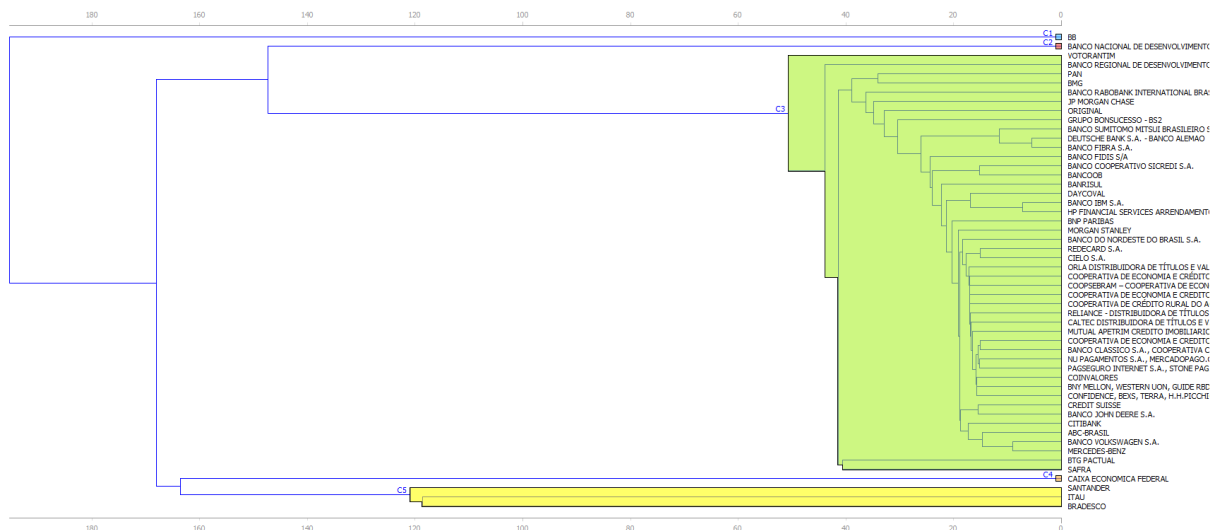


Figura 5.3: Dendograma 2 - Todas as instituições financeiras selecionadas.

Já o *cluster* C3 inclui os bancos Votorantim, Pan, Nubank e CSF no primeiro dendograma e todas as demais instituições financeiras no segundo dendograma. Nos dois casos, são instituições financeiras com menor expressividade no SFN e que possuem valores inferiores concentrados em suas carteiras de crédito. Num maior detalhamento, é possível verificar no *cluster* C3 do segundo dendograma que as instituições financeiras atuantes na mesma área tem maior proximidade dentro do *cluster*. Alguns exemplos são o Nubank, PagSeguro, MercadoPago e Stone, atuantes sobretudo no setor de pagamentos e cartões e os Bancos Volkswagen e Mercedes-Benz, que atuam prioritariamente com financiamento de veículos.

Os resultados da previsão da inadimplência total também guardam alguma relação com a clusterização. Por exemplo, os bancos Itaú, Santander e Bradesco que estão situados no *cluster* C5 tiveram resultados semelhantes na previsão da inadimplência total. Já as instituições que apresentaram baixos valores previstos e observados de inadimplência total estão agrupadas no *cluster* C3.

Sendo assim, a clusterização por si só não apresenta resultados tão significativos em um primeiro momento, entretanto, quando combinada com outra ferramenta de análise, pode contribuir na identificação de padrões que facilitam o entendimento dos dados.

# Capítulo 6

## Considerações Finais

O objetivo do trabalho foi realizar um processo de mineração e análise de dados das instituições financeiras que atuam no Brasil. Para isso, os dados do Portal IF.Data do Bacen serviram como fonte. O principal problema abordado foi a previsão da inadimplência total das 10 instituições financeiras que apresentaram maior relevância nesse indicador considerando o histórico avaliado.

O primeiro passo realizado foi a obtenção dos dados no Portal IF.data em formato *.csv*, e em seguida foi realizada a sua importação e tratamento com a utilização da linguagem *Python* e a sua carga em um banco de dados com o auxílio da ferramenta *PostgreSQL*.

Depois da importação e carga dos dados, a fase de análise foi desempenhada com a utilização do algoritmo de regressão linear múltipla, através da ferramenta *Weka* com a inclusão do pacote *timeseriesForecast*. A realização dessa etapa produziu os modelos preditivos necessários para obtenção dos valores para os quais a previsão foi proposta.

Feito isso, as previsões obtidas foram comparadas aos valores reais, tanto para o trimestre de Junho de 2020, quanto para os trimestres do mesmo mês de anos anteriores: 2017, 2018 e 2019, que eram trimestres com valores previstos que também possuíam os valores reais já divulgados pelo Bacen. Além disso, também foi feita uma comparação com a média simples da inadimplência dos últimos 3 anos. A realização das previsões e das comparações sucedeu na identificação de alguns resultados não triviais que puderam gerar algum conhecimento. Como acréscimo, também foi realizado o processo de clusterização que identificou a relação de proximidade entre algumas instituições financeiras que operam em nichos semelhantes, além de contribuir para a compreensão dos resultados da previsão de inadimplência total.

Dentre as limitações da metodologia utilizada, estão a existência de variáveis que não foram incluídas nos modelos gerados, pois não estão disponíveis na fonte escolhida. Além disso, existem diferenças entre as estratégias utilizadas pelas instituições financeiras para lidar com o problema da inadimplência, sendo que algumas instituições detalham as

estratégias empregadas e outras não, o que limita a identificação de padrões e dificulta a análise dos resultados. Há também a ocorrência da pandemia da COVID-19 durante o período de referência dos dados analisados, que representa uma variação no cenário político e econômico que possui baixa previsibilidade.

Uma possibilidade de trabalho futuro, seria a realização da previsão da inadimplência total com a adição de outras fontes de dados, como indicadores socioeconômicos, macroeconômicos e pesquisas de mercado. A adição de outras variáveis, como o endividamento das famílias brasileiras, a taxa de desemprego, a renda média dos trabalhadores, o histórico de inflação e de taxas de juros, por exemplo, poderiam reduzir significativamente o erro nas previsões, já que essas variáveis podem ter impacto relevante nos valores de inadimplência.

Outra sugestão é a realização de um processo semelhante para outras variáveis de instituições financeiras que podem ser previstas. Como exemplo: os valores da carteira de crédito, a ampliação das carteiras de setores específicos e a projeção do resultado em função das operações contratadas.

# Referências

- [1] Davenport, Thomas H.: *Big data no trabalho: derrubando mitos e descobrindo oportunidades*. Alta Books, RJ, Brasil, 2017. 1
- [2] FEBRABAN: *Pesquisa FEBRABAN de Tecnologia Bancária 2020*. <https://cmsportal.febraban.org.br/Arquivos/documentos/PDF/Pesquisa%20Febraban%20de%20Tecnologia%20Banc%C3%A1ria%202020%20VF.pdf>, acesso em 2020-12-07. 1
- [3] Group, The PostgreSQL Global Development: *PostgreSQL: About*. <https://www.postgresql.org/about/>, acesso em 2020-12-07. 3, 14
- [4] Foundation, Python Software: *General Python FAQ*. <https://docs.python.org/3.9/faq/general.html>, acesso em 2020-12-07. 3, 13
- [5] Waikato, University of: *Weka*. <https://www.cs.waikato.ac.nz/ml/weka/>, acesso em 2020-12-12. 3, 14
- [6] *Orange Data Mining - License*. <https://orange.biolab.si/license/>, acesso em 2020-12-07. 3, 15
- [7] *Banco Central do Brasil*. <https://www.bcb.gov.br/>, acesso em 2020-12-12. 4
- [8] BRASIL: *Constituição da República Federativa do Brasil*, 1988. [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm), acesso em 2020-12-04. 4
- [9] Bacen: *Composição e segmentos do Sistema Financeiro Nacional*. <https://www.bcb.gov.br/pre/composicao/composicao.asp?frame=1>, acesso em 2020-12-04. 5
- [10] *Estrutura do Sistema Financeiro Nacional - SFN*. [https://www.investidor.gov.br/menu/Menu\\_Academico/O\\_Mercado\\_de\\_valores\\_mobiliarios\\_brasileiro/Estrutura\\_Funcionamento.html](https://www.investidor.gov.br/menu/Menu_Academico/O_Mercado_de_valores_mobiliarios_brasileiro/Estrutura_Funcionamento.html), acesso em 2020-12-04. 5, 6
- [11] BRASIL: *Lei 4595 - Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, Cria o Conselho Monetário Nacional e dá outras providências*, 1964. [http://www.planalto.gov.br/ccivil\\_03/leis/14595.htm](http://www.planalto.gov.br/ccivil_03/leis/14595.htm), acesso em 2020-12-04. 5
- [12] Bacen: *Manual do Plano Contábil das Instituições do Sistema Financeiro Nacional - Cosif*. <https://www3.bcb.gov.br/aplica/cosif/completo>, acesso em 2020-12-04. 6

- [13] Bacen: *Sistema de Informações de Créditos (SCR)*. <https://www.bcb.gov.br/estabilidadefinanceira/scr>, acesso em 2020-12-04. 6
- [14] Bacen: *Portal IF.data do Banco Central do Brasil*. <https://www3.bcb.gov.br/ifdata/>, acesso em 2020-09-02. 6, 16
- [15] Bacen: *Plano de Dados Abertos do Banco Central do Brasil*. [https://www.bcb.gov.br/acessoainformacao/Plano\\_Dados\\_Abertos\\_Banco\\_Central.pdf](https://www.bcb.gov.br/acessoainformacao/Plano_Dados_Abertos_Banco_Central.pdf), acesso em 2020-12-05. 6
- [16] Amaral, Fernando: *Introdução à Ciência de Dados: Mineração de Dados e Big Data*. Alta Books, RJ, Brasil, 2016. 8
- [17] Goldschmidt, Ronaldo e Emmanuel Passos: *Data mining : um guia prático*. Elsevier, RJ, Brasil, 2005. 8, 9, 12, 13, 15
- [18] Han, Jiawei e Micheline Kamber: *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, MA, EUA, 2000. 9
- [19] Hand, David, Heikki Mannila e Padhraic Smyth: *Principles of Data Mining*. The MIT Press, MA, EUA, 2001. 9
- [20] IBM: *Guia do IBM SPSS Modeler CRISP-DM*. [ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.1/br\\_po/ModelerCRISPDM.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.1/br_po/ModelerCRISPDM.pdf), acesso em 2020-12-06. 10
- [21] Hoffmann, Rodolfo: *Análise de Regressão: Uma Introdução à Econometria*. O Autor, SP, Brasil, 2016. 12
- [22] Joseph, Rohan: *Learn with an example : Hierarchical Clustering*. [https://miro.medium.com/max/564/1\\*n1ZUkx\\_PCjoy9pSU0a\\_e7A.png](https://miro.medium.com/max/564/1*n1ZUkx_PCjoy9pSU0a_e7A.png), acesso em 2020-12-07. 14
- [23] Pandas: *About Pandas*. <https://pandas.pydata.org/about/>, acesso em 2020-12-07. 14
- [24] *Time series analysis and forecasting with weka*. <https://wiki.pentaho.com/display/datamining/time+series+analysis+and+forecasting+with+weka>, acesso em 2020-09-02. 15, 22
- [25] Bond, Letycia: *Estudo da FGV aponta que pandemia provocou queda de renda de 20,1%*. <https://agenciabrasil.ebc.com.br/geral/noticia/2020-09/estudo-da-fgv-aponta-que-pandemia-provocou-queda-de-renda-de-201>, acesso em 2020-12-09. 58
- [26] *Febraban - conheça as iniciativas do setor bancário para amenizar os efeitos do coronavírus*. <https://portal.febraban.org.br/noticia/3428/pt-br/>, acesso em 2020-09-05. 58
- [27] *Banco Pan - Press Release 2T20*. <https://ri.bancopan.com.br/informacoes-financeiras/central-de-resultados/>, acesso em 2020-09-15. 58



# Anexo I

## Script de importação e tratamento dos arquivos CSV

## Importação das bibliotecas

```
In [7]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from functools import reduce
from sqlalchemy import create_engine
import psycopg2
import io
```

## Função para definir endereço das tabelas

```
In [8]: def path(trimestre, tabela):
endereco = "C:/Users/grama/Desktop/TCC - Computação/Notebook/Dados/IF.data/"
arquivo = "/dados ("
final = ").csv"
resultado = endereco + trimestre + arquivo + tabela + final
return resultado
```

## Importação da tabela Resumo

```
In [9]: tabela = str(0)
dados = []

for i in range (25):
    if i < 3:
        dados.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col
=False, thousands = '.', names = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF',
'DATA', 'ATIV_TOT', 'CC_CLASS', 'PASS_CIRC', 'CAPT', 'PATR_LIQ', 'LUCR_LIQ', 'PATR_RE
F', 'IND_BAS', 'IND_IMOB', 'NUM_AGE', 'NUM_PA']))
    else:
        dados.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col
=False, thousands = '.', names = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF',
'DATA', 'ATIV_TOT', 'CC_CLASS', 'PASS_CIRC', 'CAPT', 'PATR_LIQ', 'LUCR_LIQ', 'NUM_AG
E', 'NUM_PA']))

resumo = pd.concat(dados, axis=0, ignore_index=True)
```

## Importação da tabela Ativo

```
In [10]: tabela = str(1)
dados1 = []

for i in range (25):
    dados1.append(pd.read_csv(path(str(i), tabela), sep=';', header = 1, index_col=Fa
lse, thousands='.', names = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DAT
A', 'DISP', 'APL_INTER', 'TVM_INST', 'OPER_CRE', 'OPER_CRE_PROV', 'OPER_CRE_LIQ', 'AR
RE_MERC_REC', 'ARRE_MERC_IMOB', 'ARRE_MERC_CRED', 'ARRE_MERC_PROV', 'ARRE_MERC_LIQ',
'OUT_CRE_LIQ', 'OUT_ATIV_REAL', 'PERM_AJUS', 'ATIV_TOT_AJUST', 'CRED_ATENC', 'ATIV_TO
T']))

ativo = pd.concat(dados1, axis=0, ignore_index=True)
```

## Importação da tabela Passivo

```
In [14]: tabela = str(2)
dados2 = []

for i in range (25):
    if i < 24:
        dados2.append(pd.read_csv(path(str(i), tabela), sep=';', header = 2, index_col=False, thousands='.', names = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DATA', 'DEP_A_VIS', 'DEP_POUP', 'DEP_INT', 'DEP_PRAZ', 'DEP_OUT', 'DEP_TOT', 'OBR_OPER_COMP', 'LET_CRED_IMOB', 'LET_CRED_AGRO', 'LET_FIN', 'OBR_TIT_EXT', 'OUT_REC_ACEI', 'REC_ACEI_EM_TIT', 'OBRIG_EMP_REP', 'CAPT', 'INST_DER', 'OUT_OBRIG', 'PASS_CIRC_EXIG', 'RES_EXEC_FUT', 'PATR_LIQ', 'PASS_TOT']))
    else:
        dados2.append(pd.read_csv(path(str(i), tabela), sep=';', header = 2, index_col=False, thousands='.', names = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DATA', 'DEP_A_VIS', 'DEP_POUP', 'DEP_INT', 'DEP_PRAZ', 'CONT_PAG_PRE_PAG', 'DEP_OUT', 'DEP_TOT', 'OBR_OPER_COMP', 'LET_CRED_IMOB', 'LET_CRED_AGRO', 'LET_FIN', 'OBR_TIT_EXT', 'OUT_REC_ACEI', 'REC_ACEI_EM_TIT', 'OBRIG_EMP_REP', 'CAPT', 'INST_DER', 'OUT_OBRIG', 'PASS_CIRC_EXIG', 'RES_EXEC_FUT', 'PATR_LIQ', 'PASS_TOT']))

passivo = pd.concat(dados2, axis=0, ignore_index=True)
```

## Importação da tabela Demonstração de Resultado

```
In [15]: tabela = str(3)
dados3 = []
nomes3 = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DATA', 'REND_OPER_CRED', 'REND_OPER_ARRE', 'REND_OPER_TVM', 'REND_OPER_INST_FIN', 'REN_OPER_CAMB', 'REND_APLIC_COMP', 'REC_INTER_FIN', 'DESP_CAPT', 'DESP_OBRIG_EMP', 'DESP_OPER_ARRE', 'DESP_OPER_CAMB', 'RES_PROV_CRED', 'DESP_INTER_FIN', 'RES_INTER_FIN', 'REN_PREST_SERV', 'REN_TAR_BANC', 'DESP_PESS', 'DESP_ADM', 'DESP_TRIB', 'RES_PART', 'OUT_REC_OPER', 'OUT_DESP_OPER', 'OUT_REC_DESP_OPER', 'RES_OPER', 'RES_N_OPER', 'RES_ANT_TRIB', 'IMP_REN_CON_SOC', 'PART_LUC', 'LUC_LIQ', 'JUR_SOB_CAP']

for i in range (25):
    dados3.append(pd.read_csv(path(str(i), tabela), sep=';', header = 2, index_col=False, thousands='.', names = nomes3))

dem_res = pd.concat(dados3, axis=0, ignore_index=True)
```

## Importação da tabela Carteira de crédito ativa Pessoa Física - modalidade e prazo de vencimento

```
In [16]: tabela = str(4)
dados4 = []
nomes4 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_CAR
_PF', 'CONSIG_VENC_15', 'CONSIG_VINC_90', 'CONSIG_VINC_360', 'CONSIG_VINC_1080', 'CON
SIG_VINC_1800', 'CONSIG_VINC_5400', 'CONSIG_VINC_ACI', 'CONSIG_TOT', 'N_CONSIG_VENC1
5', 'N_CONSIG_VINC90', 'N_CONSIG_VINC_360', 'N_CONSIG_VINC_1080', 'N_CONSIG_VINC_180
0', 'N_CONSIG_VINC_5400', 'N_CONSIG_VINC_ACI', 'N_CONSIG_TOT', 'VEIC_VENC15', 'VEIC_V
INC90', 'VEIC_VINC_360', 'VEIC_VINC_1080', 'VEIC_VINC_1800', 'VEIC_VINC_5400', 'VEIC_
VINC_ACI', 'VEIC_TOT', 'HAB_VENC15', 'HAB_VINC90', 'HAB_VINC_360', 'HAB_VINC_1080',
'HAB_VINC_1800', 'HAB_VINC_5400', 'HAB_VINC_ACI', 'HAB_TOT', 'CART_CRED_VENC15', 'CAR
T_CRED_VINC90', 'CART_CRED_VINC_360', 'CART_CRED_VINC_1080', 'CART_CRED_VINC_1800',
'CART_CRED_VINC_5400', 'CART_CRED_VINC_ACI', 'CART_CRED_TOT', 'RUR_VENC15', 'RUR_VINC
90', 'RUR_VINC_360', 'RUR_VINC_1080', 'RUR_VINC_1800', 'RUR_VINC_5400', 'RUR_VINC_AC
I', 'RUR_TOT', 'OUT_PF_VENC15', 'OUT_PF_VINC90', 'OUT_PF_VINC_360', 'OUT_PF_1080', 'O
UT_PF_1800', 'OUT_PF_5400', 'OUT_PF_ACI', 'OUT_PF_TOT', 'TOT_EXT_PF']

for i in range (25):
    dados4.append(pd.read_csv(path(str(i), tabela), sep=';', header = 1, index_col=Fa
lse, thousands='.', names = nomes4))

cred_pf = pd.concat(dados4, axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa Pessoa Jurídica - modalidade e prazo de vencimento

```
In [17]: tabela = str(5)
dados5 = []
nomes5 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_CAR
_PJ', 'CAP_GIR_VENC_15', 'CAP_GIR_VINC_90', 'CAP_GIR_VINC_360', 'CAP_GIR_VINC_1080',
'CAP_GIR_VINC_1800', 'CAP_GIR_VINC_5400', 'CAP_GIR_VINC_ACI', 'CAP_GIR_TOT', 'INVEST_
VENC15', 'INVEST_VINC90', 'INVEST_VINC_360', 'INVEST_VINC_1080', 'INVEST_VINC_1800',
'INVEST_VINC_5400', 'INVEST_VINC_ACI', 'INVEST_TOT', 'GIR_ROT_VENC15', 'GIR_ROT_VINC9
0', 'GIR_ROT_VINC_360', 'GIR_ROT_VINC_1080', 'GIR_ROT_VINC_1800', 'GIR_ROT_VINC_5400',
'GIR_ROT_VINC_ACI', 'GIR_ROT_TOT', 'REC_VENC15', 'REC_VINC90', 'REC_VINC_360', 'REC
_VINC_1080', 'REC_VINC_1800', 'REC_VINC_5400', 'REC_VINC_ACI', 'REC_TOT', 'COM_EX_VEN
C15', 'COM_EX_VINC90', 'COM_EX_VINC_360', 'COM_EX_VINC_1080', 'COM_EX_VINC_1800', 'CO
M_EX_VINC_5400', 'COM_EX_VINC_ACI', 'COM_EX_TOT', 'OUT_PJ_VENC15', 'OUT_PJ_VINC90',
'OUT_PJ_VINC_360', 'OUT_PJ_VINC_1080', 'OUT_PJ_VINC_1800', 'OUT_PJ_VINC_5400', 'OUT_P
J_VINC_ACI', 'OUT_PJ_TOT', 'FIN_INF_DES_VENC15', 'FIN_INF_DES_VINC90', 'FIN_INF_DES_V
INC_360', 'FIN_INF_DES_1080', 'FIN_INF_DES_1800', 'FIN_INF_DES_5400', 'FIN_INF_DES_AC
I', 'FIN_INF_DES_TOT', 'RUR_PJ_VENC_15', 'RUR_PJ_VINC_90', 'RUR_PJ_VINC_360', 'RUR_PJ
_VINC_1080', 'RUR_PJ_VINC_1800', 'RUR_PJ_VINC_5400', 'RUR_PJ_VINC_ACI', 'RUR_PJ_TOT',
'HAB_PJ_VENC_15', 'HAB_PJ_VINC_90', 'HAB_PJ_VINC_360', 'HAB_PJ_VINC_1080', 'HAB_PJ_VI
NC_1800', 'HAB_PJ_VINC_5400', 'HAB_PJ_VINC_ACI', 'HAB_PJ_TOT', 'TOT_EXT_PJ']

for i in range (25):
    dados5.append(pd.read_csv(path(str(i), tabela), sep=';', header = 1, index_col=Fa
lse, thousands='.', names = nomes5))

cred_pj = pd.concat(dados5, axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa Pessoa Jurídica - por atividade econômica (CNAE)

```
In [18]: tabela = str(6)
dados6 = []
nomes6 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_CAR_PJ', 'AGR_PEC_VENC_15', 'AGR_PEC_VINC_90', 'AGR_PEC_VINC_360', 'AGR_PEC_VINC_1080', 'AGR_PEC_VINC_1800', 'AGR_PEC_VINC_5400', 'AGR_PEC_VINC_ACI', 'AGR_PEC_TOT', 'IND_TRANS_VENC15', 'IND_TRANS_VINC90', 'IND_TRANS_VINC_360', 'IND_TRANS_VINC_1080', 'IND_TRANS_VINC_1800', 'IND_TRANS_VINC_5400', 'IND_TRANS_VINC_ACI', 'IND_TRANS_TOT', 'CONST_VENC15', 'CONST_VINC90', 'CONST_VINC_360', 'CONST_VINC_1080', 'CONST_VINC_1800', 'CONST_VINC_5400', 'CONST_VINC_ACI', 'CONST_TOT', 'SER_IND_VENC15', 'SER_IND_VINC90', 'SER_IND_VINC_360', 'SER_IND_VINC_1080', 'SER_IND_VINC_1800', 'SER_IND_VINC_5400', 'SER_IND_VINC_ACI', 'SER_IND_TOT', 'IND_EXTRAT_VENC15', 'IND_EXTRAT_VINC90', 'IND_EXTRAT_VINC_360', 'IND_EXTRAT_VINC_1080', 'IND_EXTRAT_VINC_1800', 'IND_EXTRAT_VINC_5400', 'IND_EXTRAT_VINC_ACI', 'IND_EXTRAT_TOT', 'COM_REP_VEIC_VENC15', 'COM_REP_VEIC_VINC90', 'COM_REP_VEIC_VINC_360', 'COM_REP_VEIC_VINC_1080', 'COM_REP_VEIC_VINC_1800', 'COM_REP_VEIC_VINC_5400', 'COM_REP_VEIC_VINC_ACI', 'COM_REP_VEIC_TOT', 'ADM_PUB_VENC15', 'ADM_PUB_VINC90', 'ADM_PUB_VINC_360', 'ADM_PUB_1080', 'ADM_PUB_1800', 'ADM_PUB_5400', 'ADM_PUB_ACI', 'ADM_PUB_TOT', 'TRANS_ARMAZ_VENC_15', 'TRANS_ARMAZ_VINC_90', 'TRANS_ARMAZ_VINC_360', 'TRANS_ARMAZ_VINC_1080', 'TRANS_ARMAZ_VINC_1800', 'TRANS_ARMAZ_VINC_5400', 'TRANS_ARMAZ_VINC_ACI', 'TRANS_ARMAZ_TOT', 'OUT_CNAE_VENC_15', 'OUT_CNAE_VINC_90', 'OUT_CNAE_VINC_360', 'OUT_CNAE_VINC_1080', 'OUT_CNAE_VINC_1800', 'OUT_CNAE_VINC_5400', 'OUT_CNAE_VINC_ACI', 'OUT_CNAE_TOT', 'ATIV_N_INFOR', 'TOT_N_INDIV_PJ', 'TOT_EXT_PJ']

for i in range (25):
    dados6.append(pd.read_csv(path(str(i), tabela), sep=';', header = 1, index_col=False, thousands='.', names = nomes6))

pj_cnae = pd.concat(dados6,axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa Pessoa Jurídica - por porte do tomador

```
In [19]: tabela = str(7)
dados7 = []
nomes7 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_CAR_PJ', 'MICRO', 'PEQ', 'MEDIA', 'GRANDE', 'INDISP', 'TOT_N_INDIV_PJ', 'TOT_EXT_PJ']

for i in range (25):
    dados7.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col=False, thousands='.', names = nomes7))

pj_porte = pd.concat(dados7,axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa - quantidade de clientes e de operações

```
In [20]: tabela = str(8)
dados8 = []
nomes8 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'QTD_CLIENTE', 'QTD_OPER']

for i in range (25):
    dados8.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col=False, thousands='.', names = nomes8))

cred_qtd = pd.concat(dados8,axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa - por nível de risco da operação

```
In [21]: tabela = str(9)
dados9 = []
nomes9 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_GER', 'RISCO_AA', 'RISCO_A', 'RISCO_B', 'RISCO_C', 'RISCO_D', 'RISCO_E', 'RISCO_F', 'RISCO_G', 'RISCO_H', 'TOT_EXT']

for i in range (25):
    dados9.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col=False, thousands='.', names = nomes9))

cred_risco = pd.concat(dados9,axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa - por indexador

```
In [22]: tabela = str(10)
dados10 = []
nomes10 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_GER', 'PREFIX', 'TR_TBF', 'TJLP', 'TLP', 'LIBOR', 'OUTR_TAX_POS_FIX', 'CART_ATIV_CDI', 'SELIC', 'OUT_TAX_FLUT', 'IGPM', 'IPCA', 'IPCC', 'OUT_IND_PREC', 'OUT_INDEX', 'TOT_N_INDIV', 'TOT_EXT']

for i in range (25):
    dados10.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col=False, thousands='.', names = nomes10))

cred_index = pd.concat(dados10,axis=0, ignore_index=True)
```

### Importação da tabela Carteira de crédito ativa - por região geográfica

```
In [23]: tabela = str(11)
dados11 = []
nomes11 = ['IF', 'COD', 'TCB', 'TD', 'TC', 'SR', 'SEGM', 'CID', 'UF', 'DATA', 'TOT_GER', 'SUDESTE', 'CENTRO_OESTE', 'NORDESTE', 'NORTE', 'SUL', 'REG_N_INFOR', 'TOT_EXT']

for i in range (1,25):
    dados11.append(pd.read_csv(path(str(i), tabela), sep=';', header = 0, index_col=False, thousands='.', names = nomes11))

cred_reg_geog = pd.concat(dados11,axis=0, ignore_index=True)
```

### Remoção de linhas com COD NaN e substituição de NI por NaN

```
In [24]: tabelas = [resumo, ativo, passivo, dem_res, cred_pf, cred_pj, pj_cnae, pj_porte, cred_qtd, cred_risco, cred_index, cred_reg_geog]

for i in range(11):
    tabelas[i] = tabelas[i].dropna(subset=['COD', 'DATA'])
    tabelas[i] = tabelas[i].replace({'NI': np.nan})
```

### Junção das tabelas pela chave 'IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DATA'

```
In [25]: chave = ['IF', 'COD', 'TCB', 'SR', 'TD', 'TC', 'CID', 'UF', 'DATA']

juncao = reduce(lambda left,right: pd.merge(left,right,on=chave, suffixes=('', '_y'
)), tabelas)
juncao.drop(juncao.filter(regex='_y$').columns.tolist(),axis=1, inplace=True)
```

### Criação da coluna INAD\_TOT (soma das inadimplências PF e PJ)

```
In [26]: colunas_inad = ['CONSIG_VENC_15', 'N_CONSIG_VENC15', 'VEIC_VENC15', 'HAB_VENC15', 'CA
RT_CRED_VENC15', 'RUR_VENC15', 'OUT_PF_VENC15', 'CAP_GIR_VENC_15', 'INVEST_VENC15',
'GIR_ROT_VENC15', 'REC_VENC15', 'COM_EX_VENC15', 'OUT_PJ_VENC15', 'FIN_INF_DES_VENC1
5', 'RUR_PJ_VENC_15', 'HAB_PJ_VENC_15']
juncao['INAD_TOT'] = juncao[colunas_inad].sum(axis=1)
```

### Conversão do campo DATA para o formato datetime

```
In [30]: juncao.DATA = pd.to_datetime(juncao.DATA)
```

### Exportação da tabela final em formato csv

```
In [27]: juncao.to_csv('juncao_06.csv')
```

### Carrega tabela em banco de dados PostgreSQL

```
In [31]: engine = create_engine('postgresql+psycopg2://postgres:admin@localhost:5432/base_ifda
ta')

juncao.head(0).to_sql('ifdata_juncao_06', engine, if_exists='replace',index=True)

conn = engine.raw_connection()
cur = conn.cursor()
output = io.StringIO()
juncao.to_csv(output, sep='\t', header=False, index=True)
output.seek(0)
contents = output.getvalue()
cur.copy_from(output, 'ifdata_juncao_06', null='')
conn.commit()
```

## Anexo II

### Relação de abreviaturas utilizadas para as colunas da tabela



<b>Nome da Coluna</b>	<b>Descrição</b>
IF	Nome da Instituição Financeira
COD	Código da Instituição Financeira
TCB	Tipo de Consolidado Bancário
SR	SR - Segmento Resolução nº 4.553/2017
TD	Tipo de Consolidação
TC	Tipo de Controle
CID	Cidade
UF	Unidade Federativa
DATA	Trimestre Referência
ATIV_TOT	Ativo Total
CC_CLASS	Carteira de Crédito Classificada
PASS_CIRC	Passivo Circulante e Exigível a Longo Prazo e Resultados de Exercícios Futuros
CAPT	Captações
PATR_LIQ	Patrimônio Líquido
LUCR_LIQ	Lucro Líquido
PATR_REF	Patrimônio de Referência para Comparação com o RWA
IND_BAS	Índice de Basileia
IND_IMOB	Índice de Imobilização
NUM_AGE	Número de Agências
NUM_PA	Número de Postos de Atendimento
DISP	Disponibilidades
APL_INTER	Aplicações Interfinanceiras de Liquidez
TVM_INST	TVM e Instrumentos Financeiros Derivativos
OPER_CRE	Operações de Crédito
OPER_CRE_PROV	Provisão sobre Operações de Crédito
OPER_CRE_LIQ	Operações de Crédito Líquidas de Provisão
ARRE_MERC_REC	Arrendamento Mercantil a Receber
ARRE_MERC_IMOB	Imobilizado de Arrendamento
ARRE_MERC_CRED	Arrendamento Mercantil - Credores por Antecipação de Valor Residual
ARRE_MERC_PROV	Provisão sobre Arrendamento Mercantil
ARRE_MERC_LIQ	Arrendamento Mercantil Líquido de Provisão
OUT_CRE_LIQ	Outros Créditos Líquido de Provisão
OUT_ATIV_REAL	Outros Ativos Realizáveis
PERM_AJUS	Permanente Ajustado
ATIV_TOT_AJUST	Ativo Total Ajustado
CRED_ATENC	Credores por Antecipação de Valor Residual
DEP_A_VIS	Depósitos à Vista
DEP_POUP	Depósitos de Poupança
DEP_INT	Depósitos Interfinanceiros
DEP_PRAZ	Depósitos a Prazo
DEP_OUT	Outros Depósitos
DEP_TOT	Depósito Total
OBR_OPER_COMP	Obrigações por Operações Compromissadas
LET_CRED_IMOB	Letras de Crédito Imobiliário
LET_CRED_AGRO	Letras de Crédito do Agronegócio
LET_FIN	Letras Financeiras
OBR_TIT_EXT	Obrigações por Títulos e Valores Mobiliários no Exterior
OUT_REC_ACEI	Outros Recursos de Aceites e Emissão de Títulos
REC_ACEI_EM_TIT	Recursos de Aceites e Emissão de Títulos
OBRIG_EMP_REP	Obrigações por Empréstimos e Repasses
INST_DER	Instrumentos Derivativos
OUT_OBRIG	Outras Obrigações
PASS_CIRC_EXIG	Passivo Circulante e Exigível a Longo Prazo
RES_EXEC_FUT	Resultados de Exercícios Futuros
PASS_TOT	Passivo Total
REND_OPER_CRED	Rendas de Operações de Crédito
REND_OPER_ARRE	Rendas de Operações de Arrendamento Mercantil
REND_OPER_TVM	Rendas de Operações com TVM
REND_OPER_INST_FIN	Rendas de Operações com Instrumentos Financeiros Derivativos
REN_OPER_CAMB	Rendas de Operações de Câmbio
REND_APLIC_COMP	Rendas de Aplicações Compulsórias
REC_INTER_FIN	Receitas de Intermediação Financeira
DESP_CAPT	Despesas de Captação
DESP_OBRIG_EMP	Despesas de Obrigações por Empréstimos e Repasses
DESP_OPER_ARRE	Despesas de Operações de Arrendamento Mercantil
DESP_OPER_CAMB	Despesas de Operações de Câmbio
RES_PROV_CRED	Resultado de Provisão para Créditos de Difícil Liquidação
DESP_INTER_FIN	Despesas de Intermediação Financeira

RES_INTER_FIN	Resultado de Intermediação Financeira
REN_PREST_SERV	Rendas de Prestação de Serviços
REN_TAR_BANC	Rendas de Tarifas Bancárias
DESP_PESS	Despesas de Pessoal
DESP_ADM	Despesas Administrativas
DESP_TRIB	Despesas Tributárias
RES_PART	Resultado de Participações
OUT_REC_OPER	Outras Receitas Operacionais
OUT_DESP_OPER	Outras Despesas Operacionais
OUT_REC_DESP_OPER	Outras Receitas/Despesas Operacionais
RES_OPER	Resultado Operacional
RES_N_OPER	Resultado não Operacional
RES_ANT_TRIB	Resultado antes da Tributação, Lucro e Participação
IMP_REN_CON_SOC	Imposto de Renda e Contribuição Social
PART_LUC	Participação nos Lucros
LUC_LIQ	Lucro Líquido
JUR_SOB_CAP	Juros Sobre Capital Social de Cooperativas
SEGM	Segmento
TOT_CAR_PF	Total da Carteira de Pessoa Física
CONSIG_VENC_15	Empréstimo com Consignação em Folha - Vencido a Partir de 15 dias
CONSIG_VINC_90	Empréstimo com Consignação em Folha - A Vencer em até 90 dias
CONSIG_VINC_360	Empréstimo com Consignação em Folha - A Vencer entre 91 e 360 dias
CONSIG_VINC_1080	Empréstimo com Consignação em Folha - A Vencer entre 361 e 1080 dias
CONSIG_VINC_1800	Empréstimo com Consignação em Folha - A Vencer entre 1081 e 1800 dias
CONSIG_VINC_5400	Empréstimo com Consignação em Folha - A Vencer entre 1801 e 5400 dias
CONSIG_VINC_ACI	Empréstimo com Consignação em Folha - A Vencer acima de 5400 dias
CONSIG_TOT	Empréstimo com Consignação em Folha - Total
N_CONSIG_VENC15	Empréstimo sem Consignação em Folha - Vencido a Partir de 15 dias
N_CONSIG_VINC90	Empréstimo sem Consignação em Folha - A Vencer em até 90 dias
N_CONSIG_VINC_360	Empréstimo sem Consignação em Folha - A Vencer entre 91 e 360 dias
N_CONSIG_VINC_1080	Empréstimo sem Consignação em Folha - A Vencer entre 361 e 1080 dias
N_CONSIG_VINC_1800	Empréstimo sem Consignação em Folha - A Vencer entre 1081 e 1800 dias
N_CONSIG_VINC_5400	Empréstimo sem Consignação em Folha - A Vencer entre 1801 e 5400 dias
N_CONSIG_VINC_ACI	Empréstimo sem Consignação em Folha - A Vencer acima de 5400 dias
N_CONSIG_TOT	Empréstimo sem Consignação em Folha - Total
VEIC_VENC15	Veículos - Vencido a Partir de 15 dias
VEIC_VINC90	Veículos - A Vencer em até 90 dias
VEIC_VINC_360	Veículos - A Vencer entre 91 e 360 dias
VEIC_VINC_1080	Veículos - A Vencer entre 361 e 1080 dias
VEIC_VINC_1800	Veículos - A Vencer entre 1081 e 1800 dias
VEIC_VINC_5400	Veículos - A Vencer entre 1801 e 5400 dias
VEIC_VINC_ACI	Veículos - A Vencer acima de 5400 dias
VEIC_TOT	Veículos - Total
HAB_VENC15	Habitação - Vencido a Partir de 15 dias
HAB_VINC90	Habitação - A Vencer em até 90 dias
HAB_VINC_360	Habitação - A Vencer entre 91 e 360 dias
HAB_VINC_1080	Habitação - A Vencer entre 361 e 1080 dias
HAB_VINC_1800	Habitação - A Vencer entre 1081 e 1800 dias
HAB_VINC_5400	Habitação - A Vencer entre 1801 e 5400 dias
HAB_VINC_ACI	Habitação - A Vencer acima de 5400 dias
HAB_TOT	Habitação - Total
CART_CRED_VENC15	Cartão de Crédito - Vencido a Partir de 15 dias
CART_CRED_VINC90	Cartão de Crédito - A Vencer em até 90 dias
CART_CRED_VINC_360	Cartão de Crédito - A Vencer entre 91 e 360 dias
CART_CRED_VINC_1080	Cartão de Crédito - A Vencer entre 361 e 1080 dias
CART_CRED_VINC_1800	Cartão de Crédito - A Vencer entre 1081 e 1800 dias
CART_CRED_VINC_5400	Cartão de Crédito - A Vencer entre 1801 e 5400 dias
CART_CRED_VINC_ACI	Cartão de Crédito - A Vencer acima de 5400 dias
CART_CRED_TOT	Cartão de Crédito - Total
RUR_VENC15	Rural e Agroindustrial - Vencido a Partir de 15 dias
RUR_VINC90	Rural e Agroindustrial - A Vencer em até 90 dias
RUR_VINC_360	Rural e Agroindustrial - A Vencer entre 91 e 360 dias
RUR_VINC_1080	Rural e Agroindustrial - A Vencer entre 361 e 1080 dias
RUR_VINC_1800	Rural e Agroindustrial - A Vencer entre 1081 e 1800 dias
RUR_VINC_5400	Rural e Agroindustrial - A Vencer entre 1801 e 5400 dias
RUR_VINC_ACI	Rural e Agroindustrial - A Vencer acima de 5400 dias
RUR_TOT	Rural e Agroindustrial - Total
OUT_PF_VENC15	Outros Créditos Pessoa Física - Vencido a Partir de 15 dias
OUT_PF_VINC90	Outros Créditos Pessoa Física - A Vencer em até 90 dias

OUT_PF_VINC_360	Outros Créditos Pessoa Física - A Vencer entre 91 e 360 dias
OUT_PF_1080	Outros Créditos Pessoa Física - A Vencer entre 361 e 1080 dias
OUT_PF_1800	Outros Créditos Pessoa Física - A Vencer entre 1081 e 1800 dias
OUT_PF_5400	Outros Créditos Pessoa Física - A Vencer entre 1801 e 5400 dias
OUT_PF_ACI	Outros Créditos Pessoa Física - A Vencer acima de 5400 dias
OUT_PF_TOT	Outros Créditos Pessoa Física - Total
TOT_EXT_PF	Total Exterior Pessoa Física
TOT_CAR_PJ	Total da Carteira de Pessoa Jurídica
CAP_GIR_VENC_15	Capital de Giro - Vencido a Partir de 15 dias
CAP_GIR_VINC_90	Capital de Giro - A Vencer em até 90 dias
CAP_GIR_VINC_360	Capital de Giro - A Vencer entre 91 e 360 dias
CAP_GIR_VINC_1080	Capital de Giro - A Vencer entre 361 e 1080 dias
CAP_GIR_VINC_1800	Capital de Giro - A Vencer entre 1081 e 1800 dias
CAP_GIR_VINC_5400	Capital de Giro - A Vencer entre 1801 e 5400 dias
CAP_GIR_VINC_ACI	Capital de Giro - A Vencer acima de 5400 dias
CAP_GIR_TOT	Capital de Giro - Total
INVEST_VENC15	Investimento - Vencido a Partir de 15 dias
INVEST_VINC90	Investimento - A Vencer em até 90 dias
INVEST_VINC_360	Investimento - A Vencer entre 91 e 360 dias
INVEST_VINC_1080	Investimento - A Vencer entre 361 e 1080 dias
INVEST_VINC_1800	Investimento - A Vencer entre 1081 e 1800 dias
INVEST_VINC_5400	Investimento - A Vencer entre 1801 e 5400 dias
INVEST_VINC_ACI	Investimento - A Vencer acima de 5400 dias
INVEST_TOT	Investimento - Total
GIR_ROT_VENC15	Capital de Giro Rotativo - Vencido a Partir de 15 dias
GIR_ROT_VINC90	Capital de Giro Rotativo - A Vencer em até 90 dias
GIR_ROT_VINC_360	Capital de Giro Rotativo - A Vencer entre 91 e 360 dias
GIR_ROT_VINC_1080	Capital de Giro Rotativo - A Vencer entre 361 e 1080 dias
GIR_ROT_VINC_1800	Capital de Giro Rotativo - A Vencer entre 1081 e 1800 dias
GIR_ROT_VINC_5400	Capital de Giro Rotativo - A Vencer entre 1801 e 5400 dias
GIR_ROT_VINC_ACI	Capital de Giro Rotativo - A Vencer acima de 5400 dias
GIR_ROT_TOT	Capital de Giro Rotativo - Total
REC_VENC15	Operações com Recebíveis - Vencido a Partir de 15 dias
REC_VINC90	Operações com Recebíveis - A Vencer em até 90 dias
REC_VINC_360	Operações com Recebíveis - A Vencer entre 91 e 360 dias
REC_VINC_1080	Operações com Recebíveis - A Vencer entre 361 e 1080 dias
REC_VINC_1800	Operações com Recebíveis - A Vencer entre 1081 e 1800 dias
REC_VINC_5400	Operações com Recebíveis - A Vencer entre 1801 e 5400 dias
REC_VINC_ACI	Operações com Recebíveis - A Vencer acima de 5400 dias
REC_TOT	Operações com Recebíveis - Total
COM_EX_VENC15	Comércio Exterior - Vencido a Partir de 15 dias
COM_EX_VINC90	Comércio Exterior - A Vencer em até 90 dias
COM_EX_VINC_360	Comércio Exterior - A Vencer entre 91 e 360 dias
COM_EX_VINC_1080	Comércio Exterior - A Vencer entre 361 e 1080 dias
COM_EX_VINC_1800	Comércio Exterior - A Vencer entre 1081 e 1800 dias
COM_EX_VINC_5400	Comércio Exterior - A Vencer entre 1801 e 5400 dias
COM_EX_VINC_ACI	Comércio Exterior - A Vencer acima de 5400 dias
COM_EX_TOT	Comércio Exterior - Total
OUT_PJ_VENC15	Outros Créditos Pessoa Jurídica - Vencido a Partir de 15 dias
OUT_PJ_VINC90	Outros Créditos Pessoa Jurídica - A Vencer em até 90 dias
OUT_PJ_VINC_360	Outros Créditos Pessoa Jurídica - A Vencer entre 91 e 360 dias
OUT_PJ_VINC_1080	Outros Créditos Pessoa Jurídica - A Vencer entre 361 e 1080 dias
OUT_PJ_VINC_1800	Outros Créditos Pessoa Jurídica - A Vencer entre 1081 e 1800 dias
OUT_PJ_VINC_5400	Outros Créditos Pessoa Jurídica - A Vencer entre 1801 e 5400 dias
OUT_PJ_VINC_ACI	Outros Créditos Pessoa Jurídica - A Vencer acima de 5400 dias
OUT_PJ_TOT	Outros Créditos Pessoa Jurídica - Total
FIN_INF_DES_VENC15	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - Vencido a Partir de 15 dias
FIN_INF_DES_VINC90	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer em até 90 dias
FIN_INF_DES_VINC_360	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer entre 91 e 360 dias
FIN_INF_DES_1080	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer entre 361 e 1080 dias
FIN_INF_DES_1800	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer entre 1081 e 1800 dias
FIN_INF_DES_5400	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer entre 1801 e 5400 dias
FIN_INF_DES_ACI	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - A Vencer acima de 5400 dias
FIN_INF_DES_TOT	Financiamento de Infraestrutura/Desenvolvimento/Projeto e Outros Créditos - Total
RUR_PJ_VENC_15	Rural e Agroindustrial Pessoa Jurídica - Vencido a Partir de 15 dias
RUR_PJ_VINC_90	Rural e Agroindustrial Pessoa Jurídica - A Vencer em até 90 dias
RUR_PJ_VINC_360	Rural e Agroindustrial Pessoa Jurídica - A Vencer entre 91 e 360 dias
RUR_PJ_VINC_1080	Rural e Agroindustrial Pessoa Jurídica - A Vencer entre 361 e 1080 dias
RUR_PJ_VINC_1800	Rural e Agroindustrial Pessoa Jurídica - A Vencer entre 1081 e 1800 dias

RUR_PJ_VINC_5400	Rural e Agroindustrial Pessoa Jurídica - A Vencer entre 1801 e 5400 dias
RUR_PJ_VINC_ACI	Rural e Agroindustrial Pessoa Jurídica - A Vencer acima de 5400 dias
RUR_PJ_TOT	Rural e Agroindustrial Pessoa Jurídica - Total
HAB_PJ_VENC_15	Habitacional Pessoa Jurídica - Vencido a Partir de 15 dias
HAB_PJ_VINC_90	Habitacional Pessoa Jurídica - A Vencer em até 90 dias
HAB_PJ_VINC_360	Habitacional Pessoa Jurídica - A Vencer entre 91 e 360 dias
HAB_PJ_VINC_1080	Habitacional Pessoa Jurídica - A Vencer entre 361 e 1080 dias
HAB_PJ_VINC_1800	Habitacional Pessoa Jurídica - A Vencer entre 1081 e 1800 dias
HAB_PJ_VINC_5400	Habitacional Pessoa Jurídica - A Vencer entre 1801 e 5400 dias
HAB_PJ_VINC_ACI	Habitacional Pessoa Jurídica - A Vencer acima de 5400 dias
HAB_PJ_TOT	Habitacional Pessoa Jurídica - Total
TOT_EXT_PJ	Total Exterior Pessoa Jurídica
AGR_PEC_VENC_15	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - Vencido a Partir de 15 dias
AGR_PEC_VINC_90	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer em até 90 dias
AGR_PEC_VINC_360	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer entre 91 e 360 dias
AGR_PEC_VINC_1080	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer entre 361 e 1080 dias
AGR_PEC_VINC_1800	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer entre 1081 e 1800 dias
AGR_PEC_VINC_5400	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer entre 1801 e 5400 dias
AGR_PEC_VINC_ACI	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - A Vencer acima de 5400 dias
AGR_PEC_TOT	Agricultura, Pecuária, Produção Floresta, Pesca e Aquicultura - Total
IND_TRANS_VENC15	Indústrias de Transformação - Vencido a Partir de 15 dias
IND_TRANS_VINC90	Indústrias de Transformação - A Vencer em até 90 dias
IND_TRANS_VINC_360	Indústrias de Transformação - A Vencer entre 91 e 360 dias
IND_TRANS_VINC_1080	Indústrias de Transformação - A Vencer entre 361 e 1080 dias
IND_TRANS_VINC_1800	Indústrias de Transformação - A Vencer entre 1081 e 1800 dias
IND_TRANS_VINC_5400	Indústrias de Transformação - A Vencer entre 1801 e 5400 dias
IND_TRANS_VINC_ACI	Indústrias de Transformação - A Vencer acima de 5400 dias
IND_TRANS_TOT	Indústrias de Transformação - Total
CONST_VENC15	Construção - Vencido a Partir de 15 dias
CONST_VINC90	Construção - A Vencer em até 90 dias
CONST_VINC_360	Construção - A Vencer entre 91 e 360 dias
CONST_VINC_1080	Construção - A Vencer entre 361 e 1080 dias
CONST_VINC_1800	Construção - A Vencer entre 1081 e 1800 dias
CONST_VINC_5400	Construção - A Vencer entre 1801 e 5400 dias
CONST_VINC_ACI	Construção - A Vencer acima de 5400 dias
CONST_TOT	Construção - Total
SER_IND_VENC15	Serviços Industriais de Utilidade Pública - Vencido a Partir de 15 dias
SER_IND_VINC90	Serviços Industriais de Utilidade Pública - A Vencer em até 90 dias
SER_IND_VINC_360	Serviços Industriais de Utilidade Pública - A Vencer entre 91 e 360 dias
SER_IND_VINC_1080	Serviços Industriais de Utilidade Pública - A Vencer entre 361 e 1080 dias
SER_IND_VINC_1800	Serviços Industriais de Utilidade Pública - A Vencer entre 1081 e 1800 dias
SER_IND_VINC_5400	Serviços Industriais de Utilidade Pública - A Vencer entre 1801 e 5400 dias
SER_IND_VINC_ACI	Serviços Industriais de Utilidade Pública - A Vencer acima de 5400 dias
SER_IND_TOT	Serviços Industriais de Utilidade Pública - Total
IND_EXTRAT_VENC15	Indústrias Extrativas - Vencido a Partir de 15 dias
IND_EXTRAT_VINC90	Indústrias Extrativas - A Vencer em até 90 dias
IND_EXTRAT_VINC_360	Indústrias Extrativas - A Vencer entre 91 e 360 dias
IND_EXTRAT_VINC_1080	Indústrias Extrativas - A Vencer entre 361 e 1080 dias
IND_EXTRAT_VINC_1800	Indústrias Extrativas - A Vencer entre 1081 e 1800 dias
IND_EXTRAT_VINC_5400	Indústrias Extrativas - A Vencer entre 1801 e 5400 dias
IND_EXTRAT_VINC_ACI	Indústrias Extrativas - A Vencer acima de 5400 dias
IND_EXTRAT_TOT	Indústrias Extrativas - Total
COM_REP_VEIC_VENC15	Comércio, Reparação de Veículos Automotores e Motocicletas - Vencido a Partir de 15 dias
COM_REP_VEIC_VINC90	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer em até 90 dias
COM_REP_VEIC_VINC_360	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer entre 91 e 360 dias
COM_REP_VEIC_VINC_1080	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer entre 361 e 1080 dias
COM_REP_VEIC_VINC_1800	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer entre 1081 e 1800 dias
COM_REP_VEIC_VINC_5400	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer entre 1801 e 5400 dias
COM_REP_VEIC_VINC_ACI	Comércio, Reparação de Veículos Automotores e Motocicletas - A Vencer acima de 5400 dias
COM_REP_VEIC_TOT	Comércio, Reparação de Veículos Automotores e Motocicletas - Total
ADM_PUB_VENC15	Administração Pública, Defesa e Seguridade Social - Vencido a Partir de 15 dias
ADM_PUB_VINC90	Administração Pública, Defesa e Seguridade Social - A Vencer em até 90 dias
ADM_PUB_VINC_360	Administração Pública, Defesa e Seguridade Social - A Vencer entre 91 e 360 dias
ADM_PUB_1080	Administração Pública, Defesa e Seguridade Social - A Vencer entre 361 e 1080 dias
ADM_PUB_1800	Administração Pública, Defesa e Seguridade Social - A Vencer entre 1081 e 1800 dias
ADM_PUB_5400	Administração Pública, Defesa e Seguridade Social - A Vencer entre 1801 e 5400 dias
ADM_PUB_ACI	Administração Pública, Defesa e Seguridade Social - A Vencer acima de 5400 dias
ADM_PUB_TOT	Administração Pública, Defesa e Seguridade Social - Total
TRANS_ARMAZ_VENC_15	Transporte, Armazenagem e Correio - Vencido a Partir de 15 dias

TRANS_ARMAZ_VINC_90	Transporte, Armazenagem e Correio - A Vencer em até 90 dias
TRANS_ARMAZ_VINC_360	Transporte, Armazenagem e Correio - A Vencer entre 91 e 360 dias
TRANS_ARMAZ_VINC_1080	Transporte, Armazenagem e Correio - A Vencer entre 361 e 1080 dias
TRANS_ARMAZ_VINC_1800	Transporte, Armazenagem e Correio - A Vencer entre 1081 e 1800 dias
TRANS_ARMAZ_VINC_5400	Transporte, Armazenagem e Correio - A Vencer entre 1801 e 5400 dias
TRANS_ARMAZ_VINC_ACI	Transporte, Armazenagem e Correio - A Vencer acima de 5400 dias
TRANS_ARMAZ_TOT	Transporte, Armazenagem e Correio - Total
OUT_CNAE_VENC_15	Outros CNAE - Vencido a Partir de 15 dias
OUT_CNAE_VINC_90	Outros CNAE - A Vencer em até 90 dias
OUT_CNAE_VINC_360	Outros CNAE - A Vencer entre 91 e 360 dias
OUT_CNAE_VINC_1080	Outros CNAE - A Vencer entre 361 e 1080 dias
OUT_CNAE_VINC_1800	Outros CNAE - A Vencer entre 1081 e 1800 dias
OUT_CNAE_VINC_5400	Outros CNAE - A Vencer entre 1801 e 5400 dias
OUT_CNAE_VINC_ACI	Outros CNAE - A Vencer acima de 5400 dias
OUT_CNAE_TOT	Outros CNAE - Total
ATIV_N_INFOR	Atividade não Informada ou Não se Aplica
TOT_N_INDIV_PJ	Total não Individualizado Pessoa Jurídica
MICRO	Microempresa
PEQ	Pequena Empresa
MEDIA	Média Empresa
GRANDE	Grande Empresa
INDISP	Porte Indisponível
QTD_CLI	Quantidade de Clientes
QTD_OPER	Quantidade de Operações
TOT_GER	Total Geral
RISCO_AA	Risco AA
RISCO_A	Risco A
RISCO_B	Risco B
RISCO_C	Risco C
RISCO_D	Risco D
RISCO_E	Risco E
RISCO_F	Risco F
RISCO_G	Risco G
RISCO_H	Risco H
TOT_EXT	Total Exterior
PREFIX	Prefixado
TR_TBF	TR/TBF
TJLP	TJLP
TLP	TLP
LIBOR	Libor
OUTR_TAX_POS_FIX	Outras Taxas Pós-fixadas
CART_ATIV_CDI	Carteira ativa com indexador CDI
SELIC	SELIC
OUT_TAX_FLUT	Outras Taxas Flutuantes
IGPM	IGPM
IPCA	IPCA
IPCC	IPCC
OUT_IND_PREC	Outros Índices de Preço
OUT_INDEX	Outros Indexadores
TOT_N_INDIV	Total não Individualizado
SUDESTE	Sudeste
CENTRO_OESTE	Centro-oeste
NORDESTE	Nordeste
NORTE	Norte
SUL	Sul
REG_N_INFOR	Região não Informada
INAD_TOT	Inadimplência Total (gerada a partir da soma das colunas de atraso de PF e PJ)

## **Anexo III**

# **Modelos preditivos gerados com o Weka**

## Modelo BB:

INAD\_TOT =

```
163438.9384 * DATA-remapped +
  0.0641 * Lag_INAD_TOT-1 +
 -0.0266 * Lag_INAD_TOT-3 +
 -0.0012 * Lag_ATIV_TOT-1 +
  0.0027 * Lag_ATIV_TOT-4 +
  0.0037 * Lag_CC_CLASS-4 +
 -0.002 * Lag_CAPT-2 +
  0.0024 * Lag_CAPT-3 +
  0.0035 * Lag_CAPT-4 +
 -0.0115 * Lag_PATR_LIQ-1 +
 -0.0072 * Lag_PATR_LIQ-2 +
 -0.0167 * Lag_PATR_LIQ-3 +
 -0.0142 * Lag_PATR_LIQ-4 +
 -0.0357 * Lag_OPER_CRE_PROV-2 +
  0.001 * Lag_PASS_TOT-3 +
  0.0027 * Lag_PASS_TOT-4 +
 -0.0702 * Lag_DESP_PESS-1 +
  0.0307 * Lag_DESP_PESS-2 +
 -0.0282 * Lag_DESP_PESS-4 +
 -0.0679 * Lag_DESP_ADM-1 +
 -0.1321 * Lag_DESP_TRIB-1 +
 -0.2393 * Lag_DESP_TRIB-3 +
 -0.1469 * Lag_DESP_TRIB-4 +
 -0.0024 * Lag_TOT_CAR_PF-1 +
 -0.0025 * Lag_TOT_CAR_PF-4 +
  0.0018 * Lag_TOT_CAR_PJ-4 +
 -0.0028 * Lag_TOT_GER-3 +
  0.003 * Lag_TOT_GER-4 +
 -0.0051 * Lag_RISCO_AA-1 +
  0.0046 * Lag_RISCO_A-2 +
 -0.0046 * Lag_RISCO_A-3 +
  0.0063 * Lag_RISCO_A-4 +
 -0.0144 * Lag_RISCO_B-3 +
  0.0133 * Lag_RISCO_C-1 +
  0.0093 * Lag_RISCO_C-3 +
 -0.0179 * Lag_RISCO_D-1 +
  0.1165 * Lag_RISCO_D-3 +
  0.0439 * Lag_RISCO_E-1 +
  0.0285 * Lag_RISCO_E-2 +
  0.0938 * Lag_RISCO_E-3 +
  0.0393 * Lag_RISCO_F-1 +
 -0.0445 * Lag_RISCO_F-2 +
  0.3818 * Lag_RISCO_F-4 +
  0.0797 * Lag_RISCO_G-1 +
  0.0582 * Lag_RISCO_G-2 +
 -0.0589 * Lag_RISCO_G-3 +
  0.0848 * Lag_RISCO_G-4 +
 -0.0764 * Lag_RISCO_H-1 +
  0.1175 * Lag_RISCO_H-2 +
 -0.0533 * Lag_RISCO_H-4 +
  0.0054 * Lag_TOT_EXT-1 +
  0.0077 * Lag_TOT_EXT-2 +
 -0.0123 * Lag_TOT_EXT-3 +
4010.4718 * DATA-remapped^2 +
121.8956 * DATA-remapped^3 +
 -0.001 * DATA-remapped*Lag_INAD_TOT-3 +
 -0.0015 * DATA-remapped*Lag_INAD_TOT-4 +
 -0.0001 * DATA-remapped*Lag_PATR_LIQ-1 +
 -0.0001 * DATA-remapped*Lag_PATR_LIQ-2 +
 -0.0001 * DATA-remapped*Lag_PATR_LIQ-3 +
 -0.0002 * DATA-remapped*Lag_PATR_LIQ-4 +
  0.0004 * DATA-remapped*Lag_OPER_CRE_PROV-4 +
```

0.0015 \* DATA-remapped\*Lag\_DESP\_PESS-3 +  
 -0.0014 \* DATA-remapped\*Lag\_DESP\_PESS-4 +  
 -0.0016 \* DATA-remapped\*Lag\_DESP\_ADM-2 +  
 0.0014 \* DATA-remapped\*Lag\_DESP\_ADM-3 +  
 0.0049 \* DATA-remapped\*Lag\_DESP\_TRIB-1 +  
 -0.0053 \* DATA-remapped\*Lag\_DESP\_TRIB-4 +  
 -0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-4 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_AA-1 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_AA-4 +  
 -0.0002 \* DATA-remapped\*Lag\_RISCO\_A-1 +  
 0.0002 \* DATA-remapped\*Lag\_RISCO\_A-2 +  
 -0.0004 \* DATA-remapped\*Lag\_RISCO\_A-3 +  
 0.0004 \* DATA-remapped\*Lag\_RISCO\_A-4 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_B-3 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_B-4 +  
 0.0002 \* DATA-remapped\*Lag\_RISCO\_C-1 +  
 -0.0002 \* DATA-remapped\*Lag\_RISCO\_C-4 +  
 -0.0012 \* DATA-remapped\*Lag\_RISCO\_D-1 +  
 -0.0009 \* DATA-remapped\*Lag\_RISCO\_D-2 +  
 0.0023 \* DATA-remapped\*Lag\_RISCO\_D-3 +  
 0.0011 \* DATA-remapped\*Lag\_RISCO\_E-1 +  
 -0.0011 \* DATA-remapped\*Lag\_RISCO\_E-4 +  
 -0.0004 \* DATA-remapped\*Lag\_RISCO\_G-3 +  
 -0.0006 \* DATA-remapped\*Lag\_RISCO\_H-1 +  
 -0.0011 \* DATA-remapped\*Lag\_RISCO\_H-4 +  
 -0.0013 \* DATA-remapped\*Lag\_TOT\_EXT-3 +  
 -0.001 \* DATA-remapped\*Lag\_TOT\_EXT-4 +  
 -2692566.4343

## Modelo BNDES:

INAD\_TOT =

44911.5611 \* DATA-remapped +  
 -0.0719 \* Lag\_INAD\_TOT-1 +  
 -0.053 \* Lag\_INAD\_TOT-2 +  
 -0.1019 \* Lag\_INAD\_TOT-3 +  
 0.0008 \* Lag\_ATIV\_TOT-3 +  
 0.002 \* Lag\_ATIV\_TOT-4 +  
 0.0013 \* Lag\_CC\_CLASS-2 +  
 0.0012 \* Lag\_CC\_CLASS-4 +  
 0.0006 \* Lag\_CAPT-4 +  
 -0.0023 \* Lag\_PATR\_LIQ-2 +  
 0.0088 \* Lag\_OPER\_CRE\_PROV-2 +  
 0.0008 \* Lag\_PASS\_TOT-3 +  
 0.002 \* Lag\_PASS\_TOT-4 +  
 -0.2192 \* Lag\_DESP\_PESS-1 +  
 -0.277 \* Lag\_DESP\_TRIB-2 +  
 -0.6374 \* Lag\_DESP\_TRIB-4 +  
 137.7199 \* Lag\_TOT\_CAR\_PF-1 +  
 -211.8367 \* Lag\_TOT\_CAR\_PF-2 +  
 126.9095 \* Lag\_TOT\_CAR\_PF-3 +  
 0.0013 \* Lag\_TOT\_CAR\_PJ-2 +  
 0.0012 \* Lag\_TOT\_CAR\_PJ-4 +  
 0.0013 \* Lag\_TOT\_GER-2 +  
 0.0012 \* Lag\_TOT\_GER-4 +  
 0.0025 \* Lag\_RISCO\_AA-1 +  
 0.0033 \* Lag\_RISCO\_AA-2 +  
 -0.0016 \* Lag\_RISCO\_AA-3 +  
 -0.0039 \* Lag\_RISCO\_A-1 +  
 0.0039 \* Lag\_RISCO\_A-3 +  
 -0.0021 \* Lag\_RISCO\_A-4 +  
 0.0034 \* Lag\_RISCO\_B-1 +  
 0.0015 \* Lag\_RISCO\_B-2 +  
 -0.0018 \* Lag\_RISCO\_B-3 +  
 0.0038 \* Lag\_RISCO\_B-4 +



```

-0.0067 * Lag_RISCO_C-1 +
-0.0191 * Lag_RISCO_D-4 +
-0.0108 * Lag_RISCO_E-2 +
-0.0671 * Lag_RISCO_F-1 +
 0.1069 * Lag_RISCO_F-2 +
 0.0852 * Lag_RISCO_F-4 +
-0.0506 * Lag_RISCO_G-2 +
 0.069 * Lag_RISCO_G-3 +
-0.0112 * Lag_RISCO_H-2 +
-0.0105 * Lag_RISCO_H-3 +
1128.9096 * DATA-remapped^2 +
35.6158 * DATA-remapped^3 +
-0.0051 * DATA-remapped*Lag_INAD_TOT-2 +
-0.0001 * DATA-remapped*Lag_PATR_LIQ-2 +
 0.0004 * DATA-remapped*Lag_OPER_CRE_PROV-2 +
-0.0384 * DATA-remapped*Lag_DESP_TRIB-3 +
-0.0339 * DATA-remapped*Lag_DESP_TRIB-4 +
 4.6783 * DATA-remapped*Lag_TOT_CAR_PF-1 +
-12.5207 * DATA-remapped*Lag_TOT_CAR_PF-2 +
 8.1972 * DATA-remapped*Lag_TOT_CAR_PF-3 +
 0.0009 * DATA-remapped*Lag_RISCO_AA-1 +
-0.0004 * DATA-remapped*Lag_RISCO_AA-2 +
 0.0006 * DATA-remapped*Lag_RISCO_AA-3 +
-0.0001 * DATA-remapped*Lag_RISCO_A-1 +
-0.0001 * DATA-remapped*Lag_RISCO_A-2 +
 0.0003 * DATA-remapped*Lag_RISCO_A-3 +
-0.0001 * DATA-remapped*Lag_RISCO_A-4 +
 0.0001 * DATA-remapped*Lag_RISCO_B-1 +
 0.0001 * DATA-remapped*Lag_RISCO_B-4 +
-0.0002 * DATA-remapped*Lag_RISCO_C-1 +
-0.0002 * DATA-remapped*Lag_RISCO_C-2 +
 0.0002 * DATA-remapped*Lag_RISCO_C-3 +
-0.0006 * DATA-remapped*Lag_RISCO_E-2 +
 0.008 * DATA-remapped*Lag_RISCO_F-2 +
 0.0064 * DATA-remapped*Lag_RISCO_F-4 +
 0.0034 * DATA-remapped*Lag_RISCO_G-1 +
-0.0004 * DATA-remapped*Lag_RISCO_H-1 +
-0.0005 * DATA-remapped*Lag_RISCO_H-2 +
-0.0004 * DATA-remapped*Lag_RISCO_H-3 +
-9229562.9573

```

## Modelo Bradesco:

INAD\_TOT =

```

138329.2846 * DATA-remapped +
 0.1695 * Lag_INAD_TOT-1 +
 0.0686 * Lag_INAD_TOT-2 +
-0.0518 * Lag_INAD_TOT-4 +
-0.0011 * Lag_ATIV_TOT-4 +
-0.005 * Lag_CC_CLASS-3 +
-0.0015 * Lag_CAPT-2 +
-0.0029 * Lag_CAPT-3 +
-0.0025 * Lag_CAPT-4 +
-0.0011 * Lag_PASS_TOT-4 +
-0.175 * Lag_DESP_TRIB-3 +
-0.2091 * Lag_DESP_TRIB-4 +
-0.007 * Lag_TOT_CAR_PJ-2 +
 0.0164 * Lag_TOT_CAR_PJ-4 +
-0.0035 * Lag_TOT_GER-2 +
 0.0053 * Lag_TOT_GER-4 +
-0.0033 * Lag_RISCO_AA-4 +
-0.0169 * Lag_RISCO_A-2 +
-0.0199 * Lag_RISCO_A-3 +
 0.0087 * Lag_RISCO_B-4 +
-0.0316 * Lag_RISCO_C-1 +

```

```

0.0291 * Lag_RISCO_C-3 +
0.0232 * Lag_RISCO_C-4 +
0.0953 * Lag_RISCO_D-1 +
0.0643 * Lag_RISCO_D-2 +
-0.108 * Lag_RISCO_D-4 +
0.2028 * Lag_RISCO_E-1 +
-0.102 * Lag_RISCO_E-3 +
0.1361 * Lag_RISCO_F-3 +
-0.1313 * Lag_RISCO_F-4 +
0.1915 * Lag_RISCO_G-1 +
0.0445 * Lag_RISCO_H-1 +
-0.0353 * Lag_RISCO_H-3 +
-0.0387 * Lag_RISCO_H-4 +
-0.0139 * Lag_TOT_EXT-1 +
-0.0133 * Lag_TOT_EXT-2 +
0.0254 * Lag_TOT_EXT-4 +
3396.6178 * DATA-remapped^2 +
105.1974 * DATA-remapped^3 +
-0.0001 * DATA-remapped*Lag_PATR_LIQ-1 +
0.0023 * DATA-remapped*Lag_DESP_PESS-4 +
0.0023 * DATA-remapped*Lag_DESP_ADM-2 +
0.0018 * DATA-remapped*Lag_DESP_ADM-4 +
-0.0001 * DATA-remapped*Lag_TOT_CAR_PJ-1 +
-0.0001 * DATA-remapped*Lag_RISCO_A-1 +
0.0014 * DATA-remapped*Lag_RISCO_B-4 +
-0.0005 * DATA-remapped*Lag_RISCO_C-1 +
-0.0027 * DATA-remapped*Lag_RISCO_D-4 +
-0.0043 * DATA-remapped*Lag_RISCO_F-1 +
-0.0011 * DATA-remapped*Lag_RISCO_H-3 +
-0.0008 * DATA-remapped*Lag_RISCO_H-4 +
-0.0002 * DATA-remapped*Lag_TOT_EXT-1 +
0.0003 * DATA-remapped*Lag_TOT_EXT-4 +
16531492.7021

```

## Modelo CEF:

INAD\_TOT =

```

59576.3553 * DATA-remapped +
0.0677 * Lag_INAD_TOT-1 +
0.035 * Lag_INAD_TOT-2 +
0.0432 * Lag_INAD_TOT-3 +
-0.0004 * Lag_ATIV_TOT-1 +
0.001 * Lag_ATIV_TOT-2 +
-0.0004 * Lag_ATIV_TOT-3 +
-0.0007 * Lag_ATIV_TOT-4 +
-0.0016 * Lag_CC_CLASS-2 +
-0.0008 * Lag_CC_CLASS-3 +
0.0005 * Lag_CAPT-1 +
0.0031 * Lag_CAPT-2 +
-0.0008 * Lag_CAPT-3 +
-0.0013 * Lag_CAPT-4 +
-0.0029 * Lag_PATR_LIQ-1 +
0.007 * Lag_PATR_LIQ-4 +
-0.0451 * Lag_OPER_CRE_PROV-1 +
0.0309 * Lag_OPER_CRE_PROV-2 +
0.0261 * Lag_OPER_CRE_PROV-4 +
-0.0004 * Lag_PASS_TOT-1 +
0.001 * Lag_PASS_TOT-2 +
-0.0004 * Lag_PASS_TOT-3 +
-0.0007 * Lag_PASS_TOT-4 +
0.0301 * Lag_DESP_PESS-1 +
0.0285 * Lag_DESP_PESS-3 +
-0.0359 * Lag_DESP_ADM-2 +
0.0277 * Lag_DESP_ADM-3 +
0.1173 * Lag_DESP_TRIB-1 +

```

-0.132 \* Lag\_DESP\_TRIB-2 +  
0.0599 \* Lag\_DESP\_TRIB-3 +  
0.1343 \* Lag\_DESP\_TRIB-4 +  
-0.0008 \* Lag\_TOT\_CAR\_PF-1 +  
-0.0012 \* Lag\_TOT\_CAR\_PF-2 +  
-0.0006 \* Lag\_TOT\_CAR\_PF-3 +  
0.0041 \* Lag\_TOT\_CAR\_PJ-1 +  
0.003 \* Lag\_TOT\_CAR\_PJ-2 +  
0.0012 \* Lag\_TOT\_CAR\_PJ-3 +  
-0.0017 \* Lag\_TOT\_CAR\_PJ-4 +  
-0.0006 \* Lag\_TOT\_GER-2 +  
-0.0037 \* Lag\_RISCO\_AA-1 +  
-0.0034 \* Lag\_RISCO\_AA-4 +  
0.0014 \* Lag\_RISCO\_A-1 +  
0.0029 \* Lag\_RISCO\_A-4 +  
0.0164 \* Lag\_RISCO\_B-1 +  
0.0022 \* Lag\_RISCO\_B-2 +  
-0.0032 \* Lag\_RISCO\_B-3 +  
0.0033 \* Lag\_RISCO\_B-4 +  
0.0116 \* Lag\_RISCO\_C-1 +  
-0.014 \* Lag\_RISCO\_C-2 +  
-0.0113 \* Lag\_RISCO\_C-3 +  
0.054 \* Lag\_RISCO\_C-4 +  
0.0271 \* Lag\_RISCO\_D-1 +  
0.0655 \* Lag\_RISCO\_D-3 +  
0.0156 \* Lag\_RISCO\_D-4 +  
-0.2113 \* Lag\_RISCO\_E-1 +  
0.0245 \* Lag\_RISCO\_E-3 +  
0.0345 \* Lag\_RISCO\_E-4 +  
0.0408 \* Lag\_RISCO\_F-1 +  
0.0258 \* Lag\_RISCO\_F-2 +  
-0.0292 \* Lag\_RISCO\_F-3 +  
0.0544 \* Lag\_RISCO\_F-4 +  
0.0547 \* Lag\_RISCO\_G-1 +  
0.331 \* Lag\_RISCO\_G-2 +  
0.493 \* Lag\_RISCO\_G-3 +  
-0.2513 \* Lag\_RISCO\_G-4 +  
0.1151 \* Lag\_RISCO\_H-1 +  
-0.1329 \* Lag\_RISCO\_H-2 +  
-0.0113 \* Lag\_RISCO\_H-3 +  
-0.2357 \* Lag\_RISCO\_H-4 +  
1489.1967 \* DATA-remapped^2 +  
48.4826 \* DATA-remapped^3 +  
0.0007 \* DATA-remapped\*Lag\_INAD\_TOT-1 +  
-0.0008 \* DATA-remapped\*Lag\_INAD\_TOT-2 +  
-0.0005 \* DATA-remapped\*Lag\_INAD\_TOT-3 +  
-0.0013 \* DATA-remapped\*Lag\_INAD\_TOT-4 +  
-0.0001 \* DATA-remapped\*Lag\_PATR\_LIQ-1 +  
-0.0001 \* DATA-remapped\*Lag\_PATR\_LIQ-3 +  
0.0004 \* DATA-remapped\*Lag\_OPER\_CRE\_PROV-2 +  
0.0003 \* DATA-remapped\*Lag\_OPER\_CRE\_PROV-3 +  
0.0003 \* DATA-remapped\*Lag\_OPER\_CRE\_PROV-4 +  
0.0018 \* DATA-remapped\*Lag\_DESP\_PESS-2 +  
0.0017 \* DATA-remapped\*Lag\_DESP\_PESS-4 +  
-0.0012 \* DATA-remapped\*Lag\_DESP\_ADM-1 +  
0.003 \* DATA-remapped\*Lag\_DESP\_ADM-2 +  
0.0036 \* DATA-remapped\*Lag\_DESP\_ADM-4 +  
0.006 \* DATA-remapped\*Lag\_DESP\_TRIB-2 +  
-0.0016 \* DATA-remapped\*Lag\_DESP\_TRIB-3 +  
0.0123 \* DATA-remapped\*Lag\_DESP\_TRIB-4 +  
-0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-2 +  
-0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-3 +  
-0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-4 +  
0.0001 \* DATA-remapped\*Lag\_RISCO\_A-1 +  
-0.0001 \* DATA-remapped\*Lag\_RISCO\_A-2 +  
-0.0002 \* DATA-remapped\*Lag\_RISCO\_A-3 +  
0.0007 \* DATA-remapped\*Lag\_RISCO\_A-4 +

-0.0001 \* DATA-remapped\*Lag\_RISCO\_B-2 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_B-3 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_B-4 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_C-1 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_C-2 +  
 -0.0002 \* DATA-remapped\*Lag\_RISCO\_C-3 +  
 -0.0002 \* DATA-remapped\*Lag\_RISCO\_D-1 +  
 -0.0005 \* DATA-remapped\*Lag\_RISCO\_D-2 +  
 0.0001 \* DATA-remapped\*Lag\_RISCO\_D-3 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_D-4 +  
 -0.0024 \* DATA-remapped\*Lag\_RISCO\_E-1 +  
 -0.0009 \* DATA-remapped\*Lag\_RISCO\_E-2 +  
 -0.0017 \* DATA-remapped\*Lag\_RISCO\_E-4 +  
 0.0005 \* DATA-remapped\*Lag\_RISCO\_F-1 +  
 0.0002 \* DATA-remapped\*Lag\_RISCO\_F-2 +  
 -0.0001 \* DATA-remapped\*Lag\_RISCO\_F-3 +  
 0.0001 \* DATA-remapped\*Lag\_RISCO\_F-4 +  
 0.0028 \* DATA-remapped\*Lag\_RISCO\_G-1 +  
 0.0017 \* DATA-remapped\*Lag\_RISCO\_G-2 +  
 0.0003 \* DATA-remapped\*Lag\_RISCO\_G-3 +  
 -0.0075 \* DATA-remapped\*Lag\_RISCO\_G-4 +  
 -0.0002 \* DATA-remapped\*Lag\_RISCO\_H-1 +  
 -0.0014 \* DATA-remapped\*Lag\_RISCO\_H-2 +  
 -0.0008 \* DATA-remapped\*Lag\_RISCO\_H-3 +  
 -0.0012 \* DATA-remapped\*Lag\_RISCO\_H-4 +

20635511.2127

## Modelo Banco CSF:

INAD\_TOT =

16178.6308 \* DATA-remapped +  
 0.1145 \* Lag\_INAD\_TOT-1 +  
 0.0644 \* Lag\_INAD\_TOT-2 +  
 0.0506 \* Lag\_INAD\_TOT-3 +  
 0.0529 \* Lag\_INAD\_TOT-4 +  
 -0.0033 \* Lag\_ATIV\_TOT-1 +  
 -0.0048 \* Lag\_ATIV\_TOT-3 +  
 0.0043 \* Lag\_CC\_CLASS-1 +  
 0.0053 \* Lag\_CC\_CLASS-2 +  
 -0.0214 \* Lag\_CAPT-1 +  
 -0.0461 \* Lag\_CAPT-2 +  
 -0.0601 \* Lag\_CAPT-3 +  
 -0.0591 \* Lag\_PATR\_LIQ-1 +  
 -0.0405 \* Lag\_PATR\_LIQ-2 +  
 -0.0279 \* Lag\_PATR\_LIQ-3 +  
 -0.0334 \* Lag\_PATR\_LIQ-4 +  
 -0.094 \* Lag\_OPER\_CRE\_PROV-1 +  
 0.1804 \* Lag\_OPER\_CRE\_PROV-3 +  
 0.105 \* Lag\_OPER\_CRE\_PROV-4 +  
 -0.0033 \* Lag\_PASS\_TOT-1 +  
 -0.0048 \* Lag\_PASS\_TOT-3 +  
 0.6898 \* Lag\_DESP\_PESS-1 +  
 0.498 \* Lag\_DESP\_PESS-2 +  
 0.3945 \* Lag\_DESP\_PESS-3 +  
 0.8204 \* Lag\_DESP\_PESS-4 +  
 0.1462 \* Lag\_DESP\_ADM-1 +  
 0.0739 \* Lag\_DESP\_ADM-2 +  
 -0.1894 \* Lag\_DESP\_TRIB-1 +  
 -0.2069 \* Lag\_DESP\_TRIB-3 +  
 0.0043 \* Lag\_TOT\_CAR\_PF-1 +  
 0.0053 \* Lag\_TOT\_CAR\_PF-2 +  
 0.0043 \* Lag\_TOT\_GER-1 +  
 0.0053 \* Lag\_TOT\_GER-2 +  
 0.0048 \* Lag\_RISCO\_A-1 +  
 0.0068 \* Lag\_RISCO\_A-2 +

0.1669 \* Lag\_RISCO\_B-1 +  
 0.4068 \* Lag\_RISCO\_B-2 +  
 1.0491 \* Lag\_RISCO\_B-3 +  
 0.8945 \* Lag\_RISCO\_C-1 +  
 0.6773 \* Lag\_RISCO\_C-2 +  
 0.4304 \* Lag\_RISCO\_C-3 +  
 0.4195 \* Lag\_RISCO\_C-4 +  
 0.2815 \* Lag\_RISCO\_D-1 +  
 0.3453 \* Lag\_RISCO\_D-4 +  
 -0.2367 \* Lag\_RISCO\_E-1 +  
 -0.1819 \* Lag\_RISCO\_E-2 +  
 -0.3203 \* Lag\_RISCO\_E-3 +  
 -0.3517 \* Lag\_RISCO\_E-4 +  
 -0.2899 \* Lag\_RISCO\_F-2 +  
 -0.3142 \* Lag\_RISCO\_F-3 +  
 0.7617 \* Lag\_RISCO\_G-3 +  
 0.0831 \* Lag\_RISCO\_H-1 +  
 441.4021 \* DATA-remapped^2 +  
 16.0509 \* DATA-remapped^3 +  
 0.0001 \* DATA-remapped\*Lag\_CC\_CLASS-4 +  
 -0.0005 \* DATA-remapped\*Lag\_CAPT-2 +  
 -0.001 \* DATA-remapped\*Lag\_CAPT-3 +  
 0.0014 \* DATA-remapped\*Lag\_CAPT-4 +  
 -0.0007 \* DATA-remapped\*Lag\_PATR\_LIQ-1 +  
 -0.0005 \* DATA-remapped\*Lag\_PATR\_LIQ-2 +  
 0.0018 \* DATA-remapped\*Lag\_OPER\_CRE\_PROV-3 +  
 0.0209 \* DATA-remapped\*Lag\_DESP\_PESS-1 +  
 0.0235 \* DATA-remapped\*Lag\_DESP\_PESS-2 +  
 0.019 \* DATA-remapped\*Lag\_DESP\_PESS-4 +  
 0.0035 \* DATA-remapped\*Lag\_DESP\_ADM-1 +  
 0.0026 \* DATA-remapped\*Lag\_DESP\_ADM-2 +  
 0.0101 \* DATA-remapped\*Lag\_DESP\_TRIB-2 +  
 0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PF-4 +  
 0.0001 \* DATA-remapped\*Lag\_TOT\_GER-4 +  
 0.0002 \* DATA-remapped\*Lag\_RISCO\_A-4 +  
 0.0107 \* DATA-remapped\*Lag\_RISCO\_B-3 +  
 -0.0033 \* DATA-remapped\*Lag\_RISCO\_E-1 +  
 0.0019 \* DATA-remapped\*Lag\_RISCO\_H-4 +  
 389671.3628

## Modelo Itaú:

INAD\_TOT =

62788.0702 \* DATA-remapped +  
 -0.1025 \* Lag\_INAD\_TOT-2 +  
 0.1073 \* Lag\_INAD\_TOT-3 +  
 0.0568 \* Lag\_INAD\_TOT-4 +  
 -0.0022 \* Lag\_CC\_CLASS-1 +  
 -0.0099 \* Lag\_PATR\_LIQ-3 +  
 -0.0079 \* Lag\_PATR\_LIQ-4 +  
 -0.0322 \* Lag\_OPER\_CRE\_PROV-2 +  
 0.0491 \* Lag\_DESP\_PESS-3 +  
 0.0417 \* Lag\_DESP\_ADM-4 +  
 -0.2394 \* Lag\_DESP\_TRIB-1 +  
 0.0899 \* Lag\_DESP\_TRIB-2 +  
 0.1233 \* Lag\_DESP\_TRIB-3 +  
 -0.0025 \* Lag\_TOT\_CAR\_PJ-1 +  
 0.0068 \* Lag\_TOT\_CAR\_PJ-3 +  
 -0.0018 \* Lag\_TOT\_GER-1 +  
 0.0133 \* Lag\_RISCO\_AA-2 +  
 0.0098 \* Lag\_RISCO\_AA-3 +  
 0.0242 \* Lag\_RISCO\_A-4 +  
 0.0223 \* Lag\_RISCO\_B-1 +  
 0.0394 \* Lag\_RISCO\_B-2 +  
 0.039 \* Lag\_RISCO\_B-3 +

```

0.0284 * Lag_RISCO_B-4 +
0.057 * Lag_RISCO_C-1 +
0.0615 * Lag_RISCO_C-2 +
0.0876 * Lag_RISCO_C-4 +
0.1127 * Lag_RISCO_D-2 +
0.2072 * Lag_RISCO_D-4 +
-0.1461 * Lag_RISCO_E-2 +
-0.1645 * Lag_RISCO_E-4 +
0.2915 * Lag_RISCO_F-2 +
-0.2449 * Lag_RISCO_F-3 +
-0.0866 * Lag_RISCO_F-4 +
-0.2512 * Lag_RISCO_G-2 +
-0.2604 * Lag_RISCO_G-3 +
0.1207 * Lag_RISCO_H-1 +
-0.005 * Lag_TOT_EXT-1 +
-0.0089 * Lag_TOT_EXT-4 +
1629.5985 * DATA-remapped^2 +
54.2551 * DATA-remapped^3 +
0.0013 * DATA-remapped*Lag_INAD_TOT-4 +
0.0015 * DATA-remapped*Lag_DESP_PESS-3 +
-0.0056 * DATA-remapped*Lag_DESP_TRIB-1 +
0.0059 * DATA-remapped*Lag_DESP_TRIB-2 +
0.005 * DATA-remapped*Lag_DESP_TRIB-3 +
0.0002 * DATA-remapped*Lag_RISCO_AA-4 +
0.0001 * DATA-remapped*Lag_RISCO_A-4 +
0.0015 * DATA-remapped*Lag_RISCO_B-4 +
0.0019 * DATA-remapped*Lag_RISCO_D-4 +
-0.0031 * DATA-remapped*Lag_RISCO_E-4 +
-0.0045 * DATA-remapped*Lag_RISCO_F-1 +
-0.0037 * DATA-remapped*Lag_RISCO_F-3 +
-0.0016 * DATA-remapped*Lag_RISCO_H-3 +
2352670.9942

```

### Modelo Nubank:

INAD\_TOT =

```

38109.1388 * DATA-remapped +
-0.0678 * Lag_INAD_TOT-1 +
-0.0059 * Lag_CC_CLASS-1 +
-0.0047 * Lag_CAPT-1 +
-0.0601 * Lag_PATR_LIQ-1 +
-0.0348 * Lag_PATR_LIQ-4 +
-0.0143 * Lag_RISCO_A-3 +
-0.4453 * Lag_RISCO_B-2 +
-1.8502 * Lag_RISCO_E-4 +
1879.0636 * DATA-remapped^2 +
115.9801 * DATA-remapped^3 +
-0.0027 * DATA-remapped*Lag_INAD_TOT-1 +
-0.0002 * DATA-remapped*Lag_CC_CLASS-1 +
-0.0002 * DATA-remapped*Lag_CAPT-1 +
-0.0023 * DATA-remapped*Lag_PATR_LIQ-1 +
0.0664 * DATA-remapped*Lag_DESP_TRIB-4 +
235897.2825

```

### Modelo Banco Pan:

INAD\_TOT =

```

19929.5505 * DATA-remapped +
-0.0712 * Lag_INAD_TOT-1 +
0.0568 * Lag_INAD_TOT-3 +
-0.1075 * Lag_INAD_TOT-4 +
-0.0078 * Lag_ATIV_TOT-2 +
0.0157 * Lag_ATIV_TOT-4 +
-0.0149 * Lag_CC_CLASS-3 +

```

0.0081 \* Lag\_CAPT-1 +  
 -0.0586 \* Lag\_PATR\_LIQ-2 +  
 -0.0592 \* Lag\_PATR\_LIQ-4 +  
 0.0988 \* Lag\_OPER\_CRE\_PROV-3 +  
 0.0072 \* Lag\_PASS\_TOT-3 +  
 0.0157 \* Lag\_PASS\_TOT-4 +  
 -0.1565 \* Lag\_DESP\_PESS-2 +  
 0.4582 \* Lag\_DESP\_PESS-3 +  
 -0.0747 \* Lag\_DESP\_ADM-1 +  
 -0.0582 \* Lag\_DESP\_ADM-4 +  
 -0.8985 \* Lag\_DESP\_TRIB-1 +  
 -0.0053 \* Lag\_TOT\_CAR\_PF-3 +  
 -0.0154 \* Lag\_TOT\_GER-3 +  
 -0.0074 \* Lag\_RISCO\_A-3 +  
 0.0447 \* Lag\_RISCO\_C-2 +  
 0.0417 \* Lag\_RISCO\_C-3 +  
 0.1642 \* Lag\_RISCO\_D-3 +  
 0.0847 \* Lag\_RISCO\_D-4 +  
 0.1692 \* Lag\_RISCO\_E-1 +  
 0.3058 \* Lag\_RISCO\_E-2 +  
 0.2852 \* Lag\_RISCO\_F-1 +  
 0.1926 \* Lag\_RISCO\_F-2 +  
 -0.7034 \* Lag\_RISCO\_G-1 +  
 1.4687 \* Lag\_RISCO\_G-2 +  
 -0.8161 \* Lag\_RISCO\_G-4 +  
 -0.1633 \* Lag\_RISCO\_H-3 +  
 499.0451 \* DATA-remapped^2 +  
 15.5999 \* DATA-remapped^3 +  
 -0.0018 \* DATA-remapped\*Lag\_INAD\_TOT-1 +  
 -0.002 \* DATA-remapped\*Lag\_INAD\_TOT-4 +  
 -0.0003 \* DATA-remapped\*Lag\_PATR\_LIQ-2 +  
 0.0141 \* DATA-remapped\*Lag\_DESP\_PESS-3 +  
 -0.0015 \* DATA-remapped\*Lag\_DESP\_ADM-1 +  
 -0.0026 \* DATA-remapped\*Lag\_DESP\_ADM-4 +  
 -0.0179 \* DATA-remapped\*Lag\_DESP\_TRIB-1 +  
 0.0211 \* DATA-remapped\*Lag\_DESP\_TRIB-2 +  
 0.0313 \* DATA-remapped\*Lag\_DESP\_TRIB-3 +  
 -0.0171 \* DATA-remapped\*Lag\_DESP\_TRIB-4 +  
 0.0016 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-1 +  
 0.0047 \* DATA-remapped\*Lag\_RISCO\_B-4 +  
 0.0075 \* DATA-remapped\*Lag\_RISCO\_C-3 +  
 0.0064 \* DATA-remapped\*Lag\_RISCO\_C-4 +  
 -0.0053 \* DATA-remapped\*Lag\_RISCO\_D-1 +  
 0.0054 \* DATA-remapped\*Lag\_RISCO\_D-3 +  
 -0.0093 \* DATA-remapped\*Lag\_RISCO\_G-1 +  
 0.0065 \* DATA-remapped\*Lag\_RISCO\_G-2 +  
 -0.009 \* DATA-remapped\*Lag\_RISCO\_G-4 +  
 216372.7035

## Modelo Santander:

INAD\_TOT =

23677.3395 \* DATA-remapped +  
 -0.0365 \* Lag\_INAD\_TOT-2 +  
 -0.1057 \* Lag\_INAD\_TOT-4 +  
 0.0005 \* Lag\_ATIV\_TOT-1 +  
 0.0003 \* Lag\_ATIV\_TOT-2 +  
 0.0007 \* Lag\_CC\_CLASS-1 +  
 0.0009 \* Lag\_CAPT-1 +  
 0.0009 \* Lag\_CAPT-2 +  
 0.0005 \* Lag\_CAPT-4 +  
 0.0048 \* Lag\_PATR\_LIQ-2 +  
 -0.0098 \* Lag\_PATR\_LIQ-4 +  
 -0.0309 \* Lag\_OPER\_CRE\_PROV-1 +  
 -0.0293 \* Lag\_OPER\_CRE\_PROV-2 +

```

0.0538 * Lag_OPER_CRE_PROV-3 +
-0.0261 * Lag_OPER_CRE_PROV-4 +
0.0005 * Lag_PASS_TOT-1 +
0.0003 * Lag_PASS_TOT-2 +
0.0806 * Lag_DESP_TRIB-1 +
0.1957 * Lag_DESP_TRIB-2 +
-0.0916 * Lag_DESP_TRIB-3 +
0.0025 * Lag_TOT_CAR_PJ-1 +
0.0063 * Lag_TOT_CAR_PJ-2 +
-0.0072 * Lag_TOT_CAR_PJ-4 +
0.0007 * Lag_TOT_GER-1 +
0.0037 * Lag_RISCO_AA-1 +
-0.0049 * Lag_RISCO_A-1 +
0.0053 * Lag_RISCO_A-2 +
0.0044 * Lag_RISCO_A-4 +
0.0078 * Lag_RISCO_C-1 +
-0.009 * Lag_RISCO_C-4 +
0.0309 * Lag_RISCO_D-1 +
0.0527 * Lag_RISCO_D-3 +
-0.0249 * Lag_RISCO_D-4 +
0.0566 * Lag_RISCO_E-1 +
0.0754 * Lag_RISCO_E-2 +
-0.1798 * Lag_RISCO_E-3 +
0.0681 * Lag_RISCO_E-4 +
-0.0974 * Lag_RISCO_F-1 +
-0.0509 * Lag_RISCO_F-2 +
-0.2112 * Lag_RISCO_F-4 +
-0.3767 * Lag_RISCO_G-2 +
0.3063 * Lag_RISCO_G-3 +
0.6254 * Lag_RISCO_G-4 +
0.0576 * Lag_RISCO_H-3 +
0.01 * Lag_TOT_EXT-1 +
0.0054 * Lag_TOT_EXT-2 +
607.5076 * DATA-remapped^2 +
18.8 * DATA-remapped^3 +
-0.0018 * DATA-remapped*Lag_DESP_PESS-2 +
-0.0021 * DATA-remapped*Lag_DESP_PESS-4 +
-0.0011 * DATA-remapped*Lag_DESP_ADM-2 +
0.0011 * DATA-remapped*Lag_DESP_ADM-3 +
-0.0011 * DATA-remapped*Lag_DESP_ADM-4 +
0.0019 * DATA-remapped*Lag_DESP_TRIB-1 +
0.002 * DATA-remapped*Lag_DESP_TRIB-2 +
0.0001 * DATA-remapped*Lag_RISCO_C-1 +
-0.0002 * DATA-remapped*Lag_RISCO_C-4 +
0.0004 * DATA-remapped*Lag_RISCO_D-1 +
0.0004 * DATA-remapped*Lag_RISCO_D-3 +
0.0009 * DATA-remapped*Lag_RISCO_E-2 +
-0.0007 * DATA-remapped*Lag_RISCO_E-3 +
0.0004 * DATA-remapped*Lag_RISCO_H-2 +
0.0002 * DATA-remapped*Lag_TOT_EXT-1 +
2509920.3427

```

## Modelo Banco Votorantim:

INAD\_TOT =

```

-30823.9805 * DATA-remapped +
0.0525 * Lag_INAD_TOT-2 +
-0.0738 * Lag_INAD_TOT-4 +
0.0028 * Lag_ATIV_TOT-2 +
0.0024 * Lag_ATIV_TOT-3 +
0.0035 * Lag_CC_CLASS-1 +
0.0073 * Lag_CC_CLASS-2 +
0.0057 * Lag_CC_CLASS-3 +
-0.0027 * Lag_CAPT-1 +
0.005 * Lag_CAPT-2 +

```



0.0046 \* Lag\_CAPT-3 +  
 0.0102 \* Lag\_PATR\_LIQ-1 +  
 0.0143 \* Lag\_PATR\_LIQ-3 +  
 -0.0765 \* Lag\_OPER\_CRE\_PROV-2 +  
 0.0819 \* Lag\_OPER\_CRE\_PROV-4 +  
 0.0028 \* Lag\_PASS\_TOT-2 +  
 0.0024 \* Lag\_PASS\_TOT-3 +  
 -0.1284 \* Lag\_DESP\_PESS-2 +  
 -0.0749 \* Lag\_DESP\_ADM-4 +  
 0.0064 \* Lag\_TOT\_CAR\_PF-1 +  
 0.0062 \* Lag\_TOT\_CAR\_PF-2 +  
 0.0069 \* Lag\_TOT\_CAR\_PF-3 +  
 0.0108 \* Lag\_TOT\_CAR\_PF-4 +  
 0.0035 \* Lag\_TOT\_GER-1 +  
 0.0067 \* Lag\_TOT\_GER-2 +  
 0.0033 \* Lag\_TOT\_GER-3 +  
 -0.0283 \* Lag\_RISCO\_AA-1 +  
 0.0292 \* Lag\_RISCO\_AA-3 +  
 -0.0356 \* Lag\_RISCO\_AA-4 +  
 0.012 \* Lag\_RISCO\_A-1 +  
 0.0126 \* Lag\_RISCO\_A-2 +  
 0.0054 \* Lag\_RISCO\_A-4 +  
 0.0148 \* Lag\_RISCO\_B-1 +  
 -0.0211 \* Lag\_RISCO\_C-1 +  
 -0.0358 \* Lag\_RISCO\_C-2 +  
 0.0149 \* Lag\_RISCO\_C-3 +  
 0.0278 \* Lag\_RISCO\_C-4 +  
 -0.0474 \* Lag\_RISCO\_D-1 +  
 -0.0703 \* Lag\_RISCO\_D-4 +  
 -0.1521 \* Lag\_RISCO\_E-1 +  
 0.1948 \* Lag\_RISCO\_E-2 +  
 -0.3086 \* Lag\_RISCO\_E-3 +  
 -0.1559 \* Lag\_RISCO\_E-4 +  
 0.4066 \* Lag\_RISCO\_F-2 +  
 0.1723 \* Lag\_RISCO\_F-3 +  
 -0.1749 \* Lag\_RISCO\_F-4 +  
 0.0499 \* Lag\_RISCO\_G-2 +  
 -0.1074 \* Lag\_RISCO\_H-1 +  
 -0.0834 \* Lag\_RISCO\_H-2 +  
 0.1189 \* Lag\_RISCO\_H-4 +  
 0.0133 \* Lag\_TOT\_EXT-3 +  
 -0.0243 \* Lag\_TOT\_EXT-4 +  
 -769.5589 \* DATA-remapped^2 +  
 -23.5338 \* DATA-remapped^3 +  
 0.0025 \* DATA-remapped\*Lag\_INAD\_TOT-2 +  
 0.0029 \* DATA-remapped\*Lag\_INAD\_TOT-3 +  
 -0.007 \* DATA-remapped\*Lag\_INAD\_TOT-4 +  
 0.0001 \* DATA-remapped\*Lag\_CC\_CLASS-3 +  
 0.0002 \* DATA-remapped\*Lag\_PATR\_LIQ-3 +  
 -0.001 \* DATA-remapped\*Lag\_OPER\_CRE\_PROV-3 +  
 -0.0033 \* DATA-remapped\*Lag\_DESP\_PESS-2 +  
 -0.003 \* DATA-remapped\*Lag\_DESP\_PESS-3 +  
 -0.0027 \* DATA-remapped\*Lag\_DESP\_ADM-3 +  
 -0.0123 \* DATA-remapped\*Lag\_DESP\_TRIB-3 +  
 0.0001 \* DATA-remapped\*Lag\_TOT\_CAR\_PF-3 +  
 0.0003 \* DATA-remapped\*Lag\_TOT\_CAR\_PJ-3 +  
 0.0001 \* DATA-remapped\*Lag\_TOT\_GER-3 +  
 -0.0005 \* DATA-remapped\*Lag\_RISCO\_AA-1 +  
 0.0004 \* DATA-remapped\*Lag\_RISCO\_AA-3 +  
 0.0001 \* DATA-remapped\*Lag\_RISCO\_A-3 +  
 0.0003 \* DATA-remapped\*Lag\_RISCO\_B-3 +  
 0.0003 \* DATA-remapped\*Lag\_RISCO\_C-3 +  
 0.0015 \* DATA-remapped\*Lag\_RISCO\_D-3 +  
 0.0048 \* DATA-remapped\*Lag\_RISCO\_F-3 +  
 0.0009 \* DATA-remapped\*Lag\_RISCO\_G-2 +  
 0.002 \* DATA-remapped\*Lag\_RISCO\_G-3 +  
 0.0013 \* DATA-remapped\*Lag\_RISCO\_G-4 +

0.002 \* DATA-remapped\*Lag\_RISCO\_H-3 +  
-0.0041 \* DATA-remapped\*Lag\_TOT\_EXT-1 +  
0.0072 \* DATA-remapped\*Lag\_TOT\_EXT-3 +  
-0.0039 \* DATA-remapped\*Lag\_TOT\_EXT-4 +  
-3068035.5008

## Anexo IV

### Script de geração dos gráficos de inadimplência total

## Importação das bibliotecas

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import dates
import datetime
```

## Função para definir endereço das tabelas

```
In [2]: def pathg(banco):
endereco = "C:/Users/grama/Desktop/TCC - Computação/Projeções/Inad "
final = ".csv"
resultado = endereco + banco + final
return resultado
```

## Importação do arquivo CSV

```
In [7]: nome_banco = "BB"

dados = pd.read_csv(pathg(nome_banco), sep = ';', thousands = '.', parse_dates=True)
#itau.Time = pd.to_datetime(itau.Time)
dados.set_index("Time", inplace=True)
```

## Geração do Gráfico da série temporal

```
In [ ]: fig, ax = plt.subplots(figsize=(15, 6))

ax.plot(dados.index,
        dados['INAD_TOT_MIL'].str.replace(',', '.').astype(float),
        marker='o',
        color = 'purple')

ax.set(xlabel="Trimestre",
        ylabel="Valor em milhões de R$",
        title="Inadimplência Total BB"
)

plt.gcf().autofmt_xdate(rotation=90)
plt.ticklabel_format(style='plain', axis='y')

#plt.show()
plt.savefig('inad_bb.png')
```