

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Previsão de Pontos Críticos de Crime

*Uma modelagem por aprendizado de
máquina para prevenção de roubos de
celulares em São Paulo*

Julio Kenji Ueda

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE
FORMATURA SUPERVISIONADO

Orientador: Prof. Dr. Alfredo Goldman

São Paulo
10 de Março de 2021

Previsão de Pontos Críticos de Crime
Uma modelagem por aprendizado de máquina para prevenção de roubos de celulares em São Paulo

Julio Kenji Ueda

Esta é a versão original da monografia elaborada pelo candidato Julio Kenji Ueda, tal como submetida à Comissão Julgadora.

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Resumo

Julio Kenji Ueda. **Previsão de Pontos Críticos de Crime: Uma modelagem por aprendizado de máquina para prevenção de roubos de celulares em São Paulo.** Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

O roubo de celulares é um problema que vem crescendo atualmente. Concentrações espaciais de atividades criminosas sugerem que crimes não são aleatórios no espaço e no tempo, podem ser estudados e padrões podem ser modelados. O objetivo deste trabalho é o desenvolvimento e avaliação de um modelo de previsão de pontos críticos de crimes por aprendizado de máquina, alimentados com dados dos boletins de ocorrência disponíveis no portal da transparência da Secretaria da Segurança Pública de São Paulo. O modelo realiza uma tarefa de classificação binária, prevendo se uma determinada área será um ponto crítico de crime ou não no próximo mês. Resultados dos experimentos mostram que o modelo obtém uma acurácia em torno de 75%, podendo alcançar até 85%, na previsão dos pontos críticos de crimes no Estado de São Paulo.

Palavras-chave: Modelo de Previsão, Ponto Crítico de Crime, Aprendizado de Máquina, Prevenção de Roubo

Abstract

Julio Kenji Ueda. **Crime Hotspots Forecasting: *A machine learning model for mobile phone robbery prevention in São Paulo***. Capstone Project Report (Bachelor).
Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2021.

Mobile phone robbery is a problem that is growing nowadays. Spatial concentrations of criminal activity suggest that crime is not random in time and space, can be studied and patterns can be modeled. The objective of this work is the development and evaluation of a machine learning crime hotspots forecasting model, with data from the police reports available on the transparency portal of the São Paulo Department of Public Security. The model performs a binary classification task, forecasting whether a particular area will be a crime hotspot or not in the next month. Results of the experiments show that the model obtains an accuracy of around 75%, and can reach up to 85%, of the crime hotspots forecasts in the State of São Paulo.

Keywords: Forecasting Model, Crime Hotspot, Machine Learning, Robbery Prevention

Lista de Abreviaturas

SSP-SP	Secretaria da Segurança Pública de São Paulo
IMEI	Identidade Internacional de Equipamentos Móveis
DOPE	Departamento de Operações Policiais Estratégicas
V	Verdadeiro
F	Falso
VP	Verdadeiros positivos
VN	Verdadeiros negativos
FP	Falsos Positivos
FN	Falsos Negativos
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo

Lista de Figuras

1.1	Cartaz de alerta em áreas de risco de roubo de celular	2
2.1	Exemplo de Série Temporal.	5
2.2	Exemplo de Diagrama de <i>Voronoi</i>	6
2.3	Exemplo de Complexidade Computacional.	7
2.4	Exemplos de distribuição das amostras para demonstração dos comportamentos dos algoritmos.	7
2.5	Exemplo de Regressão Linear.	8
2.6	Exemplo da estrutura da Árvore de Decisão.	9
2.7	Exemplo do resultado de regressão da Árvore de Decisão e da Floresta Aleatória.	10
2.8	Exemplos de Perceptron e Perceptron Multicamadas.	10
2.9	Exemplo de Regressão com Perceptron Multicamadas.	12
2.10	Comparação dos modelos de Regressão.	12
2.11	Exemplo do agrupamento K-Means e suas fronteiras de decisão.	13
2.12	Exemplo de funcionamento do algoritmo <i>K-Means</i>	14
2.13	Matriz de confusão	15
3.1	Exemplos de divisões de uma região.	18
4.1	Exemplo de distribuição dos dados de treinamento e teste.	19
4.2	Exemplo de agrupamento dos dados de treinamento e teste.	20
4.3	Exemplo de criação de séries temporais com os dados de treinamento e teste.	20
4.4	Exemplo dos números da previsão e teste.	21
4.5	Exemplo dos números da previsão e teste.	21
4.6	Visão Geral da metodologia.	22
4.7	Proporção de dados disponíveis para cada atributo.	24
4.8	Exemplo de distribuição de crimes em uma cidade fictícia, composta por 8 bairros quadrados.	26

4.9	Exemplo de agrupamento com 8 centróides, inicializados aleatoriamente.	27
4.10	Exemplo de agrupamento com 8 centróides, inicializados pela posição média das amostras de cada bairro.	27
4.11	Exemplo de classificação dos resultados.	30
4.12	Exemplo de avaliação dos resultados.	32
4.13	Matrizes de confusão	32
5.1	Diagrama dos testes e resultados dos experimentos.	36
5.2	Resultado e comparação das acurácias mínimas dos regressores com número de agrupamentos definido.	37
5.3	Resultado e comparação das acurácias médias dos regressores com número de agrupamentos definido.	38
5.4	Resultado e comparação dos desvios padrão das acurácias dos regressores com número de agrupamentos definido.	39
5.5	Resultado e comparação das medidas F_1 mínimas dos regressores com número de agrupamentos definido.	40
5.6	Resultado e comparação das medidas F_1 médias dos regressores com número de agrupamentos definido.	41
5.7	Resultado e comparação dos desvios padrão das medidas F_1 dos regressores com número de agrupamentos definido.	42
5.8	Resultado e comparação do tempo de treinamento máximo dos regressores com número de agrupamentos definido.	43
5.9	Resultado e comparação do tempo de treinamento médio dos regressores com número de agrupamentos definido.	44
5.10	Resultado e comparação dos desvios padrão do tempo de treinamento dos regressores com número de agrupamentos definido.	45
5.11	Resultado e comparação das acurácias mínimas dos classificadores com número de agrupamentos definido.	46
5.12	Resultado e comparação das acurácias médias dos classificadores com número de agrupamentos definido.	47
5.13	Resultado e comparação dos desvios padrão das acurácias dos classificadores com número de agrupamentos definido.	48
5.14	Resultado e comparação das medidas F_1 mínimas dos classificadores com número de agrupamentos definido.	49
5.15	Resultado e comparação das medidas F_1 médias dos classificadores com número de agrupamentos definido.	50
5.16	Resultado e comparação dos desvios padrão das medidas F_1 dos classificadores com número de agrupamentos definido.	51

5.17	Resultado e comparação dos tempos de treinamento máximo dos classificadores com número de agrupamentos definido.	52
5.18	Resultado e comparação dos tempos de treinamento médio dos classificadores com número de agrupamentos definido.	53
5.19	Resultado e comparação dos desvios padrão dos tempos de treinamento dos classificadores com número de agrupamentos definido.	54
5.20	Resultado e comparação das acurácias médias do regressor e classificador, com número de agrupamentos definido.	55
5.21	Resultado e comparação das medidas F_1 médias do regressor e classificador, com número de agrupamentos definido.	56
5.22	Resultado e comparação dos tempos de treinamento médio do regressor e classificador, com número de agrupamentos definido.	57
5.23	Resultado e comparação das acurácias mínimas dos regressores com número de agrupamentos dinâmico.	57
5.24	Resultado e comparação das acurácias médias dos regressores com número de agrupamentos dinâmico.	58
5.25	Resultado e comparação dos desvios padrão das acurácias dos regressores com número de agrupamentos dinâmico.	58
5.26	Resultado e comparação das medidas F_1 mínimas dos regressores com número de agrupamentos dinâmico.	59
5.27	Resultado e comparação das medidas F_1 médias dos regressores com número de agrupamentos dinâmico.	59
5.28	Resultado e comparação dos desvios padrão das medidas F_1 dos regressores com número de agrupamentos dinâmico.	60
5.29	Resultado e comparação dos tempos de treinamento máximo dos regressores com número de agrupamentos dinâmico.	60
5.30	Resultado e comparação dos tempos de treinamento médio dos regressores com número de agrupamentos dinâmico.	61
5.31	Resultado e comparação dos desvios padrão dos tempos de treinamento dos regressores com número de agrupamentos dinâmico.	61
5.32	Resultado e comparação das acurácias mínimas dos classificadores com número de agrupamentos dinâmico.	62
5.33	Resultado e comparação das acurácias médias dos classificadores com número de agrupamentos dinâmico.	62
5.34	Resultado e comparação dos desvios padrão das acurácias dos classificadores com número de agrupamentos dinâmico.	63

5.35	Resultado e comparação das medidas F_1 mínimas dos classificadores com número de agrupamentos dinâmico.	63
5.36	Resultado e comparação das medidas F_1 médias dos classificadores com número de agrupamentos dinâmico.	64
5.37	Resultado e comparação dos desvios padrão das medidas F_1 dos classificadores com número de agrupamentos dinâmico.	64
5.38	Resultado e comparação dos tempos de treinamento máximo dos classificadores com número de agrupamentos dinâmico.	65
5.39	Resultado e comparação dos tempos de treinamento médio dos classificadores com número de agrupamentos dinâmico.	65
5.40	Resultado e comparação dos desvios padrão dos tempos de treinamento dos classificadores com número de agrupamentos dinâmico.	66
5.41	Resultado e comparação das acurácias médias do regressor e classificador, com número de agrupamentos dinâmico.	66
5.42	Resultado e comparação das medidas F_1 médias do regressor e classificador, com número de agrupamentos dinâmico.	67
5.43	Resultado e comparação dos tempos de treinamento médio do regressor e classificador, com número de agrupamentos dinâmico.	67
6.1	Resultados das duas abordagens de agrupamentos	70
6.2	Resultados dos modelos na região metropolitana de São Paulo.	71
6.3	Resultados dos modelos na região interior do estado de São Paulo.	72
6.4	Número médio de agrupamentos para a abordagem de agrupamentos dinâmico.	73

Lista de Tabelas

4.1	Avaliação do modelo de previsão usando acurácia e medida F_1	21
4.2	Atributos com mais informações e exemplo de uma amostra.	25
4.3	Exemplo de dados de treinamento e teste.	28
4.4	Exemplo dados de treinamento e teste para modelo de classificação.	29
4.5	Exemplo de previsão dos números de crimes.	30
6.1	Resultados dos trabalhos relacionados.	74

Sumário

1	Introdução	1
1.1	Motivação	1
1.2	Objetivo	3
1.3	Capítulos	3
2	Conceitos	5
2.1	Previsão e Séries Temporais	5
2.2	Diagrama de <i>Voronoi</i>	6
2.3	Complexidade Computacional	6
2.4	Aprendizado de Máquina	7
	2.4.1 Aprendizado Supervisionado	7
	2.4.2 Aprendizado não Supervisionado	13
2.5	Métricas de Desempenho	15
	2.5.1 Acurácia	15
	2.5.2 Medida F_1	16
3	Trabalhos Relacionados	17
3.1	Geradores e Atratores de Crimes	17
3.2	Metodologia	17
4	Metodologia	19
4.1	Visão Geral	19
4.2	Aquisição dos Dados	23
4.3	Modelo de Agrupamento	26
	4.3.1 Número de agrupamentos definido	27
	4.3.2 Número de agrupamentos dinâmico	27
4.4	Série Temporal	28
4.5	Modelos de Previsões	30
	4.5.1 Modelo Fictício	30

4.5.2	Modelo Linear	31
4.5.3	Floresta Aleatória	31
4.5.4	Perceptron Multicamadas	31
4.6	Avaliação dos Modelos	32
4.6.1	Acurácia	32
4.6.2	Medida F_1	33
5	Resultados dos Experimentos	35
5.1	Regressores (agrupamento definido)	37
5.1.1	Acurácia	37
5.1.2	Medida F_1	40
5.1.3	Complexidade Computacional	43
5.2	Classificadores (agrupamento definido)	46
5.2.1	Acurácia	46
5.2.2	Medida F_1	49
5.2.3	Complexidade Computacional	52
5.3	Melhor abordagem (agrupamento definido)	55
5.3.1	Acurácia	55
5.3.2	Medida F_1	56
5.3.3	Complexidade Computacional	56
5.4	Regressores (agrupamento dinâmico)	57
5.4.1	Acurácia	57
5.4.2	Medida F_1	59
5.4.3	Complexidade Computacional	60
5.5	Classificadores (agrupamento dinâmico)	62
5.5.1	Acurácia	62
5.5.2	Medida F_1	63
5.5.3	Complexidade Computacional	65
5.6	Melhor abordagem (agrupamento dinâmico)	66
5.6.1	Acurácia	66
5.6.2	Medida F_1	67
5.6.3	Complexidade Computacional	67
6	Discussão dos Resultados	69
6.1	Modelo de Previsão	69
6.2	Modelo de Agrupamento	71
6.3	Desempenho Geral	74

7 Conclusão	75
--------------------	-----------

Referências	77
--------------------	-----------

Capítulo 1

Introdução

1.1 Motivação

O roubo de aparelhos celulares é um problema crescente nos dias atuais. Quando o usuário tem medo de atender na rua o seu celular, este deixa de ser móvel e passa a ser portátil, utilizado somente em ilhas onde a pessoa se sente segura, como a sua casa, seu trabalho, dentro de estabelecimentos comerciais, etc. Portanto, a falta de segurança pública afeta diretamente o propósito central do serviço de telefonia móvel e toda a sua infraestrutura (Box, 2019).

De acordo com a Secretaria da Segurança Pública de São Paulo (SSP-SP), cerca de 25% dos crimes de roubo e furto no estado de São Paulo em 2019 têm como alvo o aparelho celular (SSP-SP, 2019). Os alvos preferidos são as pessoas que andam distraídos pela rua, falando ao celular ou com o aparelho no bolso de trás (JN, 2016). Os aparelhos roubados vão parar no mercado clandestino (SP1, 2017).

As vítimas são orientadas a informar a Identidade Internacional de Equipamentos Móveis (IMEI) do aparelho no boletim de ocorrência, para que as operadoras de telefonia celular bloqueiem os aparelhos. Porém, investigações acerca dos destinos dos aparelhos celulares realizado pelo Departamento de Operações Policiais Estratégicas (DOPE) apuraram que grande parte dos aparelhos seriam levados para Dacar, em Senegal, na África, onde não têm o controle do IMEI (BAND, 2020b). Portanto, muitos aparelhos não retornam às vítimas.

Os danos causados, além da perda do aparelho, incluem a utilização dos dados das vítimas para realizar transferências de dinheiro (SP2, 2020) e latrocínio (SP1, 2018b; SP1, 2018a; SP2, 2018). Para alertar a população, a Polícia Militar instalou cartazes em áreas de risco de roubo de celular (BAND, 2020c). Estes cartazes, financiados pela Vigilância Solidária (Figura 1.1), têm o objetivo de reduzir das chances de ações dos criminosos. (BAND, 2020a).



Figura 1.1: Cartaz de alerta em áreas de risco de roubo de celular

1.2 Objetivo

O objetivo deste trabalho é o desenvolvimento e avaliação de um modelo capaz de prever os pontos críticos de crime (áreas com alto índice de crimes). A modelagem será feita por aprendizado de máquina, com os dados dos boletins de ocorrência de roubo de celulares disponíveis no portal da transparência da Secretaria da Segurança Pública de São Paulo. A construção será guiada pela metodologia dos trabalhos relacionados, com a meta de obter desempenho similar, mas optando pela simplicidade para que o modelo seja fácil de entender e implementar.

1.3 Capítulos

Este trabalho é estruturado da seguinte forma:

- O capítulo 2 descreve, da forma mais simples possível e por meio de exemplos, os conceitos necessários para o entendimento deste trabalho.
- O capítulo 3 aborda os trabalhos relacionados, com uma breve descrição dos conceitos, métodos e resultados.
- O capítulo 4 apresenta uma visão geral, por meio de exemplos, e a descrição dos métodos utilizados nos experimentos.
- O capítulo 5 apresenta e compara todos os resultados dos experimentos.
- O capítulo 6 analisa e discute os resultados dos melhores modelos.

Por fim, o capítulo 7 apresenta a conclusão do trabalho.

Capítulo 2

Conceitos

Este capítulo abordará todos os conceitos necessários para o entendimento deste trabalho, da forma mais simples possível. Os tópicos abordados são: [Previsão e Séries Temporais](#), [Diagrama de Voronoi](#), [Complexidade Computacional](#), [Aprendizado de Máquina](#) e [Métricas de Desempenho](#).

2.1 Previsão e Séries Temporais

Uma previsão é uma predição de algum evento futuro. A maioria dos problemas de previsão envolvem o uso de dados de séries temporais. Uma série temporal é uma sequência cronológica de observações de uma variável de interesse, onde as observações são coletadas em períodos de tempo igualmente espaçados ([MONTGOMERY *et al.*, 2015](#)) (Figura 2.1).

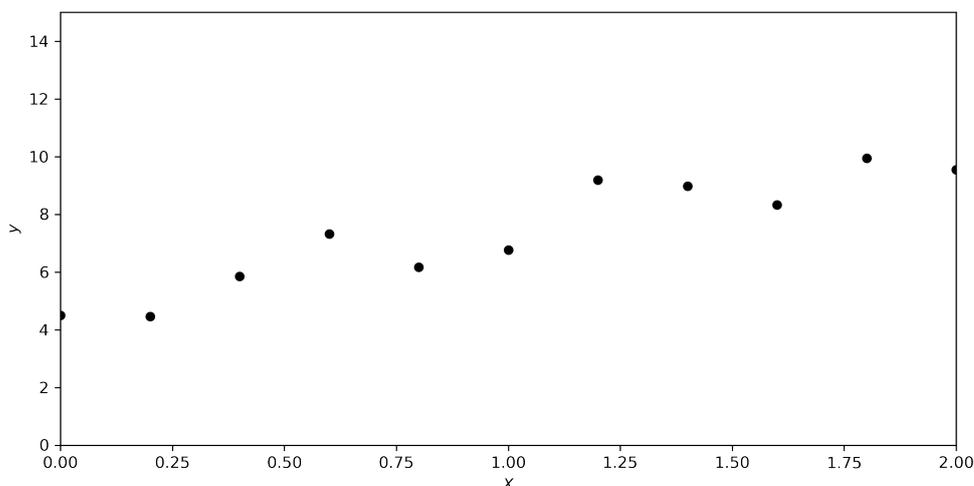


Figura 2.1: Exemplo de *Série Temporal*.

2.2 Diagrama de Voronoi

O conceito é simples, mas intuitivamente atraente. Dado um conjunto finito de pontos distintos e isolados em um espaço contínuo, associamos todos os locais nesse espaço ao membro mais próximo do conjunto de pontos. O resultado é uma divisão do espaço em um conjunto de regiões (OKABE *et al.*, 2000) (Figura 2.2).

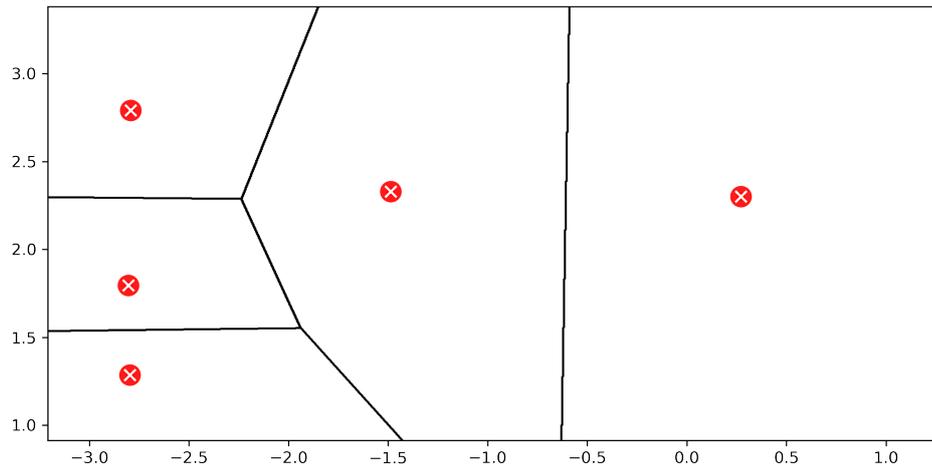


Figura 2.2: Exemplo de Diagrama de Voronoi.

2.3 Complexidade Computacional

Neste trabalho refere-se ao limite superior da complexidade computacional assintótica, utilizando a notação O (O-grande). Em outras palavras, ter uma noção do tempo de execução “no pior caso” e dos recursos computacionais que serão necessários. A maioria dos algoritmos possui um parâmetro que afeta o tempo de execução de forma mais significativa, que pode ser o número de atributos, número de amostras, etc. Seja n tal parâmetro:

- Um algoritmo $O(\log_2 n)$ é dito ter complexidade logarítmica: para dobrar o valor de $\log_2 n$ temos que considerar o quadrado de n .
- Um algoritmo $O(n)$ é dito ter complexidade linear: cada vez que n dobra de tamanho o tempo de execução dobra.
- Um algoritmo $O(n^2)$ é dito ter complexidade quadrática: sempre que n dobra o tempo de execução é multiplicado por 4.
- Um algoritmo $O(2^n)$ é dito ter complexidade exponencial: quando n dobra, o tempo de execução fica elevado ao quadrado.

Programas podem ser avaliados através da comparação de suas funções de complexidade (ZIVIANI, 2004) (Figura 2.3).

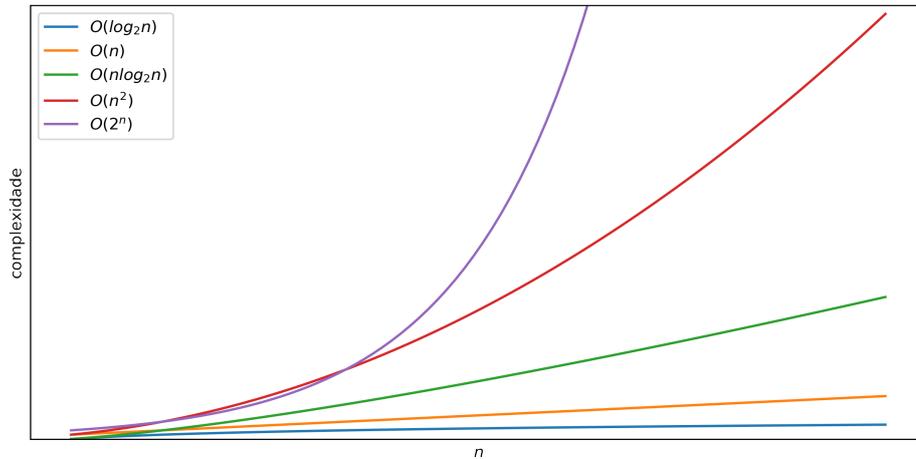


Figura 2.3: Exemplo de Complexidade Computacional.

2.4 Aprendizado de Máquina

Algoritmos de aprendizado de máquina constroem um modelo matemático baseado em dados amostrais, a fim de fazer previsões ou decisões sem serem explicitamente programados para isso (SAMUEL, 1959). O objetivo principal é permitir que os computadores aprendam automaticamente sem assistência ou intervenção humana. Neste trabalho são utilizados as técnicas de **aprendizado supervisionado** e **não supervisionado**. A fim de demonstrar o comportamento dos algoritmos, considere as distribuições das amostras da figura 2.4:

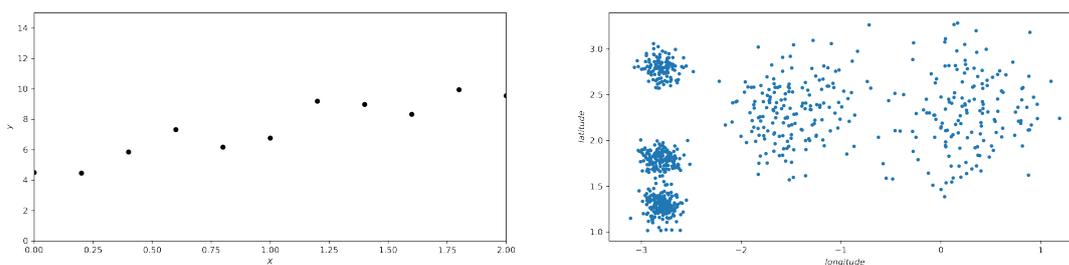


Figura 2.4: Exemplos de distribuição das amostras para demonstração dos comportamentos dos algoritmos.

2.4.1 Aprendizado Supervisionado

A modelagem por aprendizado supervisionado é feita com um conjunto de dados de treinamento e as soluções desejadas. O algoritmo “aprende” uma função que faz o mapeamento da entrada para a saída. Os algoritmos de aprendizado supervisionado abordados neste trabalho são: **Regressão Linear**, **Floresta Aleatória** e **Perceptron Multicamadas**.

Regressão Linear e Logística

Regressão Linear assume que o relacionamento entre o vetor de valores dos atributos e a solução desejada é aproximadamente linear, com a variável resposta contínua. O modelo linear realiza uma predição calculando a soma ponderada dos atributos de entrada, além de um termo constante chamado intercepto. A predição do modelo é mostrada na equação 2.1:

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2.1)$$

Onde:

- \hat{y} é o valor predito.
- n é o número de atributos.
- x_i é o valor do i -ésimo atributo.
- w_j é o j -ésimo parâmetro do modelo (incluindo o intercepto w_0).

A figura 2.5 mostra um exemplo para $\hat{y} = 4.6 + 2.77x$, onde $x \in \mathbb{R}$, $n = 1$, $w_0 = 4.6$ e $w_1 = 2.77$:

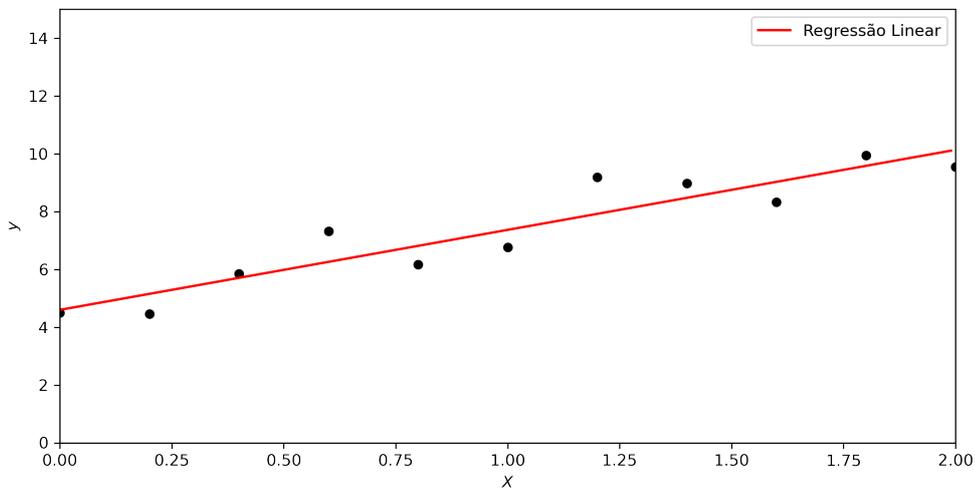


Figura 2.5: Exemplo de *Regressão Linear*.

A Regressão Logística é utilizada para classificação, com a variável resposta categórica. O modelo calcula a soma ponderada dos atributos de entrada (incluindo o termo intercepto), e o resultado passa por uma função sigmoide, que estima a probabilidade da amostra pertencer à uma dada categoria. Se a probabilidade exceder 50%, então retorna Verdadeiro (V), caso contrário retorna Falso (F).

A **complexidade computacional** do treinamento é aproximadamente $O(n^2)$, e da predição é $O(n)$, onde n é o número de atributos (GÉRON, 2019).

Árvore de Decisão e Floresta Aleatória

Árvore de Decisão representa uma função que toma como entrada um vetor de valores de atributos e retorna uma “decisão” — um valor de saída único (RUSSELL e NORVIG, 2009). Os valores de entrada e saída podem ser discretos ou contínuos e a árvore alcança sua decisão executando uma sequência de testes. Cada nó interno na árvore corresponde a um teste do valor de um dos atributos de entrada, e as ramificações dos nós são definidas a partir dos valores possíveis. Cada nó de folha na árvore especifica o valor a ser retornado pela função. A figura 2.6 mostra um exemplo da estrutura de uma Árvore de Decisão:

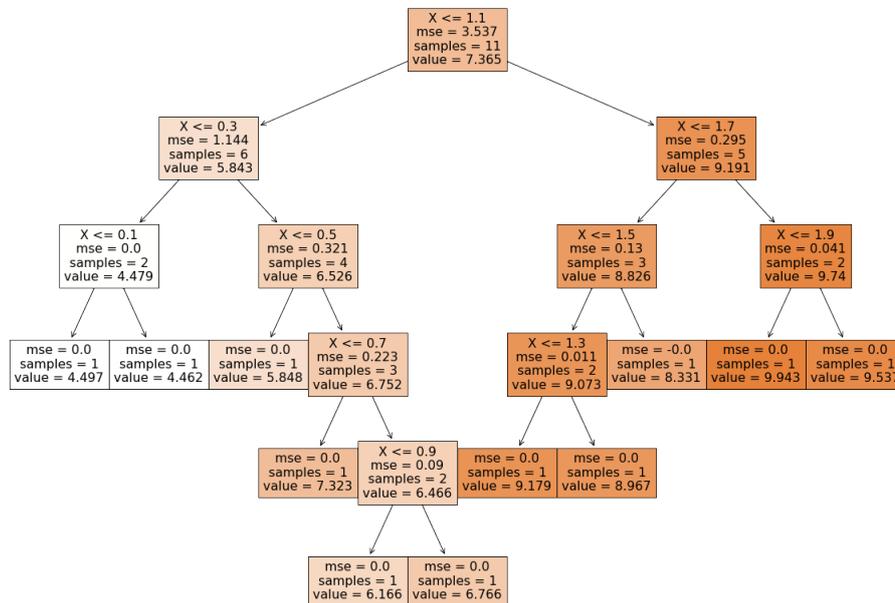


Figura 2.6: Exemplo da estrutura da Árvore de Decisão.

A **complexidade computacional** do treinamento é aproximadamente $O(n \times m \log_2 m)$ e da predição é $O(\log_2 m)$, onde n é o número de atributos e m é o número de amostras (GÉRON, 2019).

Floresta Aleatória é um método de aprendizado em conjunto, que utiliza vários modelos de Árvores de Decisão. Cada modelo é treinado com um conjunto de amostras aleatória (com reposição) dos dados de treinamento, e o resultado é a classe majoritária das predições (no caso de classificação) ou a média das predições de todas as árvores (no caso de regressão). A figura 2.7 mostra um exemplo do resultado de regressão com Árvore de Decisão e da Floresta Aleatória, com 100 árvores.

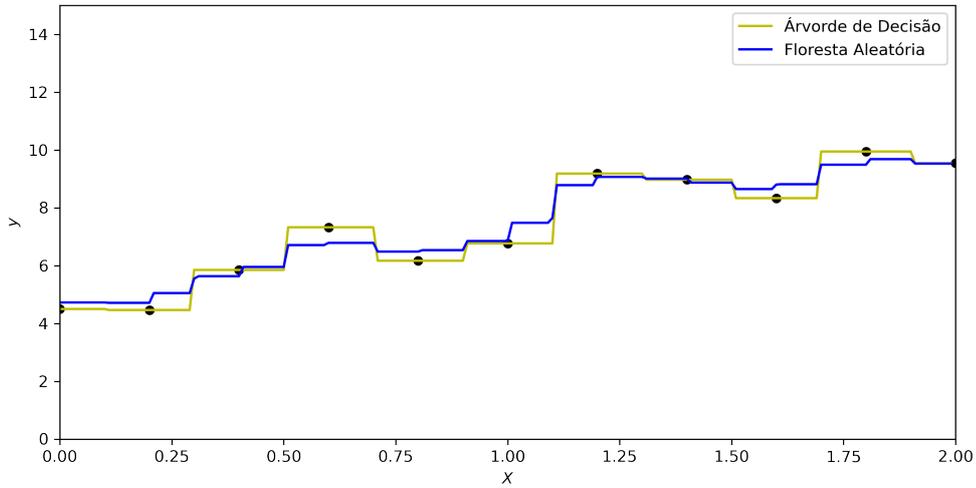


Figura 2.7: Exemplo do resultado de regressão da *Árvore de Decisão* e da *Floresta Aleatória*.

Perceptron e Perceptron Multicamadas

Perceptron é um algoritmo de classificação e também é uma das estruturas mais simples de uma rede neural artificial. O algoritmo calcula a soma ponderada de suas entradas

$$z = w_1x_1 + w_2x_2 + \dots + w_nx_n \tag{2.2}$$

e depois aplica uma função de ativação $f(z)$, que “dispara” quando a soma excede um certo limiar. A figura 2.8a mostra um exemplo da estrutura de um Perceptron com três entradas.

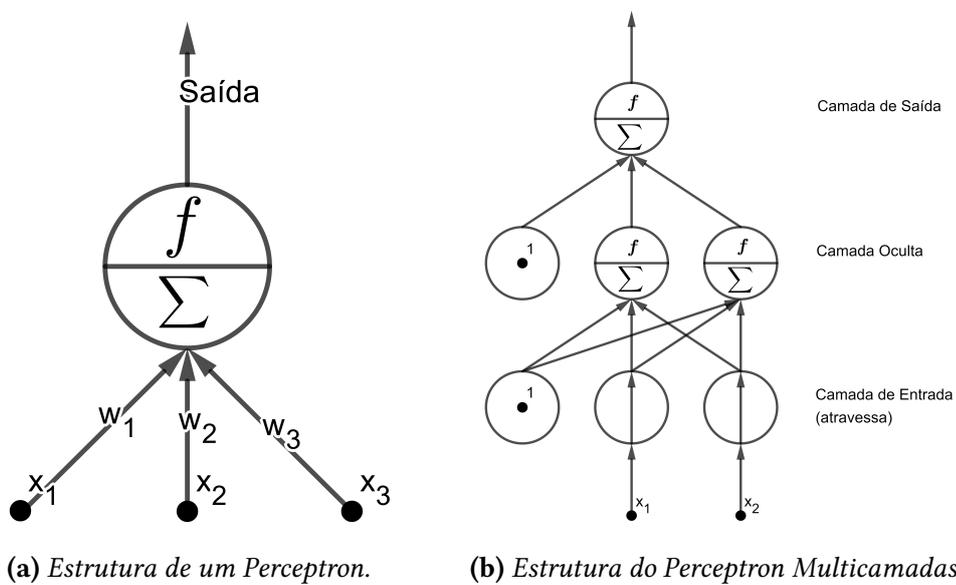


Figura 2.8: Exemplos de *Perceptron* e *Perceptron Multicamadas*.

As funções de ativação de um Perceptron (também chamada de função degrau) mais comuns são:

- Heaviside:

$$f(z) = \begin{cases} 0 & \text{se } z < 0 \\ 1 & \text{se } z \geq 0 \end{cases} \quad (2.3)$$

- Sinal:

$$f(z) = \begin{cases} -1 & \text{se } z < 0 \\ 0 & \text{se } z = 0 \\ 1 & \text{se } z > 0 \end{cases} \quad (2.4)$$

Perceptron Multicamadas é composto por pelo menos 3 camadas de Perceptrons: uma camada de entrada, uma ou mais camadas ocultas (camadas intermediárias) e uma camada de saída. A figura 2.8b mostra a topologia com a camada de entrada com três neurônios (um de viés), uma camada oculta com três neurônios (um de viés) e uma camada de saída com um neurônio. O neurônio de viés $x_0 = 1$ serve para adicionar um intercepto w_0 na equação linear.

As funções de ativação de um Perceptron Multicamadas podem variar para executar regressões ou classificações. As mais comuns são:

- Identidade: retorna o resultado da soma

$$f(z) = z$$

- Sigmoid: retorna um resultado entre 0 e 1

$$f(z) = \text{sigmoide}(z)$$

- Tangente hiperbólica: retorna um resultado entre -1 e 1

$$f(z) = \text{tanh}(z)$$

- Unidade linear retificada: retorna o resultado da soma se for positivo, ou 0 caso contrário

$$f(z) = \text{max}(0, z)$$

O modelo com Perceptron Multicamadas é geralmente treinado com o algoritmo de retro-propagação de erro: para cada amostra de treinamento, o algoritmo faz uma predição (processamento direto) e mede o erro (entre o valor predito e a solução desejada). Em seguida, retorna por cada camada no sentido contrário para medir a contribuição de erro de cada conexão (processamento reverso). Por fim, ajusta levemente os pesos das conexões para reduzir o erro (GÉRON, 2019).

A figura 2.9 mostra o exemplo do resultado da regressão com Perceptron Multicamadas, com uma camada oculta com 100 “neurônios”:

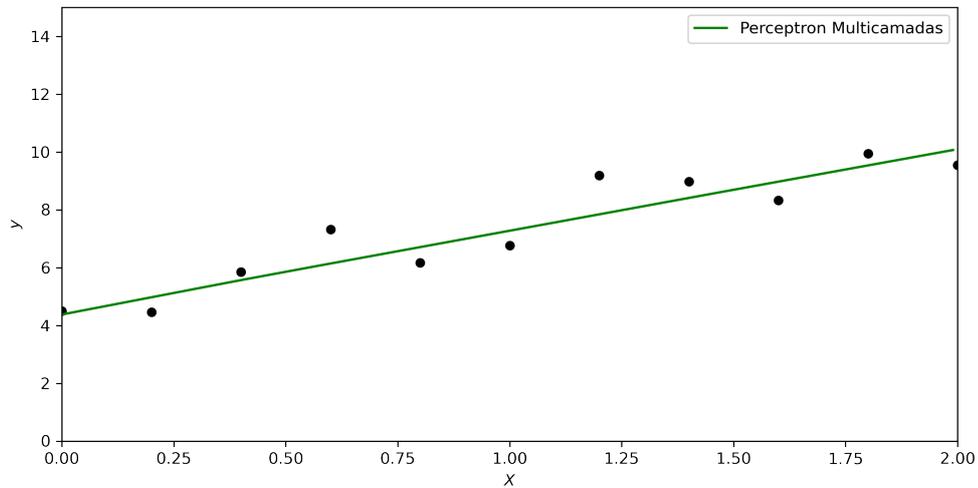


Figura 2.9: Exemplo de Regressão com Perceptron Multicamadas.

A complexidade computacional é $O(n \times m \times h^k \times o \times i)$, onde n é o número de amostras, m é o número de atributos, k é o número de camadas ocultas, cada um contendo h neurônios, o é o número de neurônios de saída e i é o número de iterações de reajuste.

Comparação

A figura 2.10 mostra a comparação dos modelos de regressão, utilizando o mesmos dados de treinamento:

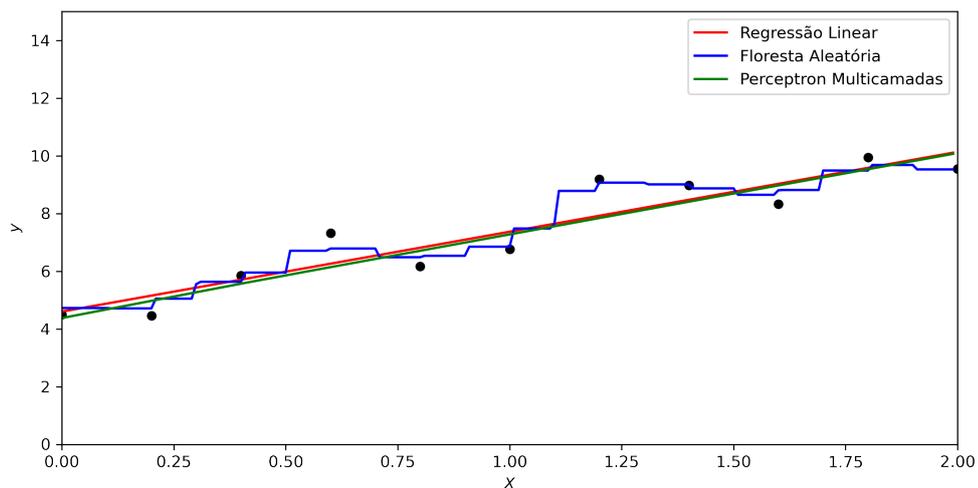


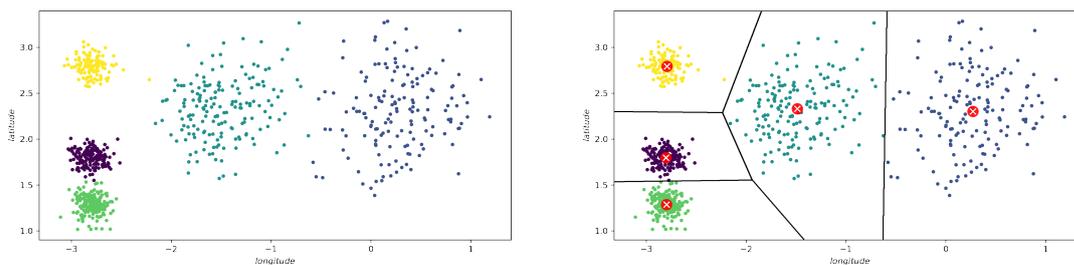
Figura 2.10: Comparação dos modelos de Regressão.

2.4.2 Aprendizado não Supervisionado

A modelagem por aprendizado não supervisionado é feita somente com o conjunto de dados de treinamento, sem as soluções desejadas. Os algoritmos mais importantes são de agrupamento, detecção de anomalias, visualização e redução de dimensões. Algoritmos de agrupamento classificam cada amostra em grupos específicos, de modo que amostras de um mesmo grupo tenham propriedades semelhantes. Neste trabalho é utilizado o algoritmo de agrupamento *K-Means*.

K-Means

K-Means é um algoritmo simples, capaz de agrupar m amostras em k grupos de maneira rápida e eficiente, de modo que cada amostra pertença ao grupo com a média (centróide) mais próxima. O número de grupos k é especificado pelo usuário, como parâmetro. Os agrupamentos possuem formato convexo, definidos a partir dos centróides, resultando em um particionamento do espaço de dados em células *Voronoi* (Figura 2.11).



(a) Partição de 781 amostras em 5 agrupamentos. (b) Visualização das fronteiras de decisão.

Figura 2.11: Exemplo do agrupamento *K-Means* e suas fronteiras de decisão.

O algoritmo é inicializado colocando k centróides aleatoriamente no espaço, ou escolhendo a posição inicial. Assim, classificam-se as amostras (pelo centróide mais próximo), atualizam-se os centróides (a posição da média do grupo), classificam-se as amostras, atualizam-se os centróides, e assim por diante até que os centróides parem de se mover (Figura 2.12). O algoritmo tem garantia de convergir em um número finito de passos, mas pode não convergir para uma solução ótima (pode convergir para um ótimo local), dependendo da posição de inicialização dos centróides (GÉRON, 2019).

A **complexidade computacional** do algoritmo é geralmente linear em relação ao número de amostras m , o número de agrupamentos k e o número de dimensões n . No entanto, isso só é verdade quando os dados têm uma estrutura de agrupamento. Caso contrário, no pior cenário a complexidade pode aumentar exponencialmente com o número de amostras. Na prática, isso raramente acontece e *K-Means* é geralmente um dos algoritmos de agrupamentos mais rápidos (GÉRON, 2019).

David Sculley (SCULLEY, 2010) propôs uma variante do algoritmo, denominada *mini-batch K-Means*, que produz excelentes resultados de agrupamento com baixo custo computacional em grandes conjunto de dados: em vez de utilizar todo o conjunto de dados a

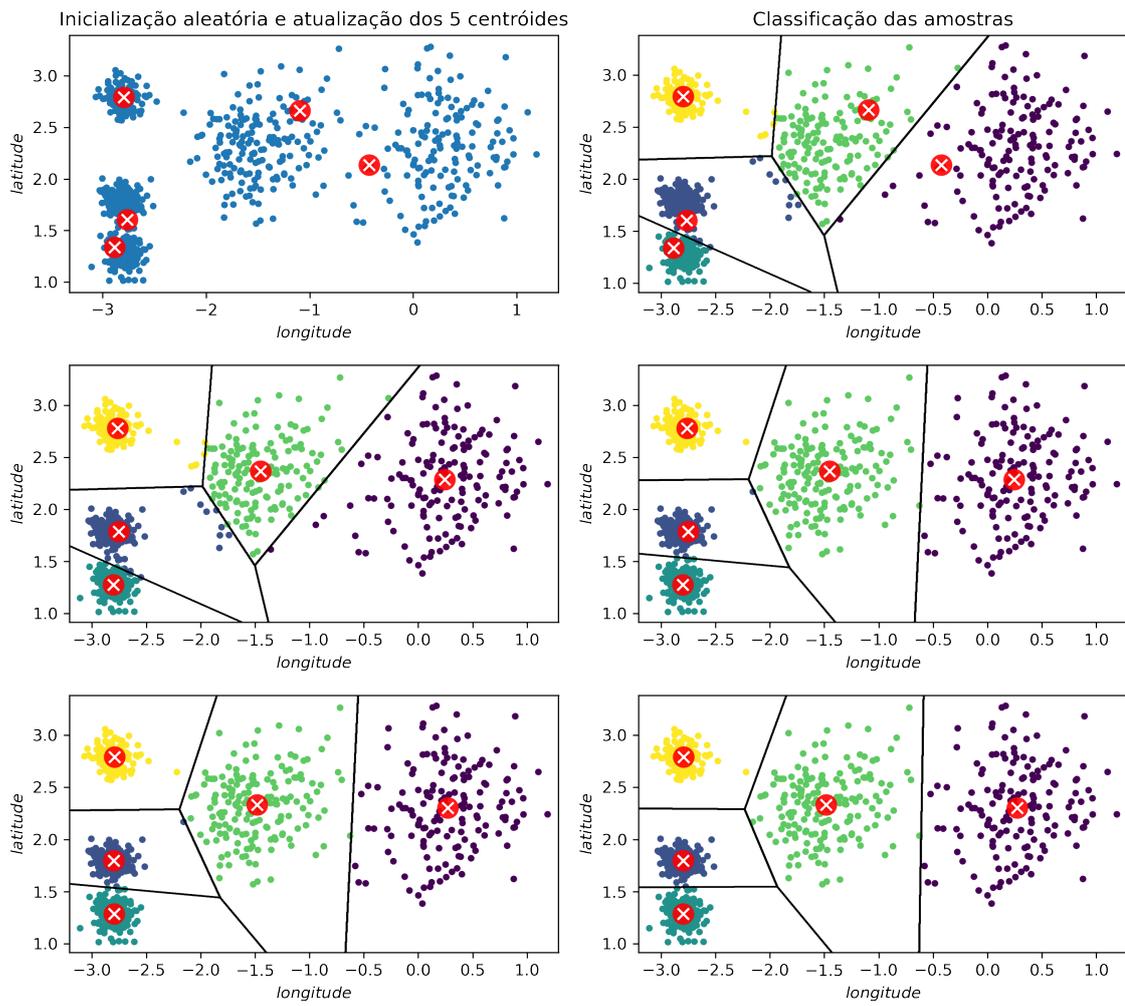


Figura 2.12: Exemplo de funcionamento do algoritmo *K-Means*.

cada iteração, o algoritmo utiliza pequenos lotes, movendo os centróides apenas um pouco a cada iteração. Isso acelera o algoritmo por um fator de 3 ou 4 e torna possível agrupar grandes conjuntos de dados que não cabem na memória (GÉRON, 2019).

2.5 Métricas de Desempenho

Avaliar um modelo de classificação costuma ser significativamente mais complicado do que avaliar um modelo de regressão. A melhor maneira de avaliar o desempenho é através da matriz de confusão. No contexto de um classificador binário (duas classes, positivo e negativo), considere a matriz da figura 2.13:

		Predição	
		N	P
Real	N	VN	FP
	P	FN	VP

Figura 2.13: Matriz de confusão

Onde:

- *VP* é a quantidade de verdadeiros positivos (amostras positivas classificadas corretamente)
- *VN* é a quantidade de verdadeiros negativos (amostras negativas classificadas corretamente)
- *FP* é a quantidade de falsos positivos (amostras negativas classificadas como positivos)
- *FN* é a quantidade de falsos negativos (amostras positivas classificadas como negativos)

As métricas de desempenho abordados neste trabalho são a [Acurácia](#) e [Medida \$F_1\$](#) .

2.5.1 Acurácia

Acurácia é uma métrica de desempenho habitual. É simplesmente a proporção de amostras classificadas corretamente:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

Contudo, no mundo real, frequentemente nossos dados possuem classes desbalanceadas. Na presença de classes desbalanceadas, a acurácia sofre de um paradoxo em que o modelo é altamente acurado, mas carece de poder preditivo. Por exemplo, imagine que estamos tentando prever a presença de um câncer muito raro que ocorre em 0.1% da população. Depois de treinar nosso modelo, descobrimos que a acurácia é de 95%. Porém, 99,9% das pessoas não têm câncer: se simplesmente criássemos um modelo que “previsse” que

ninguém tinha essa forma de câncer, nosso modelo ingênuo seria 4,9% mais acurado, mas claramente não é capaz de prever nada. Por esse motivo, muitas vezes somos motivados a utilizar outras métricas, como precisão, revocação e **medida F_1** (ALBON, 2018).

2.5.2 Medida F_1

Precisão é a proporção de todas as amostras classificadas como positivas que são realmente positivas. Em outras palavras, é a acurácia das predições positivas:

$$\text{Precisão} = \frac{VP}{VP + FP}$$

Revocação é a proporção de todas as amostras positivas que são verdadeiramente positivas. Em outras palavras, é a proporção das classes positivas corretamente detectada pelo classificador:

$$\text{Revocação} = \frac{VP}{VP + FN}$$

Medida F_1 é a média harmônica (um tipo de média utilizada para proporções) da precisão e revocação. É uma medida de correção alcançada na classificação positiva — isto é, das amostras classificadas como positivas, quantas realmente são positivas:

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

Enquanto a média regular trata todos os valores igualmente, a média harmônica dá mais peso à baixos valores. Assim, o classificador só obterá uma medida F_1 alta se a precisão e a revocação forem altas.

Capítulo 3

Trabalhos Relacionados

Este capítulo aborda os trabalhos relacionados, com uma breve descrição dos conceitos, métodos e resultados.

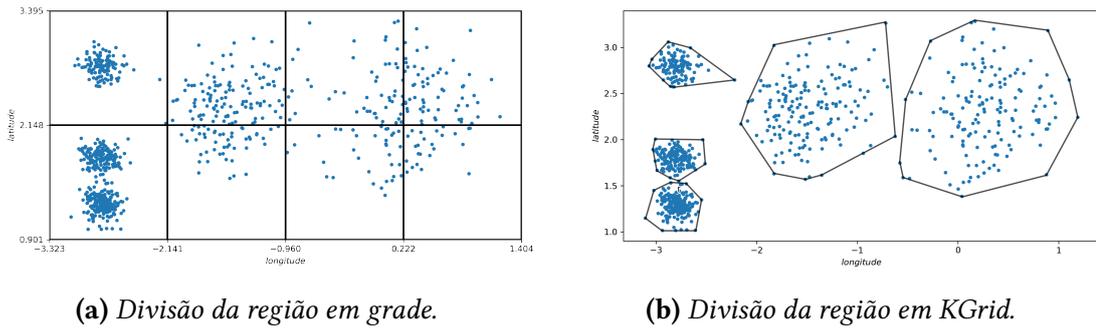
3.1 Geradores e Atratores de Crimes

Os geradores de crimes são locais de fácil acesso ao público. Eles podem se tornar pontos críticos de crime porque a presença de grandes grupos de pessoas cria ocasiões para o crime. Exemplos típicos são centros comerciais, escolas secundárias e estações de transporte público. Alguns lugares oferecem oportunidades muito específicas e se tornam atratores de crimes. São lugares que não reúnem necessariamente grandes grupos de pessoas ao mesmo tempo, mas sua função os torna adequados para que infratores motivados encontrem vítimas ou alvos atraentes e mal protegidos (BERNASCO e BLOCK, 2011).

Geradores e atratores de crimes são dois conceitos semelhantes usados para explicar as concentrações espaciais das atividades criminosas, sugerindo que crimes não ocorrem aleatoriamente no espaço e tempo. Se não são aleatórios, então podem ser estudados e, padrões podem ser modelados. Como consequência, as teorias de criminologia ambiental foram testadas cientificamente e, na última década, vários campos de pesquisa fizeram muito progresso no desenvolvimento de métodos para previsão e avaliação de crimes (espaço-temporais).

3.2 Metodologia

Os trabalhos a seguir estudam os métodos para a previsão de pontos críticos de crimes sobre o espaço e tempo. A região escolhida é dividida em áreas menores (Figura 3.1), e cada área é classificada como um ponto crítico se, em algum momento futuro, o número previsto de crimes ultrapassar um limiar, definido pelo autor.



(a) Divisão da região em grade.

(b) Divisão da região em KGrid.

Figura 3.1: Exemplos de divisões de uma região.

Yu (YU *et al.*, 2011) utilizou dados dos registros de crimes de uma cidade do nordeste dos Estados Unidos e a região foi dividida sobre uma *grade* uniforme (como um tabuleiro de xadrez). O autor realizou seu estudo com duas resoluções (24×20 e 41×40) e teve o cuidado de remover as áreas onde crimes não ocorrem (como pista de aeroportos, represas, etc), pois o uso destas aumentam o desempenho dos classificadores (é fácil para um dado modelo prever que nada irá acontecer onde nada acontece). Uma área é considerada um ponto crítico de crime se há pelo menos um caso de roubo residencial e a distribuição do número de pontos críticos não é balanceado (há mais pontos não-críticos do que pontos críticos). O modelo treinado com redes neurais utilizando dados de 9 meses obteve o melhor resultado, com *acurácia* e *medida F_1* próximo a 91% e 75% respectivamente. Os melhores resultados foram obtidos com a grade de menor resolução (um total de 480 áreas).

Bogomolov (BOGOMOLOV *et al.*, 2014) utilizou dados de comportamento humano, derivado da combinação da atividade de redes móveis e informações demográficas, em conjunto com dados abertos de crimes de uma região metropolitana de Londres. A região também foi dividida em forma de *grade* (um total de 124119 áreas) e cada área é considerada um ponto crítico de crime se o número de crimes previsto em um momento futuro é maior do que a mediana dos números de crimes. Assim, o número de pontos críticos e pontos não-críticos é aproximadamente balanceado. O modelo treinado com o algoritmo *Florestas Aleatórias* obteve o melhor resultado, com *acurácia* próximo a 70% e *medida F_1* em torno de 67% dos pontos críticos de crime.

Araújo (ARAÚJO *et al.*, 2018) utilizou dados dos registros de crimes da cidade de Natal, no Rio Grande do Norte, obtido pela parceria entre o projeto de pesquisa *Smart Metropolis* e a Secretaria da Segurança Pública do Rio Grande do Norte. A região foi dividida em polígonos convexos baseado em *K-Means*, denominado *KGrid*, ajustados com as coordenadas das ocorrências. Cada área poligonal é considerado um ponto crítico se a previsão da quantidade de crimes for maior do que a média das quatro últimas observações. A distribuição dos números de pontos críticos não é balanceado. O modelo treinado com *Perceptron Multicamadas* obteve o melhor resultado, com *acurácia* próximo a 80%, com um total de 70 áreas.

Capítulo 4

Metodologia

Este capítulo abordará os métodos utilizados neste trabalho para a previsão dos pontos críticos de crime. A seção 4.1 apresenta uma visão geral da metodologia e as seções seguintes detalham cada método utilizado.

4.1 Visão Geral

O objetivo é criar e avaliar um modelo capaz de prever se uma área específica de uma região será um ponto crítico de crime no mês m utilizando os dados de n meses anteriores a m . Para isso, é proposto os seguintes passos:

1. Os dados dos n meses anteriores serão os dados de treinamento (para a criação do modelo) e os dados de m serão os dados de teste (para a avaliação do modelo). Para simplificar a explicação, suponha que desejamos prever as áreas de uma região que serão pontos críticos de crime para o mês de Abril (considere que possuímos os dados deste mês), com os dados de 3 meses anteriores (Janeiro, Fevereiro e Março). A figura 4.1 mostra um exemplo:

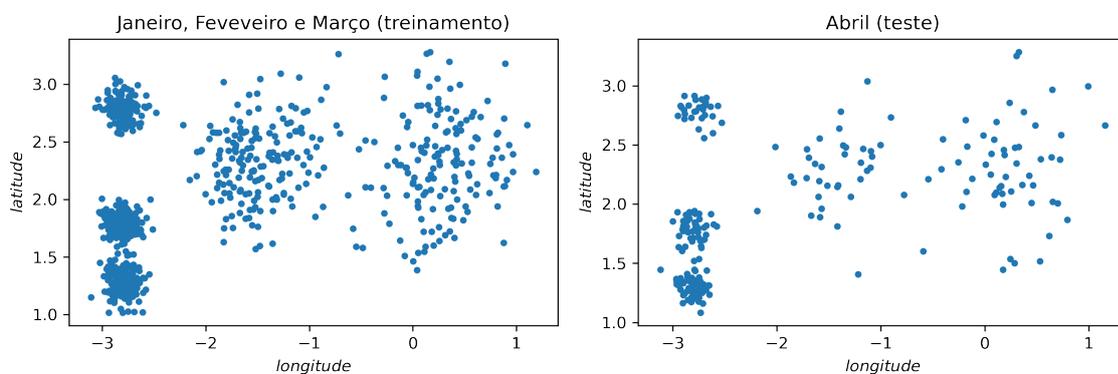


Figura 4.1: Exemplo de distribuição dos dados de treinamento e teste.

2. Criação de um modelo de agrupamento utilizando o algoritmo de **aprendizado não supervisionado *K-Means***. Este modelo será treinado para dividir as amostras dos dados de treinamento em k grupos, e todas elas serão assimiladas ao seu respectivo grupo. Todas as amostras dos dados de teste também serão assimilados de acordo com o resultado do modelo. A figura 4.2 mostra um exemplo da divisão dos dados em 5 grupos:

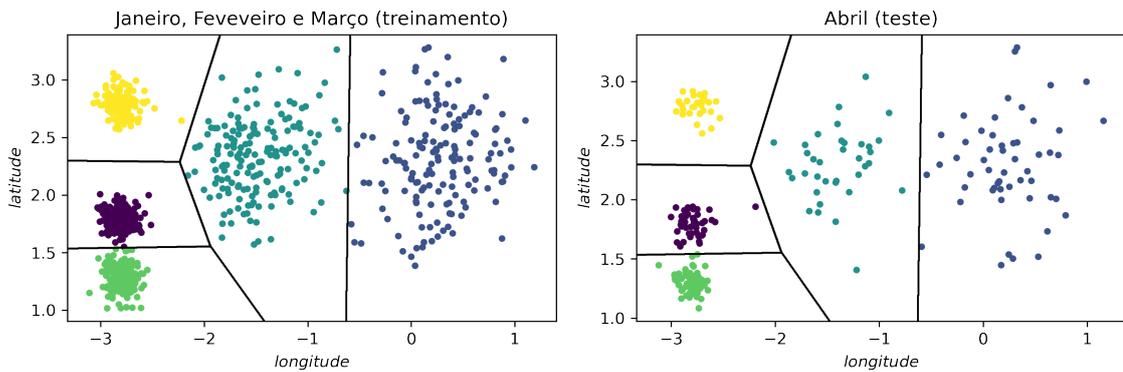


Figura 4.2: Exemplo de agrupamento dos dados de treinamento e teste.

3. Criação de uma **série temporal** com o número de crimes mensais de cada grupo. A figura 4.3 mostra os números de crimes mensais dos 5 grupos, para os dados de treinamento e teste:

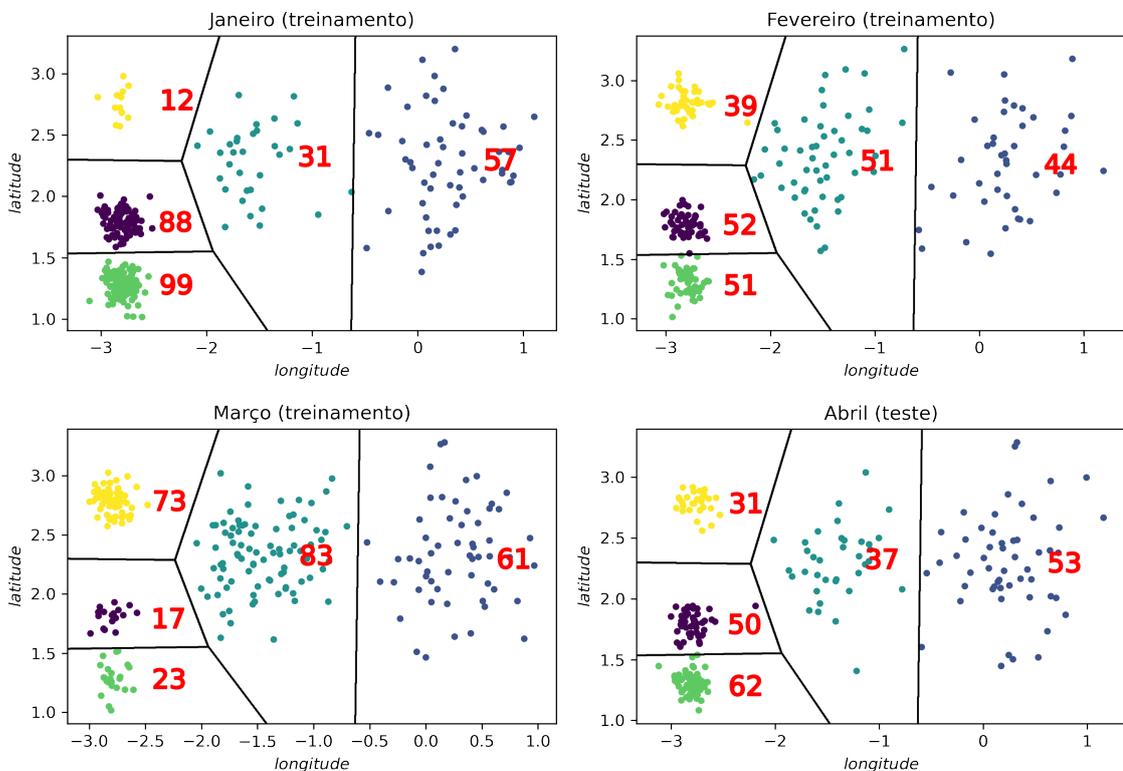


Figura 4.3: Exemplo de criação de séries temporais com os dados de treinamento e teste.

4. Criação de um modelo de regressão utilizando algoritmo de **aprendizado supervisionado** com os novos dados de treinamento (da série temporal), e previsões dos números de crimes para o mês m para todos os grupos. A figura 4.4 mostra os números da previsão, utilizando o modelo com **Regressão Linear**, e os números verdadeiros, dos dados de teste, para o mês de Abril:

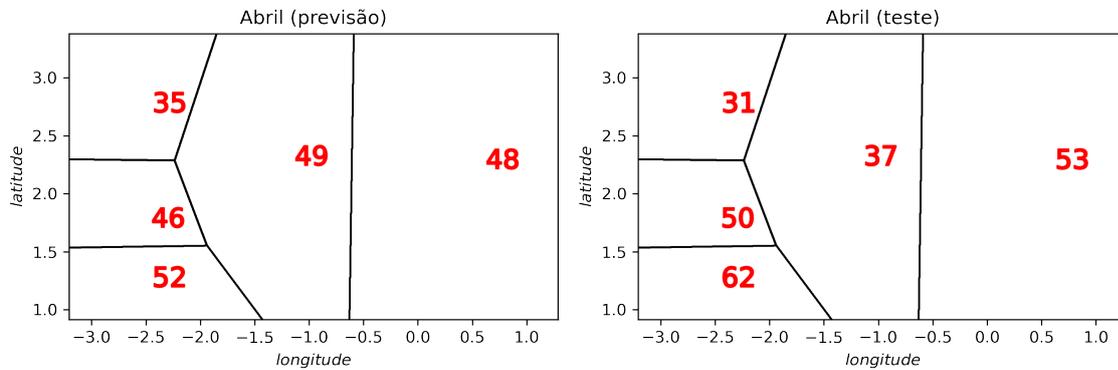


Figura 4.4: Exemplo dos números da previsão e teste.

5. Classificação dos resultados. Cada grupo será considerado um ponto crítico se o número previsto for maior do que o limiar escolhido. Os dados de teste também serão classificados desta maneira. A figura 4.5 mostra o exemplo de classificação, considerando o limiar como a mediana dos números de crimes dos dados de treinamento:

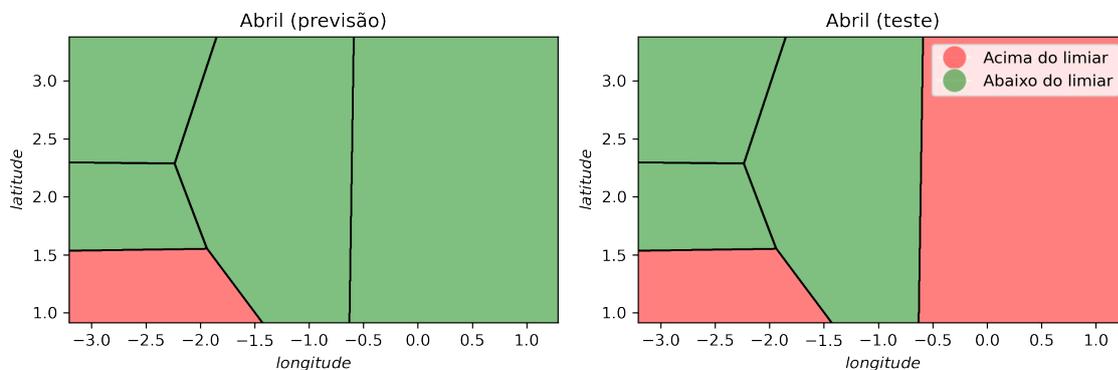


Figura 4.5: Exemplo dos números da previsão e teste.

6. Avaliação do modelo, comparando os resultados da previsão com os verdadeiros, utilizando a **acurácia** e **medida F_1** . A tabela 4.1 mostra o resultado da avaliação do modelo de exemplo:

Acurácia	Medida F_1
80%	78%

Tabela 4.1: Avaliação do modelo de previsão usando acurácia e medida F_1 .

A metodologia pode ser visualizado na figura 4.6:

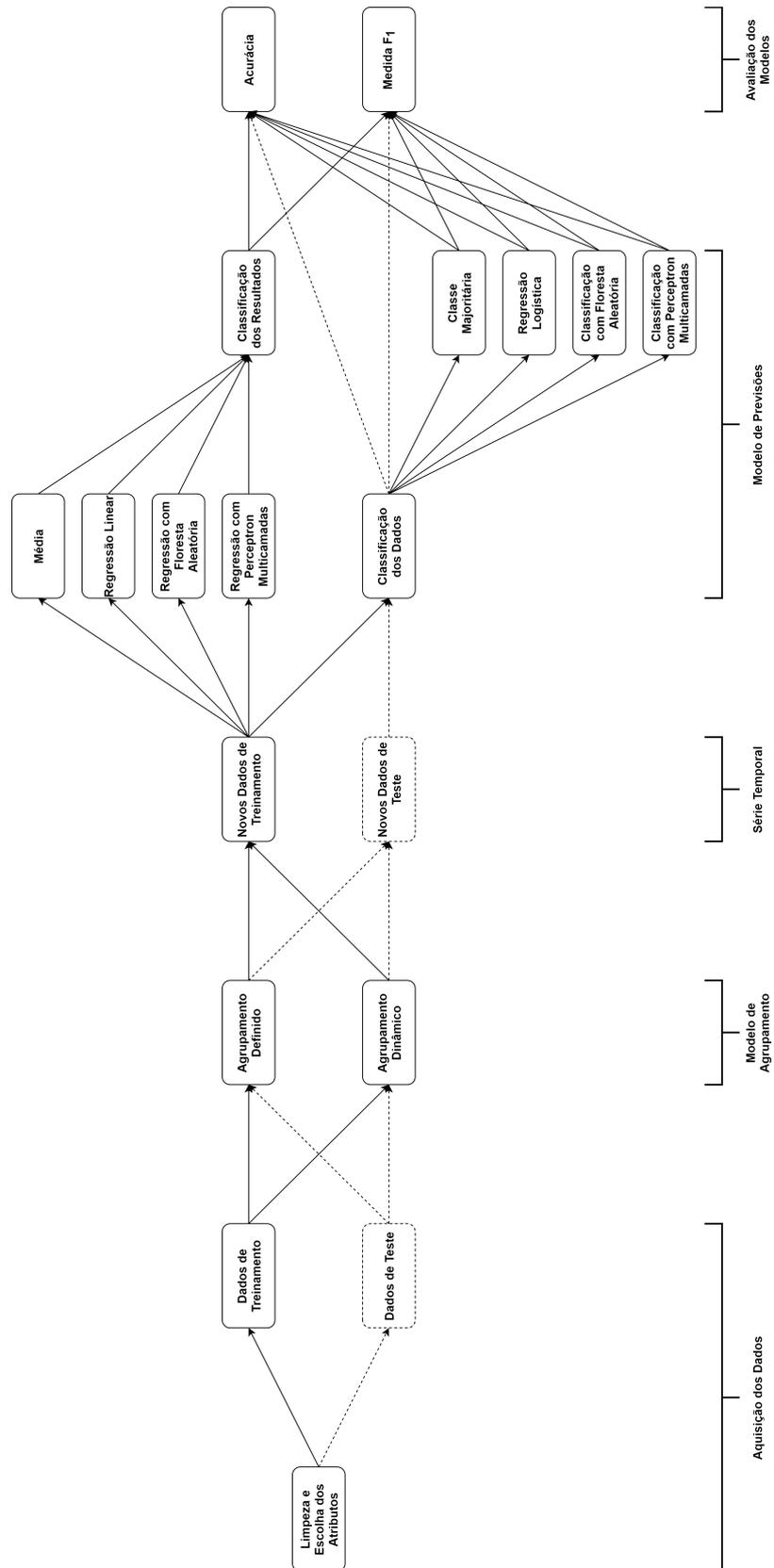


Figura 4.6: Visão Geral da metodologia.

4.2 Aquisição dos Dados

A SSP-SP disponibiliza no seu portal da transparência os dados dos boletins de ocorrências de diversos tipos de roubos (SSP-SP, 2020). Este trabalho irá focar especificamente em roubo de celulares no Estado de São Paulo.

A figura 4.7 mostra a proporção de informações disponíveis para cada atributo de todas as amostras (crimes) entre Janeiro de 2018 a Dezembro de 2019 (24 meses). É possível verificar que há atributos com pouca ou nenhuma informação, e como não podemos trabalhar com dados de atributos ausentes, normalmente temos três opções:

1. Remover as amostras com dados de atributos ausentes.
2. Preencher os dados ausentes dos atributos (com zeros, média, mediana, etc.).
3. Remover os atributos com pouca ou nenhuma informação.

A primeira opção não é possível, pois há atributos sem nenhuma informação. A segunda opção também não é possível, pois muitos atributos possuem valor qualitativo (como cor, nome, sexo, etc.) e não quantitativo. Logo, seguimos com a terceira opção, mantendo os atributos que possuem pelo menos 80% de informações. A tabela 4.2 mostra quais são esses atributos, com um exemplo.

Visualizando os dados, é possível entender o tipo de informação que contém cada atributo. As informações sobre a Delegacia responsável, instantes da criação, emissão e número dos boletins de ocorrência são irrelevantes para o nosso modelo, assim como a data de comunicação. Dentre os atributos, iremos escolher LATITUDE e LONGITUDE para agrupamento *K-Means* e BAIRRO e CIDADE para uma proposta de inicialização dos centróides.

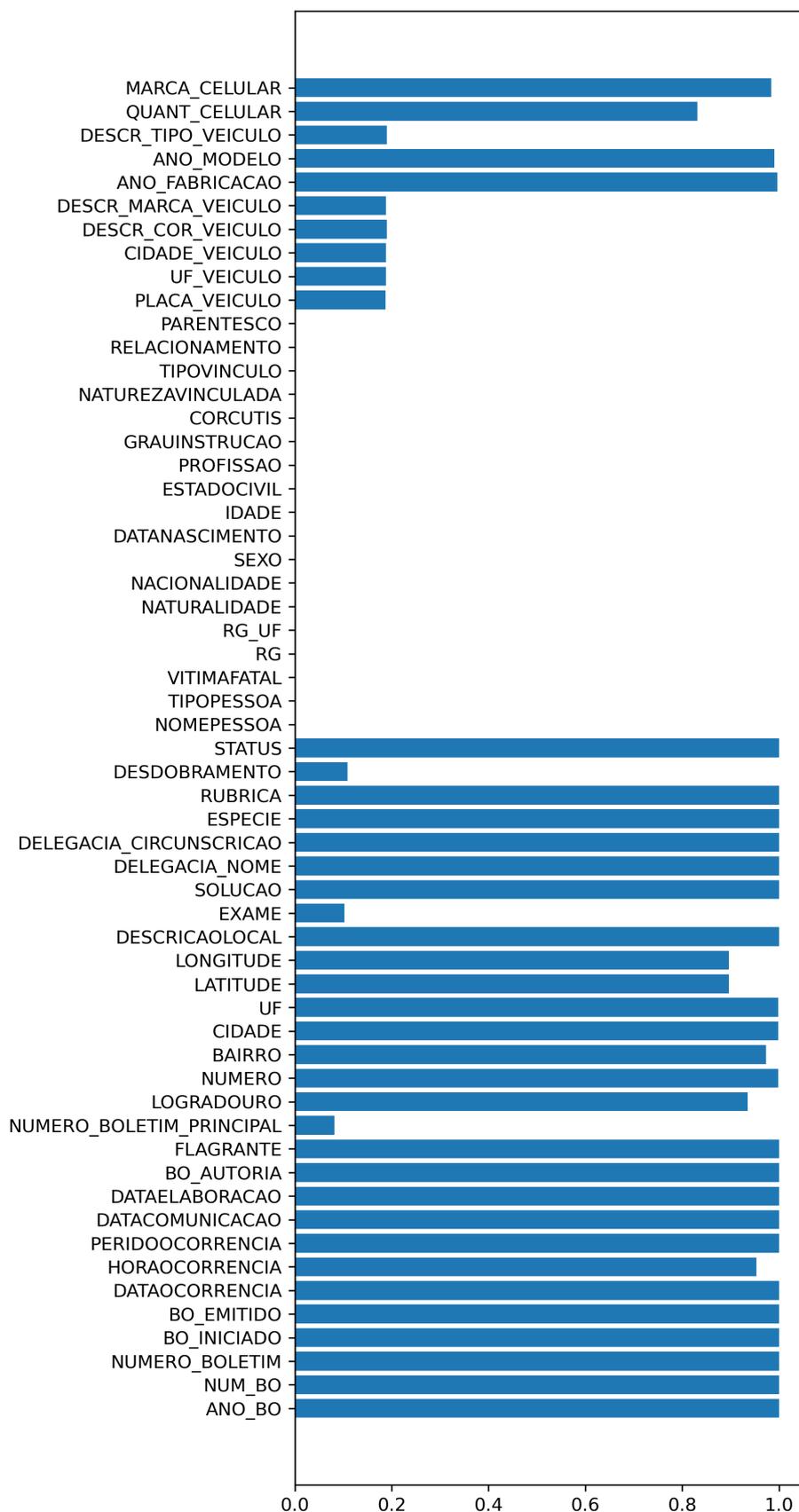


Figura 4.7: Proporção de dados disponíveis para cada atributo.

Atributo	Dado
ANO_BO	2017
NUM_BO	4233
NUMERO_BOLETIM	4233/2017
BO_INICIADO	31/12/2017 23:57:15
BO_EMITIDO	01/01/2018 00:41:28
DATAOCORRENCIA	31/12/2017
HORAOCORRENCIA	21:40
PERIDOOCCORRENCIA	A NOITE
DATACOMUNICACAO	31/12/2017
DATAELABORACAO	31/12/2017 23:57:15
BO_AUTORIA	Desconhecida
FLAGRANTE	Não
LOGRADOURO	RUA CONCHAL
NUMERO	30.0
BAIRRO	PARQUE NOVA VENEZA
CIDADE	SUMARE
UF	SP
LATITUDE	-22.812321
LONGITUDE	-47.232128
DESCRICAOLocal	Via pública
SOLUCAO	ENCAMINHAMENTO DP ÁREA DO FATO
DELEGACIA_NOME	DEL.POL.PLANTÃO SUMARE
DELEGACIA_CIRCUNSCRICAO	03° D.P. SUMARE
ESPECIE	Título II - Patrimônio (arts. 155 a 183)
RUBRICA	Roubo (art. 157) - VEICULO
STATUS	Consumado
ANO_FABRICACAO	2014.0
ANO_MODELO	2015.0
QUANT_CELULAR	1.0
MARCA_CELULAR	Samsung

Tabela 4.2: Atributos com mais informações e exemplo de uma amostra.

4.3 Modelo de Agrupamento

Utilizaremos um algoritmo de agrupamento para capturar as **regiões geradores e atratores de crimes**, que podem se tornar pontos críticos de crimes, discutido no capítulo 3. Araújo (ARAÚJO *et al.*, 2018) utilizou um método semelhante em seu trabalho.

A escolha do algoritmo *K-Means* para o modelo de agrupamento é devido à eficiência e capacidade de trabalhar com grande volume de amostras. As fronteiras de decisões podem ser visualizadas criando um **diagrama de Voronoi** a partir dos centróides resultantes.

O algoritmo *K-Means* requer que o usuário especifique o número de agrupamentos k (que equivale ao número de centróides) e, opcionalmente, a posição inicial dos centróides. Se o último não for especificado, o algoritmo pode definir a posição inicial dos centróides aleatoriamente. Portanto, são propostas duas abordagens para agrupamentos das amostras:

- **Número de agrupamentos definido** (posição inicial dos centróides aleatória).
- **Número de agrupamentos dinâmico** (número e posição inicial dos centróides definida pelos bairros afetados pelos crimes).

Para cada abordagem, serão realizados treinamentos do modelo com m (mês da previsão) entre 1 a 12 (Janeiro de 2019 a Dezembro de 2019) e n (número de meses anteriores) entre de 1 a 12 (de um mês a um ano).

Para facilitar a explicação nos tópicos seguintes, suponha que os dados da figura 4.8 representem a distribuição de crimes de uma cidade fictícia, composta por 8 bairros quadrados.

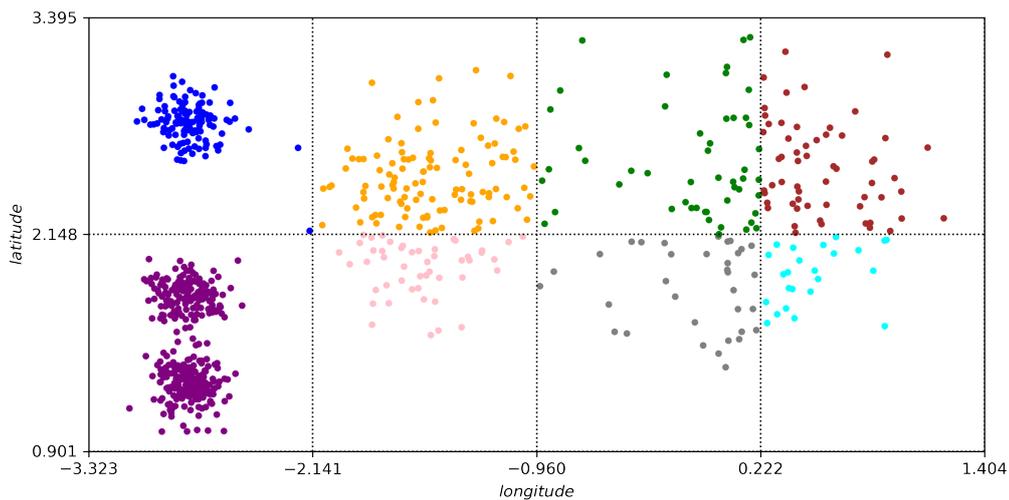


Figura 4.8: Exemplo de distribuição de crimes em uma cidade fictícia, composta por 8 bairros quadrados.

4.3.1 Número de agrupamentos definido

O número de áreas possui relação com o desempenho das previsões, como observado nos trabalhos de Yu (Yu *et al.*, 2011) e Araújo (ARAÚJO *et al.*, 2018) no capítulo 3. Para verificar essa relação realizamos os agrupamentos para valores definidos de k , variando entre 500 a 5000 com incrementos de 500 (500, 1000, 1500, ..., 5000). Todos os k centróides são iniciados aleatoriamente. A figura 4.9 mostra um exemplo de agrupamento com 8 centróides.

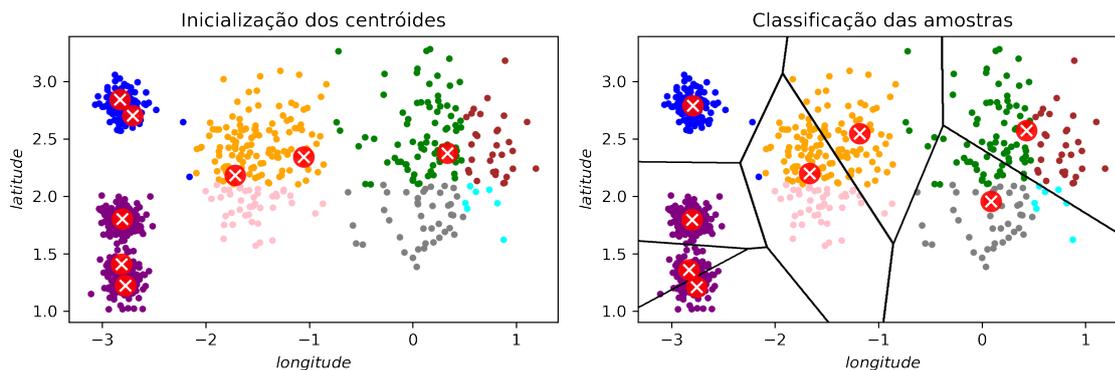


Figura 4.9: Exemplo de agrupamento com 8 centróides, inicializados aleatoriamente.

4.3.2 Número de agrupamentos dinâmico

O número de agrupamentos será definido pelo número de bairros afetados para cada n (número de meses anteriores). Por exemplo, se $m = 4$ (previsão para Abril) e $n = 3$ (dados de 3 meses anteriores), então o número de agrupamentos será o número de bairros afetados com pelo menos um crime, entre Janeiro a Março. A posição inicial dos centróides será a posição média dos crimes de cada bairro. A figura 4.10 mostra um exemplo de agrupamento com 8 centróides (número de bairros afetados), iniciados a partir da posição média das amostras de cada bairro.

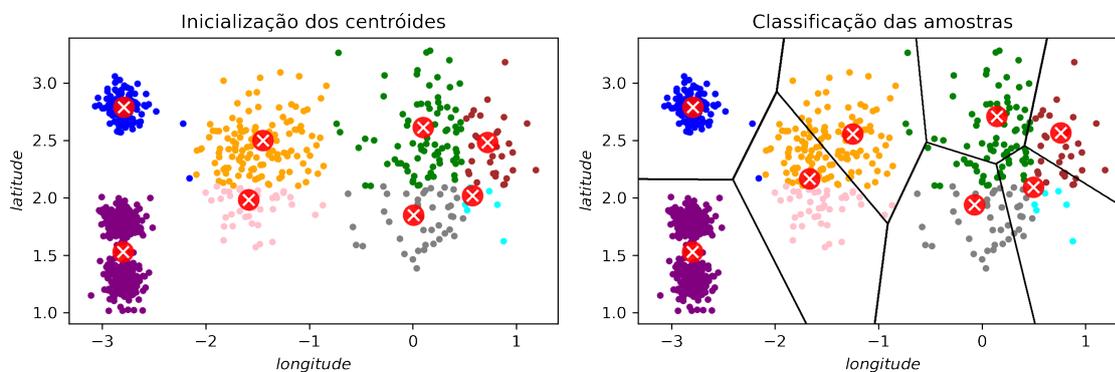


Figura 4.10: Exemplo de agrupamento com 8 centróides, inicializados pela posição média das amostras de cada bairro.

4.4 Série Temporal

Os novos dados de treinamento e teste serão criados como visto no seção 4.1: a partir de uma *série temporal* com o número de crimes mensais de cada grupo. A tabela 4.3 mostra os novos dados de treinamento e teste para os modelos de regressão de acordo com o exemplo da figura 4.3.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Mês	Crimes
1	0	0	0	0	1	12
1	0	0	0	0	2	39
1	0	0	0	0	3	73
0	1	0	0	0	1	88
0	1	0	0	0	2	52
0	1	0	0	0	3	17
0	0	1	0	0	1	99
0	0	1	0	0	2	51
0	0	1	0	0	3	23
0	0	0	1	0	1	31
0	0	0	1	0	2	51
0	0	0	1	0	3	83
0	0	0	0	1	1	57
0	0	0	0	1	2	44
0	0	0	0	1	3	61

(a) Exemplo de dados de treinamento.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Mês	Crimes
1	0	0	0	0	4	31
0	1	0	0	0	4	50
0	0	1	0	0	4	62
0	0	0	1	0	4	37
0	0	0	0	1	4	53

(b) Exemplo de dados de teste.

Tabela 4.3: Exemplo de dados de treinamento e teste.

O limiar para a classificação dos pontos críticos de crime será definida pelo valor da mediana dos números de crime dos dados de treinamento. A escolha da mediana como limiar resulta na distribuição aproximadamente balanceada das classes, assim como no trabalho de Bogomolov (BOGOMOLOV *et al.*, 2014) apresentada no capítulo 3, a fim de evitar o paradoxo da acurácia.

Para treinar os modelos de classificação, é necessário transformar a coluna da solução desejada (dos número de crimes) em uma classe binária (ponto crítico de crime) nos dados de treinamento e teste. Para isso, definimos o novo atributo como V (Verdadeiro) se o número de crimes for maior do que o limiar, ou F (Falso) caso contrário. Supondo que a tabela 4.3 represente os dados de treinamento e teste para o modelo de regressão, a tabela 4.4 representa os dados de treinamento e teste para o modelo de classificação.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Mês	Crítico
1	0	0	0	0	1	F
1	0	0	0	0	2	F
1	0	0	0	0	3	V
0	1	0	0	0	1	V
0	1	0	0	0	2	V
0	1	0	0	0	3	F
0	0	1	0	0	1	V
0	0	1	0	0	2	F
0	0	1	0	0	3	F
0	0	0	1	0	1	F
0	0	0	1	0	2	F
0	0	0	1	0	3	V
0	0	0	0	1	1	V
0	0	0	0	1	2	F
0	0	0	0	1	3	V

(a) Exemplo de dados de treinamento.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Mês	Crítico
1	0	0	0	0	4	F
0	1	0	0	0	4	F
0	0	1	0	0	4	V
0	0	0	1	0	4	F
0	0	0	0	1	4	V

(b) Exemplo de dados de teste.

Tabela 4.4: Exemplo dados de treinamento e teste para modelo de classificação.

4.5 Modelos de Previsões

Os modelos de regressão e classificação serão treinados com os dados criados conforme a seção 4.4. As previsões dos modelos de classificação serão os resultados finais. Para os modelos de regressão, será necessário transformar o resultado da previsão (Tabela 4.5) em uma classe binária, definindo a nova variável como V (Verdadeiro) se a previsão for maior do que o limiar, ou F (Falso) caso contrário (Figura 4.11).

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Mês	Crimes
1	0	0	0	0	4	35.33
0	1	0	0	0	4	46.33
0	0	1	0	0	4	51.67
0	0	0	1	0	4	49.00
0	0	0	0	1	4	48.00

Tabela 4.5: Exemplo de previsão dos números de crimes.

Grupo	Real	Previsão		Grupo	Real	Previsão
1	31	35.33	Classificação → Acima do Limiar	1	F	F
2	50	46.33		2	F	F
3	62	51.67		3	V	V
4	37	49.00		4	F	F
5	53	48.00		5	V	F

Figura 4.11: Exemplo de classificação dos resultados.

Os modelos de previsões utilizados nos experimentos são: Modelo Fictício, [Regressão Linear](#) e [Logística](#), [Floresta Aleatória](#) e [Perceptron Multicamadas](#).

4.5.1 Modelo Fictício

Em seu trabalho, Bogomolov ([BOGOMOLOV et al., 2014](#)) incluiu nos experimentos um classificador fictício, que sempre retorna a classe majoritária. Tal classificador serve como linha de base, a fim de comparar os resultados e verificar a habilidade preditiva dos modelos. Neste trabalho serão utilizados dois modelos fictícios, um para regressão e outro para classificação, com as seguintes regras:

- **Regressão:** realiza a predição retornando apenas a média das soluções desejadas dos dados de treinamento. Supondo que o modelo utilize os dados de treinamento da tabela 4.3a, a predição será sempre 52.07 (média dos crimes).
- **Classificação:** realiza a predição retornando a classe majoritária das soluções desejadas dos dados de treinamento. Supondo que o modelo utilize os dados de treinamento da tabela 4.4a, a predição será sempre F (classe majoritária).

4.5.2 Modelo Linear

Trata-se de um dos modelos mais simples, que apresenta bom desempenho e baixa complexidade computacional. Neste trabalho é utilizado o algoritmo de Regressão Linear para o modelo de regressão, e o algoritmo de Regressão Logística para o modelo de classificação. Ambos os modelos irão utilizar o intercepto (termo constante) na equação linear.

4.5.3 Floresta Aleatória

Em seu trabalho, Bogomolov ([BOGOMOLOV *et al.*, 2014](#)) obteve o melhor desempenho em seu modelo utilizando o algoritmo de Floresta Aleatória, porém não especificou o número de estimadores. Neste trabalho, para os modelos de regressão e classificação, serão utilizados 100 estimadores (árvores de decisão).

4.5.4 Perceptron Multicamadas

Yu ([YU *et al.*, 2011](#)) obteve melhor resultado por meio de Redes Neurais com duas camadas, mas não especificou o nome do algoritmo e os parâmetros (como número de neurônios utilizados por camada). Supomos que o algoritmo utilizado é o Perceptron Multicamadas. Araújo ([ARAÚJO *et al.*, 2018](#)) também obteve melhor desempenho em seu modelo utilizando o algoritmo de Perceptron Multicamadas, e também não especificou os parâmetros. Neste trabalho, para os modelos de regressão e classificação, serão utilizados uma camada oculta com 100 neurônios, e a função de ativação utilizada para esta camada será a unidade linear retificada.

4.6 Avaliação dos Modelos

Os resultados das previsões dos modelos de regressão e classificação serão avaliados por meio da **acurácia** e **medida F_1** . A figura 4.12 mostra o exemplo da avaliação:

Grupo	Real	Previsão
1	F	F
2	F	F
3	V	V
4	F	F
5	V	F

$\xrightarrow{\text{Avaliação}}$
 $\left\{ \begin{array}{l} 80\% \text{ Acurácia} \\ 78\% \text{ Medida } F_1 \end{array} \right.$

Figura 4.12: Exemplo de avaliação dos resultados.

Para o cálculo da acurácia e da medida F_1 do exemplo da figura 4.12, considere as **matrizes de confusão** resultantes (Figura 4.13):

		Predição					
		N	P			N	P
Real	N	3	0	Real	N	1	1
	P	1	1		P	0	3
$\xrightarrow{\text{Onde}}$		$\left\{ \begin{array}{l} VP = 1 \\ VN = 3 \\ FP = 0 \\ FN = 1 \end{array} \right.$		$\xrightarrow{\text{Onde}}$		$\left\{ \begin{array}{l} VP = 3 \\ VN = 1 \\ FP = 1 \\ FN = 0 \end{array} \right.$	
(a) Classe V como classe positiva				(b) Classe F como classe positiva			

Figura 4.13: Matrizes de confusão

4.6.1 Acurácia

A acurácia é a proporção de amostras classificadas corretamente. Considerando as matrizes de confusão da figura 4.13 como exemplo, temos:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{3 + 1}{3 + 1 + 1 + 0} = \frac{4}{5} = 80\%$$

4.6.2 Medida F_1

Neste trabalho, é utilizado a medida F_1 ponderada, onde é calculado a medida F_1 para cada classe e então a média ponderada por suporte (proporção de instâncias verdadeiras para cada classe). Considerando as matrizes de confusão da figura 4.13 como exemplo, primeiro calculamos a medida F_1 da classe V (consideramos a classe como positiva):

$$\text{Precisão} = \frac{VP}{VP + FP} = \frac{1}{1 + 0} = 1$$

$$\text{Revocação} = \frac{VP}{VP + FN} = \frac{1}{1 + 1} = \frac{1}{2}$$

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} = 2 \times \frac{1 \times \frac{1}{2}}{1 + \frac{1}{2}} = \frac{2}{3} \approx 67\%$$

Depois, calculamos a medida F_1 da classe F (consideramos a classe como positiva):

$$\text{Precisão} = \frac{VP}{VP + FP} = \frac{3}{3 + 1} = \frac{3}{4}$$

$$\text{Revocação} = \frac{VP}{VP + FN} = \frac{3}{3 + 0} = 1$$

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} = 2 \times \frac{\frac{3}{4} \times 1}{\frac{3}{4} + 1} = \frac{6}{7} \approx 86\%$$

A proporção da classe V é 40% e da classe F é 60%. Logo, a medida F_1 ponderada é:

$$F_1 = 40\% \times 67\% + 60\% \times 86\% \approx 78\%$$

Capítulo 5

Resultados dos Experimentos

Este capítulo apresenta e compara todos os resultados dos experimentos, e é composto pelas seguintes seções:

1. Regressores (agrupamento definido)
2. Classificadores (agrupamento definido)
3. Melhor abordagem (agrupamento definido)
4. Regressores (agrupamento dinâmico)
5. Classificadores (agrupamento dinâmico)
6. Melhor abordagem (agrupamento dinâmico)

As três primeiras seções abordam os resultados dos experimentos com o modelo de agrupamento com o **número de agrupamento definido**, e as três últimas seções abordam os resultados dos experimentos com o modelo de agrupamento com o **número de agrupamento dinâmico**. Para cada modelo de agrupamento, temos as comparações dos modelos de previsões de:

- Regressores: comparação dos resultados dos experimentos entre Modelo Fictício, Regressão Linear, Regressão com Floresta Aleatória e Regressão com Perceptron Multicamadas.
- Classificadores: comparação dos resultados dos experimentos entre Modelo Fictício, Regressão Logística, Classificação com Floresta Aleatória e Classificação com Perceptron Multicamadas.
- Melhor abordagem: comparação dos resultados do melhor regressor e melhor classificador.

Cada seção é composta pelas seguintes avaliações de desempenho:

- Acurácia: proporção de amostras classificadas corretamente (maiores valores são melhores).
- Medida F_1 : média harmônica da precisão e revocação (maiores valores são melhores).

- Complexidade Computacional: valores representam o tempo de treinamento em segundos (menores valores são melhores).

Para cada avaliação de desempenho, é apresentado as seguintes medidas:

- Pior caso: representa o menor resultado encontrado na previsão de algum mês do ano de 2019, para a acurácia e medida F_1 , e o maior valor para o item Complexidade Computacional. O objetivo é identificar o pior resultado possível para a configuração do modelo.
- Média: representa o resultado médio do ano, com o objetivo de medir a qualidade geral do modelo. Médias altas para acurácia e medida F_1 e médias baixas para a complexidade computacional, para uma dada configuração, indicam que o modelo possui um bom desempenho.
- Desvio Padrão: representa a estabilidade do modelo. Baixos valores indicam que o modelo é capaz de manter o mesmo desempenho durante todo o ano.

A figura 5.1 mostra o diagrama dos testes e resultados dos experimentos:

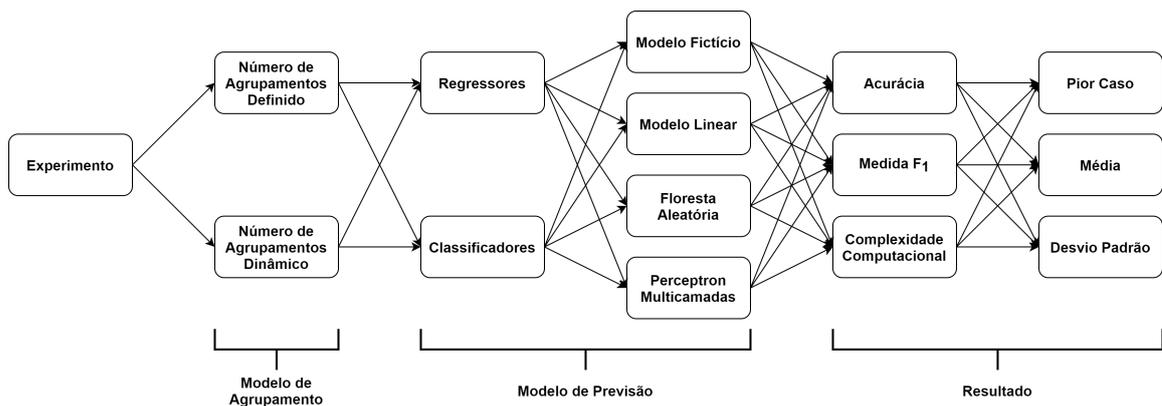


Figura 5.1: Diagrama dos testes e resultados dos experimentos.

5.1 Regressores (agrupamento definido)

5.1.1 Acurácia

Acurácia (pior caso)

A figura 5.2 mostra que o modelo Fictício (MF) representa a maior parte os menores resultados. O modelo com Perceptron Multicamadas possui resultados semelhantes ao modelo Fictício e o modelo com Floresta Aleatória possui resultados semelhantes ao modelo com Regressão Linear (ML), mas este último representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a melhor acurácia, no pior caso.

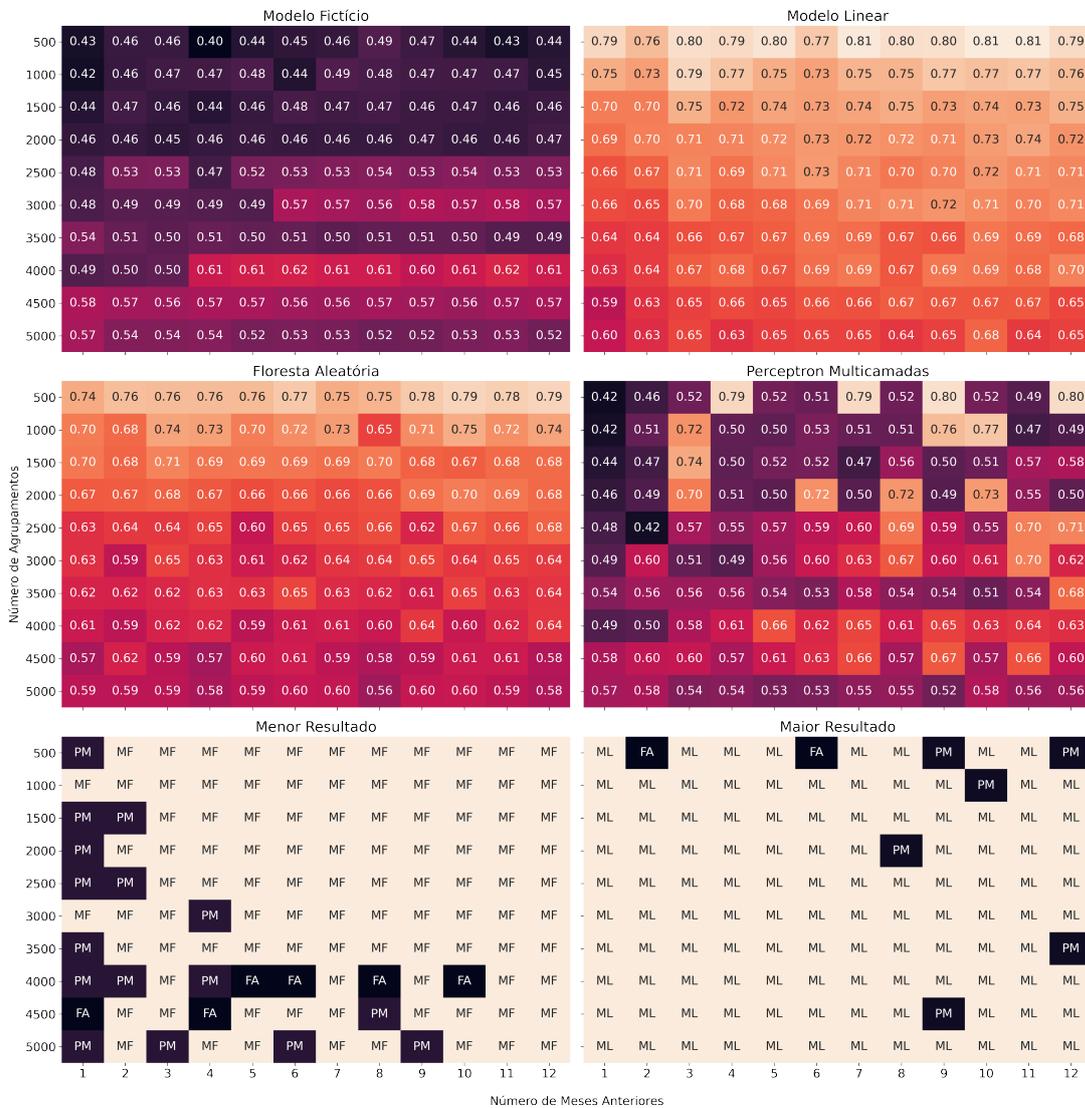


Figura 5.2: Resultado e comparação das acurácias mínimas dos regressores com número de agrupamentos definido.

Acurácia (média)

A figura 5.3 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. Os resultados dos modelos com Regressão Linear, Floresta Aleatória e Perceptron Multicamadas são semelhantes, mas o modelo com Regressão Linear (ML) representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a melhor acurácia, de forma geral.

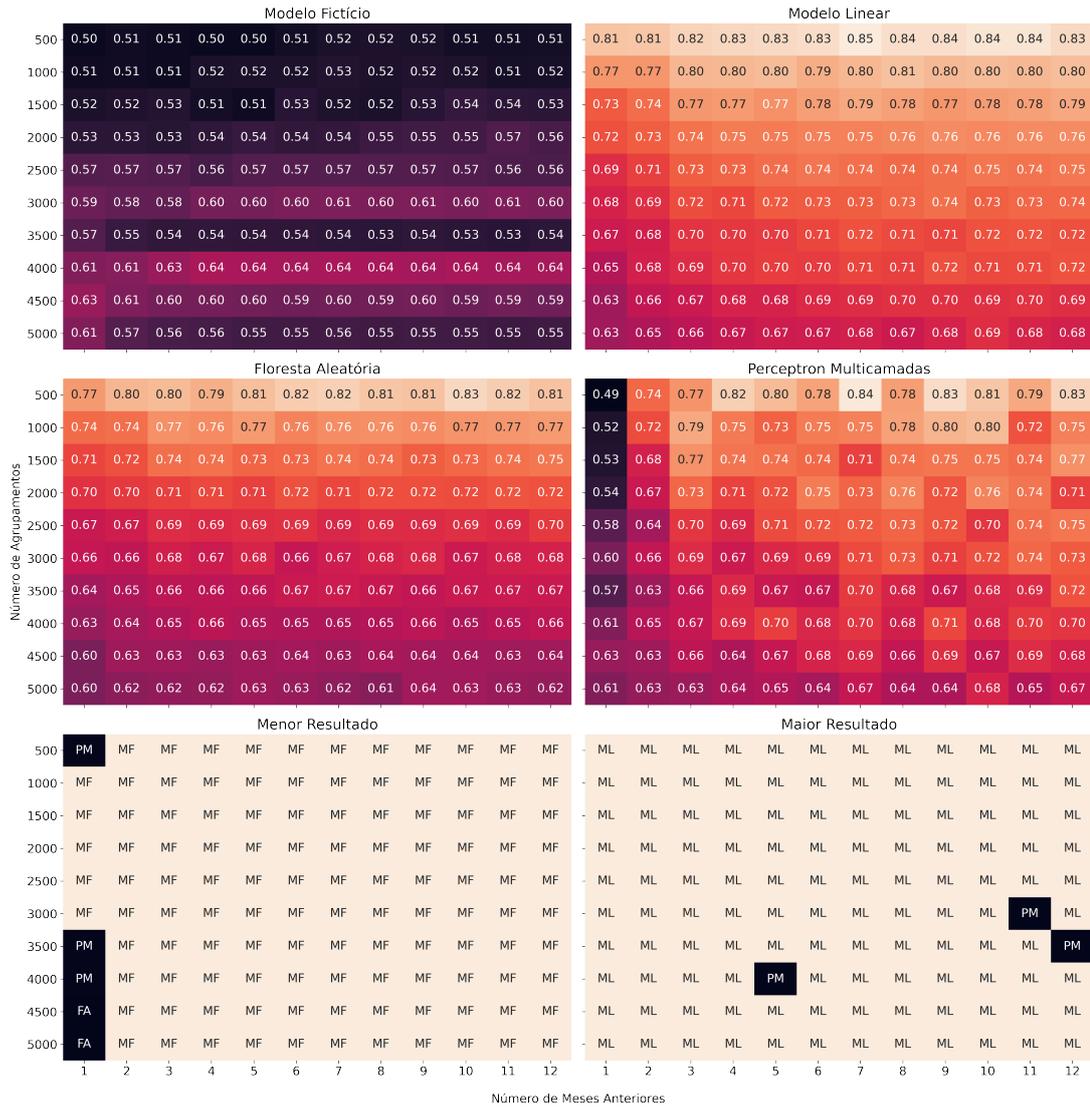


Figura 5.3: Resultado e comparação das acurácias médias dos regressores com número de agrupamentos definido.

5.1.2 Medida F_1

Medida F_1 (pior caso)

A figura 5.5 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados, e uma pequena parte representado pelo modelo com Perceptron Multicamadas (PM), indicando que esses modelos não classificam corretamente as duas classes. O modelo com Floresta Aleatória possui resultados semelhantes ao modelo com Regressão Linear (ML), onde este último representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear apresenta a melhor medida F_1 , no pior caso.



Figura 5.5: Resultado e comparação das medidas F_1 mínimas dos regressores com número de agrupamentos definido.

Medida F_1 (média)

A figura 5.6 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados, indicando que este modelo não consegue classificar corretamente as duas classes. Os modelos com Regressão Linear, Floresta Aleatória e Perceptron Multicamadas possuem resultados semelhantes, mas o modelo com Regressão Linear (ML) representa a maior parte dos melhores resultados. Portanto, o modelo com Regressão Linear possui a melhor medida F_1 , de forma geral.

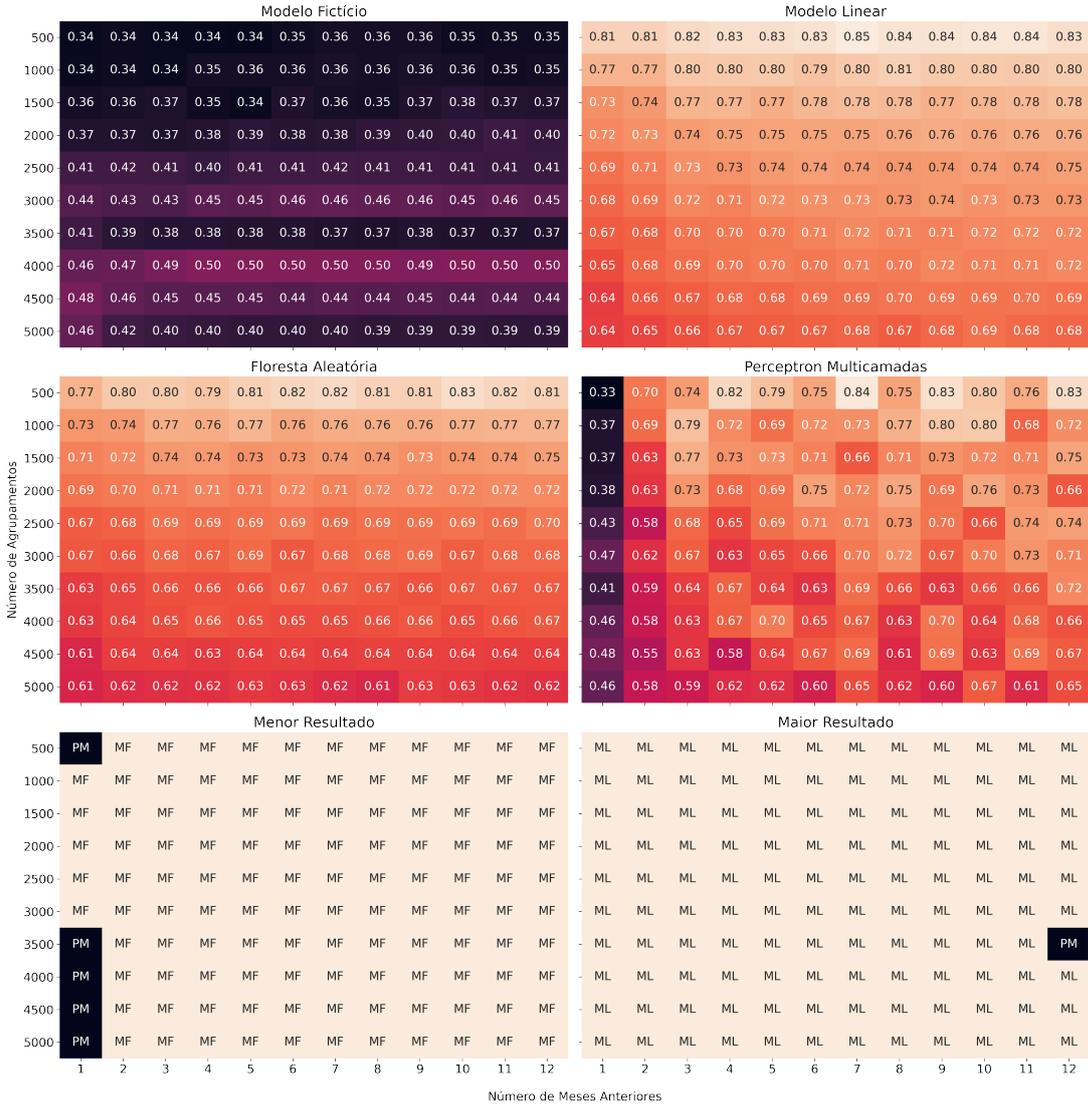


Figura 5.6: Resultado e comparação das medidas F_1 médias dos regressores com número de agrupamentos definido.

Medida F_1 (desvio padrão)

A figura 5.7 mostra que os modelos com Regressão Linear e Floresta Aleatória possuem resultados baixos e semelhantes, e o modelo com Perceptron Multicamadas possui resultados elevados, em alguns casos chegando a 20%. A maior parte dos menores resultados são compostos pelo modelo com Regressão Linear (ML), enquanto os maiores resultados são compostos pelo modelo com Perceptron Multicamadas (PM). Logo, o modelo com Regressão Linear possui a medida F_1 mais estável durante o ano.

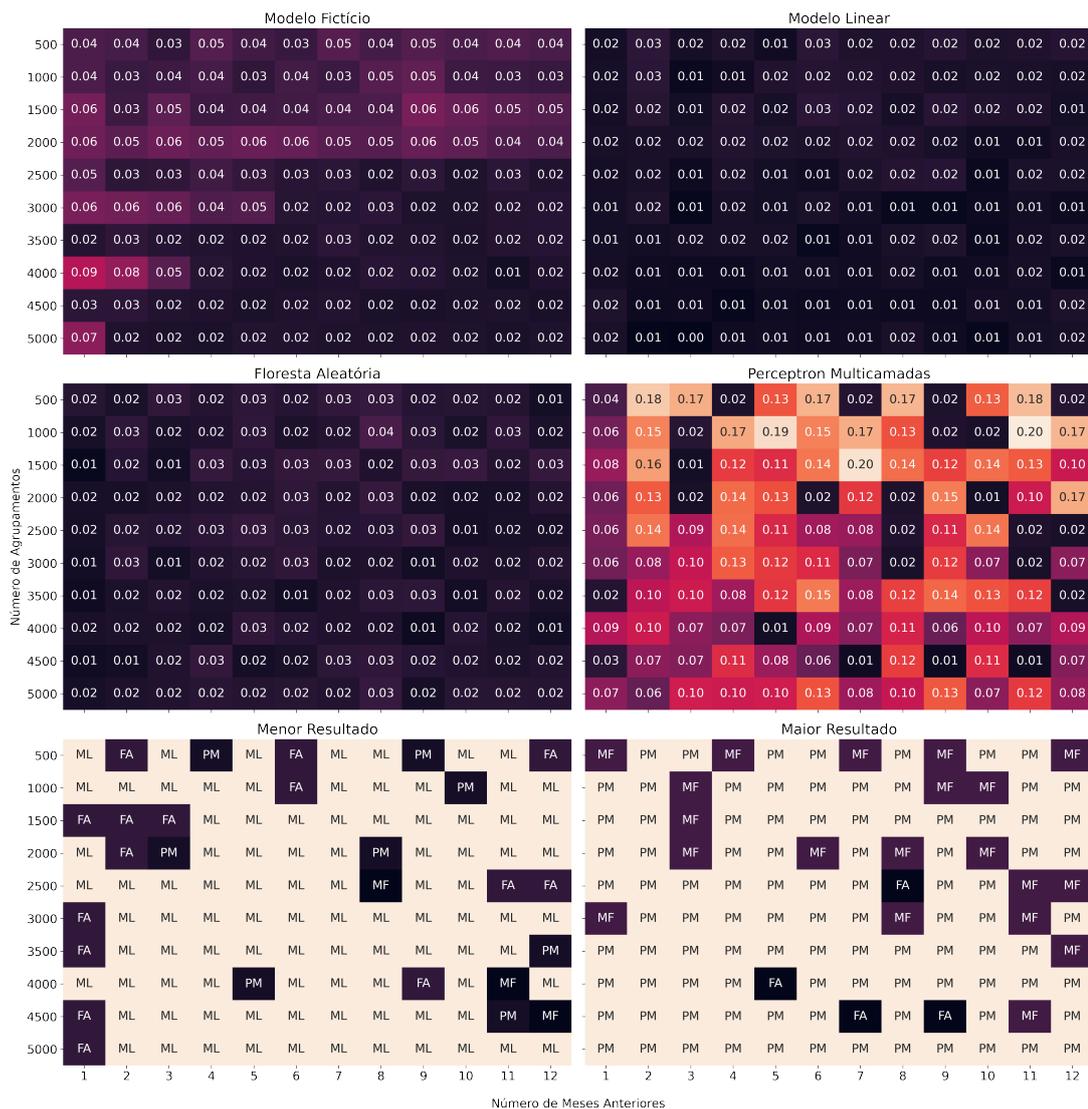


Figura 5.7: Resultado e comparação dos desvios padrão das medidas F_1 dos regressores com número de agrupamentos definido.

5.1.3 Complexidade Computacional

Complexidade Computacional (pior caso)

A figura 5.8 mostra que o modelo com Regressão Linear (ML) representa todos os menores resultados. O modelo com Perceptron Multicamadas (PM) representa os maiores resultados, com o maior resultado em torno de 180 segundos, ou 3 minutos. Os resultados do modelo Fictício não foram incluídos na comparação, pois este modelo não possui uma etapa de treinamento. Portanto, o modelo com Regressão Linear possui o menor tempo de treinamento, no pior caso.

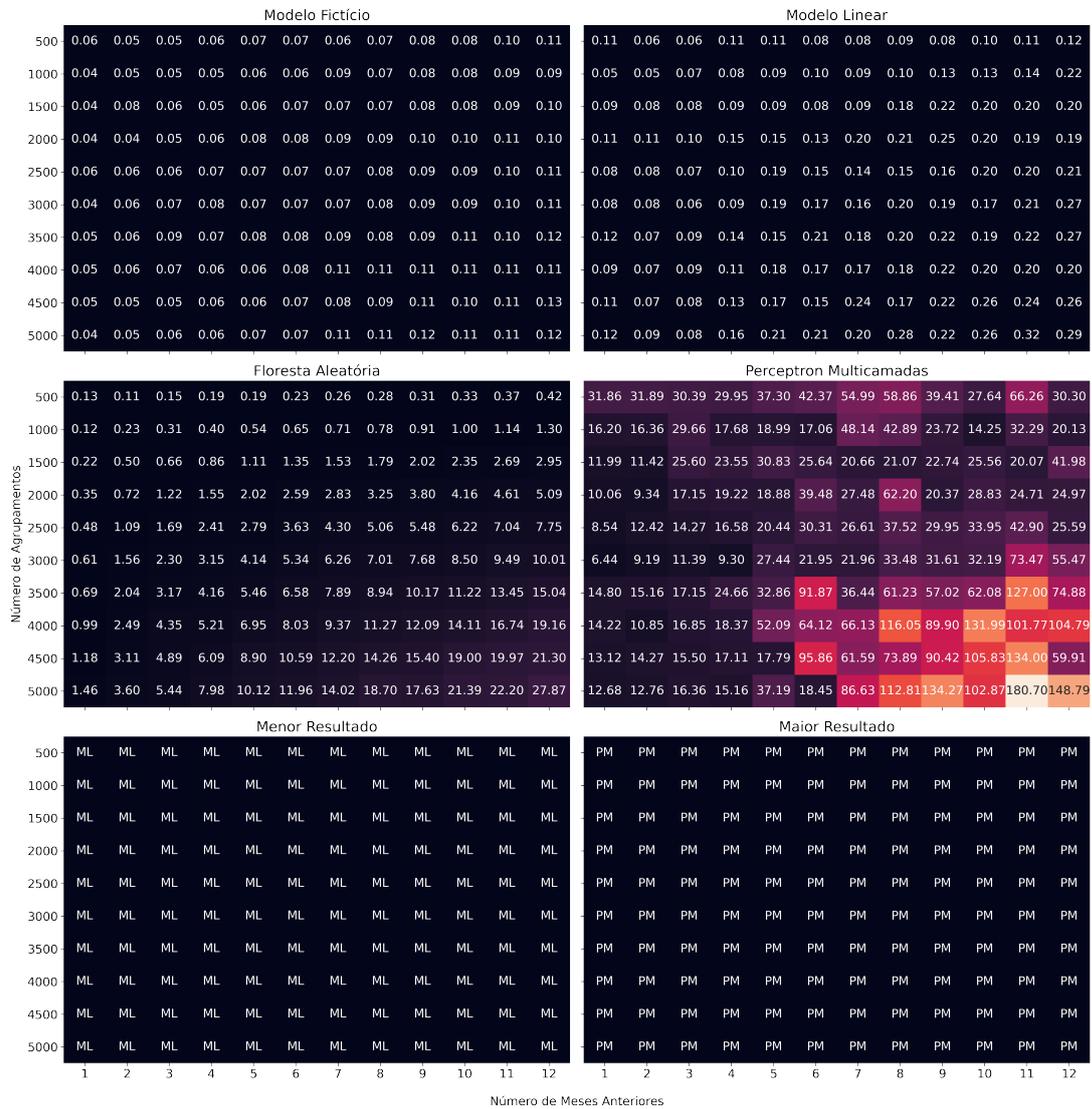


Figura 5.8: Resultado e comparação do tempo de treinamento máximo dos regressores com número de agrupamentos definido.

Complexidade Computacional (média)

A figura 5.9 mostra que o modelo com Regressão Linear (ML) representa todos os menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa os maiores resultados. Os resultados do modelo Fictício não foi incluso na comparação, pois este modelo não possui uma etapa de treinamento. Logo, o modelo com Regressão Linear possui o menor tempo de treinamento, em média.

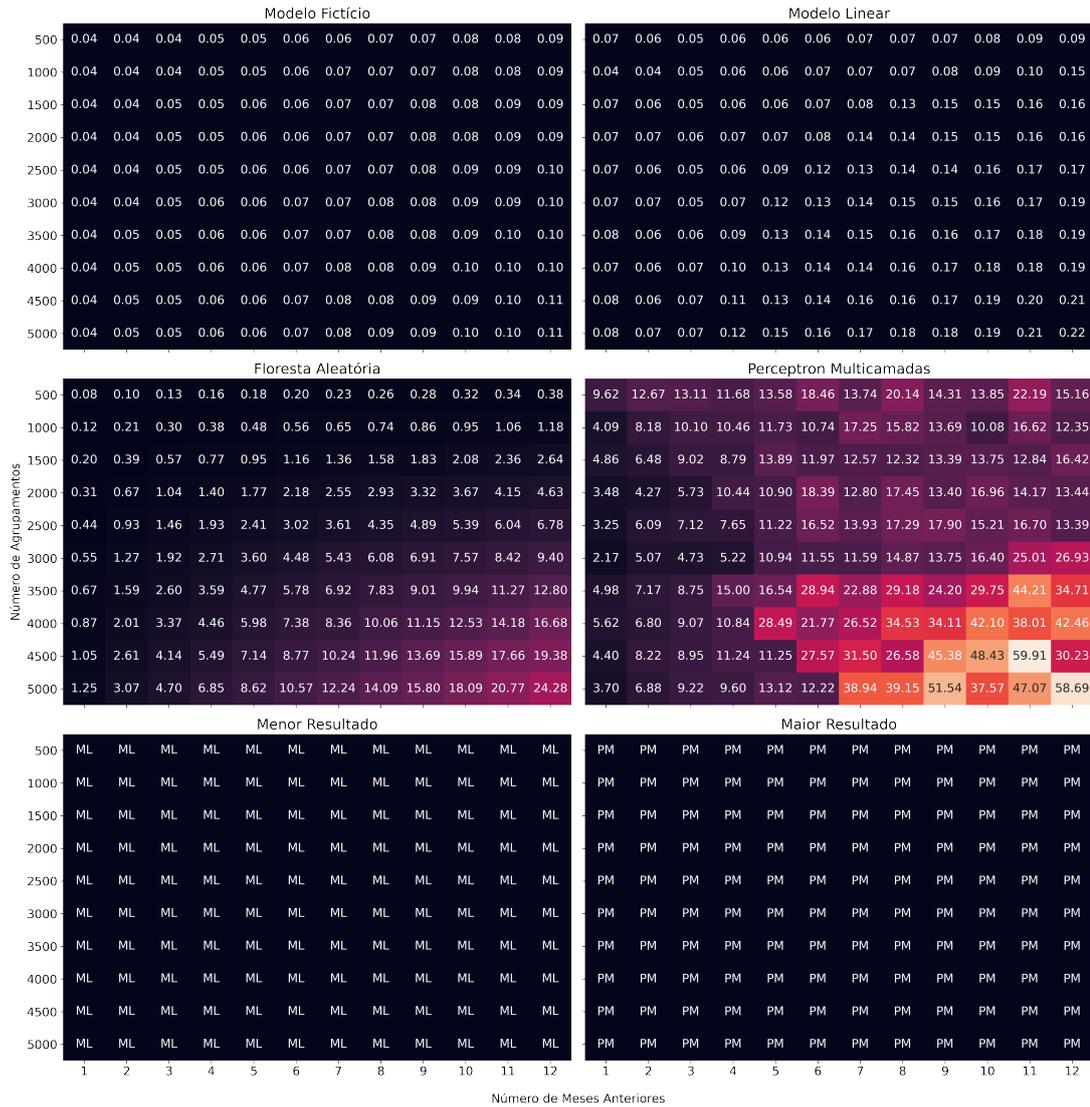


Figura 5.9: Resultado e comparação do tempo de treinamento médio dos regressores com número de agrupamentos definido.

Complexidade Computacional (desvio padrão)

A figura 5.10 mostra que o modelo com Regressão Linear (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa todos os maiores resultados. Os resultados do modelo com Floresta Aleatória está entre os dois modelos. Portanto, o modelo com Regressão Linear possui o tempo de treinamento mais estável.

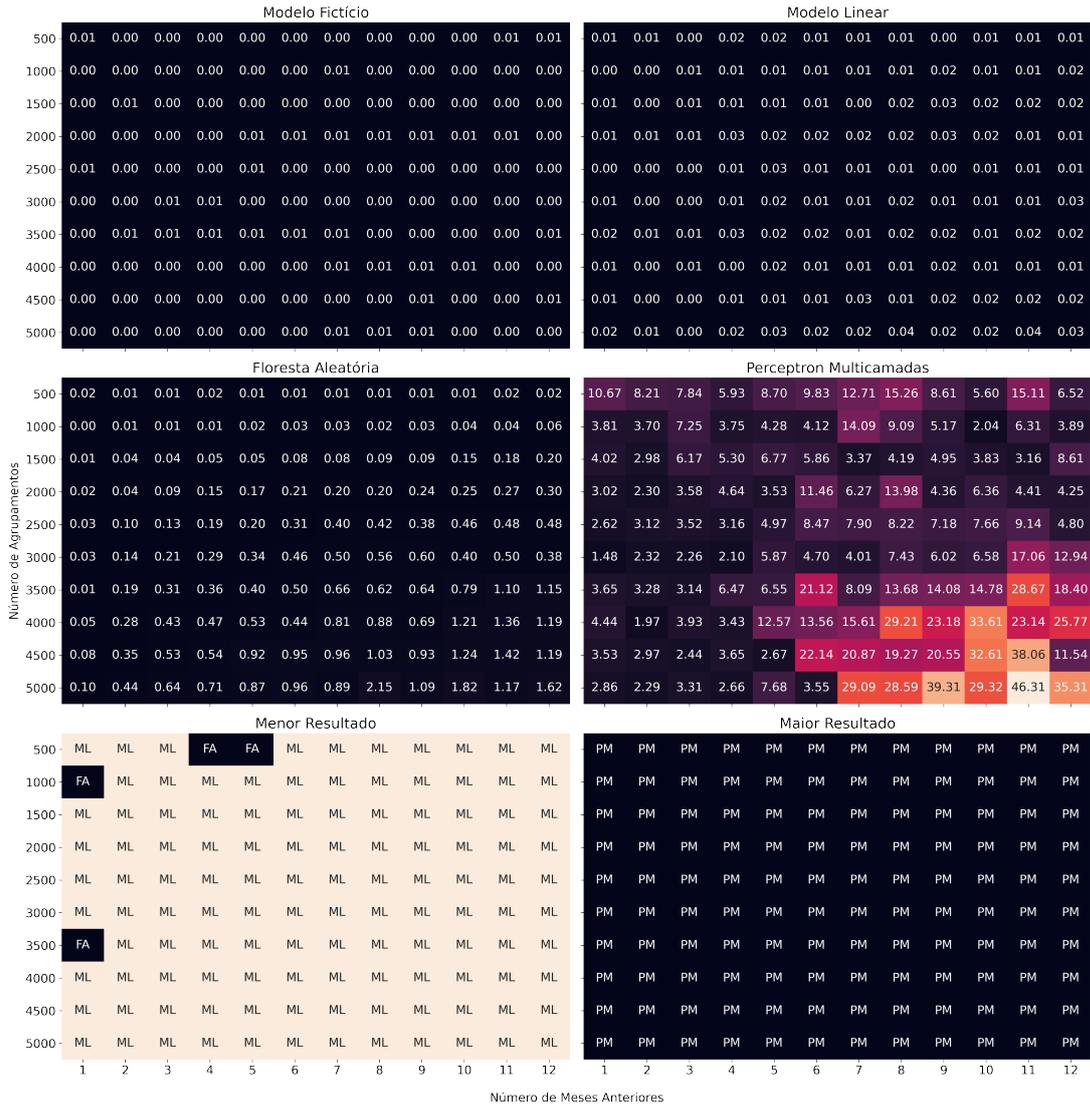


Figura 5.10: Resultado e comparação dos desvios padrão do tempo de treinamento dos regressores com número de agrupamentos definido.

5.2 Classificadores (agrupamento definido)

5.2.1 Acurácia

Acurácia (piores caso)

A figura 5.11 mostra que os modelos Fictício (MF) e com Perceptron Multicamadas (PM) compartilham os menores resultados. O modelo com Floresta Aleatória possui resultados semelhantes ao modelo com Regressão Logística (ML), mas este último possui a maior parte dos maiores resultados. Assim, o modelo com Regressão Logística possui a melhor acurácia, no pior caso.

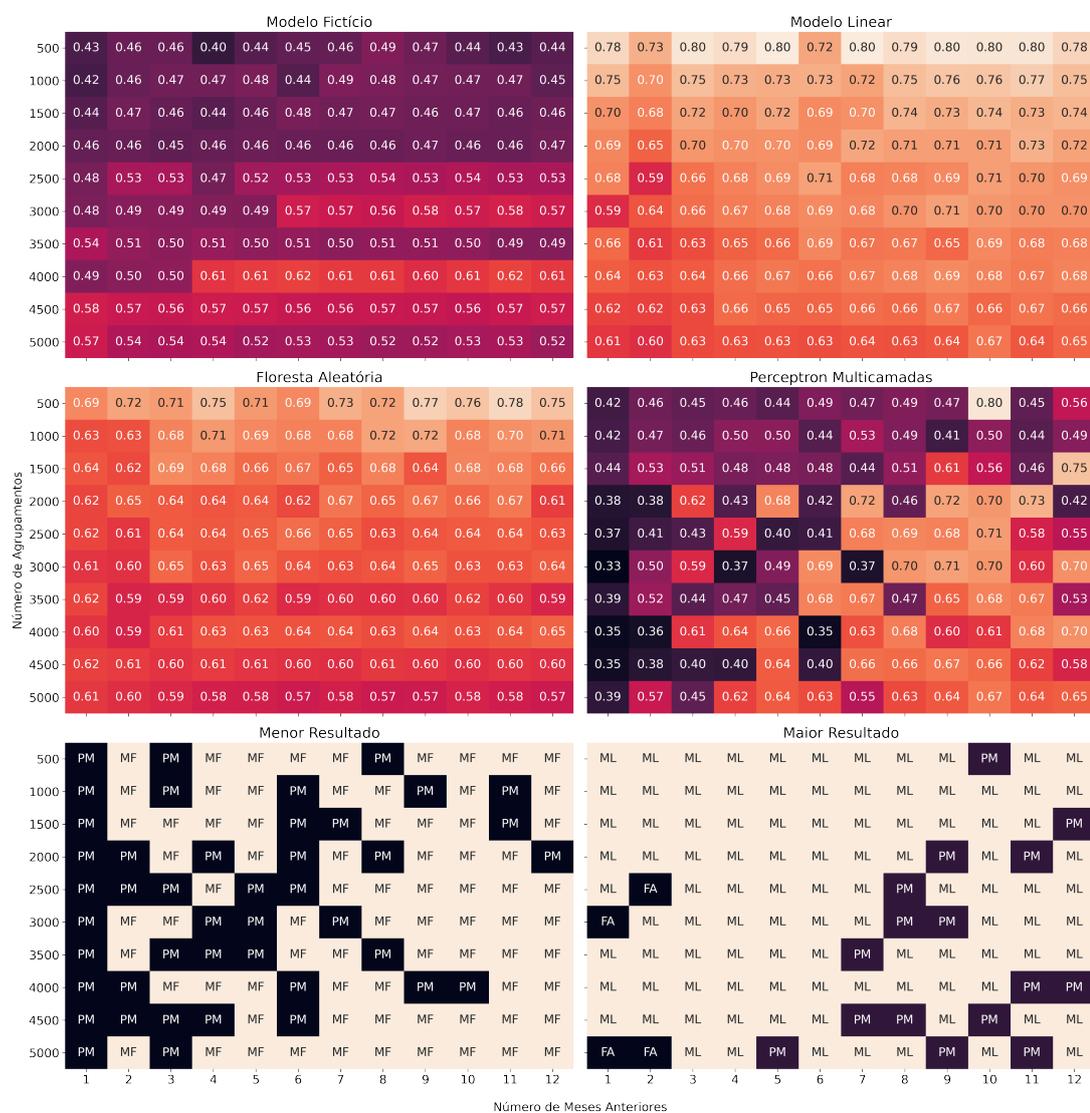


Figura 5.11: Resultado e comparação das acurácias mínimas dos classificadores com número de agrupamentos definido.

Acurácia (média)

A figura 5.12 mostra que o modelo Fictício(MF) representa a maior parte dos menores resultados e o modelo com Perceptron Multicamadas (PM) possui alguns resultados baixos, com um deles chegando a apenas 49%. O modelo com Floresta Aleatória possui resultados semelhantes ao modelo com Regressão Logística (ML), mas este último representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Logística possui a melhor acurácia, de forma geral.

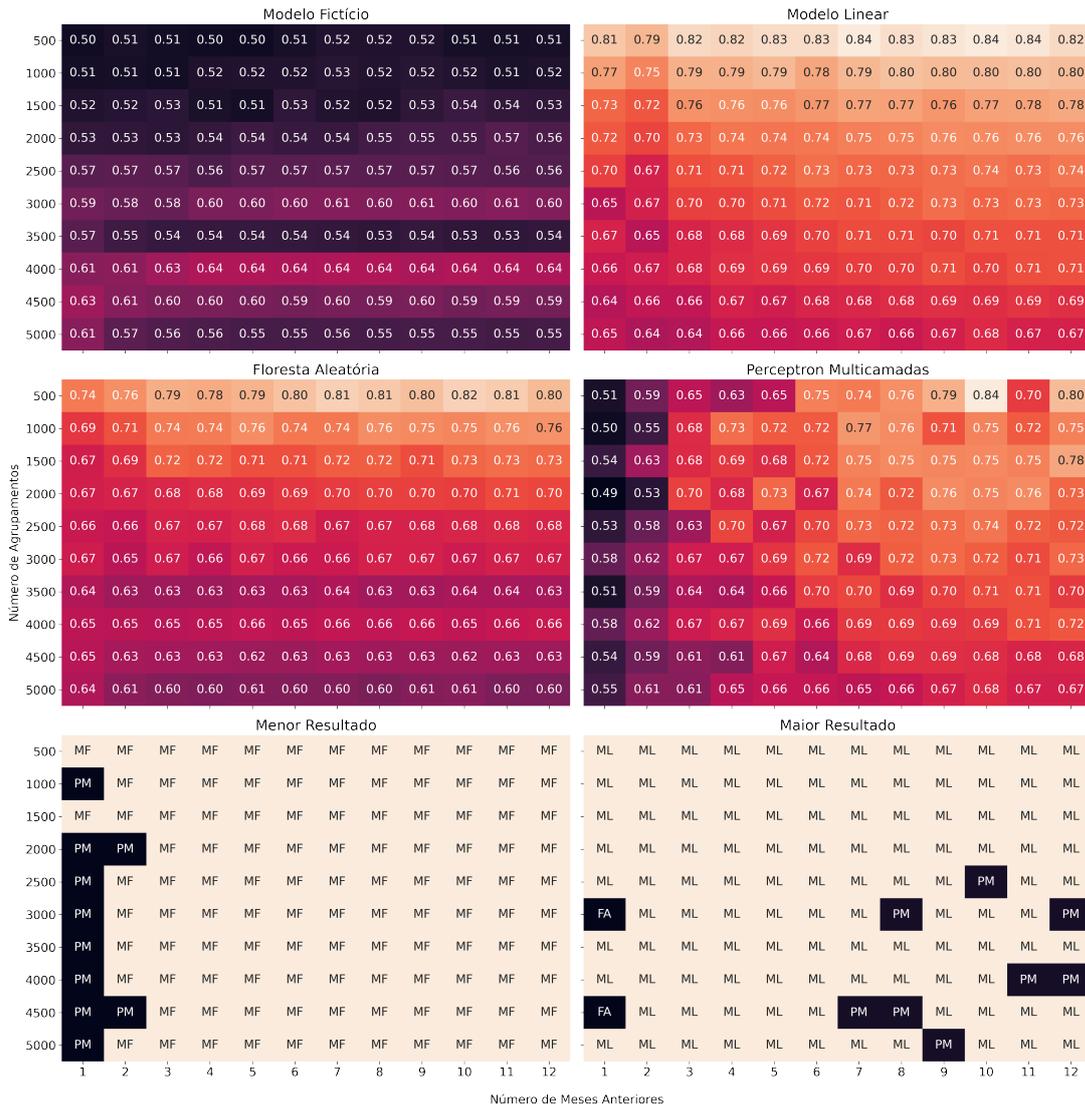


Figura 5.12: Resultado e comparação das acurácias médias dos classificadores com número de agrupamentos definido.

Acurácia (desvio padrão)

A figura 5.13 mostra que os modelos com Regressão Logística (ML), Floresta Aleatória (FA) e Perceptron Multicamadas (PM) compartilham os menores resultados, com uma pequena vantagem para o modelo com Regressão Logística. Por outro lado, o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados, seguido pelo modelo Fictício (MF) e Floresta Aleatória (FA). Assim, o modelo com Regressão Logística possui a acurácia mais estável.

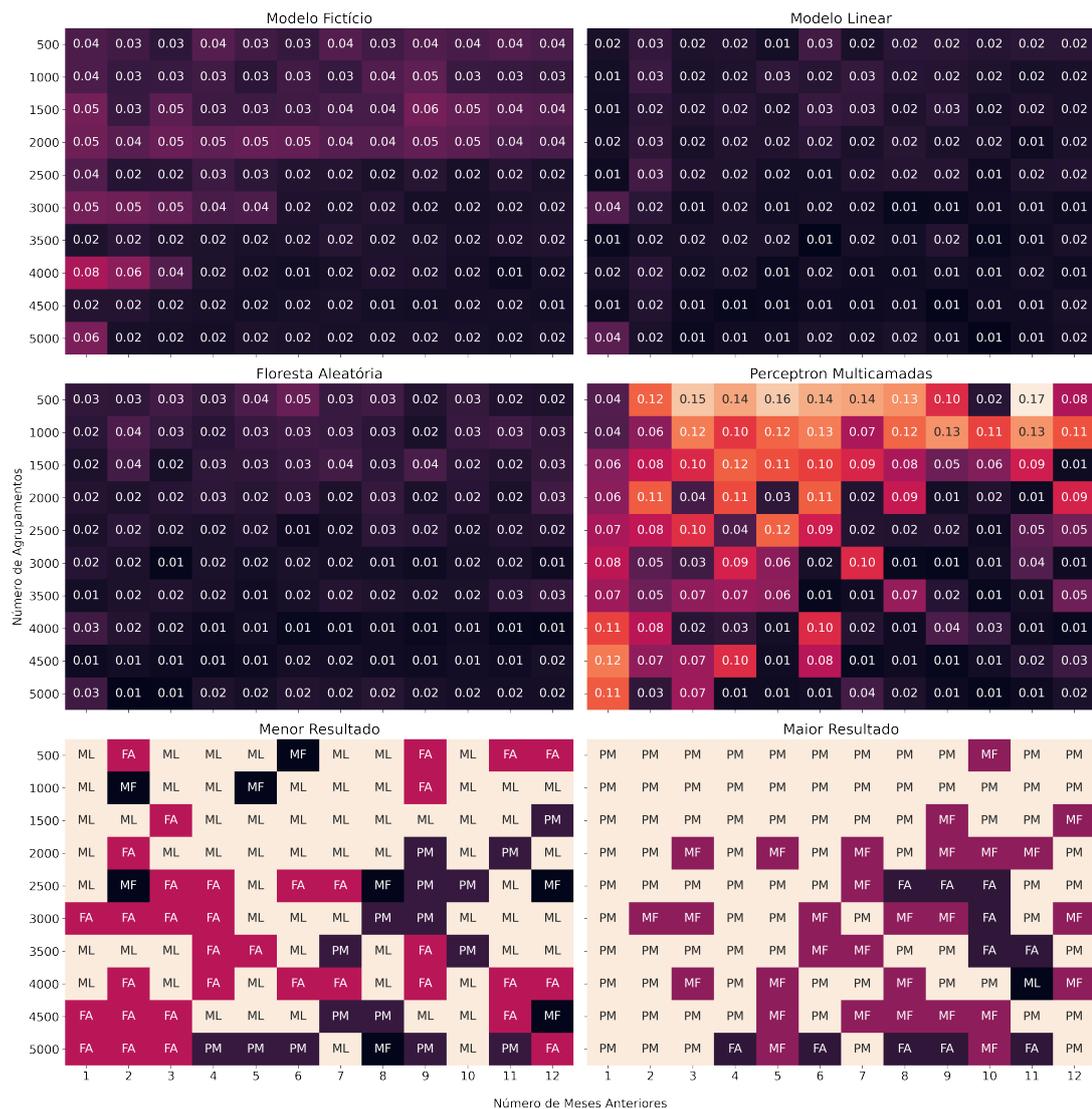


Figura 5.13: Resultado e comparação dos desvios padrão das acurácias dos classificadores com número de agrupamentos definido.

5.2.2 Medida F_1

Medida F_1 (pior caso)

A figura 5.14 mostra que os modelos Fictício (MF) e com Perceptron Multicamadas (PM) compartilham os menores resultados. O modelo com Floresta Aleatória (FA) possui resultados semelhantes ao modelo com Regressão Logística (ML), mas este último representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Logística possui a melhor medida F_1 , no pior caso.



Figura 5.14: Resultado e comparação das medidas F_1 mínimas dos classificadores com número de agrupamentos definido.

Medida F_1 (média)

A figura 5.15 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados e o modelo com Perceptron Multicamadas (PM) possui alguns resultados baixos, com um deles chegando a apenas 34%. O modelo com Floresta Aleatória (FA) possui resultados semelhantes ao modelo com Regressão Logística (ML), mas este último representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Logística possui a melhor medida F_1 , de forma geral.

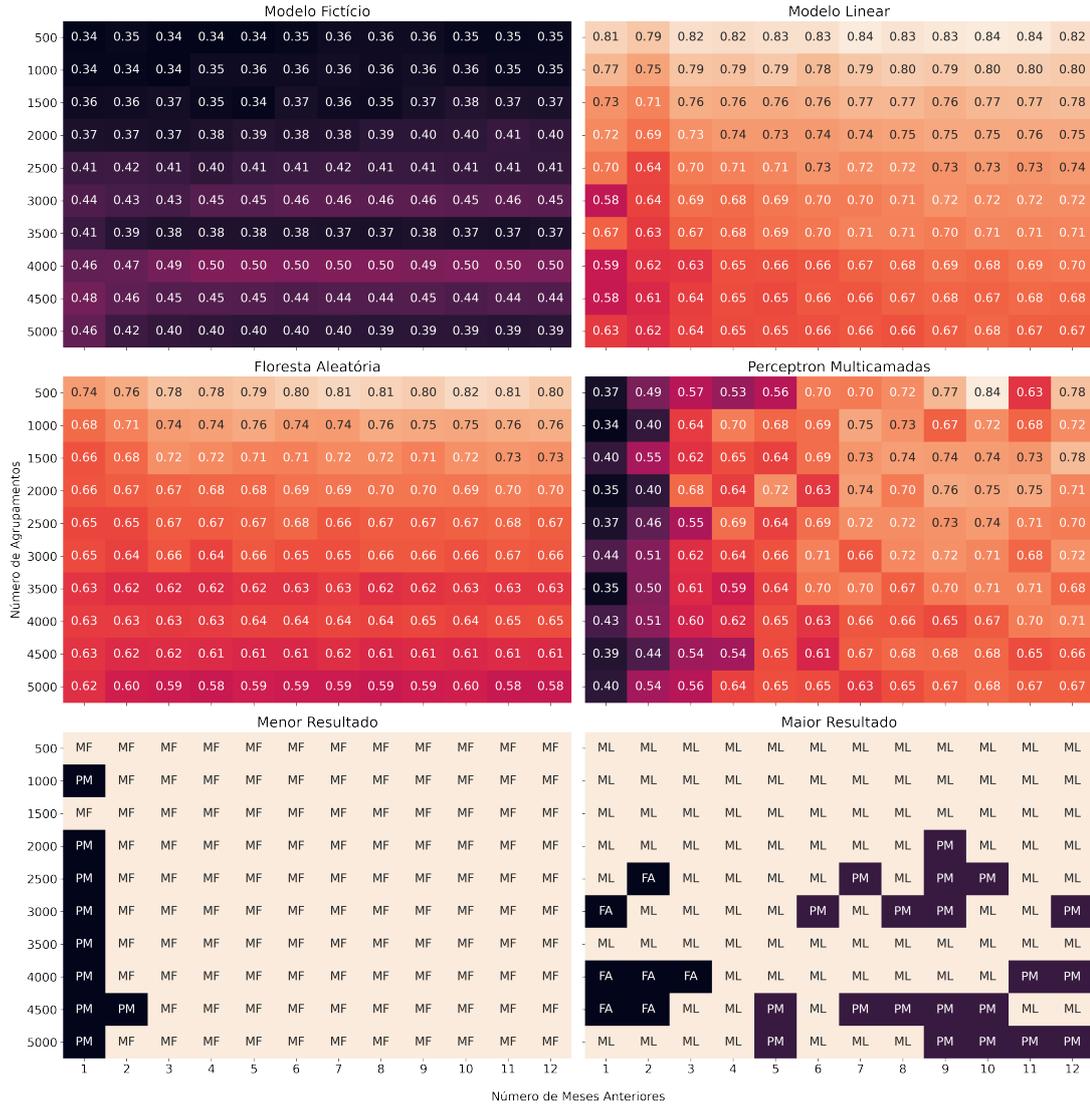


Figura 5.15: Resultado e comparação das medidas F_1 médias dos classificadores com número de agrupamentos definido.

Medida F_1 (desvio padrão)

A figura 5.16 mostra que os modelos com Regressão Logística (ML) e Floresta Aleatória (FA) compartilham os menores resultados, com uma pequena vantagem para o modelo com Regressão Logística. Por outro lado, o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados, seguido pelos modelos Fictício (MF) e com Floresta Aleatória (FA). Portanto, o modelo com Regressão Logística possui a medida F_1 mais estável.

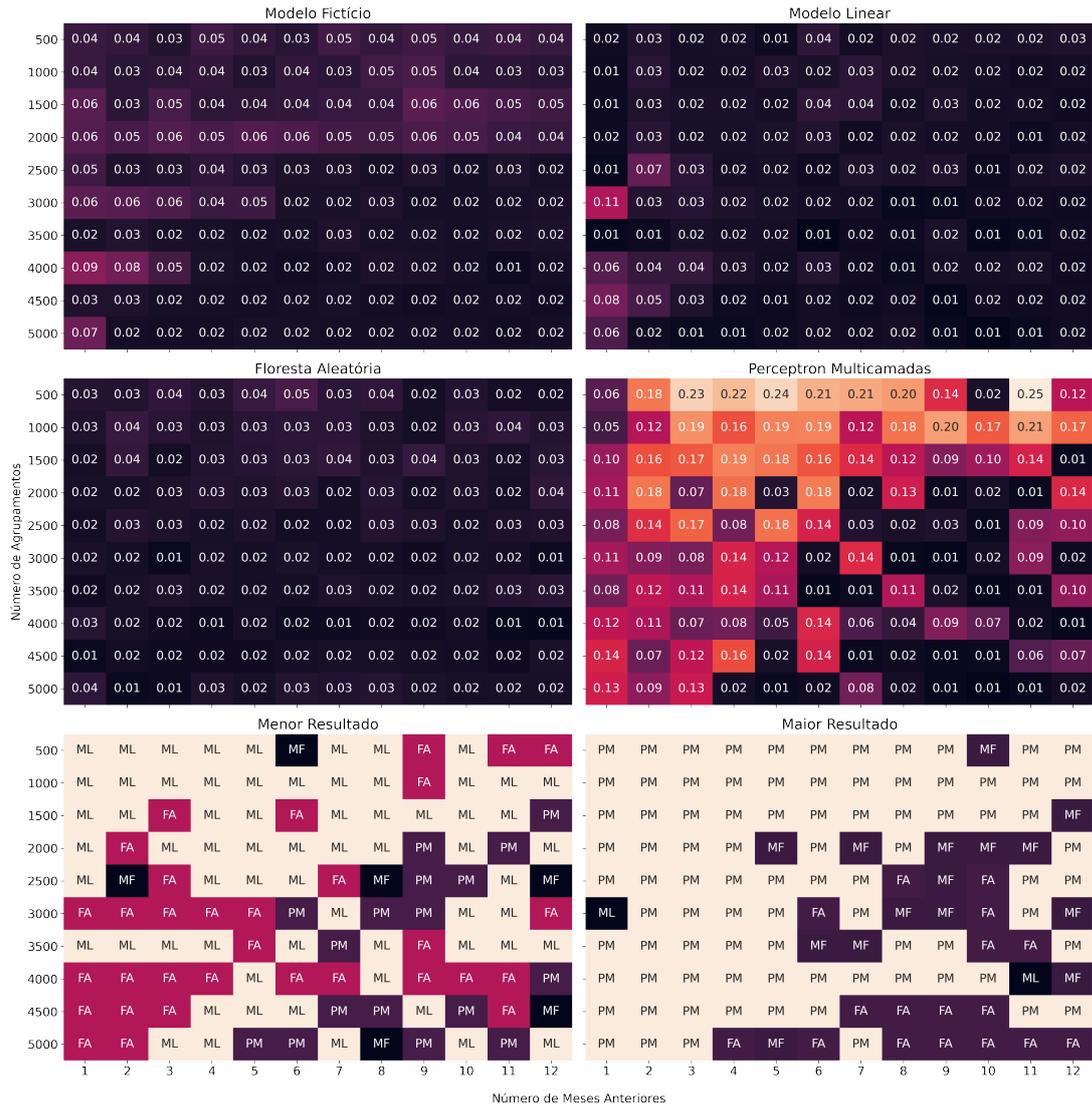


Figura 5.16: Resultado e comparação dos desvios padrão das medidas F_1 dos classificadores com número de agrupamentos definido.

5.2.3 Complexidade Computacional

Complexidade Computacional (pior caso)

A figura 5.17 mostra que o modelo com Regressão Logística (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados. É possível observar que os resultados do modelo com Floresta Aleatória estão entre os dois modelos. O modelo Fictício não foi incluído na comparação, pois o mesmo não possui etapa de treinamento. Portanto, o modelo com Regressão Logística possui o menor tempo de treinamento, no pior caso.

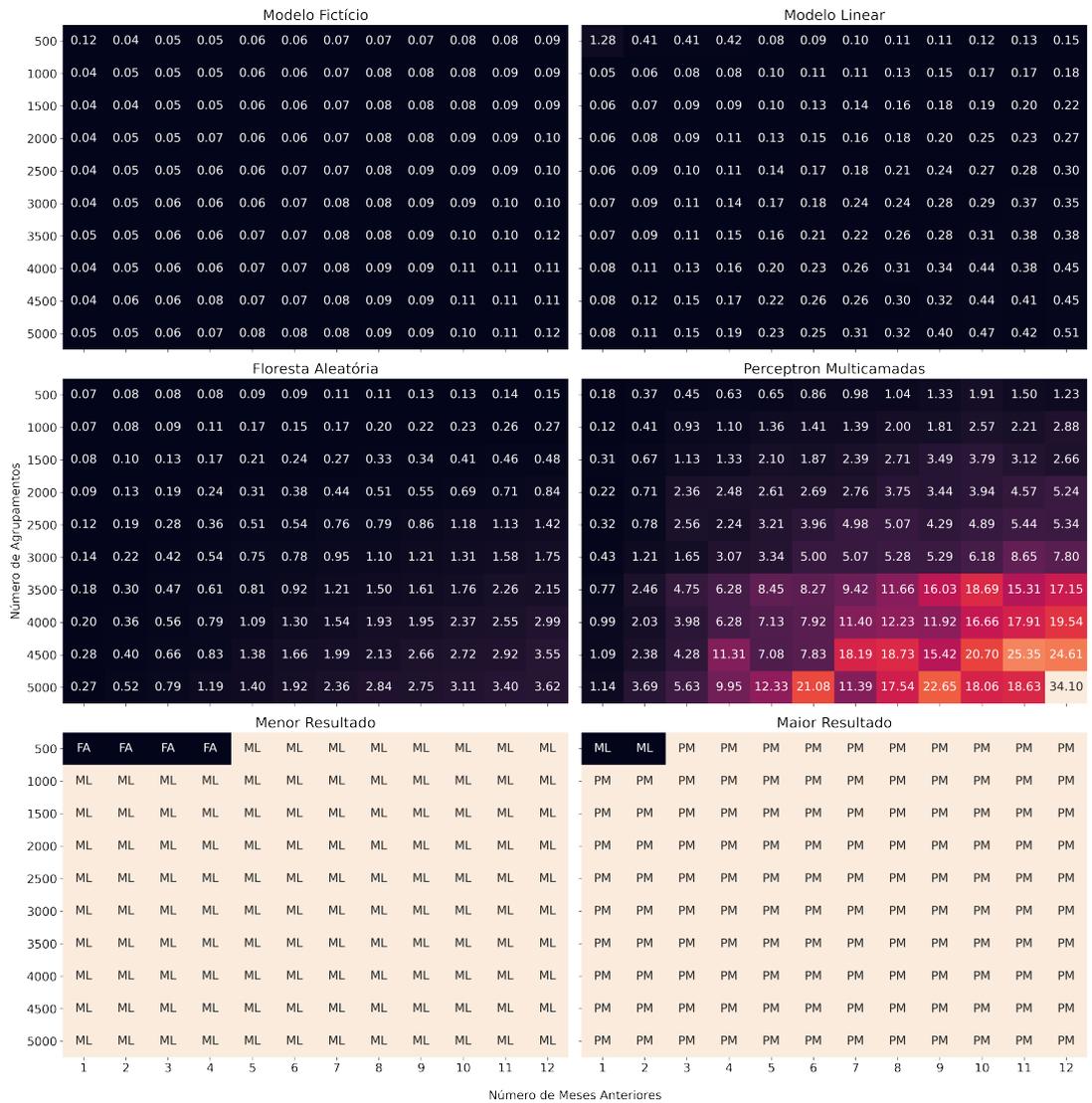


Figura 5.17: Resultado e comparação dos tempos de treinamento máximo dos classificadores com número de agrupamentos definido.

Complexidade Computacional (média)

A figura 5.18 mostra que o modelo com Regressão Logística (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados. É possível observar que os resultados do modelo com Floresta Aleatória encontram-se entre os dois modelos. O modelo Fictício não foi incluído na comparação, pois o mesmo não possui uma etapa de treinamento. Logo, o modelo com Regressão Logística apresenta o menor tempo de treinamento, de forma geral.

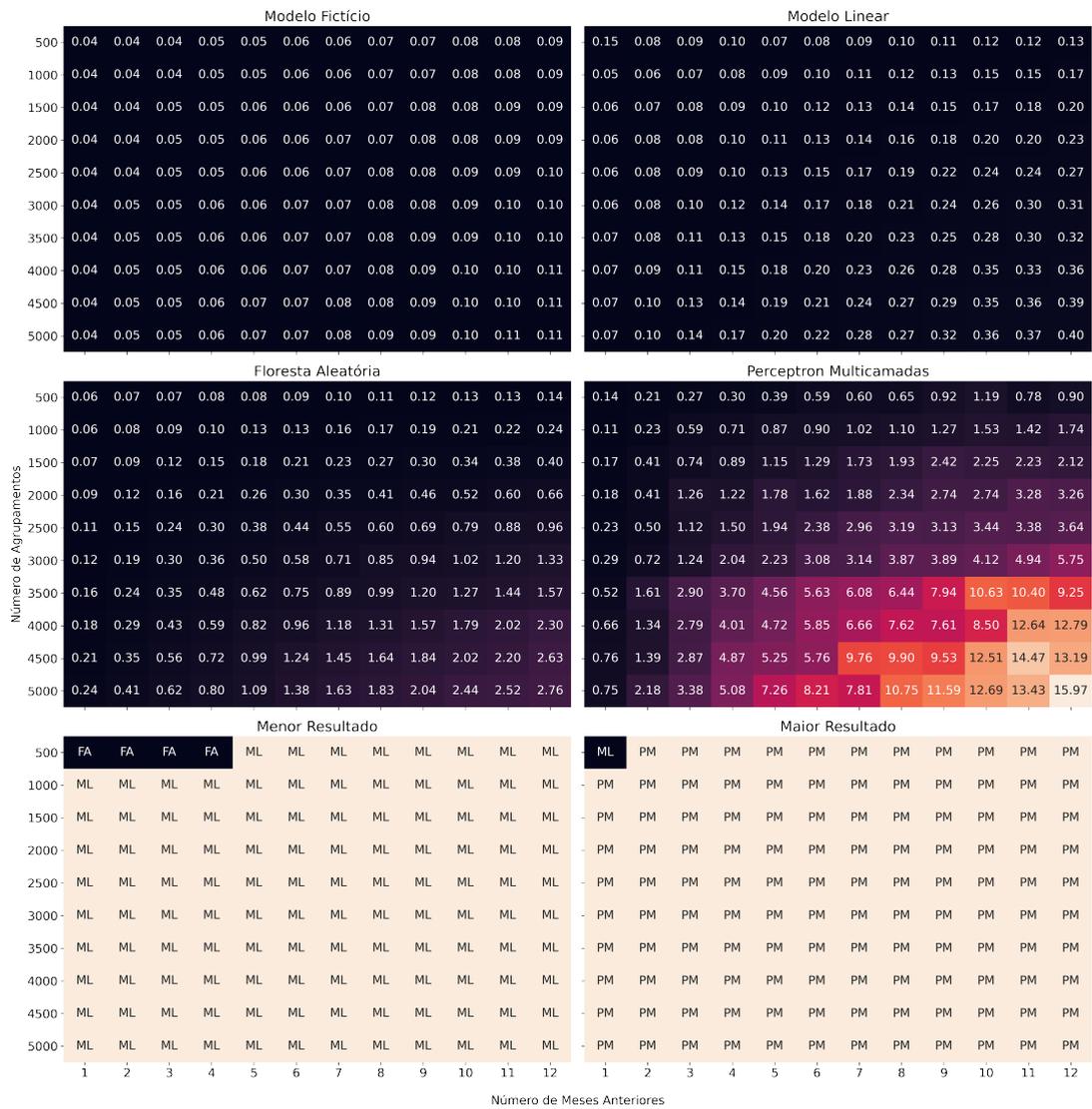


Figura 5.18: Resultado e comparação dos tempos de treinamento médio dos classificadores com número de agrupamentos definido.

Complexidade Computacional (desvio padrão)

A figura 5.19 mostra que o modelo com Regressão Logística (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados. É possível observar que os resultados do modelo com Floresta Aleatória encontram-se entre os dois modelos. O modelo Fictício não foi incluído na comparação, pois o mesmo não possui uma etapa de treinamento. Portanto, o modelo com Regressão Logística apresenta o tempo de treinamento mais estável.

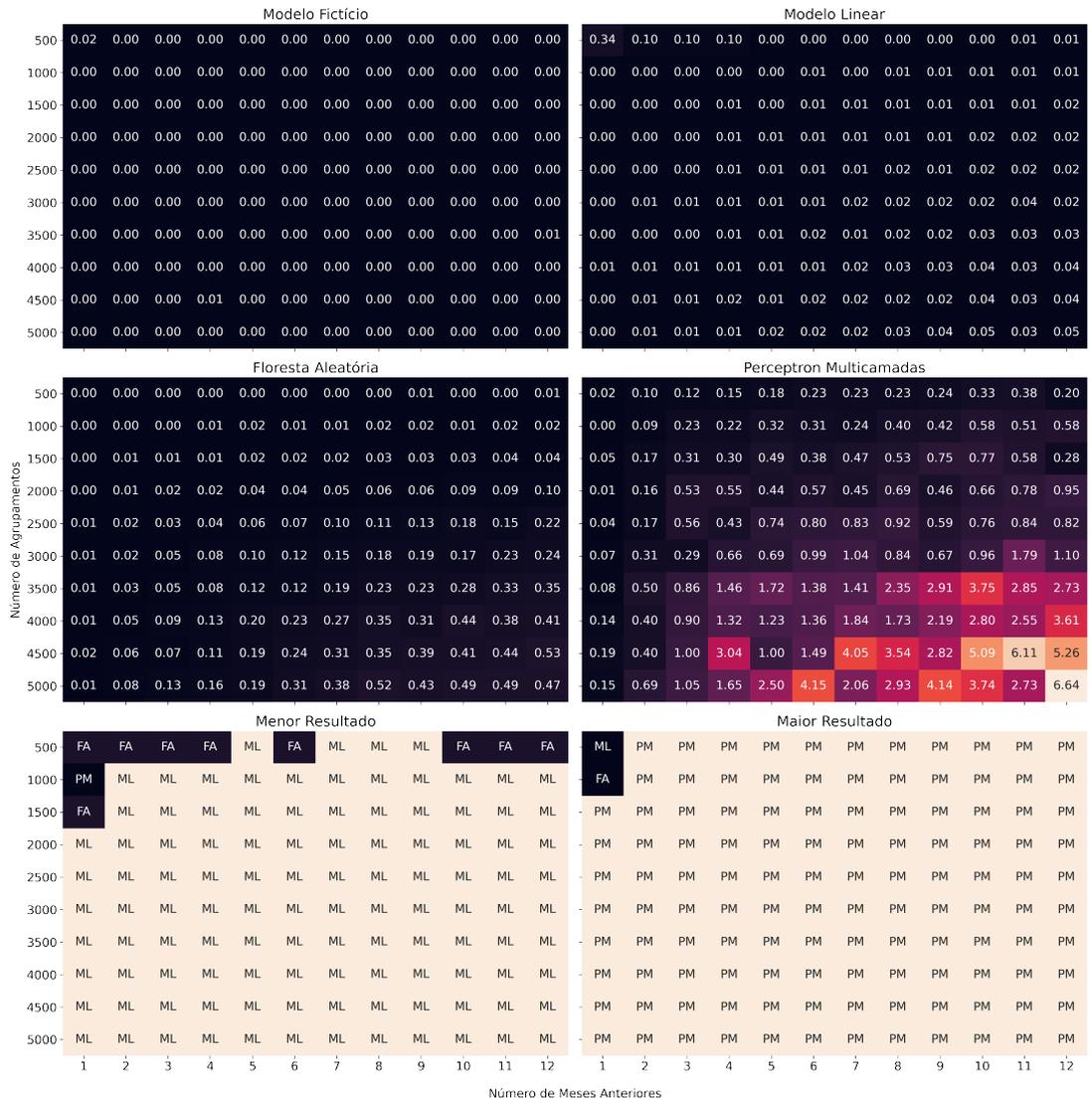


Figura 5.19: Resultado e comparação dos desvios padrão dos tempos de treinamento dos classificadores com número de agrupamentos definido.

5.3 Melhor abordagem (agrupamento definido)

Os comparativos das seções 5.1 e 5.2 mostram que o melhores resultados são obtidos com o modelo com Regressão Linear (entre os regressores) e o modelo com Regressão Logística (entre os classificadores). Esta seção irá comparar os resultados para definir a melhor abordagem, com número definido de agrupamentos.

5.3.1 Acurácia

A figura 5.20 mostra que os modelos com Regressão Logística (C) e Regressão Linear (R) possuem resultados semelhantes, mas este último representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Linear possui a melhor acurácia, de forma geral.

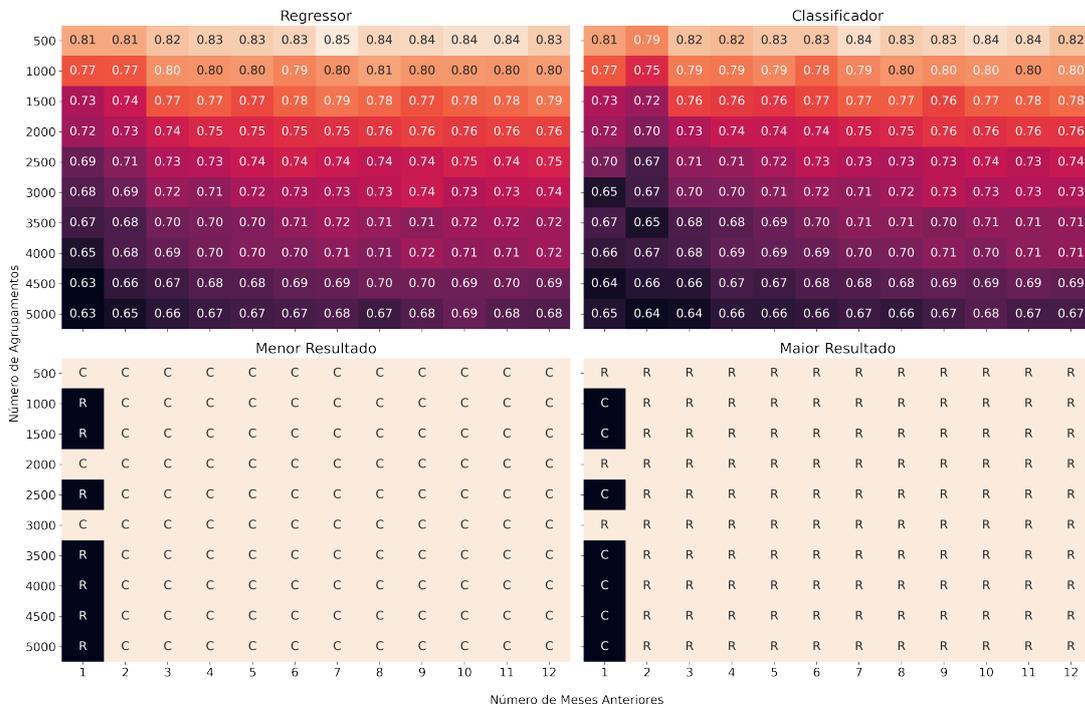


Figura 5.20: Resultado e comparação das acurácias médias do regressor e classificador, com número de agrupamentos definido.

5.3.2 Medida F_1

A figura 5.21 mostra que os modelos com Regressão Logística (C) e Regressão Linear (R) possuem resultados semelhantes, mas este último representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a melhor medida F_1 , de forma geral.

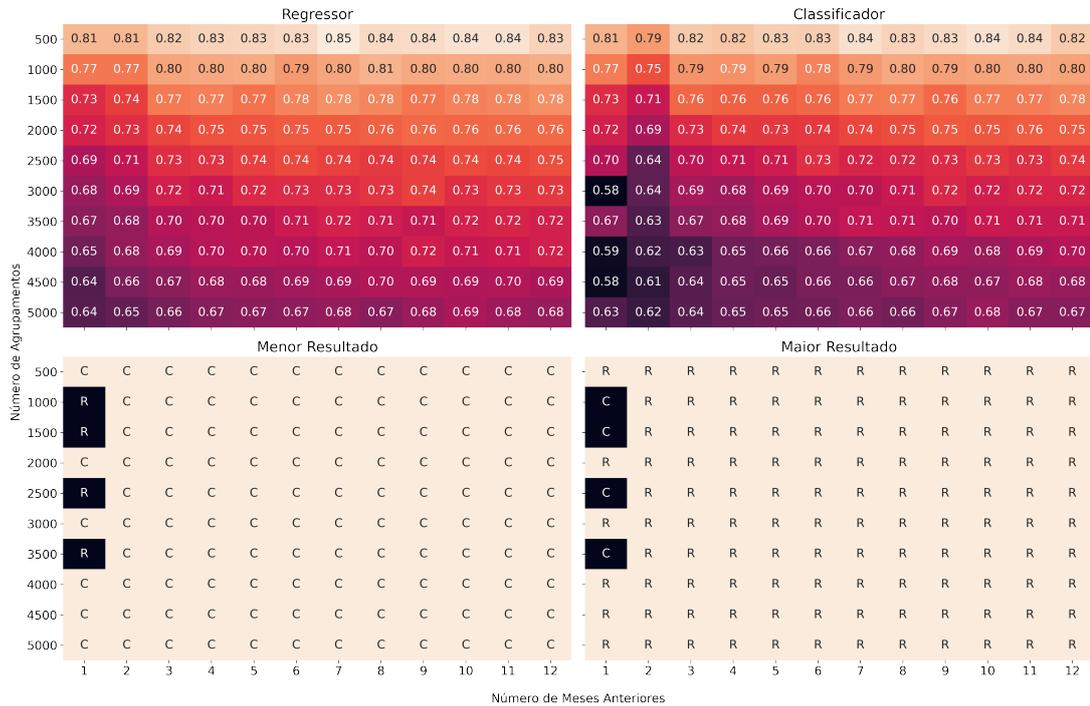


Figura 5.21: Resultado e comparação das medidas F_1 médias do regressor e classificador, com número de agrupamentos definido.

5.3.3 Complexidade Computacional

A figura 5.22 mostra que os modelos com Regressão Logística (C) e Regressão Linear (R) possuem resultados semelhantes, mas este último representa a maior parte dos menores resultados. Assim, o modelo com Regressão Linear possui o menor tempo de treinamento, de forma geral.

5.4 | REGRESSORES (AGRUPAMENTO DINÂMICO)

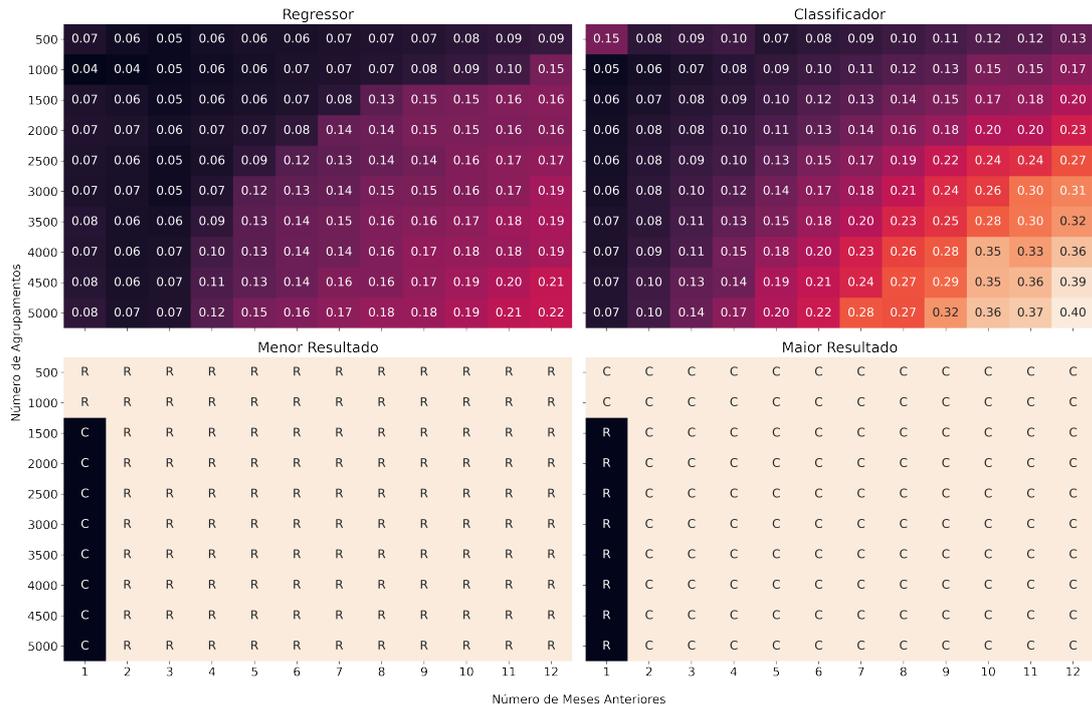


Figura 5.22: Resultado e comparação dos tempos de treinamento médio do regressor e classificador, com número de agrupamentos definido.

5.4 Regressores (agrupamento dinâmico)

5.4.1 Acurácia

Acurácia (pior caso)

A figura 5.23 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Floresta Aleatória possui resultados semelhantes ao modelo com Regressão Linear (ML), mas este último representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a melhor acurácia, no pior caso.



Figura 5.23: Resultado e comparação das acurácias mínimas dos regressores com número de agrupamentos dinâmico.

Acurácia (média)

A figura 5.24 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. Os três modelos possuem resultados semelhantes, mas o modelo com Regressão Linear (ML) representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Linear possui a melhor acurácia, de forma geral.



Figura 5.24: Resultado e comparação das acurácias médias dos regressores com número de agrupamentos dinâmico.

Acurácia (desvio padrão)

A figura 5.25 mostra que o modelo com Regressão Linear (ML) e Floresta Aleatória (FA) compartilham os menores resultados, com a maior parte representada pelo modelo com Regressão Linear. O modelo com Perceptron Multicamadas representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a acurácia mais estável durante o ano.

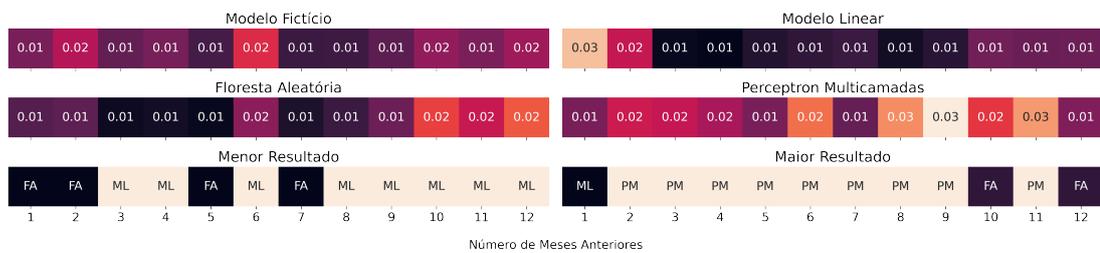


Figura 5.25: Resultado e comparação dos desvios padrão das acurácias dos regressores com número de agrupamentos dinâmico.

5.4.2 Medida F_1

Medida F_1 (piores caso)

A figura 5.26 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Floresta Aleatória (FA) possui resultados semelhantes ao modelo com Regressão Linear (ML), com este último representando a maior parte dos maiores resultados. Portanto, o modelo com Regressão Linear possui a maior medida F_1 , no pior caso.

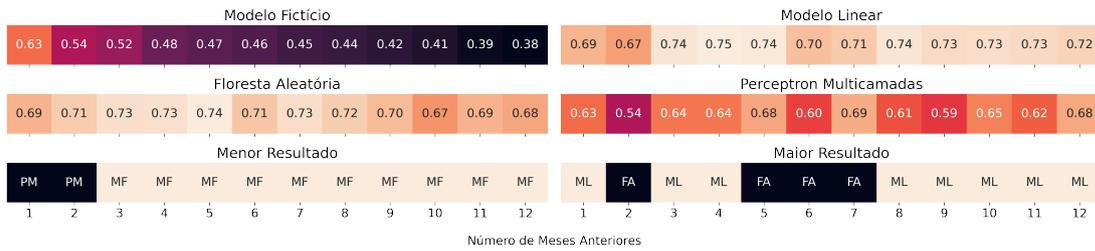


Figura 5.26: Resultado e comparação das medidas F_1 mínimas dos regressores com número de agrupamentos dinâmico.

Medida F_1 (média)

A figura 5.27 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Floresta Aleatória (FA) possui resultados levemente inferiores ao modelo com Regressão Linear (ML), que possui a maior parte dos maiores resultados. Assim, o modelo com Regressão Linear possui a maior medida F_1 , de forma geral.

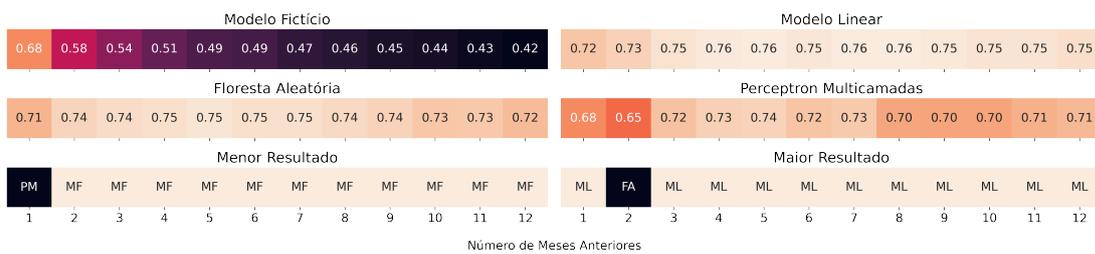


Figura 5.27: Resultado e comparação das medidas F_1 médias dos regressores com número de agrupamentos dinâmico.

Medida F_1 (desvio padrão)

A figura 5.28 mostra que o modelo com Floresta Aleatória (FA) e Regressão Linear (ML) compartilham os menores resultados, com uma pequena vantagem para o modelo com Regressão Linear. O modelo com Perceptron Multicamadas representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Linear possui a medida F_1 mais estável durante o ano.

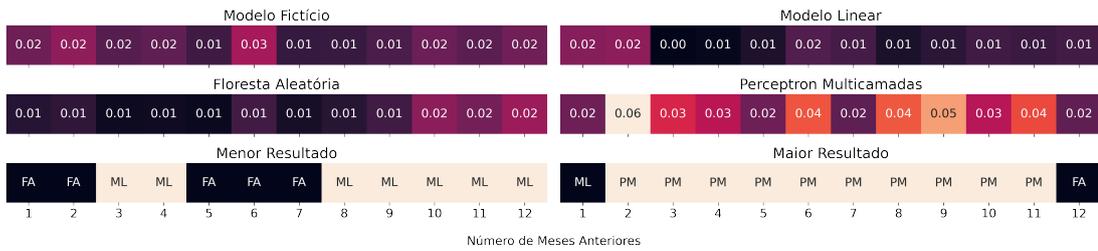


Figura 5.28: Resultado e comparação dos desvios padrão das medidas F_1 dos regressores com número de agrupamentos dinâmico.

5.4.3 Complexidade Computacional

Complexidade Computacional (pior caso)

A figura 5.29 mostra que o modelo com Regressão Linear (ML) representa todos os menores resultados, enquanto que o modelo com Floresta Aleatória (FA) representa todos os maiores resultados. Portanto, o modelo com Regressão Linear possui o menor tempo de treinamento, no pior caso.

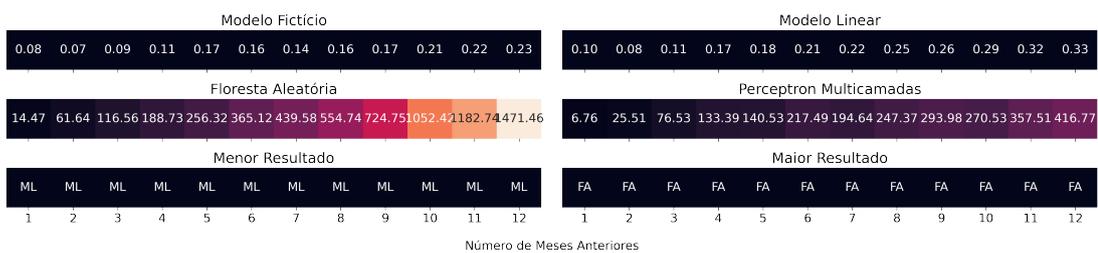


Figura 5.29: Resultado e comparação dos tempos de treinamento máximo dos regressores com número de agrupamentos dinâmico.

Complexidade Computacional (média)

A figura 5.30 mostra que o modelo com Regressão Linear (ML) representa todos os menores resultados, enquanto que o modelo com Floresta Aleatória (FA) representa todos os maiores resultados. Logo, o modelo com Regressão Linear possui o menor tempo de treinamento, de forma geral.



Figura 5.30: Resultado e comparação dos tempos de treinamento médio dos regressores com número de agrupamentos dinâmico.

Complexidade Computacional (desvio padrão)

A figura 5.31 mostra que o modelo com Regressão Linear (ML) representa todos os menores resultados, enquanto que os modelos com Floresta Aleatória (FA) e Perceptron Multicamadas (PM) compartilham os maiores resultados. Portanto, o modelo com Regressão Linear possui o tempo de treinamento mais estável.

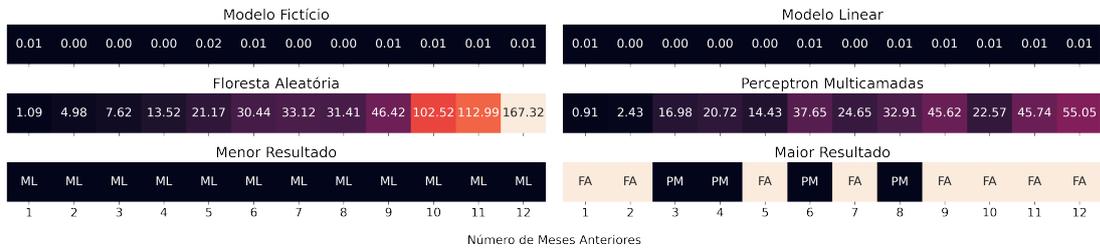


Figura 5.31: Resultado e comparação dos desvios padrão dos tempos de treinamento dos regressores com número de agrupamentos dinâmico.

5.5 Classificadores (agrupamento dinâmico)

5.5.1 Acurácia

Acurácia (pior caso)

A figura 5.32 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Regressão Logística (ML) possui resultados semelhantes ao modelo com Perceptron Multicamadas, mas este último representa a maior parte dos maiores resultados. Portanto, o modelo com Perceptron Multicamadas possui a melhor acurácia, no pior caso.

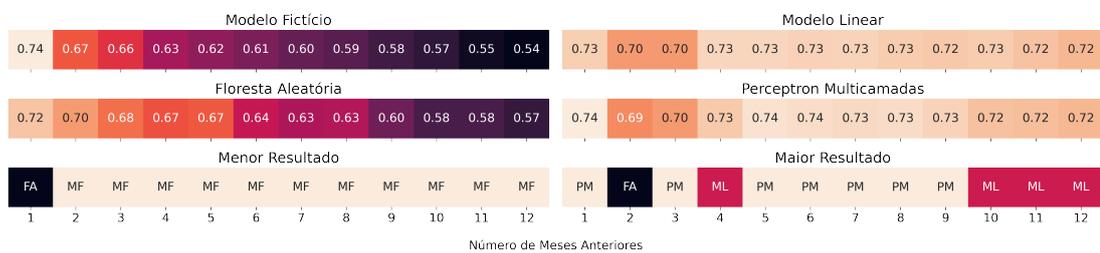


Figura 5.32: Resultado e comparação das acurácias mínimas dos classificadores com número de agrupamentos dinâmico.

Acurácia (média)

A figura 5.33 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Regressão Logística (ML) possui resultados semelhantes ao modelo com Perceptron Multicamadas, e ambos os modelos compartilham igualmente os maiores resultados. Assim, os modelos com Regressão Logística e Perceptron Multicamadas possuem as melhores acurácias, de forma geral.

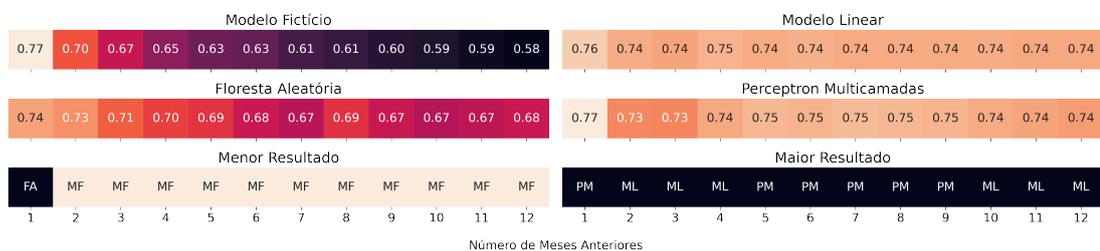


Figura 5.33: Resultado e comparação das acurácias médias dos classificadores com número de agrupamentos dinâmico.

Acurácia (desvio padrão)

A figura 5.34 mostra que o modelo com Perceptron Multicamadas (PM) possui resultados semelhantes ao modelo com Regressão Logística (ML), com este último representando a maior parte dos menores resultados. O modelo com Floresta Aleatória representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Logística possui a acurácia mais estável.

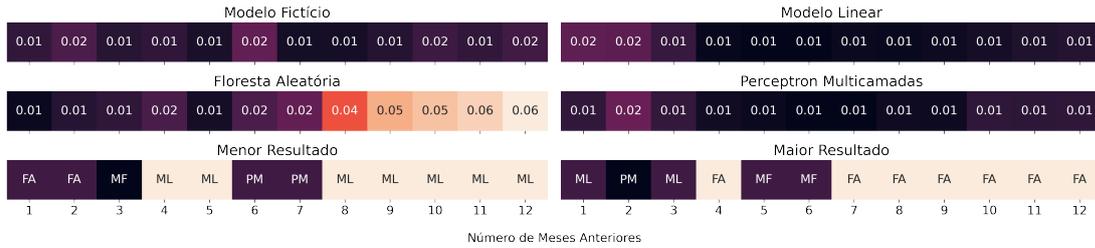


Figura 5.34: Resultado e comparação dos desvios padrão das acurácias dos classificadores com número de agrupamentos dinâmico.

5.5.2 Medida F_1

Medida F_1 (piores resultados)

A figura 5.35 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Regressão Logística (ML) possui resultados semelhantes ao modelo com Perceptron Multicamadas (PM), que possui resultados melhores e representa a maior parte dos maiores resultados. Logo, o modelo com Perceptron Multicamadas possui a maior medida F_1 , no pior caso.

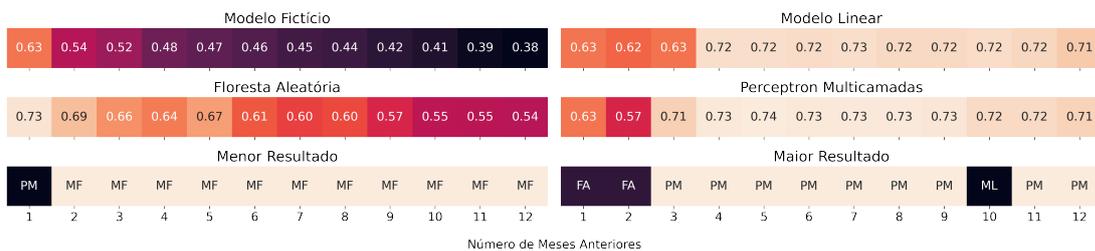


Figura 5.35: Resultado e comparação das medidas F_1 mínimas dos classificadores com número de agrupamentos dinâmico.

Medida F_1 (média)

A figura 5.36 mostra que o modelo Fictício (MF) representa a maior parte dos menores resultados. O modelo com Regressão Logística (ML) possui resultados semelhantes ao modelo com Perceptron Multicamadas (PM), e este último representa a maior parte dos maiores resultados. Portanto, o modelo com Perceptron Multicamadas possui a maior medida F_1 , de forma geral.

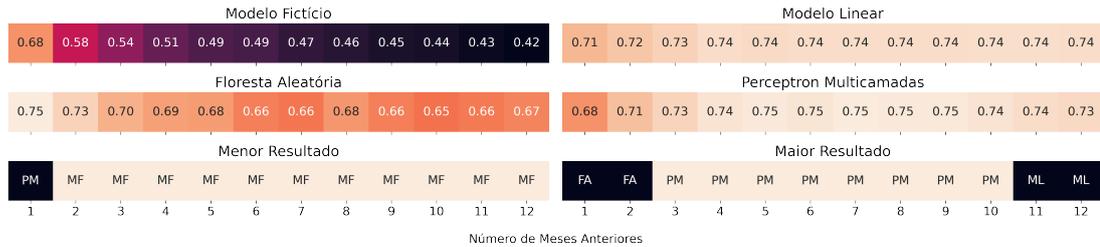


Figura 5.36: Resultado e comparação das medidas F_1 médias dos classificadores com número de agrupamentos dinâmico.

Medida F_1 (desvio padrão)

A figura 5.37 mostra que o modelo com Regressão Logística (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Floresta Aleatória (FA) representa a maior parte dos maiores resultados. Logo, o modelo com Regressão Logística possui as medidas F_1 mais estáveis.

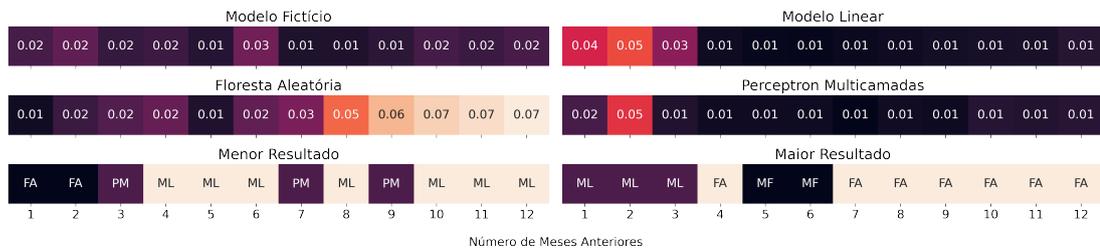


Figura 5.37: Resultado e comparação dos desvios padrão das medidas F_1 dos classificadores com número de agrupamentos dinâmico.

5.5.3 Complexidade Computacional

Complexidade Computacional (pior caso)

A figura 5.38 mostra que o modelo com Regressão Logística (ML) representa todos os menores resultados, enquanto que o modelo com Perceptron Multicamadas representa todos os maiores resultados. Assim, o modelo com Regressão Logística possui o menor tempo de treinamento, no pior caso.

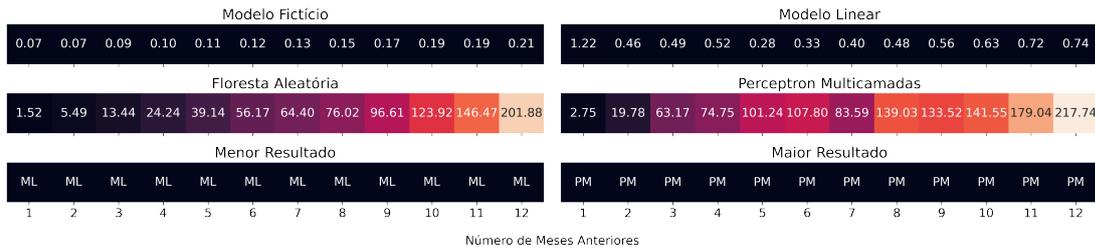


Figura 5.38: Resultado e comparação dos tempos de treinamento máximo dos classificadores com número de agrupamentos dinâmico.

Complexidade Computacional (média)

A figura 5.39 mostra que o modelo com Regressão Logística (ML) representa todos os menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa todos os maiores resultados. Logo, o modelo com Regressão Logística possui o menor tempo de treinamento, de forma geral.

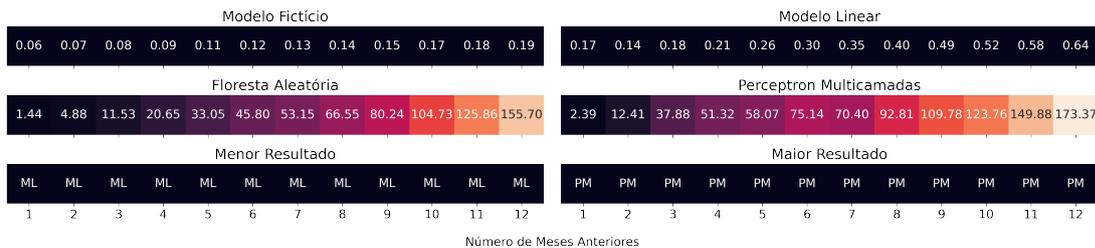


Figura 5.39: Resultado e comparação dos tempos de treinamento médio dos classificadores com número de agrupamentos dinâmico.

Complexidade Computacional (desvio padrão)

A figura 5.40 mostra que o modelo com Regressão Logística (ML) representa a maior parte dos menores resultados, enquanto que o modelo com Perceptron Multicamadas (PM) representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Logística possui o tempo de treinamento mais estável.



Figura 5.40: Resultado e comparação dos desvios padrão dos tempos de treinamento dos classificadores com número de agrupamentos dinâmico.

5.6 Melhor abordagem (agrupamento dinâmico)

Os comparativos das seções 5.4 e 5.5 mostram que os melhores resultados são obtidos pelos modelos com Regressão Linear (entre os regressores) e Regressão Logística e Perceptron Multicamadas (entre os classificadores). O modelo com Regressão Logística possui o melhor tempo de treinamento, enquanto que o modelo com Perceptron Multicamadas possui a melhor acurácia e medida F_1 . Esta seção irá comparar os melhores resultados para definir a melhor abordagem, com número dinâmico de agrupamentos.

5.6.1 Acurácia

A figura 5.41 mostra que o modelo com Perceptron Multicamadas (C) representa a maior parte dos menores resultados, enquanto que o modelo com Regressão Linear (R) representa a maior parte dos maiores resultados. Portanto, o modelo com Regressão Linear possui a maior acurácia, de forma geral.

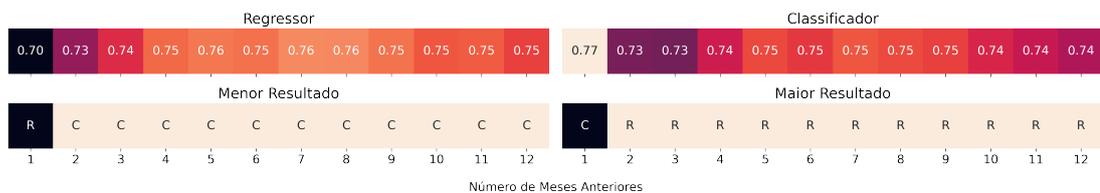


Figura 5.41: Resultado e comparação das acurácias médias do regressor e classificador, com número de agrupamentos dinâmico.

5.6.2 Medida F_1

A figura 5.42 mostra que o modelo com Perceptron Multicamadas (C) representa todos os menores resultados, enquanto que o modelo com Regressão Linear (R) representa todos os maiores resultados. Logo, o modelo com Regressão Linear possui a maior medida F_1 , de forma geral.

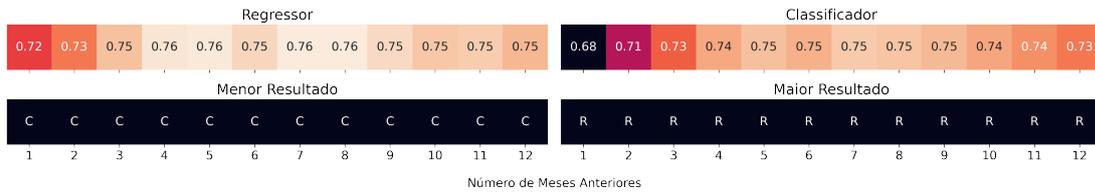


Figura 5.42: Resultado e comparação das medidas F_1 médias do regressor e classificador, com número de agrupamentos dinâmico.

5.6.3 Complexidade Computacional

A figura 5.43 mostra que o modelo com Regressão Logística (C) representa todos os maiores resultados, enquanto que o modelo com Regressão Linear (R) representa todos os menores resultados. Assim, o modelo com Regressão Linear possui o menor tempo de treinamento, de forma geral.

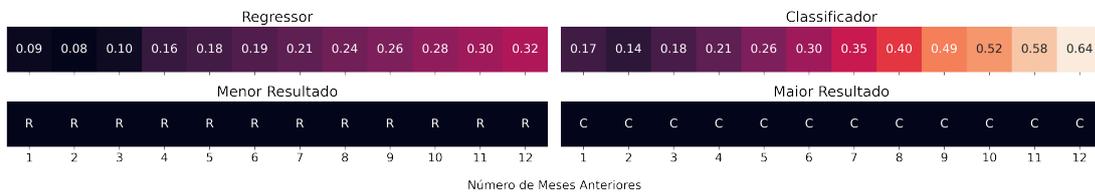


Figura 5.43: Resultado e comparação dos tempos de treinamento médio do regressor e classificador, com número de agrupamentos dinâmico.

Capítulo 6

Discussão dos Resultados

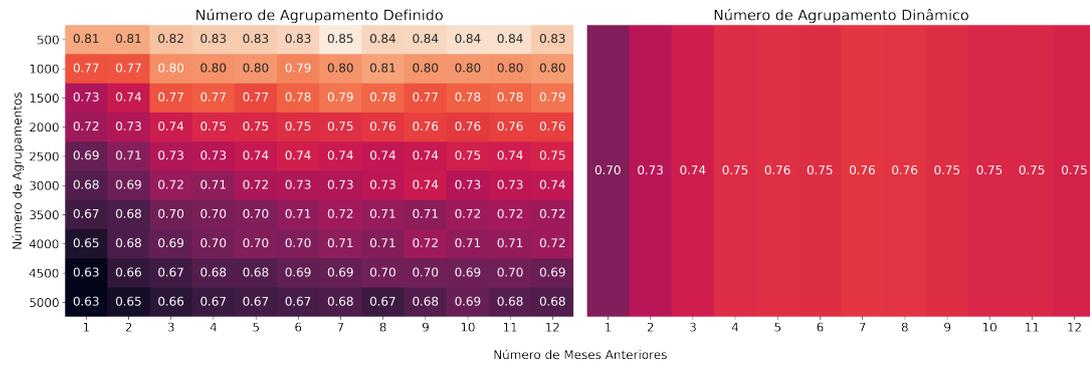
Este capítulo analisa e discute os resultados dos experimentos.

6.1 Modelo de Previsão

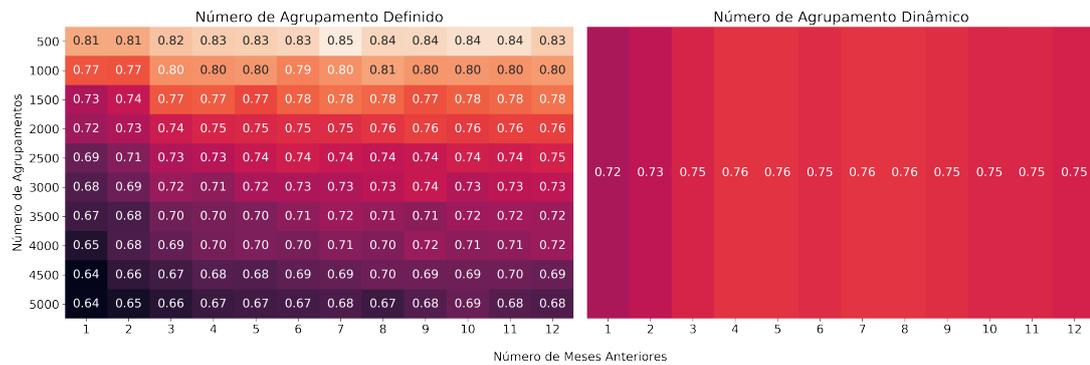
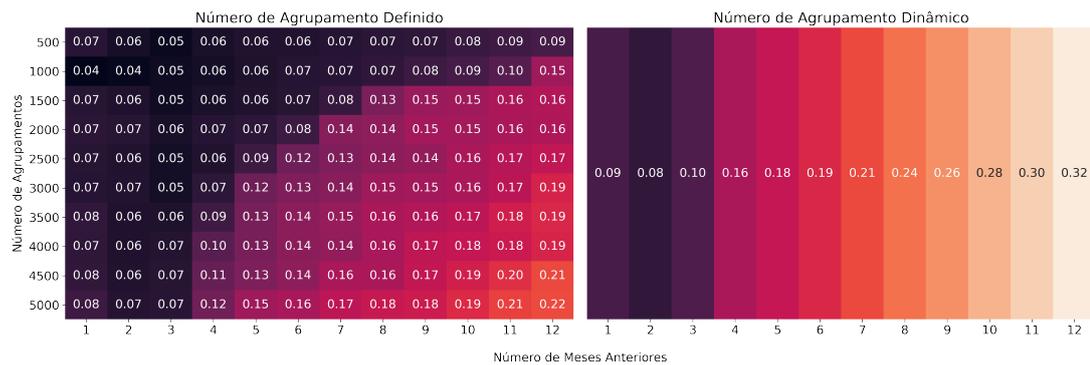
Os modelos criados com algoritmos Lineares, Florestas Aleatórias e Perceptron Multi-camadas apresentaram resultados melhores do que os modelos fictícios (linha de base), confirmando a habilidade preditiva dos modelos criados com algoritmos de aprendizado de máquina, como no trabalho de Bogomolov (BOGOMOLOV *et al.*, 2014).

Nas duas abordagens de agrupamentos, os modelos criados com Regressão Linear obtiveram os melhores resultados na previsão dos pontos críticos de crime para todos os meses de 2019 e também possuem o menor custo computacional, de acordo com as seções 5.3 e 5.6. A figura 6.1 mostra os resultados médios das acurácias, medidas F_1 e tempos de treinamentos.

Com os resultados das acurácias e medidas F_1 para o modelo com número definido de agrupamentos, é possível observar que o desempenho diminui em função do aumento do número de agrupamentos, demonstrando comportamento similar ao trabalho de Yu (YU *et al.*, 2011), descrito no capítulo 3. Assim, o modelo alcança desempenho (máximo) de 85% para a acurácia e medida F_1 com uso de dados de 7 meses anteriores e 500 agrupamentos (mínimo). Porém, números de agrupamentos menores particionam a região do estudo em poucos grupos, resultando em áreas maiores, mais generalistas e menos informativos. Portanto, é necessário escolher um número de tal modo que o resultado seja informativo, e com desempenho satisfatório.



(a) Resultados das acurácias médias dos modelos

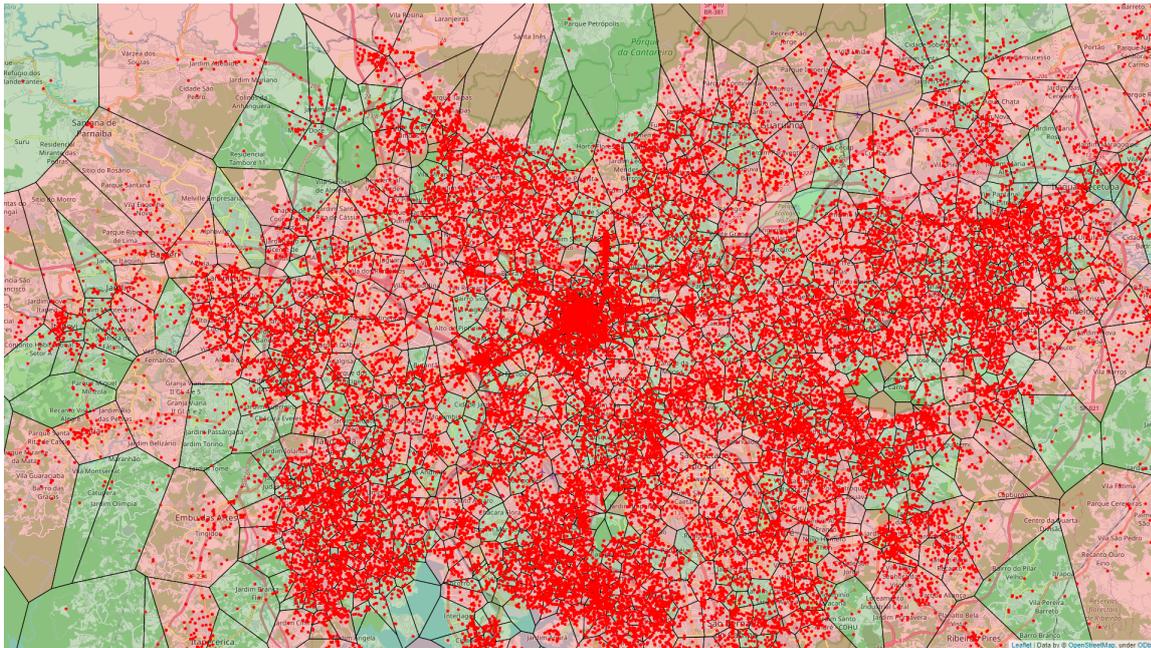
(b) Resultados das medidas F_1 médias dos modelos

(c) Resultados dos tempos de treinamentos médios dos modelos

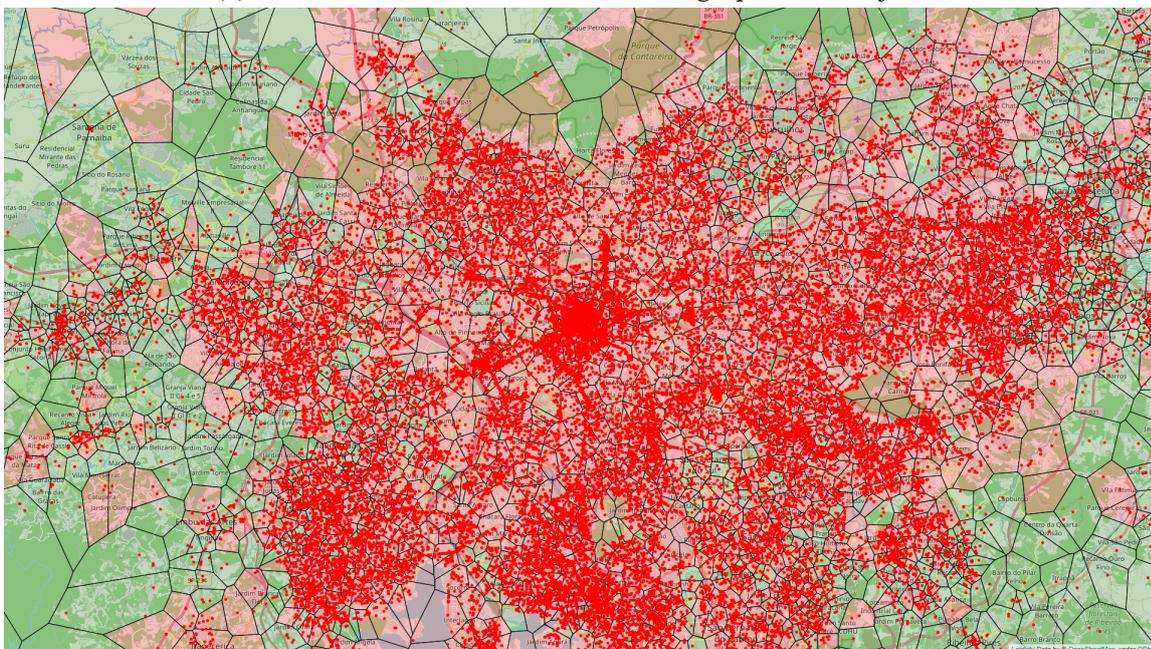
Figura 6.1: Resultados das duas abordagens de agrupamentos

6.2 Modelo de Agrupamento

Os resultados das acurácias e medidas F_1 para o modelo com número dinâmico de agrupamentos equivalem ao modelo com 2000 agrupamentos (definido). A figura 6.2 mostra a visualização das previsões do mês de Janeiro de 2020, da região metropolitana de São Paulo, com dados dos 4 meses anteriores (Setembro, Outubro, Novembro e Dezembro de 2019) para as duas abordagens de agrupamento:



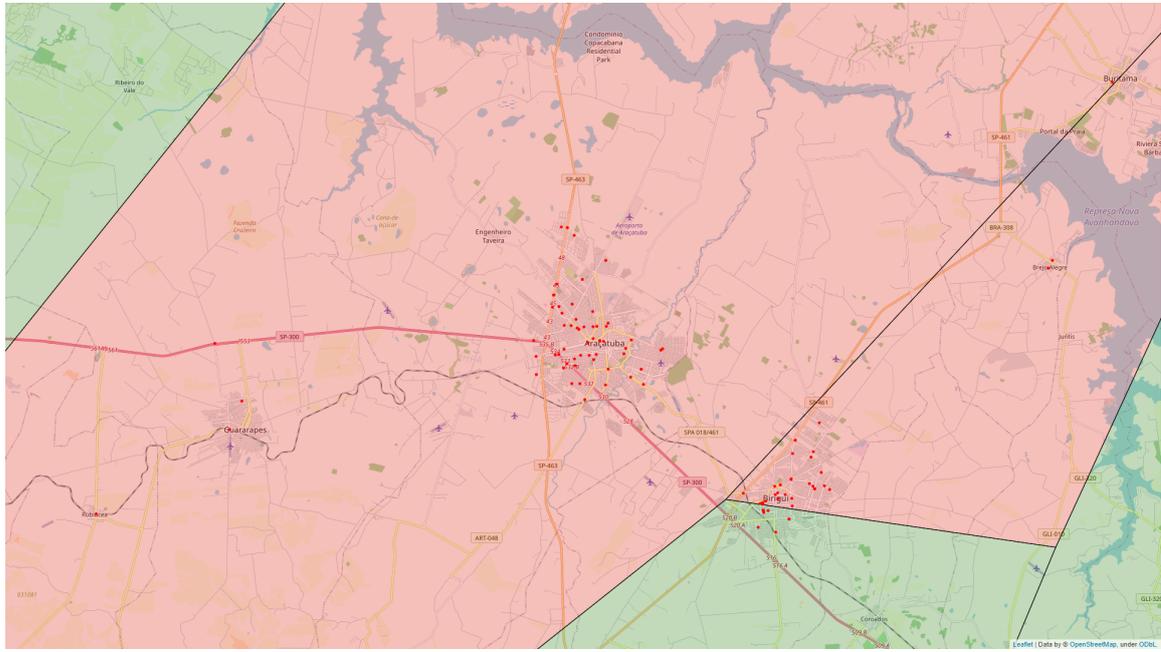
(a) Resultado do modelo com número de agrupamentos definido



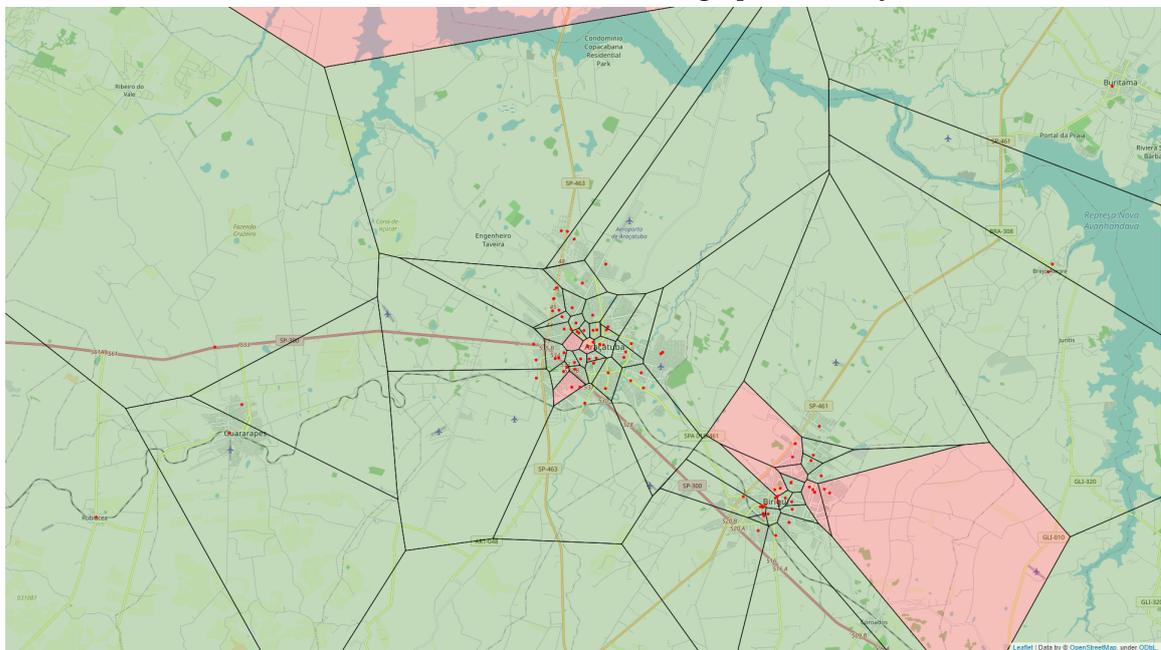
(b) Resultado do modelo com número de agrupamentos dinâmico

Figura 6.2: Resultados dos modelos na região metropolitana de São Paulo.

Os pontos vermelhos representam roubos ocorridos nos 4 meses, e as células *Voronoi* vermelhas representam os pontos críticos de crime. A região da capital paulista possui células de tamanhos semelhantes, enquanto que as suas extensões possuem células maiores, na abordagem com 2000 agrupamentos (definido). A figura 6.3 mostra os resultados da região do interior do estado de São Paulo, onde esta diferença é aparente:



(a) Resultado do modelo com número de agrupamentos definido



(b) Resultado do modelo com número de agrupamentos dinâmico

Figura 6.3: Resultados dos modelos na região interior do estado de São Paulo.

A abordagem com 2000 agrupamentos (definido) apresenta células maiores, classificando uma cidade inteira como um ponto crítico de crime, de forma generalista e pouco informativo. Já a abordagem com o número dinâmico de agrupamentos apresenta células menores, com resultados mais específicos e, portanto, mais informativo. A diferença está relacionada com a posição de inicialização dos centróides.

No primeiro caso (número de agrupamentos definido), cada centróide é inicializado na posição de alguma amostra, escolhida aleatoriamente. Probabilisticamente, a maior parte das posições iniciais dos centróides são distribuídas nas regiões com maior densidade de amostras (crimes), correspondentes às grandes cidades, e o restante realizam os agrupamentos de grandes regiões. Já no segundo caso (número de agrupamentos dinâmico), elas são inicializadas em qualquer bairro afetado com pelo menos um crime, independente da cidade e de maneira igualitária. Este comportamento pode ser visualizado nos exemplos da seção 4.3.

Outro fator que implica na quantidade de informação é o número de agrupamentos do modelo. Os números médios para a abordagem com número dinâmico de agrupamentos é mostrada na figura 6.4:

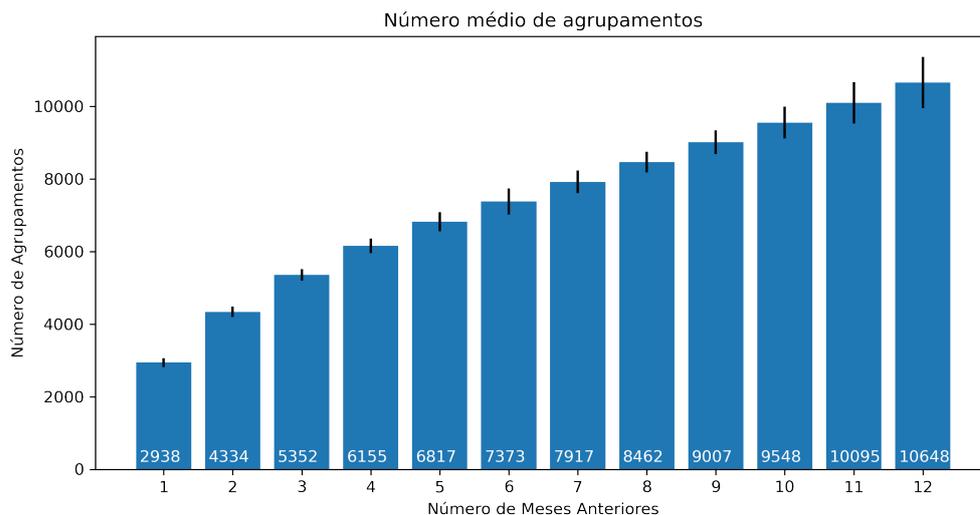


Figura 6.4: Número médio de agrupamentos para a abordagem de agrupamentos dinâmico.

Os números de agrupamentos (dinâmico) variam entre 2938 a 10648, dependendo da quantidade de dados dos meses anteriores, e muito superiores ao modelo com 2000 agrupamentos (definido). Como ambas as abordagens possuem acurácias e medidas F_1 em torno de 75%, o modelo com número de agrupamentos dinâmico é a melhor opção nesta faixa de desempenho. O modelo com número definido de agrupamentos seria a opção para desempenhos superiores a esta faixa, em detrimento da quantidade de informação dos resultados.

6.3 Desempenho Geral

Apesar de que a comparação justa entre os modelos dos **trabalhos relacionados** (incluindo este) não é possível, pois os modelos utilizam conjunto de dados distintos, é possível ter uma noção do desempenho de uma forma geral. A tabela 6.1 mostra os resultados e características dos modelos dos trabalhos:

Trabalhos Relacionados	Particionamento da Região		Desempenho		Classes Balanceadas
	Quantidade	Tipo	Acurácia	Medida F_1	
Yu	< 480	Grade Fixa	91%	75%	Não
Bogomolov	< 124119	Grade Fixa	70%	67%	Sim
Araújo	70	Polígono	80%	-	Não
Este trabalho	500 a 10648	Polígono	75% a 85%	75% a 85%	Sim

Tabela 6.1: Resultados dos trabalhos relacionados.

No contexto de particionamento da região, números menores aumentam o desempenho e reduzem a capacidade informativa. Nos trabalhos com particionamento em formato de grade fixa, a quantidade de células removidas, onde crimes não ocorrem, não foram mencionadas. Assim, o número de células utilizadas na modelagem é desconhecido, porém menor do que a quantidade total de células. Já nos trabalhos com particionamento em formato de polígono, as células se ajustam dinamicamente aos dados e todas são utilizadas na modelagem.

No contexto de desempenho, a acurácia é favorecida quando a distribuição das classes não são balanceadas e, devido ao **paradoxo da acurácia**, a medida F_1 torna-se mais confiável nesse caso. Apenas um trabalho não utilizou esta medida, mesmo com a distribuição não balanceada das classes.

O modelo desenvolvido neste trabalho alcança desempenho (acurácia e medida F_1) máximo de 85% com a quantidade mínima de agrupamentos e 75% com a quantidade máxima, sempre com a distribuição balanceada de classes. O desempenho geral é superior a dois dos três trabalhos relacionados, para cada contexto.

Capítulo 7

Conclusão

O objetivo deste trabalho foi o desenvolvimento e avaliação de um modelo capaz de prever os pontos críticos de crime (áreas com alto índice de crimes). A modelagem foi feita por aprendizado de máquina, utilizando os algoritmos *K-Means* (não supervisionado) e *Regressão Linear* (supervisionado), alimentados com os dados dos boletins de ocorrência de roubo de celulares disponíveis no portal da transparência da Secretaria da Segurança Pública de São Paulo. A construção foi guiada pelas metodologias dos trabalhos relacionados, e obteve resultados bem satisfatórios, com desempenho em torno de 75% (alcançando até 85%) na previsão dos pontos críticos de crimes no Estado de São Paulo.

Os modelos criados por aprendizado de máquina apresentaram resultados melhores do que os modelos fictícios (linha de base), confirmando a habilidade preditiva da modelagem por este tipo de técnica. O modelo de previsão deste trabalho pode se tornar a nova linha de base para experimentos futuros.

A modelagem, de modo geral, é simples, fácil de entender e os algoritmos de aprendizado de máquina utilizados são muito conhecidos e possuem baixo custo computacional. A implementação pode ser realizada por qualquer linguagem de programação com biblioteca de aprendizado de máquina, como *Python*, *R*, *Javascript* e outros.

A implementação do modelo pode beneficiar a prática do policiamento preditivo, permitindo a alocação eficiente de recursos policiais, que são limitados. A presença ostensiva da polícia nos pontos críticos de crimes pode dificultar a ação dos criminosos e reduzir o número de crimes. Outro benefício da implementação do modelo seria de caráter informativo para alertar a população, como os cartazes instalados pela Polícia Militar em áreas de risco de roubo de celular, discutido na seção 1.1. A previsão dos pontos críticos de crime abre a possibilidade de evitar esses locais ou auxiliar na tomada de decisão de utilizar o celular, mesmo que a área tenha um alto índice de crime.

Referências

- [ALBON 2018] Chris ALBON. *Machine Learning with Python Cookbook*. O’Reilly Media, Inc., 2018. ISBN: 9781491989388 (citado na pg. 16).
- [ARAÚJO *et al.* 2018] A. ARAÚJO, N. CACHO, L. BEZERRA, C. VIEIRA e J. BORGES. “Towards a crime hotspot detection framework for patrol planning”. Em: *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. 2018, pgs. 1256–1263. DOI: [10.1109/HPCC/SmartCity/DSS.2018.00211](https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00211) (citado nas pgs. 18, 26, 27, 31).
- [BAND 2020a] Jornal da BAND. *SP: Polícia alerta para roubos de celulares no Centro*. 2020. URL: <https://videos.band.uol.com.br/16743297/sp-policia-alerta-para-roubos-de-celulares-no-centro.html> (acesso em 06/04/2020) (citado na pg. 1).
- [BAND 2020b] Jornal da BAND. *SP: Polícia descobre comércio clandestino de celulares*. 2020. URL: <https://videos.band.uol.com.br/16770853/sp-policia-descobre-comercio-clandestino-de-celulares.html> (acesso em 06/04/2020) (citado na pg. 1).
- [BAND 2020c] Jornal da BAND. *SP: Polícia usa cartazes para alertar população sobre roubos*. 2020. URL: <https://videos.band.uol.com.br/16743263/sp-policia-usa-cartazes-para-alertar-populacao-sobre-roubos.html> (acesso em 06/04/2020) (citado na pg. 1).
- [BERNASCO e BLOCK 2011] Wim BERNASCO e Richard BLOCK. “Robberies in chicago: a block-level analysis of the influence of crime generators, crime attractors, and offender anchor points”. Em: *Journal of Research in Crime and Delinquency* 48 (fev. de 2011), pgs. 33–57. DOI: [10.1177/0022427810384135](https://doi.org/10.1177/0022427810384135) (citado na pg. 17).
- [BOGOMOLOV *et al.* 2014] Andrey BOGOMOLOV *et al.* “Once upon a crime: towards crime prediction from demographics and mobile data”. Em: *Proceedings of the 16th International Conference on Multimodal Interaction*. ICMI ’14. Istanbul, Turkey: Association for Computing Machinery, 2014, pgs. 427–434. ISBN: 9781450328852. DOI: [10.1145/2663204.2663254](https://doi.org/10.1145/2663204.2663254). URL: <https://doi.org/10.1145/2663204.2663254> (citado nas pgs. 18, 28, 30, 31, 69).

- [Box 2019] Panorama Mobile Time / Opinion Box. *Roubo de Celulares no Brasil*. 2019. URL: <https://panoramamobiletime.com.br/pesquisa-celulares-roubados-julho-de-2019/> (acesso em 06/04/2020) (citado na pg. 1).
- [GÉRON 2019] Aurélien GÉRON. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2nd. O'Reilly Media, Inc., 2019. ISBN: 9781492032649 (citado nas pgs. 8, 9, 11, 13, 14).
- [MONTGOMERY *et al.* 2015] Douglas C. MONTGOMERY, Cheryl L. JENNINGS e Murat KULAHCI. *Introduction to Time Series Analysis and Forecasting*. 2nd. Wiley, 2015. ISBN: 9781118745151 (citado na pg. 5).
- [JN 2016] Jornal NACIONAL. *JN flagra ação de ladrões em viaduto de São Paulo*. 2016. URL: <https://globoplay.globo.com/v/5400197/> (acesso em 06/04/2020) (citado na pg. 1).
- [OKABE *et al.* 2000] Atsuyuki OKABE, Barry BOOTS, Kokichi SUGIHARA, Sung Nok CHIU e D. G. KENDALL. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. 2nd. Wiley, 2000. ISBN: 9780470317013 (citado na pg. 6).
- [RUSSELL e NORVIG 2009] Stuart RUSSELL e Peter NORVIG. *Artificial Intelligence: A Modern Approach*. 3rd. USA: Prentice Hall Press, 2009. ISBN: 0136042597 (citado na pg. 9).
- [SAMUEL 1959] A. L. SAMUEL. “Some studies in machine learning using the game of checkers”. Em: *IBM Journal of Research and Development* 3.3 (1959), pgs. 210–229. DOI: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210) (citado na pg. 7).
- [SCULLEY 2010] D. SCULLEY. “Web-scale k-means clustering”. Em: WWW '10. Raleigh, North Carolina, USA: Association for Computing Machinery, 2010, pgs. 1177–1178. ISBN: 9781605587998. DOI: [10.1145/1772690.1772862](https://doi.org/10.1145/1772690.1772862). URL: <https://doi.org/10.1145/1772690.1772862> (citado na pg. 13).
- [SSP-SP 2020] Secretaria da SEGURANÇA PÚBLICA DE SÃO PAULO. *Transparência*. URL: <https://www.ssp.sp.gov.br/transparenciassp/> (acesso em 06/04/2020) (citado na pg. 23).
- [SSP-SP 2019] Secretaria da SEGURANÇA PÚBLICA DE SÃO PAULO. *Distribuição das ocorrências de roubo no estado de São Paulo, segundo o contexto, o tipo e o subtipo de objeto roubado*. 2019. URL: http://www.ssp.sp.gov.br/estatistica/perf_roubo/2019.htm (acesso em 06/04/2020) (citado na pg. 1).
- [SP1 2017] SPTV. *SP1 flagra venda de celulares roubados em feiras*. 2017. URL: <https://globoplay.globo.com/v/6114067/> (acesso em 06/04/2020) (citado na pg. 1).
- [SP1 2018a] SPTV. *Em menos de dez dias, três casos de roubo de celular terminam em morte na Grande São Paulo*. 2018. URL: <https://globoplay.globo.com/v/6957569/> (acesso em 06/04/2020) (citado na pg. 1).

REFERÊNCIAS

- [SP1 2018b] SPTV. *Polícia tenta identificar assassinos do estudante Robert Henrique*. 2018. URL: <https://globoplay.globo.com/v/6943896/> (acesso em 06/04/2020) (citado na pg. 1).
- [SP2 2018] SPTV. *São Paulo registra mais de 1.200 roubos e furtos de celular por dia*. 2018. URL: <https://globoplay.globo.com/v/6633037/> (acesso em 06/04/2020) (citado na pg. 1).
- [SP2 2020] SPTV. *Quadrilha que roubava celulares em São Paulo é presa*. 2020. URL: <https://globoplay.globo.com/v/8996397/> (acesso em 11/04/2020) (citado na pg. 1).
- [Yu *et al.* 2011] C. YU, M. W. WARD, M. MORABITO e W. DING. “Crime forecasting using data mining techniques”. Em: *2011 IEEE 11th International Conference on Data Mining Workshops*. 2011, pgs. 779–786. DOI: [10.1109/ICDMW.2011.56](https://doi.org/10.1109/ICDMW.2011.56) (citado nas pgs. 18, 27, 31, 69).
- [ZIVIANI 2004] Nivio ZIVIANI. *Projeto de algoritmos: com implementações em Pascal e C*. Vol. 2. 2004 (citado na pg. 6).