

# Grupo 3

Lucas de Miranda Oliveira<sup>1</sup>  
Pedro Henrique Sebe Rodrigues<sup>2</sup>  
Renata Massami Hirota<sup>3</sup>  
Rodolfo Riani Sundfeld<sup>4</sup>  
Rubens Santos Andrade Filho<sup>5</sup>

Setembro de 2020

## Sumário

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Descrição do Problema . . . . .</b> | <b>2</b> |
| <b>2</b> | <b>Conjunto de dados . . . . .</b>     | <b>2</b> |
| <b>3</b> | <b>Avaliação do problema . . . . .</b> | <b>3</b> |
| <b>4</b> | <b>Endereço . . . . .</b>              | <b>3</b> |
| <b>5</b> | <b>Referências . . . . .</b>           | <b>3</b> |

---

<sup>1</sup>Número USP: 11577209

<sup>2</sup>Número USP:

<sup>3</sup>Número USP: 7165654

<sup>4</sup>Número USP:

<sup>5</sup>Número USP: 10370336

# 1 Descrição do Problema

Brasil tem segunda maior taxa de homicídios da América do Sul, segundo relatório das Nações Unidas em 2017, e ocupa a 12<sup>o</sup> posição no ranking global de violência, com uma taxa de 27,5% de homicídios a cada 100 000 habitantes. Em 2016, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) que as vítimas de assaltos no Brasil, nos últimos 12 meses a pesquisa, foram de 7,9% dos entrevistados, quase o dobro da média dos países entrevistados. Este preocupante quadro mostra a situação da violência e a necessidade de políticas públicas adequadas para segurança pública.

A partir de dados de boletins de ocorrência (BO) da Secretaria de Segurança Pública (SSP) do estado de São Paulo, o objetivo deste estudo é prever o horário do crime de acordo com algumas características, por exemplo, localização, sexo da vítima, cor da vítima, entre outros. Visando, dessa forma, identificar quais os horários das ocorrências de acordo com a tipologia do crime e assim fornecer insights para o policiamento.

# 2 Conjunto de dados

A base de dados que será trabalhado refere-se a boletins de ocorrência na Grande São Paulo.

| Variável                | Descrição   |
|-------------------------|---|
| ID_DELEGACIA            | Código da delegacia responsável pelo registro da ocorrência |
| NOME_DEPARTAMENTO       | Departamento responsável pelo registro                      |
| NOME_SECCIONAL          | Delegacia Seccional responsável pelo registro               |
| NOME_DELEGACIA          | Delegacia responsável pelo registro                         |
| CIDADE                  | Cidade de Registro  |
| ANO_BO                  | Ano da ocorrência   |
| NUM_BO                  | Número do BO  |
| NOME_DEPARTAMENTO_CIRC  | Departamento de Circunscrição                               |
| NOME_SECCIONAL_CIRC     | Seccional de Circunscrição                                  |
| NOME_DELEGACIA_CIRC     | Delegacia de Circunscrição                                  |
| NOME_MUNICIPIO_CIRC     | Município de Circunscrição                                  |
| DESCR_TIPO_BO           | Tipo de Documento   |
| DATA_OCORRENCIA_BO      | Data da Ocorrência  |
| HORA_OCORRENCIA_BO      | Hora da Ocorrência  |
| DATAHORA_COMUNICACAO_BO | Data Hora da Comunicação da Ocorrência                      |
| FLAG_STATUS             | Status da Ocorrência  |
| RUBRICA                 | Natureza jurídica da ocorrência                             |
| DESCR_CONDUTA           | Conduta na Ocorrência                                       |
| DESDOBRAMENTO           | Desdobramento na Ocorrência                                 |
| DESCR_TIPOLOCAL         | Tipo de Local   |
| DESCR_SUBTIPOLOCAL      | Descrição do subTipo de local                               |
| LOGRADOURO              | Logradouro dos fatos  |
| NUMERO_LOGRADOURO       | Numero do Logradouro dos fatos                              |
| LATITUDE                | Latitude da Ocorrência                                      |
| LONGITUDE               | Longitude da Ocorrência                                     |
| DESCR_TIPO_PESSOA       | Qualificação do envolvido na ocorrência                     |
| FLAG_VITIMA_FATAL       | Condição do Autor / Vítima na ocorrência                    |
| SEXO_PESSOA             | Sexo  |
| IDADE_PESSOA            | Idade   |
| COR_CUTIS               | Cor da Pele   |

### 3 Avaliação do problema

A variável HORA\_OCORRENCIA\_BO será transformada de forma a produzir quatro classes de acordo com o horário do crime: manhã, tarde, noite e madrugada. O problema é de classificação.

Os métodos de aprendizagem supervisionada para classificação utilizados serão:

- Multinomial logistic regression;
- Support Vector Machine Multiclass;
- K-Nearest Neighbor(KNN) Classification;
- Feedforward Neural Network For Multiclass Classification;
- Random Forest;
- Boosting;
- Bagging;
- Gradient Boosting.

### 4 Endereço

A base dados que será trabalhada é a Crime Data in Brazil e pode ser acessada **aqui**.

### 5 Referências

- BREIMAN, Leo. Bagging predictors. Machine learning, v. 24, n. 2, p. 123-140, 1996.
- BREIMAN, Leo. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.
- FRIEDMAN, Jerome H. Stochastic gradient boosting. Computational statistics & data analysis, v. 38, n. 4, p. 367-378, 2002.
- HEARST, Marti A.. et al. Support vector machines. IEEE Intelligent Systems and their applications, v. 13, n. 4, p. 18-28, 1998
- SVOZIL, Daniel; KVASNICKA, Vladimir; POSPICHAL, Jiri. Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, v. 39, n. 1, p. 43-62, 1997.