

# Aprendizagem Automática I

## Projeto de Avaliação

-

Mestrado em Engenharia Informática  
Universidade do Minho  
Relatório

### **Grupo**

---

PG41081	José Alberto Martins Boticas
PG41091	Nelson José Dias Teixeira

27 de Dezembro de 2019

## **Resumo**

Este projeto de avaliação relativo à unidade curricular de Aprendizagem Automática I consiste, globalmente, na aplicação de uma das técnicas abordadas durante as aulas sobre um conjunto de dados. O conjunto de dados mencionado previamente é escolhido sem qualquer tipo de restrição por parte dos elementos do grupo por forma a despoletar o interesse dos mesmos durante a análise estatística dos dados presentes. Como tal, durante a execução deste trabalho prático (cuja unidade curricular integra o perfil de Ciência de Dados), surge uma motivação extra na interpretação dos resultados obtidos.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Apresentação da base de dados escolhida . . . . .	2
1.2	Contextualização . . . . .	2
1.3	Definição das variáveis . . . . .	2
1.4	Objetivo de análise . . . . .	3
<b>2</b>	<b>Metodologia</b>	<b>5</b>
2.1	Modelo adoptado . . . . .	5
<b>3</b>	<b>Resultados</b>	<b>6</b>
<b>4</b>	<b>Conclusão</b>	<b>7</b>
<b>5</b>	<b>Webgrafia</b>	<b>8</b>

# Capítulo 1

## Introdução

### 1.1 Apresentação da base de dados escolhida

A base de dados escolhida pelos dois elementos que constituem este grupo diz respeito aos relatórios de incidentes criminosos reportados pelo departamento policial de Boston (BPD - *Boston Police Department*) desde 14 de Junho de 2015 até ao momento. Estes documentos registam os detalhes em torno de um determinado incidente que foi respondido pela polícia de Boston.

### 1.2 Contextualização

Hoje em dia grande parte dos cidadãos que integram a sociedade mundial questionam a intervenção da polícia nas suas cidades ou países. Esta incerteza reside no facto de não só a população considerar que a força policial não é suficientemente adequada para um determinado tipo de incidente criminoso como também a abordagem adotada pela mesma ser demasiado violenta. Consequentemente, surge a seguinte questão: "*De que forma é que podemos melhorar a intervenção das autoridades?*". Por forma a dar resposta a esta pergunta, vamos estudar as intervenções policiais na cidade de Boston, nos Estados Unidos da América, com o intuito de perceber em que tipo de ocorrências estes mais intervêm, verificando se a sua intervenção é ou não eficaz. Desta forma, é possível avaliar se de facto a polícia de Boston foi ou não correta no tratamento dos incidentes criminosos reportados e, consequentemente, afirmar se é até possível melhorar a sua intervenção.

### 1.3 Definição das variáveis

Quanto às incógnitas presentes na base de dados foi possível identificar tanto as variáveis quantitativas como as variáveis qualitativas ou categóricas. É exibido de seguida as mesmas:

- **INCIDENT\_NUMBER**: número do incidente;
- **OFFENSE\_CODE**: código do crime/incidente;
- **OFFENSE\_CODE\_GROUP**: nome do grupo/categoria associado(a) ao código do crime/incidente;
- **OFFENSE\_DESCRIPTION**: descrição associada ao código do crime/incidente;
- **DISTRICT**: distrito de Boston onde ocorreu o incidente;

- **REPORTING\_AREA**: número que identifica a área onde foi reportado o crime;
- **SHOOTING**: variável que indica se houve ou não um tiroteio num determinado crime;
- **OCCURRED\_ON\_DATE**: data e hora do incidente;
- **YEAR**: ano do incidente;
- **MONTH**: mês do incidente;
- **DAY\_OF\_WEEK**: dia da semana do incidente;
- **HOURL**: hora do incidente;
- **UCR\_PART**: *Universal Crime Reporting Part number*;
- **STREET**: rua onde ocorreu o crime.

Apresenta-se de seguida o tipo de cada uma das variáveis presentes:

Nome da variável	Tipo de variável
INCIDENT_NUMBER	Qualitativa ordinal
OFFENSE_CODE	Quantitativa discreta
OFFENSE_CODE_GROUP	Qualitativa nominal
OFFENSE_DESCRIPTION	Qualitativa nominal
DISTRICT	Qualitativa nominal
REPORTING_AREA	Quantitativa discreta
SHOOTING	Qualitativa ordinal
OCCURRED_ON_DATE	Qualitativa ordinal
YEAR	Quantitativa discreta
MONTH	Quantitativa discreta
DAY_OF_WEEK	Qualitativa nominal
HOURL	Quantitativa discreta
UCR_PART	Qualitativa nominal
STREET	Qualitativa nominal

De salientar que foram removidas 3 variáveis que representavam as coordenadas do local do incidente criminoso pois não acrescentavam grande interesse na análise estatística.

## 1.4 Objetivo de análise

Dado o enorme número de registos presentes na base de dados surgiram algumas perguntas ou curiosidades sobre as quais queremos tomar conhecimento. Apresentam-se de seguida as mesmas:

1. Quais os tipos de crime são mais comuns?
2. Onde e quando é que os diferentes tipos de crimes têm maior probabilidade de ocorrer?
3. A frequência dos crimes cometidos muda ao longo do dia? E ao longo da semana? E durante o ano?
4. Qual é a distribuição de ocorrências de incidentes criminosos observados por hora?

5. O dia da semana influencia a quantidade de ocorrências de crimes?

Consequentemente, por forma a responder a estas questões, é necessário especificar um modelo estatístico que se adequa a este contexto. Como tal, na próxima secção deste documento, é apresentado o modelo requerido.

## Capítulo 2

# Metodologia

### 2.1 Modelo adoptado

## Capítulo 3

# Resultados



## Capítulo 4

## Conclusão

## Capítulo 5

# Webgrafia

- *Website* indicado pela docente:  
*<https://www.kaggle.com/datasets>*
- *Website* com o resumo da base de dados escolhida:  
*<https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>*
- Informação oficial acerca da base de dados:  
*<https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>*
- Informação mais detalhada acerca do campo *UCR\_PART* presente na base de dados:  
*[https://en.wikipedia.org/wiki/Uniform\\_Crime\\_Reports](https://en.wikipedia.org/wiki/Uniform_Crime_Reports)*