

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

19 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

1	Introdução	2
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
2	Metodologia	4
2.1	Modelo adoptado	4
3	Resultados	5
4	Conclusão	6
5	Webgrafia	7
A	Observações	8

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 do incidente em questão e que foi respondido pelos polícias de Boston.

1.2 Contextualização

1.3 Definição das variáveis

Quanto às incógnitas presentes na base de dados foi possível identificar tanto variáveis quantitativas como 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***: (...) **ACABAR !**

- ***STREET***: rua onde ocorreu o crime.

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 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?

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/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>

Apêndice A

Observações