**Data Science for Business –**  2DTSR

Conceitos

Atividade 1:

Um analista tem objetivo de estudar os vinhos de uma certa região de Portugal. Ele está com uma amostra de vinhos para estudá-la. (Base descrita abaixo). A variável Quality é uma variável sensorial, isto é, alguém provou e deu uma nota para o vinho.

**Etapa 1:**

Com objetivo de estimar a nota de qualidade dos vinhos (variável “Quality”) em função das características físico-químicas dos vinhos.

1. Qual é a técnica adequada para essa resolver esse objetivo:

|  |  |
| --- | --- |
|  | Técnicas Supervisionadas |
|  | Técnicas Não Supervisionadas |

Explique a escolha:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Há variáveis dependentes (D) e independentes (I)? Quais:

|  |  |
| --- | --- |
| Variáveis Dependentes: |  |
| Variáveis Independentes: |  |

1. Qual (is) a tarefa(s) de Mining é (são) adequada (s) para solução desse objetivo:

|  |  |
| --- | --- |
|  | Predição |
|  | Classificação |
|  | Regras de Associação |
|  | Agrupamento |
|  | Sumarização |
|  | Detecção de Anomalias |

1. Você pode indicar alguns algoritmos possíveis para essa tarefa?

|  |  |
| --- | --- |
|  | Regressão Linear |
|  | Árvore de Decisão |
|  | Regras de Associação |
|  | Séries Temporais |
|  | Análise de Clusters |
|  | Componentes Principais |
|  | Redes Neurais |
|  | Random Forest, |
|  | Outras: |

1. Você pode indicar algumas medidas de acurácia para essa tarefa?

|  |  |
| --- | --- |
|  | Multiple R2 |
|  | Adjusted R2 |
|  | RMSE |
|  | MAPE |
|  | Matriz de confusão |
|  | Curva ROC |
|  | Medida AUC |
|  | Estatística KS |
|  | Outras: |

1. Quais cuidados que são indicados a fazer antes da etapa de aplicar os algoritmos/técnicas?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Etapa 2:**

Com o objetivo de classificar os vinhos dessa região em vinhos em “Bom” ou “Ruim” em função de suas características físico-químicas;

1. Qual é a técnica adequada para essa resolver esse objetivo:

|  |  |
| --- | --- |
|  | Técnicas Supervisionadas |
|  | Técnicas Não Supervisionadas |

Explique a escolha:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Há variáveis dependentes (D) e independentes (I)? Quais?

|  |  |
| --- | --- |
| Variáveis Dependentes: |  |
| Variáveis Independentes: |  |

1. Qual (is) a tarefa(s) de Mining é (são) adequada (s) para solução desse objetivo:

|  |  |
| --- | --- |
|  | Predição |
|  | Classificação |
|  | Regras de Associação |
|  | Agrupamento |
|  | Sumarização |
|  | Detecção de Anomalias |

1. Você pode indicar alguns algoritmos possíveis para essa tarefa?

|  |  |
| --- | --- |
|  | Regressão Linear |
|  | Árvore de Decisão |
|  | Regras de Associação |
|  | Séries Temporais |
|  | Análise de Clusters |
|  | Componentes Principais |
|  | Redes Neurais |
|  | Random Forest, |
|  | Outras: |

1. Você pode indicar algumas medidas de acurácia para essa tarefa?

|  |  |
| --- | --- |
|  | Multiple R2 |
|  | Adjusted R2 |
|  | RMSE |
|  | MAPE |
|  | Matriz de confusão |
|  | Curva ROC |
|  | Medida AUC |
|  | Estatística KS |
|  | Outras: |

1. Quais cuidados que são indicados a fazer antes da etapa de aplicar os algoritmos/técnicas?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Etapa 3:**

Uma importadora tem por objetivo fazer importação de vinhos dessa região a pergunta e contrata esse analista para ajuda-lo na importação : A pergunta dele é quantos diferentes tipos de vinhos encontramos essa região ou quantas tipologias de vinhos ele deve contemplar na importação?

1. Qual é a técnica adequada para essa resolver esse objetivo:

|  |  |
| --- | --- |
|  | Técnicas Supervisionadas |
|  | Técnicas Não Supervisionadas |

Explique a escolha:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Há variáveis dependentes (D) e independentes (I)? Quais?

|  |  |
| --- | --- |
| Variáveis Dependentes: |  |
| Variáveis Independentes: |  |

1. Qual (is) a tarefa(s) de Mining é (são) adequada (s) para solução desse objetivo:

|  |  |
| --- | --- |
|  | Predição |
|  | Classificação |
|  | Regras de Associação |
|  | Agrupamento |
|  | Sumarização |
|  | Detecção de Anomalias |

1. Você pode indicar alguns algoritmos possíveis para essa tarefa?

|  |  |
| --- | --- |
|  | Regressão Linear |
|  | Árvore de Decisão |
|  | Regras de Associação |
|  | Séries Temporais |
|  | Análise de Clusters |
|  | Componentes Principais |
|  | Redes Neurais |
|  | Random Forest, |
|  | Outras: |

1. Você pode indicar algumas medidas de acurácia para essa tarefa?

|  |  |
| --- | --- |
|  | Multiple R2 |
|  | Adjusted R2 |
|  | RMSE |
|  | MAPE |
|  | Matriz de confusão |
|  | Curva ROC |
|  | Medida AUC |
|  | Estatística KS |
|  | Outras: |

1. Quais cuidados que são indicados a fazer antes da etapa de aplicar os algoritmos/técnicas?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Atividade 2: Escolha uma das etapas para construir o script Pyhon para responder ao objetivo do problema.

|  |  |
| --- | --- |
| **Anexos:**  **Wine Quality Data Set**  *Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality.names)  **Abstract**: Two datasets are included, related to red and white “vinho verde” wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], [[Web Link]](http://www3.dsi.uminho.pt/pcortez/wine/)). | https://archive.ics.uci.edu/ml/assets/MLimages/Large186.jpg |
|  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Set Characteristics: | Multivariate | Number of Instances: | 6.497 | Area: | Business |
| Attribute Characteristics: | Real | Number of Attributes: | 12 | Date Donated | 2009-10-07 |
| Associated Tasks: | Classification, Regression | Missing Values? | N/A | Number of Web Hits: | 381176 |

**Source:**

Paulo Cortez, University of Minho, Guimarães, Portugal, <http://www3.dsi.uminho.pt/pcortez>   
A. Cerdeira, F. Almeida, T. Matos and J. Reis, Viticulture Commission of the Vinho Verde Region(CVRVV), Porto, Portugal @2009

**Data Set Information:**

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult: [[Web Link]](http://www.vinhoverde.pt/en/) or the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).   
  
**Relevant Papers:**

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties.   
In Decision Support Systems, Elsevier, 47(4):547-553, 2009.   
  
Available at: [[Web Link]](http://dx.doi.org/10.1016/j.dss.2009.05.016)

|  |  |  |
| --- | --- | --- |
| **Característica** | **Descrição** | |
| * **fixedacidity**  *(Acidez fixa)* | | a maioria dos ácidos presentes no vinho ou fixos ou não voláteis (não evaporaram prontamente) |
| * **volatileacidity** *(Acidez volátil)* | | a quantidade de ácido acético no vinho, que em níveis muito altos pode levar a um gosto desagradável de vinagre. |
| * **Citricacid** (Ácido cítrico) | | encontrado em pequenas quantidades, o ácido cítrico pode adicionar “frescura” e sabor aos vinhos. |
| * **Residualsugar**   *(Açúcar residual)* | | a quantidade de açúcar restante depois que o processo de fermentação (em contato com a levedura, o açúcar da uva vai se transformando em álcool, ou seja, a fermentação vai ocorrendo) |
| * **Chlorides**  *(*Cloretos) | | a quantidade de sal no vinho. |
| * **freesulfurdioxide**   *(Dióxido de enxofre livre)* | | a forma livre de SO2 existe em equilíbrio entre o SO2 molecular (como gás dissolvido) e o íon bissulfito; impede o crescimento microbiano e a oxidação do vinho. Garante condições melhores para os processos de vinificação da bebida, elimina bactérias e leveduras frágeis e indesejáveis, o que permite que apenas as melhores prossigam com o processo fermentativo. Além do mais, melhora o aroma e afina a cor da bebida |
| * **totalsulfurdioxide**   *(Dióxido de enxofre total)* | | quantidade de formas livres e encadernadas de S02; em baixas concentrações, o SO2 é quase indetectável no vinho, mas nas concentrações de SO2 livre acima de 50 ppm, o SO2 se torna evidente no nariz e no sabor do vinho. |
| * **Density**  (*Densidade)* | | a densidade é próxima à da água, dependendo do percentual de álcool e teor de açúcar. |
| * **pH** (*PH)* | | descreve como o vinho é acídico ou básico numa escala de 0 (muito ácido) a 14 (muito básico); a maioria dos vinhos tem entre 3-4 na escala de pH. |
| * **Sulphates** *(Sulfatos)* | | um aditivo de vinho que pode contribuir para os níveis de gás de dióxido de enxofre (S02), que age como um antimicrobiano e antioxidante. |
| * **alcohol** *(Álcool)* | | o teor alcoólico percentual do vinho. |
| * **quality** *(Qualidade)* | | variável de saída (com base em dados sensoriais) que poderiam ser de 0 a 10 sendo zero muito ruim e 10 muito excelente. |

Exemplo da Base (10 linhas)



Data de entrega: 11/08/2023

Formato: Word

Regina Bernal

02/08/2023