

Competição 1

Inteligência computacional

Isabella Beatriz da Silva

201802727

2022

INF

INSTITUTO DE
INFORMÁTICA



Sumário

1. Descrição do problema
2. Análise exploratória dos dados
3. Algoritmos utilizados
4. Avaliação



Descrição do problema

O custo para uma operadora de plano de saúde manter prestadores de serviços para fazer a autorização de exames é bem alto.

A partir desse problema foi proposto desenvolver um modelo de predição a partir de um dataset real de pedidos de autorização de exames em planos de saúde.



Análise exploratória dos dados

Análise exploratória dos dados

1. Entendimento dos dados (tipos de dados, seu significado, tipo de variáveis).
2. Analisar a quantidade de valores ausentes no dataset.
3. Analisar a quantidade de valores únicos de um atributo.

```
df_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 227122 entries, 0 to 227121  
Data columns (total 32 columns):
```

| # | Column | Non-Null Count | Dtype |
|----|---------------------------|-----------------|---------|
| 0 | Unnamed: 0 | 227122 non-null | int64 |
| 1 | NR_SEQ_REQUISICAO | 227122 non-null | int64 |
| 2 | NR_SEQ_ITEM | 227122 non-null | int64 |
| 3 | DT_REQUISICAO | 227122 non-null | int64 |
| 4 | DS_TIPO_GUIA | 227122 non-null | object |
| 5 | DT_NASCIMENTO | 227112 non-null | float64 |
| 6 | NR_PRODUTO | 227122 non-null | int64 |
| 7 | DS_TIPO_PREST_SOLICITANTE | 227122 non-null | object |
| 8 | DS_CBO | 227122 non-null | object |
| 9 | DS_TIPO_CONSULTA | 10511 non-null | object |
| 10 | QT_TEMPO_DOENCA | 266 non-null | float64 |
| 11 | DS_UNIDADE_TEMPO_DOENCA | 266 non-null | object |
| 12 | DS_TIPO_DOENCA | 531 non-null | object |
| 13 | DS_INDICACAO_ACIDENTE | 209539 non-null | object |
| 14 | DS_TIPO_SAIDA | 0 non-null | float64 |
| 15 | DS_TIPO_INTERNACAO | 59863 non-null | object |
| 16 | DS_REGIME_INTERNACAO | 59863 non-null | object |
| 17 | DS_CARATER_ATENDIMENTO | 227122 non-null | object |
| 18 | DS_TIPO_ACOMODACAO | 59781 non-null | object |
| 19 | QT_DIA_SOLICITADO | 58995 non-null | float64 |
| 20 | CD_GUIA_REFERENCIA | 37463 non-null | float64 |
| 21 | DS_TIPO_ATENDIMENTO | 168045 non-null | object |
| 22 | CD_CID | 131250 non-null | object |
| 23 | DS_INDICACAO_CLINICA | 179944 non-null | object |
| 24 | DS_TIPO_ITEM | 227122 non-null | object |
| 25 | CD_ITEM | 227122 non-null | int64 |
| 26 | DS_ITEM | 227122 non-null | object |

```
uniqueValues = df_train.nunique()  
print(uniqueValues)
```

| | |
|---------------------------|--------|
| Unnamed: 0 | 227122 |
| NR_SEQ_REQUISICAO | 80699 |
| NR_SEQ_ITEM | 227122 |
| DT_REQUISICAO | 357 |
| DS_TIPO_GUIA | 3 |
| DT_NASCIMENTO | 16557 |
| NR_PRODUTO | 1 |
| DS_TIPO_PREST_SOLICITANTE | 12 |
| DS_CBO | 59 |
| DS_TIPO_CONSULTA | 4 |
| QT_TEMPO_DOENCA | 17 |
| DS_UNIDADE_TEMPO_DOENCA | 3 |
| DS_TIPO_DOENCA | 2 |
| DS_INDICACAO_ACIDENTE | 4 |
| DS_TIPO_SAIDA | 0 |
| DS_TIPO_INTERNACAO | 6 |
| DS_REGIME_INTERNACAO | 3 |
| DS_CARATER_ATENDIMENTO | 2 |
| DS_TIPO_ACOMODACAO | 8 |
| QT_DIA_SOLICITADO | 34 |
| CD_GUIA_REFERENCIA | 4610 |
| DS_TIPO_ATENDIMENTO | 13 |
| CD_CID | 1626 |
| DS_INDICACAO_CLINICA | 40428 |
| DS_TIPO_ITEM | 2 |
| CD_ITEM | 6220 |
| DS_ITEM | 6146 |
| DS_CLASSE | 460 |
| DS_SUBGRUPO | 72 |

Atributos escolhidos

Variáveis discretas


- IDADE: variável criada a partir de um pré-processamento que subtrai a 'DT_REQUISICAO' pela 'DT_NASCIMENTO'
 - A idade é importante porque geralmente pessoas mais velhas tem uma necessidade maior que seus exames sejam autorizados.

Atributos escolhidos

Variáveis categóricas

- DS_TIPO_PREST_SOLICITANTE: caracteriza onde foi feita a solicitação do exame (hospital, clínica, laboratório...)
- DS_CBO: tipo de médico que atendeu o paciente.
- DS_INDICACAO_ACIDENTE: tipo do acidente que levou ao atendimento, inclusive mostra quando não houve acidente.
- DS_CARATER_ATENDIMENTO: se foi uma consulta de rotina ou emergência, geralmente os exames são autorizados em casos de emergência.

- DS_TIPO_INTERNACAO: o tipo da internação do paciente que levou a solicitação.
- DS_TIPO_ACOMODACAO: onde o paciente estava (enfermaria, UTI...)
- DS_TIPO_ATENDIMENTO: qual tipo de procedimento foi solicitado (exame, cirurgia, terapia)



Algoritmos utilizados e resultados da avaliação

Decision tree

- Modelo de simples entendimento e com bons resultados para predição.

```
[[27245 3551]
 [ 8083 6546]]
      precision    recall  f1-score   support

  Autorizado      0.77      0.88      0.82     30796
    Negado       0.65      0.45      0.53     14629

 accuracy      0.74     45425
  macro avg     0.71     45425
weighted avg     0.73     45425

0.7438855255916346
```

Random forest

- Escolhido para validar os resultados obtidos utilizando a árvore de decisão sozinha.

```
[[26852 3944]
 [ 7749 6880]]
precision    recall  f1-score   support

 Autorizado    0.78    0.87    0.82    30796
   Negado     0.64    0.47    0.54    14629

 accuracy              0.74    45425
 macro avg           0.71    0.67    0.68    45425
 weighted avg        0.73    0.74    0.73    45425

0.7425866813428729
```

MLP sklearn

- Rede neural simples que aprenderia com os erros de um dataset desbalanceado e teria teoricamente um resultado melhor.

```
[[27261  3535]
 [ 9727  4902]]
      precision    recall  f1-score   support

  Autorizado      0.74      0.89      0.80     30796
    Negado       0.58      0.34      0.43     14629








 accuracy      0.71     45425
 macro avg      0.66      0.61      0.61     45425
weighted avg      0.69      0.71      0.68     45425

0.7080462300495322
```

Score final da competição

Public Private

The private leaderboard is calculated with approximately 70% of the test data.
This competition has completed. This leaderboard reflects the final standings.

| # | △ | Team | Members | Score |
|---|-----|----------------------------------|--|---------|
| 1 | ▲ 1 | Matheus Paz |  | 0.71070 |
| 2 | ▼ 1 | To tentando |   | 0.71014 |
| 3 | — | Mateus Pereira |  | 0.70829 |
| 4 | — | HUMBERTO P T SILVA |   | 0.70616 |
| 5 | — | ISABELLA BEATRIZ DA SILVA |  | 0.70581 |