



UFOP

Universidade Federal de Ouro Preto – UFOP
Instituto de Ciências Exatas e Biológicas – ICEB
Departamento de Computação – DECOM
Disciplina: BCC325 – Inteligência Artificial
Professor: Jadson Castro Gertrudes

Avaliação 02 - Especificação

A segunda avaliação da disciplina consiste na aplicação e avaliação de algoritmos de classificação em uma base de dados contendo informações de empresas com e sem potencial para decretar falência. Para tal avaliação, devem ser utilizadas as implementações dos algoritmos vistos em aula e disponíveis no pacote scikit-learn da linguagem de programação Python.

1. Classificadores

Os seguintes algoritmos de classificação devem ser considerados para aplicação e avaliação dos resultados:

- Regressão Logística (*Logistic Regression*)
- Árvore de Decisão (*Decision Tree*)
- Árvores aleatórias (*Random Forest*)

Considere a acurácia e a medida F-measure na avaliação dos resultados. Por fim, compare os algoritmos, comente e apresente sua interpretação para os resultados obtidos.

Observação: Outros algoritmos de classificação podem ser utilizados, desde que apresentem uma boa explicação do seu processo de aprendizado.

2. Problema

O gerente de uma agência financeira coletou dados de quatrocentas empresas localizadas em um estado. Todas as empresas analisadas continham informações ou não de falência naquele ano (rótulo). Logo em seguida, foi enviada para o gerente uma nova lista com 100 empresas, as quais ele deveria informar se elas possuíam ou não risco de falência. Para resolver esse problema de forma rápida a equipe de alunos de BCC325 foi acionada para tentar prever da melhor forma possível a situação das 100 empresas coletadas.

2.1 Base de dados

Falência de empresa (falencia.csv). O conjunto de dados **falencia** representa o resultado da falência ou não de uma empresa a partir de uma série de medidas e índices coletados. Originalmente, possui 21 atributos, onde os 20 primeiros são as medidas e índices financeiros e o último atributo apresenta a situação (rótulo) da empresa, classificada como **Falência (1)** ou **Não falência (0)**. O conjunto possui um total de **500** observações, sendo que **100** delas serão utilizadas para teste. Não houve nenhum pré-processamento dos dados.

Algumas observações sobre o conjunto de dados:

- Os dados não estão balanceados. Existe uma proporção menor de empresas que **decretaram falência**.



UFOP

- Dentre os atributos (informações) coletados das empresas, existe um subconjunto deles que pode melhorar sua tarefa de classificação. Uma boa estratégia seria descobrir quais são os atributos mais relevantes.

3. Avaliação

Este trabalho será avaliado em três partes:

- Avaliação do relatório: **Notas de 0 a 100% - Peso 45%**
 - Utilize gráficos e tabelas com o fim de facilitar a comparação e obtenção de conclusões. A análise e conclusões do grupo devem ser apresentados em forma de um relatório com no máximo **6 páginas**. Relatórios contendo somente gráficos e/ou tabelas sem análises profundas e fundamentadas receberão nota zero.
- **Valor da métrica do modelo** (F-measure ponderada): **Notas de 0% a 100% - Peso 10%**
- Apresentação da metodologia implementada: **Notas de 0% a 100% - Peso 45%**
 - As apresentações serão **gravadas**, carregadas no Youtube (ou google Drive) e disponibilizadas para o professor da disciplina até o dia 20/mar/2023.
 - Cada apresentação deve ser de, no máximo, 20 minutos. Neste é incluído o tempo da demonstração da implementação.
 - A apresentação deverá ser feita por **todos** os integrantes do grupo. Todos os integrantes do grupo devem ser capazes de explicar sobre o trabalho.
 - Os critérios de avaliação da apresentação serão:
 - I. Clareza e organização da apresentação.
 - II. Capacidade de justificar cada passo metodologia implementada.
 - O trabalho deve ser feito em grupos de 5 pessoas.
 - Os grupos devem ser definidos pelos alunos até o **dia 24/fev/2023**. Os nomes dos integrantes de cada grupo deverão ser enviados para o e-mail jadson.castro@ufop.edu.br.
 - Os grupos serão sorteados aleatoriamente via software caso os alunos não atendam o prazo especificado no item anterior.
 - As submissões das apresentações serão utilizadas para contabilização da presença naquela semana.
 - Todos os grupos devem utilizar a plataforma **Google Colab** para apresentação dos experimentos.
 - Casos excepcionais serão tratados pelo professor.