

Lista #5

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Profª. Cristiane Neri Nobre

Data de entrega: 23/09

Valor: 1,5 pontos

Questão 01

Implemente a *Árvore de decisão*, *Naive Bayes* e *Random Forest*, em Python, no dataset de **breast-cancer.csv**, disponível no CANVAS.

Faça uma comparação destes três algoritmos utilizando as métricas, recall, precision e F1-Score.

Além disso, discuta as métricas para cada uma das classes. Veja que o desempenho da classe minoritária é menor (problema muito comum em bases desbalanceadas). O que você fará para resolver este problema? Investigue, pelo menos, um métodos para balanceamento de dados.

Bibliotecas para os algoritmos de aprendizado:

Naive Bayes: `From sklearn.naive_bayes import GaussianNB`

Árvore de Decisão: `From sklearn.tree import DecisionTreeClassifier`

Random Forest: `from sklearn.ensemble import RandomForestClassifier`

Bibliotecas/funções para balanceamento:

`From imblearn.under_sampling import TomekLinks`

`from imblearn.over_sampling import SMOTE`

Investigue o funcionamento do RandomUnderSampler: https://imbalanced-learn.org/stable/references/generated/imblearn.under_sampling.RandomUnderSampler.html

Investigue outros métodos para oversampling e undersampling em:

https://imbalanced-learn.org/stable/under_sampling.html

Questão 02

Implemente a *Árvore de decisão*, *Naive Bayes* e *Random Forest*, em Python, no dataset de **labor.csv**, disponível no CANVAS.

Faça uma comparação destes três algoritmos utilizando as métricas, recall, precision e F1score. Veja que esta base de dados possui dados ausentes. O que você fará para imputar estes valores? Investigue pelo menos 1 método de imputação.

Comandos para verificação de ausência:

IsNull

Métodos para imputação de dados ausentes: Média, moda, missforest, KNNImputer

Investigue a biblioteca missingpy: <https://pypi.org/project/missingpy/>