

Lista #4

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Prof^a. Cristiane Neri Nobre

Data de entrega: 17/03

Valor: 1 ponto

Questão 01

Considerando-se a base de dados de **breast-cancer.csv** disponível no CANVAS, pede-se:

- 1) Implemente a *Árvore de decisão*, *Naive Bayes* e *Random Forest*, em Python. Faça uma comparação destes três algoritmos utilizando as métricas, recall, precision e F1-Score, para cada uma das classes (**câncer recorrente/Não recorrente**). Discuta os resultados obtidos e veja se o desbalanceamento da base está prejudicando o treinamento.
- 2) Investigue, pelo menos, 1 algoritmo de balanceamento do tipo **undersampling** e 1 do tipo **oversampling** para melhorar o desempenho da classe **minoritária**.
- 3) Investigue, pelo menos, 1 algoritmo de **imputação 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

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/>