**Trabalho de classificação de dados (09/12/2020)**

Inicialmente cada dupla deve procurar uma base de dados na qual seja possível realizar um trabalho de classificação. **A base deve possuir mais de 200 instâncias e cada uma contendo pelo menos 10 atributos**. Duas duplas não podem utilizar a mesma base, portanto, em caso de duas duplas selecionarem a mesma, o que me enviar o nome e uma breve descrição primeiro ficará com ela.

**A aula de sexta (11/12/2020) será reservada para que cada um me apresente a base**, em seguida, analisarei se ela será permitida ou se você deve escolher outra. Principais pontos a serem apresentados:

- De onde a base foi obtida;

- Qual o objetivo da base e qual a tarefa de classificação que será realizada;

- Número de instância e quantidade de atributos, juntamente com a descrição de cada atributo.

Sugestões de locais que vocês podem procurar bases:

- <https://toolbox.google.com/datasetsearch>

- <https://www.kaggle.com>

- <http://mlr.cs.umass.edu/ml/>

-<https://towardsdatascience.com/the-50-best-public-datasets-for-machine-learning-d80e9f030279>

**Após a escolha da base, vocês irão realizar a tarefa de classificação com um classificador selecionado por mim. Vocês selecionam a dupla e eu seleciono o classificador. Tarefas que vocês devem realizar:**

- **Escolher uma técnica de normalização de dados e aplicar na base**. Na apresentação em sala deve ser explicado o funcionamento dessa técnica e qual a diferença que ela faz nos resultados de classificação obtidos;

- **Escolher uma técnica para a seleção de atributos e aplicar na base**. Na apresentação em sala deve ser explicado o funcionamento dessa técnica e qual a diferença que ela faz nos resultados obtidos. Mostrar os resultados com todos os atributos da base e somente com os selecionados.

- **Escolher as métricas de avaliação dos resultados obtidos.** Na apresentação em sala deve ser explicado porque você escolheu elas.

A apresentação dos trabalhos será realizada na **próxima quarta (16/12/2020)** e cada dupla terá até 15 minutos. A apresentação deve descrever todos os passos realizados, também não esquecer de descrever o funcionamento do algoritmo de classificação utilizado (pode usar alguns slides), seus principais parâmetros e como a variação dos mesmos afetam os resultados obtidos. Por fim, também devem ser descritos os principais problemas e dificuldades encontrados na base e como você solucionou isso.

Support Vector Machines (SVM)

* from sklearn import svm

Multi-layer Perceptron (MLP)

* from sklearn.neural\_network import MLPClassifier

Stochastic Gradient Descent (SGD)

* from sklearn.linear\_model import SGDClassifier

K Nearest Neighbors (KNN)

* from sklearn.neighbors import KNeighborsClassifier

Gaussian Processes Classification (GPC)

* from sklearn.gaussian\_process import GaussianProcessClassifier

Naive Bayes (NB)

* from sklearn.naive\_bayes import GaussianNB

Decision Trees (DT)

* from sklearn.tree import DecisionTreeClassifier

Gradient Tree Boosting

* from sklearn.ensemble import GradientBoostingClassifier

Random Forests

* from sklearn.ensemble import RandomForestClassifier

Bagging meta-estimator (BME)

* from sklearn.ensemble import BaggingClassifier

AdaBoost Classifier (ABC)

* from sklearn.ensemble import AdaBoostClassifier