



Projeto do aluno - Ensaio de Machine Learning

☰ Ciclo	Ciclo 09 - Projeto do aluno
# Aula	72
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A empresa Data Money
- ☐ O Ensaio
- ☐ Na próxima aula

Conteúdo:

▼ 1. A empresa Data Money

A empresa Data Money fornece serviços de consultoria de Análise e Ciência de Dados para grandes empresas no Brasil e no exterior.

O seu principal diferencial de mercado em relação aos concorrentes é o alto retorno financeiro para as empresas clientes, graças a performance de seus algoritmos de Machine Learning.

A Data Money acredita que a expertise no treinamento e ajuste fino dos algoritmos, feito pelos Cientistas de Dados da empresa, é a principal motivo dos ótimos resultados que as consultorias vem entregando aos seus clientes.

Para continuar crescendo a expertise do time, os Cientistas de Dados acreditam que é extremamente importante realizar ensaios nos algoritmos de Machine Learning para adquirir uma experiência cada vez maior sobre o seu funcionamento e em quais cenários as performances são máximas e mínimas, para que a escolha do algoritmo para cada situação seja a mais correta possível.

Como Cientista de Dados recém contratado pela empresa, a sua principal tarefa será realizar 3 ensaios com algoritmos de Classificação, Regressão e Clusterização, a fim de extrair aprendizados sobre o seu funcionamento em determinados cenário e conseguir transmitir esse conhecimento para o restante do time.

▼ 2. O Ensaio de Machine Learning

▼ 2.1. Descrição do Ensaio

O ensaio de Machine Learning ajuda os Cientistas de Dados a ganhar mais experiência na aplicação dos algoritmos. Nesse ensaio, em específico, cada algoritmo será treinado com os dados de treinamento e sua performance será medida usando 3 conjuntos de dados:

1. Os próprios dados de treinamento
2. Os dados de validação
3. Os dados de teste.

A performance de cada algoritmo será medida, utilizando diferentes métricas de performance.

O seu trabalho nesse ensaio será construir uma tabela mostrando os valores das métricas de performance para cada algoritmo de Machine Learning, conforme o exemplo abaixo.

▼ **Tabela do ensaio de Machine Learning**

	Nome do algoritmo	Métrica 1	Métrica 2	Métrica 3	Métrica 4
1					
2					
3					
4					

Cada tabela vai armazenar os resultados da performance sobre um conjunto de dados diferentes, ou seja, você precisa criar 3 tabelas: 1) Performance sobre os dados de treinamento, 2) Performance sobre os dados de validação e 3) Performance sobre os dados de teste para o Ensaio de classificação, regressão e clusterização

▼ **2.2. Os algoritmos e métricas do ensaio**

Classificação:

Algoritmos: KNN, Decision Tree, Random Forest e Logistic Regression

Métricas de performance: Accuracy, Precision, Recall e F1-Score

Regressão:

Algoritmos: Linear Regression, Decision Tree Regressor, Random Forest Regressor, Polinomial Regression, Linear Regression Lasso, Linear Regression Ridge, Linear Regression Elastic Net, Polinomial Regression Lasso, Polinomial Regression Ridge e Polinomial Regression Elastic Net

Métricas de performance: R2, MSE, RMSE, MAE e MAPE

Agrupamento:

Algoritmos: K-Means e Affinity Propagation

Métricas de performance: Silhouette Score

▼ **2.3. Os parâmetros do ensaio**

▼ 2.3.1. Classificação

Algoritmo	Parâmetros
Random Forest Classifier	n_estimators
	max_depth
K-Neighbors Classifier	n_neighbors
Logistic Regression	C
	solver
	max_iter
Decision Tree Classifier	max_depth

▼ 2.3.1. Regressão

Algoritmo	Parâmetros
Decision Tree Regressor	max_depth
Polinomial Regression	degree
Polinomial Regression Lasso e Ridge	degree
	alpha
	max_iter
Polinomial Regression Elastic Net	degree
	alpha
	l1_ratio
	max_iter

Algoritmo	Parâmetros
Random Forest Regressor	n_estimators
	max_depth
Linear Regression Lasso e Ridge	alpha
	max_iter
Linear Regression Elastic Net	alpha
	l1_ratio
	max_iter

▼ 2.3.1. Agrupamento

Algoritmo	Parâmetros
K-Means	k
Affinity Propagation	preference

▼ 3. Na próxima aula

Resultado esperado do Ensaio