



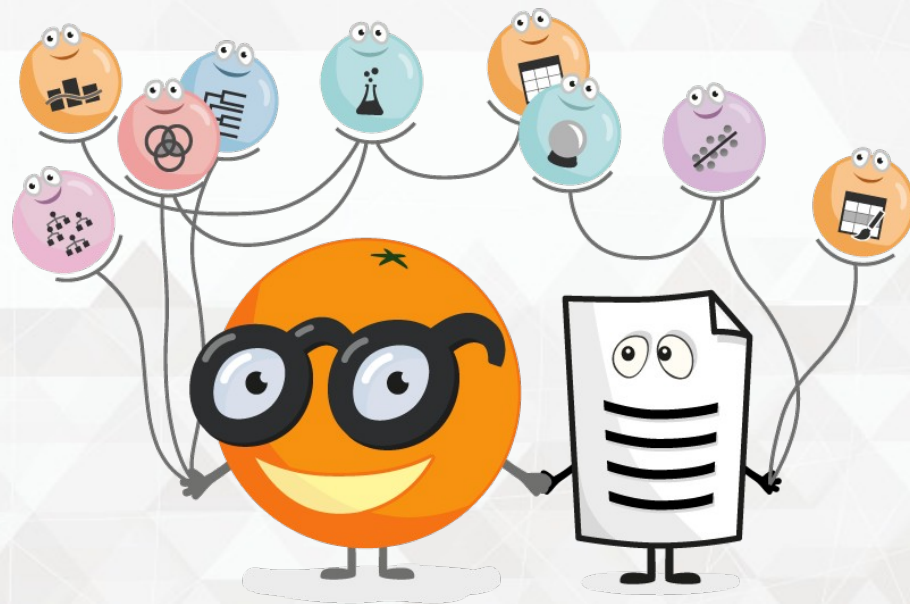
Orange

Modelos de Machine Learning

2023

Conteúdo

- Modelos supervisionados
- Avaliação e predição
- Modelos não-supervisionados
- Add-ons



Conteúdo

- Modelos supervisionados
- Avaliação e predição
- Modelos não supervisionados
- Add-on

SEMPRE LEMBRANDO PARA
QUAIS
OBJETIVOS O ORANGE FOI
PROJETADO...



Modelos Supervisionados

- Seleção de widgets para modelos supervisionados
- Parâmetros simplificados
 - Sempre lembrando a intenção da ferramenta
- Ferramentas de visualização
- Ótimo auxílio visual para comparação de possíveis algoritmos para solução de determinado problema
 - Se parâmetros simplificados puderem ser aceitos

Modelos Supervisionados



Constant



CN2 Rule Induction



Calibrated Learner



kNN



Tree



Random Forest



Gradient Boosting



SVM



Linear Regression



Logistic Regression



Naive Bayes



AdaBoost



Neural Network



Stochastic Gradient
Descent



Stacking



Save Model



Load Model

Modelos Supervisionados

- Classificação por árvores



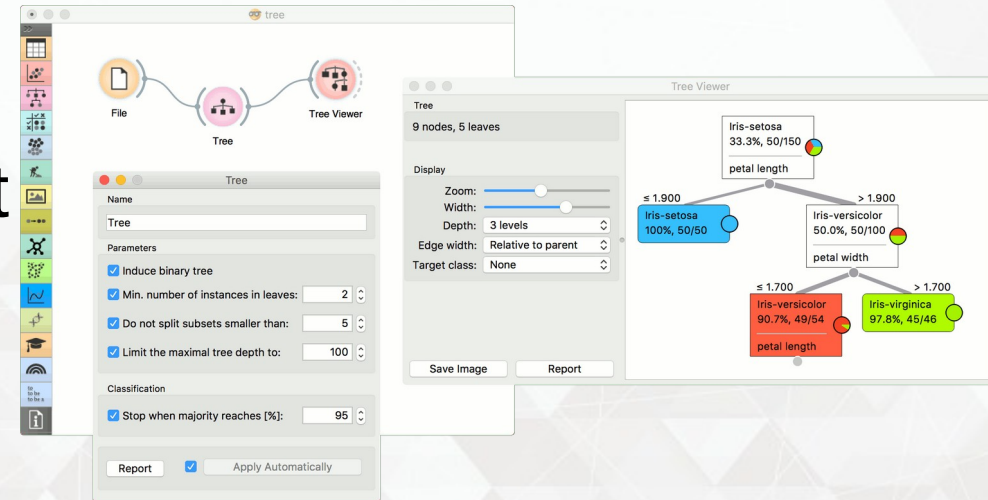
- Decision Tree



- Random Decision Forest

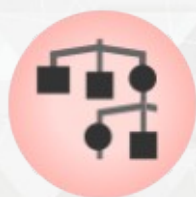


- Gradient Boosting



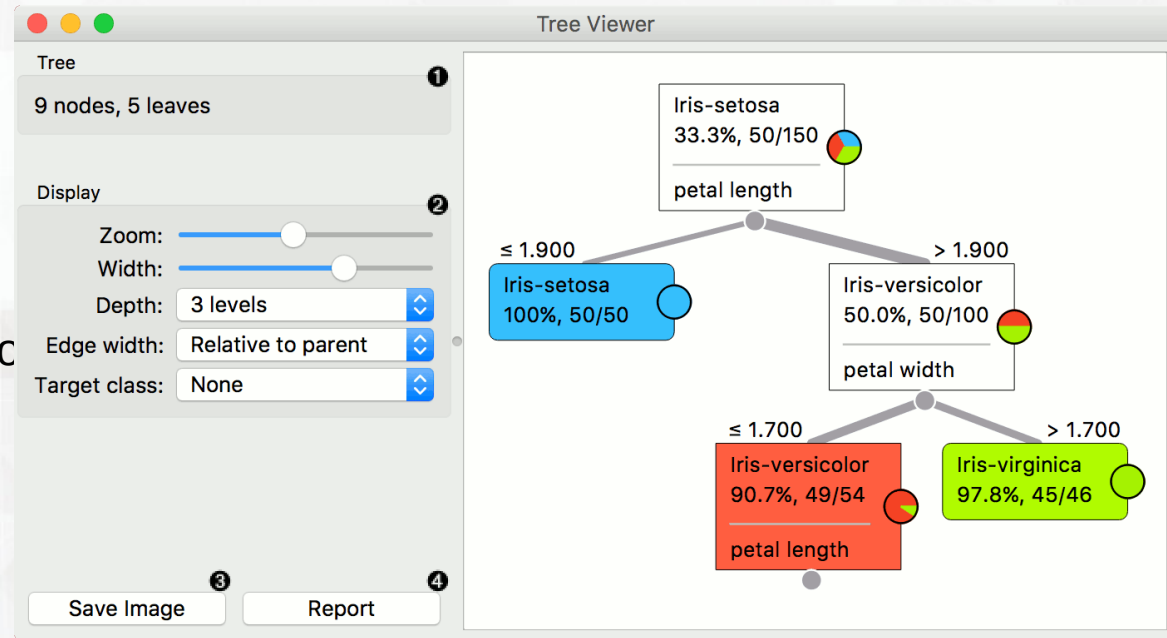
Modelos Supervisionados

- Visualização de árvores
 - Tree Viewer
 - Pythagorean Tree
 - Pythagorean Forest
 - Generalized Pythagoras Trees for Visualizing Hierarchies



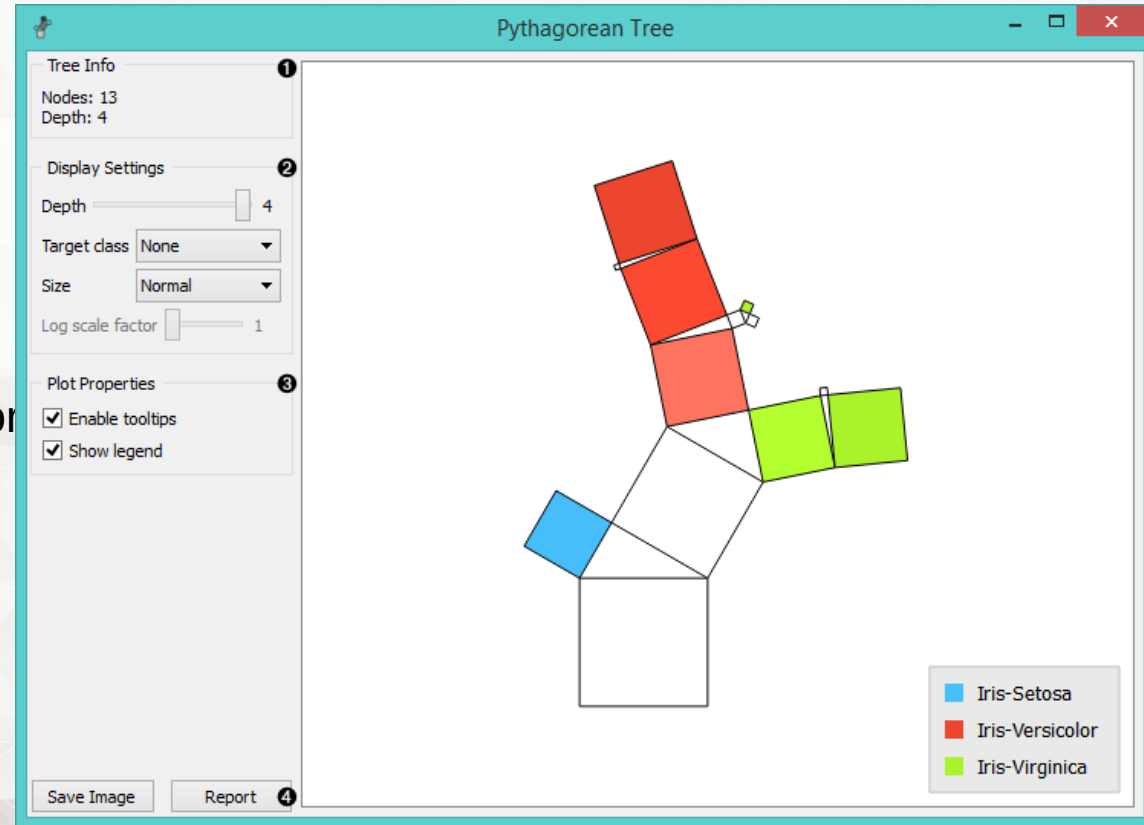
Modelos Supervisionados

- Visualização de árvores
 - Tree Viewer
 - Pythagorean Tree
 - Pythagorean Forest
 - Generalized Pythagoras Trees for Visualizing Hierarchies



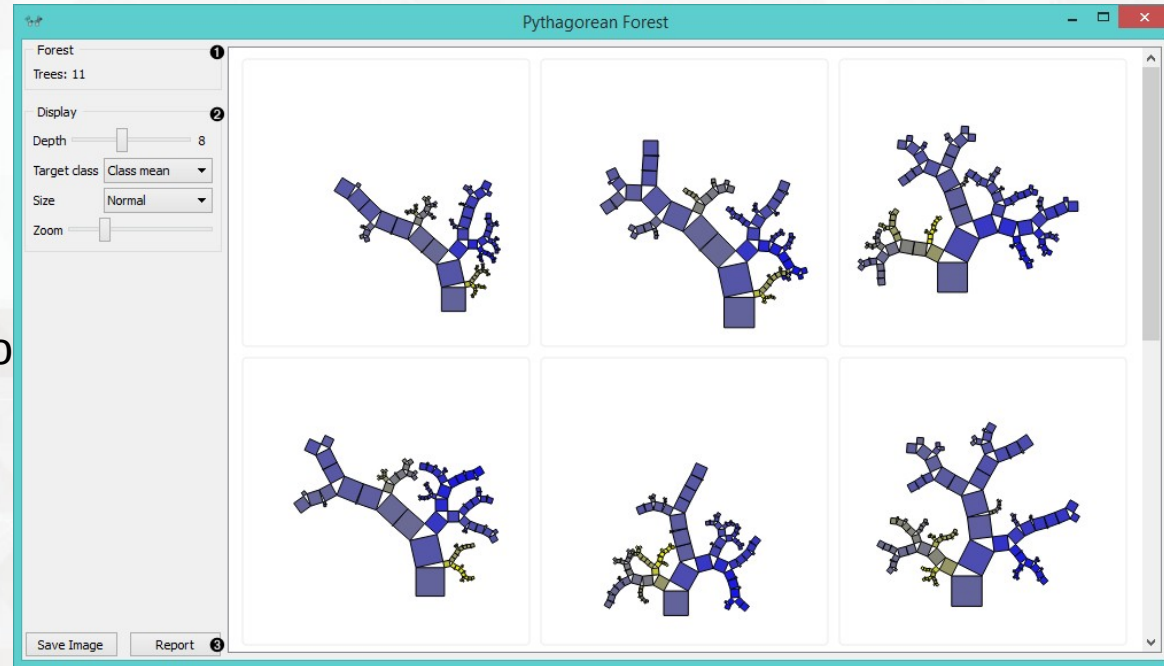
Modelos Supervisionados

- Visualização de árvores
 - Tree Viewer
 - Pythagorean Tree
 - Pythagorean Forest
 - Generalized Pythagoras Trees for Visualizing Hierarchies



Modelos Supervisionados

- Visualização de árvores
 - Tree Viewer
 - Pythagorean Tree
 - Pythagorean Forest
 - Generalized Pythagoras Trees for Visualizing Hierarchies



Modelos Supervisionados

- Modelos podem ser salvos e carregados
 - Para fluxos mais complexos
 - Treinamento já realizado
 - Formato pkcls
- Save Model
- Load Model



Avaliação de Modelos

- Comparação de modelos treinados
 - Pode mostrar métricas de diversos modelos em comparação
 - Para um mesmo dataset
 - Métricas convencionais
 - Accuracy, precision, recall, F1, AUC (area under curve)
- Dataset é dividido dentro do componente
 - Pode ser particionado previamente, se necessário
 - Random sampling, cross-validation

Avaliação de Modelos



- Test and Score

- Avalia e compara modelos



- Predictions

- Realiza predições com um dataset auxiliar



- Confusion Matrix



- ROC Analysis

Avaliação de Modelos

- Test and Score

The screenshot displays the 'Test and Score' application interface, which is used for evaluating machine learning models. The interface is divided into several sections:

- Sampling (1):** This section allows users to choose between 'Cross validation' and 'Random sampling'. Under 'Cross validation', users can specify the 'Number of folds' (set to 10) and whether to use 'Stratified' sampling (checked). Under 'Random sampling', users can specify the 'Repeat train/test' (set to 10) and the 'Training set size' (set to 66%). There are also options for 'Stratified' sampling and 'Leave one out'.
- Target Class (2):** This section allows users to select the 'Target Class' (set to '(Average over classes)').
- Model Comparison (4):** This section allows users to select the 'Model Comparison' method (set to 'Area under ROC curve'). There is also a checkbox for 'Negligible difference' (set to 0.1).
- Evaluation Results (3):** This section displays a table of evaluation results for four models: Logistic Regression, Naive Bayes, SVM, and Tree. The table includes columns for AUC, CA, F1, Precision, and Recall.
- Model Comparison by AUC (5):** This section displays a table comparing the AUC values for the four models. The table includes columns for Tree, SVM, Naive Bayes, and Logistic Regression.

Below the 'Model Comparison by AUC' table, there is a note: 'Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.'

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.732	0.776	0.762	0.770	0.776
Naive Bayes	0.700	0.776	0.762	0.770	0.776
SVM	0.500	0.531	0.546	0.581	0.531
Tree	0.737	0.783	0.749	0.816	0.783

	Tree	SVM	Naive Bay...	Logistic R...
Tree		1.000	0.993	0.802
SVM	0.000		0.000	0.000
Naive Bayes	0.007	1.000		0.015
Logistic Regression	0.198	1.000	0.985	

Modelos Não-Supervisionados



Distance File



Distance Matrix



t-SNE



Distance Map



Hierarchical
Clustering



k-Means



Louvain Clustering



DBSCAN



Manifold Learning



PCA



Correspondence
Analysis



Distances



Distance
Transformation



MDS



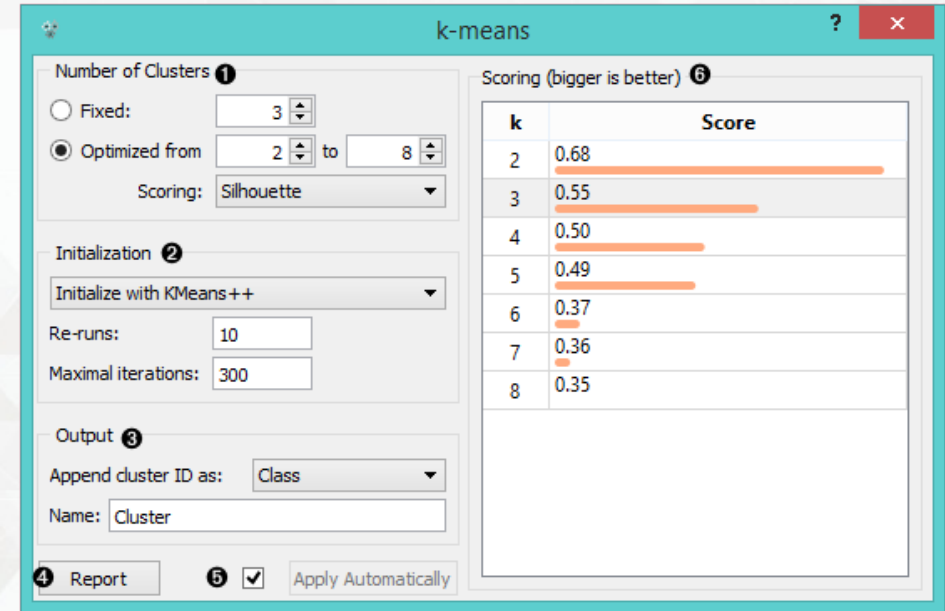
Save Distance Matrix



Self-Organizing Map

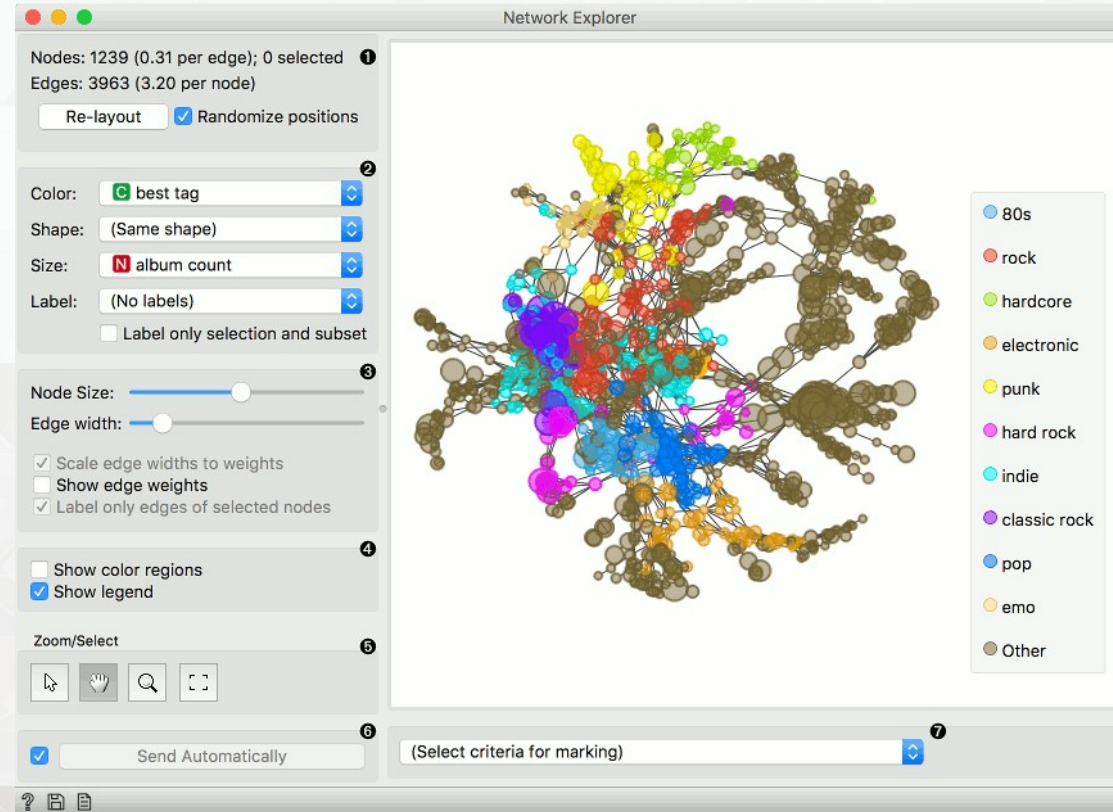
Modelos Não-Supervisionados

- K-Means
 - Avaliação simplificada de número de clusters
 - Silhouette



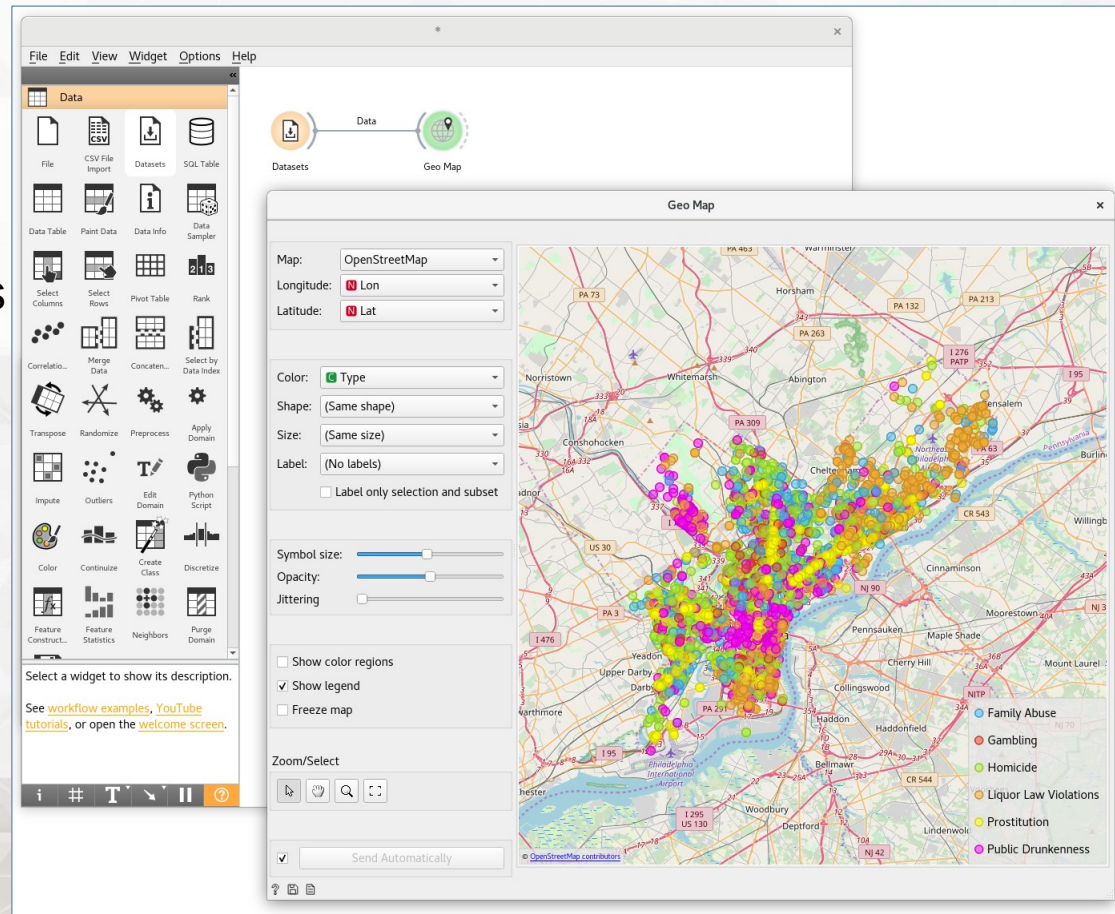
Add-ons

- Recursos incorporados por tema
 - Algoritmos e técnicas com uma menor utilização
 - Pela avaliação dos desenvolvedores
- API para desenvolvimento de add-ons
- Categorias interessantes
 - Text / Textable
 - Associate
 - Image Analytics
 - Network
 - Geo



Add-ons

- Recursos incorporados por tema
 - Algoritmos e técnicas com uma menor utilização
 - Pela avaliação dos desenvolvedores
- API para desenvolvimento de add-ons
- Categorias interessantes
 - Text / Textable
 - Associate
 - Image Analytics
 - Network
 - Geo



E ainda tem muito mais...

- Orange é um software complexo e em constante evolução
 - Evolução já deixou de depender apenas dos desenvolvedores originais
- Ainda assim é importante lembrar dos objetivos do projeto
 - E das limitações quando a qualidade do código, testes, processo de desenvolvimento, etc
 - Como muitos esforços colaborativos
- Existe pouca documentação disponível
 - Podemos até colaborar ;-)



Obrigado

leandro@utfpr.edu.br

<http://lapti.ct.utfpr.edu.br>