



**Desmistificando
Automated Machine
Learning com
Qlik AutoML e Qlik Sense**

**Pablo Labbe
Junho /2023**





Pablo Labbe 

Pablo Labbe



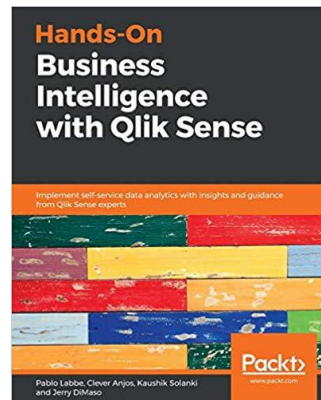
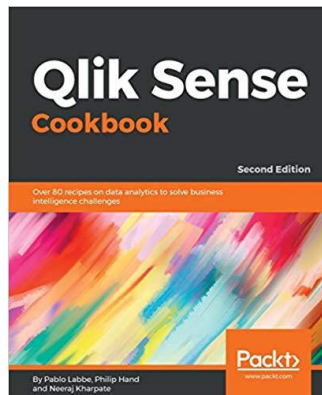
Arquiteto de Soluções SR @iMaps

+20 anos projetos de Analytics / 13 Qlik

Qlik Luminary 2019,2020

Qlik Partner Ambassador 2021,2022,2023


Co-autor de 2 livros sobre Qlik Sense



Diamante



AGGRANDIZE

COMMVAULT 



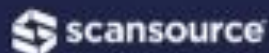
TD SYNnex

Platina



DISCOVER

Ouro



VERTICA
by opentext

Prata

TRACES



Rox

We take care
of your data

Apoio

FIAP

GRUPO
POSEIDON
DIGITAL

Agenda

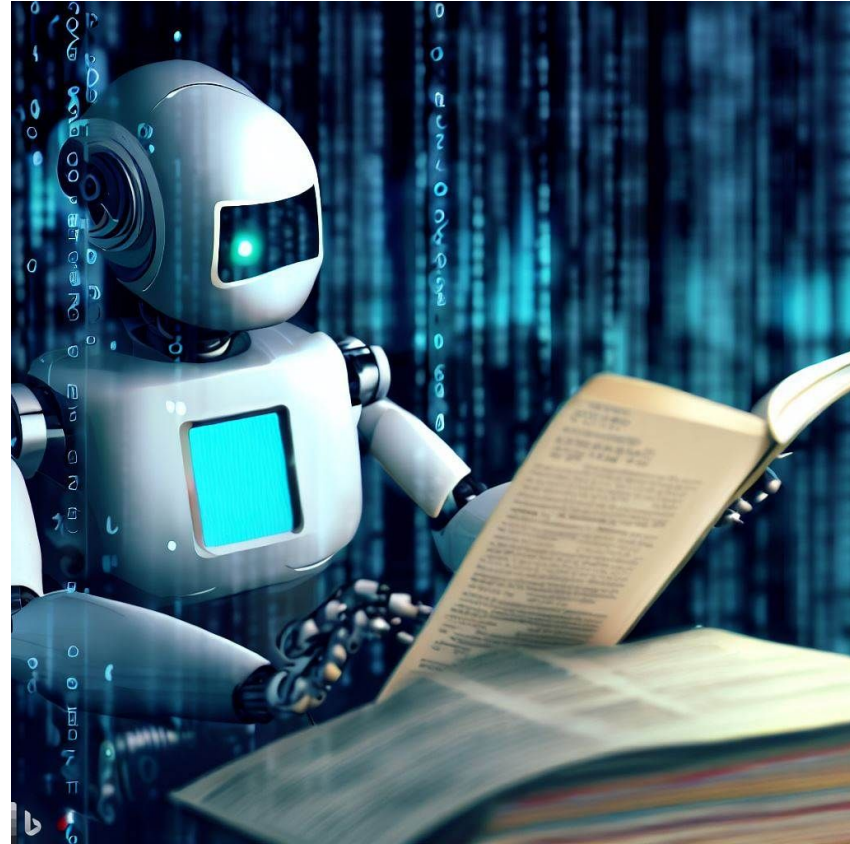
- O que é
- Introdução ao Automated Machine Learning (AutoML)
- O Processo Tradicional de Escolha de Modelos de Machine Learning
- Vantagens do Automated Machine Learning
- Exemplos de Ferramentas de AutoML
- Conclusão



Conceitos de ML e AutoML

Definição de Machine Learning

“É uma área de conhecimento específica da I.A. que trata do processo de reconhecimento de padrões em dados históricos através de algoritmos para previsão de resultados futuros”



O Processo Tradicional de Machine Learning



- Pré-processamento manual de dados: limpeza, normalização, tratamento de valores ausentes, etc.
- Seleção manual de algoritmos: experimentação com diferentes algoritmos para encontrar o melhor desempenho.
- Otimização manual de hiperparâmetros: ajuste manual dos parâmetros para otimizar o desempenho do modelo.

Aquisição
de Dados

Exploração
de Dados

Preparação
de Dados

Feature
Engineering

Seleção do
Modelo

Treinamento
do Modelo

Otimização
Hiper
parâmetros

Projeções



Criar modelos de ML no método tradicional pode ser complexo e custar muito tempo !!

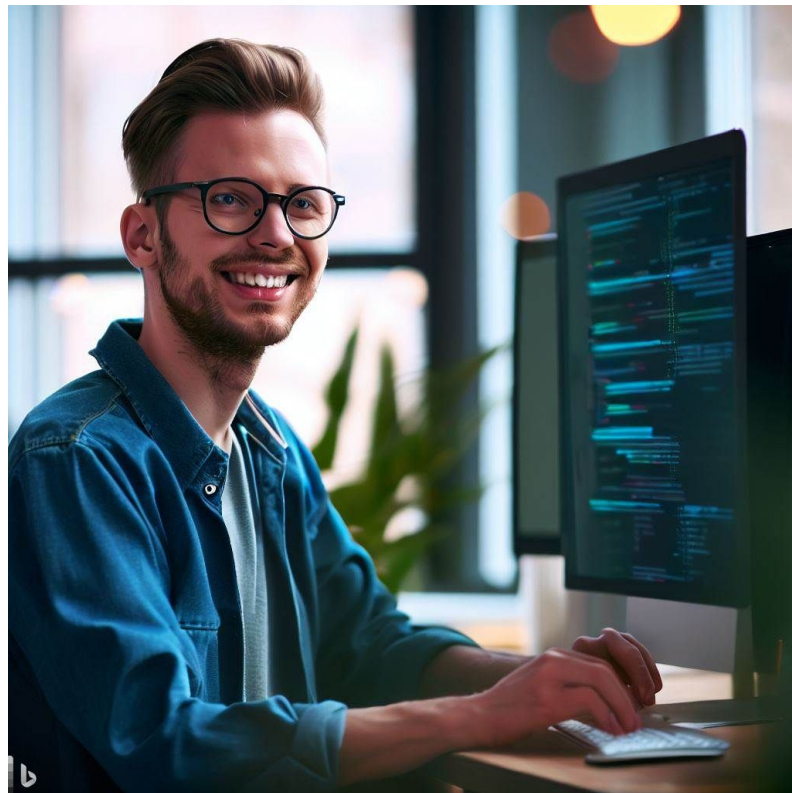


Automated Machine Learning (AutoML)

O que é Automated Machine Learning (AutoML)

Automated Machine Learning (AutoML) é o processo de automatizar tarefas de seleção de modelos, pré-processamento de dados e otimização de hiperparâmetros

Isso facilita a construção de modelos de ML para profissionais que não são Cientistas de Dados.



Vantagens do AutoML

- Economia de tempo: o AutoML automatiza tarefas que costumam ser demoradas e repetitivas.
- Acesso mais fácil: permite que pessoas com menos conhecimento técnico construam modelos de machine learning.
- Maior desempenho: o AutoML pode explorar várias combinações de algoritmos e hiperparâmetros para encontrar o melhor modelo.
- Redução de erros: minimiza erros humanos durante o processo de seleção de modelos.



Qlik AutoML

Qlik AutoML (parte da plataforma Qlik Cloud)

Data Integration


Data Movement
& Streaming


Data
Warehouse
Automation


Data
Transformation


Application
Automation

Analytics


Visualization
& Dashboards


Augmented
Analytics & AutoML


Embedded
Analytics


Alerting
& Action

FOUNDATIONAL SERVICES



Catalog & Lineage



Artificial Intelligence



Associative Engine



Orchestration



Governance & Security



Reporting & Collaboration



Developer & API



Hybrid Cloud

Universal Connectivity



RDBMS



Data
Warehouse



Data Lake



SaaS



Apps



Mainframe



Stream

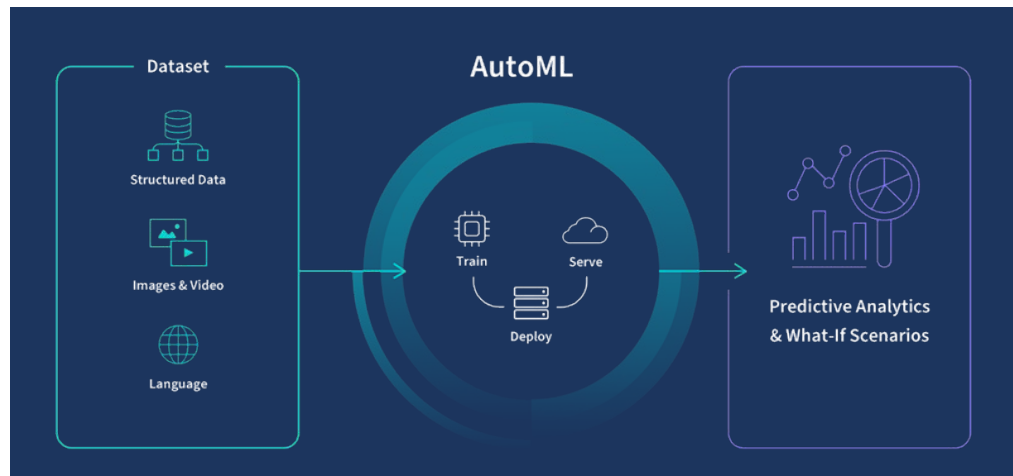


Files



Qlik Auto ML

- Treine modelos de ML automaticamente sem código
- Desenhado para times de analytics
- Experimentos ilimitados
- Projeções e explicabilidade do modelo
- Endpoints de API para projeções em tempo real
- Agendamento da execução dos modelos



Problemas de negócio

Sales

- Sales Pipeline – Win / Loss Forecasting
- Customer Churn / Retention
- Customer Prospecting / Targeting

Marketing

- Demand / Revenue Forecasting
- Customer Lifetime Value
- Customer Next Best Offer

Finance

- Capital Investment Optimization
- Expense Management
- Risk Management / Reduction

Operations

- Workforce Demand Prediction
- Capacity Allocation
- Appointment Cancellations

HR

- Employee Retention / Attrition Prediction
- Employee Satisfaction
- Recruiting / Candidate Profiling

IT

- Software / Licensing Usage
- Infrastructure Performance Prediction

Supply Chain

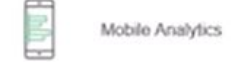
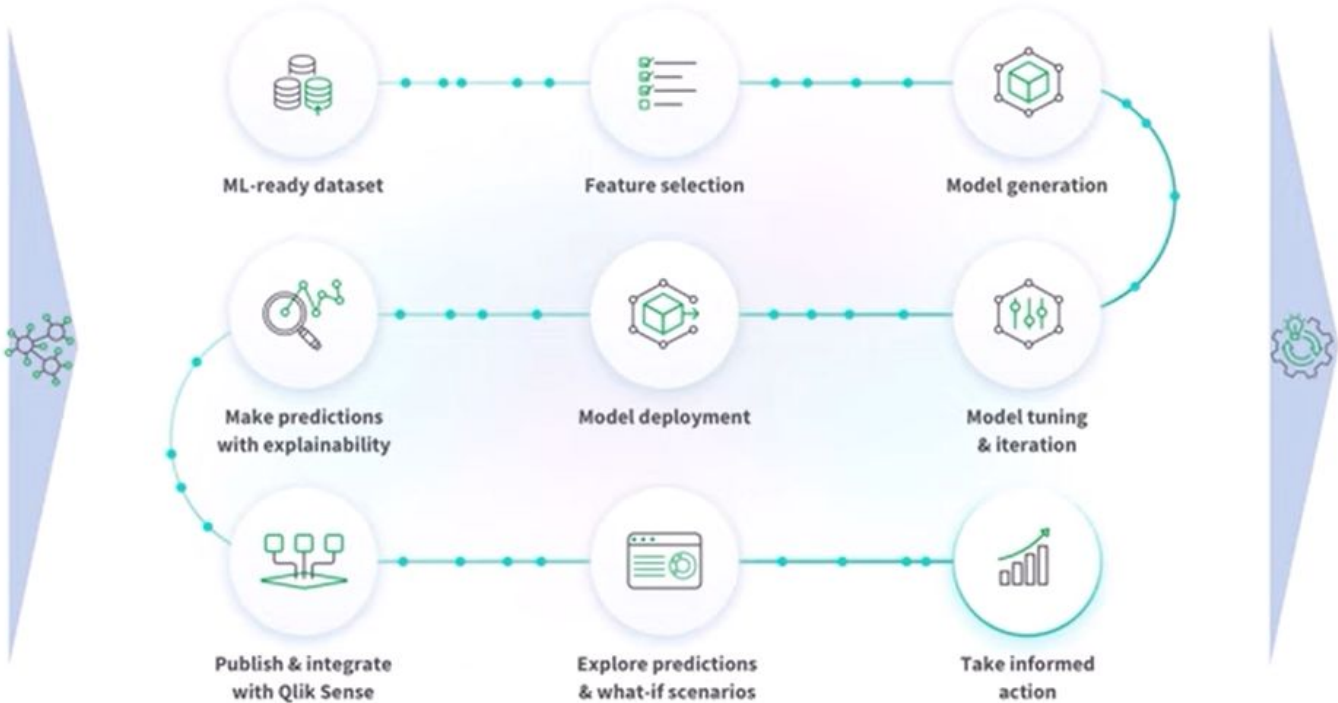
- Inventory Stock-Outs Prediction
- Supply Chain Performance / Bottlenecks
- Transportation Optimization

Service & Support

- Support Case Prediction
- Predictive Maintenance



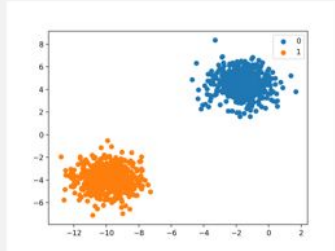
Workflow de Projeto



Tipos de problemas que o Qlik AutoML resolve

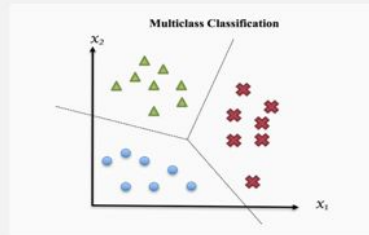
Classificação Binária

Qualquer questão que a resposta seja **Sim** ou **Não**



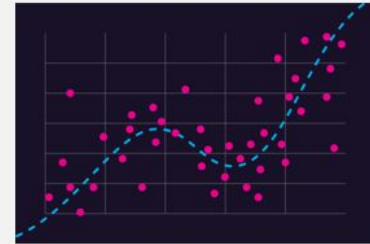
Classificação Multi-Class

Questões que o resultado seja de múltipla escolha



Regressão

Prever um número em algum ponto no futuro



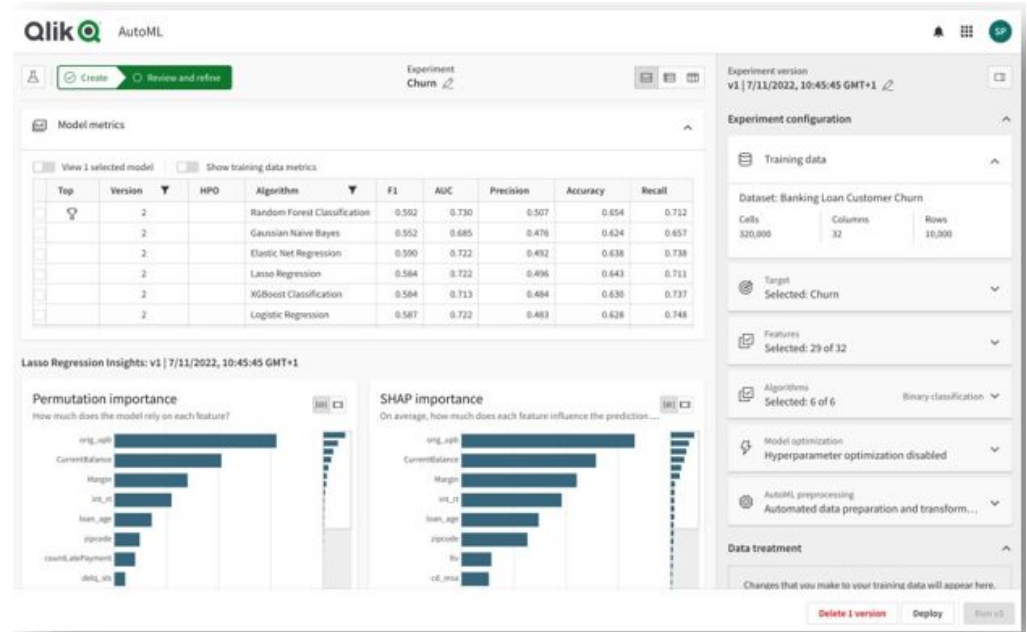
Algoritmos disponíveis

Binary & Multiclass

- CatBoost Classification
- Elastic Net Regression
- Gaussian Naive Bayes
- Lasso Regression
- LightGBM Classification
- Logistic Regression
- Random Forest Classification
- XGBoost Classification

Regressão

- CatBoost Regression
- LightGBM Regression
- Linear Regression
- Random Forest Regression
- SGD Regression
- XGBoost Regression





Slide Demo

DataSet - Base de Treinamento

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp
<p>1 891</p>	<p>0 1</p>	<p>1 3</p>	Braund, Mr. Owen Harris Cumings, Mrs. John Bradley (Florence Briggs Thayer) Heikkinen, Miss. Laina	male 577 female 314	<p>0,42 80</p>	<p>0</p>
Tipo de dados: INTEGER Numérico	Tipo de dados: INTEGER Numérico	Tipo de dados: INTEGER Numérico	Tipo de dados: STRING Categórico	Tipo de dados: STRING Categórico	Tipo de dados: DOUBLE Numérico	Tipo de dados: INT Numérico
Valores distintos: 891 Valores nulos: 0	Valores distintos: 2 Valores nulos: 0	Valores distintos: 3 Valores nulos: 0	Valores distintos: 891 Valores nulos: 0	Valores distintos: 2 Valores nulos: 0	Valores distintos: 89 Valores nulos: 0	Valores distintos: 7 Valores nulos: 0
1	0	3	Braund, Mr. Owen Harris	male	22	1
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1
3	1	3	Heikkinen, Miss. Laina	female	26	0
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1

DataSet - Base de Treinamento

Bas

Parch	Ticket	Fare	Cabin	Embarked
<p>Avg (0.38)</p>	<p>Avg (260.32k)</p>	<p>Avg (32.2)</p>	<p>null 687</p> <p>null 687</p> <p>Outro -483</p>	<p>S 644</p> <p>C 168</p> <p>Outro 79</p>
<p>Tipo de dados: INTEGER</p> <p>Numérico</p> <p>Valores distintos: 7</p> <p>Valores nulos: 0</p>	<p>Tipo de dados: STRING</p> <p>Catagórico</p> <p>Valores distintos: 681</p> <p>Valores nulos: 0</p>	<p>Tipo de dados: DOUBLE</p> <p>Numérico</p> <p>Valores distintos: 248</p> <p>Valores nulos: 0</p>	<p>Tipo de dados: STRING</p> <p>Catagórico</p> <p>Valores distintos: 148</p> <p>Valores nulos: 0</p>	<p>Tipo de dados: STRING</p> <p>Catagórico</p> <p>Valores distintos: 4</p> <p>Valores nulos: 0</p>
0	A/5 21171	7.25		S
0	PC 17599	71.2833	C85	C
0	STON/O2. 3101282	7.925		S
0	113803	53.1	C123	S

Avaliação de Desempenho



👤 Criar **Revisar e refinar**

Experimento de ML
Experimento Titanic ✎



Métricas do modelo

Versão ▾ Algoritmo ▾ Mais filtros de modelo ▾

Mostrar métricas de dados de treinamento

	Superior	Versão	HPO	Algoritmo	F1	AUC	Precisão	Exatidão	Recall	Hiperparâmetros	
<input checked="" type="checkbox"/>	🏆	1		Classificação Catboost	0,819	0,902	0,787	0,855	0,855	📄	
<input type="checkbox"/>		1		Classificação do LightGBM	0,763	0,903	0,699	0,799	0,841	📄	
<input type="checkbox"/>		1		Gaussian Naive Bayes	0,783	0,893	0,783	0,832	0,783	📄	
<input type="checkbox"/>		1		Regressão de rede elástica	0,787	0,895	0,709	0,816	0,884	📄	
<input type="checkbox"/>		1		Regressão lasso	0,787	0,893	0,709	0,816	0,884	📄	
<input type="checkbox"/>		1		Classificação XGBoost	0,774	0,879	0,779	0,827	0,768	📄	

Classificação Catboost Ideias: v1 | 2023-06-23 03:21:15

Permutation importance

SHAP importance

Avaliação de Desempenho



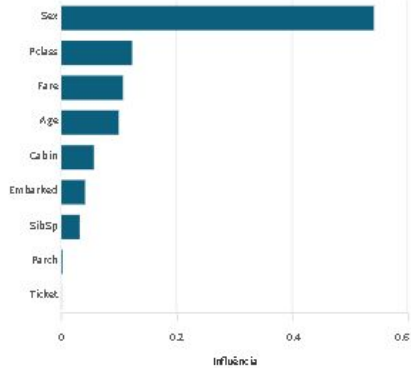
	Superior	Versão	HPO	Algoritmo	F1	AUC	Precisão	Exatidão	Recall
<input checked="" type="checkbox"/>		1		Classificação Catboost	0,819	0,902	0,787	0,855	0,855
<input type="checkbox"/>		1		Classificação do LightGBM	0,763	0,903	0,699	0,799	0,841
<input type="checkbox"/>		1		Gaussian Naive Bayes	0,783	0,893	0,783	0,832	0,783
<input type="checkbox"/>		1		Regressão de rede elástica	0,787	0,895	0,709	0,816	0,884
<input type="checkbox"/>		1		Regressão lasso	0,787	0,893	0,709	0,816	0,884
<input type="checkbox"/>		1		Classificação XGBoost	0,774	0,879	0,779	0,827	0,768
<input type="checkbox"/>		.			-----	-----	-----	-----	-----

Analise Gráfica



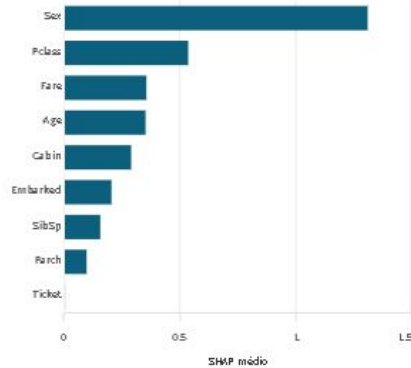
Permutation importance

Quanto a modelo depende de cada recurso?



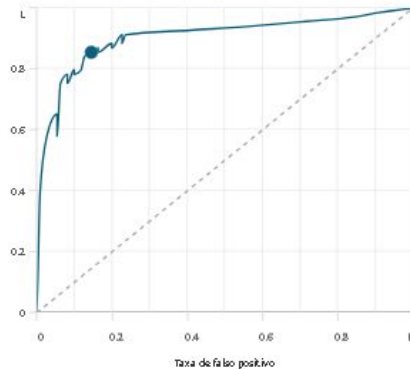
SHAP importance

Em média, quanto cada recurso influencia a previsão de 'Survived'?



Curva ROC

Quão precisa é a modelo em prever a classe positiva?



Confusion matrix

Quão precisa é a modelo para prever as classes alvo?

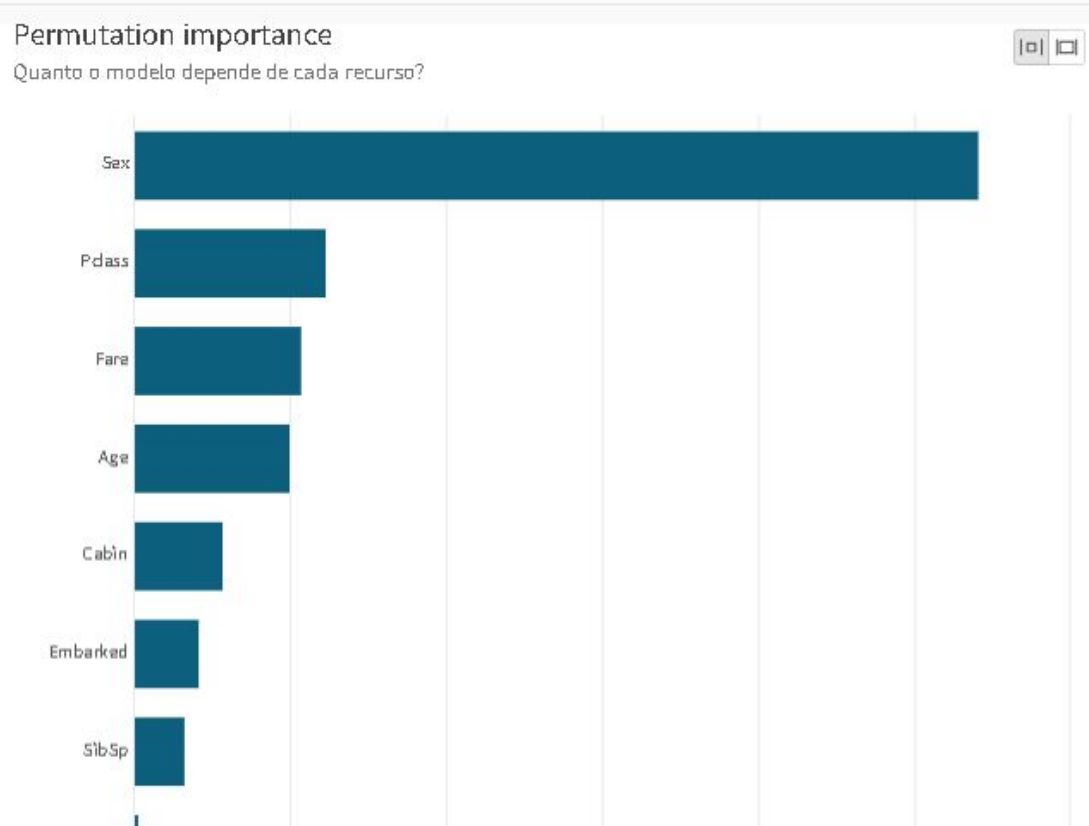


	Verdadeira real [1]	Falsa real [0]	
Verdadeira prevista [1]	Verdadeiro positivo 59 Recall	Falso positivo 16 Fallout	75
Falsa prevista [0]	Falso negativo 10 Miss rate	Verdadeiro negativo 34 Specificity	104
	69	110	Total 179

Análise Gráfica



Quanto o modelo depende cada recurso.

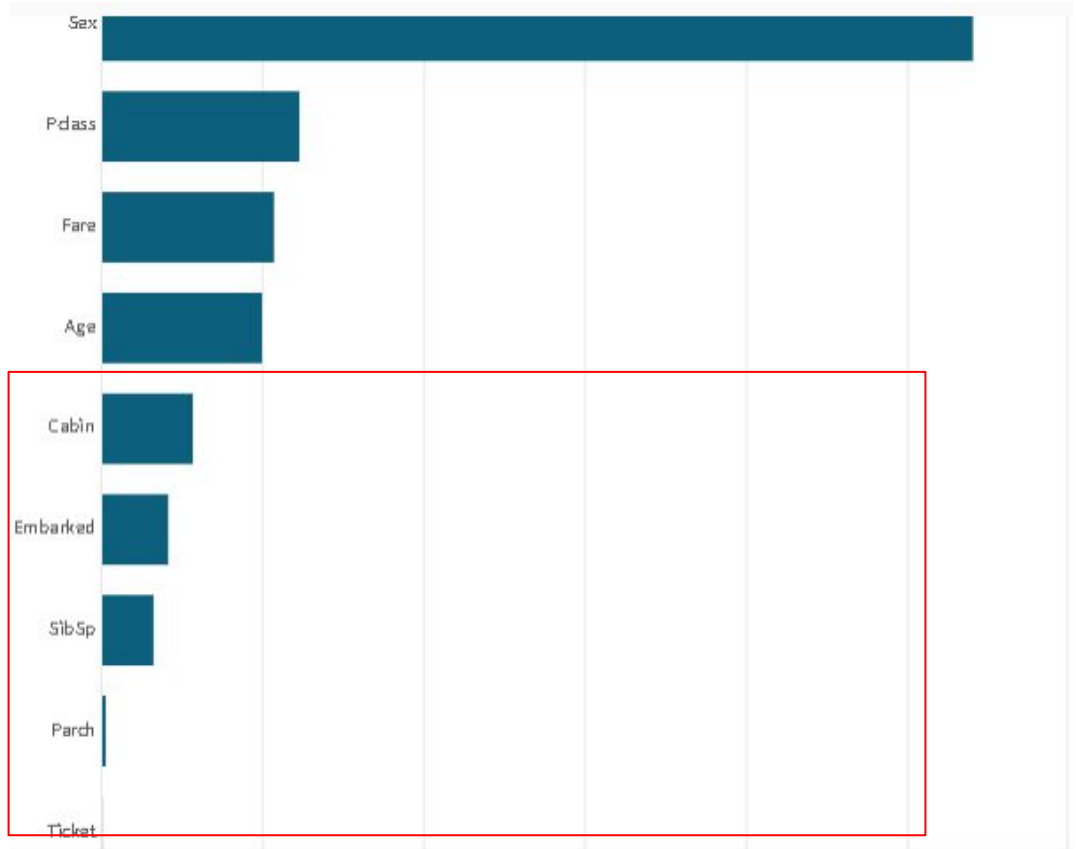


Análise Gráfica



Identificamos features que não são relevantes para o modelo.

Vamos analisar e entender os dados

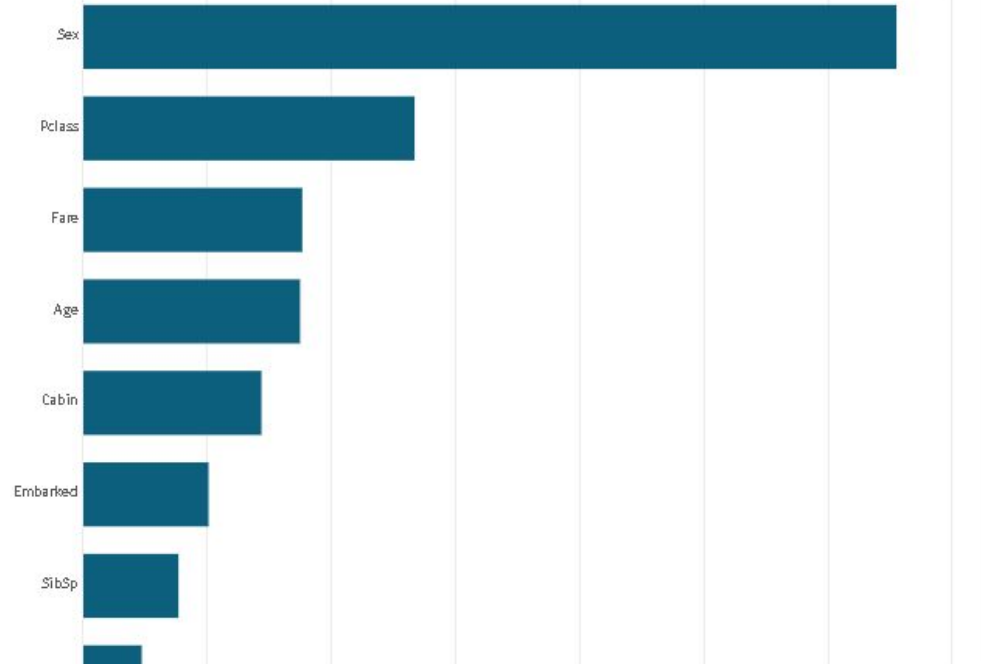


Analise Gráfica

Podemos avaliar em média o quanto cada recurso influencia a previsão de 'Survived'?

SHAP importance

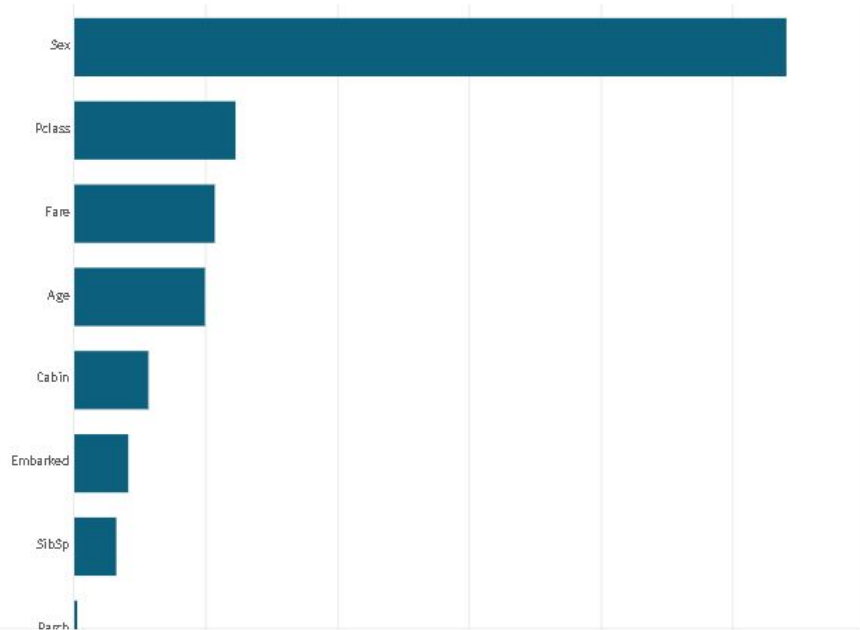
Em média, quanto cada recurso influencia a previsão de 'Survived'?



Análise Gráfica

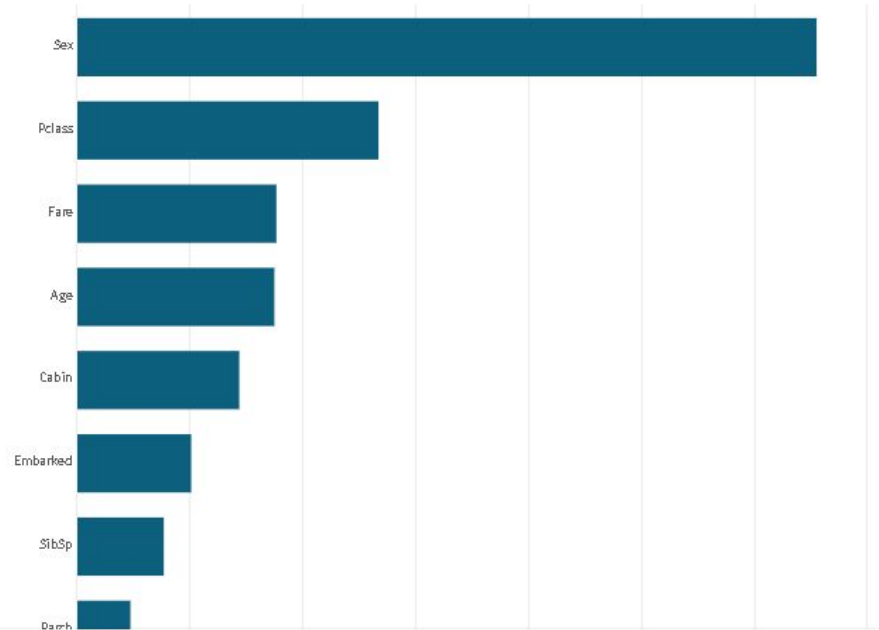
Permutation importance

Quanto o modelo depende de cada recurso?



SHAP importance

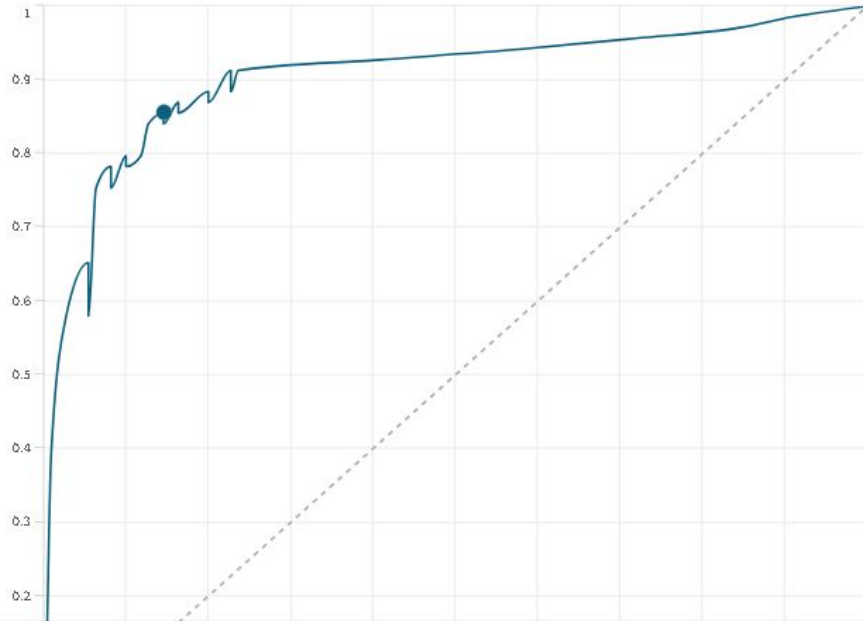
Em média, quanto cada recurso influencia a previsão de 'Survived'?



Análise Gráfica

Curva ROC

Quão preciso é o modelo em prever a classe positiva?



Confusion matrix

Quão preciso é o modelo para prever as classes alvo?

	Verdadeiro real (1)	Falso real (0)	
Verdadeira prevista (1)	Verdadeiro positivo 59 <i>Recall</i>	Falso positivo 16 <i>Fallout</i>	75
Falsa prevista (0)	Falso negativo 10 <i>Miss rate</i>	Verdadeiro negativo 94 <i>Specificity</i>	104

Depois de algumas iterações escolhemos o modelo

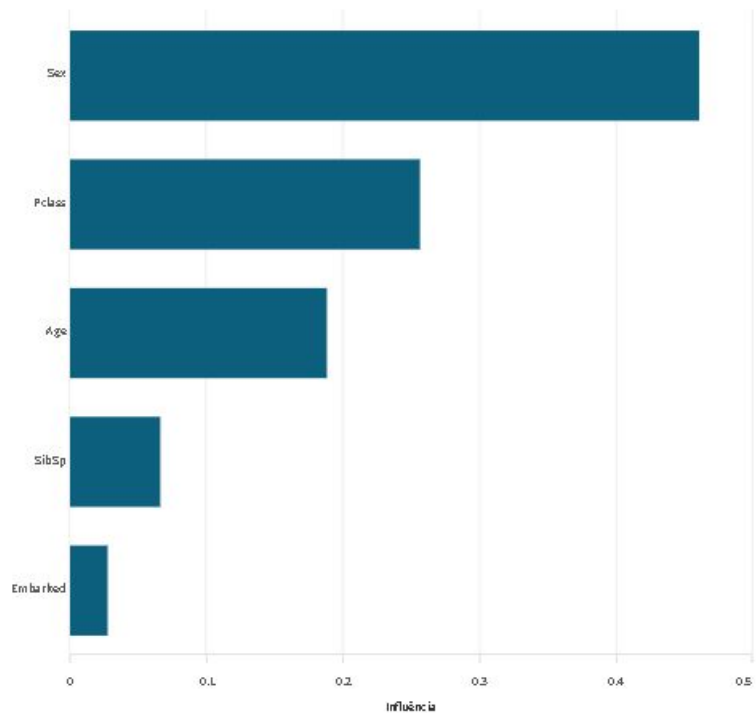


	Superior	Versão	HPO	Algoritmo	F1	AUC	Precisão	Exatidão	Recall
<input checked="" type="checkbox"/>		7		Classificação do LightGBM	0,803	0,917	0,809	0,849	0,797
<input type="checkbox"/>		7		Classificação Catboost	0,783	0,905	0,685	0,804	0,913
<input type="checkbox"/>		7		Gaussian Naive Bayes	0,769	0,897	0,820	0,832	0,725
<input type="checkbox"/>		7		Regressão de rede elástica	0,765	0,892	0,713	0,804	0,826
<input type="checkbox"/>		7		Regressão lasso	0,765	0,892	0,713	0,804	0,826
<input type="checkbox"/>		7		Classificação XGBoost	0,789	0,910	0,744	0,827	0,841

Análise Gráfica

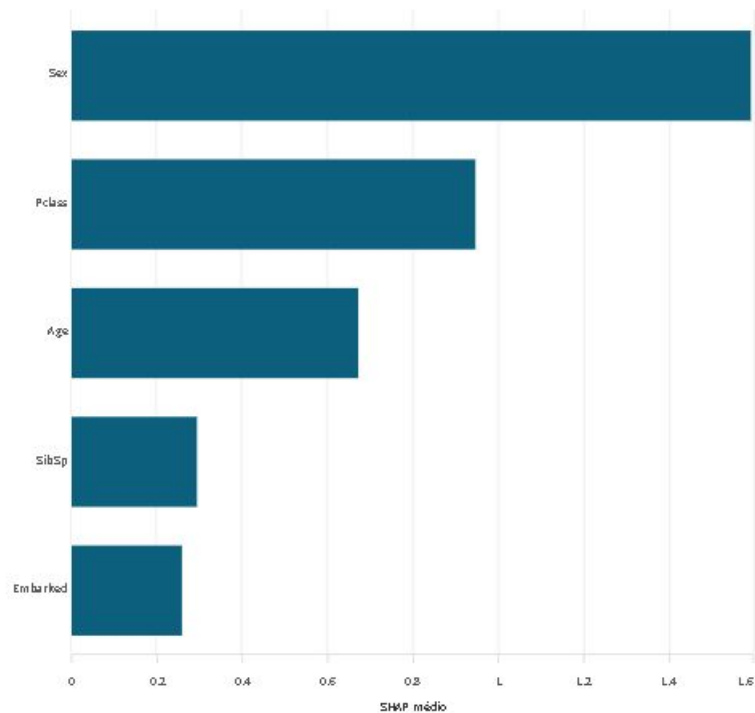
Permutation importance

Quanto a modelo depende de cada recurso?



SHAP importance

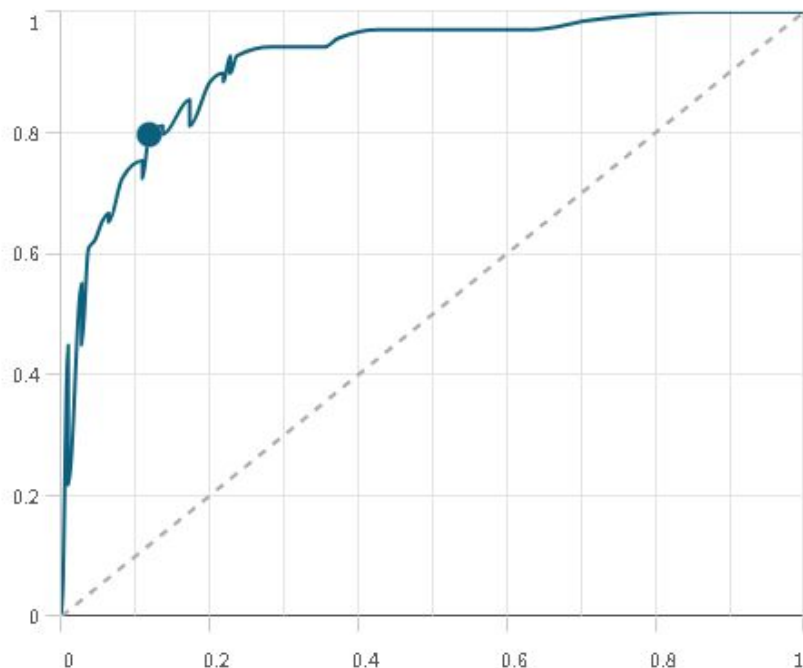
Em média, quanto cada recurso influencia a previsão de Survived?



Análise Gráfica

Curva ROC

Quão preciso é o modelo em prever a classe positiva?



Confusion matrix

Quão preciso é o modelo para prever as classes alvo?

	Verdadeiro real (1)	Falso real (0)	
Verdadeiro previsto (1)	Verdadeiro positivo 55 <i>Recall</i>	Falso positivo 13 <i>Fallout</i>	68
Falso previsto (0)	Falso negativo 14 <i>Miss rate</i>	Verdadeiro negativo 97 <i>Specificity</i>	111
	69	110	Total 179

Rodando o modelo

Visão geral da implementação

Previsões do conjunto de dados

Previsões em tempo real

 Exibir experimento ... 

Esquema de modelo



Conjunto de dados de treinamento
train

Recurso	Tipo de recurso
Pclass	Numérico
Age	Numérico
Embarked	Categórico
Sex	Categórico
SibSp	Numérico

Detalhes da implementação



Experimento Titanic_v7-LGBMC

O nome do experimento é 'Experimento Titanic'

A versão do experimento é v7

O tipo de experimento é Classificação binária

O algoritmo é Classificação do LightGBM

O conjunto de dados de treinamento é 'train'

Dados de treinamento


train

Nome do experimento

Experimento Titanic 

Criar previsão











Rodando o modelo

Implementação de ML
Experimento Titanic_v7-LGBMC 

Pesquisar conjuntos de dados

Espaços AutoML - Titanic Proprietário + conjunto de dados de Última modificação

AutoML - Titanic Apagar tudo

Nome	Proprietário	Última modificação	Espaço
 train.csv	 Pablo Labbe	há um dia	 AutoML - Titanic
 gender_submission.csv	 Pablo Labbe	há um dia	 AutoML - Titanic
 test.csv 	 Pablo Labbe	há um dia	 AutoML - Titanic

Configuração de previsão

Proprietário da configuração de previsão
Pablo Labbe

Aplicar dados

Conjunto de dados: Nenhum conjunto de dados de a...

Selecionar conjunto de dados de aplicação

Conjunto de dados de previsão
Conjunto de dados de previsão não criado

Opções de previsão
Opções de previsão não configuradas

Rodando o modelo

Esquema de modelo



Conjunto de dados de treinamento
train

Recurso	Tipo de recurso
---------	-----------------

Pclass	Numérico
--------	----------

Age	Numérico
-----	----------

Embarked	Categórico
----------	------------

Sex	Categórico
-----	------------

SibSp	Numérico
-------	----------

Aplicar esquema de conjunto de dados



Aplicar conjunto de dados
test.csv

Recurso	Tipo de recurso
---------	-----------------

Pclass	Numérico
--------	----------

Age	Numérico
-----	----------

Embarked	Categórico
----------	------------

Sex	Categórico
-----	------------

SibSp	Numérico
-------	----------

Configuração de previsão



Proprietário da configuração de previsão
Pablo Labbe



Aplicar dados
test.csv



Conjunto de dados de previsão

Conjunto de dados: Não criado
Espaço: Não selecionado

Nomear conjunto de dados de previsão

Opções de previsão

Rodando o modelo

Nomear conjunto de dados de previsão

[Saiba mais sobre a nomenclatura dinâmica de conjuntos de dados](#) ⓘ

Nome

Output_Previsão .csv

Espaço

AutoML - Titanic

- .csv ✓
- .qvd
- .parquet

Cancelar Confirmar

Aplicar dados
test.csv

Conjunto de dados de previsão

Conjunto de dados: Não criado

Arquivo: Não selecionado

Nomear conjunto de dados de previsão

Opções de previsão

Opções de previsão não configuradas

Programação

Programação não configurada

Cancelar


Salvar e fechar

Salvar e visualizar agora

Rodando o modelo

Implementação de ML

Experimento Titanic_v7-LGBMC 

Esquema de modelo	
 Conjunto de dados de treinamento train	
Recurso	Tipo de recurso
Pclass	Numérico
Age	Numérico
Embarked	Categórico
Sex	Categórico
SibSp	Numérico


Aplicar esquema de conjunto de dados	
 Aplicar conjunto de dados test.csv	
Recurso	Tipo de recurso
Pclass	Numérico
Age	Numérico
Embarked	Categórico
Sex	Categórico
SibSp	Numérico

Erros

Output_Previsão_Erros .csv

Coluna de índice:

- Índice de linha do AutoML (gerado automaticamente)
- PassengerId
- Name

 Programação

Programação não configurada

[Criar programação](#)


Cancelar

Salvar e fechar

Salvar e visualizar agora

Rodando o modelo

Implementação de ML
Experimento Titanic_v7-LGBMC 

Proprietário	Aplicar conjunto de dados	Conjunto de dados de previsão	Espaço	Incluído nas previsões	Última execução	Próxima execução	Último status	Ações
Pablo Labbe	test.csv	Output_Previsão.csv	AutoML - Titanic	Erros, SHAP	há segundos		<i>Executando</i>	

Linhas por página

50 

1-1 de 1



[Criar previsão](#)

Visão geral da previsão

Testing Overview



Age

Cabin

Embarked

Pclass

SibSp

Parch

Survived_predicted

Passengers

418

Avg(Age)

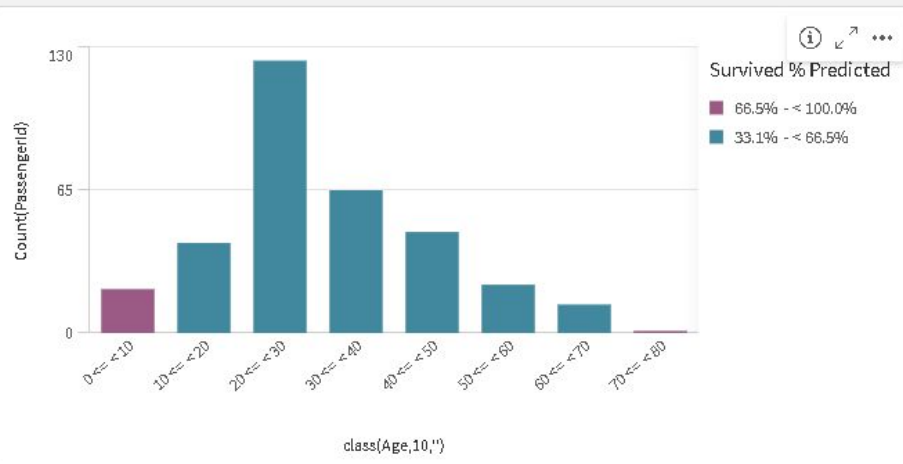
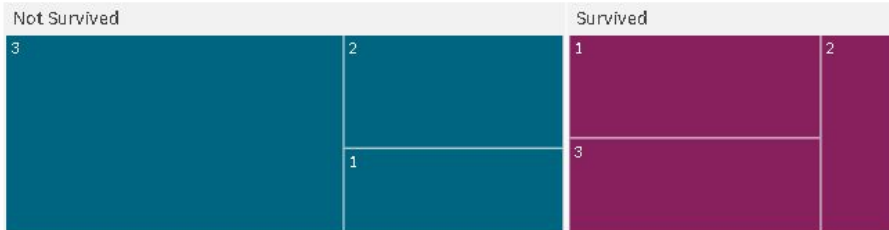
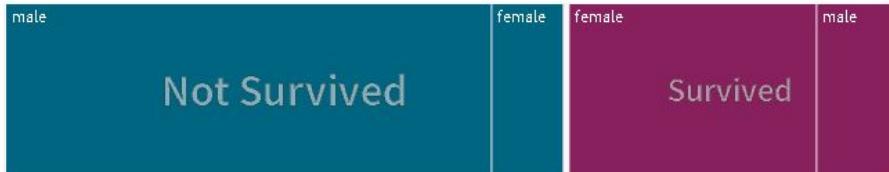
30.3

Survived Predicted

154

Survived % Predicted

36.8%



Crianças

Testing Overview



Age

Cabin

Embarked

Pclass

SibSp

Parch

Survived_predicted

Passengers

20

Avg(Age)

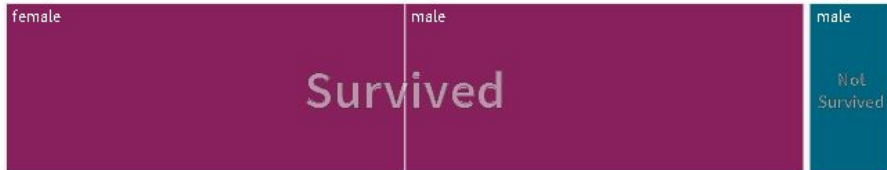
3.95

Survived Predicted

18

Survived % Predicted

90.0%



Faixa de 20 a 30 anos

Testing Overview



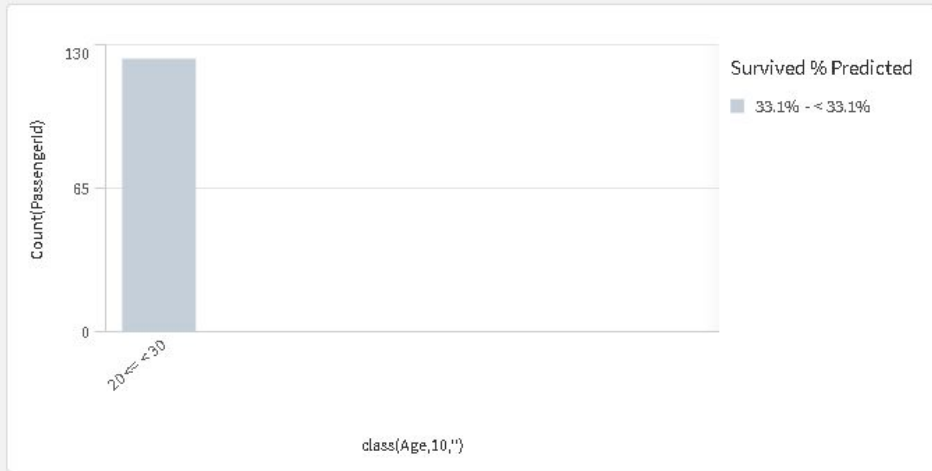
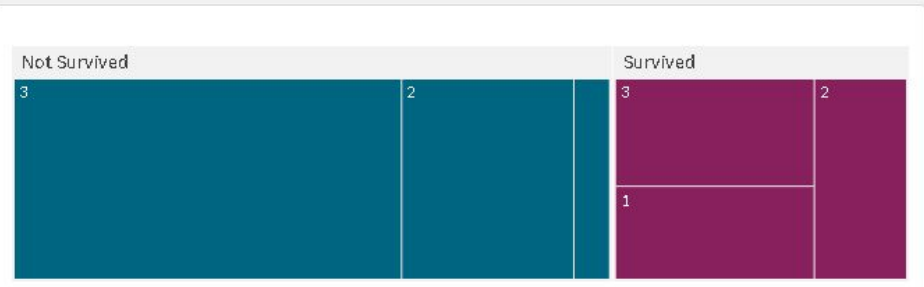
Age Cabin Embarked Pclass SibSp Parch Survived_predicted

Passengers
124

Avg(Age)
24.24

Survived Predicted
41

Survived % Predicted
33.1%



Taxa sobrevivência de quem estava na 3ª Classe

Testing Overview



Age Cabin Embarked Pclass SibSp Parch Survived_predicted

Passengers

218

Avg(Age)

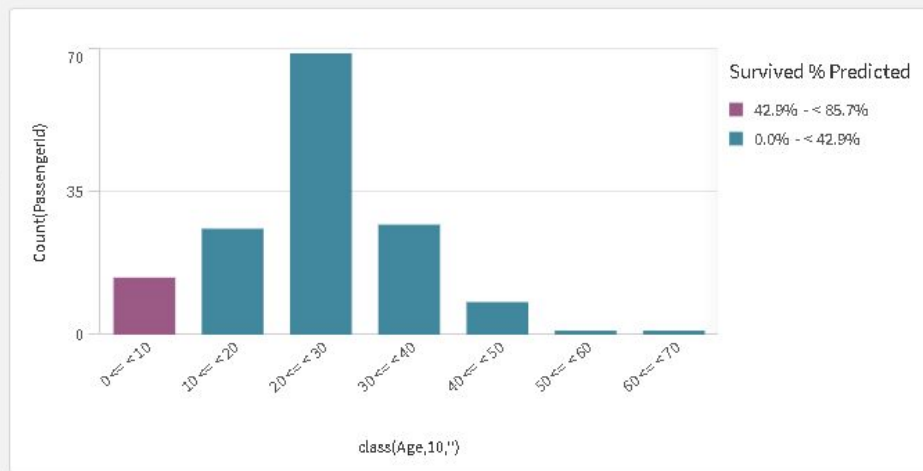
24.08

Survived Predicted

58

Survived % Predicted

26.6%



Taxa de sobrevivência de filhos e Pais

Testing Overview



Age Cabin Embarked Pclass SibSp **Parch** Survived_predicted

Passengers

94

Avg(Age)

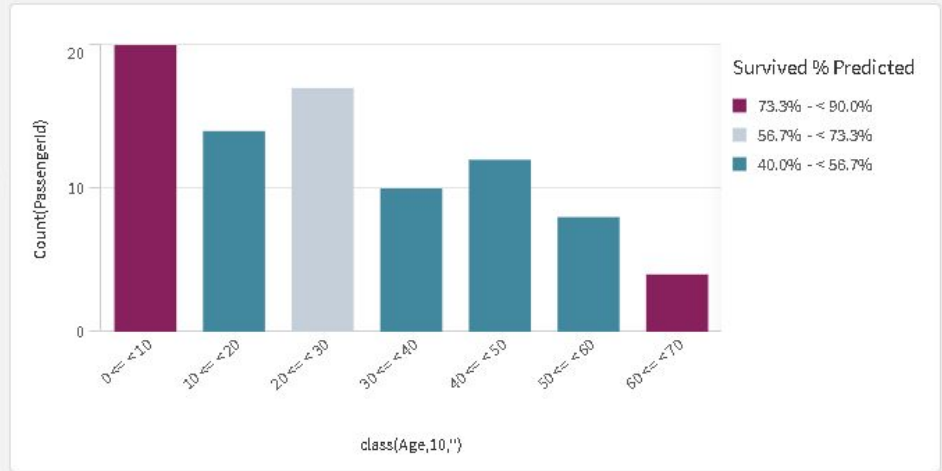
26.66

Survived Predicted

56

Survived % Predicted

59.6%



Taxa de sobrevivência de quem não tem filhos

Passengers

324

Avg(Age)

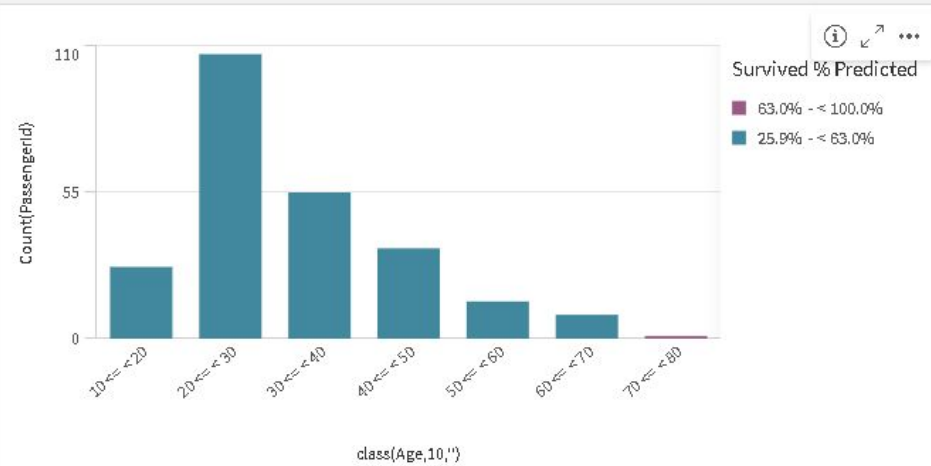
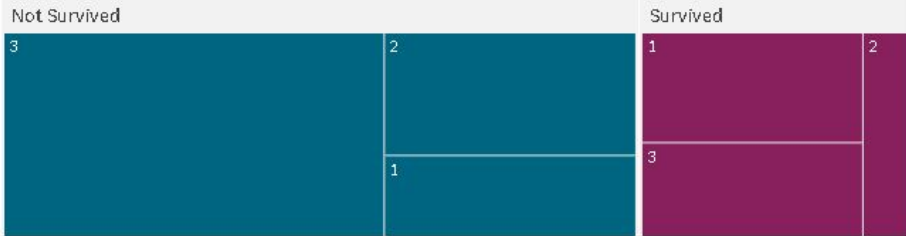
31.55

Survived Predicted

98

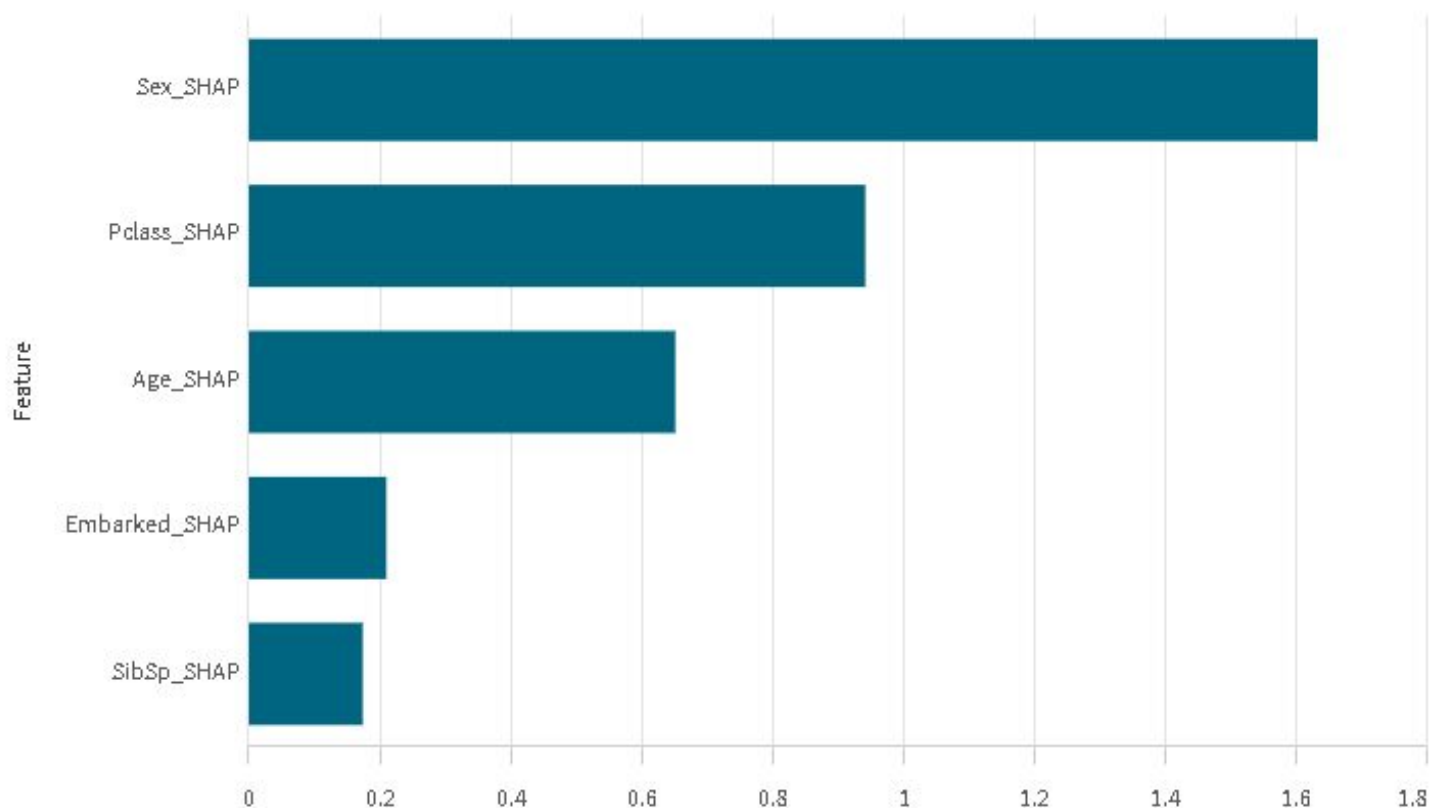
Survived % Predicted

30.2%



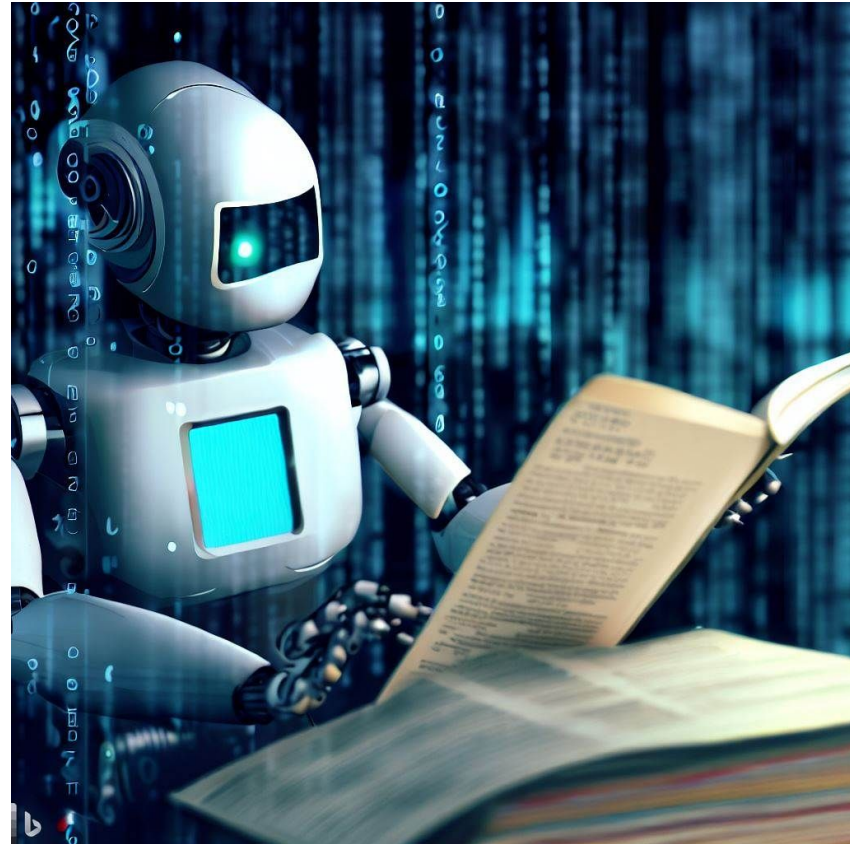
Visualização

Feature Importance



Conclusão

- O Automated Machine Learning simplifica o processo de escolha de modelos de machine learning, tornando-o mais acessível para os times de Analytics.
- Ele oferece vantagens significativas, como economia de tempo, melhor desempenho e redução de erros.
- Qlik AutoML se destaca pela facilidade, velocidade e experimentos ilimitados
- Qualidade no dado segue sendo muito importante !



Obrigado !

Contatos:

pablo.labbe@imaps.com.br

Linkedin: pablolabbe



iMaps