

Aplicação de Análise de Textos com o Orange Canvas

Douglas Hermes Fonseca Silva

Karine França de Souza

Neander Ferreira Almeida

Rodrigo da Silva Marques

INOVA
2ª Semana de Inovação

 **prodemge**


www.uncc.edu



Campo <descricao>

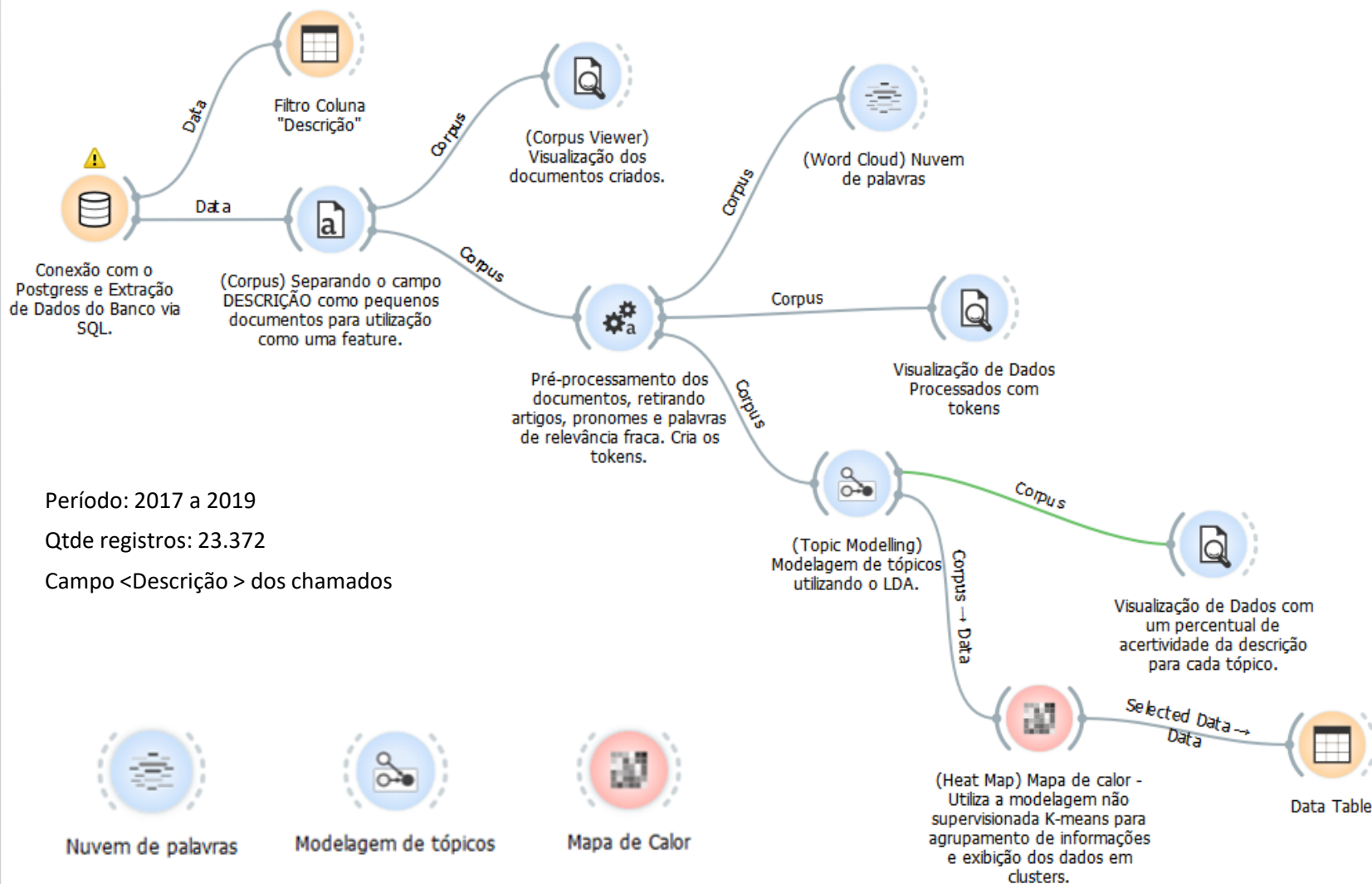


Criando o modelo



Orange

communication skills
predictive
statistics
SQL R hadoop
machine learning
python SAS
computer science
data visualization
clustering



[illegible]

Topic Modelling

Latent Dirichlet Allocation

Number of topics: 5

Topics

- 1: digital, procedimento, renovacao, solicitacao, assinatura, utilizacao, processo, retorno, cpf, validade
2: senha, pin, puk, token, erro, cac, acessar, instalacao, email, receita
3: token, a3, cartao, cnpj, cpf, aquisicao, certificados, compra, informacao, informacoes
4: validacao, social, pedido, emissao, titular, certificados, empresa, numero, presencial, assunto
5: suporte, nota, sistema, fiscal, problema, erro, programas, protocolo, emitir, chamado

Topico	Palavras	Assunto
1	digital, procedimento, renovacao, solicitacao, assinatura, utilizacao, processo, retorno, cpf, validade	RENOVAÇÃO
2	senha, pin, puk, token, erro, cac, acessar, instalacao, email, receita	SENHA
3	token, a3, cartao, cnpj, cpf, aquisicao, certificados, compra, informacao, informações	AQUISIÇÃO
4	validacao, social, pedido, emissao, titular, certificados, empresa, numero, presencial, assunto	INFORMAÇÕES SOBRE DOCUMENTAÇÃO PARA VALIDAÇÃO
5	suporte, nota, sistema, fiscal, problema, erro, programas, protocolo, emitir, chamado	ERRO NO USO

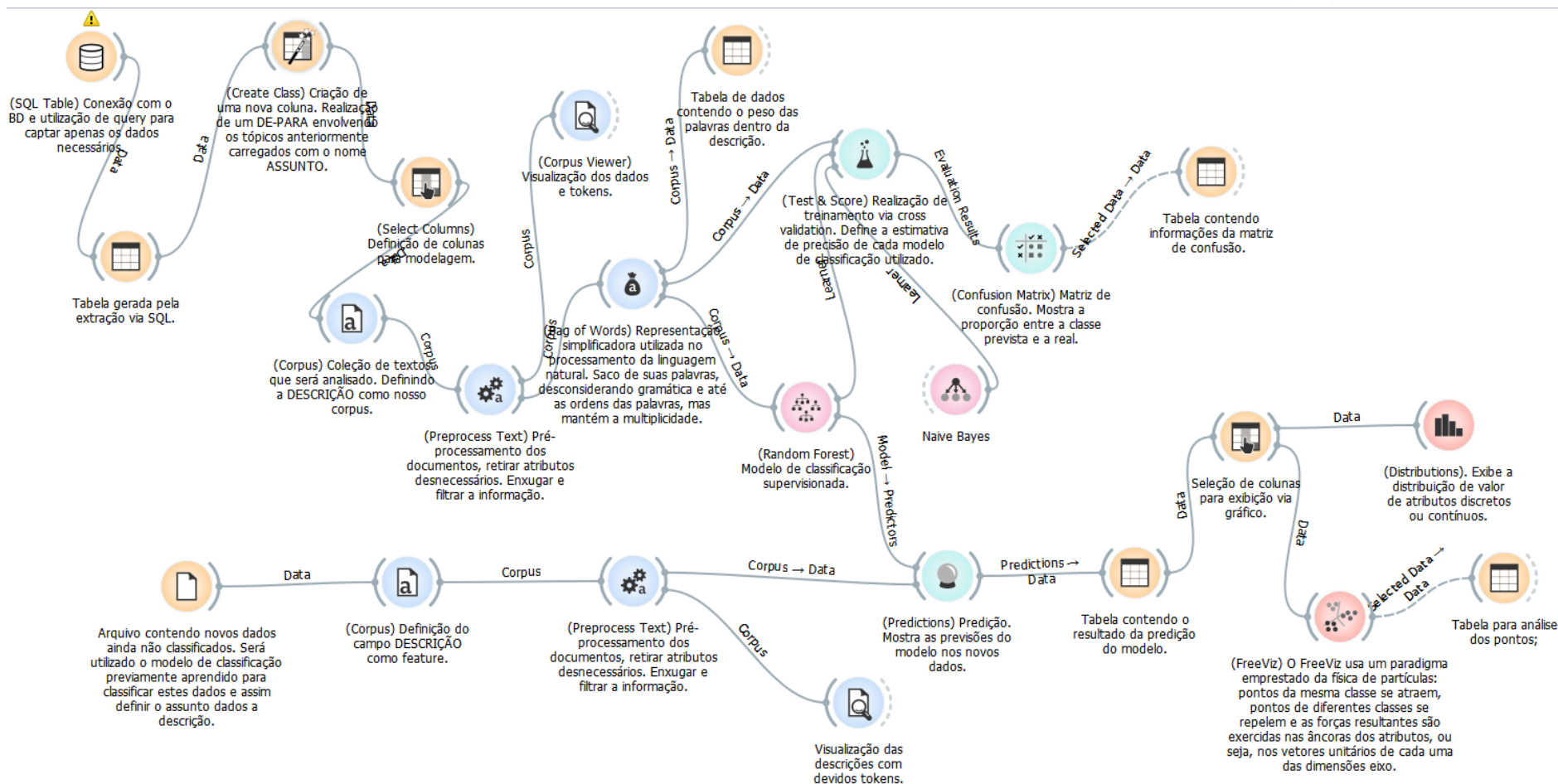


Colocando o modelo a prova !

communication skills
predictive
statistics
SQL R hadoop
machine learning
python SAS
computer science
data visualization
actual
cluster
classification

Teste: 1ª semana de Setembro/19

Qtde registros: 195



Resultado dos modelos de classificação

communication skills
predictive
matlab
statistics
SQL R hadoop
python SAS
scalable
computer science
data visualization
clustering
machine learning
python
statistics
SQL
R
hadoop
scalable
computer science
data visualization
clustering

Test & Score

Sampling

☒ Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

☐ Random sampling

Repeat train/test: 10

Training set size: 70 %

☒ Stratified

☐ Leave one out

☐ Test on train data

☐ Test on test data

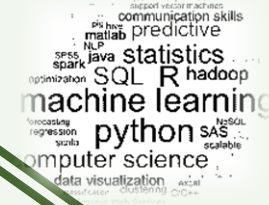
Target Class

(Average over classes)

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.969	0.851	0.851	0.851	0.851
Naive Bayes	0.957	0.810	0.810	0.811	0.810

Resultado da Matriz de Confusão

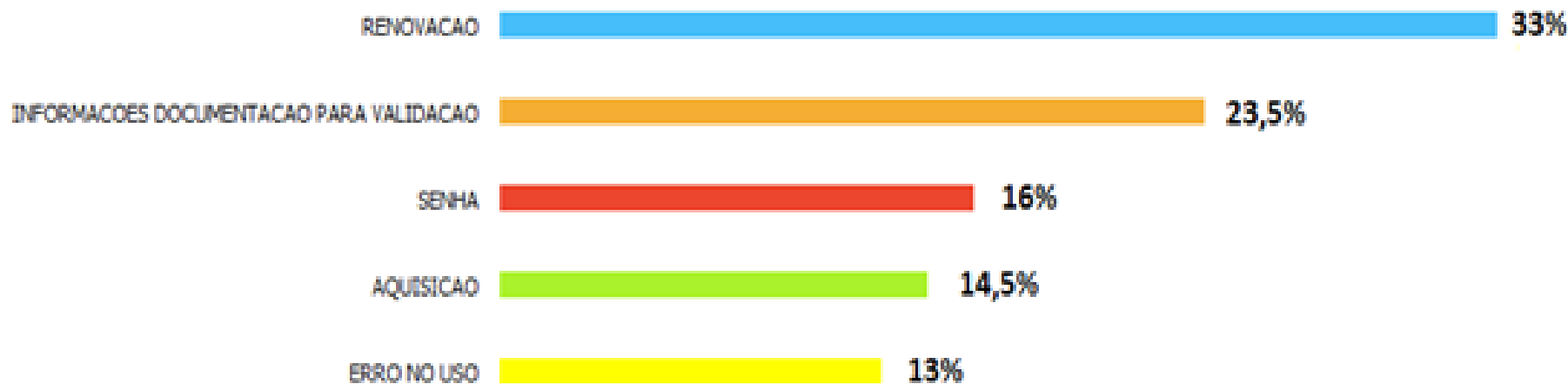


		Predicted						
		RENOVACAO	SENHA	AQUISICAO	INFORMACOES DOCUMENTACAO PARA VALIDACAO	ERRO NO USO		Σ
Actual	RENOVACAO	3412	133	249		165	154	4113
	SENHA	102	3959	120		84	173	4438
	AQUISICAO	195	110	4392		339	128	5164
	INFORMACOES DOCUMENTACAO PARA VALIDACAO	147	68	351	4740		103	5409
	ERRO NO USO	155	210	189		148	3546	4248
Σ		4011	4480	5301		5476	4104	23372
		Predicted						
		RENOVACAO	SENHA	AQUISICAO	INFORMACOES DOCUMENTACAO PARA VALIDACAO	ERRO NO USO		Σ
Actual	RENOVACAO	85.1 %	3.0 %	4.7 %		3.0 %	3.8 %	4113
	SENHA	2.5 %	88.4 %	2.3 %		1.5 %	4.2 %	4438
	AQUISICAO	4.9 %	2.5 %	82.9 %		6.2 %	3.1 %	5164
	INFORMACOES DOCUMENTACAO PARA VALIDACAO	3.7 %	1.5 %	6.6 %	86.6 %		2.5 %	5409
	ERRO NO USO	3.9 %	4.7 %	3.6 %		2.7 %	86.4 %	4248
Σ		4011	4480	5301		5476	4104	23372

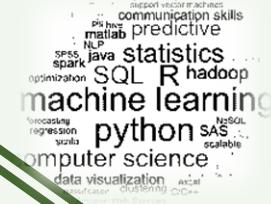
Resultado

1ª semana de Setembro/19: 195 registros

Percentual de assuntos encontrados



Resultado



assuntos
mais recorrentes

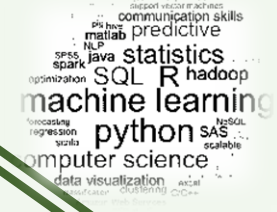
atendimento virtual
(Tokenho)

contratação do
certificado

melhoria do
atendimento

chamado melhor direcionado

CHAMADO	RESUMO	DESCRIÇÃO	ABERTURA	ASSUNTO SUGERIDO
INP1768040	Validação Fora da AR	Sr. Fernanda gostaria de solicitar informações para validação fora da AR, inicialmente para 2 certificados.	05/09/2019 11:15	INFORMACOES DOCUMENTACAO PARA VALIDACAO
INP1768669	Aproveitamento de mídia/ inclusão CAEPF	Sr. Alfredo informa que o certificado venceu e já possui token. Quer saber se pode aproveitá-lo na nova aquisição. Também relata que fez um pedido de A3 somente certificado e-CPF com CEI vinculado e já está agendado dia 10. Quer ser acompanhado na compra porque gostaria de incluir o CAEPF no certificado.	06/09/2019 09:21	AQUISICAO
INP1768756	SUPORTE	Cliente entrou em contato informando que no processo de importação do certificado pede uma senha privada. Cliente não sabe qual senha é esta.	06/09/2019 10:37	SENHA
INP1768870	Funcionamento	Cliente informa que tem o certificado e-CNPJ e relata que esta tentando acessar o sistema conectivida social mas ocorre erro ao assinar o documento .	06/09/2019 12:48	ERRO NO USO
INP1772047	Renovação A3 on-line Pessoa Fisica.	Cliente solicita ajuda na renovação de seu certificado.	11/09/2019 09:27	RENOVACAO



Obrigado !