



teradata.

Inteligência Artificial Aplicada à Simulação de Cenários da Substituição Tributária

Reunião GT34

16/10/2018

José Márcio Faier
Cientista de Dados

Gilsomar Rezende
Engenheiro de Dados

Diego Fernandes
Arquiteto de Soluções

Nathalia Pietro
Gerente de Projetos

Vinícius Fonseca
Diretor de Serviços

Carlos Bokor
Diretor Comercial

AGENDA

- CONTEXTO E MOTIVAÇÃO
- HISTÓRICO DA ESTRUTURAÇÃO DOS DADOS NA SEFA/PR
- COMO ABORDAMOS O PROBLEMA
- PERGUNTA DE NEGÓCIO
- VISÃO DA SOLUÇÃO
- CLASSIFICADOR CEST / NCM
- ESTUDOS
- RESULTADOS
- PRÓXIMOS PASSOS

CONTEXTO E MOTIVAÇÃO

SUBSTITUIÇÃO TRIBUTÁRIA

ANÁLISE DE IMPACTO - EXCLUSÃO DE PRODUTOS



Parametrização
Responsáveis
- CNAES

Parametrização
dos Produtos
- NCM
- Descrição
- Finalidade
- Fabricação

Parametrização Arrecadação
- Arrecadação ICMS ST
- Recup/Ressarc. ICMS ST
- Saldo credor em GIA
- Estoque
- % Vendas para RN e SN

UM POUCO DO HISTÓRICO...



teradata.

CONTRATO 001/2015-CRE, DE 27 DE JULHO DE 2015

SERVIÇOS DE APOIO E MENTORIA EM
DATA WAREHOUSING



MARÇO
2012

- INFRA DW
- SERVIÇOS DW

SETEMBRO
2013

- LICENÇAS BI
- SERVIÇOS BI

JULHO
2015

- MENTORIA DW

JULHO
2017

- NOVA INFRA DW

JANEIRO
2013

- LICENÇAS ETL

OUTUBRO
2013

- SERVIÇOS DW

AGOSTO
2016

- SERVIÇOS BI

teradata.

GRANDES VOLUMES DE DADOS

NF-e, NFC-e, EFD, CV115 ...

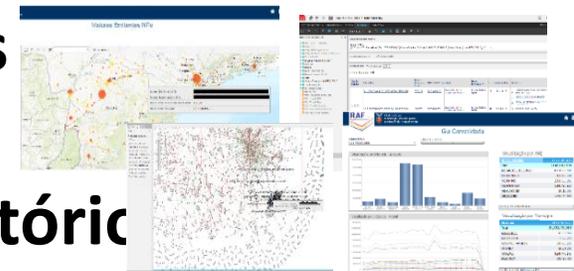
ORIGENS DE DADOS INTEGRADOS



MAIS ORIGENS
DE DADOS ESTÃO
EM FASE DE
IMPLEMENTAÇÃO
(E OUTRAS NA FILA!)

335 milhões | 195M documentos fiscais
140M documentos escriturados
de documentos por mês

250 auditores fiscais
20 mil consultas
2 mil diferentes relatórios
4 bilhões de registros extraídos



MicroStrategy [Business Intelligence]

Usuários Avançados

Aplicações



teradata. [Data Warehouse]

15 mil processos de carga

(workflows, worklets, sessions)

21 mil objetos de banco de dados

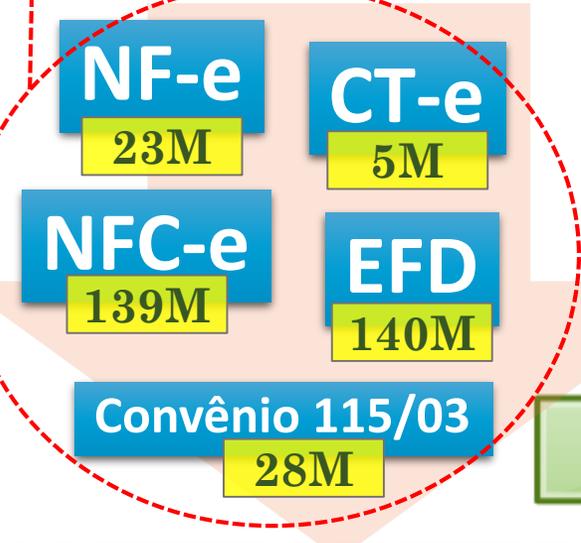
(databases, tabelas, views, procedures, macros, triggers)

17,5 milhões de consultas por mês

(cargas de dados, malhas fiscais, autorregularização, BI MicroStrategy)

44 TB de espaço ocupado

teradata.



Informatica
[Integração de Dados]

COMO ABORDAMOS O PROBLEMA DE NEGÓCIO

RACE

NOSSA ABORDAGEM

METODOLOGIA

- Uma abordagem que **integra ciência de dados, conhecimento de negócio, e criatividade** para maximizar o Retorno sobre o Investimento (ROI)

RACE (Engajamento Rápido de Consultoria Analítica)



QUEM ESTÁ ENVOLVIDO?



Cientista de Dados
Engenheiro de Dados
Especialistas do Negócio

DURAÇÃO

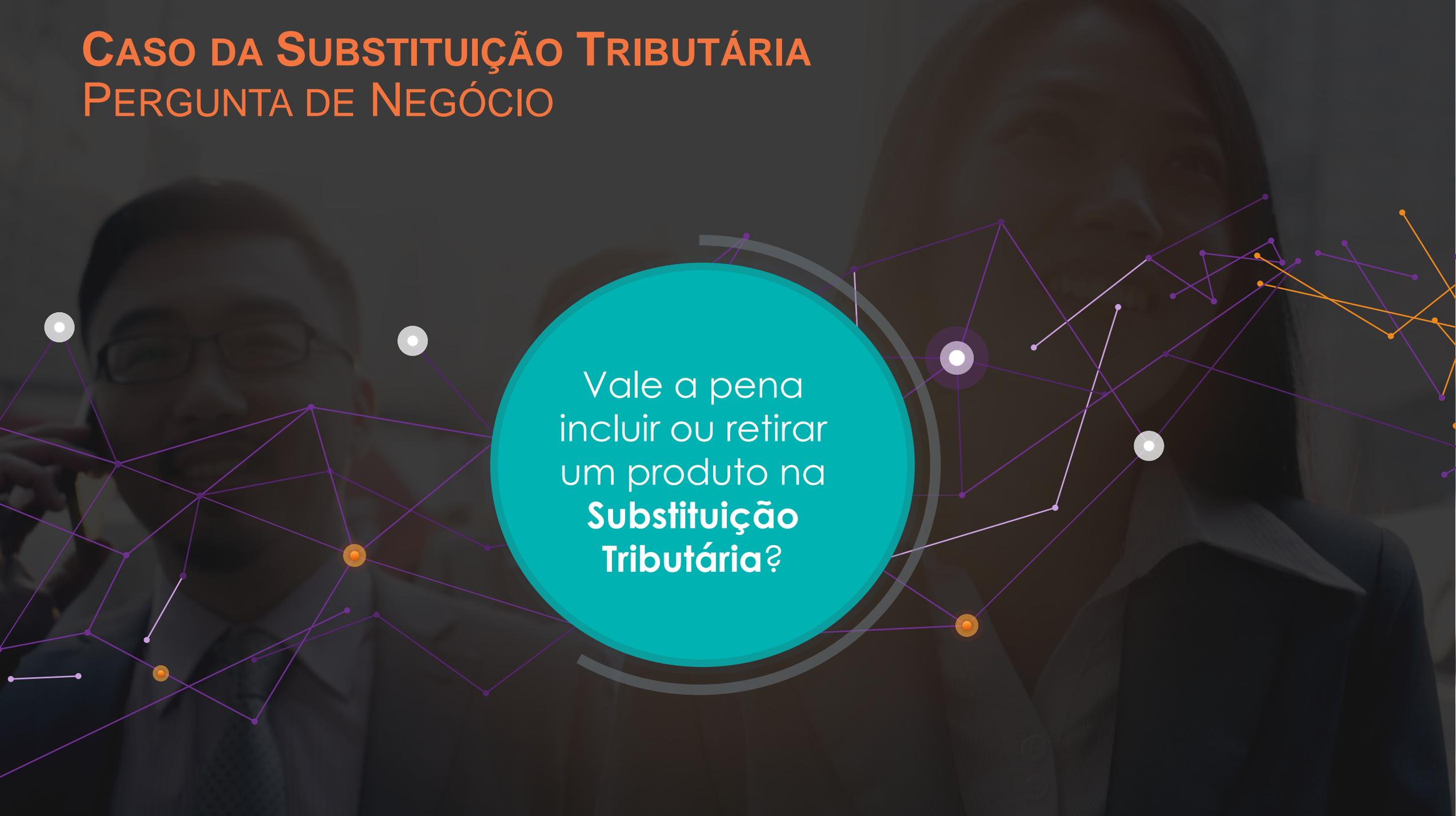


8 -10 semanas
por ciclo*

1 – Roadmap de valor de negócio (preferido), mas o Caso de Uso é requerido
5 - Escopo e extensão possível para incluir uma fase de implantação
*Duração pode variar devido à disponibilidade de dados e complexidade

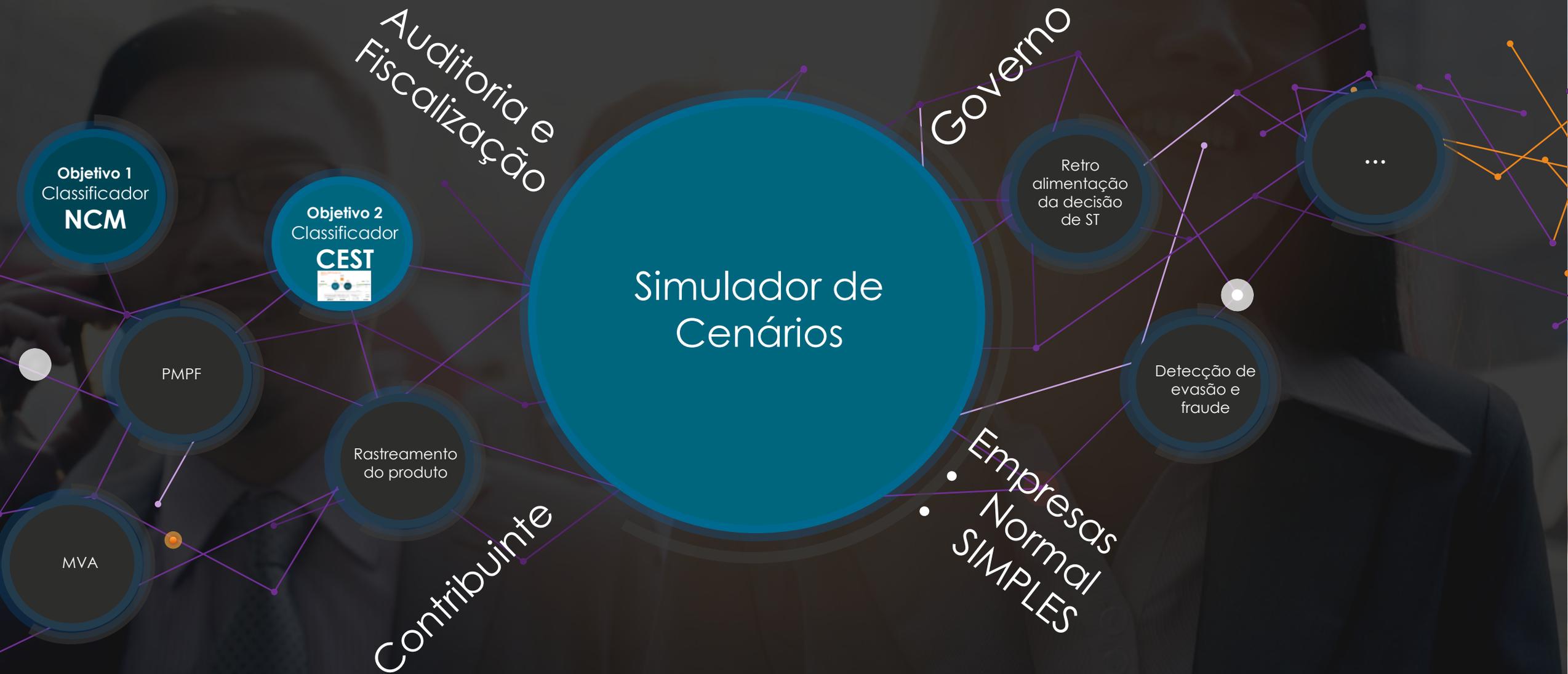
CASO DA SUBSTITUIÇÃO TRIBUTÁRIA

PERGUNTA DE NEGÓCIO



Vale a pena
incluir ou retirar
um produto na
**Substituição
Tributária?**

VISÃO DA SOLUÇÃO



MODELO CEST ALIMENTOS

CEST 17

Modelo

Entrada

Saída



Cest 17

DESCRIÇÃO	CEST PREDITO	BINARIO PREDITO
60-1193 - BEBIDA DE MARACUJA 27X200ML V - 1193	1711100	1.0
SUFRESH ABACAXI NECTAR 1L	1701000	1.0
Maio Liza 500g	1703900	1.0

Não Cest 17

REFRIGERANTE 2 LTS NEON LIMAO	1711100	0.0
-------------------------------	---------	-----

91%
Acurácia
itens

Modelo CEST
Alimentos 17
Binário

Dados com
CEST
Alimentos

79%
Acurácia
itens

87%
Acurácia
itens

Modelo CEST
Alimentos 17
Multiclasse

Conhecimento especialista
de 5 inspetorias/unidades
de negócio (IGF, IGT, IGA,
AGAA e Celepar)



5 Bilhões de itens de Nota
Fiscal prontos para análise
(Data Warehouse)

teradata.

+ 100 estudos analíticos
avancados



Detalhes



teradata.

RESULTADOS CEST

EXEMPLOS



DESCRIÇÃO	CEST PREDITO	BINARIO PREDITO	DESCRIÇÃO DA TABELA CEST
60-1193 - BEBIDA DE MARACUJA 27X200ML V - 1193	1711100	1.0	Refrescos e outras bebidas não alcoólicas, exceto os refrigerantes e as demais bebidas nos CEST 03.007.00 e 17.110.00
SUFRESH ABACAXI NECTAR 1L	1701000	1.0	Sucos de frutas ou de produtos hortícolas; mistura de sucos
Maio Liza 500g	1703900	1.0	Maionese em embalagens imediatas de conteúdo inferior ou igual a 650 g, exceto as embalagens contendo envelopes individualizados (sachês) de conteúdo inferior ou igual a
CHOCOLATE LACTA 20GR A	1700300	1.0	Chocolate em barras, tabletes ou blocos ou no estado líquido, em pasta, em pó, grânulos ou formas semelhantes, em recipientes ou embalagens imediatas de conteúdo inferior ou igual a 2 kg
REFRIGERANTE 2 LTS NEON LIMAO	1711100	0.0	Refrescos e outras bebidas não alcoólicas, exceto os refrigerantes e as demais bebidas nos CEST 03.007.00 e 17.110.00
BANANA MANIA LIGHT 150GR	1704200	1.0	Barra de cereais
Choc Trento 32g Pret	1705800	1.0	"Waffles" e "wafers"- com cobertura
BISCOITO RECHEADO CLUB 115G CHOCOLATE	1705300	1.0	Biscoitos e bolachas derivados de farinha de trigo; (exceto dos tipos "cream cracker", "água e sal", "maisena", "maria" e outros de consumo popular que não sejam adicionados de cacau, nem recheados, cobertos ou amanteigados, independentemente de sua denominação comercial)
M REN EXPRESS MAIS 87G	1704700	1.0	Massas alimentícias tipo instantânea
SNACK ELMA CHIPS DORITOS MEDIO PCT 55G	1703000	1.0	Produtos à base de cereais, obtidos por expansão ou torrefação

Resultados NCM

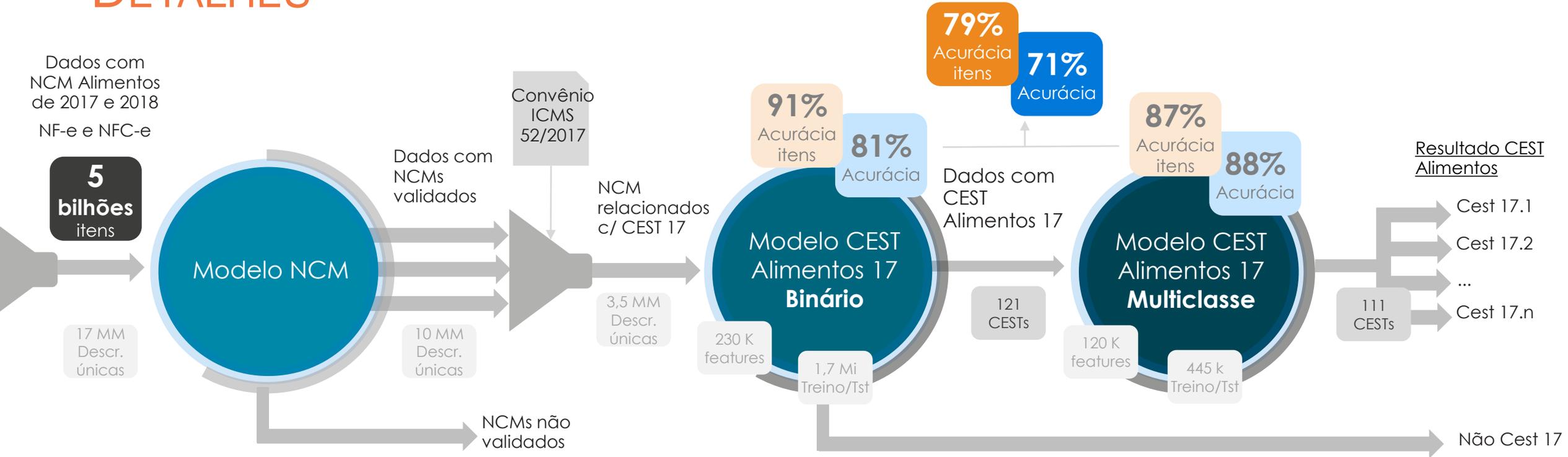
MODELO NCM

EXEMPLOS

Descrição Item	NCM Original	Descrição NCM Original	NCM Predito	Descrição NCM Predito
Ruffles Original 57g			20052000	BATATAS Ñ CONS.VINAGRE/ÁCIDO ACÉTICO, Ñ CONG.
SORVETE TRUFA 70G			21050010	SORVETES MESMO C/CACAU, D/C/INF A 1KG INC
VINHO TINTO CREVELIM 880ML	10061092	ARROZ ("PADDY") COM CASCA, NÃO PARBOILIZADO	22042100	VINHOS E MOSTOS D/UVAS, FERM.IMP.P/ALC.INF
Achoc. Toddy 400grs Un			18069000	OUTRAS PREPS.C/CACAU Ñ CIT.ANTERIORMENTE
Bisc Panco Golden Black Acai 140g Bisc Panco Golden Bl			19053100	BOLACHAS/BISC.ADICIONADOS DE EDULCORANTES
CHOCO GAROTO AO LEITE 100G	17049010	CHOCOLATE BRANCO SEM CACAU	18063210	CHOCOLATES Ñ RECHEADS, TABLETE, BARRAS, PAUS
Reserva Casillero del Diablo Concha y To	17049020	BOMBONS, CARAMELOS, CONFEIT.PASTILH.S/ CACAU	22042100	VINHOS E MOSTOS D/UVAS, FERM.IMP.P/ALC.INF
Arroz Sabor Sul Bco 2kg	18069000	OUTRAS PREPS.C/CACAU Ñ CIT.ANTERIORMENTE	10063021	ARROZ SEMIBRANQ.OU BRANQ. Ñ PARBOIL., POLIDO
CHICLETES TRIDENT SABORES			21069050	GOMAS DE MASCAR, SEM AÇÚCAR
PURATOS PANETONE DELUXE KG NOVO (10)	19012000	MISTUR. PASTA PREP.PROD.PADARIA, PAST.BISC	19052010	PANETONE

MODELO CEST

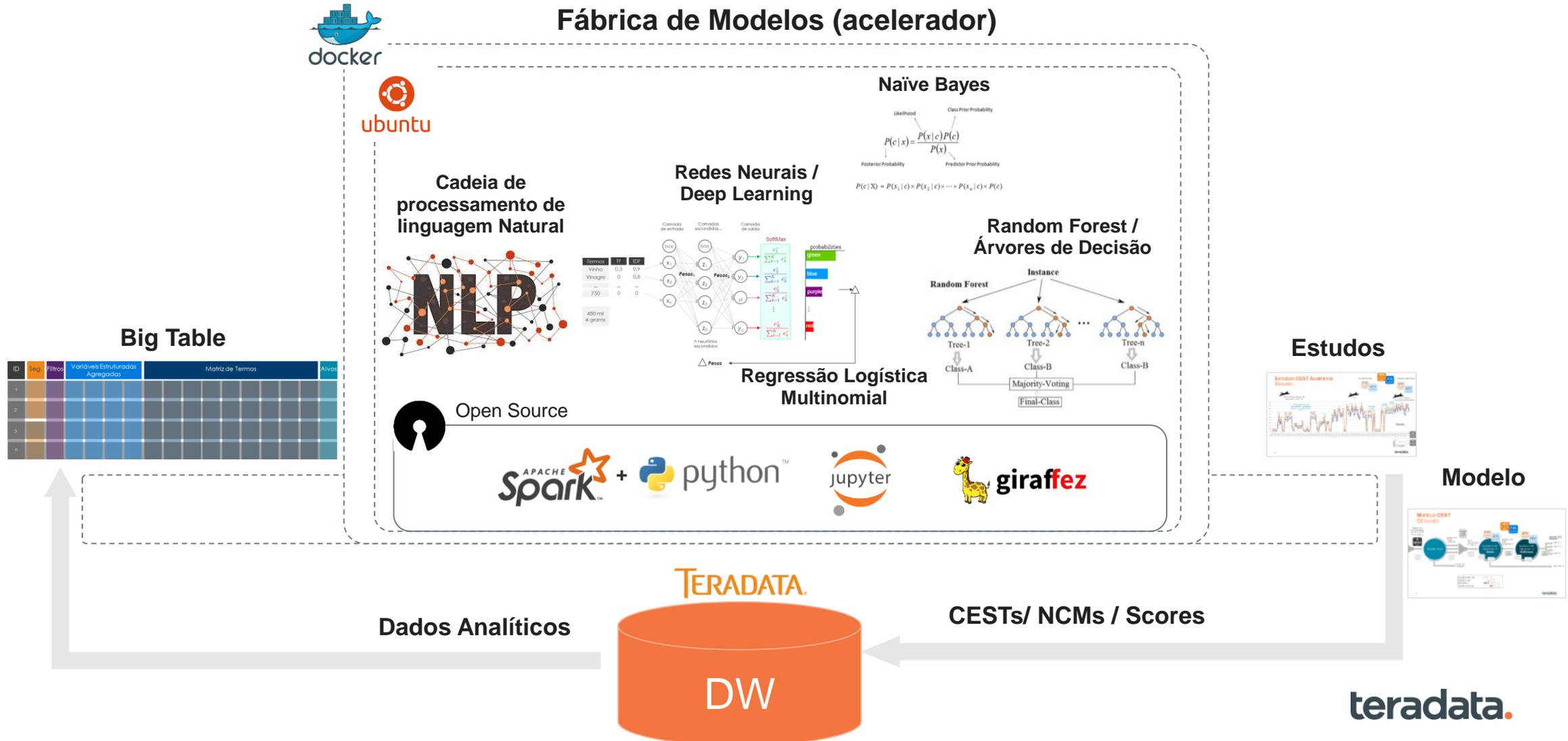
DETALHES



Acelerador da
Fábrica de
Modelos
(open source)



FÁBRICA DE MODELOS - ACELERADOR ARQUITETURA



ESTUDOS CEST ALIMENTOS

RESUMO



Base analítica estruturada
(itens de NF-e/NFC-e)



CST



NCM Predito e
refinamento de modelos

Modelo binário

79%
Acurácia
itens

71%
Acurácia

Modelo Multi-Classe

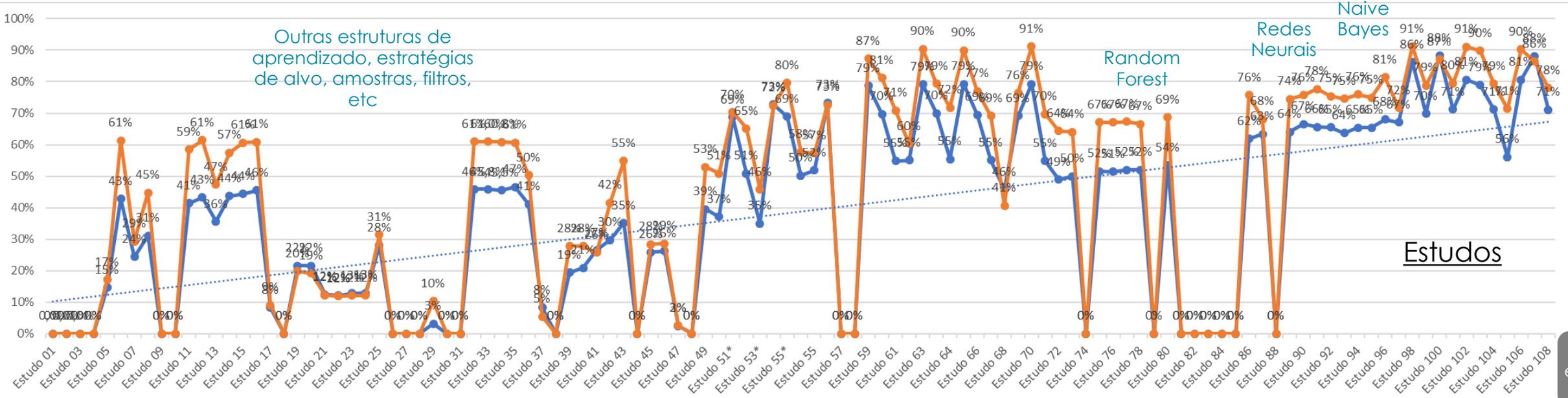
91%
Acurácia
itens

81%
Acurácia

87%
Acurácia
itens

88%
Acurácia

Outras estruturas de
aprendizado, estratégias
de alvo, amostras, filtros,
etc



Estudos

108
estudos
CEST

26
estudos
NCM

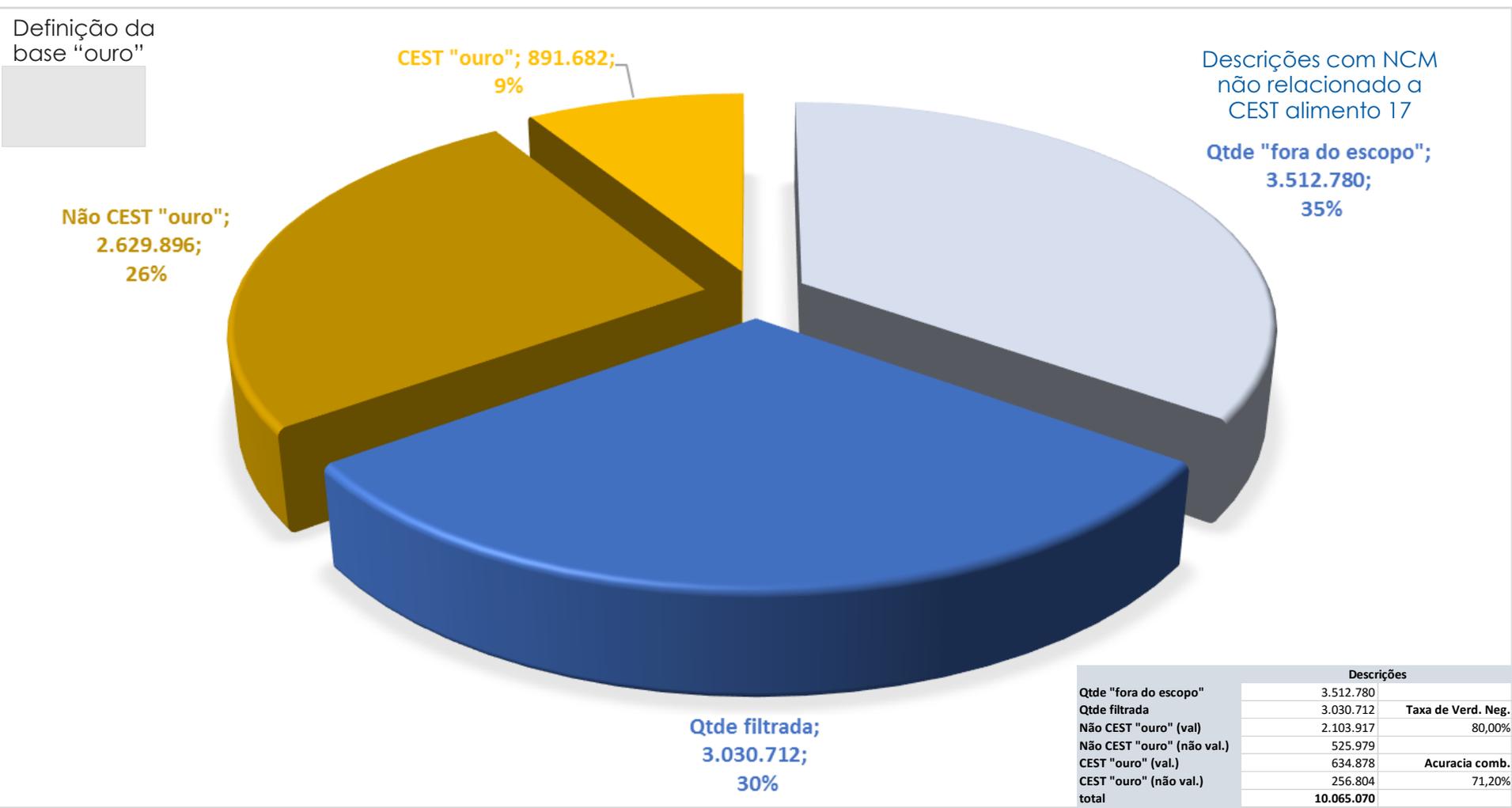


BASE DE APRENDIZADO (CEST 17)

VOLUMETRIA DE DESCRIÇÕES ÚNICAS

10 milhões
descrições
únicas

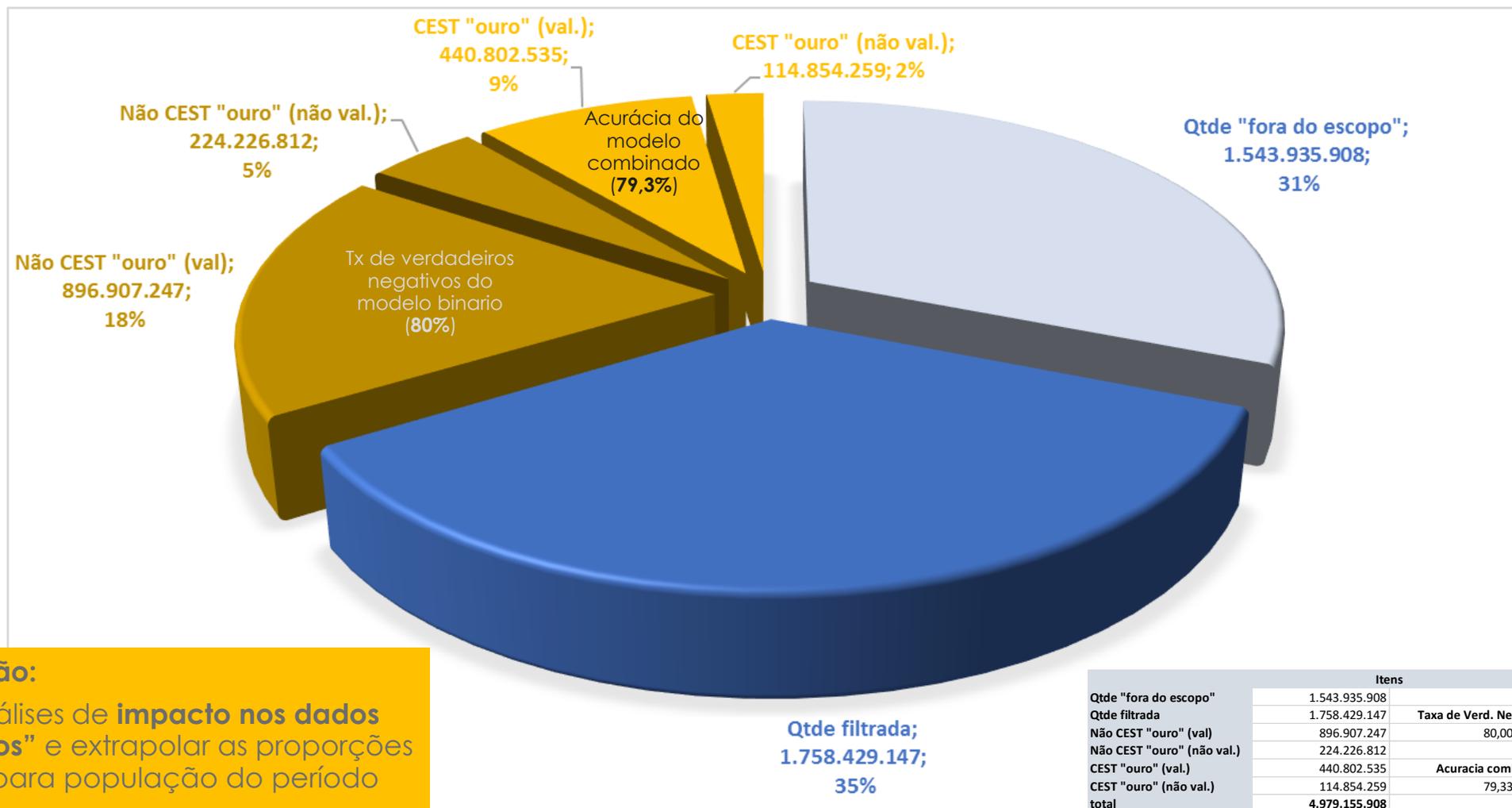
Base de
aprendizado
1 ano
(2017/2018)



RESULTADO DO MODELO (CEST 17)

VOLUMETRIA DE ITENS

5
bilhões
itens

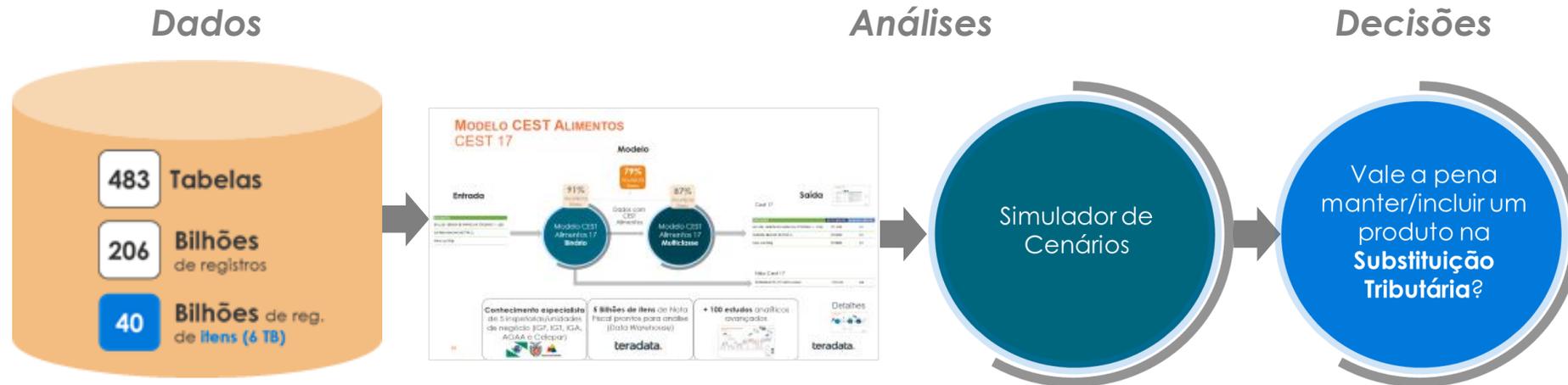


Recomendação:

Realizar as análises de **impacto nos dados "ouro validados"** e extrapolar as proporções encontradas para população do período analisado

PRÓXIMOS PASSOS

- Análise das **estimativas de impacto**



- “**Produtização**”, ampliação do modelo para **outros segmentos** e **melhoramento contínuo**
- Análise de **evasão** (falta de classificação ou erro no preenchimento)
- Análise de **fraude** (classificação inconsistente com o produto vendido)
- Outras aplicações

Obrigado.

José Márcio Faier
Cientista de Dados
jose.faiier@teradata.com

Gilsomar Rezende
Engenheiro de Dados
gilsomar.resende@teradata.com

Diego Fernandes
Arquiteto de Soluções
diego.fernandes@teradata.com

Nathalia Pietro
Gerente de Projetos
nathalia.pietro@teradata.com

Vinícius Fonseca
Diretor de Serviços
vinicius.fonseca@teradata.com

Carlos Bokor
Diretor Comercial
carlos.bokor@teradata.com

teradata.

©2018 Teradata