TensorFlow
User Group São Paulo

TensorFlow

TFX1|E1

Episódio 1: Machine Learning produtivo e componentes TFX de ingestão de dados

Alex Mansano

Pedro Gengo

Vinicius Caridá

Made by ML GDEs

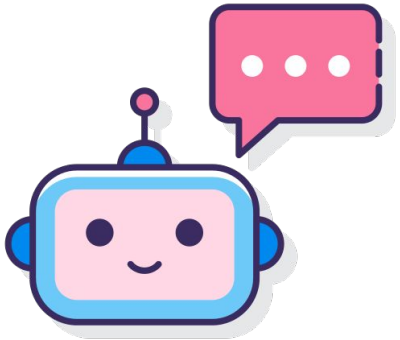
Agenda

- Machine learning em produção
- MLOps - Definição
- TFX
- Componentes TFX - Visão geral
- Componentes de ingestão e validação de dados:
    - ExampleGen
    - StatisticsGen
    - SchemaGen

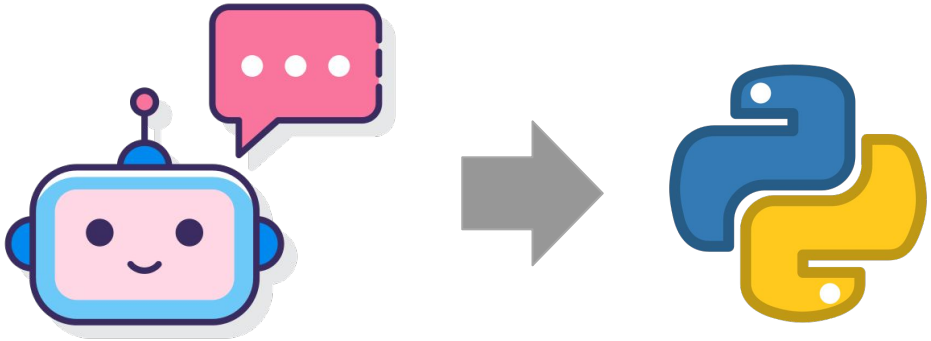
Machine Learning em produção

TensorFlow

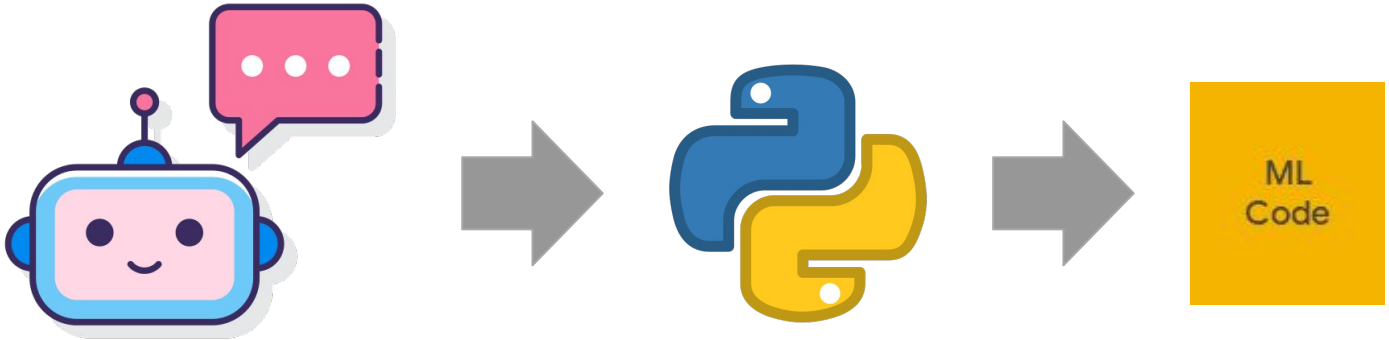
Machine learning é fácil, não é?



Machine learning é fácil, não é?



Machine learning é fácil, não é?

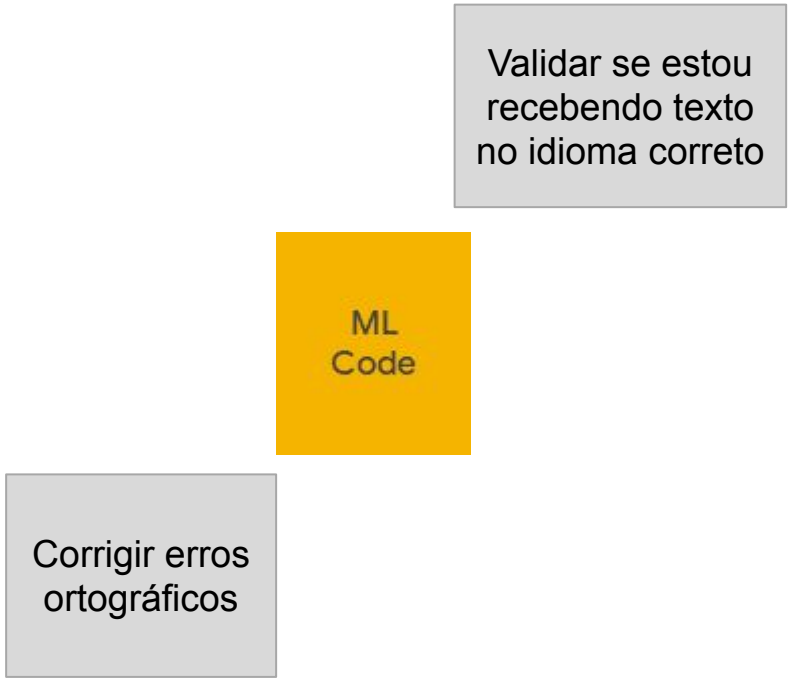


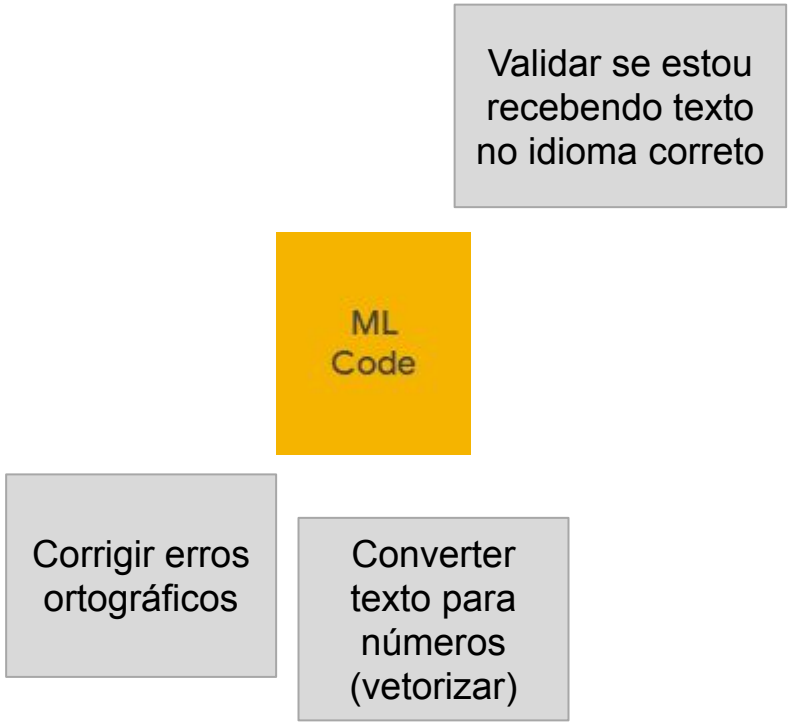
É só isso que precisamos para nossa solução?

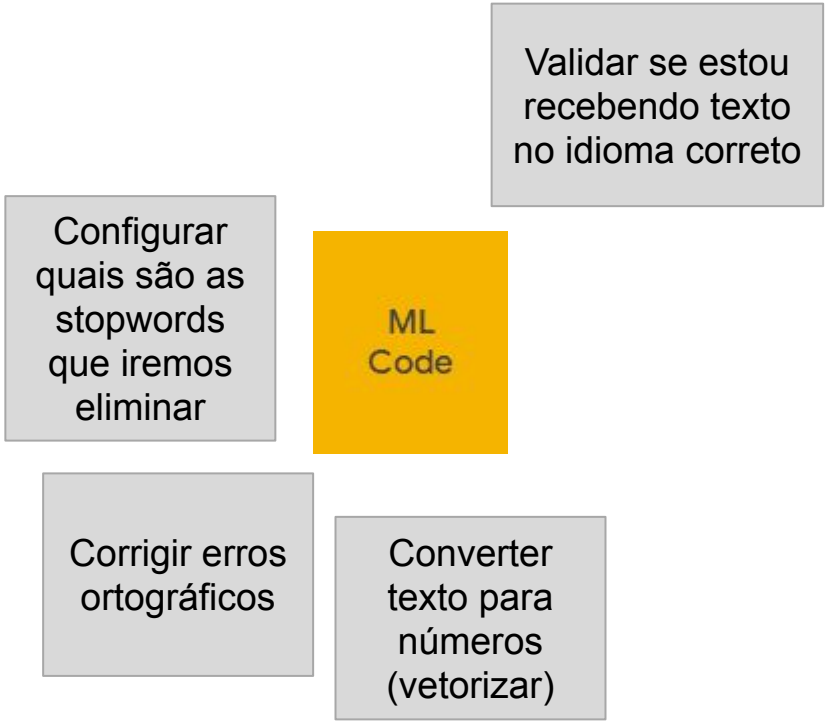

ML
Code

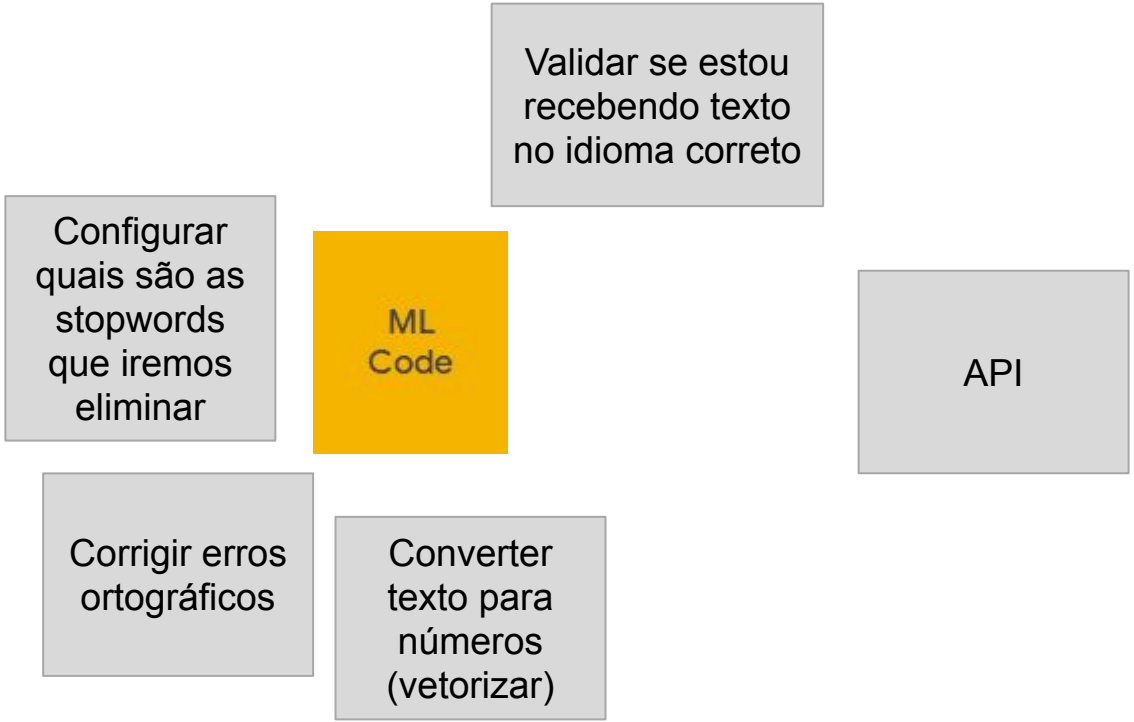

ML
Code

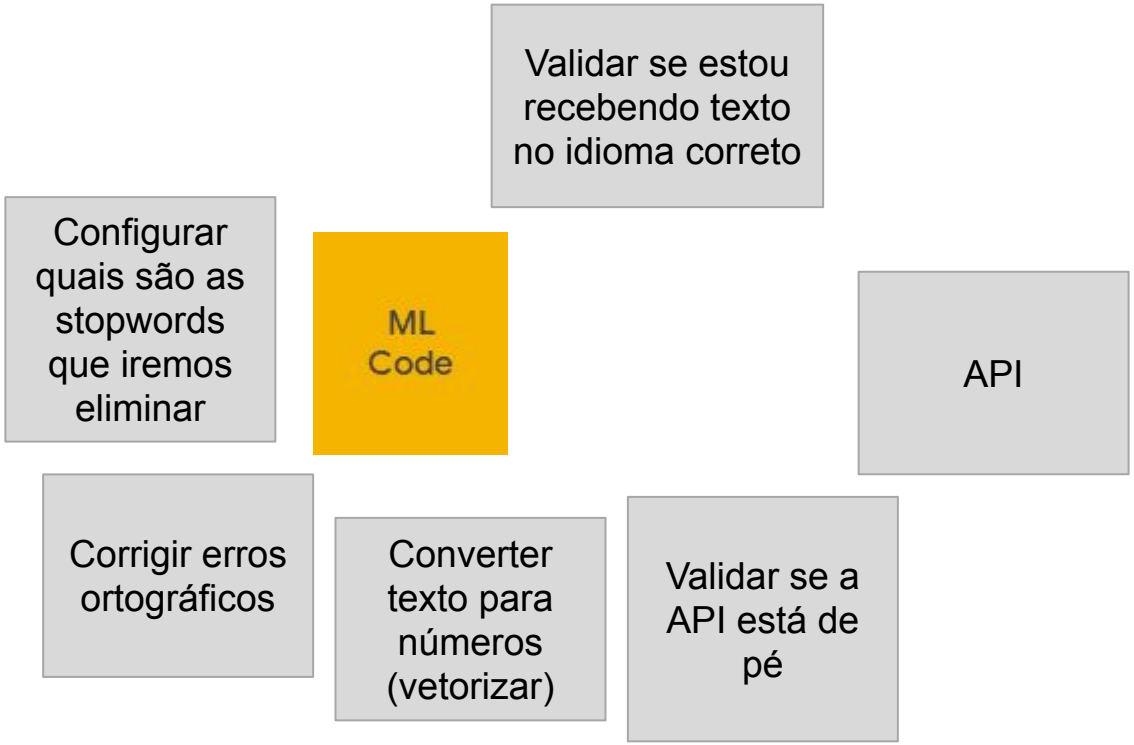
Corrigir erros ortográficos

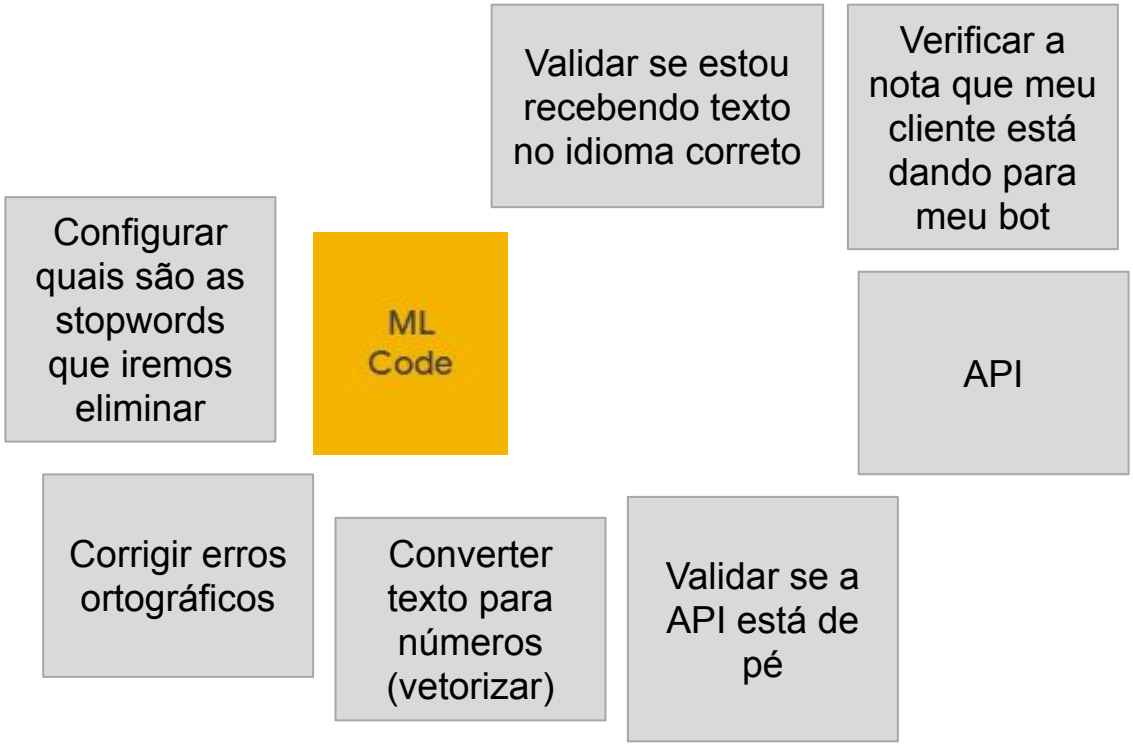

Validar se estou recebendo texto no idioma correto
ML Code
Corrigir erros ortográficos

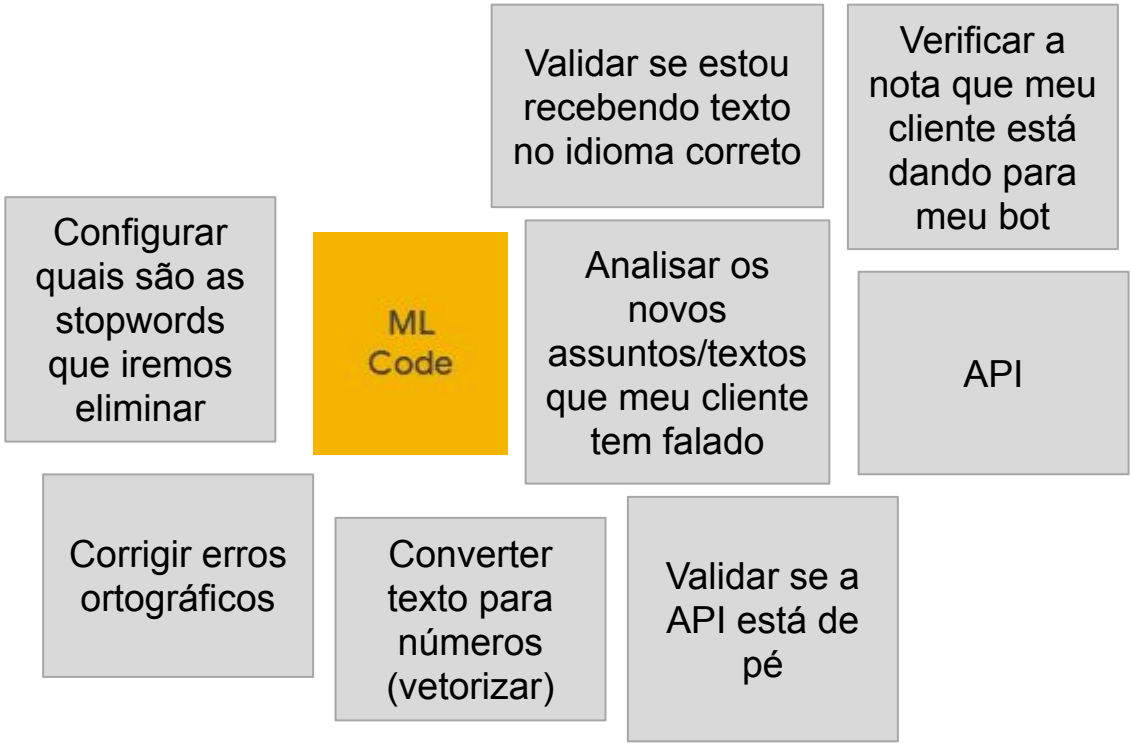

Validar se estou recebendo texto no idioma correto
ML Code
Corrigir erros ortográficos
Converter texto para números (vetorizar)

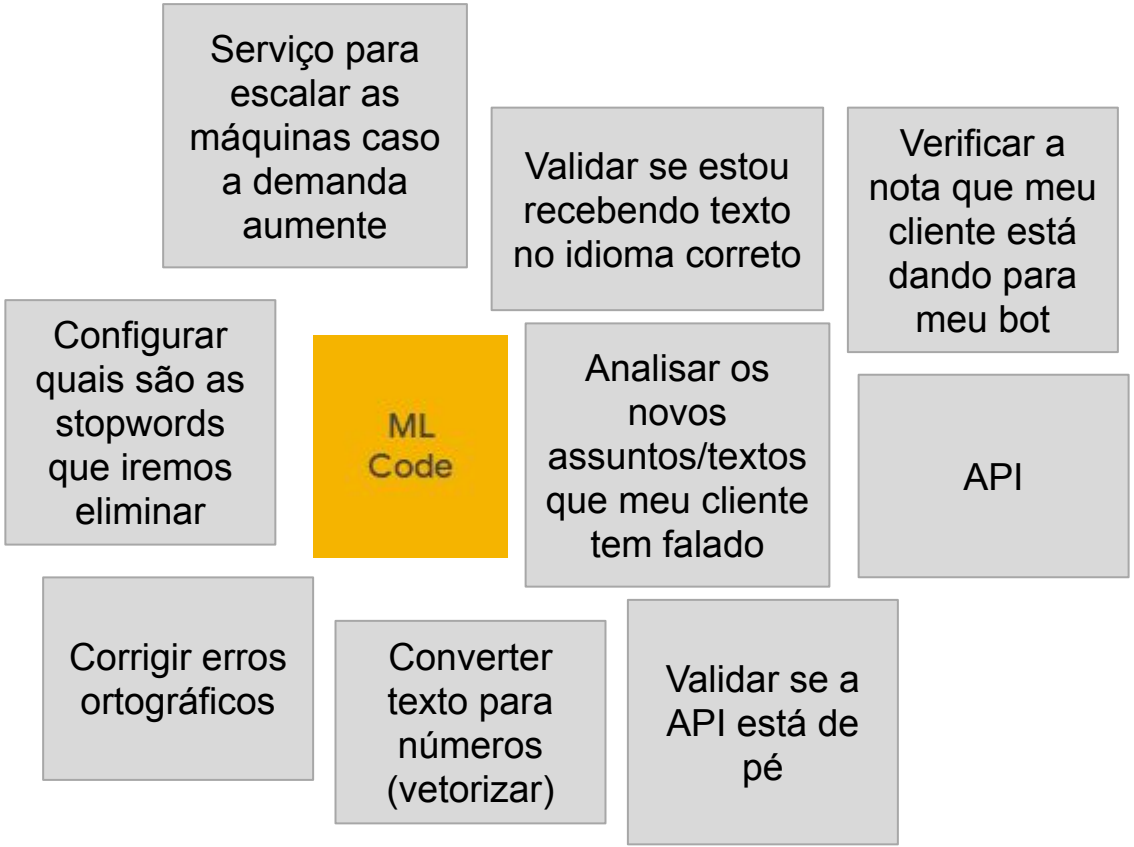

Validar se estou recebendo texto no idioma correto
Configurar quais são as stopwords que iremos eliminar
ML Code
Corrigir erros ortográficos
Converter texto para números (vetorizar)


Validar se estou recebendo texto no idioma correto
Configurar quais são as stopwords que iremos eliminar
ML Code
API
Corrigir erros ortográficos
Converter texto para números (vetorizar)


Validar se estou recebendo texto no idioma correto
Configurar quais são as stopwords que iremos eliminar
ML Code
API
Corrigir erros ortográficos
Converter texto para números (vetorizar)
Validar se a API está de pé


Validar se estou recebendo texto no idioma correto
Verificar a nota que meu cliente está dando para meu bot
Configurar quais são as stopwords que iremos eliminar
ML Code
API
Corrigir erros ortográficos
Converter texto para números (vetorizar)
Validar se a API está de pé


Validar se estou recebendo texto no idioma correto
Verificar a nota que meu cliente está dando para meu bot
Configurar quais são as stopwords que iremos eliminar
ML Code
Analisar os novos assuntos/textos que meu cliente tem falado
API
Corrigir erros ortográficos
Converter texto para números (vetorizar)
Validar se a API está de pé

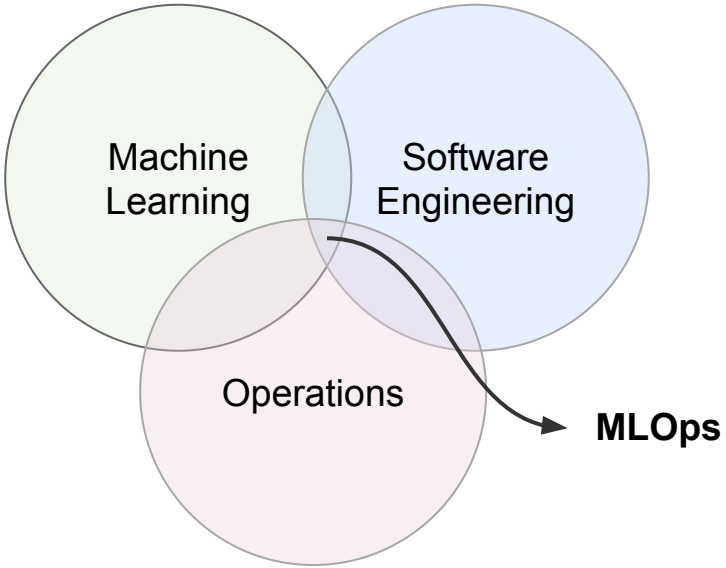

Serviço para escalar as máquinas caso a demanda aumente
Validar se estou recebendo texto no idioma correto
Verificar a nota que meu cliente está dando para meu bot
Configurar quais são as stopwords que iremos eliminar
ML Code
Analisar os novos assuntos/textos que meu cliente tem falado
API
Corrigir erros ortográficos
Converter texto para números (vetorizar)
Validar se a API está de pé

Hidden Technical Debt in Machine Learning Systems, NeurIPS 2015

MLOps

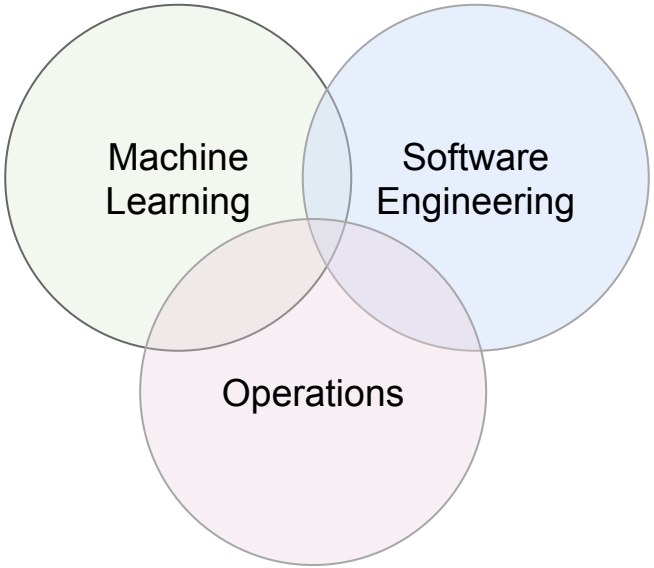
TensorFlow

MLOps



MLOps

- Desenvolvimento de modelos
- Avaliação de modelos
- Tuning de hiperparâmetros



- Desenvolvimento de pipelines
- Boas práticas de código

- CI/CD (deploy)
- Logging e monitoramento
- Gerenciamento de artefatos (metadata)

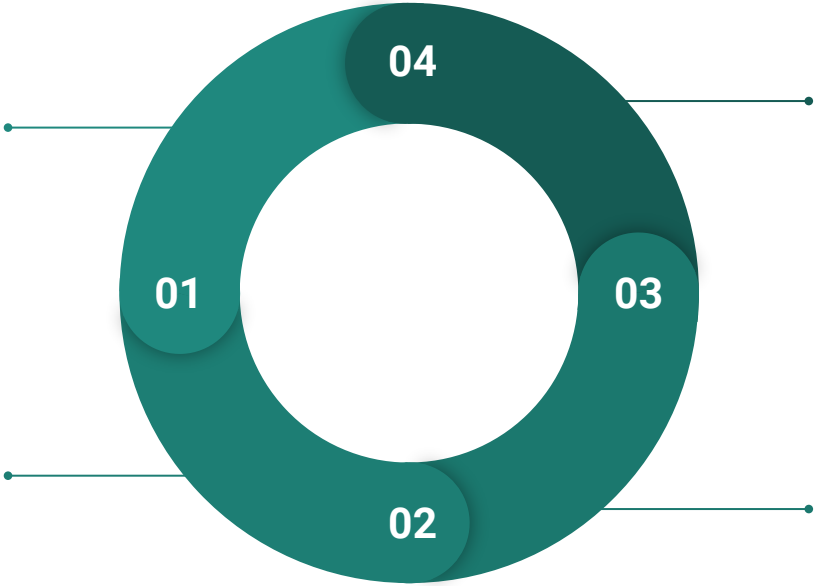
Ciclo de vida de produtos de ML

Experimentação/Desenvolvimento

Testar diferentes modelos, diferentes processamentos de dados, diferentes funções de perda, etc

Treinamento de um novo modelo

Após experimentar, escolher um modelo e treinar para colocar em produção



Monitoramento

Monitoramento constante para verificar mudanças na distribuição dos dados de entrada, piora na qualidade das predições, etc

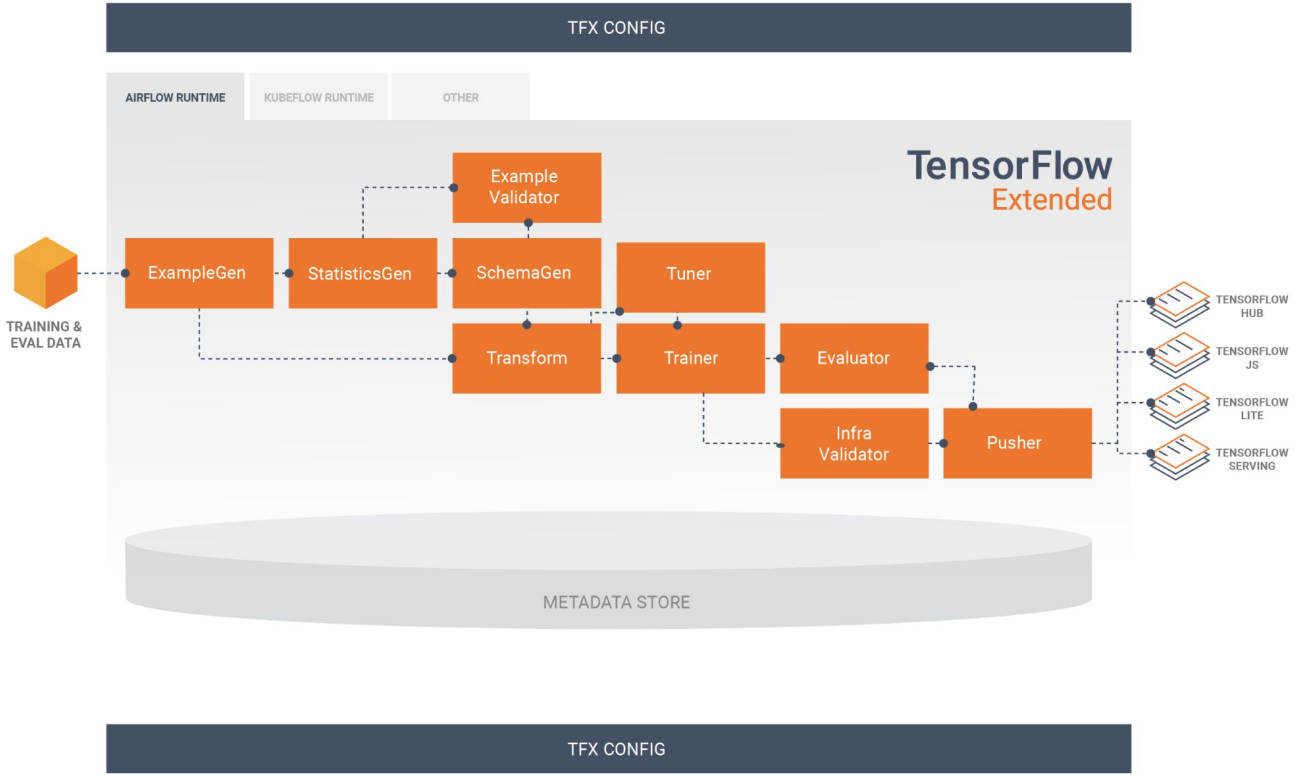
CI/CD

Processo automatizado para validar o novo modelo e colocá-lo em produção

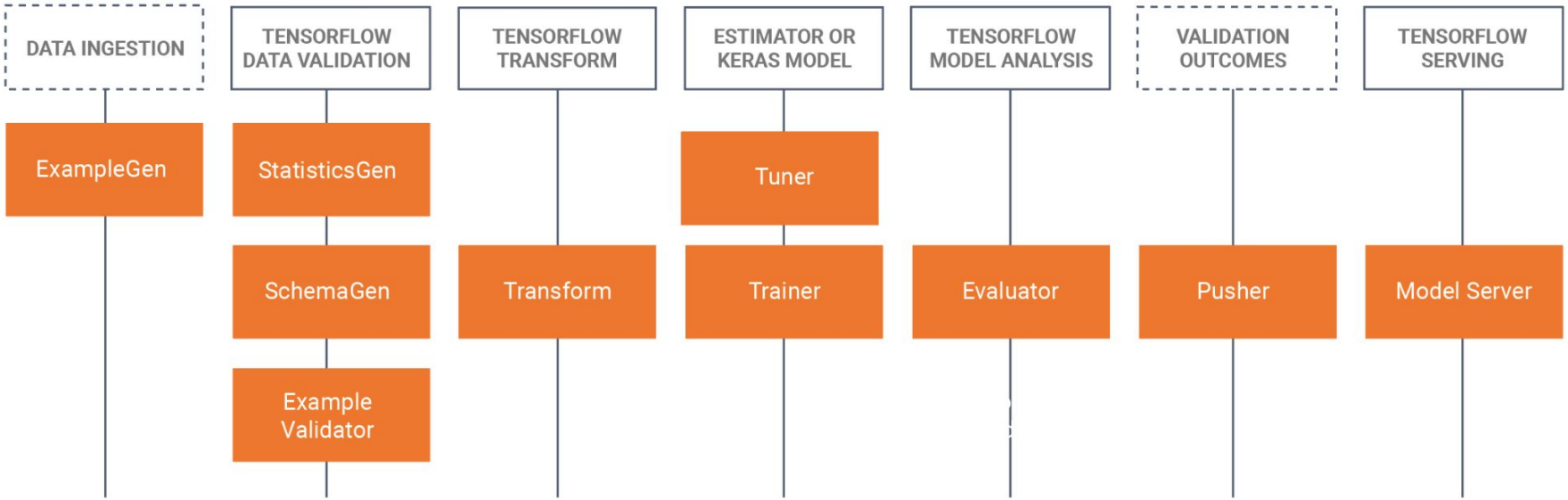
TFX

TensorFlow

TFX


TFX CONFIG
AIRFLOW RUNTIME
KUBEFLOW RUNTIME
OTHER
TensorFlow
Extended
Example Validator
ExampleGen
StatisticsGen
SchemaGen
Tuner
Transform
Trainer
Evaluator
Infra Validator
Pusher
TRAINING & EVAL DATA
TENSORFLOW HUB
TENSORFLOW JS
TENSORFLOW LITE
TENSORFLOW SERVING
METADATA STORE
TFX CONFIG

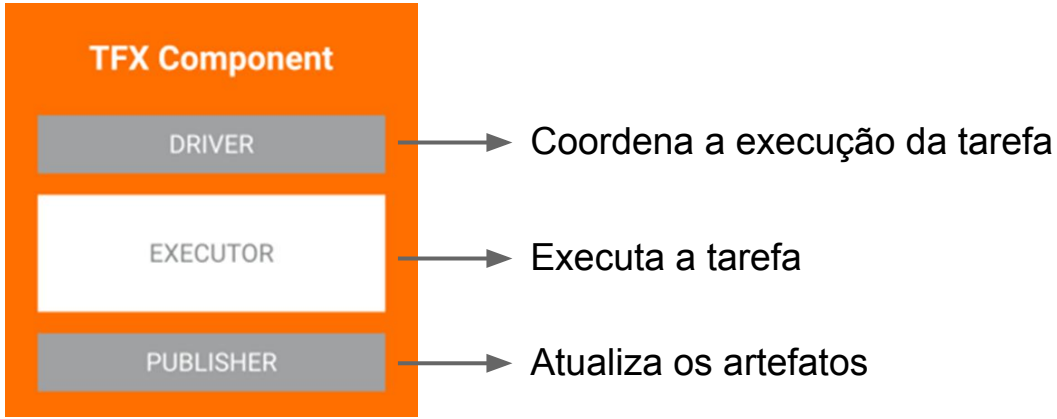
Bibliotecas do TFX



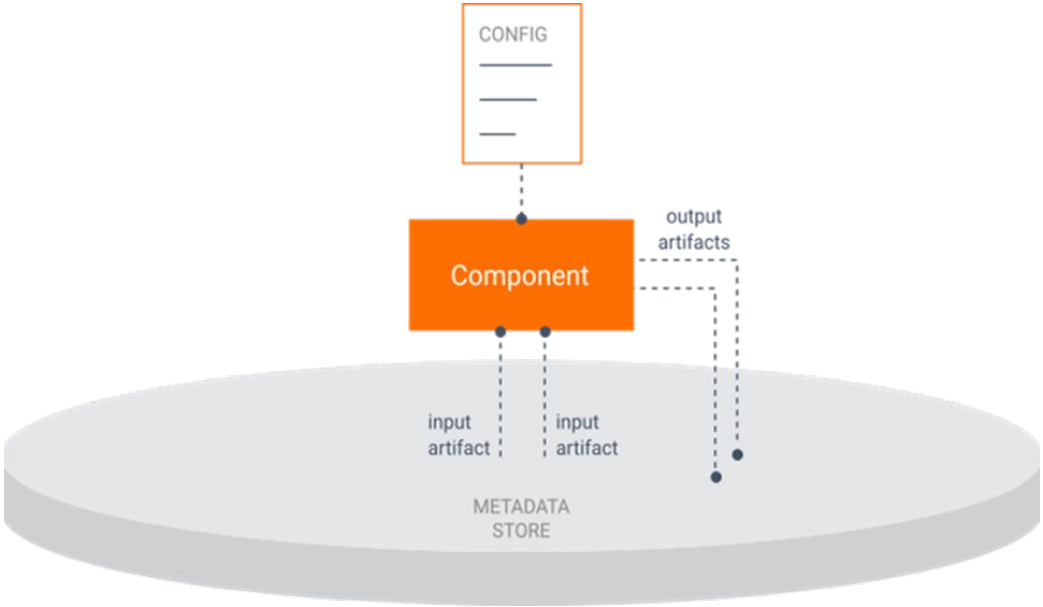
Componentes TFX

TensorFlow

Visão individual de um componente



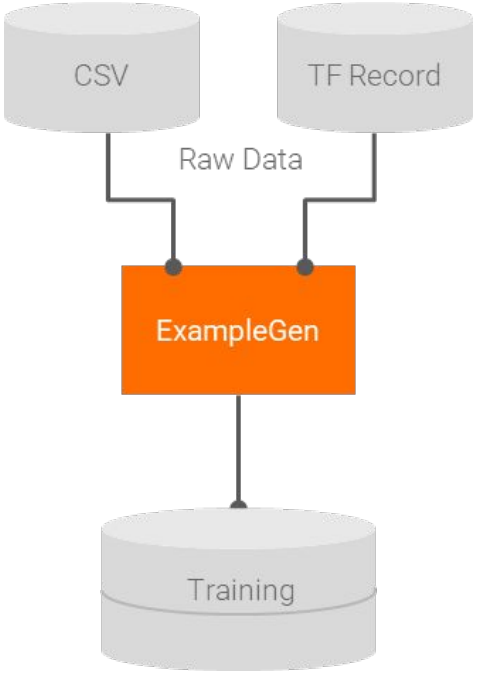
Definição de um componente na pipeline



Ingestão e Validação de Dados no TFX

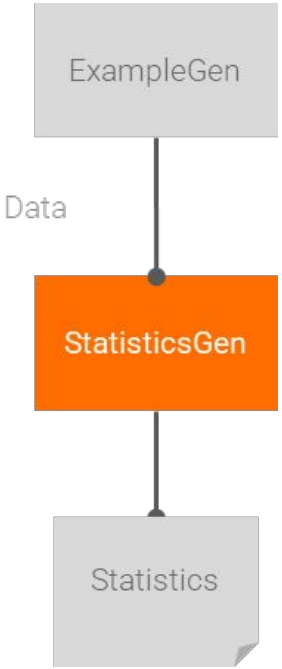
TensorFlow

ExampleGen



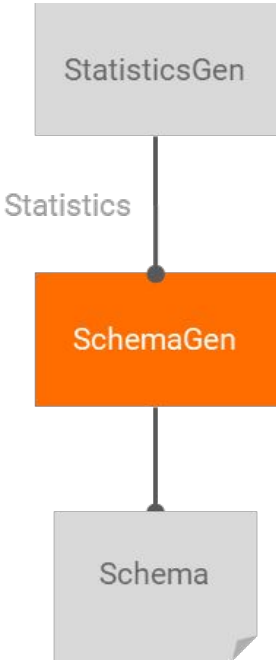
- Componente de ingestão de dados
- Ingere dados externos para gerar Examples que serão utilizados por outros componentes da pipeline
- Aceita dados em diferentes formatos como: CSV, BigQuery, tf.Record, Parquet, etc

StatisticsGen



- Usa o output do ExampleGen como input
- Gera estatísticas dos dados, que serão utilizados por outros componentes
- Usa a biblioteca tfdv (TensorFlow Data Validation)

SchemaGen



- Usa o output do StatisticsGen como input
- Infere os tipos, intervalos de valores, valores permitidos, etc
- É recomendado revisar o schema gerado (usar componente ImportSchemaGen)