

INT260

# Data Classification with Python SDK and SAP AI Business Services.

Michael Haas, Senior Software Developer  
Samuel Seah, Product Owner

PUBLIC



# Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP's intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

# Agenda

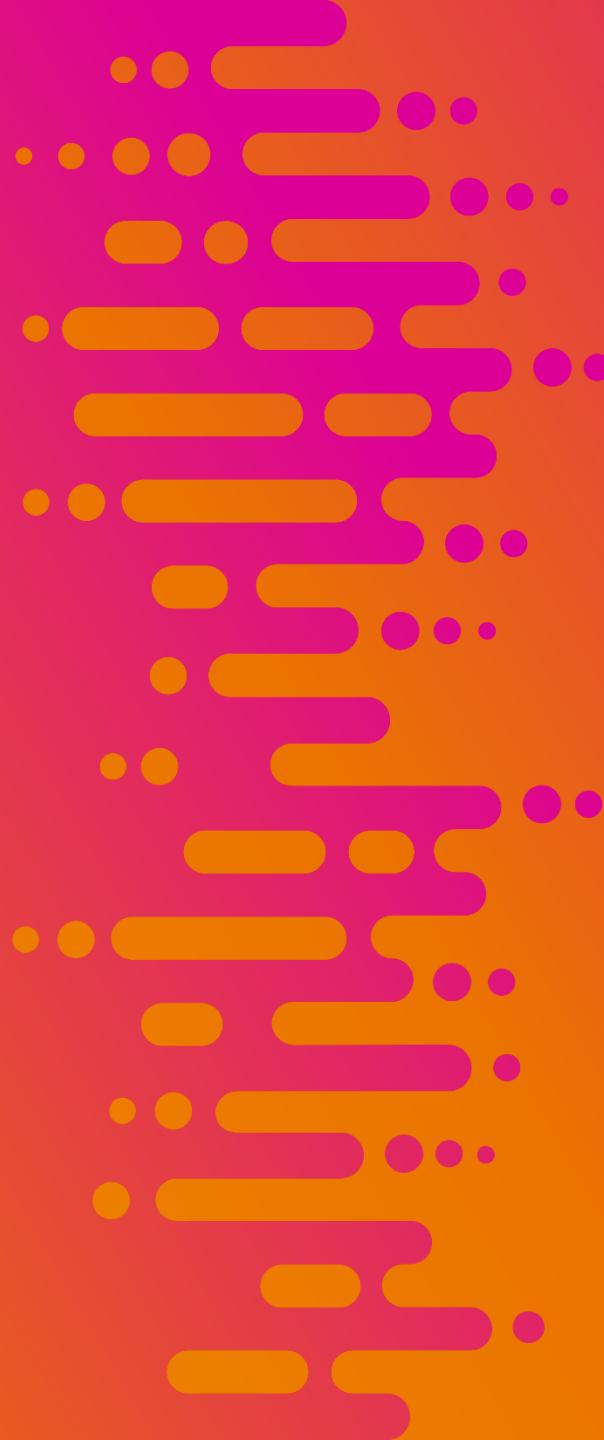
## Data classification with Data Attribute Recommendation

- Introduction
- Use cases
- Workshop hands-on

## Data classification with Service Ticket Intelligence

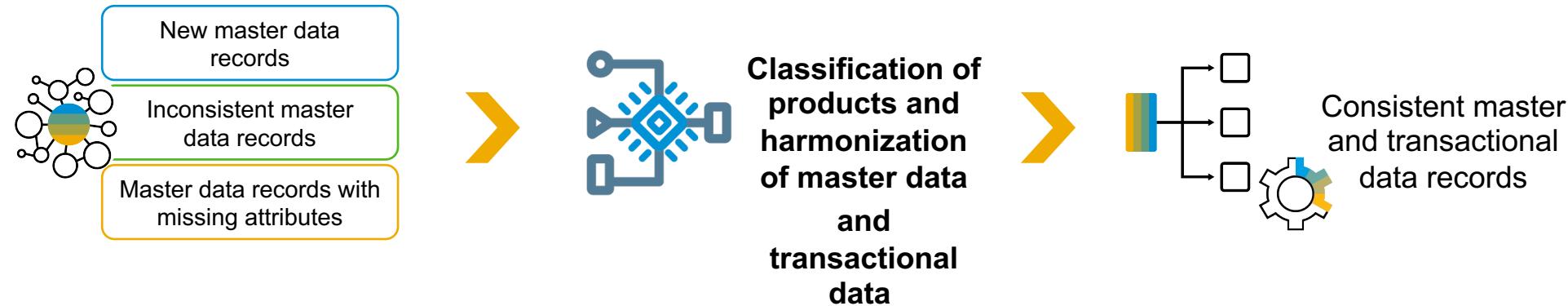
- Introduction
- Use cases
- Workshop hands-on

# Data classification with Data Attribute Recommendation



# Data Attribute Recommendation

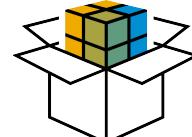
Automate master data / transactional data management tasks



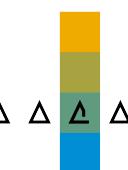
**Data Attribute Recommendation helps to classify entities such as products, stores and users into multiple classes, using free text, numbers and categories as input.**



Automate and speed up master data creation and maintenance

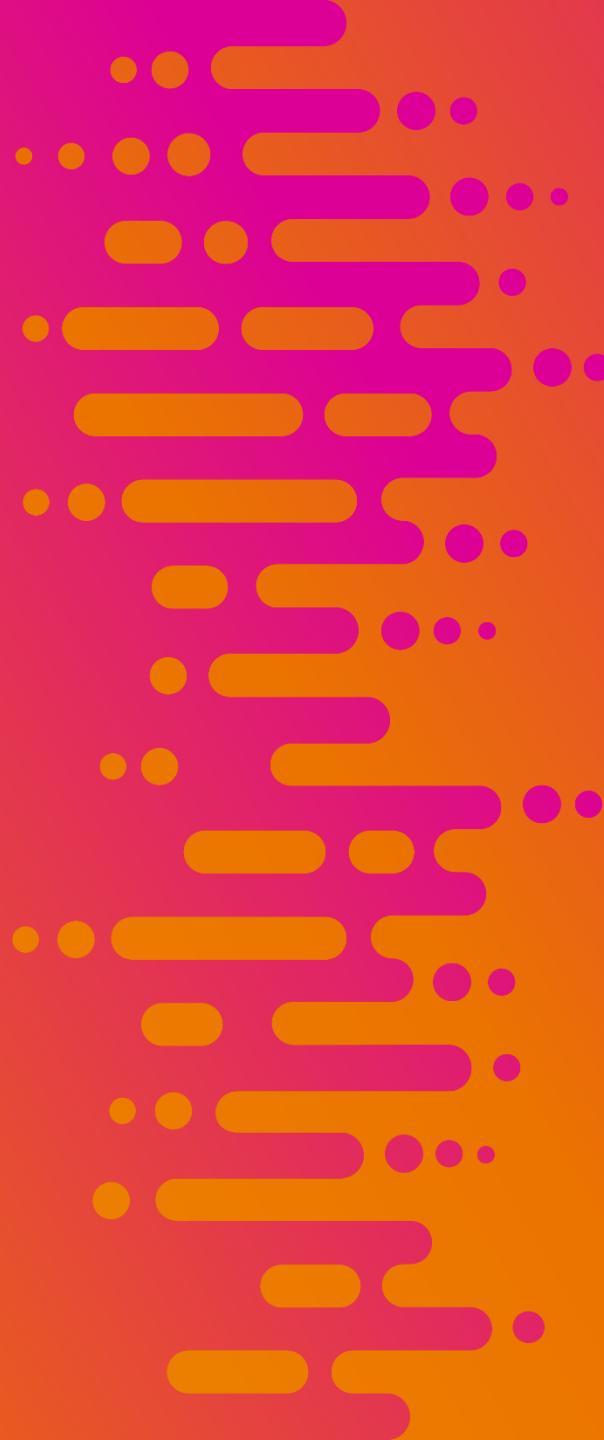


Gain easier and faster master data insights

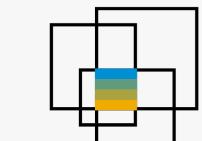


Reduce errors and manual efforts when maintaining master data

# Use cases for Data Attribute Recommendation



# Scenario 1 – Match point of sales information to your own product hierarchy



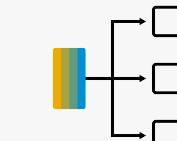
Point of Sales  
information



Your own product  
hierarchy



Data Attribute  
Recommendation



Consistent  
Master Data

## BUSINESS PROBLEM

- Struggle with the **organization and management of large volumes of documents**
- Large amounts of unorganized, yet business critical documents (like contracts, products) are **preventing the workforce to work efficiently**



## SOLUTION

### Data Attribute Recommendation:

- Can be used for a **prediction performance**: bulk and individual classifications of materials and its characteristics are done online
- **Enables intelligent master data governance** at companies struggling with the creation of new material requests



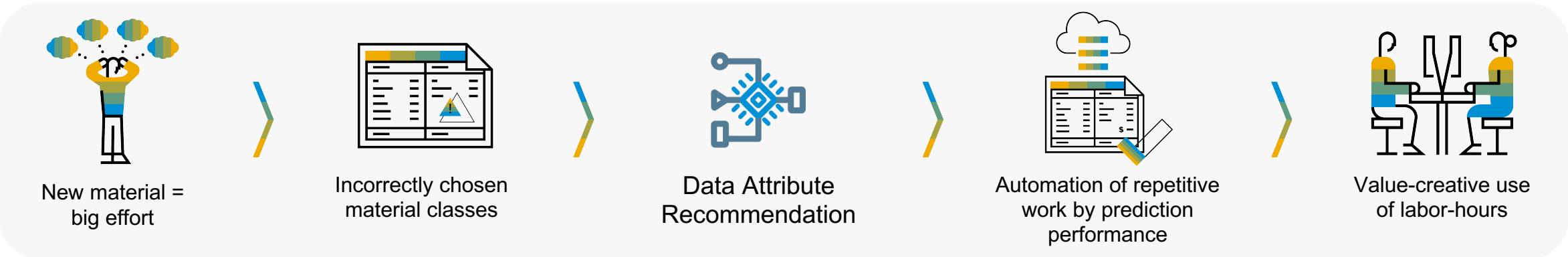
## BENEFITS

- **The only way** to master the chaos is by using automation
- **Reducing** the manual work
- **Speeding up** and **improving the quality**
- **Easily Customizable**
- **Improve quality of data**

### SAP SOLUTIONS IN THIS USE CASE:

- *Data Attribute Recommendation via SAP Cloud Platform Enterprise Agreement*
- *SAP Demand Signal Management*
- *SAP Cloud Platform Enterprise Agreement*

## Scenario 2 – Get suggestions of material class and its characteristics



### BUSINESS PROBLEM

#### Creating new material requests:

- Big, time consuming effort
- **Many requests coming back** due to incorrectly chosen material class

### SOLUTION

#### Data Attribute Recommendation:

- Identifies relevant matching patterns
- Following these patterns it predicts to which representation in the product hierarchy the incoming product record belongs which helps to gain direct insights in the incoming point of sales information

### BENEFITS

- Enables organizations to analyze valuable information about their competitors in comparison to themselves
- **Acceleration** of the process
- **Reduction of more than 20% of the effort** for new material creation
- **Automation of the repetitive work** and use labor hours on **high value tasks**

SAP SOLUTIONS IN THIS  
USE CASE:

• Data Attribute Recommendation via SAP Cloud Platform Enterprise Agreement

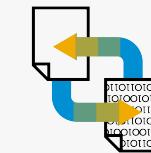
# Scenario 3 – Commodity Code Prediction



New Commodity,  
no Commodity Code



Data Attribute  
Recommendation



Commodity  
Code



New Commodity is  
Imported / Exported

## BUSINESS PROBLEM

- International trade demands commodity codes
- No correct commodity code
  - Rejection of assignment goods by customs
  - Waiting on borders = high costs for the organization
- Maintaining such master data
  - is supported by rules
  - leads to inconsistencies & left-overs for manual processing

## SOLUTION

**Data Attribute Recommendation** can predict:

- Correct values by learning the logic behind complex fields (such as the commodity code)
- Dependencies to multiple data attributes

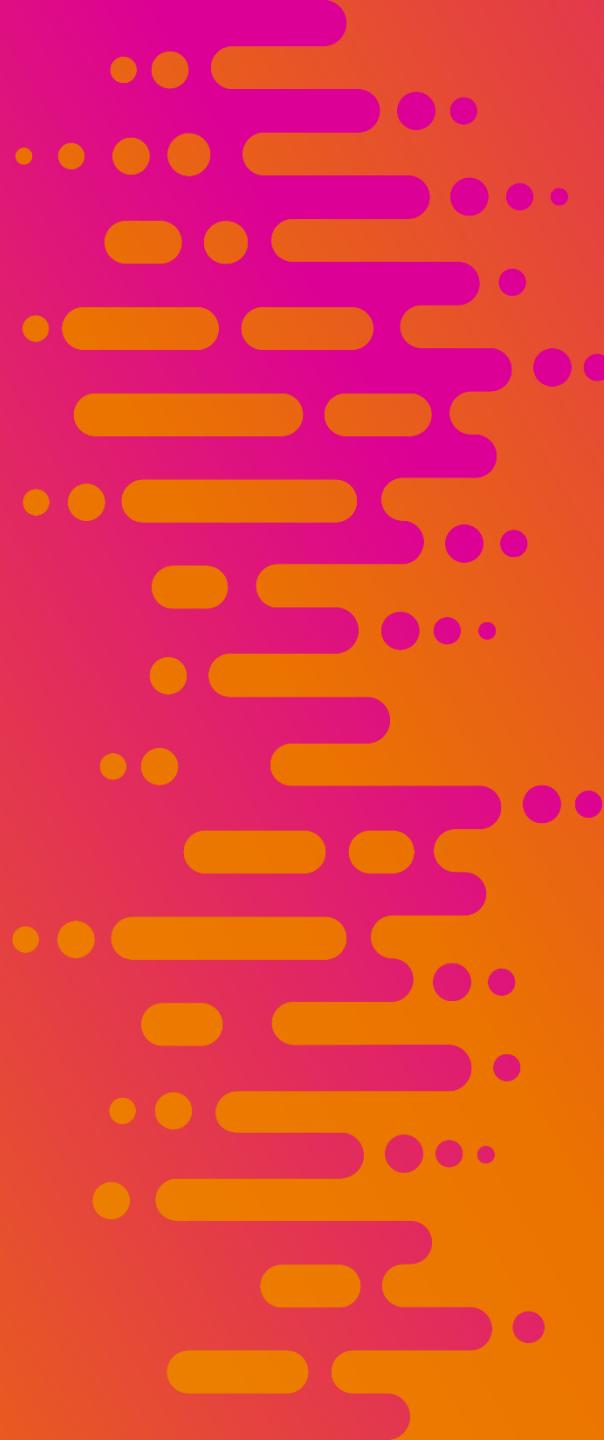
## BENEFITS

- Reduces manual efforts in commodity code assignment
- Increases the validity of assigned commodity codes
- Decreased costs:  
caused by labor & mal-assignment of commodity codes

SAP SOLUTIONS IN THIS  
USE CASE:

Data Attribute Recommendation via SAP Cloud Platform Enterprise Agreement

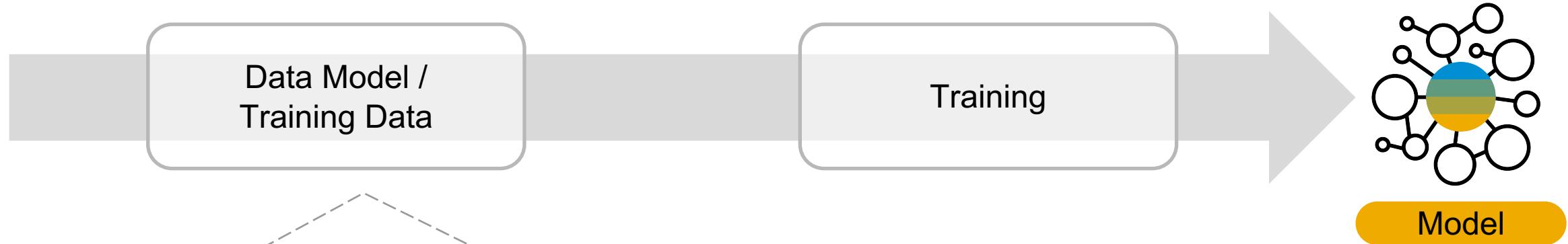
# Hands-on for Data Attribute Recommendation



# Data Attribute Recommendation **Hands-On**

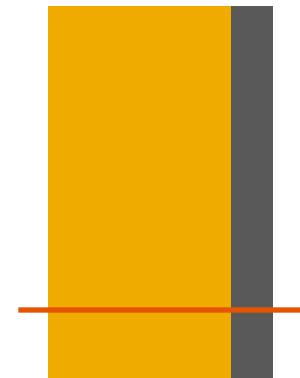
- Follow the exercises in the [github.com/SAP-samples/teched2020-INT260](https://github.com/SAP-samples/teched2020-INT260) repository

# Training-Process



## Data Requirements:

- The data set should be **>3,000 documents**  
At best: use **all data that is available** on the topic
- The **higher the quality** of the data, the better the results  
**Contradictory** data worsens the result
- At least **one free text** should be used for training (the more, the better)



Training data consists of **text input** and **labels**

Model is initialized based on the categories found in the **labels**.

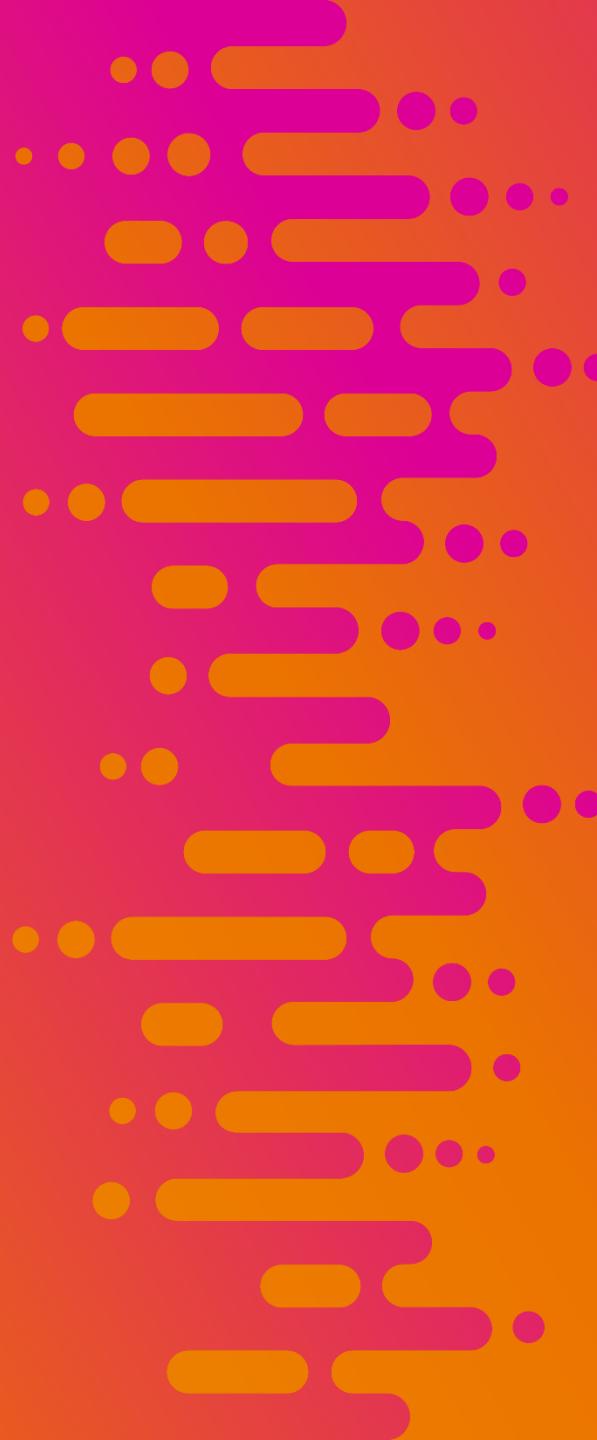
Data is **split** into training testing and validation sets in 8:1:1 ratio.

# Data Attribute Recommendation **Trial Limitations**

- The free Trial Account is subject to [usage limits](#)
- Any model deployed will be undeployed after being active for 8 hours.

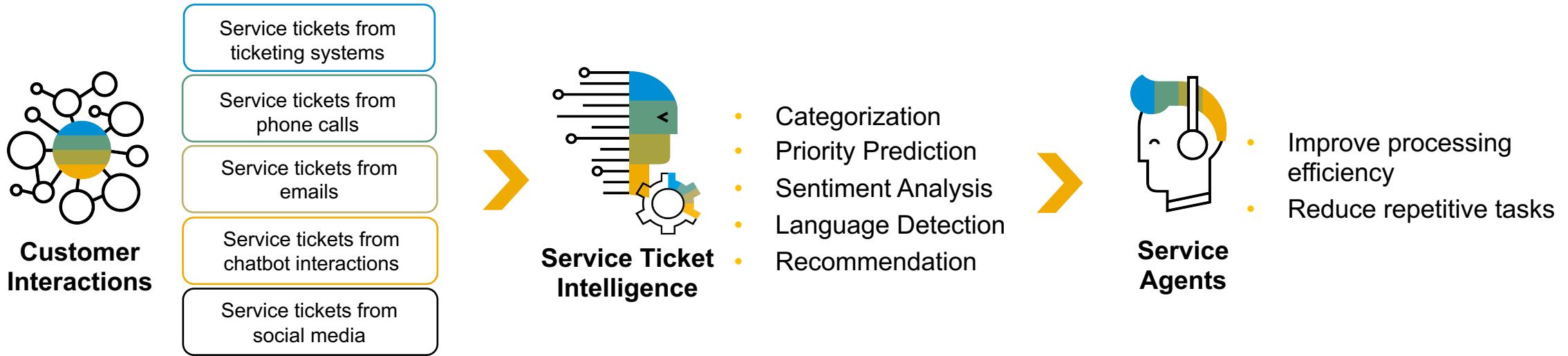
<b>Input</b>	<b>Maximum Limit</b>
Number of dataset schemas	50
Number of datasets	20
Size of dataset	10 MB
Number of pending jobs	1
Number of trained models	10
Number of simultaneously deployed models	1
Number of record predictions	2000

# Data classification with Service Ticket Intelligence

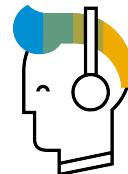


# Service Ticket Intelligence

Reimagine Customer Service with Automated Processes



**Building a scalable, self-driven customer service and yet delivering great customer experience**



Speed up service response times



Higher rates of problem resolution and ticket closure

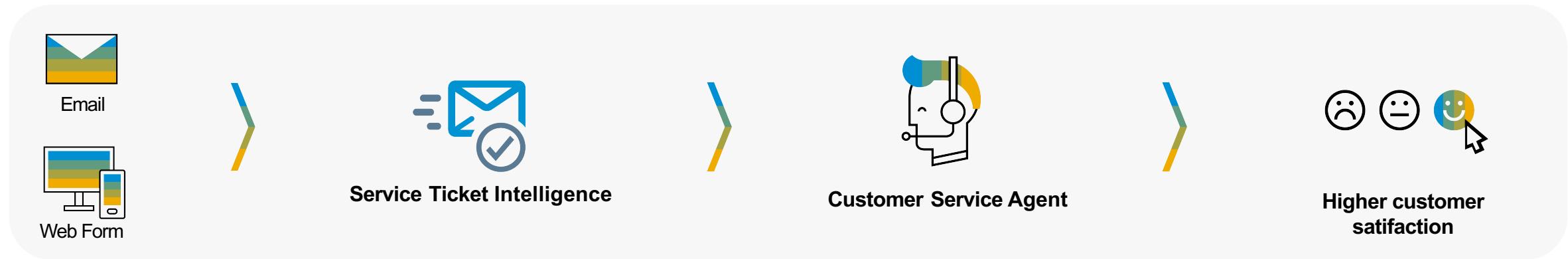


Increase customer satisfaction

# Use cases for Service Ticket Intelligence



# Scenario 1 – Integration with an Omnichannel front office



## BUSINESS PROBLEM

- Increased volume and speed of customer interactions
- Higher need in management of interactions by the agents
- Higher expectations in terms of response times
- Seamless handover across interaction channels

## SOLUTION

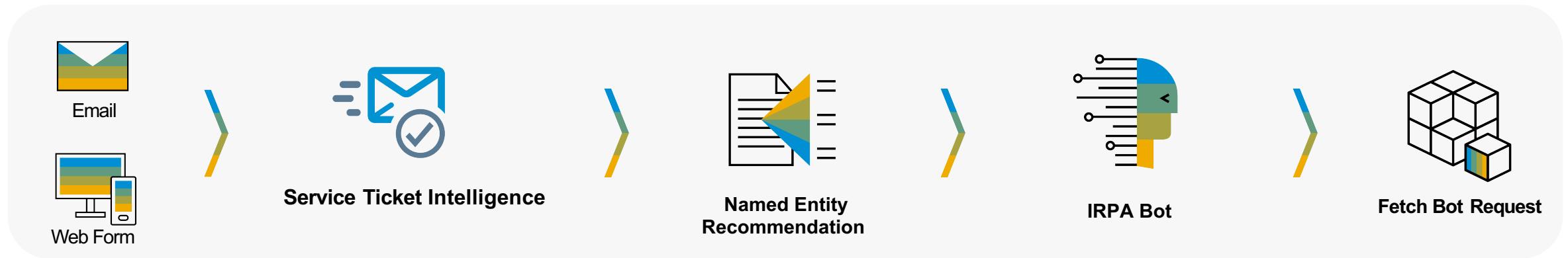
### Service Ticket Intelligence:

- Automatization of **classification** and **processing** of customer service requests
- Classification of customer inputs **from various channels** into internal service groups

## BENEFITS

- Minimization of repetitive tasks
- Focusing on **value-creative** tasks
- Provision of faster and more accurate responses to the customer

# Scenario 2 – Shared Services at back office



## BUSINESS PROBLEM

### Shared Service Centers

= multi-inbound channel for internal and external standard inquiries about various topics

→ **high workloads with queued worklists**

## SOLUTION

### Service Ticket Intelligence:

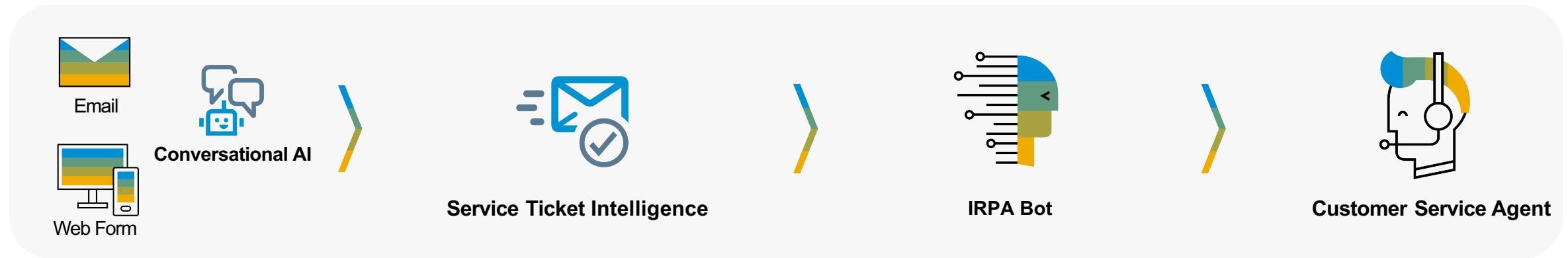
- Automatization of **classification** and **processing** of customer service requests
- Categorization of customer input **from various channels** into internal service groups
- **Redirection** of categorized inquiries to right processor

- **Prediction of category** of the incoming inquiries & redirection to right processor
- **Standard inquiries** can be fetched by **IRPA** and answered automatically
- **Minimization of repetitive tasks**
- Focusing on **value-creative** tasks
- **Provision of faster and more accurate** responses to the customer

SAP SOLUTIONS IN THIS  
USE CASE:

- Service Ticket Intelligence as part of Cloud Platform Enterprise Agreement
- Named Entity Recommendation as part of Cloud Platform Enterprise Agreement
- SAP Service Cloud Enterprise Edition or SAP CRM or SAP S/4HANA Customer Management
- SAP Intelligent Robotic Process Automation
- SAP Shared Service Framework for S/4HANA

# Scenario 3 – In combination with ConversationalAI and iRPA



## BUSINESS PROBLEM

- **Increased volume and speed** of customer interactions
- Higher need in management of interactions by the agents
- **Higher expectations** in terms of response times
- Seamless handover across **interaction channels**

## SOLUTION

### Service Ticket Intelligence:

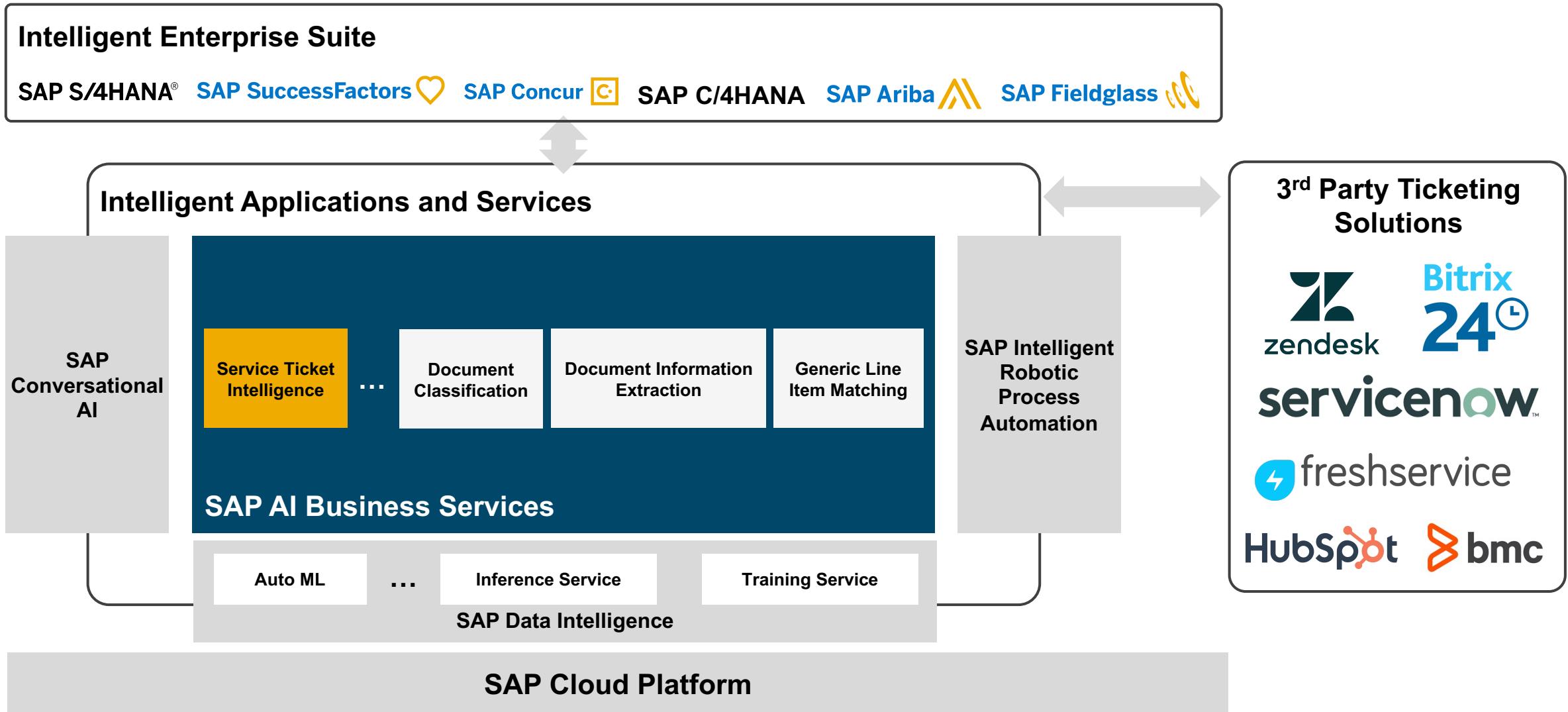
- Automatization of **classification** and **processing** of customer service requests
- Categorization of customer input **from various channels** into internal service groups

**Conversational AI** allows direct answer to the customer without including an agent

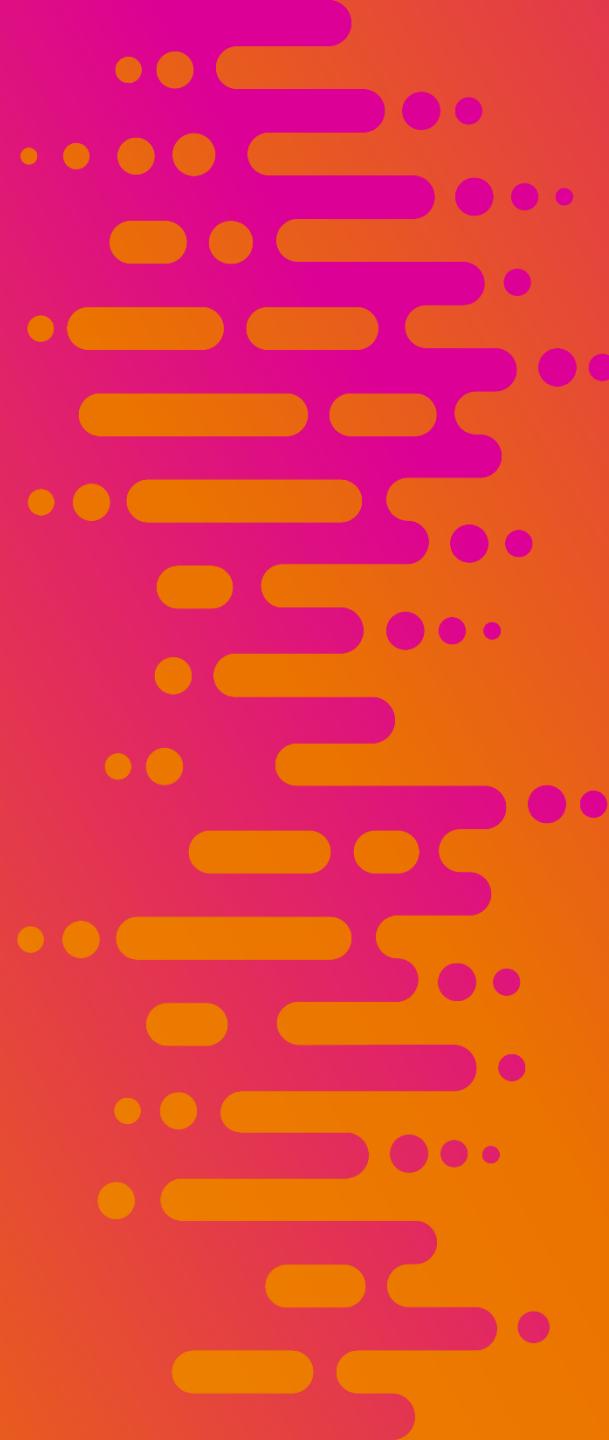
## BENEFITS

- **Prediction of category** of the incoming inquiries & redirection to right processor
- **Standard inquiries** can be fetched by **iRPA** and answered automatically
- **Minimization of repetitive** tasks
- Focusing on **value-creative** tasks
- **Provision of faster and more accurate** responses to the customer

# Service Ticket Intelligence can be integrated with various ticket handling applications

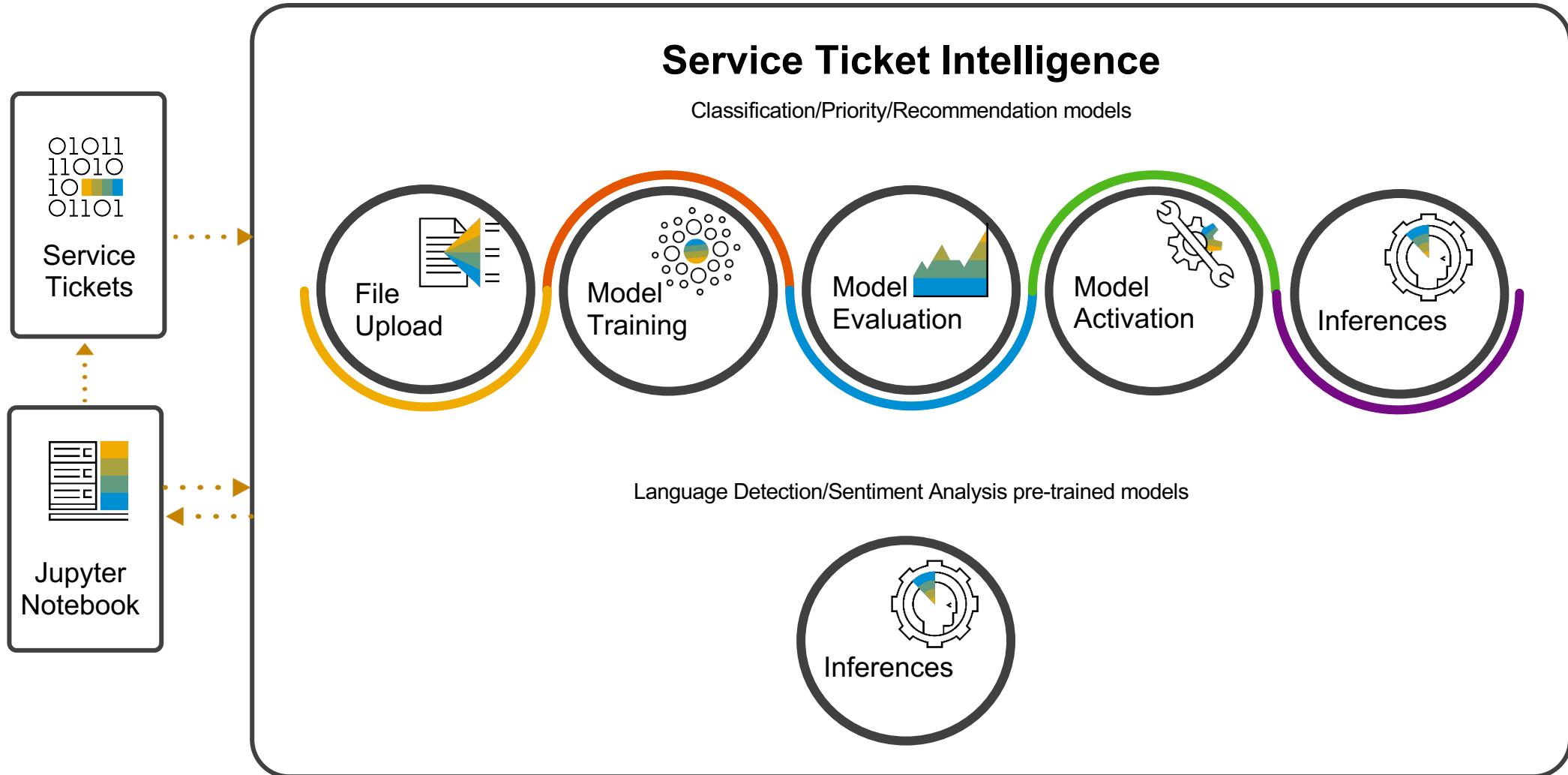


# Hands-on for Service Ticket Intelligence



# Service Ticket Intelligence Overview

## Ticket Classification and Recommendation Scenarios



# Exercise 1 – Provision and retrieve service keys for Service Ticket Intelligence

1. Register for a SAP CP trial account and login to <https://account.hanatrial.ondemand.com/cockpit/>
2. Follow **Exercise 1** to locate your service keys
3. Find your service credentials in the service keys

Service Key Parameter	Value
uaa-url	
uaa-clientid	
uaa-clientsecret	
sti_service_url	

4. Your service keys are required for **Exercise 2**

The screenshot shows the SAP Cloud Platform Cockpit interface. In the left sidebar, under 'Services', 'Service Instances' is selected. A table lists one instance: 'sti-test' (Service: Service Ticket Intelligence, Plan: standard, Status: Created). On the right, there are sections for 'Bound Applications (0)' and 'Service Keys (1)'. The 'Service Keys (1)' section shows a key named 'sti-key'.

Learning point: The client ID provided has full privileges to the service. The client application should grant different access privileges to different user roles (e.g. admin versus end user). Access and secret keys should not be given to non-admin users. Read more in [STI Security Guide](#).

# Exercise 2 – Ticket Classification Scenario

## Data preparation, model training & evaluation, inferences

1. Jupyter Notebooks are interactive playgrounds to code and are often used in data science to explore datasets
2. Open the Classification Jupyter demo notebook in [Exercise 2](#) and follow the steps
3. Dataset used in this demo will be in the context of service request complaints from the finance industry

The screenshot shows a Jupyter Notebook interface with the title "Classification Demo". The notebook contains the following text and code:

In this notebook, we will see how to prepare the data for classification, upload the data, start training and do inference.

Install pyjwt library if not already installed

```
In [1]: pip install pyjwt
Requirement already satisfied: pyjwt in c:\users\i312065\appdata\local\continuum\anaconda3\lib\site-packages (1.7.1)
```

In [2]: import os
import pandas as pd
from sklearn.model\_selection import train\_test\_split

matplotlib inline
import matplotlib.pyplot as plt
import jwt
import time
import json

In [3]: import requests
import base64

Prepare training and test data

We will use the sentiment140 dataset for this demo which is publicly available. To view/download the full dataset, you may proceed to one of the urls:

1. <https://www.kaggle.com/kazanov/sentiment140>
2. <http://help.sentiment140.com/for-students/>

In this exercise, we will load the subset of twitter sentiment analysis dataset which has two columns:

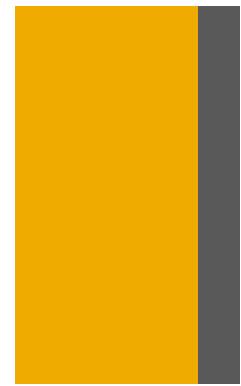
1. text (tweet) and
2. the sentiment polarity of the text (positive/negative associated with it).

Our objective will be to train a model to classify it as a positive or negative sentiment polarity with given text.

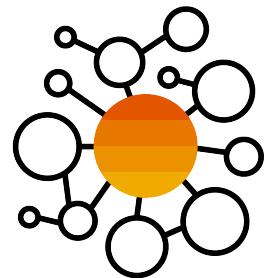


Learning point: STI Classification models are retrainable with multiple input labels for prediction.  
The input file should be UTF-8 formatted csv.  
Read more at [STI API Reference > Model Training API](#)

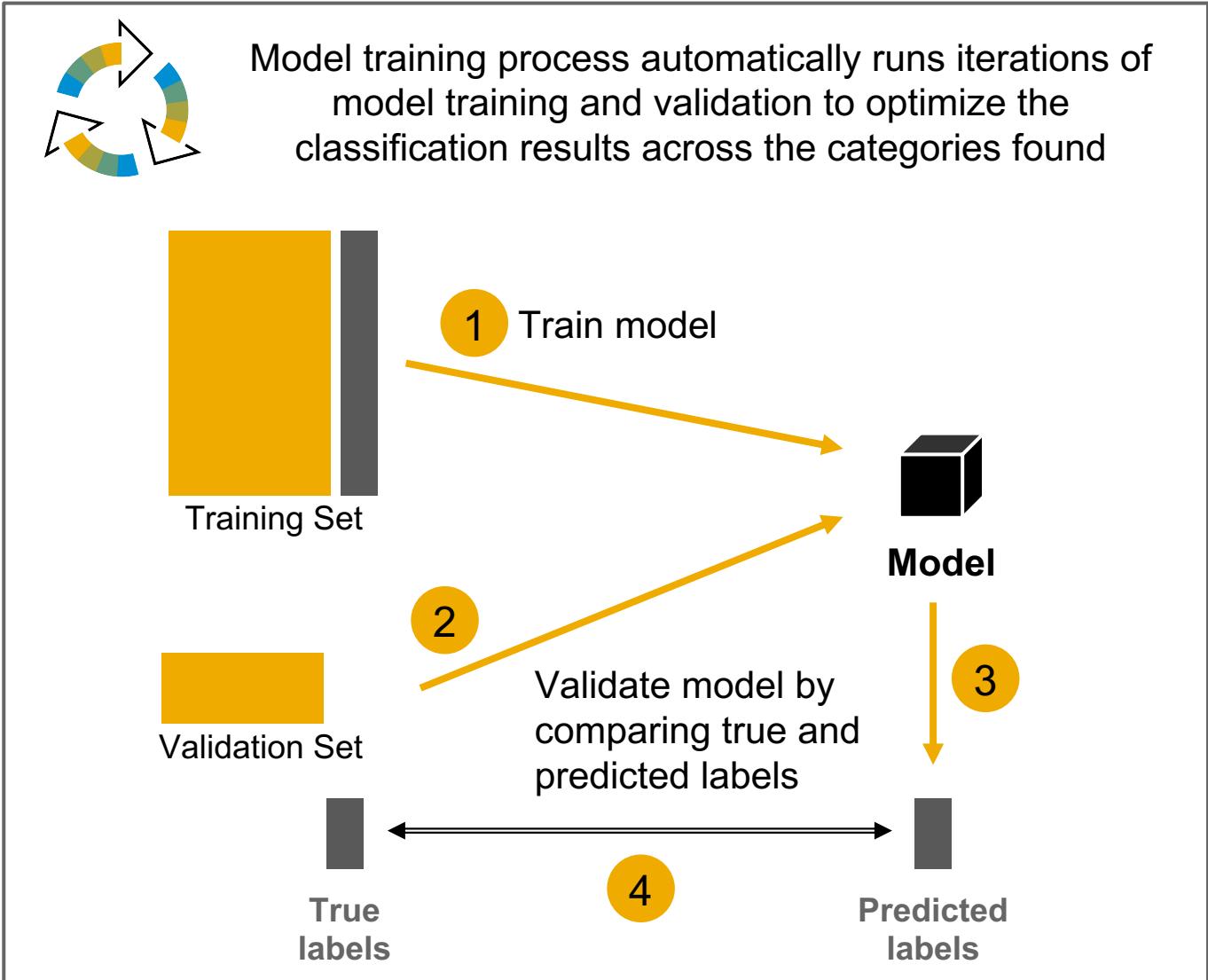
# What happens during model training?



Training data upload via file upload or oData connection with **text input** and **labels**



Model is initialized based on the categories found in the **labels**.  
Data is split into training and validation sets.



# Model Training Best Practices

## Recommended number of records

- At least 1,000 samples for every unique label value

## Minimum number of records

- At least 20 samples for every unique label (labels, e.g. categories, with less than 20 training examples are treated as outliers and are removed automatically by STI)

## Data distribution

- Uniform data distribution across all unique labels (good to have)

## Business activity level

- At least 10,000 service tickets per month

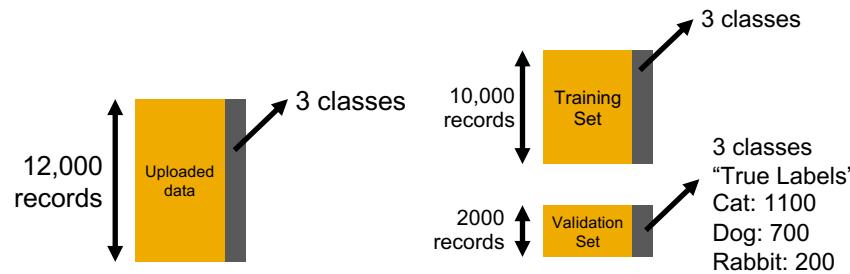
Generally speaking, the more data the better..

# Evaluating model accuracy

- A model training report in the form of a confusion matrix on the validation set, and corresponding precision, recall and f1 values for each category label in the dataset is generated at the end of the model training process. This report (in JSON format) can be retrieved from the [Get Model Accuracy API](#)
- Overall model accuracy represents the test results from using the model to make predictions on the validation set. This value can be retrieved from the [Get Model Status API](#) as “[combined\\_accuracy](#)” in the Model Status Object.

## Best practices to improve model accuracy

- Prior to model training, evaluate how much training data is available, the absolute number of categories, and the distribution of data across these categories.
- If a dataset has a large number of categories and semantically similar categories (or data which is very similar between two categories e.g. request for refund vs request for replacement), this typically makes the data noisy and thus results in poorer accuracies.



There are 1100 actual cats (true labels), of which 850 were accurately classified as cats, 200 were classified as dogs and 50 were classified as rabbits. Precision rate for the ‘cat’ label is 0.77 (850 / 1100).

Of the 1080 cats predicted by the model (predicted labels), 850 were truly cats.

Recall rate for the ‘cat’ label is 0.79 (850 / 1080).

The f1 score is a harmonic mean of the precision and recall values, which therefore gives a single score that represents both precision and recall rates for a particular label.

## Overall model accuracy

Sum of total correct predictions over all predictions made on the validation set.

$$(850 + 500 + 100) / 2000 = 0.725$$

True label	Cat	Dog	Rabbit
Predicted label	850	200	50
Cat	150	500	50
Dog	80	20	100
Rabbit			

```
confusion_matrix : {
    labels: [
        'cat',
        'dog',
        'rabbit'
    ],
    values: [
        [850, 200, 50],
        [150, 500, 50],
        [80, 20, 100]
    ]
}
```

See [term definition](#) for the Confusion Matrix Object

# Developer resources

The screenshot shows the SAP Help Portal interface for the Service Ticket Intelligence API Reference. The top navigation bar includes the SAP logo, 'SAP Help Portal', and the document title 'Service Ticket Intelligence'. Below the navigation is a search bar and a 'Download PDF' button. The main content area features a large QR code in the center. To the left is a sidebar with a 'Table of Contents' section containing links to 'Introduction', 'Term Definition', 'Recommended Workflows', 'Authorization', 'The API URL Structure', 'Model Training API', 'Classification API', 'Recommendation API', 'Common Status and Error Codes', and 'Glossary'. The main content page has a header 'Service Ticket Intelligence API Reference' and sections for 'Introduction' and 'Related Information'.

API documentation on [help.sap.com/stint](https://help.sap.com/stint) > [API Reference](#)

The screenshot shows the SAP Cloud Platform Cockpit interface. The top navigation bar includes the SAP logo, 'SAP Cloud Platform Cockpit', and user information like 'Phoebe'. The left sidebar is titled 'Services' and includes options for 'Service Marketplace', 'Service Instances', 'User-Provided Services', 'Portal', 'Routes', 'Security Groups', 'Events', and 'Members'. The main content area shows a list titled 'Space: dev - Service Marketplace' with one item: 'service-ticket-intelligence' by 'SAP Service Ticket Intelligence'. To the right of the list is a large QR code.

[Personal Trial on SAP Cloud Platform](#)

The screenshot shows the SAP Developers portal. The top navigation bar includes the SAP logo, 'Developers', and tabs for 'Products', 'Tutorials', 'Trials and Downloads', and 'Res'. The main content area features a tutorial titled 'Use Machine Learning to Process Service Requests'. It includes a brief description: 'Classify customer service requests and get solution recommendations – either with Postman or Jupyter Notebook Services in SAP Cloud Platform.', a 'Details' section, and two call-to-action buttons: 'Using Postman' (with a link to 'Get a Free Trial Account on SAP Cloud Platform') and 'Create Service Instance for Service Ticket Intelligence' (with a link to 'Create a service instance and the associated service keys for Service Ticket Intelligence, one of the SAP AI Business Services, using the SAP Cloud Platform trial cockpit'). To the right of the content is a large QR code.

[Developer Tutorials](#)

YOU ASK, I  
ANSWER



# Continue your learning experience from SAP TechEd in 2020

Your exclusive path to build and maintain SAP solution skills anytime, any place

**Get empowered with access to relevant, up-to-date digital learning for SAP TechEd participants through a complete enablement solution that drives adoption and innovation.**



## Deepen your learning experience from SAP TechEd

[Activate your free access](#) to SAP Learning Hub, event edition, for:

- **Learning Journey** illustrations to guide you through **complementary** self-paced learning content
- **Content specific to SAP TechEd** in the online **SAP Learning Room for SAP TechEd**
- Access to SAP experts in **special live sessions**



## Deepen and validate your SAP solution skills

[Subscribe](#) to SAP Learning Hub, solution editions, for:

- **Solution-specific Learning Journey guides, content, collaborative learning, and hands-on practice** for your role and goals
- Drive performance and business success with validated solution expertise from the **SAP Global Certification** program

## Your benefits

- Gain insight into the latest innovations, and master software proficiency
- Keep skills up-to-date, and enable performance and business success with help from SAP solution experts
- Achieve competitive advantages and digital transformation success with trusted certifications

**500,000+**

Learners in SAP Learning Hub

**100+**

Experts getting certified per day

**150+**

SAP Global Certifications



# Thanks for attending this session.

## Contact for further topic inquiries

Data Attribute Recommendation  
Michael Haas  
[michael.hass01@sap.com](mailto:michael.hass01@sap.com)

Service Ticket Intelligence  
Samuel Seah  
[samuel.seah@sap.com](mailto:samuel.seah@sap.com)

## Follow us



[www.sap.com/contactsap](http://www.sap.com/contactsap)

© 2020 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platforms, directions, and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See [www.sap.com/copyright](http://www.sap.com/copyright) for additional trademark information and notices.