



Data Classification with Python SDK and **SAP AI Business Services**

Michael Haas, Senior Software Developer

Samuel Seah, Product Owner

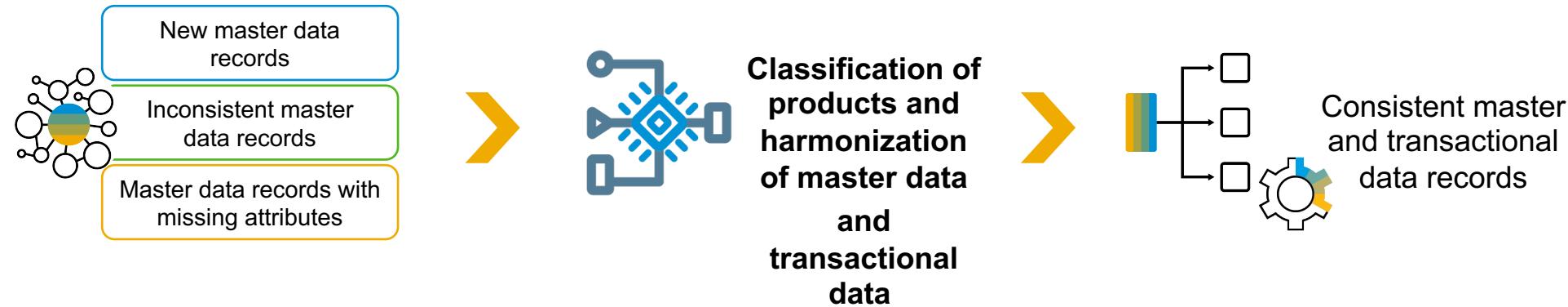
INTERNAL

DATA CLASSIFICATION WITH

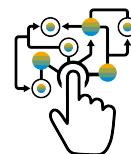
Data Attribute Recommendation

Data Attribute Recommendation

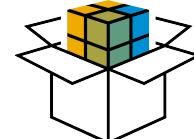
Automate master data / transactional data management tasks



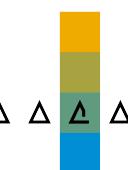
Data Attribute Recommendation helps to classify entities such as products, stores and users into multiple classes, using free text, numbers and categories as input.



Automate and speed up master data creation and maintenance



Gain easier and faster master data insights



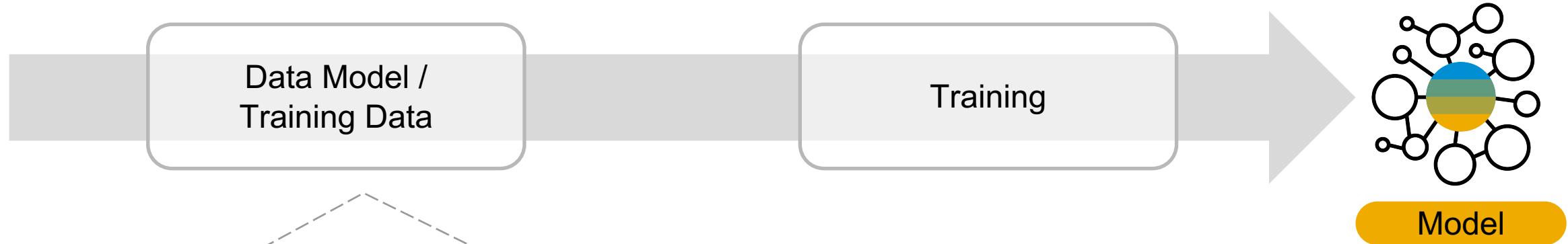
Reduce errors and manual efforts when maintaining master data

WORKSHOP HANDS-ON

Data Attribute Recommendation **Hands-On**

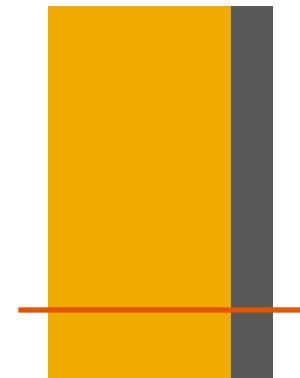
- Follow the exercises in the github.com/SAP-samples/teched2020-INT260 repository

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)



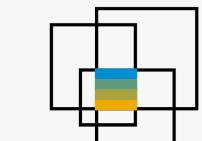
Training data consists of **text input** and **labels**

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.

Use Cases for Data Attribute Recommendation

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



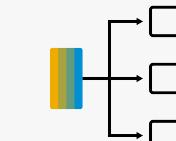
Point of Sales
information



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



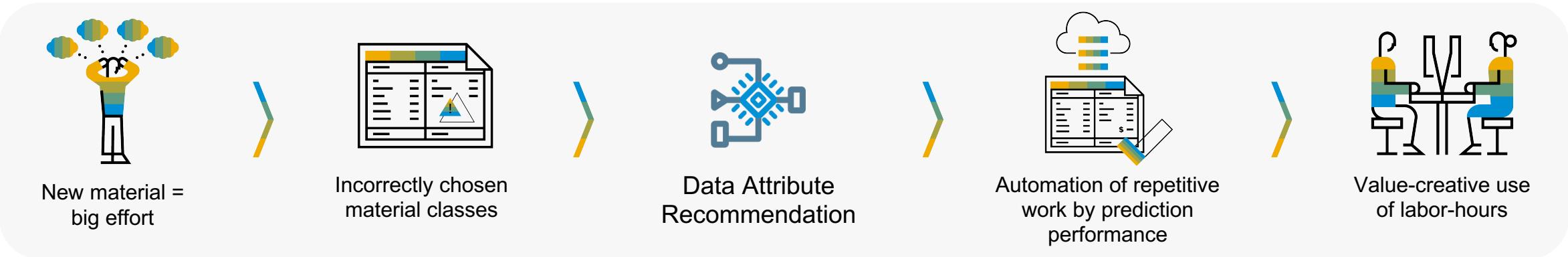
BENEFITS

- **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

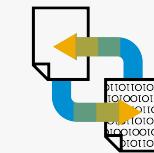
Scenario 3 – Commodity Code Prediction



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

DATA CLASSIFICATION WITH SERVICE TICKET INTELLIGENCE

Service Ticket Intelligence with SAP Artificial Intelligence

Reimagine Customer Service with Automated Processes



Process customer issues faster and deliver great customer experience



Speed up service response times



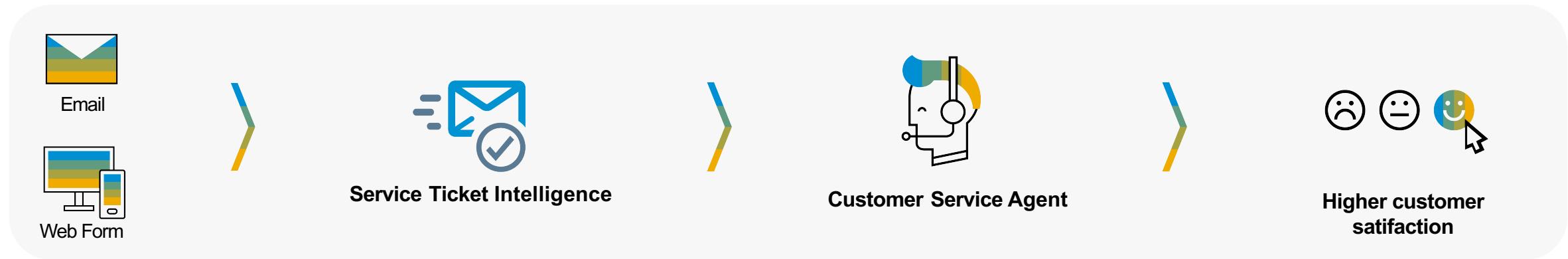
Have higher rates of problem resolution and ticket closure



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

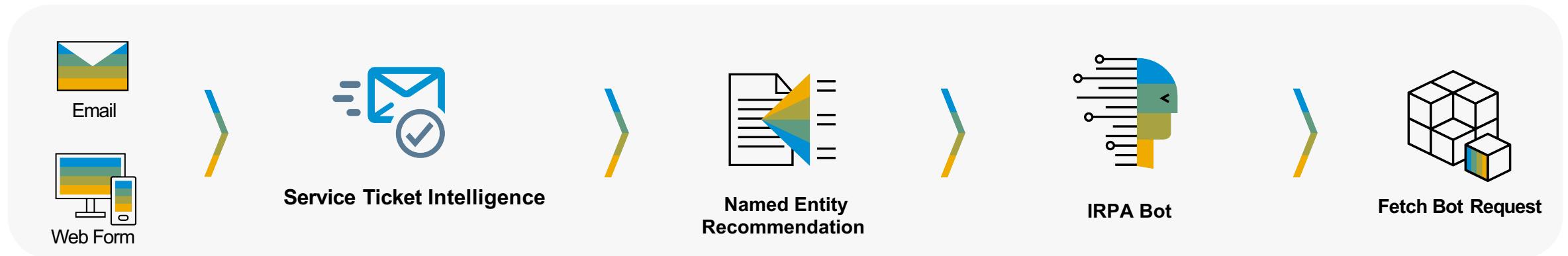
Service Ticket Intelligence:

- 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:

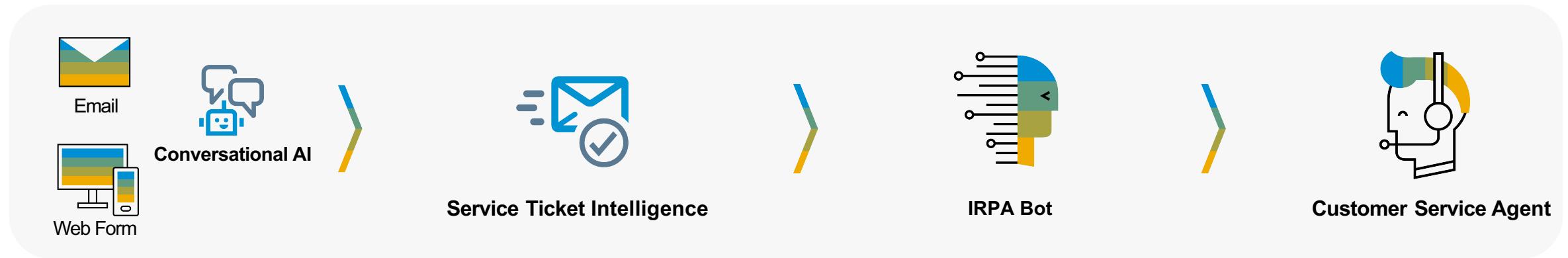
- Automatization of **classification** and **processing** of customer service requests
- Categorization of customer input **from various channels** into internal service groups
- **Redirection** of categorized inquiries to right processor

- **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:

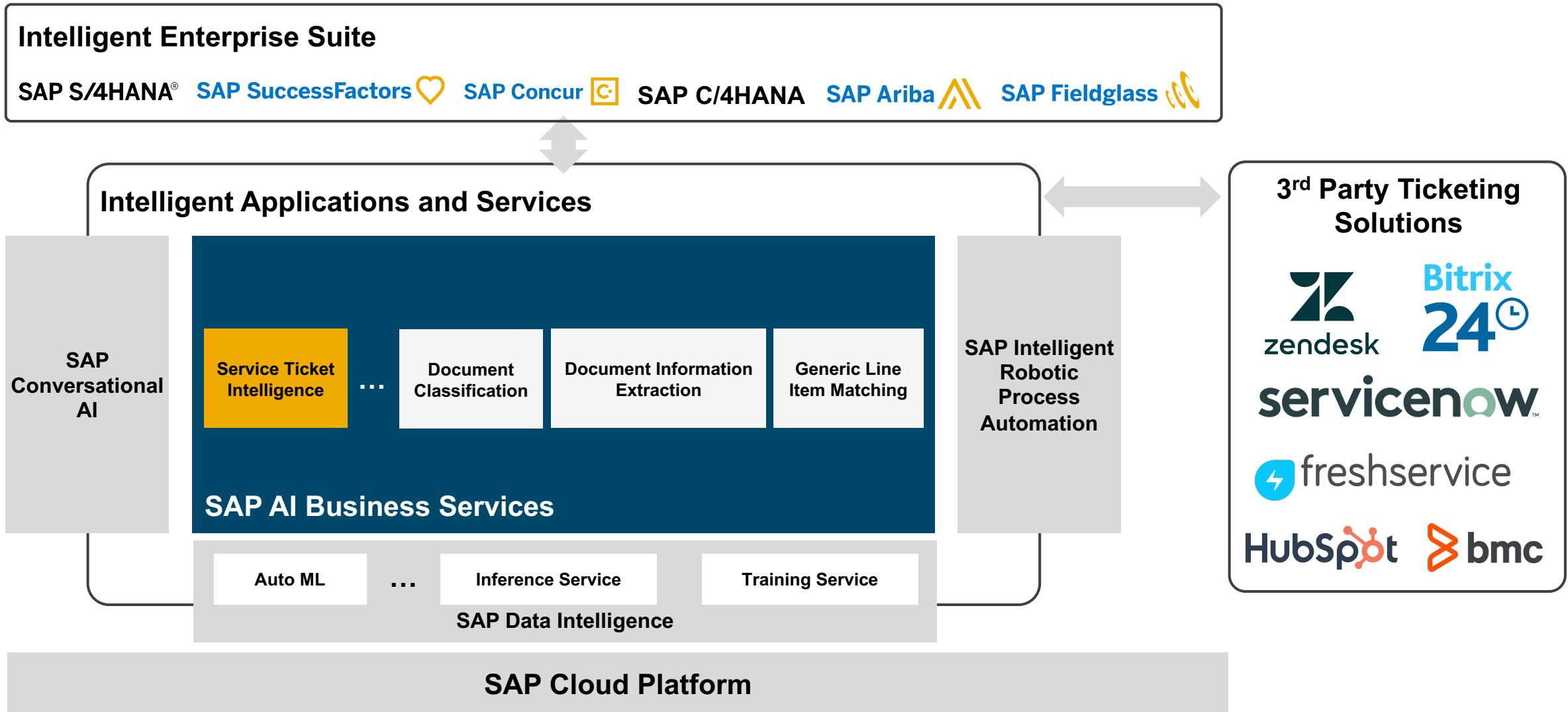
- Automatization of **classification** and **processing** of customer service requests
- Categorization of customer input **from various channels** into internal service groups

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

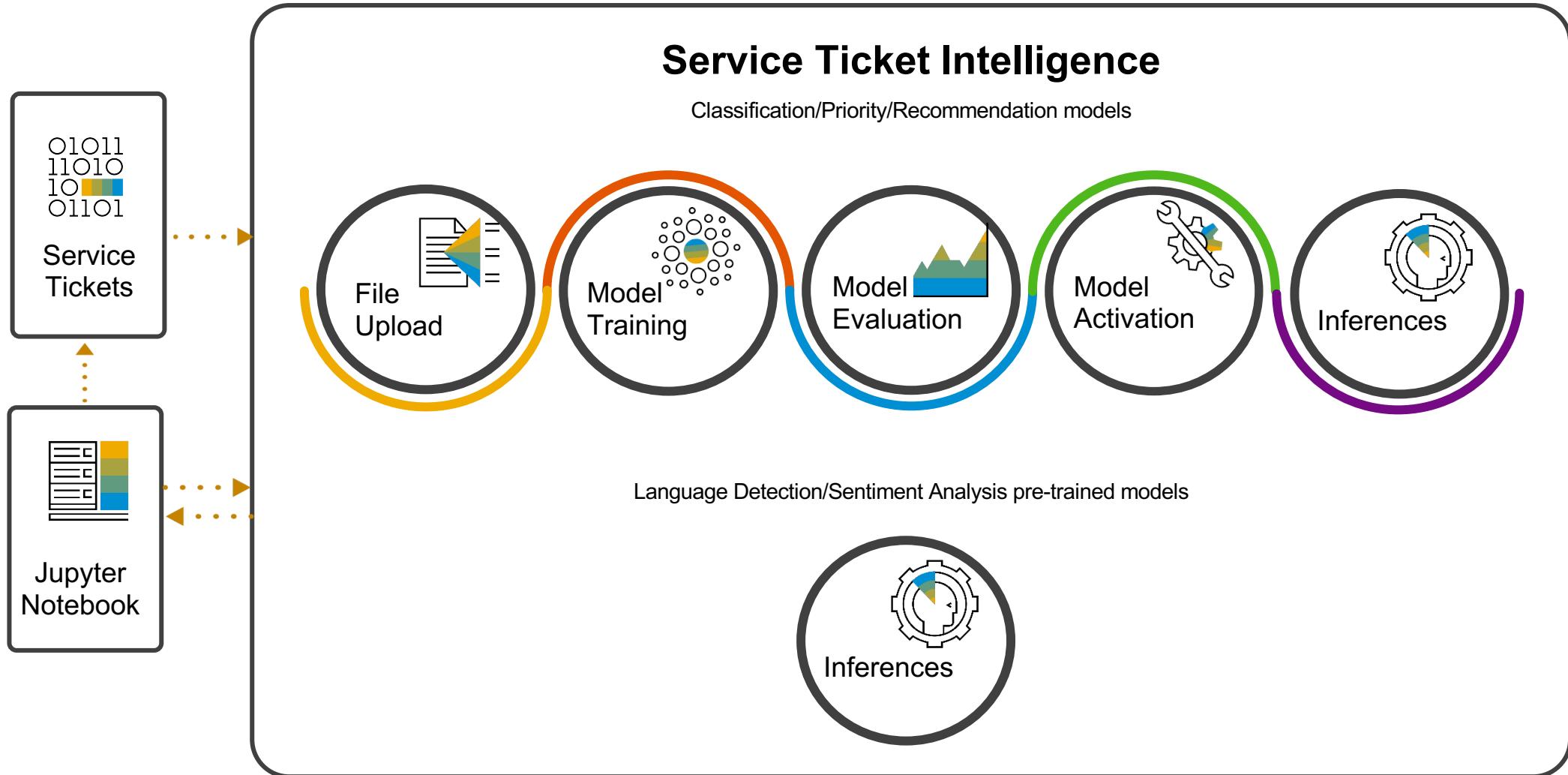
Service Ticket Intelligence can be integrated with various ticket handling applications



WORKSHOP HANDS-ON

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)'. Under 'Service Keys (1)', it 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 content:

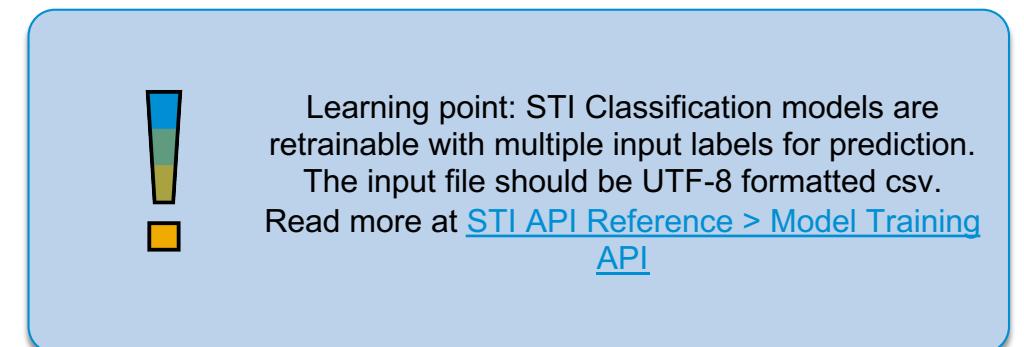
Classification Demo
In this notebook, we will see how to prepare the data for classification, upload the data, start training and do inference.

Install pyjwt library if not already installed

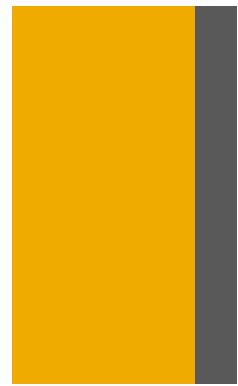
```
In [1]: pip install pyjwt
Requirement already satisfied: pyjwt in c:\users\i312065\appdata\local\continuum\anaconda3\lib\site-packages (1.7.1)
```

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/>

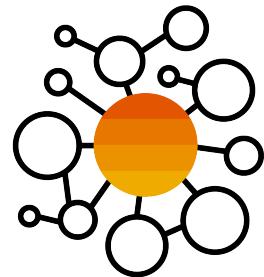
In this exercise, we will load the subset of twitter sentiment analysis dataset which has two columns:
1. text (tweet) and
2. the sentiment polarity of the text (positive/negative associated with it).
Our objective will be to train a model to classify it as a positive or negative sentiment polarity with given text.



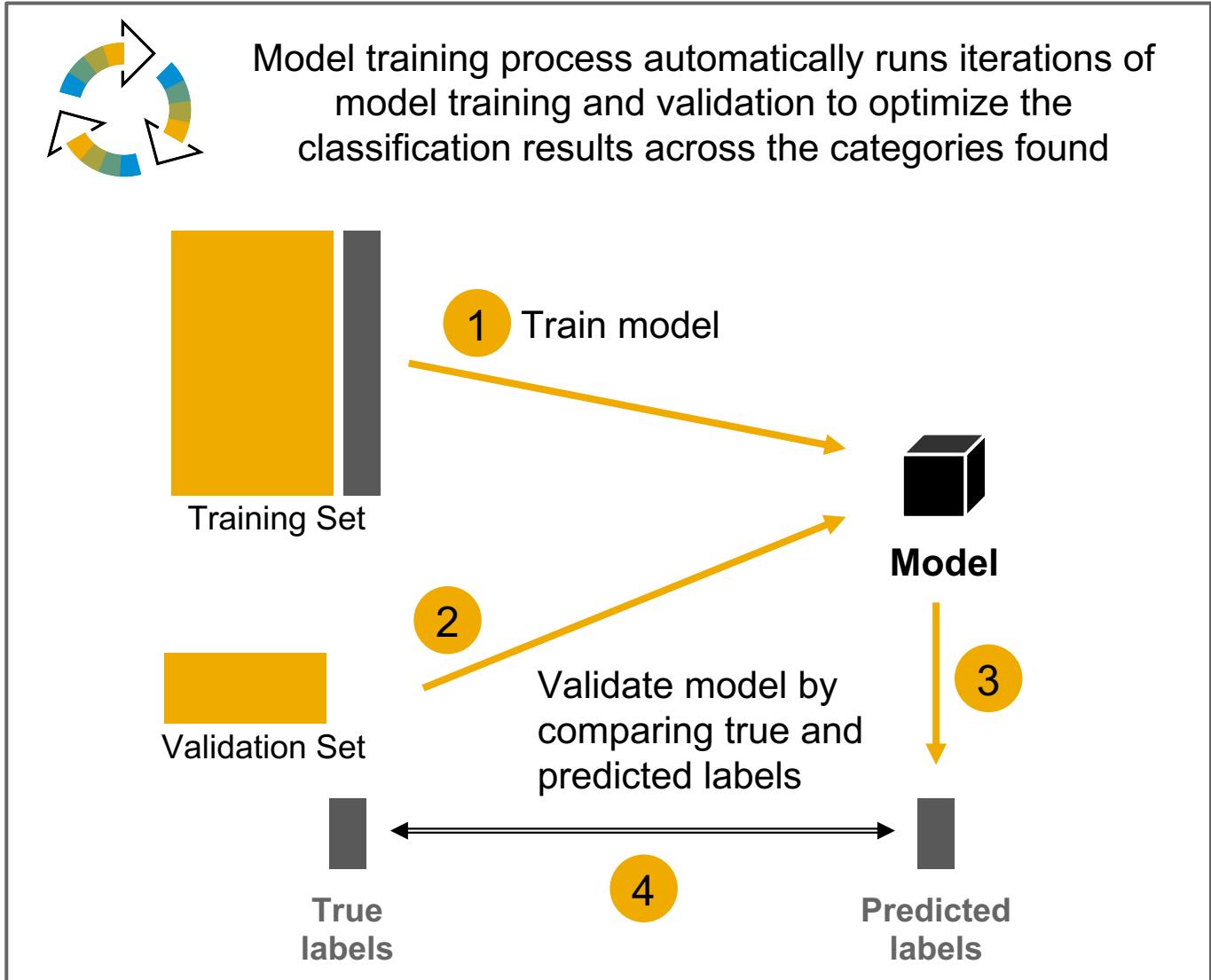
What happens during model training?



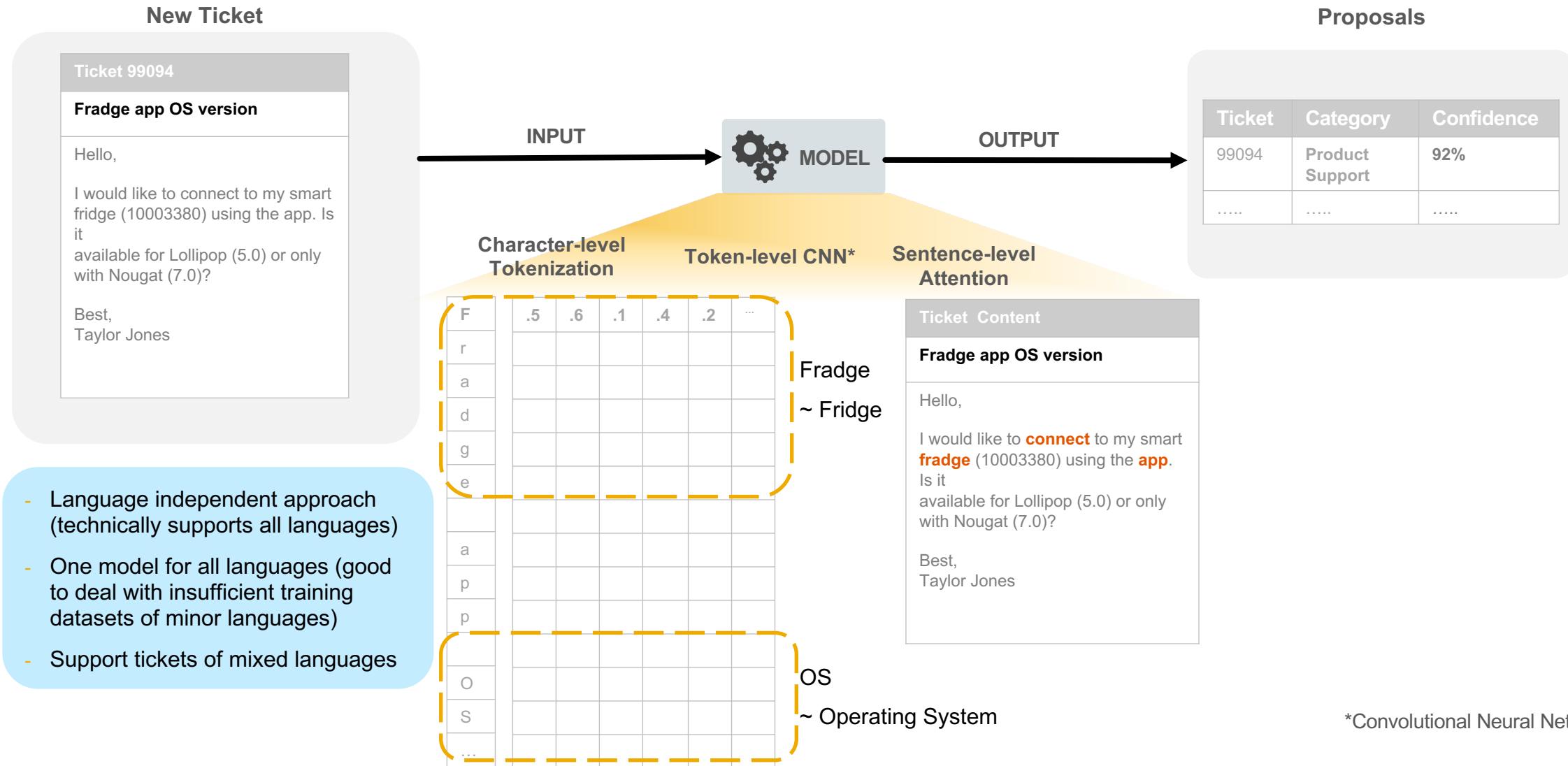
Training data upload via file upload or oData connection with **text input** and **labels**



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



STI Ticket Classification (Categorization) Model



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

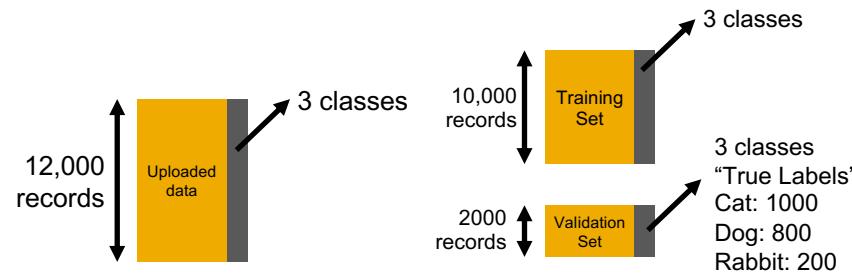
Generally speaking, the more data the better..

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.

$$(850 + 200 + 100) / 2000 = 0.575$$

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

```
confusion_matrix : {
    labels: [
        'cat',
        'dog',
        'rabbit'
    ],
    values: [
        [850, 200, 50],
        [150, 200, 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 sections for 'Introduction' (with a 'Send feedback' link) and 'Related Information' (links to 'API Version' and 'API Previous Versions').

API documentation on 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 website. 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 green circular icon with a white 'T' symbol, a brief description, and a 'Details' link. Below this are sections for '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

Thank you.

Contact information:

Michael Haas
michael.haas01@sap.com

Samuel Seah
samuel.seah@sap.com

Partner logo

THE BEST RUN  SAP®