



# Pega NLP Essentials

STUDENT GUIDE

**© Copyright 2023**  
**Pegasystems Inc., Cambridge, MA**  
All rights reserved.

This document describes products and services of Pegasystems Inc. It may contain trade secrets and proprietary information. The document and product are protected by copyright and distributed under licenses restricting their use, copying, distribution, or transmittal in any form without prior written authorization of Pegasystems Inc.

This document is current as of the date of publication only. Changes in the document may be made from time to time at the discretion of Pegasystems. This document remains the property of Pegasystems and must be returned to it upon request. This document does not imply any commitment to offer or deliver the products or services provided.

This document may include references to Pegasystems product features that have not been licensed by your company. If you have questions about whether a particular capability is included in your installation, please consult your Pegasystems service consultant.

PegaRULES, Process Commander, SmartBPM® and the Pegasystems logo are trademarks or registered trademarks of Pegasystems Inc. All other product names, logos and symbols may be registered trademarks of their respective owners.

Although Pegasystems Inc. strives for accuracy in its publications, any publication may contain inaccuracies or typographical errors. This document or Help System could contain technical inaccuracies or typographical errors. Changes are periodically added to the information herein. Pegasystems Inc. may make improvements and/or changes in the information described herein at any time.

This document is the property of:  
Pegasystems Inc.  
1 Rogers Street  
Cambridge, MA 02142  
Phone: (617) 374-9600  
Fax: (617) 374-9620  
[www.pegasystems.com](http://www.pegasystems.com)

**Mission:** Pega NLP Essentials

**Products:** Pega Platform™ '23, Pega Customer Decision Hub™, Pega Customer Service™

**URL:** <https://academy.pegasystems.com/mission/pega-nlp-essentials/v3>

**Date:** 25 August 2023

# Contents

Pega NLP overview.....	4
Text analytics.....	5
Pega NLP overview.....	10
Text analytics for email routing.....	15
Email routing.....	16
Training a topic model to improve email routing.....	22
Using entity extraction with chatbot channel.....	30
Training predictions for chatbot.....	31

# Pega NLP overview

## Description

Gain a greater understanding of the key features, capabilities, and benefits of Pega natural language processing in decision management context. Businesses are exploring ways to utilize language processing. With Pega NLP you can analyze and extract meaningful information from a large amount of text by using text analytics to improve business performance and customer experience.

## Learning objectives

- Explain text analytics in Pega NLP context.
- Describe practical use cases where text analytics can be used.

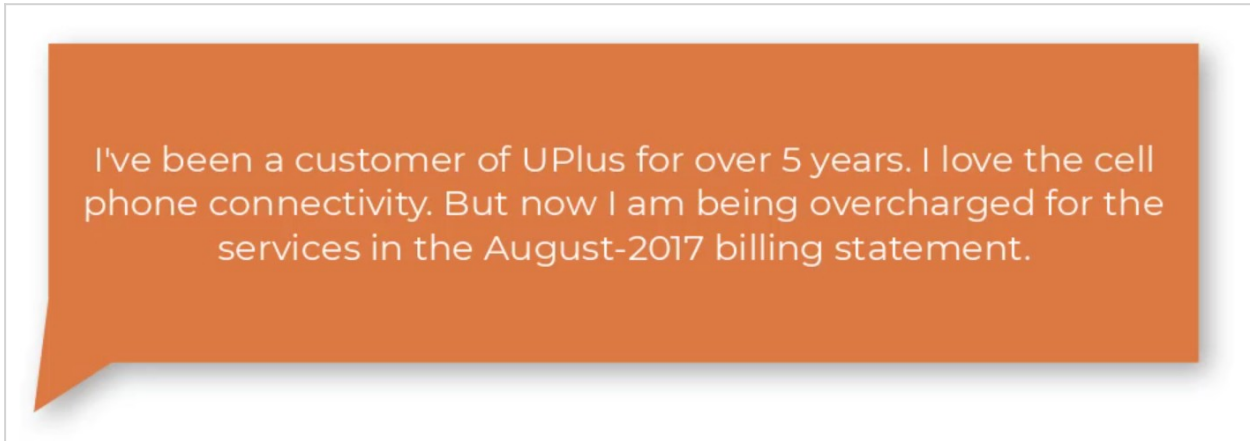
# Text analytics

Today, digital sources such as news, blogs, social media, emails, and chatbots generate continuous text. Text analytics helps to transform large amounts of text into a structured format so that your applications can use the text to optimize processes such as email routing, automatic chatbot interactions, or entity extraction.

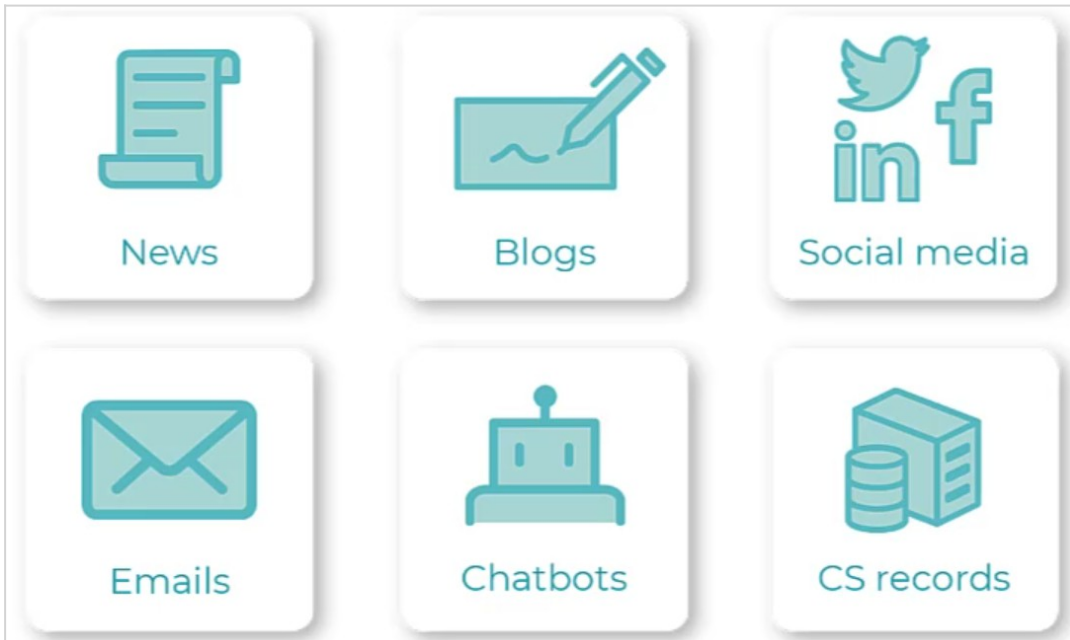
## Transcript

Learn about text analytics and how it helps determine customer sentiments and needs.

Consider this complaint from a customer to a customer service department.



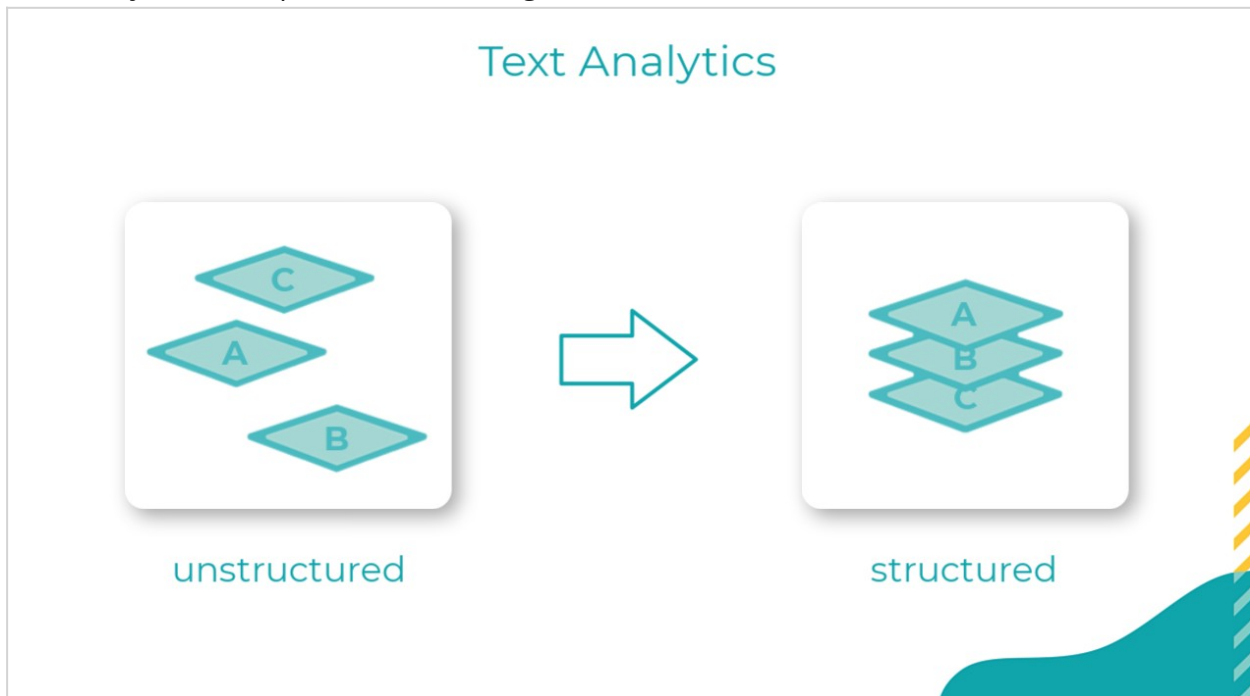
Humans can understand the implications of a piece of text by summarizing, elaborating, or even rephrasing it while preserving its original meaning. The challenge for humans comes when text exists on a large scale.



Consider a company with tens of thousands of customer service records in text form.

The company wants to route all complaints automatically to the customer service department. Then, it wants a report to understand whether customers are satisfied with various products and services. The company must find a way to process and extract the text, and text analytics is the most efficient and effective way to accomplish this task.

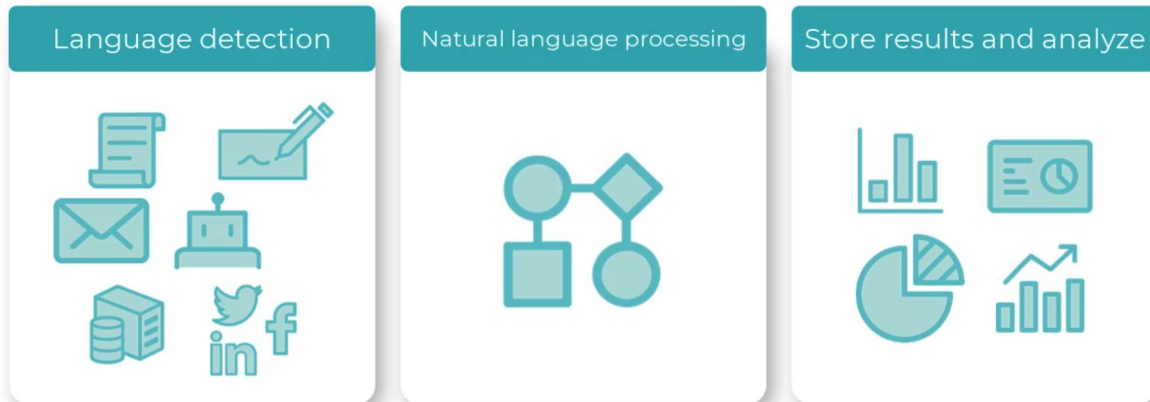
Text analytics is the process of deriving valuable information from text.



It involves converting large amounts of text into structured data that can be analyzed using statistical methods. There are three main steps involved in text analytics:

1. Language detection
2. Natural language processing
3. Store results and analyze

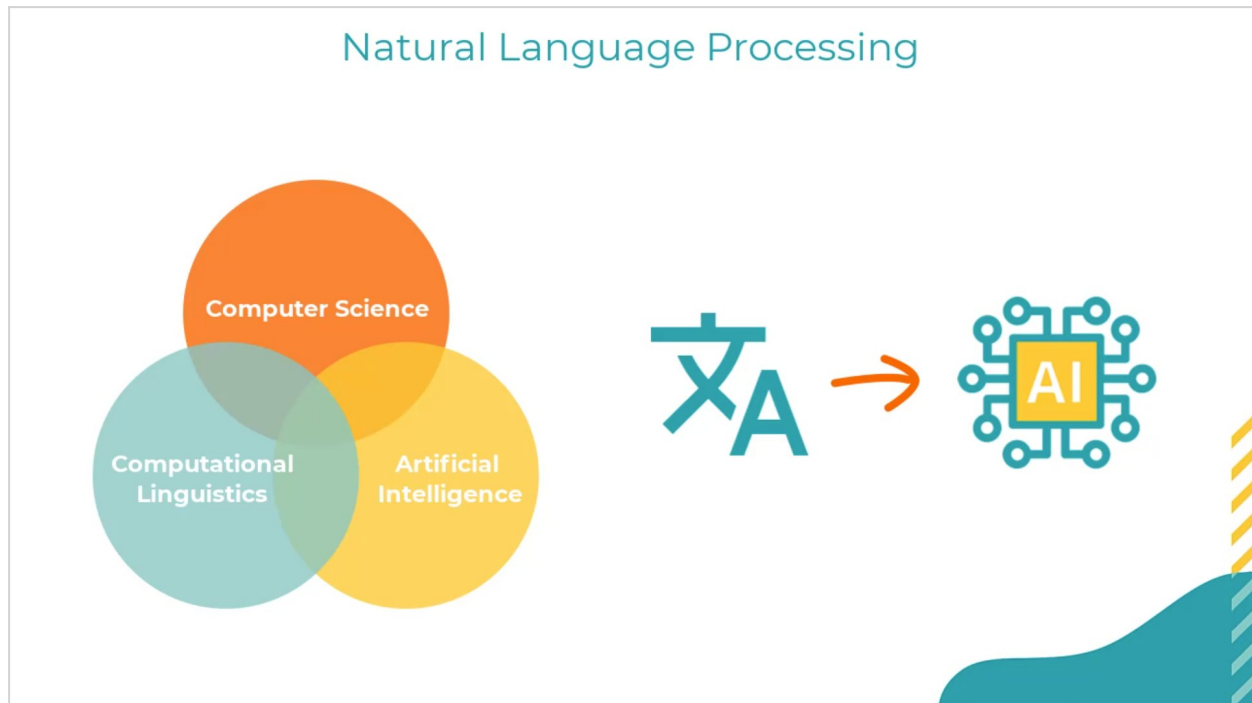
## Text Analytics Process



First, the system collects data from a source, such as social media, emails, chatbots, or customer service records. Second, natural language processing techniques are applied to extract specific attributes from the text and present them as structured data. And third, the extracted structured data is stored and analyzed.

Natural language processing is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human

language.

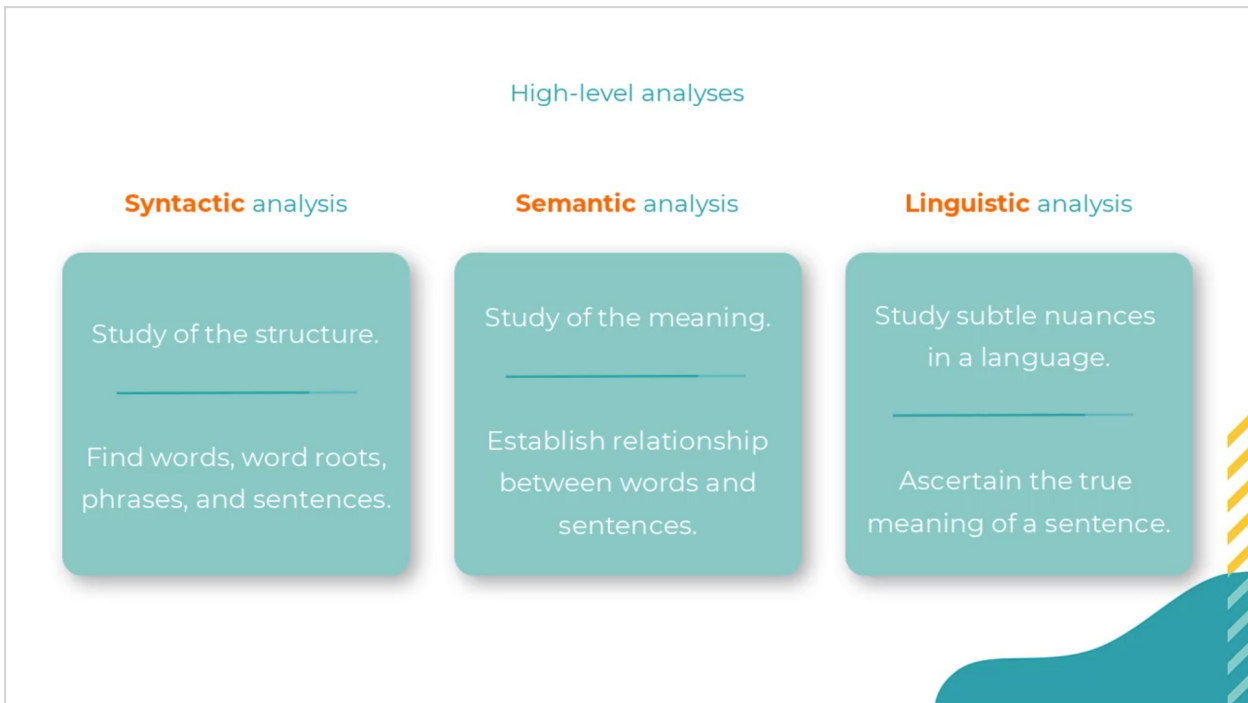


In particular, it is concerned with programming computers to process large volumes of natural language text fruitfully. At a high level, the following analyses occur during natural language processing.

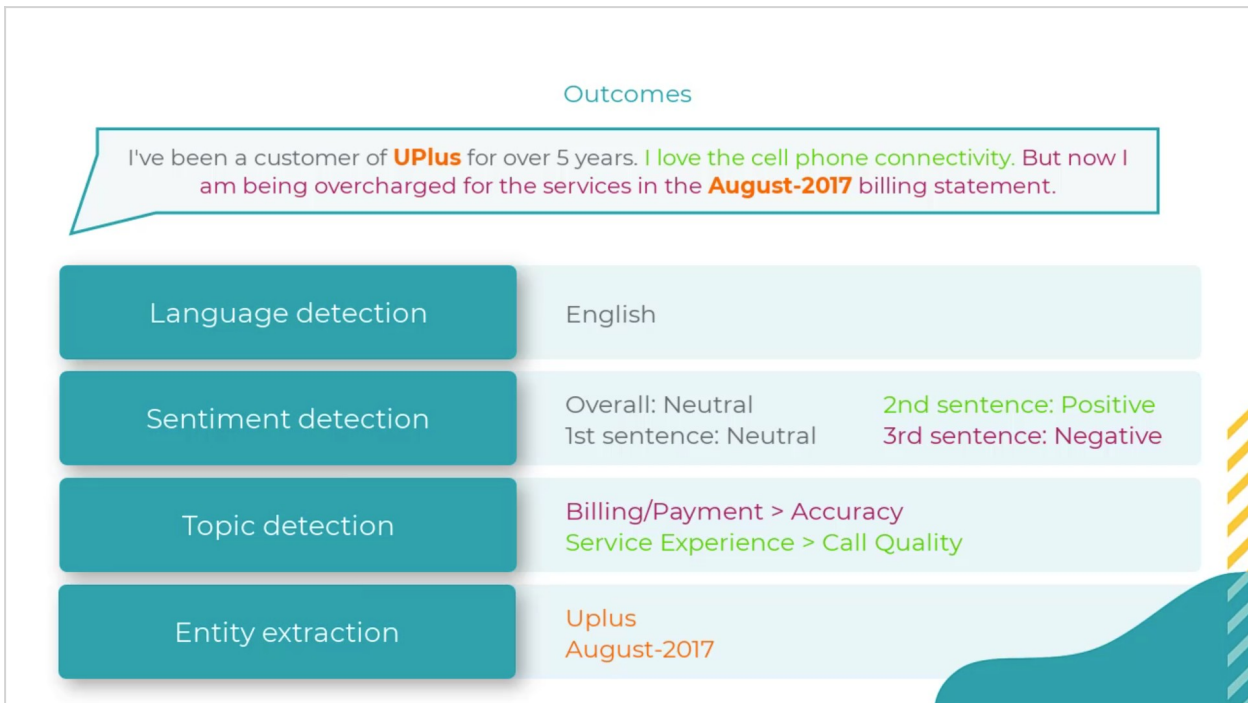
Syntactic analysis analyzes the structure of text and recognizes, for example, nouns, verbs, and adjectives. Semantic analysis establishes the meaning of a piece of text and the relationship between words and sentences. The linguistic analysis phase determines the



exact meaning of the text.



The processing of the text from the customer who claims that they were overcharged produces a series of outcomes. First, the analysis detects the language because this action is a prerequisite for the next steps.



Then, the analysis detects the sentiment of the text as positive, negative, or neutral. The sentiment is the general attitude expressed towards a subject by the author of the message. The analysis reports the overall sentiment and sentiment for each sentence.

The analysis classifies the text into one or more predefined topics, such as the business functions of a company. It also reports the sentiment of the corresponding sentence for each category that it detects. Entities refer to the proper nouns found in the text (such as names of people, places, dates and times, and organizations), which helps establish the subject of the text.

You have reached the end of this video. You have learned:

- What text analytics is.
- How text analytics process works.
- What natural language processing is and how it works in text analytics.

# Pega NLP overview

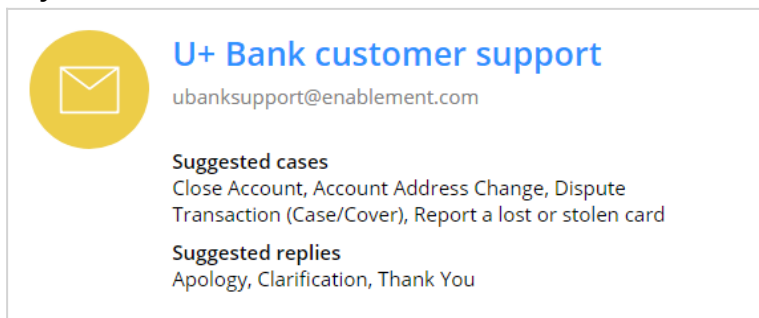
## Introduction

Pega natural language processing (NLP) focuses on text categorization and text extraction. Discover how the system uses Pega NLP to route incoming emails based on the topic of the email, and how NLP drives a chatbot to automatically detect the topic of the message, extract necessary entities, create a case, and then route it to the correct customer service workbasket.

## Transcript

This video introduces you to Pega NLP, a feature of the decision management capability of Pega Platform. You can apply decision management to any application that uses Pega Platform, which makes it highly versatile.

Pega NLP uses text analytics predictions to detect the topic and entities that are included in the message. For example, a company has an email channel where its customers can send any email.



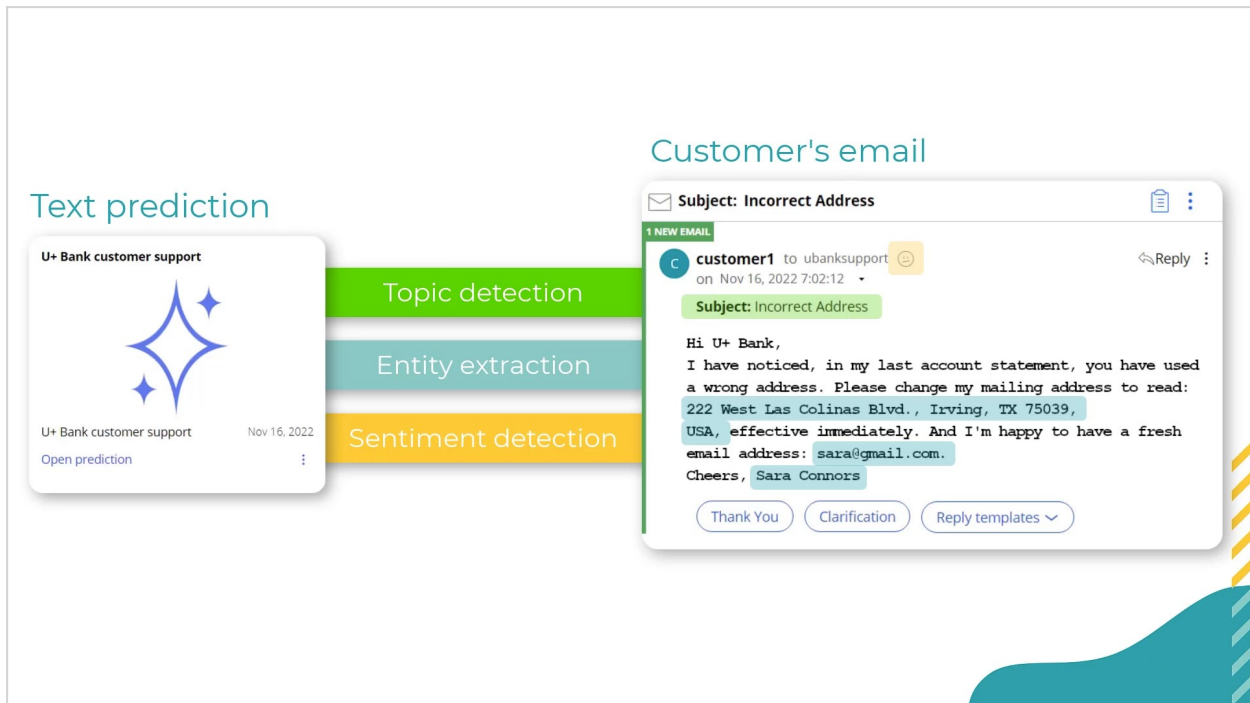
The screenshot shows a prediction interface for 'U+ Bank customer support'. It features a yellow envelope icon in a circle on the left. To the right, the text reads 'U+ Bank customer support' in blue, followed by the email address 'ubanksupport@enablement.com'. Below this, there are two sections: 'Suggested cases' with the text 'Close Account, Account Address Change, Dispute Transaction (Case/Cover), Report a lost or stolen card', and 'Suggested replies' with the text 'Apology, Clarification, Thank You'.

A text prediction drives the email channel. Consider this email sent by the customer through the company's email channel. The text prediction aims to detect the topic of the email and route it to the appropriate department.

Apart from the topic detection model, the prediction also includes an entity extraction model that can extract entities from an email. It can automatically detect and extract certain entities such as an account number, email address, or street address. This functionality allows to automatically process or prioritize emails.

In addition to topic detection and entity extraction, Pega NLP detects the sentiment of an email based on its content. The sentiment score ranges from -1 to 1 and is negative,

positive, or neutral.

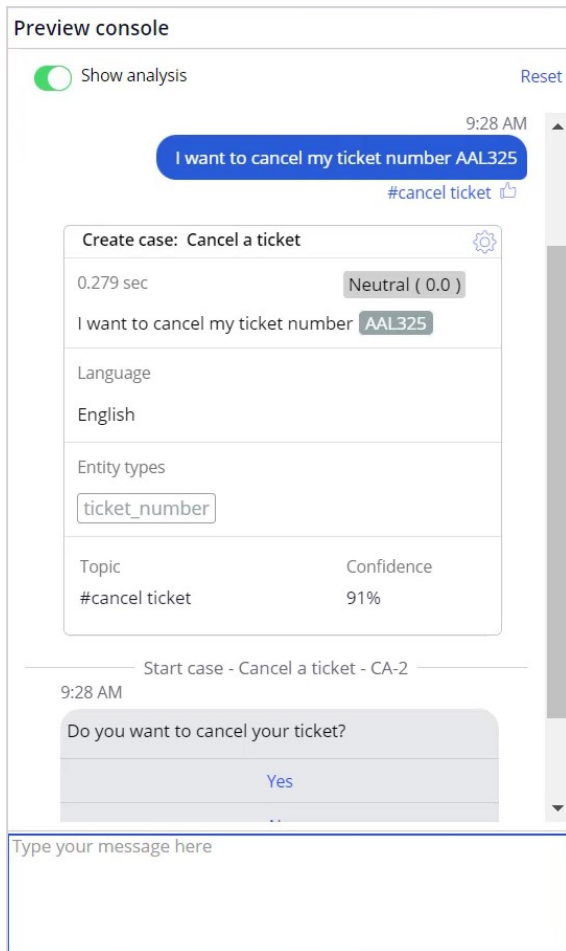


This functionality offloads work from customer service representatives, speeds up the process, and improves the customer experience.

Multiple channels can use the Pega NLP feature. For example, an airline sets up a chatbot channel to speed up customer service by automating the ticket cancellation process.

When Troy, a customer, sends a message to the Airline chatbot that he wants to cancel his ticket, the text prediction detects the topic of the message and runs the `Cancel a ticket` case type that an application developer preconfigured in the system. The case type contains the conversation flow that the chatbot uses in conversation with Troy. The chatbot uses entity extraction models to detect entities such as the ticket number, includes all necessary information in a case, and routes it to a correct customer service workbasket.

In App Studio, you can test the chatbot in the preview console.



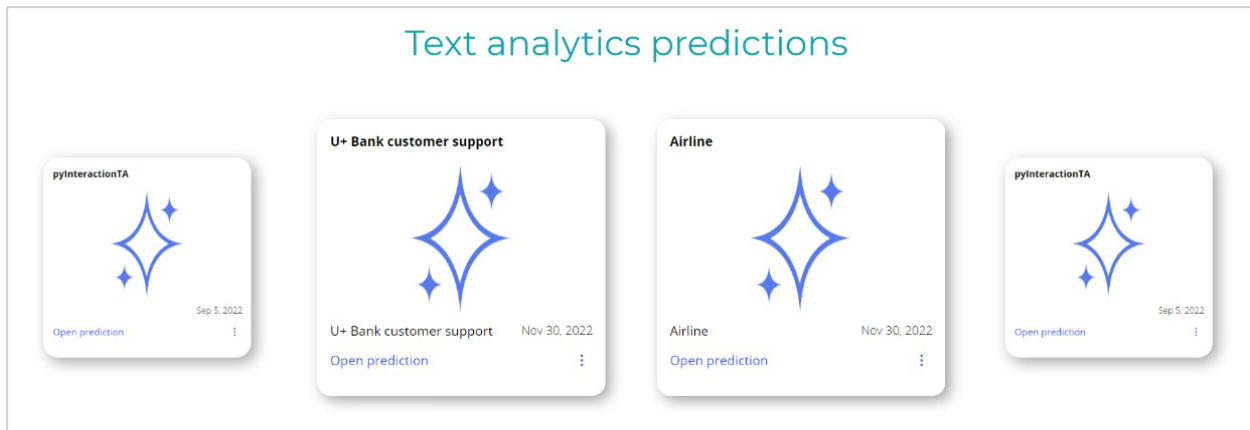
You can turn on the Show Analysis switch to see the chatbot details. Notice the customer's message. The topic detection model detects the topic of this message with 91 percent confidence, and the entity extraction model identifies the ticket number. Chatbot collects this information, includes it in a case, and then automatically routes it to a customer service representative for further processing.

Predictive models drive predictions. A data scientist manages predictions in Prediction Studio.

Prediction Studio is the dedicated workspace in Pega Platform where you manage the life cycle of predictive models and the predictions that they drive.

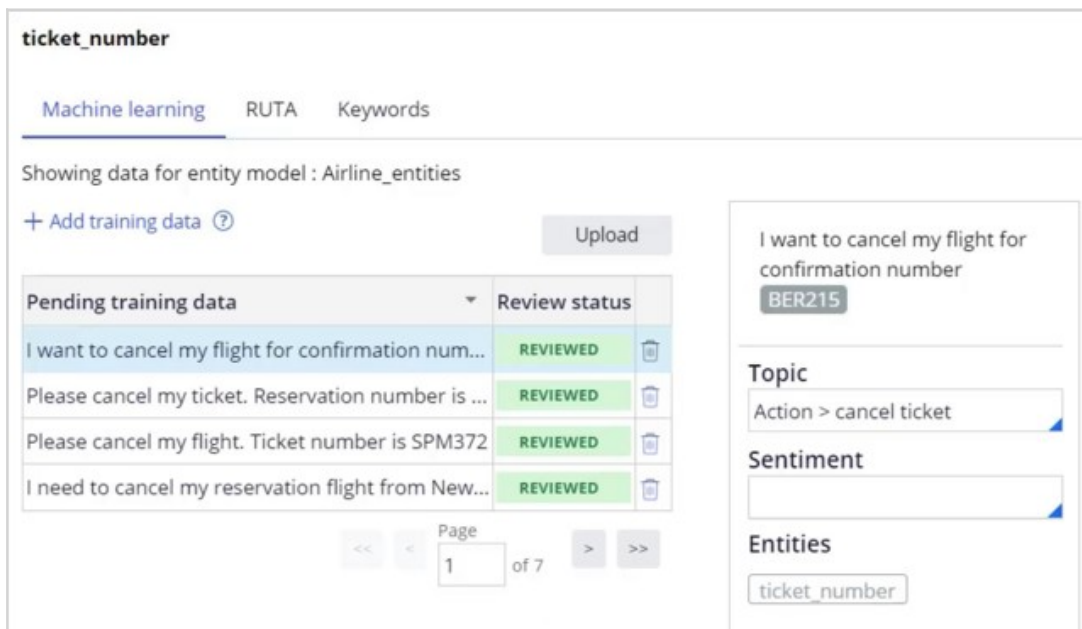
The workspace provides data scientists with everything they need to monitor and change predictions.

This is the Predictions landing page, where you manage predictions. Pega NLP uses Text analytics predictions.



Data scientists build topic detection model in a text prediction by using training data that they add manually or as a batch.

Data scientists also configure entity extraction in a text prediction. Several methods are available, including machine learning, Ruta scripts, or keyword-based entity extraction. It is also possible to use Google AutoML models. The machine learning-based model is the most versatile but requires training data to function. When adding entity extraction training data, you can also specify the topic of each message to further enhance the topic detection model.



You have reached the end of this video. You have learned:

- How Pega NLP allows you to improve business processes by using predictions.

- How topic and sentiment detection models and entity extraction models drive text analytics predictions.
- That machine learning models require training data.

# Text analytics for email routing

## Description

Humans can effortlessly interpret a single tweet but are unable to parse a large volume of information efficiently. Businesses are exploring ways to use machine learning to extract meaningful information from a large number of text messages. Learn how a text prediction can work to detect topics, extract entities, and identify the sentiment for incoming emails.

## Learning objectives

- Explain text analytics
- Describe practical applications where text analytics can be used
- Describe the role of machine learning in text analytics
- Explain how text predictions are trained on classified messages



# Email routing

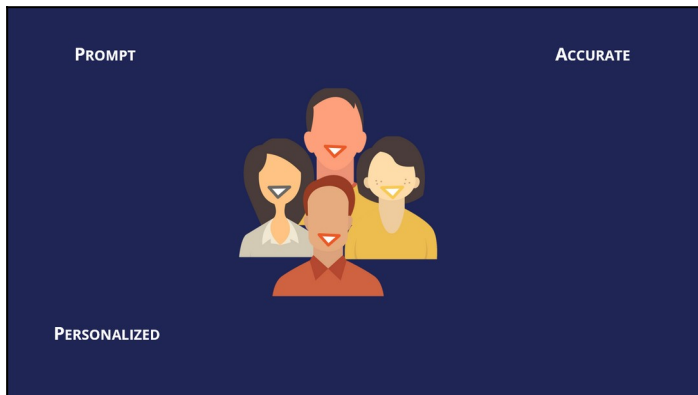
## Introduction

A company has an email channel where its customers can send any email — from service requests to compliments to sales inquiries — to the product support team. This email variety makes it difficult to provide each customer with a prompt and personalized response. Pega Infinity™ uses AI-powered text analytics to perform intelligent email routing.

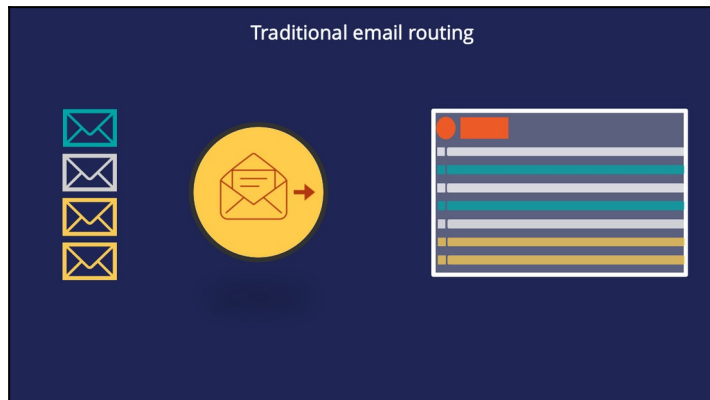
## Transcript

In this video, you will learn how Pega Infinity™ uses AI-powered text analytics to do intelligent email routing.

Customer satisfaction is a reflection of what a customer expects from a company vs. what they experience from the company. Meeting, or even better, exceeding customer expectations means addressing their service requests and complaints promptly, accurately, and with personalized solutions. Doing this will ensure the customer has a great experience.



Assume a company has an email channel in which its customers can send any type of email —from service requests to compliments to sales inquiries—to the product support team. These emails are often routed to the same container and are uncategorized. This makes it difficult to provide each customer with a prompt and personalized response.

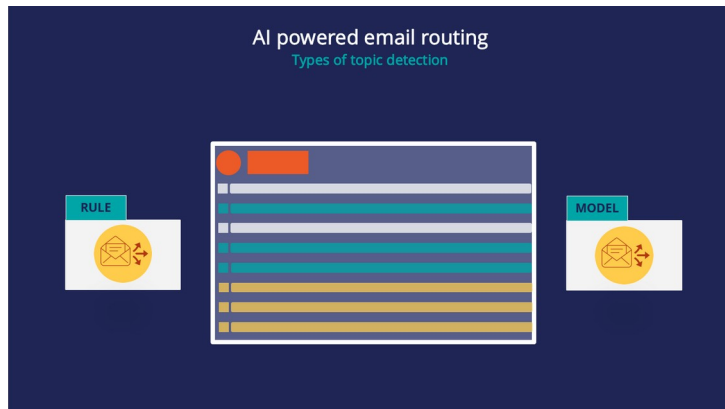


In Pega Infinity, this issue can be addressed by routing the emails using artificial intelligence. Pega Infinity uses AI-powered Natural Language Processing to detect the topic of an email and route the email to the appropriate container.

Consider emails from customers requesting an account address change, making a compliment, or requesting a new credit card. With the help of AI-powered text analytics, Pega Infinity is able to read and understand the content of each email and route it appropriately. This means customer service representatives can be alerted to any account-related service requests and resolve them quickly.

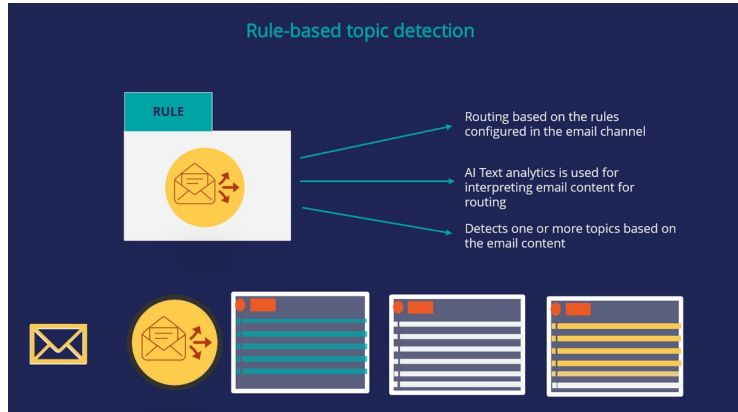


Email routing is done using the topic detection mechanism. The two types of topic detection are rule-based and model-based.



In rule-based topic detection the routing is based on the rules configured in the email channel. AI-powered text analytics is used to detect the topic of the email, and the channel rules route it to the right container. This type of topic detection may detect one or more topics if the email contains words associated with more than one topic.

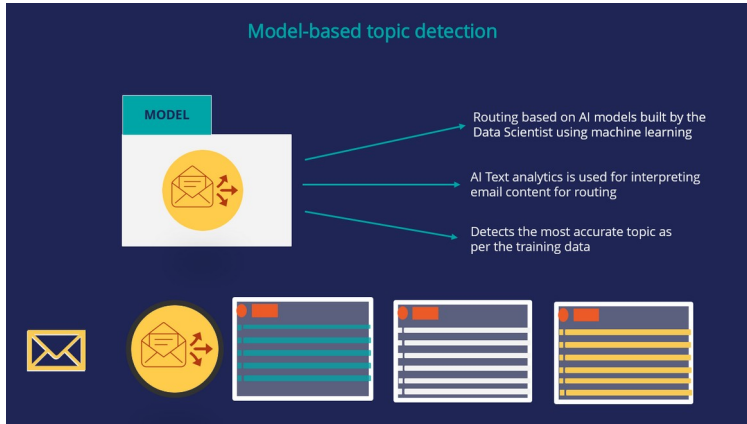
Let's consider a service request email from a customer. The email content is analyzed and routed to the right container. If an email from another customer contains words that are associated with two topics, the rule-based topic detection detects both topics. The email can then be routed to two different containers depending on how the channel rules are configured.



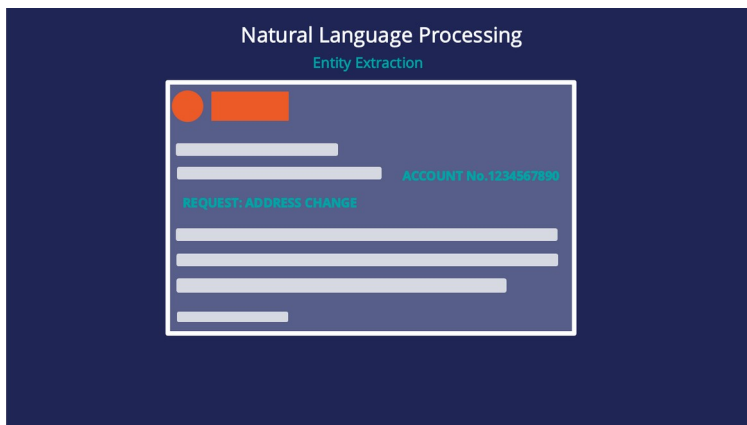
In model-based topic detection the routing is based on AI models built by a Data Scientist using machine learning. Building these models requires a training data set and a test data set. The data sets consist of a list of emails and the associated topic for each email. This type of topic detection identifies the most accurate topic based on the AI model and training set used by the Data Scientist.

Let's consider the same service request email from the customer. The email's content is analyzed and routed to the right container. If an email from another customer contains words associated with two topics during the training of the models, the model-based topic

detection detects both topics but typically with a different accuracy factor. In this case, the topic with the highest accuracy factor is chosen.



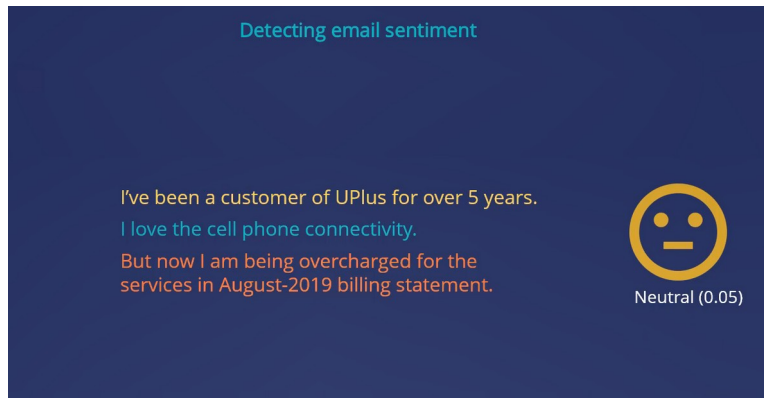
Pega Infinity also enables you to extract entities from an email. This means that when an email is sent, certain entities such as account number, email address, street address, etc. can be automatically detected and extracted. This allows certain emails to be automatically processed or given priority.



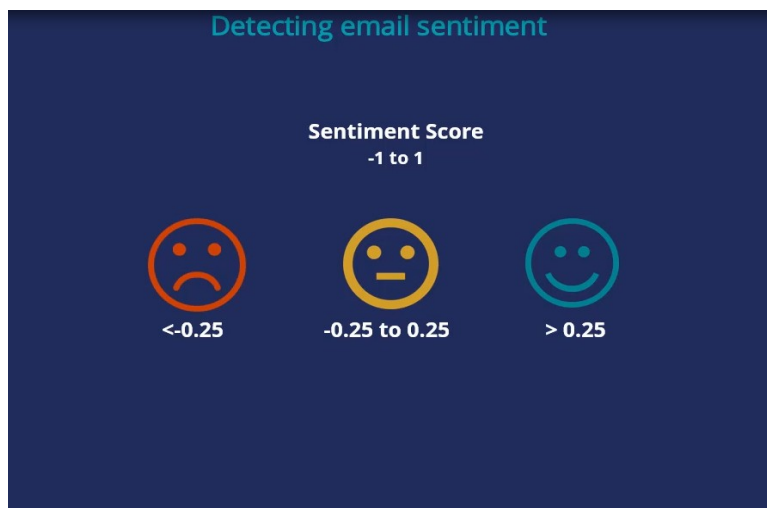
Besides Topic detection and Entity Extraction, Pega Infinity uses its AI-powered text analytics to enable you to detect the sentiment of an email based on its content.

Suppose a customer sends the following email to customer service.

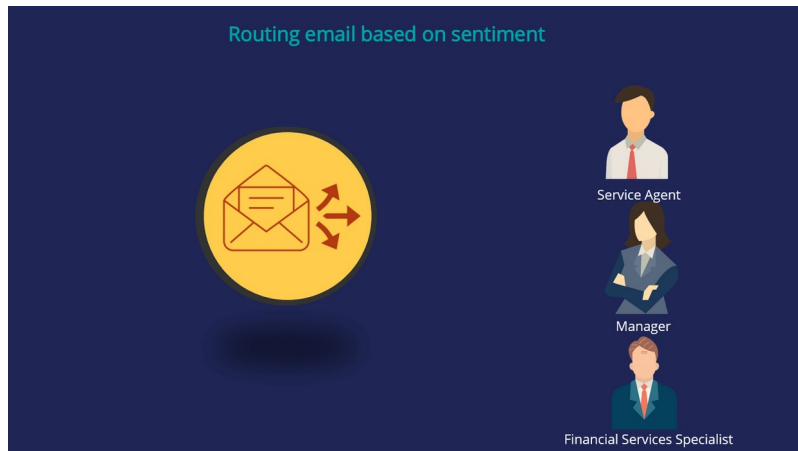
The email's content is a combination of different types of sentences. The first sentence has a neutral sentiment. In the second sentence, the customer expresses his satisfaction with the connectivity, so it has a positive sentiment. The third sentence is negative, as the customer is complaining about an overcharge. The overall sentiment of the email is determined based on the weight of the content sentiments. In this specific example the overall sentiment is neutral with a sentiment score of 0.05.



The sentiment score is a value between -1 and 1. In the out-of-the-box configuration, a sentiment score  $< -0.25$  results in a Negative sentiment, a sentiment score between  $-0.25$  and  $0.25$  results in a Neutral sentiment, and a sentiment score above  $0.25$  results in a positive sentiment.



Once the email sentiment is detected, you can configure the email channel to route a specific topic with a specific sentiment to a specialized agent for a quick and personalized response. For example, you could route an address change with a neutral sentiment to a Service Agent, a complaint email with a negative sentiment to a Manager, and a credit card inquiry with a negative sentiment to a Financial Services Specialist.



In summary, Pega Infinity's AI-based email routing capability enables customer service representatives to be more productive, reduces request processing time, and improves the customer experience by providing prompt and personalized service.



# Training a topic model to improve email routing

## Introduction

U+ Bank uses Pega Customer Service™ to route incoming emails to the appropriate department based on the topic of the email. For several use cases (for example, an address change), emails are routed based on keywords that are detected in the message. To improve the email routing, learn how to train the text prediction with a data set that contains classified messages.

## Transcript

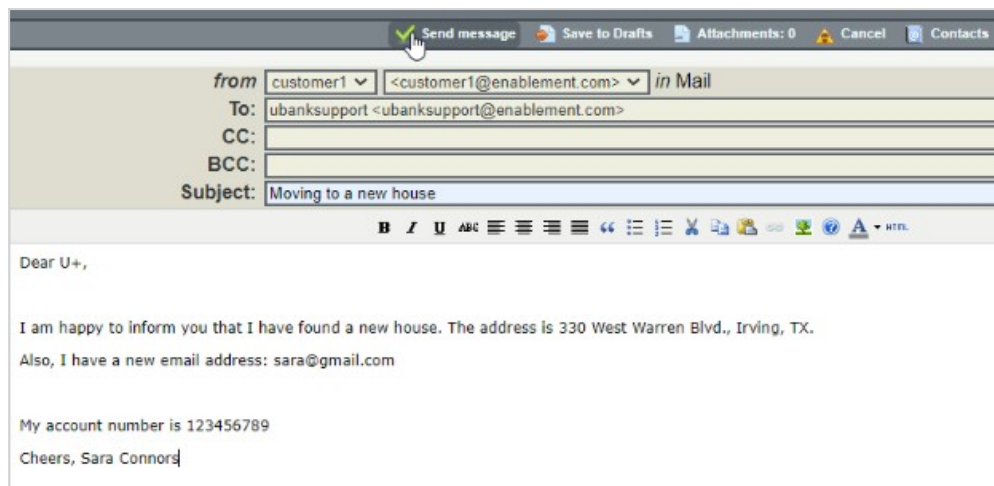
This demo shows you how to train a text prediction to improve email routing.

U+ Bank uses machine learning to route inbound messages in the email channel to the appropriate department based on the topic of the email.

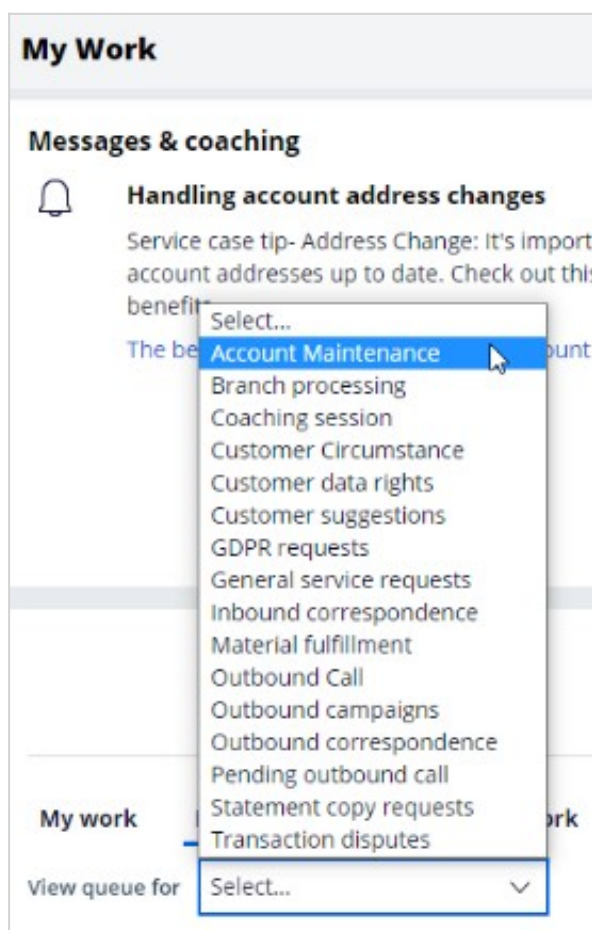
A text prediction that aims to detect the topic of the message drives the routing.



For example, when Sara writes an email to inform U+ Bank that she has moved to a new house, the text prediction detects an address change as the topic.

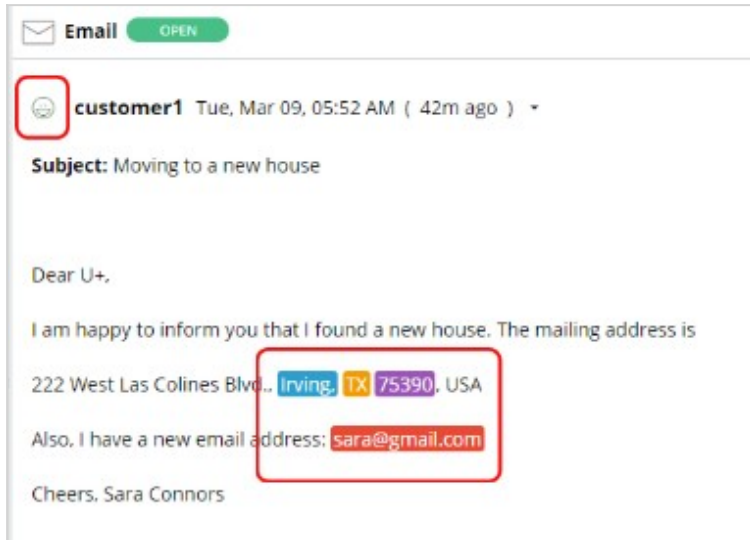


The **Account maintenance** department receives all emails where an address change is detected as the topic of the message.

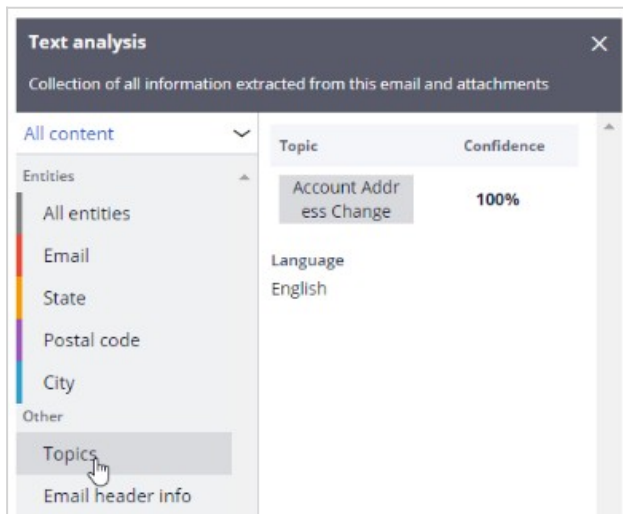


The text prediction also detects entities such as a ZIP Code and an email address and the overall sentiment of the message.

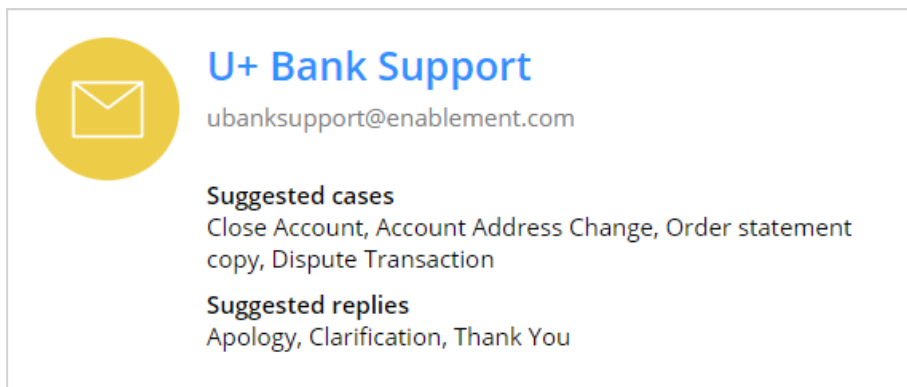




Notice that the topic Account Address Change is detected with 100% confidence.



The intelligent routing is set up in the email channel, in App Studio.



An email is routed to the work queue of the Account maintenance department when the detected topic is an address change.

Complaints are routed to the **Transaction Disputes** department.

If the address change and complaint topics are not detected, the email is routed to a default work queue.

Intelligent routing

Process:

Triggers only the first action that matches conditions.

Action:  Value:

When:

Action:  Value:

When:

+ Add condition

Otherwise:  Value:

Note that in the **When** condition of the email routing, you use the outcome of the text prediction.

Optionally, you can also use the sentiment of the email in the routing conditions.

Topic  Account Address Change

Header

NLP

Condition

Entities

Topic

Decisioning result

Sentiment

Language

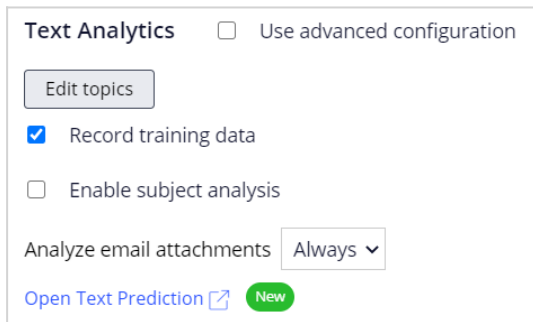
Value:

You can set a threshold for the confidence with which a topic must be detected to trigger routing.

Confidence threshold

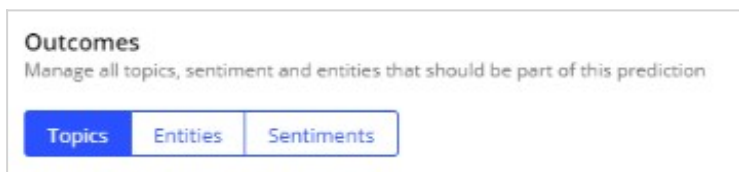
OK

Every channel is associated with a text prediction.



The screenshot shows a configuration panel for Text Analytics. At the top, it says "Text Analytics" followed by an unchecked checkbox for "Use advanced configuration". Below this is a button labeled "Edit topics". There are three checkboxes: "Record training data" (checked), "Enable subject analysis" (unchecked), and "Analyze email attachments" (unchecked) with a dropdown menu set to "Always". At the bottom, there is a link "Open Text Prediction" with an external link icon and a green "New" badge.

A text prediction is driven by predictive models that detect topics, entities, and sentiments.

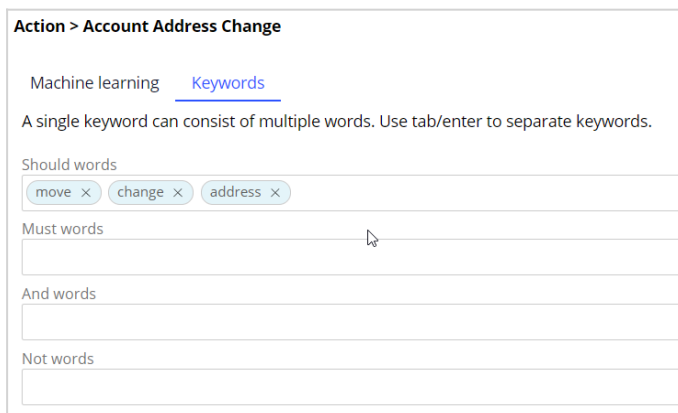


The screenshot shows an "Outcomes" management interface. The title is "Outcomes" with a subtitle "Manage all topics, sentiment and entities that should be part of this prediction". Below the subtitle are three tabs: "Topics" (selected), "Entities", and "Sentiments".

Many entity extraction models, and a sentiment model, are available out of the box.

For topic detection, Prediction Studio supports keyword models as well as machine learning models based primarily on training data.

A keyword model uses **Should words**, **Must words**, **And words**, and **Not words**.



The screenshot shows a configuration interface for keywords. The title is "Action > Account Address Change". Below the title, it says "Machine learning" and "Keywords" (underlined). A note states: "A single keyword can consist of multiple words. Use tab/enter to separate keywords." There are four input fields: "Should words" (containing "move", "change", and "address" with 'x' icons), "Must words", "And words", and "Not words".

If any of the **Should words** appear in a piece of text, topic detection assigns that text to the corresponding topic. To achieve accurate results, create an exhaustive list of **Should words**.

Only if all **Must words** appear in a piece of text, will the topic detection assign that text to the corresponding topic.

Use **And words** to distinguish between similar topics while using identical **Must words**.

If a **Not word** appears in a piece of text, the text is not assigned to the corresponding topic.

You can test the output of a prediction on a sample message.

Two topics are correctly detected: an address change and a complaint.

Notice that the confidence score for both topics is **1**. The keyword model performs a Boolean match based on the presence or absence of words and detects the topic with absolute certainty.

Topic detection					
Granularity			Analysis		
Document			Rule		
Topic	Sentiment	Sentiment score	Model name	Model type	Confidence score
Action > Account Address Change	Neutral	0	pyInteractionTaxonomy	Pega NLP	1.00

It is highly recommended to use a machine learning model for topic detection. It determines the confidence score based on evidence.

A machine learning model is based on training data that consists of categorized texts.

You can get training data from two sources. The first is training data accumulated through the application that is running in production.

When a customer service representative corrects the topic of an incoming email, this change is added to the training data.

After the messages are reviewed and approved, they are used to rebuild the topic model.

Also, you can choose to import data, provided you have accumulated data in the past.

The file must contain the message and the associated topic.

	A	B	C
1	<b>content</b>	<b>result</b>	<b>type</b>
20	Effective 11/28/17 please change mailing address to	Action > Account Address Change	Test
21	Please amend the mailing/billing address to the follow	Action > Account Address Change	Test
22	Confirm - mailing address: 1430 E 33rd St., Signal Hill	Action > Account Address Change	Test
23	Effective 8/28/17, please change the insured's mailin	Action > Account Address Change	Test
24	Per our insureds request, please issue the following cl	Action > Account Address Change	
25	Effective 8/25/17, please correct the mailing address	Action > Account Address Change	
26	Effective 8-16-17 please amend the mailing address a	Action > Account Address Change	
27	Please amend the mailing address.	Action > Account Address Change	
28	Effective 8/28/17 please change the mailing and billin	Action > Account Address Change	
29	AMEND insured's MAILING ADDRESS to read:	Action > Account Address Change	
30	Effective: 09/01/2017 Please amend mailing address	Action > Account Address Change	
31	Please update mailing address per the below:	Action > Account Address Change	
32	We have received a request to change your mailing a	Action > Account Address Change	
33	Please change mailing address to read: 197 Warren A	Action > Account Address Change	
34	Please change the mailing address as per the attache	Action > Account Address Change	
35	Please amend mailing & billing address to 3253 Phillip	Action > Account Address Change	

Notice that training data for the address change and the complaint topics are in the pending state.

You can now rebuild the models.

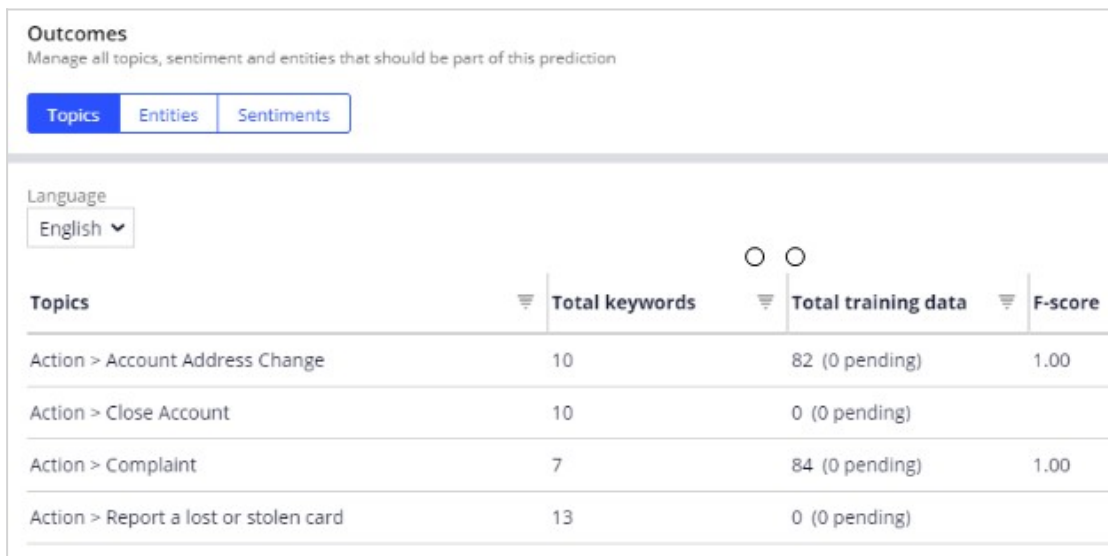
You can select individual models or rebuild all models across model types and languages.

While training is in progress, you have the option to cancel.

When the process completes, you can view the latest model report on the **Models** tab.

The report contains the validation data, the confusion matrix, and the score sheet.

The topic detection now uses machine learning models based on the training data and the keywords provided for the address change and complaint topics.



**Outcomes**  
Manage all topics, sentiment and entities that should be part of this prediction

Topics Entities Sentiments

Language  
English

Topics	Total keywords	Total training data	F-score
Action > Account Address Change	10	82 (0 pending)	1.00
Action > Close Account	10	0 (0 pending)	
Action > Complaint	7	84 (0 pending)	1.00
Action > Report a lost or stolen card	13	0 (0 pending)	

The **Should words** and **Must words** act as positive features for matching text to a topic, while the **Not words** act as negative features.

But the training and testing data have the greatest impact on your machine learning model, while keywords have a smaller impact.

Once the models are rebuilt, you can test the prediction.

Notice that the topics are detected with a confidence score below one.

In this case, the topic message is recognized as a complaint with high confidence. The address change topic is detected, but with a lower confidence score.

When multiple topics are detected in a message, comparison of the confidence scores allows selection of the one with the highest priority.

Topic	Sentiment	Sentiment score	Model name	Model type	Confidence score
Action > Complaint	Neutral	-0.25	U+ Bank customer support Pega NLP		0.75
Action > Account Address Change	Neutral	-0.25	U+ Bank customer support Pega NLP		0.25
Action > Report a lost or stolen card	Neutral	-0.25	U+ Bank customer support Pega NLP		0.00

The other outcomes of the text prediction are the entities detected, such as an email address and the sentiment of the message.

Output

I am happy to inform you that I found a new house. The address: 330 West Warren Blvd., Irving.  
 Also, I have a new email: sara@gmail.com. For the third time, I have noticed an invalid transaction.  
 The amount is small, but could you please check?

Sentiment indication

Positive                      Negative                      Neutral

You have reached the end of this demo. What did it show you?

- How Pega Customer Service routes incoming emails to the appropriate department based on the topic of the email.
- How text predictions work to predict the topic and sentiment of a message and detect entities.
- How to train a topic model and use machine learning to identify the topic correctly.

# Using entity extraction with chatbot channel

## Description

Better understand the key features and benefits of entity extraction with chatbot channel. Use this module to learn how to enable the chatbot to handle ticket cancellation requests by training a text prediction that drives the chatbot. The chatbot detects the topic of a message, extracts all relevant entities, and creates a case that the Customer Service Representative (CSR) in the CSR portal can handle later.

## Learning objectives

- Describe how Pega Chatbot uses natural language processing to determine the topic of inbound message and extract entities
- Configure the chatbot channel
- Explain the out-of-the-box entity extraction models
- Train the topic detection and entity extraction models

# Training predictions for chatbot

## Introduction

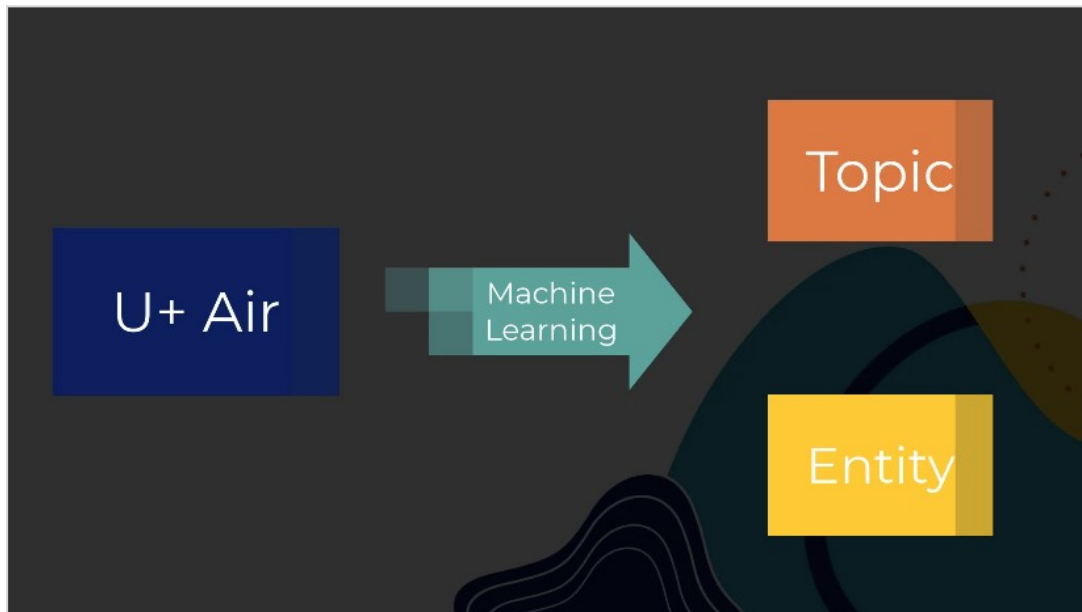
Web Chatbot is a feature in Pega Customer Service™ that allows your customers to enter questions or requests in a chat window that is accessible from a customer-facing web page, and then receive answers and guidance.

In a web chatbot session, the chatbot responds to a customer request, such as *I want to cancel a ticket*, by detecting the right topic with machine learning. The chatbot then walks the customer through a set of questions. The chatbot gathers the required information by leveraging entity extraction models. Learn how to train these topic detection and entity extraction models.

## Transcript

This demo shows you how to train a text prediction to enable the chatbot to handle ticket cancellation requests.

U+ Air uses machine learning to identify the topic of the incoming message and detect entities such as ticket numbers.

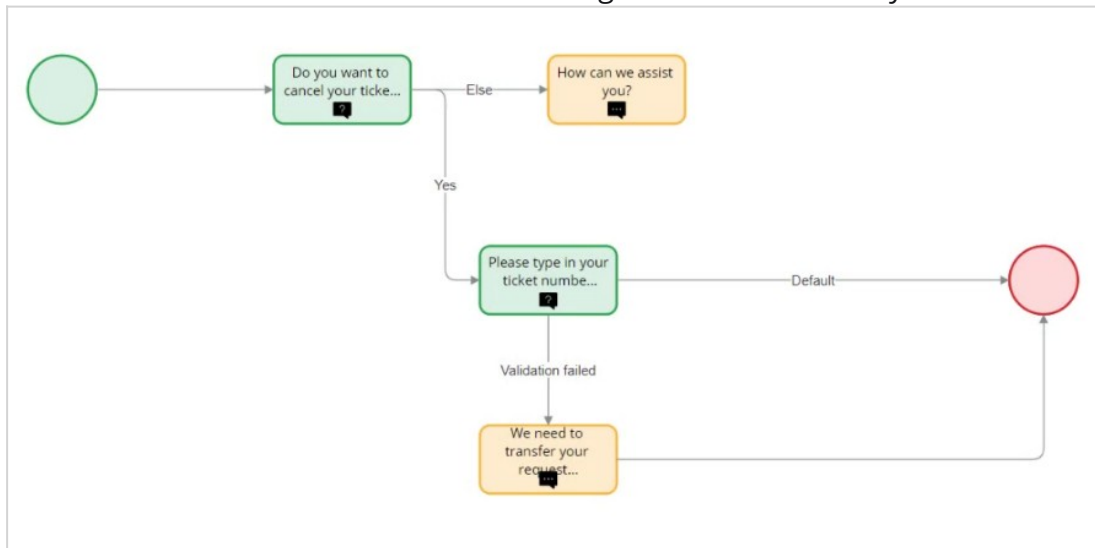


A text prediction that aims to detect the topic and entities included in the message drives the chatbot.

For example, when Troy, a customer, sends a message to the Airline chatbot that he wants to cancel his ticket, the text prediction detects the topic of the message and runs the *Cancel a ticket* case type that an application developer preconfigured in the system. The case type



contains the conversation flow that the chatbot uses in conversation with Troy. Notice the conversation elements such as *Ask a question* or *Send a message*. Observe how the chatbot follows the conversation flow during interaction with Troy.



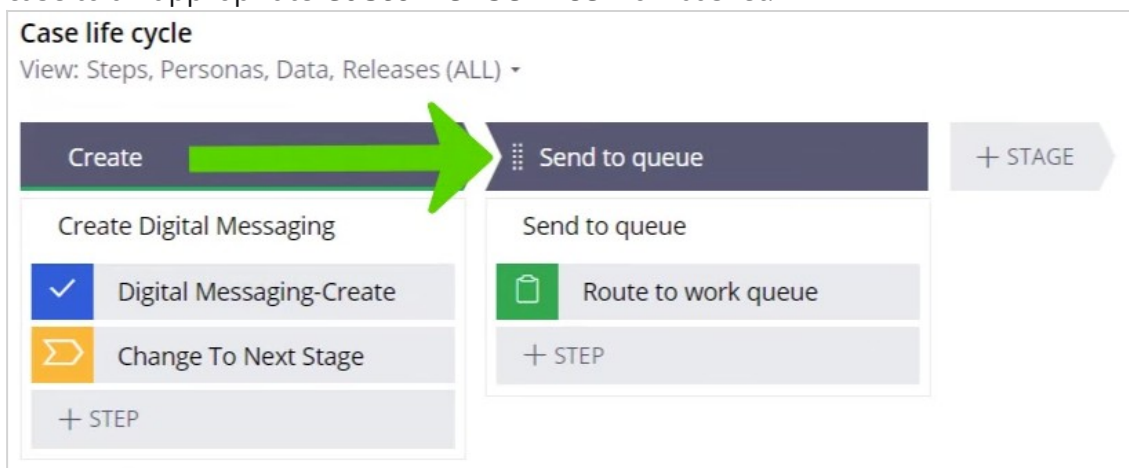
The prediction detects the ticket number entity that Troy provides, and then asks for confirmation.

Notice that the chatbot skips the *Please type in your ticket number* question because the text prediction detected that Troy provided the ticket number in the first message.

See how the chatbot reacts if Troy does not provide the ticket number in the first message.

Instead of skipping the question, the chatbot asks Troy to provide the ticket number.

Once the conversation ends, the chatbot saves the necessary information and routes the case to an appropriate **Customer Service** workbasket.



Note that the case contains the ticket number extracted by the text prediction during the conversation with Troy.

To make the chatbot available to customers, you first configure the digital messaging channel in App Studio.

Notice that the *Cancel a ticket* case type is specified. This is the same case type that contains the conversation flow that the chatbot uses in the conversation with Troy. The **Create case command** represents the topic in the text prediction. The chatbot triggers the case type creation when the text prediction detects this specific topic.

The screenshot shows the 'Create command' dialog box with the following configuration:

- Case type: Cancel a ticket
- Create case command: cancel ticket
- Can be suggested by: Chatbot to user OR AI to agent
- Requires authentication:

The topic detection and entities extraction are driven by a text prediction that is automatically generated after chatbot channel creation. Data scientist builds the models in the prediction in Prediction Studio. Models learn with training data. In this scenario, the training data contains three topics: *action book ticket*, *action cancel ticket*, and *action reschedule*. Besides topic detection models, the prediction contains entity extraction models that extract necessary information from the incoming chat message.

Notice the existing *pySystemEntities* model. It is an out-of-the-box entity extraction model that contains basic entities; you cannot modify it. To add a new entity type such as ticket number, you create a new entity extraction model.

Entity Name	Type	Included languages	Data	Status	
pySystemEntities	Pega-NLP	English	ALL <a href="#">↗</a>	ACTIVE	<a href="#">+</a> Add model <a href="#">🗑️</a>

After creation of the new model, you can create a new entity.

U+ Air determines that using Rule-based Text Annotation (Ruta) scripts is the best choice to detect the strict ticket number pattern. Ruta is a rule-based script language which detects keywords and phrases that follow specific patterns.

Notice the rulescript that detects the ticket number. This script detects ticket numbers that follow a strict pattern of two letters followed by three digits. In case of a rescheduled ticket or human error, the application also needs to detect unusual ticket number patterns. For example, a dash in between the letters and numbers, or three letters and three digits. Using Ruta with machine learning can achieve this business outcome.

**New entity**

Custom Pick from list

Entity name Entity model  
ticket\_number Airline\_entities

Machine learning **RUTA** Keywords

Enable RUTA

Template  
Select

Rulescript

```
1 CAP{REGEXP("..\")}NUM{REGEXP("...")} -> MARK(EntityType,1,3);
```

Cancel Save

A machine learning-based model must be built to enable machine learning entity extraction. You add training data to the ticket number entity to build the model; you can add training data one record at a time.

A training data record is a typical chat message that the customer sends. After adding a training data record, you manually identify the entities you want the model to recognize. In this case, JK-294 is a ticket number entity. Optionally, to improve the topic detection model, you can specify the topic of the message that provides additional training data.

You can also use a batch import from a dataset file to provide additional training data. Notice the tagging that identifies a specific entity in each record. It is possible to have

multiple tags to train multiple entity types simultaneously.

**ticket\_number**

Machine learning RUTA Keywords

Showing data for entity model : Airline\_entities

+ Add training data ? Upload

Pending training data	Review status	
I want to cancel my flight for confirmation num...	REVIEWED	
Please cancel my ticket. Reservation number is ...	REVIEWED	
Please cancel my flight. Ticket number is SPM372	REVIEWED	
I need to cancel my reservation flight from New...	REVIEWED	

Page 1 of 7

I want to cancel my flight for confirmation number **BER215**

**Topic**  
Action > cancel ticket

**Sentiment**

**Entities**  
ticket\_number

After uploading the training data file, you can review each record and ensure that it was correctly imported. The correct ticket number entity and topic are flagged.

After uploading the dataset, build the models. Notice the New F-score for the Airline and Airline\_entities models. Now, the entity extraction model can also detect unusual ticket number patterns.

**Model training report** ✕

Date  
3/24/22 9:20 AM

Language	Model name	Outcomes	Status	Previous F-score	New F-score
English	Airline	Topic	Completed	0.8	1 ↑
English	Airline_entities	Entity	Completed	0	1 ↑

Close

To enable the chatbot to use the extracted entity in the created case, an application developer maps the entity to a case property in a case command.

The system uses this property in the *Cancel a ticket* case type.

The screenshot shows a 'Response configuration' dialog box with three tabs: 'Response', 'Text analysis', and 'Entities extraction'. The 'Entities extraction' tab is active. Below the tabs, there is a heading 'Map entities types to **Cancel a ticket** case model.' Below this heading, there are two columns: 'Entity' and 'Case property'. Under 'Entity', there is a dropdown menu with 'ticket\_number' selected. Under 'Case property', there is a dropdown menu with 'ticket\_number' selected. To the right of the 'Case property' dropdown is a trash icon. Below the dropdowns, there is a '+ Add mapping' link.

With all models in the text prediction built, and all necessary properties mapped, it is possible to see in full detail how Pega natural language processing (NLP) works. The digital messaging interface has a built-in preview console that simulates the chatbot. After sending the *I want to cancel my ticket number AA562* message, you can turn on the **Show Analysis** switch to see the chatbot details. In this case, the chatbot detects the topic of the message with 94% confidence. The confidence score, which is a value from 1 to 100, represents the confidence of the prediction. The chatbot detects the ticket number entity in the first message, skips the ticket number question, and then routes the case to a correct

workbasket.

**Preview console**


Show analysis Re

---

Language	
English	
Entity types	
<b>Confirmation</b>	
Value	Type
Yes	User input


9:59 AM

Please type in your ticket number.

**Ask (skipped): ticket\_number** 

Value	Type
<b>AA562</b>	ticket_number

9:59 AM

 Thank you! The next step in this case has been routed appropriately.

The case that the chatbot creates is available for review by the customer service representative in the CSR portal. Notice that because of the detected topic, the system creates a *Cancel a ticket case*. The case contains the ticket number extracted by the text

prediction during the conversation with the customer.

<b>Create</b> (CA-3)
Label
<input type="text" value="Cancel a ticket"/>
ticket_number *
<input type="text" value="AA562"/>
City
<input type="text"/>

You have reached the end of this video. What did it show you?

- How to create a chatbot channel.
- How to create a topic detection model.
- How to create an entity extraction model with machine learning and Ruta script.