



Pega NLP Essentials

STUDENT GUIDE

© Copyright 2022
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

Mission: Pega NLP Essentials

Products: Pega Platform™ 8.7, Pega Customer Decision Hub™, Pega Customer Service™

URL: <https://academy.pegasystems.com/mission/pega-nlp-essentials/v1>

Date: 10 May 2022

Contents

Pega AI overview.....	4
Predictive models drive predictions.....	5
Prediction Studio.....	9
Text analytics for email routing.....	18
Text analytics.....	19
Email routing.....	24
Training a topic model to improve email routing.....	30
Using entity extraction with chatbot channel.....	38
Training predictions for chatbot.....	39

Pega AI overview

Description

Prediction Studio is the dedicated workspace for data scientists to control the life cycles of predictions and the predictive models that drive them. Prediction Studio offers prediction and model reports that allow users to monitor and spot predictions and models that underperform.

Learning objectives

- Describe how predictive models drive case management predictions, Pega Customer Decision Hub™ predictions, and text analytics predictions
- Describe the purpose of the control group in Customer Decision Hub predictions
- Describe the purpose of the work areas in Prediction Studio
- Recognize the transparency settings for predictive models

Predictive models drive predictions

With the decision management capability of Pega Platform™, you can enhance applications to help optimize business processes, predict customer behavior, analyze natural language, and make informed decisions to better meet customers' needs and to achieve positive business outcomes.

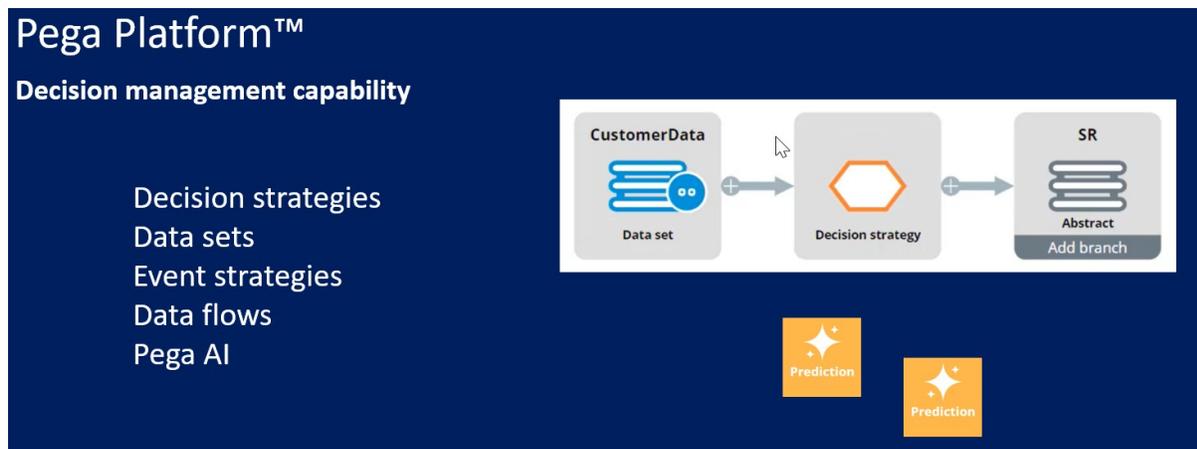
Transcript

This video introduces you to Pega AI, a feature of the decision management capability of Pega Platform™.

Other decisioning features of the Pega Platform include:

- Decision strategies to improve customer experience and deploy intelligent processes based on behavioral and operational data and data sets to read and write the data used in the decision strategies.
- You can use event strategies to detect patterns in data streams and react to them.
- And to ingest, process, and move data from one or more sources to one or more destinations, you can configure data flows as scalable and resilient data pipelines.

Decision management uses Pega AI to make predictions about customer behavior, successful case completion, the topic of an incoming message, or other subjects to make the decisions more relevant.



Decision management is a Pega Platform capability. You can apply decision management to any application that is built on Pega Platform.

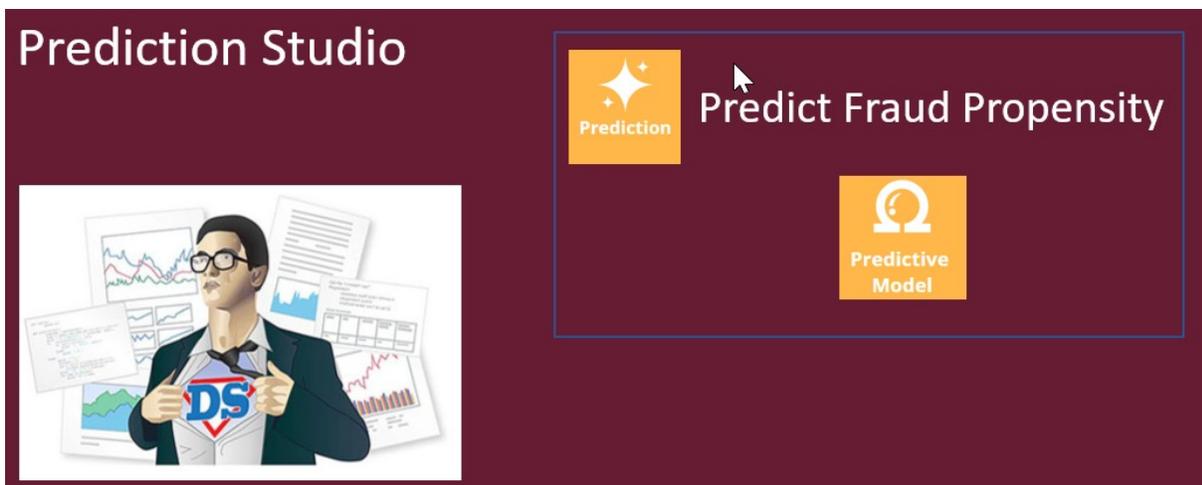
Predictions differ to suit the domain they are used in, but one or more predictive models drive them all.

A data scientist can create a predictive model in Pega Platform or an external environment that can export the model as a PMML or H2O file. Another option is to connect to a machine learning service such as Google ML or AWS SageMaker.



If an insurance company wants to use Pega Process AI™ to route incoming claims that might be fraudulent to an expert based on the outcome of a predictive model ...

... the data scientist creates a fraud model to drive a new case management prediction in Prediction Studio.



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

A prediction is a hand-off to an application developer, who can then use the prediction in a decision step in the case type to route cases more accurately. This strengthens the separation of concerns.

You can use Pega Customer Decision Hub™ to make next-best-action decisions for your customers.

Customer Decision Hub predictions can predict customer behavior, such as which customer is about to churn ...

... or predict the likelihood that a customer clicks on a web banner to support the decision on which banner to show to a customer.

Pega Adaptive Decision Manager (ADM), a key component of the decision management capability ...

... allows a data scientist to configure self-learning, adaptive models that continuously improve predictions about business processes and customer behavior.

An adaptive model rule typically represents many adaptive model instances because each unique combination of the model context generates a model.

In Customer Decision Hub, adaptive models drive many predictions that come with the product out of the box, such as the Predict Web Propensity prediction that predicts the likelihood that a customer clicks a web banner.

Customer Decision Hub predictions have several features specific for the domain ...

... such as a control group for which the prediction outputs a random propensity instead of the propensity that is generated by the adaptive model instance.

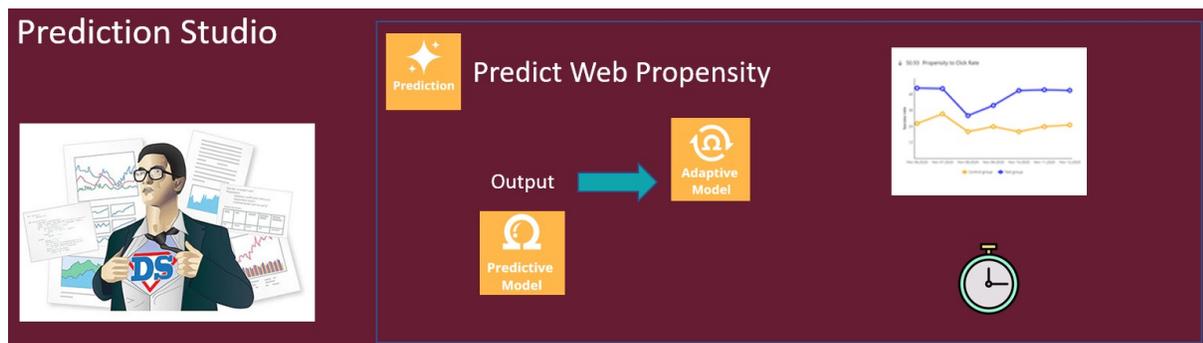
Comparison of the control group and the model propensity-based group allows you to measure the lift in a success rate that the AI generates, an important business metric.

Also, Customer Decision Hub predictions feature a response timeout setting. After the timeout expires, a negative response is recorded.

The response timeout setting depends on the use case. For example, in a web use case, several minutes suffice ...

... while in an outbound email campaign, the response timeout is set to several days to allow customers enough time to respond.

You can further enhance the prediction by using the output of a predictive model as a predictor in the adaptive model.



The Pega Customer Service™ application uses the natural language processing capability of decision management to analyze incoming text and route the messages based on the topics and entities detected.

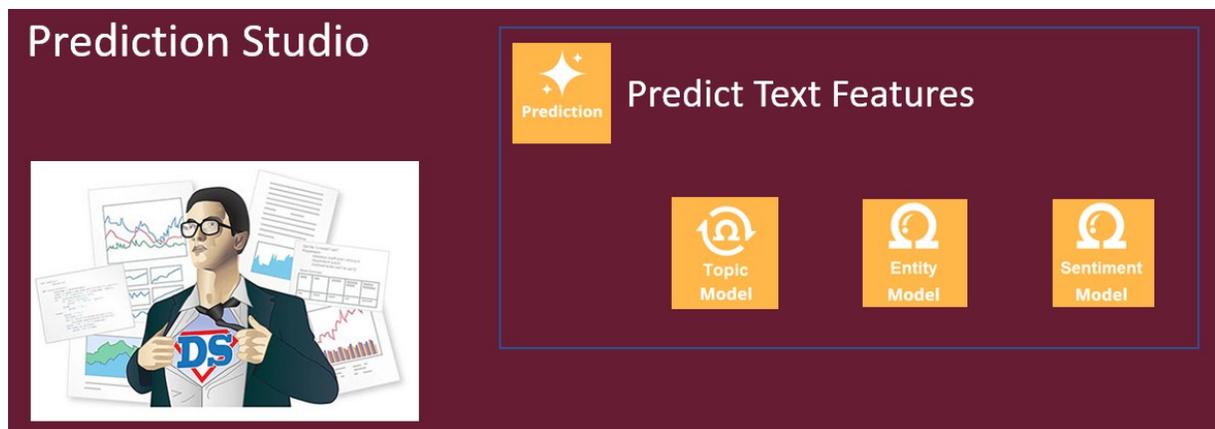
Pega Customer Service uses Text analytics predictions that are distinctly different from both case management predictions and Customer Decision Hub predictions.

Text analytics predictions use predictive models to detect the topic of an incoming message that the application can use to optimize the routing of the message to the relevant department.

Secondly, text analytics predictions use entity extraction models that qualify text as, for example, an account number, a postal ZIP code, or an address.

The application can use this information to fill relevant fields in a case automatically.

Finally, the text analytics predictions come with a sentiment model that can route or prioritize negative messages to improve the customer experience.



Feedback on the detected topics, entities, and sentiment by CSRs improves the performance of the text analytics prediction over time.

This video has concluded. What did it show you?

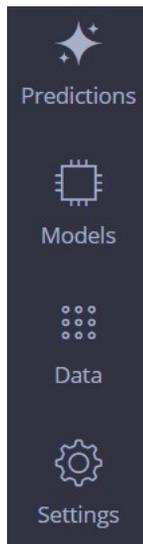
- Pega AI allows you to improve business processes and customer engagement by using predictions.
- Predictive models drive the predictions.
- The predictive models can be static or adaptive.
- Predictions are managed in Prediction Studio.

Prediction Studio

Predictions and the predictive models that drive them are created, monitored and updated in Prediction Studio, the dedicated workspace for data scientists.

Transcript

This video gives you an overview of the features of Prediction Studio. The workspace provides data scientists with everything they need to author, deploy, govern, monitor, and change predictions. Prediction Studio has four work areas: Predictions, Models, Data, and Settings.



The Predictions landing page is used to create and manage predictions. Predictions can be one of three types.

Create a prediction

Where will you be using the prediction?

- Customer Decision Hub
Optimize the engagement with your customers
- Case management
Use predictions to improve the automation in cases
- Text analytics
Analyze the text that comes through your channels

Customer Decision Hub predictions are used in the Pega Customer Decision Hub™ application to optimize 1:1 customer engagement. **Case management predictions** are used in case types to support decisions in business processes and **Text analytics**

predictions are used in the Pega Customer Service™ application to predict the topic of incoming messages. The three types of predictions differ to suit the domain they are used in, but one or more predictive models drive them all.

The Model landing page is used to create and manage the predictive models. There are four types of predictive models.

Static **predictive** models are built on historical data. A data scientist can create a predictive model in an external tool and import the model file. Another option is to connect to a machine learning service, such as Google ML or Amazon SageMaker.



Predictive model

Adaptive models continuously learn from responses and adapt to changes over time. You can configure an adaptive model rule that typically represents many adaptive model instances, because each unique combination of the model context will generate a model.



Adaptive model

Text categorization models can detect the topic of a message and the sentiment of the author. **Text extraction** models identify entities such as an email address, an account number, or a city.



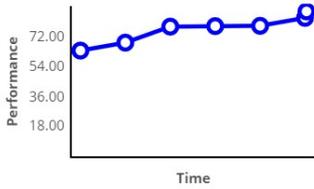
Text categorization Text extraction

Many widely used Customer Decision Hub predictions ship with the product.



One of these is the **Predict web propensity** prediction, which predicts the likelihood that a customer clicks a web banner.

Predict Web Propensity 87.06 (AUC)



Consider, for example, the cross-sell in a web scenario for U+ Bank. The bank shows a personalized credit card offer to eligible customers when they log in to the bank’s website.



Standard card

0% APR and no annual fee

[Learn more](#)

When a customer is eligible for multiple credit cards, the prediction calculates the propensity of receiving a positive response from the customer for each card. Customer Decision Hub decides which credit card to offer based on business rules, interaction context, and predictions.

The adaptive model that drives the **Predict Web Propensity** prediction is the **Web Click Through Rate** model.

Supporting models

Name	Component name	Type	Performance	Status
Web_Click_Through_Rate	Web_Click_Through_Rate_Customers	Adaptive model	68.55 AUC	ACTIVE
Web_Click_Through_Rate	Web_Click_Through_Rate_Accounts	Adaptive model	68.55 AUC	ACTIVE

You can configure several aspects of a Customer Decision Hub prediction. A control group is a small group of customers who receive random offers, as opposed to the target group.

Control group

The control group is used to measure lift by comparing the success rate in the target group with the control group. Customers in the control group will receive an action determined by a random propensity.

Percentage Field

Percentage
 %

Customers in the target group receive the offers that they are most likely to accept, based on the propensity that the prediction calculates for each customer.

The purpose of the control group is to calculate lift by comparing the success rate in the control group with the success rate in the target group.

The random offers also allow predictive models to continuously explore all actions.

Based on the lift, you can determine the effectiveness of your prediction, for example, in increasing conversion rates. The control group is typically defined as 2% of all customers, but this can be changed.

The response labels represent the possible outcomes of a prediction. The propensity is computed based on the number of outcomes registered under the target label versus the alternative label.

For example, in **Predict Web Propensity**, because you want to predict the likelihood of a customer clicking a banner, the Target label (which in this case represents the positive outcome) is **Clicked**. The alternative label that represents a negative outcome in this case, is **NoResponse**.

Response labels

Labels for the possible values of the responses.

Propensity to Click

Target label Alternative label

Clicked NoResponse

The **NoResponse** response can be captured on request or automatically depending on the response time-out setting. The response time-out defines how long to wait for the customer to respond to your offer. After the specified amount of time elapses, the system automatically records the alternative outcome for the interaction.

🕒 Response timeout

You can choose how long you want to wait for a response. If this period elapses, the alternative label will be recorded.

Propensity to Click

Indefinitely Fixed time frame

Amount

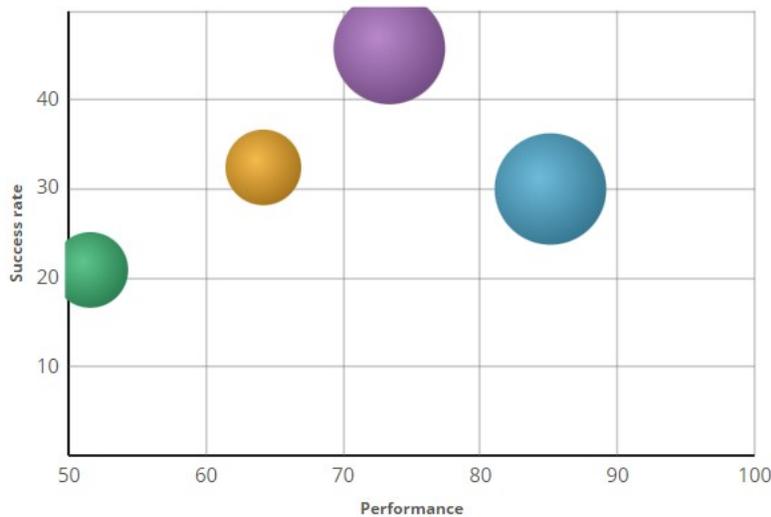
10

Time unit

minutes ▼

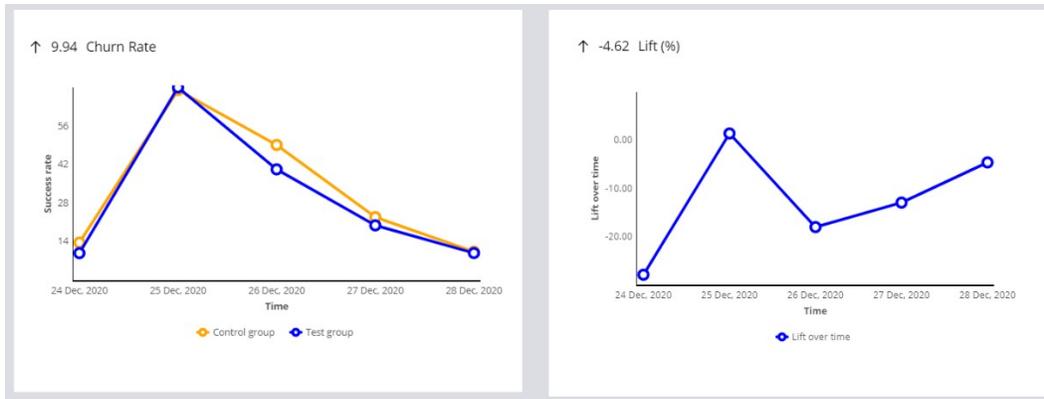
This setting depends on the use case. When predicting a click on a web banner, you typically set it to 30 minutes or less, but in an outbound email offer, a waiting time of several days is more appropriate.

The **Web Click Through Rate** model rule is the model template for each of the credit card offers. You can monitor the performance of the models in a diagram that shows the success rate versus model performance.



The models are represented by colored circles. The size of the circles indicates the number of responses captured by the model.

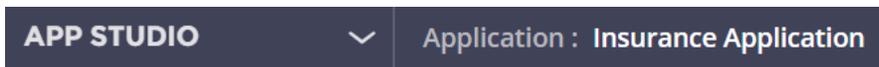
Monitor the prediction over time to analyze how successful it is. The available metrics are the success rate, the lift calculated using the control group, the prediction performance, and the total number of cases.



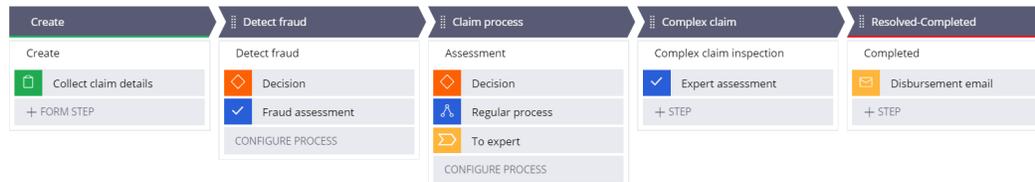
Prediction Studio generates actionable insights and notifies the user when predictions and predictive models show unexpected behavior (for example, a significant drop in success rate).



Case management predictions support decisions in a case type.



Consider this case type, which handles incoming car insurance claims.



An application developer can use the outcome of the prediction in the condition of a decision step, instead of a business rule. Based on the condition, a case is routed to a fraud expert when the prediction flags the claim as abnormal.

When

Custom condition

(Probability is equal to "abnormal")

Go to

Fraud assessment

Add path

Otherwise go to

[End]

The Pega Customer Service application can use text predictions to analyze messages that come in through various channels, such as email and chat channels.

PREDICTION STUDIO ▾

Application: Customer Service

A text prediction is automatically generated for each new channel.

A text prediction detects topics, entities, and sentiment, to improve the routing of messages to the appropriate department.

Outcomes

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

Topics

Entities

Sentiments

Topics can be detected based on keywords or machine learning. In this example, topic detection is keyword based. It is highly recommended to include machine learning in topic detection.

Consider the following message, about an address change:

Hi U+ Bank,

I have noticed, in my last account statement, you have used a wrong address. Please change my mailing address to read: 222 West Las Colinas Blvd., Irving, TX 75039, USA, effective immediately. And I'm happy to have a fresh email address: sara@gmail.com.

Cheers, Sara Connors

Based on keywords, two topics are detected. Both topics have a confidence score of 1, so it is not possible to determine the correct topic.

Topic



Action > Complaint

Action > Account Address Change

To train the topic model, use a data set with records that contain a message and the associated topic. When the trained model is tested with the same message, the model correctly generates the highest confidence score for the address change topic.

	Sentiment	Sentiment score	Model name	Model type	Confidence
Account Address Change	Positive	0.41	U+ Bank customer support Pega NLP		0.71
Complaint	Positive	0.41	U+ Bank customer support Pega NLP		0.29

The sentiment model is shipped with the product and predicts an overall positive sentiment.

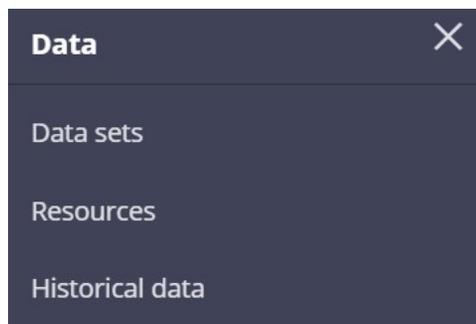
Output

Hi U+ Bank, I have noticed, in my last account statement, you have used a wrong address. Please change my mailing address to read: 222 West Las Colinas Blvd., Irving, TX 75039, USA, effective immediately. And I'm happy to have a fresh email address: sara@gmail.com. Cheers, Sara Connors

Entity extraction can be based on keywords, machine learning, or RUTA scripts.

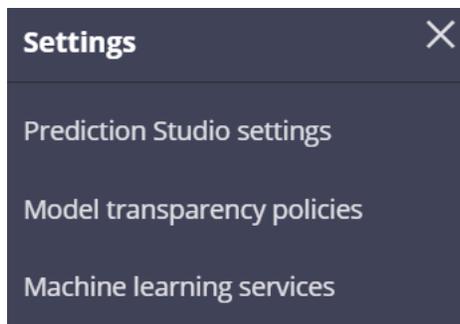
Value	Name	Type	Method	Resolved value
Irving,	City	City	Machine learning	Irving,
sara@gmail.com	Email	Email	RUTA	sara@gmail.com
TX	State	State	Machine learning	TX
75039	Zipcode	ZipCode	Machine learning	75039

The **Data** work area is used to define data sets, resources, and historical data.



A data set instance can be sourced from a database table, from stream services, or even social media, such as Twitter and YouTube. Resources include taxonomies and the default sentiment lexicon to use in building machine learning models. When enabled, historical data used for the training of adaptive models and monitoring of predictive models is recorded for offline analysis.

In the **Settings** work area, you can manage general Prediction Studio settings and connect to third-party machine learning platforms.



Also, you can review company policies regarding the transparency thresholds for different business issues. In risk management, decisions must be explainable. In marketing, more accurate models may be allowed at the expense of transparency.

Each model type is assigned a transparency score ranging from 1 to 5, where 1 means that the model is opaque, and 5 means that the model is transparent. Depending on the threshold setting, some types of models can be non-compliant for a specific business issue.

Adaptive model Pega 3 Compliance All business issues	Bivariate model Pega 3 Compliance All business issues	Genetic algorithm Pega 2 Compliance All business issues
Clustering model 3 Compliance All business issues	Ensemble model 1 Compliance All business issues	General regression 4 Compliance All business issues
Neural network 1 Compliance All business issues	Random forest 1 Compliance All business issues	Scorecard 5 Compliance All business issues

This demo has concluded. What did it show you?

- How to create and manage Customer Decision Hub predictions, case management predictions and text analytics prediction.
- How to create and manage the predictive models that drive the predictions.
- How to inspect the model transparency settings of the business.

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

Text analytics

Introduction

A company wants all incoming emails to be automatically routed to the correct department. The company must find a way to process the text and extract the information for which it is looking. Text analytics is the most efficient and effective way to obtain information.

Transcript

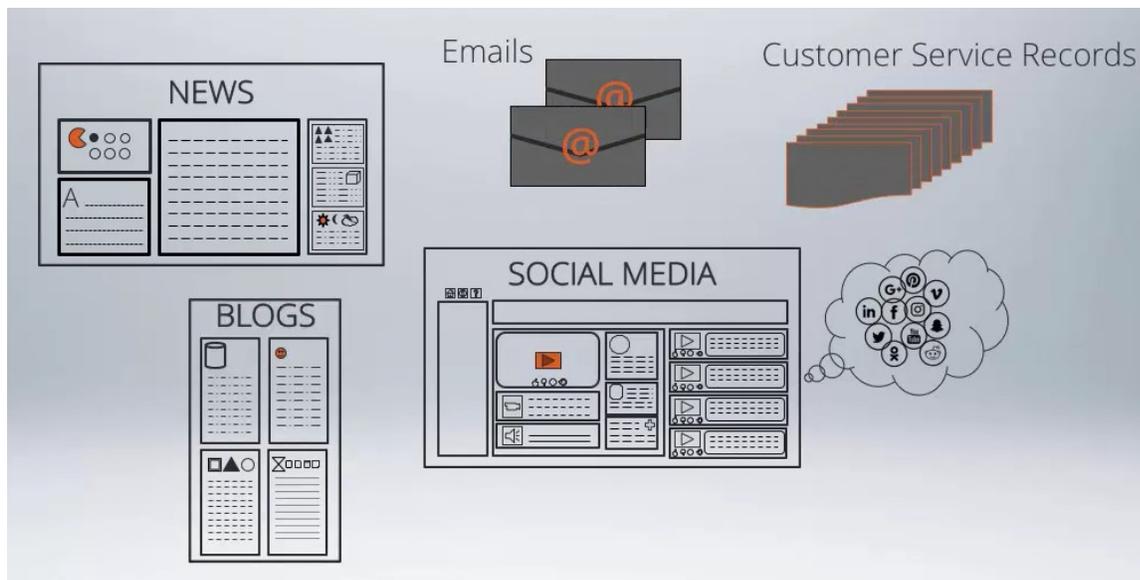
In this lesson you are going to learn what text analytics is and how it is used to help understand customer sentiments and needs.

Consider this message from a customer.

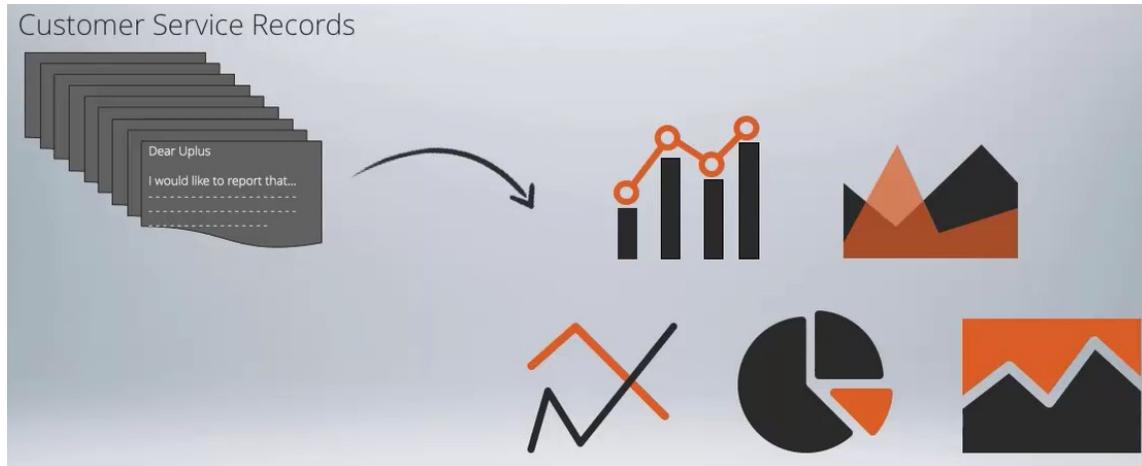
I've been a customer of UPlus for over 5 years. I love the cell phone connectivity. But now I am being overcharged for the services in the August-2017 billing statement.

Humans have the ability to understand the implications of a piece of text. We can summarize it or elaborate on it or even re-phrase it and still preserve the original meaning. In fact, we do this very well. But can we do it on a large scale?

Today, digital sources such as news, blogs, social media and emails are generating more text than ever before.

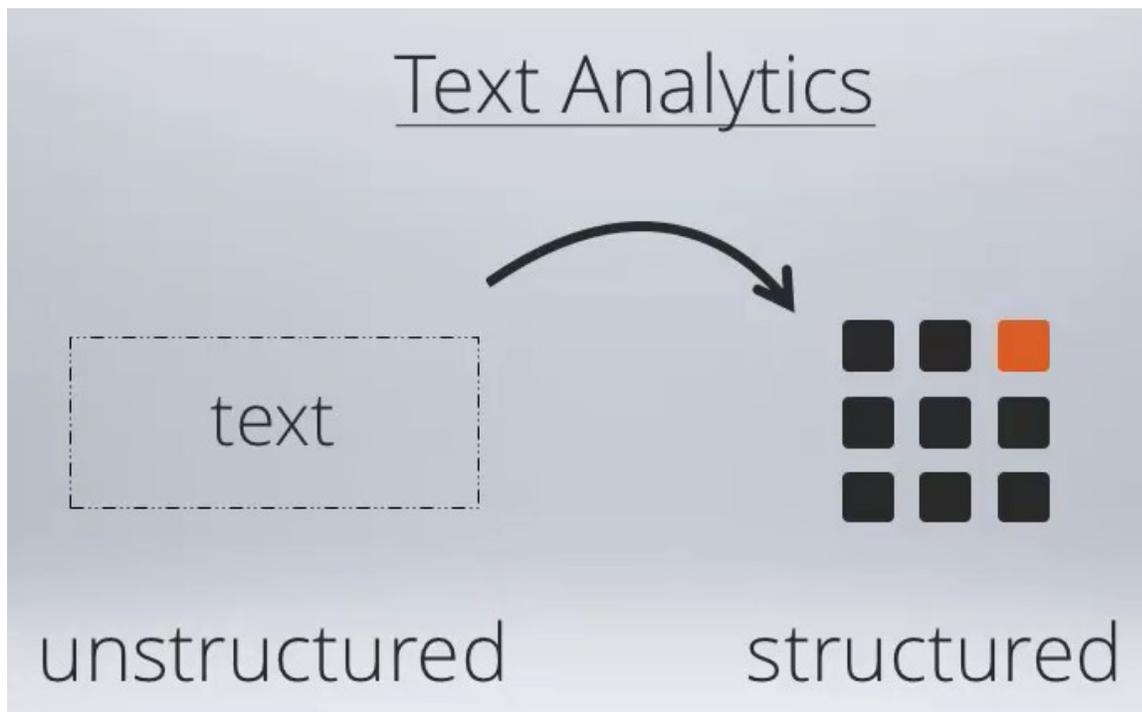


It's not possible for humans alone to analyze and extract valuable information from such a large volume of text. Consider a company that has tens of thousands of customer service records in text form.

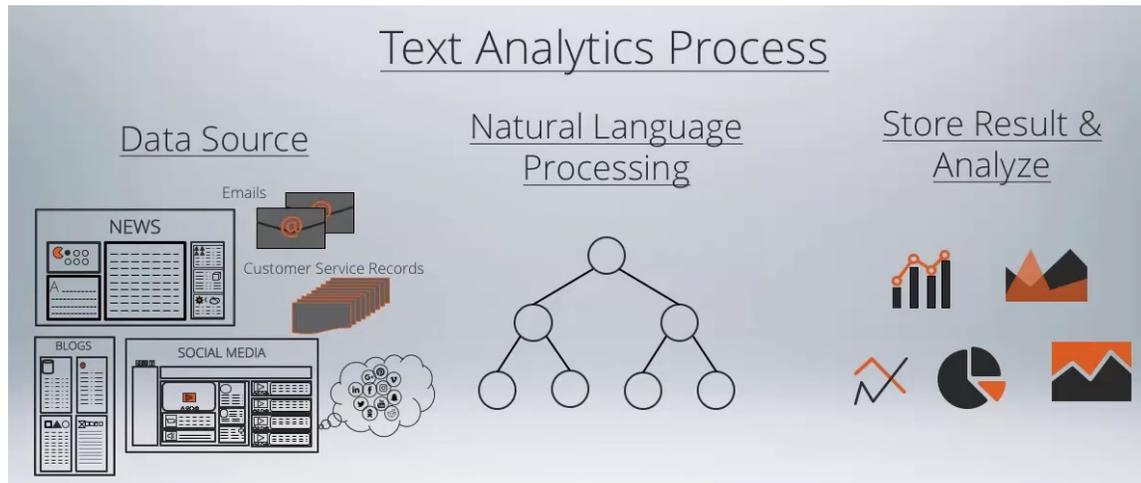


Let's say the company wants all of the complaints to be automatically routed to the right department. Then it wants a report that breaks down the complaints by department. The company also wants to use these records to understand how satisfied its customers are about various products and services. The company must find a way to process the text so that the information it's looking for can be extracted. Text analytics is the most efficient and effective way to do this.

Text Analytics is the process of deriving high-value information from text.

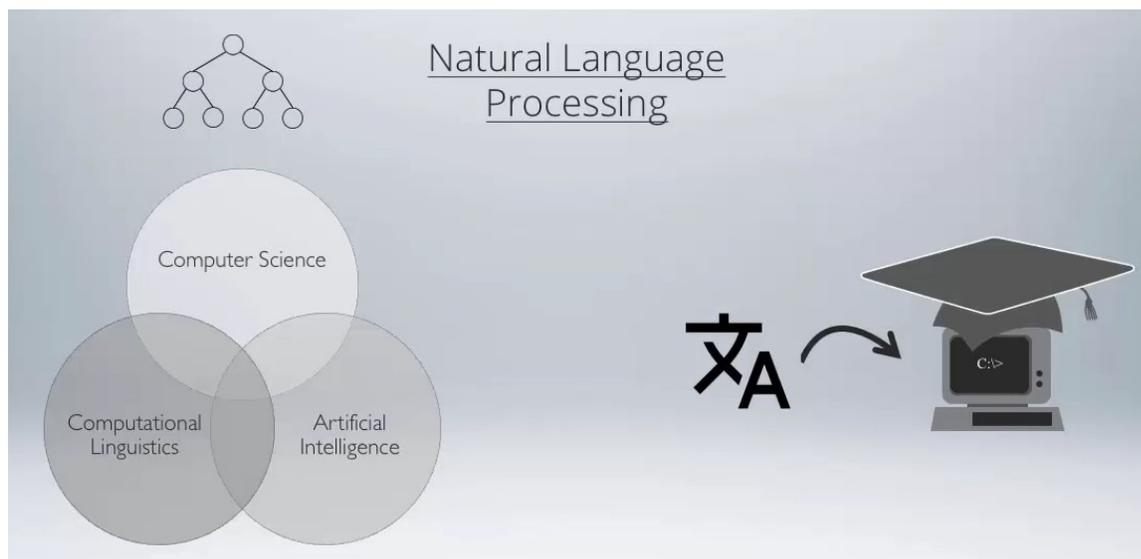


It involves converting unstructured text into structured data that can be analyzed using statistical methods. There are three main steps involved in text analytics.

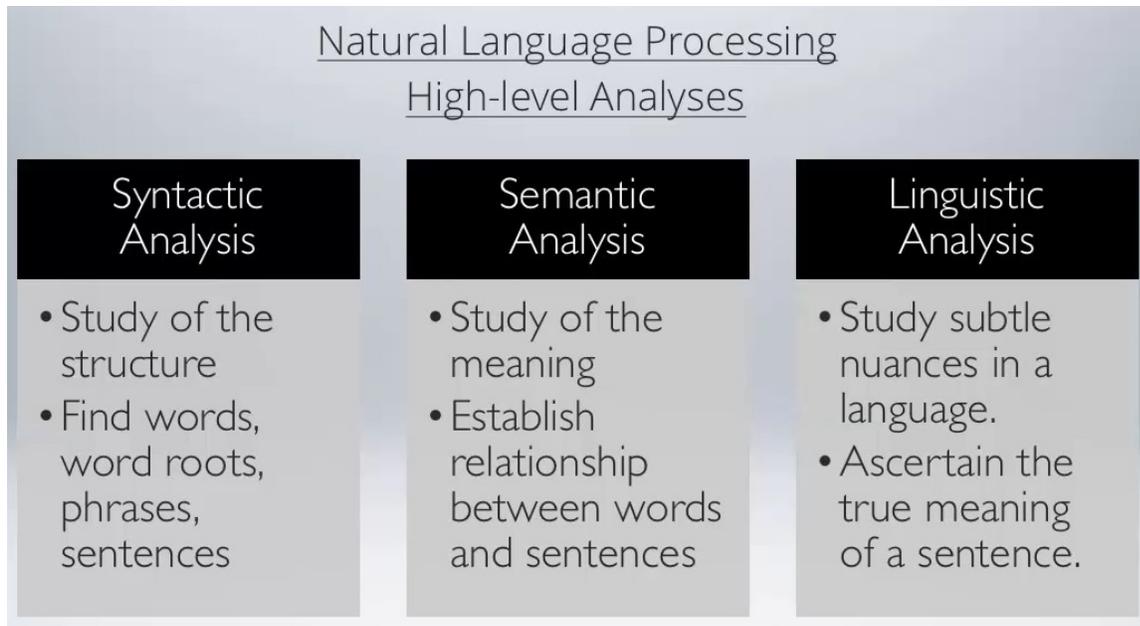


First, data is fetched from a source, such as social media, emails or customer service records. Second, natural language processing techniques are applied to extract certain attributes from the text and present them as structured data. And third, the extracted structured data is stored and analyzed.

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

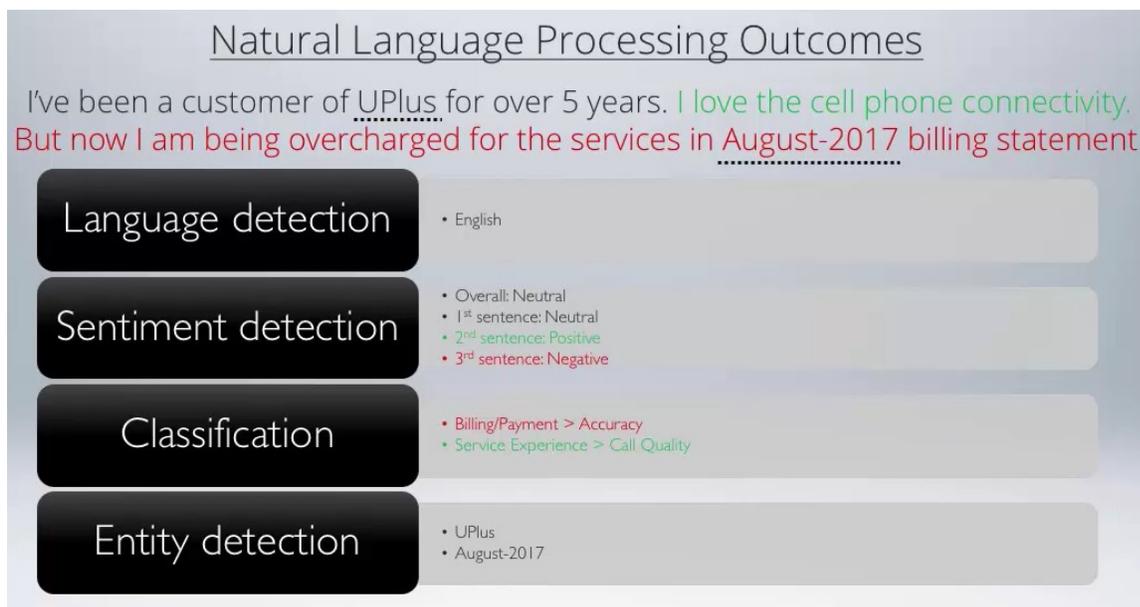


In particular, it is concerned with programming computers to fruitfully process large volumes of natural language text. At a high level, the following analyses occur during natural language processing.



Syntactic analysis analyzes the structure of text and recognizes the parts of speech, such as nouns, verbs, adjectives, etc. Semantic analysis establishes the meaning of a piece of text. So the relationship between words and sentences is analyzed. In the linguistic analysis phase, the nuances of a language are considered to ascertain the exact meaning of the text.

Here are the outcomes of natural language processing.



The language of the text is detected. The sentiment is detected as positive, negative or neutral. Sentiment is the general attitude expressed towards a subject. The overall sentiment as well as the sentiment per sentence is reported.

The text is classified into one or more pre-defined categories, such as the business functions in a company. For each category detected, the sentiment of the corresponding sentence is also reported. Entities refer to the proper nouns found in the text such as names of people, places, dates and times, organizations, etc. This helps in establishing the subject of the text.

You have now reached the end of this lesson which provided an overview of text analytics and how it works.

Email routing

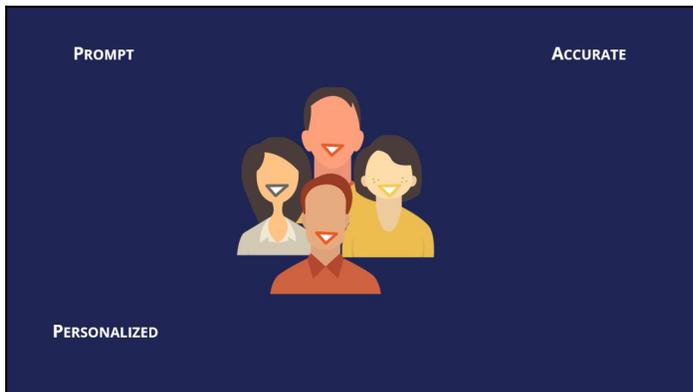
Introduction

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.

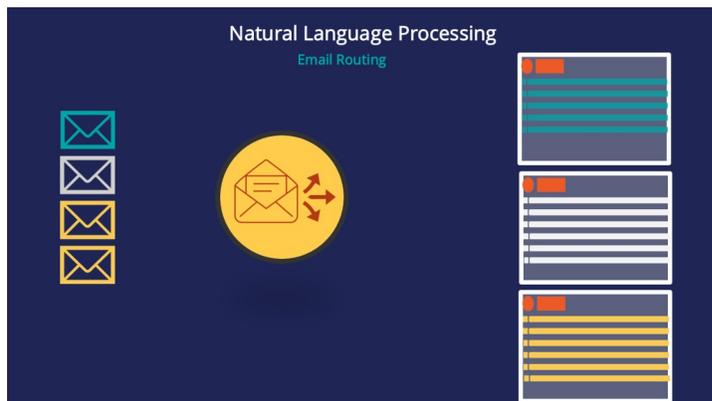


Assume a company has an email channel in which its customers can send any type of email —from service requests to compliments to sales inquiries—to the product support team. These emails are often routed to the same container and are uncategorized. This makes it difficult to provide each customer with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

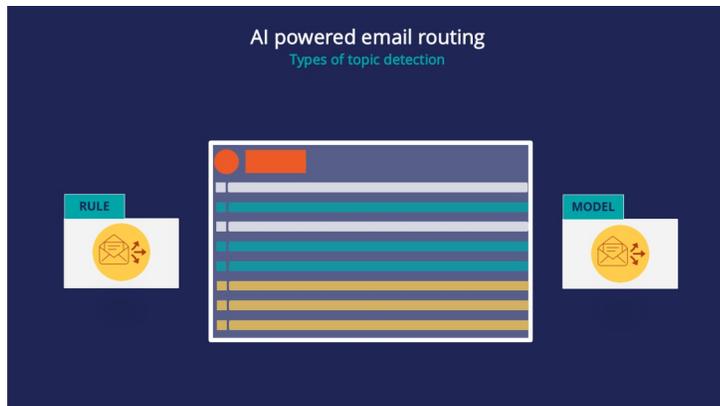


In Pega Infinity, this issue can be addressed by routing the emails using artificial intelligence. Pega Infinity uses AI-powered Natural Language Processing to detect the topic of an email and route the email to the appropriate container.

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.

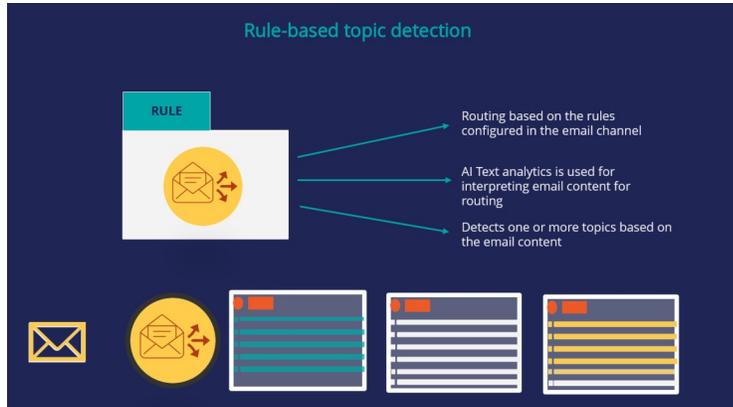


Email routing is done using the topic detection mechanism. The two types of topic detection are rule-based and model-based.



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.

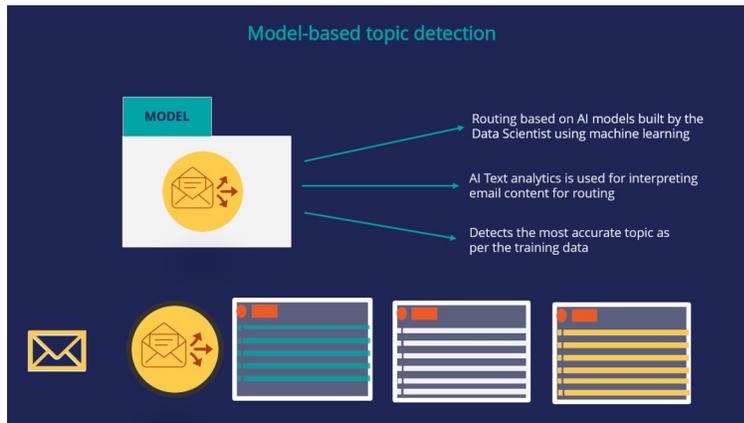
Let's consider a service request email from a customer. The email content is analyzed and routed to the right container. If an email from another customer contains words that are associated with two topics, the rule-based topic detection detects both topics. The email can then be routed to two different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



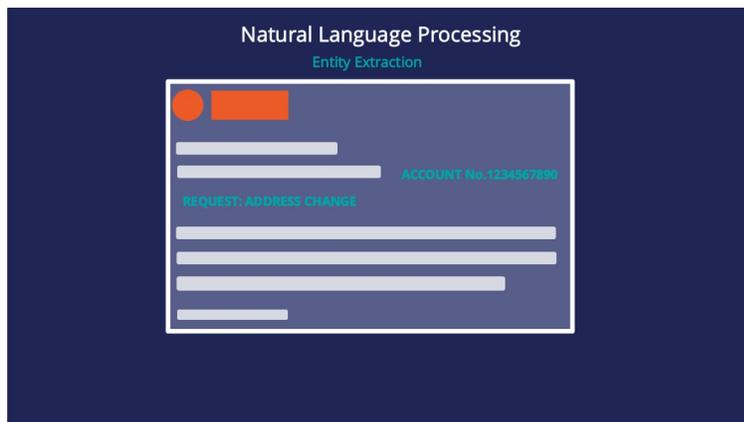
In model-based topic detection the routing is based on AI models built by a Data Scientist using machine learning. Building these models requires a training data set and a test data set. The data sets consist of a list of emails and the associated topic for each email. This type of topic detection identifies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



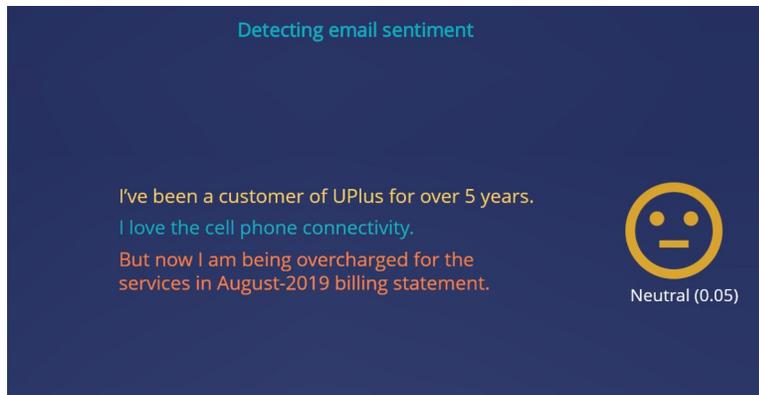
Pega Infinity also enables you to extract entities from an email. This means that when an email is sent, certain entities such as account number, email address, street address, etc. can be automatically detected and extracted. This allows certain emails to be automatically processed or given priority.



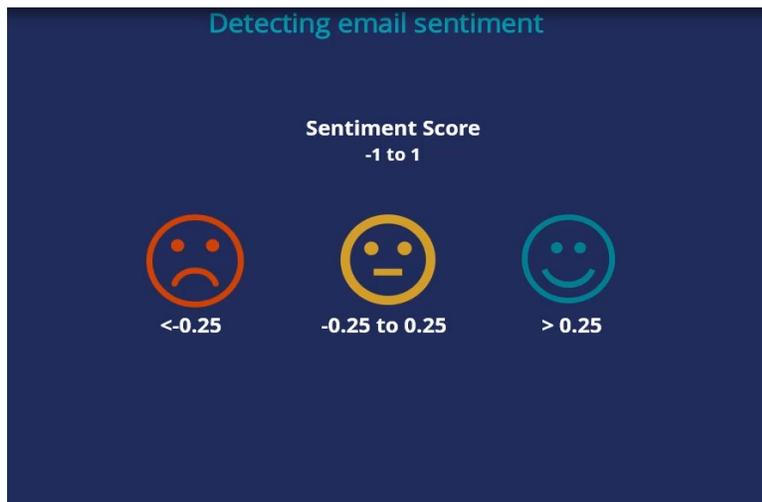
Besides Topic detection and Entity Extraction, Pega Infinity uses its AI-powered text analytics to enable you to detect the sentiment of an email based on its content.

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.

U+ Bank customer support



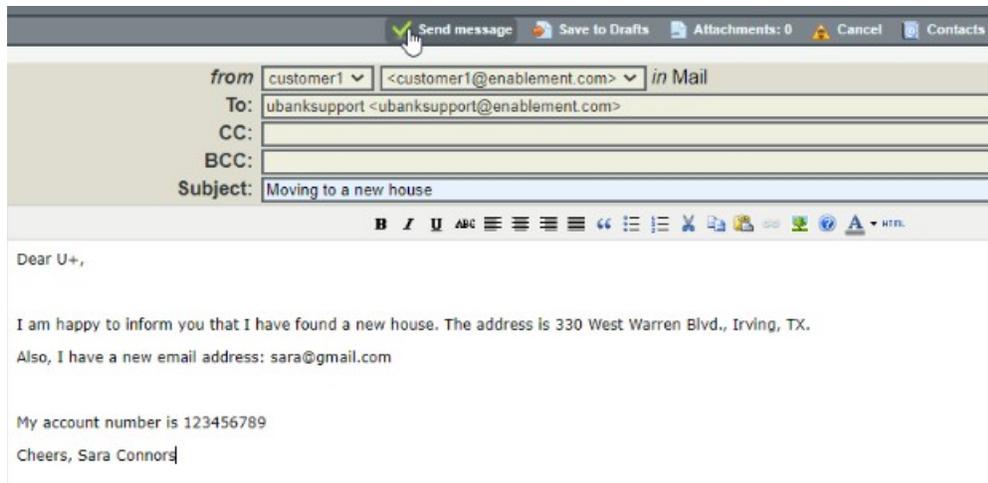
U+ Bank customer support

Mar 11, 2021

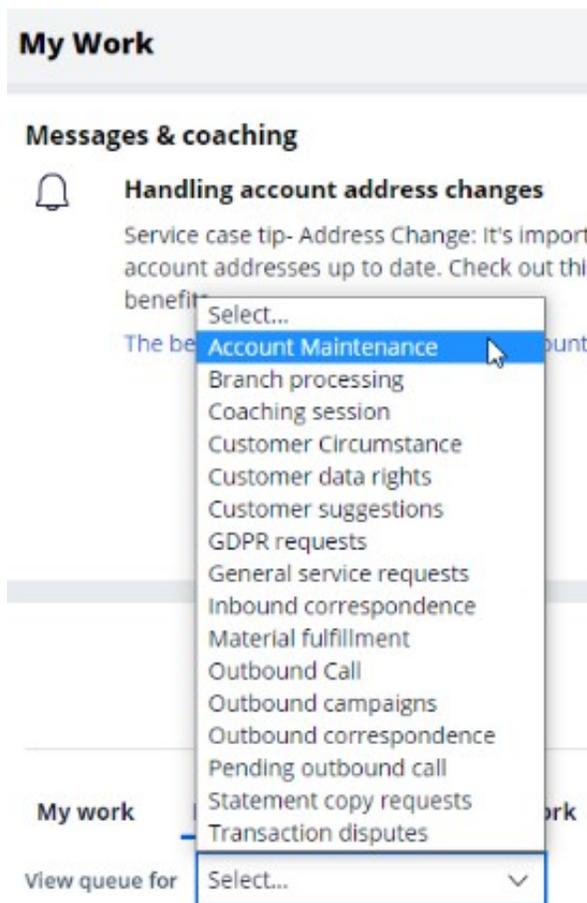
[Open prediction](#)



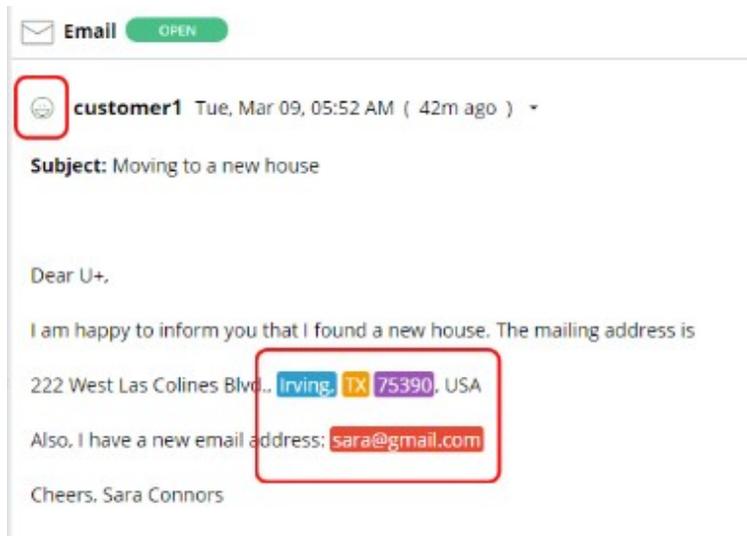
For example, when Sara writes an email to inform U+ Bank that she has moved to a new house, the text prediction detects an address change as the topic.



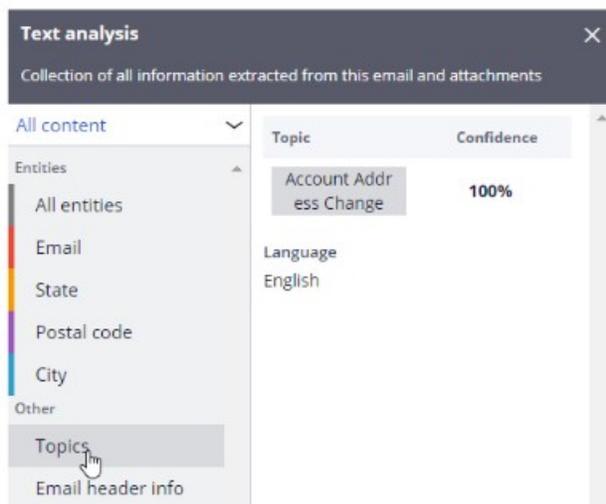
The **Account maintenance** department receives all emails where an address change is detected as the topic of the message.



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.



U+ Bank Support

ubanksupport@enablement.com

Suggested cases

Close Account, Account Address Change, Order statement copy, Dispute Transaction

Suggested replies

Apology, Clarification, Thank You

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
<input type="text" value="Route to work queue"/>	<input type="text" value="AccountMaintenance"/>

When

Action	Value
<input type="text" value="Route to work queue"/>	<input type="text" value="TransactionDisputes"/>

When

[+ Add condition](#)

Otherwise

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.

- Header
- NLP**
 - Entities
 - Topic
 - Decisioning result
 - Sentiment**
 - Language
- Condition

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

Confidence threshold

Every channel is associated with a text prediction.

Text Analytics Use advanced configuration

Edit topics

Record training data

Enable subject analysis

Analyze email attachments Always ▾

[Open Text Prediction](#) 

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

Outcomes

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

Topics

Entities

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**.

Action > Account Address Change

Machine learning [Keywords](#)

A single keyword can consist of multiple words. Use tab/enter to separate keywords.

Should words

move × change × address ×

Must words

And words

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
Document

Analysis
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	content	result	type
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 billir	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

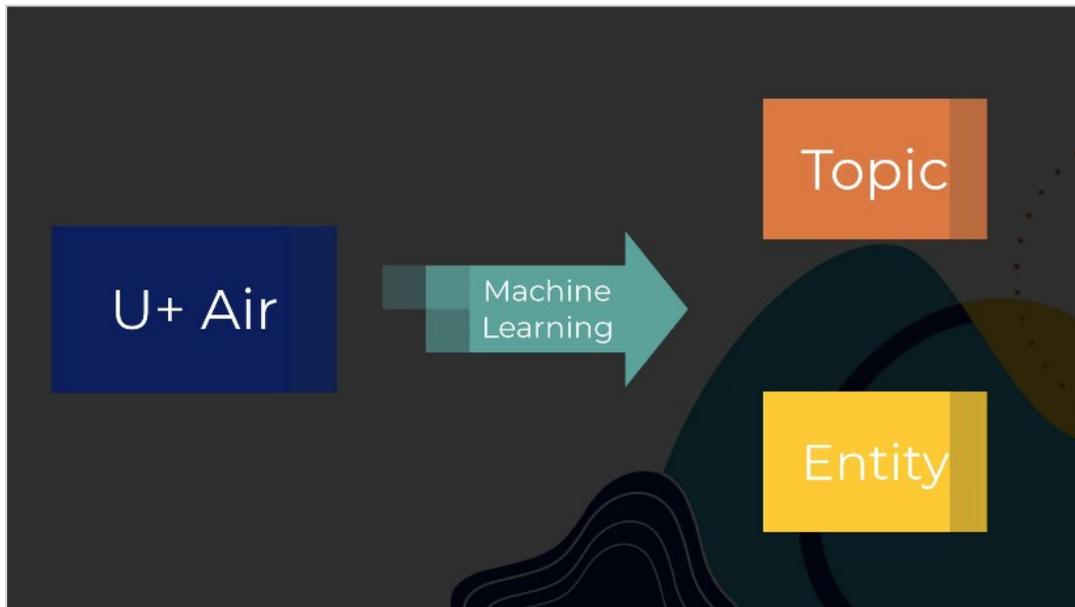
Web Chatbot is a feature in Pega Customer Service™ that allows your customers to enter questions or requests in a chat window that is accessible from a customer-facing web page, and then receive answers and guidance.

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.

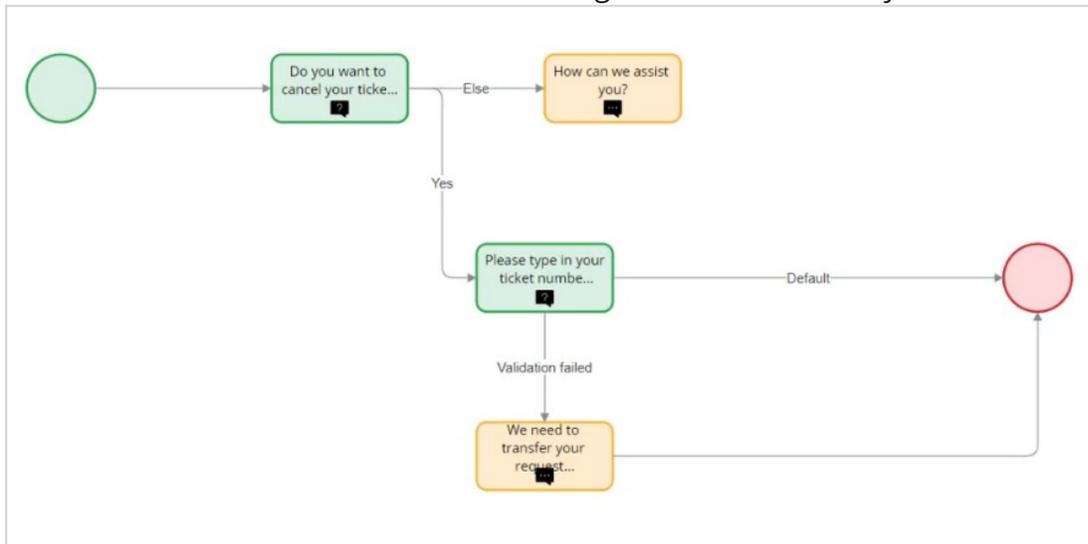
U+ Air uses machine learning to identify the topic of the incoming message and detect entities such as ticket numbers.



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.



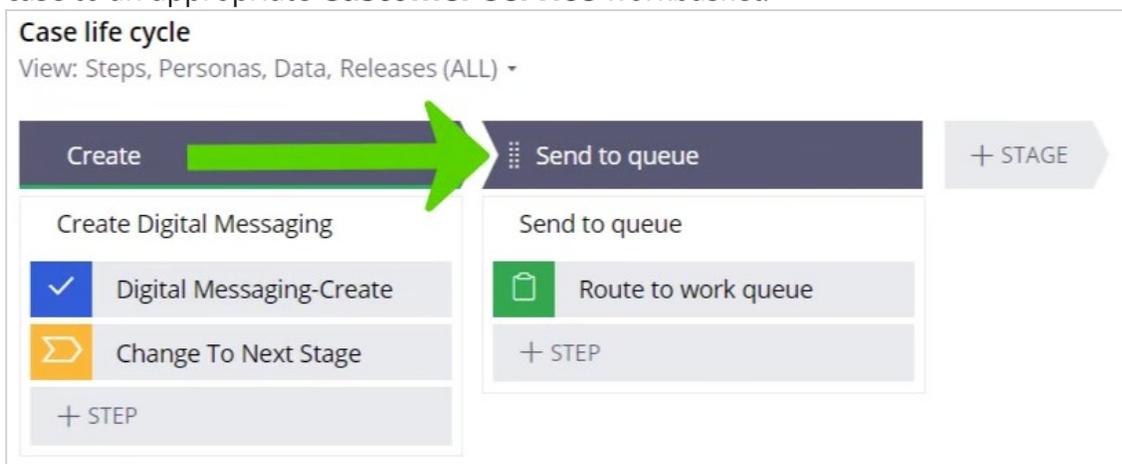
The prediction detects the ticket number entity that Troy provides, and then asks for confirmation.

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. The 'Text analysis' tab is selected. The 'Case type' dropdown is set to 'Cancel a ticket'. The 'Create case command' dropdown is set to 'cancel ticket'. The 'Can be suggested by' dropdown is set to 'Chatbot to user OR AI to agent'. The 'Requires authentication' checkbox is unchecked.

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	+ Add model
pySystemEntities	Pega-NLP	English	ALL ↗	ACTIVE	🗑️

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 inbetween 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.

Response configuration ✕

Response Text analysis Entities extraction

Map entities types to **Cancel a ticket** case model.

Entity	Case property	
ticket_number ▼	ticket_number ▼	

[+ Add mapping](#)

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	
Confirmation	
Value	Type
Yes	User input

9:59 AM

Please type in your ticket number.

Ask (skipped): ticket_number 

Value	Type
AA562	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.

Create (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.