



This PDF is generated from authoritative online content, and is provided for convenience only. This PDF cannot be used for legal purposes. For authoritative understanding of what is and is not supported, always use the online content. To copy code samples, always use the online content.

Predictive Routing Deployment and Operations Guide

Prepare Your Data

Contents

- 1 Prepare Your Data
 - 1.1 Interaction data
 - 1.2 Agent Profile data
 - 1.3 Customer Profile data
 - 1.4 Outcome Data
 - 1.5 Correctly Specifying Data Types in Your Dataset
 - 1.6 Supported Encodings
 - 1.7 Unsupported Characters in Agent and Customer Profiles and Datasets
 - 1.8 Considerations
 - 1.9 Data Size Guidelines - Data Import, Model Training, and Feature Analysis
 - 1.10 Genesys Info Mart Data

Prepare Your Data

This topic addresses issues such as data sources, what kind of data Genesys Predictive Routing (GPR) needs in order to function, how much data you need, and how best to structure it. See [Supported Encodings](#) and [Unsupported Characters](#) for information about constraints on data formats. It also includes [data size guidelines](#) and a [list of Genesys Info Mart tables](#) that provide data for GPR.

Because environments differ and the exact metrics you want to target for improvement are also dependent on your environment and business needs, this topic offers only general guidelines.

GPR can use any relevant data available in your environment. Some data is automatically imported into GPR, but the remainder must be combined into consistent schemas and saved as CSV files. You can use either the GPR web application or the API to import this CSV data. You can upload your CSV file itself or as a ZIP archive.

Important

- This *Guide* assumes you are using the GPR web application. For details about the API, see the [Predictive Routing API Reference](#).
- Genesys supports only the GPR application and the GPR API for uploading Dataset and Agent and Customer Profile data to Genesys Predictive Routing.
- Once a Dataset schema is created and accepted in AICS, you cannot change the schema. If you need to modify the schema, you must delete the Dataset and create a new one with the desired schema.

In general, you need the following types of data:

Interaction data

- You must create a CSV file containing the required data and upload it, using either the GPR web application or the GPR API.

Agent Profile data

- ASC automatically extracts Agent Profile data from Configuration Server. See [Configuring Agent Profiles](#) in the *Predictive Routing Help* for how to configure an Agent Profile schema.

Customer Profile data

- You must create a CSV file and upload it, using either the GPR web application or the GPR API. See [Configuring Customer Profiles](#) in the *Predictive Routing Help* for how to configure a Customer Profile schema.

Important

If you are running AICS 9.0.015.00 or lower and your Customer Profile CSV file is very large, split your CSV file. Upload one section, then perform as many separate append operations as necessary to upload the entire Customer Profile. For example, in a medium-sized single-server deployment, (64 GB RAM; 16 CPU), each CSV upload should be no larger than 340 MB (approximately 100 features and 100000 rows). AICS 9.0.015.03 and higher uses MinIO for uploads, which resolves this issue.

Outcome Data

- You must create a CSV file and upload it, using either the GPR web application or the GPR API.

Correctly Specifying Data Types in Your Dataset

GPR automatically determines the data types of the columns in your dataset during dataset initialization by analyzing the first 1000 rows of each column. To ensure that GPR can make a correct determination, Genesys recommends that you insert a "dummy" row at the beginning of your dataset that contains values that can be unambiguously interpreted as the expected data types for each column. This prevents cases in which the first 1000 rows may contain all NULL or 0 values, which might lead to an incorrect data type assignment (since 0 can be a valid integer, float, or Boolean value). If a column does contain meaningful values, the dummy row is analyzed along with the other values and contributes to the data type determination. Genesys recommends you use the following data type specifications in your dummy row:

- 'a_string' - Is recognized as a string.
- 2.1, or any integer or float value > 1 - Is recognized as a float.
- False or True - Is recognized as Boolean.
- Unix Timestamp (such as 1535538976) - Is recognized as a timestamp.

Supported Encodings

By default, GPR handles data using UTF-8 encoding. However, starting with release 9.0.014.00, GPR supports importing of data that uses certain legacy encodings. [Appendix: Supported Encodings](#) lists those encodings currently supported. This list is updated as new encodings are verified. If you use an encoding type that is not listed, contact your Genesys representative for assistance.

Important

All responses and returned data is provided in UTF-8 encoding.

Unsupported Characters in Agent and Customer Profiles and Datasets

The following characters are not supported for column names in Datasets or Agent and Customer Profile schemas. If GPR encounters these characters in a CSV file, it reads them as column delimiters and parses the data accordingly.

- | (the pipe character)
- \t (the TAB character)
- , (the comma)

Workaround: To use these characters in column names, add double quotation marks (" ") around the entire affected column name, except in the following situations:

- If you have a comma-delimited CSV file, add double quotations marks around commas within column names; you do *not* need quotations for the \t (TAB) character.
- If you have a TAB-delimited CSV file, add double quotations marks around TAB characters within column names; you do *not* need quotations for the , (comma) character.
- You must *always* use double quotations for the | (pipe) character.

Unsupported characters in releases prior to 9.0.014.00

In releases prior to 9.0.014.00, certain characters in column names are ignored, are unsupported, or cause an upload to fail, as explained in the following points:

- Columns with the following symbols in their column names are not added to Agent Profiles or Customer Profiles:
*, !, %, ^, (,), ', &, /, â, è, ü, ó, â, ï
- The following symbols in column names are ignored, and the column is added with the symbol dropped out as though it had not been entered:
[Space], -, <
- Non-ASCII characters are not supported. How they are handled differs depending on what data you are uploading:
 - In Agent Profiles and Customer Profiles, columns with non-ASCII characters in the column name are not added.
 - In Datasets, when a column name contains a mix of ASCII and non-ASCII characters, GPR removes the non-ASCII characters from the column name as though they had not been entered and correctly uploads all column values.
 - In Datasets, when a column name contains only non-ASCII characters, the column name is entirely omitted. All the column values are preserved, but you cannot modify or save the schema. In this scenario, GPR generates the following error message: An unhandled exception has occurred: `KeyError('name')`.

Logs for Unsupported Characters

The following Agent State Connector log messages record issues with unsupported characters:

- `<datetime> [47] ERROR <BOTTLE> schema_based.py:63 Invalid expression while parsing: <fieldname> = None`
- `<datetime> [47] ERROR <BOTTLE> agents.py:172 Fields set([u'<fieldname>']) were ignored because names were invalid.`

Considerations

Important

Genesys testing has identified certain **guidelines about data size** to keep in mind while creating datasets and planning modeling and feature analysis.

- What metrics of Contact Center operation do you want to optimize?
- What databases are available?
- Is it possible to join the data from those databases?
- Is the data clean and consistent in format? This is especially an issue if you are joining data from various sources.
- Do you have enough data to draw conclusions about agents performance?
- What data is available from the IVR and the CRM database at runtime?

Note that when you create your CSV file, you will need to include a value, a `context_id` or customer ID, that can be entered into your scoring request from your strategy and that enables you to retrieve the relevant record.

Data Size Guidelines - Data Import, Model Training, and Feature Analysis

Description	Column Count	Row Count/File Size
Size of data that can be uploaded to create a dataset in a single batch. You can append data; you can load bigger datasets in multiple batches. Data uploads successfully for a file with 250 columns but with a smaller number of total records so that the total file size is less than the 250 MB limit.	250 columns maximum	250 MB file size
Minimum number of records needed to train a DISJOINT model	Not Applicable	10

Description	Column Count	Row Count/File Size
for an agent.		
Total Cardinality limit for model training. Total Cardinality = the number of numeric columns plus the sum of the number of unique values across all string columns within a given dataset.	No specific column count; has been tested up to 250 columns.	Total Cardinality should be less than 2 to the power of 29.
Record count limit for GLOBAL model training.	Not Applicable; from a model-training perspective there is virtually no limit on the number of columns. The constraining issue is the possibility of compromising the Global model quality by ending up with a reduced number of samples for training.	The total number of records should be less than 2 to power of 29 (that is, 536870912) divided by Total Cardinality as defined above. Example 1: You are required to use ALL of the data for training the GLOBAL model (note that the GLOBAL model is trained even if you select DISJOINT, so that the scoring engine can rank agents who do not yet have data). The dataset contains 1 million records. Therefore the maximum total cardinality is 536 (536870912 divided by 1 million) . Example 2: You can <i>undersample</i> the data for training the GLOBAL model—that is, use fewer than the ideal number of records for training. You might take 10,000 as the total cardinality, but only 53,687 of your total of 1 million records will be used for training. The calculation to determine this is $10,000 * 53,687 = 536870912$ (the maximum cardinality).
Column count limitation on the Feature Analysis report, Agent Variance report, Lift Estimation report, and Model Quality report.	250	Not Applicable

Genesys Info Mart Data

The following table represents the data required from Genesys Info Mart (GIM) and the Genesys Configuration Database.

Domain	Fields	Table
Agent Profile	Agent username	RESOURCE_
Agent Profile	Employee ID	RESOURCE_
Customer Profile	CustomerID	Attached Data to map with interaction metadata record
Customer Profile	SERVICE_TYPE	Attached Data to map with interaction metadata record
Interaction Metadata	INTERACTION ID	Interaction_Fact (IF)
Interaction Metadata	INTERACTION TYPE	IRF
Interaction Metadata	MEDIA TYPE	Media_Type
Interaction Metadata	RESOURCE ROLE	TECHNICAL_DESCRIPTOR
Interaction Metadata	ROLE REASON	TECHNICAL_DESCRIPTOR
Interaction Metadata	TECHNICAL RESULT	TECHNICAL_DESCRIPTOR
Interaction Metadata	RESULT REASON	TECHNICAL_DESCRIPTOR
Interaction Metadata	MEDIA RESOURCE (VIRTUAL QUEUE)	RESOURCE_
Interaction Metadata	RESOURCE GROUP COMBINATION	RESOURCE_GROUP_COMBINATION
Interaction Metadata	ROUTING TARGET	ROUTING_TARGET
Interaction Metadata	TARGET_OBJECT_SELECTED	TARGET_OBJECT_SELECTED
Interaction Metadata	SKILL EXPRESSION	REQUESTED_SKILL
Interaction Metadata	START_TS	INTERACTION_FACT
Interaction Metadata	END_TS	INTERACTION_FACT
Interaction Metadata	LAST ROUTING POINT	RESOURCE_
Interaction Metadata	LAST VQ	
Interaction Metadata	IS_LAST_RESOURCE	IRF
Interaction Metadata	SOURCE_ADRESS (ANI,...)	SOURCE_ADDRESS
Interaction Metadata	TARGET_ADRESS (DNIS,...)	TARGET_ADDRESS
Interaction Metadata	WAITING TIME	(QUEUE DURATION, ROUTING POINT DURATION, MEDIATION DURATION)
Interaction Metadata	RINGING TIME	RING_DURATION
Interaction Metadata	HANDLE TIME	(Talk_Duration, ACW Duration,...)
Interaction Metadata	HOLD TIME	HOLD_DURATION
Interaction Metadata	HOLD COUNT	HOLD_COUNT
Interaction Metadata	ACW COUNT	AFTER_CALL_WORK_COUNT
Interaction Metadata	FOCUS TIME	FOCUS_TIME

Domain	Fields	Table
Interaction Metadata	CONSULTATION TIME (CONS_RCV_TALK_DURATION, POST_CONS_XFER_TALK_COUNT,...)	