



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.

# Genesys Knowledge Management User Guide

Large Training Objects

5/12/2025

# Large Training Objects

This topic describes part of the functionality of [Genesys Content Analyzer](#).

If your training object is very large (over 50,000 e-mails), training may consume considerable memory and time. To reduce this consumption without impacting quality, follow these recommendations when you [Scheduling training](#):

- Set Cross Validation to None.
- Set Keyword Threshold above 25.
- Set Min Samples in Category above 25.
- Set Training Quality below 4. A level of 3 or 4 is adequate for production use.

You should also allocate memory as follows:

- Ensure that the host machine of Training Server has at least 4 GB of RAM for Solaris, or 2 GB of RAM for Windows.
- In the .sh or ProcessParameters.ini file, change the parameter -Xmx800m as follows:
  - On Windows, change to -Xmx1400m. This is enough for a training object of about 40,000 e-mails, the maximum recommended size on this platform.
  - On Solaris, change to -Xmx3000m. This is enough for a training object of about 100,000 e-mails, the maximum recommended size on this platform.

For large training objects, these recommendations supersede those in the “Knowledge Manager” section on [Administration General Recommendations](#).

A successful test has been done with the following parameters:

Host: Solaris, Enterprise 450 Model 4300 with 4000 MB RAM

Training object: 100,000 e-mails in 1,000 categories

Cross Validation: None

Keyword Threshold: 25

Min Samples in Category: 25

Training Quality: 3

The expected computational time is between 12 and 18 hours.

Note that the model produced has no quality ratings because you set Cross Validation to None. Genesys strongly recommends against using cross-validation on such large training objects. To obtain quality ratings for the model, build an additional small training object and **test the model on a training object**.