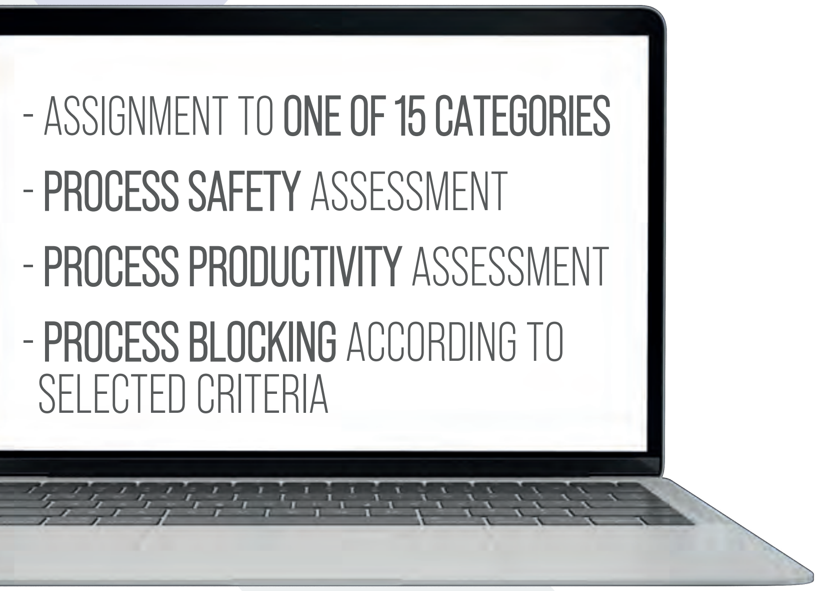
# BTC MACHINE LEARNING **PROCESS CLASSIFICATION**

DATASHEET

eAuditor V7 WEB



- ASSIGNMENT TO **ONE OF 15 CATEGORIES**
- **PROCESS SAFETY** ASSESSMENT
- **PROCESS PRODUCTIVITY** ASSESSMENT
- **PROCESS BLOCKING** ACCORDING TO SELECTED CRITERIA

## 15 CATEGORIES

**Processes are classified based on information obtained via the Bing search engine and are assigned to one of 15 categories.**

## EFFECTIVENESS > 85%

**The high efficiency of classification results from the applied artificial intelligence algorithms and the built database of keyword patterns within individual categories.**

## NUMBER OF ITEMS IN THE CATALOG

**The catalog of classified processes and applications is constantly being expanded and is available in the cloud for any use in other solutions via API.**

## HIGH SPEED

**Standard classification speed > 10 processes/sec. Can be increased as needed.**

## MACHINE LEARNING IN CLASSIFICATION OF COMPUTER PROCESSES AND APPLICATIONS

Process and application classification can be useful in any entity where supervision and control of user activity can have a real impact on security and user productivity assessment. Implementation of the machine learning algorithm allows for efficient and fast classification of any process in terms of and assignment (classification) to the appropriate category. The process classification module in the eAuditor V7 WEB system is prepared for the occurrence of various random events in such a way that despite an error on the server side, it does not interrupt operation and correctly performs its task, assigning processes to the appropriate categories.

## CLASSIFIER OF COMPUTER PROCESSES AND APPLICATIONS

The Bayesian classifier, which is based on Bayes' theorem, is particularly suitable for solving problems with very high dimensions at the input. Despite the simplicity of the method, it often works better than other, very complicated classification methods. The aforementioned classifier can be trained in supervised learning mode. This means that for the algorithm to work correctly and even better, human supervision is necessary, who will constantly analyze and correct any errors in the algorithm. The classification is correct as long as the correct category is more likely than the others.

## PROCESS CLASSIFICATION TIME

The classification of a single process takes about 1 second. In practice, high efficiency is achieved due to multi-threaded handling of classification processes (simultaneous classification of up to a hundred processes).

## CORRECTNESS OF CLASSIFICATION OF COMPUTER PROCESSES AND APPLICATIONS

As part of the machine learning test in eAuditor V7 WEB, 1000 random and unpopular processes were categorized. The correctness of assigning categories is **about 85%**. The problem with achieving better results does not lie with the algorithm, because it determines the highest probability of a given category. The problem is that each process is searched for in a web browser, which in turn does not always return appropriate search results. It also happens that the searched processes may refer to different programs from different categories.

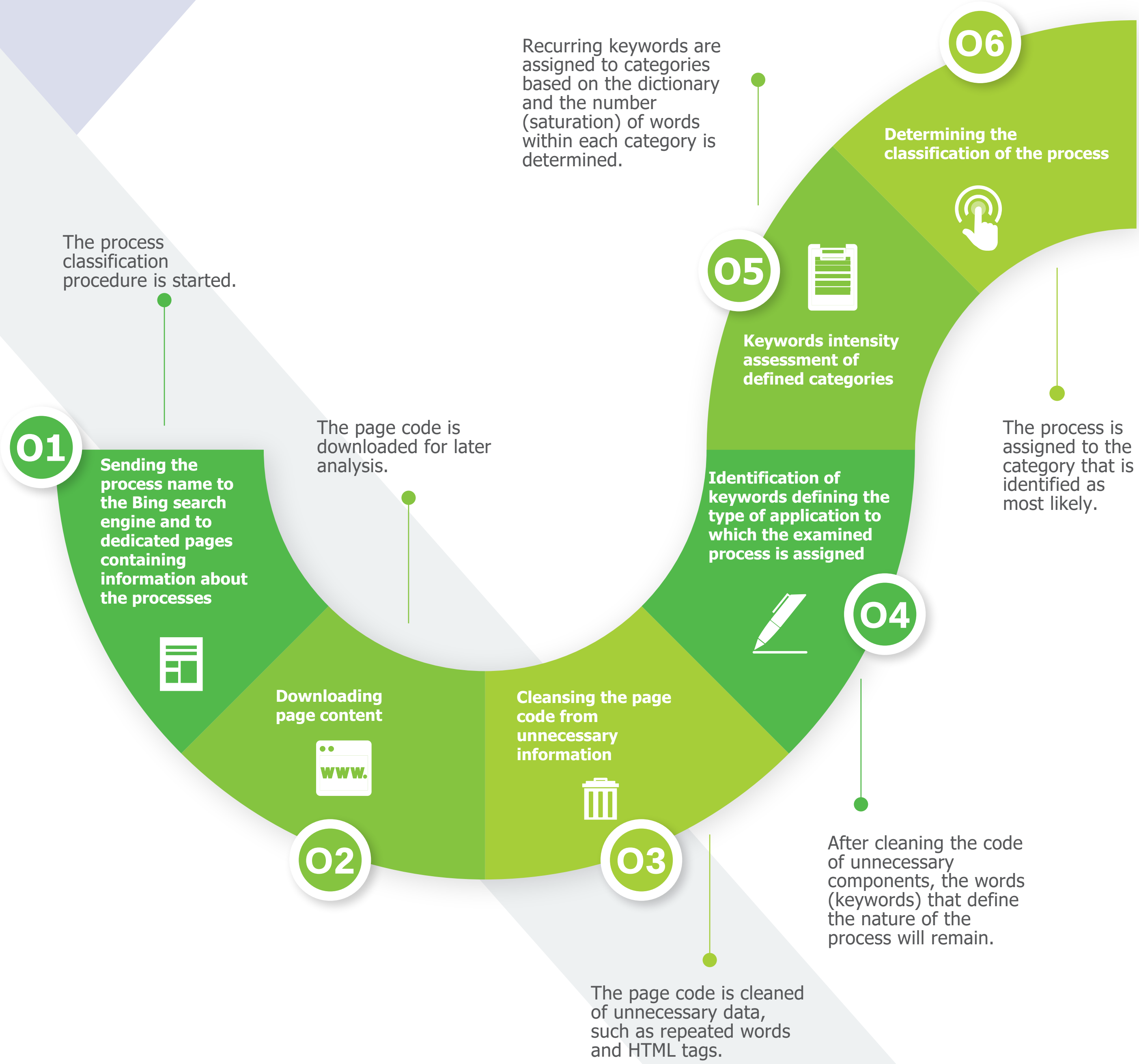
For example, the *win_driver_installer.exe* process can be categorized as both a system process and a driver.

## EFFICIENCY

The solution implements a number of mechanisms that reject processes that will most likely not return correct results. This saves time and does not involve the classification algorithm in classifying erroneous processes. The classifier engine is located in the data center and is supported by a scalable set of servers, which ensures **unlimited performance.**

# BTC MACHINE LEARNING **PROCESS CLASSIFICATION**

## HOW DOES MACHINE LEARNING WORK IN CLASSIFYING SYSTEM PROCESSES AND APPLICATIONS?



The process classification procedure is started.

**01** **Sending the process name to the Bing search engine and to dedicated pages containing information about the processes**

The page code is downloaded for later analysis.

**02** **Downloading page content**

The page code is cleaned of unnecessary data, such as repeated words and HTML tags.

**03** **Cleansing the page code from unnecessary information**

After cleaning the code of unnecessary components, the words (keywords) that define the nature of the process will remain.

**04** **Identification of keywords defining the type of application to which the examined process is assigned**

Recurring keywords are assigned to categories based on the dictionary and the number (saturation) of words within each category is determined.

**05** **Keywords intensity assessment of defined categories**

The process is assigned to the category that is identified as most likely.

**06** **Determining the classification of the process**

## WHY DID WE INTRODUCE MACHINE LEARNING TO EAUDITOR V7 WEB?

Here are some reasons why we used Machine Learning in eAuditor V7 WEB instead of a statistical process database:

- automatic creation of a database with information about processes, their categories along with an assessment of productivity and safety,
- resistance to changing categories over time (the category changes automatically),
- automatic, real-time synchronization with the database of classified processes in the data center.