

## Information About Features

In this year's project, we will be using data from PE (Portable Executable) files.

More information can be found in this [link](#). Additional research on the internet is encouraged.

Feature (Column) Name	Information
Sha256	A unique identifier for the file.
file_size	Size of file on disk
vsize	Virtual size – size of the file image when loaded into memory
imports	Number of imported functions
exports	Number of exported functions
has_debug	Whether a file has a debug section
has_tls	Whether a file has thread local storage
has_resources	Whether a file has resources
has_relocations	Whether a file has relocations
has_signature	Whether a file has a signature
symbols	Number of symbols
numstrings	The number of printable strings that are at least five printable characters long.
avlength	Average string length
printables	Number of printable characters
paths	number of strings that begin with C:\ (case insensitive) that may indicate a path
urls	Number of occurrences of http:// or https:// (case insensitive) that may indicate a URL
registry	the number of occurrences of HKEY_ that may indicate a registry key
MZ	the number of occurrences of the short string MZ that may provide weak evidence of a Windows PE dropper or bundled executables
file_type_trid	The file type with the highest probability, given by TRID*
file_type_prob_trid	The probability of mentioned file type
A,B,C	Anonymous columns – might be helpful for prediction.
Label	Label column – is a file malicious (1) or not (0).

\* **TrID** – a utility designed to identify file types from their binary signatures. [Link](#).