now i will give you the details of the data csv and then a several rows so we could begin:  
  
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). |

row1:

a5d72c609a1594621e8eb775e7324d4fc291d5569cfefd5f11e9785e85b9c80f 6744064 6799360 524 3 0 1 1 0 1 0 64882 0 0 0 28 899075 13.857079 Win32 EXE PECompact compressed (generic) 64.2 -3.978980782 5.732353959 er 1

row2:

ccaa49b76e6fd84842399c4389e2086a5a3687dfabbf016e92a5cb627995fbd2 22016 45056 2 0 0 0 1 0 0 0 89 0 0 0 1 856 9.617977528 Win32 Dynamic Link Library (generic) 38.4 7.064132731 5.738201141 er 0

row3:

b0b126adc73b980a114e38027b082d149fcf491547452f091686613ecf922817 686136 704512 133 4 1 1 1 1 0 0 2699 0 32 0 2 32767 12.14042238 Win64 Executable (generic) 82 1.052726682 5.722733498 gow 0

row4:

ff6176c53ca9b58040c93870c264843b93b5e290c28e576af5e23b13f2db599b 8192 36864 6 0 0 0 0 0 0 71 0 0 0 1 678 9.549295775 Win64 Executable (generic) 28 2.217531102 5.896053314 vh 1

row5:

59b8bc6ac157a553440978406c8aa8eb8a25c166af11754e1fc1aac3e65c8a59 396288 421888 133 0 1 1 1 0 0 0 4729 0 0 0 5 35881 7.587439205 Win64 Executable (generic) 32.2 9.401164194 vh 1

row6:

04c28622a87150c4760ccd9801d378c6e96373099bbe020512be900fe23935c8 434200 450560 1 1 1 1 0 0 623 1 15 0 3 9025 14.48635634 Generic CIL Executable (.NET, Mono, etc.) 72.5 5.760772994 5.412149429 gow 0