Dataset Column Descriptions & Malicious vs Benign

FileName

Name of the file, typically a DLL, used for identification.

md5Hash

Unique hash value for the file, used for integrity verification.

Machine

Specifies the target machine architecture (e.g., 32-bit or 64-bit).

DebugSize

Size of debugging information embedded in the file.

DebugRVA

Relative virtual address of the debug directory in memory.

MajorImageVersion

Major version number of the image file.

MajorOSVersion

Major version of the operating system the file is built for.

ExportRVA

Relative virtual address of the export table, if present.

ExportSize

Size of the export table, indicating functions exported.

IatVRA

Address of the Import Address Table (IAT) used for linking.

MajorLinkerVersion

Major version of the linker used to build the file.

MinorLinkerVersion

Minor version of the linker used.

NumberOfSections

Number of sections in the file, affecting its structure.

SizeOfStackReserve

Reserved stack memory size for execution.

DllCharacteristics

Security-related characteristics of the DLL.

ResourceSize

Size of embedded resources in the file.

BitcoinAddresses

Presence of Bitcoin wallet addresses (potential indicator of malware).

Benign

Output label: 1 = Benign (not malicious), 0 = Likely malicious.

Differences Between Malicious and Benign Files

In this dataset, the 'Benign' column indicates whether a file is malicious or safe:
- Benign (1) = The file is considered safe (not malicious).
- Benign (0) = The file is considered malicious (potential malware or threat).

Differences Between Malicious and Benign Files:
1. **File Characteristics**:
   - Malicious files might have unusual or large DebugSize, ExportSize, or ResourceSize.
   - Benign files usually follow standard DLL structures.

2. **Sections and Linker Information**:
   - Malicious files may have an irregular number of sections (NumberOfSections).
   - They might use uncommon MajorLinkerVersion values.

3. **Presence of Bitcoin Addresses**:
   - If BitcoinAddresses > 0, the file is more likely to be malicious (e.g., related to cryptojacking malware).

4. **Security Flags (DllCharacteristics)**:
   - Malicious files might lack standard security features like Data Execution Prevention (DEP) or Address Space Layout Randomization (ASLR).