**Dataset Abstract:**

So basically our dataset is divided into 2 types of files i.e. Malware and other is benign (legitimate software). As getting to the feasibility site of our work so we have collected:

* 41324 types of benign files (exe, dll) means that are legitimate.
* 96724 types of malware files from Virusshare.com

All are in Window PE (portable executable) format.

**PE File Header:**

Like other executable files, a PE file has a collection of fields that defines what the rest of file looks like. The header contains info such as the location and size of code. There are some basic sub-sections or I may say as characteristics defined in the header section itself; they are listed below:

* Signature
* Machines
* NumberOfSections
* SizeOfOptionalHeader
* Image\_Optional\_Header
* ImageBase
* SectionAlignment
* FileAlignment
* MajorSubSystemVersion
* SizeOfImage

**Execution Plan:**

1. Implementation of all dependencies.
2. Adding and training the dataset models
3. Data Preprocessing.
4. Feature extraction from binaries (PE Parameters) using PE file.
5. Feature selection to identify independent variables which will be relevant.
6. Selection of Classification machine learning model.