

Prevention of Attacks using One-Class Classification and Autoencoders

Rohan Allen 18BCI0247
Harshita Pundir 18BCI0192
Rakshith Sachdev 18BCI0109

Introduction

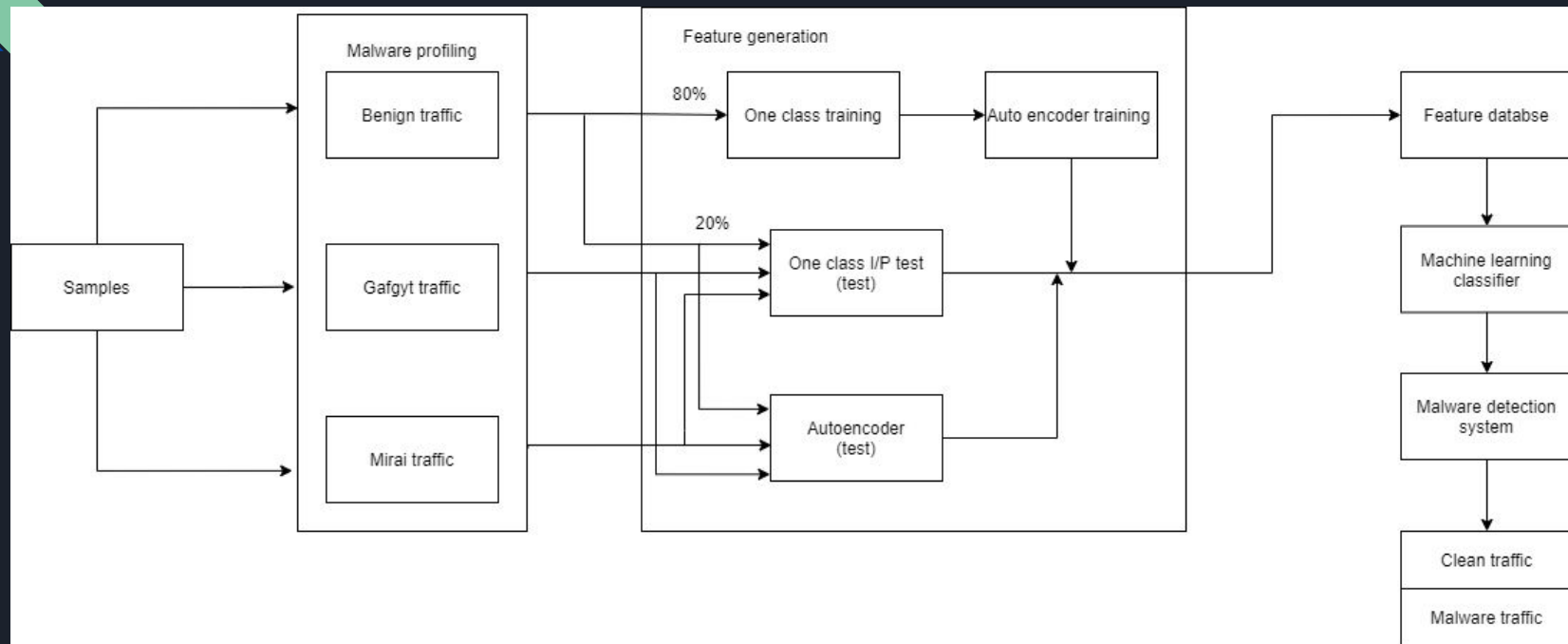
The idea of this project is to train our system to differentiate between bad network traffic which might contain some virus, malware to the system, or any other type and normal network using machine learning. The model has also been trained to predict false data and henceforth prevent the installation of any particular software during the risk of an attack, which results in an increased cost. The main key objective is to provide maximum accurate results by using unsupervised learning/ one class-based modeling approaches and reducing the processing time significantly.



- In this model, we are using one-class classifications and autoencoders so that the system can detect bad traffic more accurately. With the help of this model, we can predict false data, and therefore, we can prevent the installation of software at the time of any risks to decrease the cost.
- result. Now as this model helps us to predict false data during a risk, this model is better than other models as other models prediction is not that accurate but with this model, even with a very large dataset, we can ensure that this technique is really very beneficial for preventing real-time attacks.
- One of the merits of this model is its simplicity, it consists of only a single hidden layer of 8 neurons making it really very easy to train and especially suitable for online learning.

Architecture





Methodology





ONE CLASS SVM

One-class SVM is an unsupervised algorithm that learns a decision work for curiosity identification: ordering new information as comparative or diverse to the preparation set. One-class classification algorithms are often used for binary classification tasks with a severely skewed class distribution. These techniques are used to fit on the input examples from the huge class within the training dataset, then evaluated on the remaining test dataset. Albeit not designed for these types of problems, one-class classification methods are frequently successful for unbalanced classification datasets with no or few instances of the minority class, or datasets with no cohesive structure to distinguish the classes that a supervised algorithm would learn. The SVM algorithm, which was originally designed for classification tasks, is frequently employed for one-class classification.

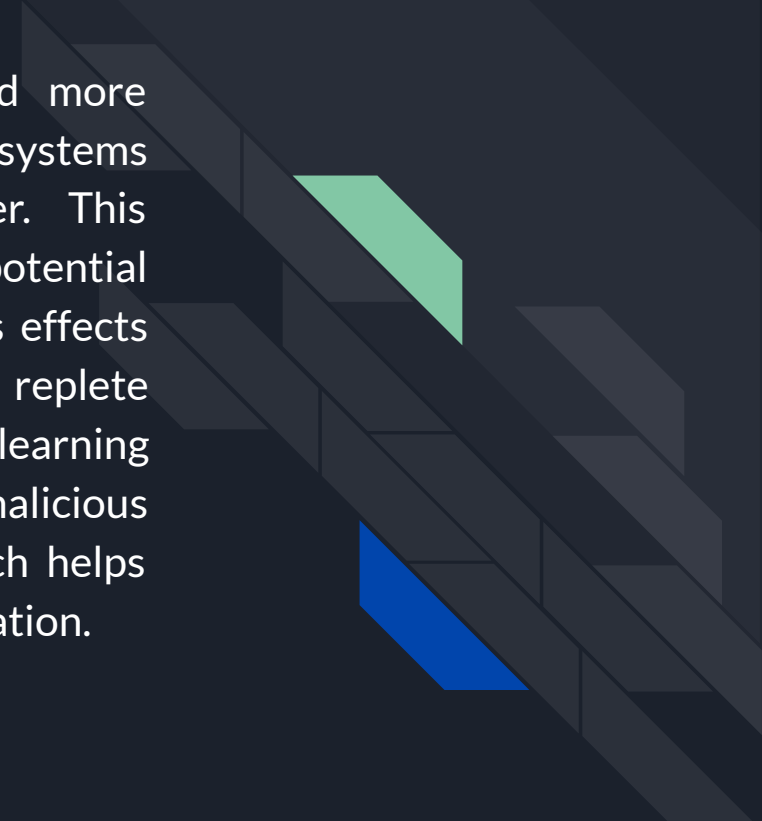



AUTOENCODERS

An autoencoder neural network is a type of unsupervised machine learning technique that uses backpropagation to adjust the target values to the inputs. Autoencoders have a habit of condensing the dimensions of our inputs into a more compact representation. If the first data is required, the condensed data will be used to recreate it. The purpose of an autoencoder is to train the model to ignore signal noise in order to discover a symbol for a set of knowledge, generally to reduce dimensions. They work by condensing the input into a latent-space description and then reconstructing the outcome from there. We'll visualize our findings and results as part of the analysis once both of these models have been implemented and trained, and we'll try and compare the trained models to see which one gives an exact result.

Disaster Recovery

Nowadays with the rise of IOT devices more and more organizations have a massive amount of computer systems and devices all interconnected to each other. This exponentially increases the surface area for any potential cyber attack which can have absolutely disastrous effects on any business. Our intrusion detection system replete with the latest machine learning and deep learning algorithms is able to proactively detect any malicious attacks with a very high degree of accuracy, which helps enhance the general security posture of the organization.



- 
- Businesses use information technology to quickly and effectively process information. Employees use electronic mail and Voice Over Internet Protocol (VOIP) telephone systems to communicate. Electronic data interchange (EDI) is used to transmit data including orders and payments from one company to another. Servers process information and store large amounts of data. Desktop computers, laptops and wireless devices are used by employees to create, process, manage and communicate information.
 - Businesses large and small create and manage large volumes of electronic information or data. Much of that data is important. Some data is vital to the survival and continued operation of the business. The impact of data loss or corruption from hardware failure, human error, hacking or malware could be significant. A plan for data backup and restoration of electronic information is essential.



Conclusion

The one class classifiers we used in the project was used for both training and testing set with which it will segregate malicious and good traffic which comes through the network. To increase the efficiency of the one class classifiers we implemented a model using auto encoders which uses deep learning neural networks. By the use of both the algorithms we increased the efficiency of the project and we also attempted to overcome the problems that exists in the datasets, namely the class imbalance issue and the data being unrealistic, by avoiding the attacks data during training, the model was trained only using normal traffic, so it was not affected by the class imbalance of the dataset. Another strength of this approach is its simplicity, it consists of only a single hidden layer of 8 neurons making it very easy to train and especially suitable for online learning. During evaluation we avoided human manipulation of the threshold in order to achieve reproducible results without human interference