# DETECTION OF SMART GRID ATTACKS USING MACHINE LEARNING TECHNIQUES

A Project Report Submitted in partial fulfillment of the requirements for the award of the degree of

## BACHELOR OF TECHNOLOGY

**IN**

## COMPUTER SCIENCE & ENGINEERING

Submitted by

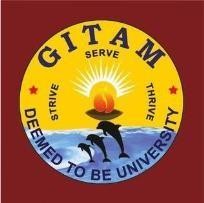
## B. Pranitha Reddy, 121810306011

## 

## Under the esteemed guidance of

**Dr. P. ANURADHA**

**Assistant professor-CSE**



## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

**GITAM**

**(Deemed to be University)**

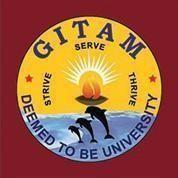
**VISAKHAPATNAM**

**MARCH 2022**

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GITAM INSTITUTE OF TECHNOLOGY

**GITAM**

**(Deemed to be University)**



## DECLARATION

We hereby declare that the project report entitled “**DETECTION OF SMART GRID ATTACKS USING MACHINE LEARNING TECHNIQUES**” is an original work done in the Department of Computer Science and Engineering, GITAM Institute of Technology, GITAM (Deemed to be University) submitted in partial fulfillment of the requirements for the award of the degree of B.Tech. in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree or diploma.

Date: 31-03-2022

|  |  |  |  |
| --- | --- | --- | --- |
| **Registration No(s).** | **Name(s)** | | **Signature(s)** |
| 121810306011 | B. PRANITHA REDDY | |  |
|  | |
|  | |
|  |  | |  |

****

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | Abstract | | 1 |
| 2. | Introduction | | 2-3 |
| 3. | Literature Review | | 4-5 |
| 4. | Problem Identification & Objectives | | 6-7 |
| 5. | System Methodology | | 8-13 |
| 6. | Overview of Technologies | | 14 |
| 7. | Implementation | | 15-21 |
|  | 8.1 | Coding | 22-24 |
|  | 8.2 | Testing |  |
| 8. | Results & Discussions | | 25-27 |
| 9. | Conclusion & Future Scope | | 28 |
| 10. | References | | 29 |

**TABLE OF CONTENTS**

# 1. ABSTRACT

## As cybersecurity is becoming a major concern in Smart Grid Technology, it is important to reduce the vulnerabilities in it. Experiments reveal that machine learning techniques outperform traditional attack detection algorithms in detecting attacks. Machine learning approaches were used to examine malicious activity and intrusion detection challenges at the network layer of smart grid communication systems. In this project, machine learning algorithms are used to classify the measurements as either being attacked or secured. The proposed system aims to predict the false data injections in the smart grid by using various machine learning algorithms such as Perceptron, Logistic Regression, Support Vector Machine and KNN algorithm.

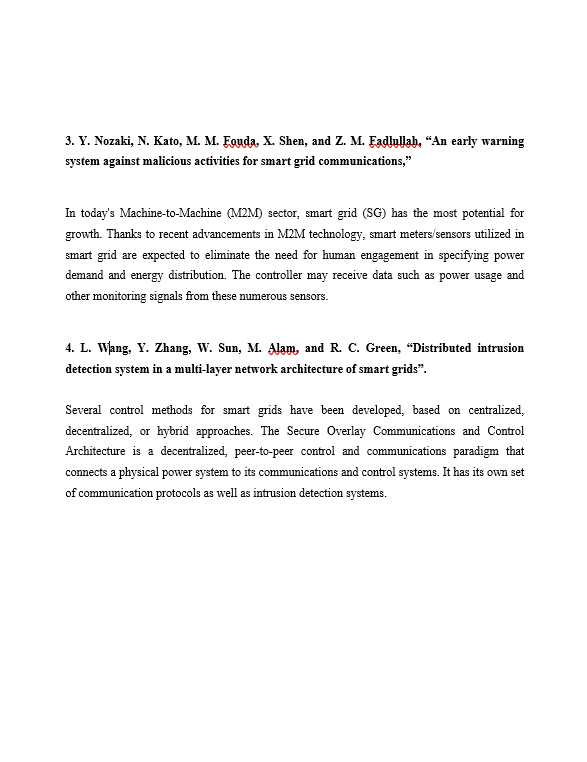
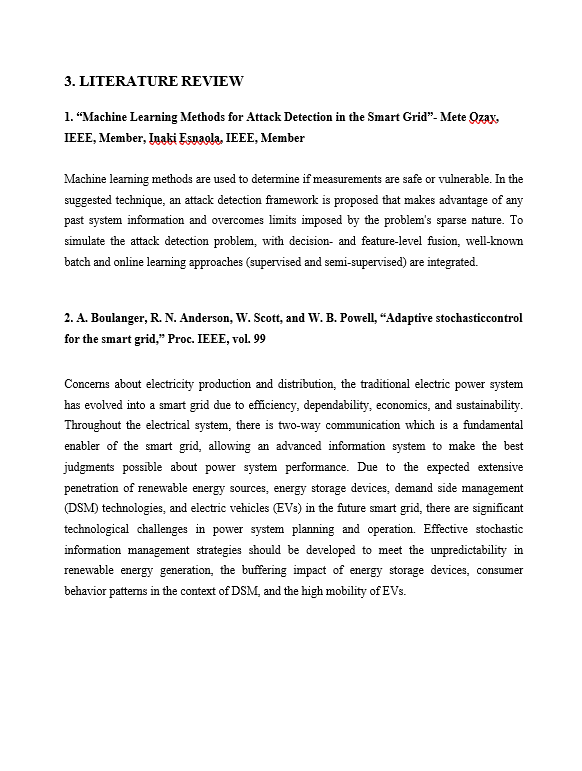
# 2. INTRODUCTION

Machine learning methods for power system monitoring and control have been frequently proposed in the literature on smart grids. Rudin presents an intelligent system design paradigm that use machine learning techniques to predict system component failures. Anderson uses machine learning techniques to control energy in smart grid networks by managing demands and sources. Machine learning approaches were used to examine malicious activity and intrusion detection challenges at the smart grid communication system’s network layer.

This project will focus on identifying fake data injection attacks on the smart grid's physical layer. In this model, the attack is guided by introducing erroneous data into the observed local measurements by either a local carrier in a hierarchical network or a smart phasor measurement unit (PMU). The measurements are organized into clusters in this way.

The system's state is determined initially from the observed data in attack detection approaches that use State Vector Estimation (SVE). The exact recovery of state vectors in sparse networks, where the Jacobian measurement matrix is sparse, is a challenge for SVE-based techniques. The problem can be solved using sparse reconstruction methods, the sparsity of the state vectors, however, limits the performance of this technique.

Ozay, who predicted fake data injection attacks using supervised learning methods, provides a full review of the methodologies. The smart grid's validity of statistical learning theory's major assumptions is also investigated. Semi-supervised and online learning techniques, as well as fusion techniques at the decision and feature levels, are then offered in a generic attack construction framework for use in hierarchical and topological networks in a variety of attack scenarios.



# 4. PROBLEM IDENTIFICATION AND OBJECTIVES

## EXISTING SYSTEM

In certain smart cities, the smart grid controls the use of power, the opening and closing of doors, and other operations. Some malicious users may attempt to target this smart grid system in order to propagate or inject fake information, and the smart grid may be activated based on this false information, potentially resulting in significant financial loss. Existing techniques for detecting such attacks used state vector estimation to check if values are within a given threshold; if values are outside of that threshold, it is considered an attack. However, this technique is not reliable for detecting attacks when values are mixed with genuine and attack data, as it is insufficient to detect unobserved data mixed with observe data.

## DISADVANTAGES OF EXISTING SYSTEM

Existing techniques for detecting attacks used state vector estimation to check if values are within a given threshold, and if they are outside of that threshold, it was considered an attack. However, this technique is not reliable for detecting attacks when values are mixed with genuine and attack data, as it is insufficient to detect unobserved data mixed with observe data.

## PROPOSED SYSTEM

We use a variety of machine learning algorithms to detect harmful attacks from both observed and unobserved data in order to avoid dangerous assaults. Because machine learning algorithms are good at noticing unobserved data mixed with observe data, they can predict fake injection attacks combined with observed data.

## ADVANTAGES OF PROPOSED SYSTEM

The alarm system allows the smart grid control center to predict various hostile occurrences, allowing SG to react ahead of time and limit the potential consequences of the malicious behavior. We use computer simulations to test the usefulness of the suggested early warning system.

# 5. SYSTEM METHODOLOGY

# SYSTEM DESIGN

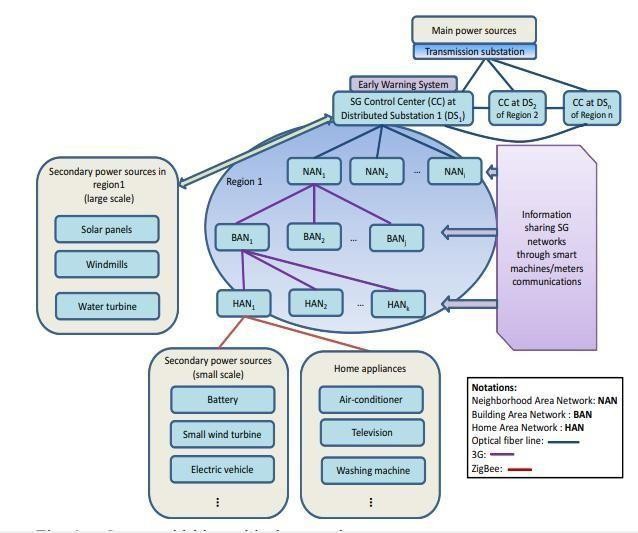
The technical heart of the software engineering process is software design, which is employed regardless of development methodology or application area. The first step in the creation of any technical product or system is design. The goal of the designer is to build a model or representation of entity which will subsequently be developed.

The aim of design is identifying the components of software and their interactions. By considering the software structure and offering the blue print, one can provide a blueprint for the phase of the document. One of the desired characteristics of large systems is modularity. It signifies that the system is divided into several parts. In this way, the relationship between the parts is presented in the simplest possible way.

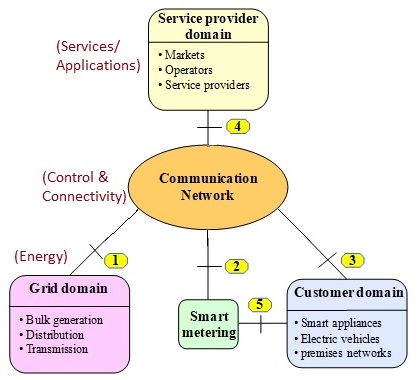
Developers are responsible for bridging the gap between the requirements specification established during requirements elicitation and analysis and the system delivered to the user during system design activities. Design is the area where excellence is promoted through progress. Software design is the process of translating the requirements into the software’s visual representation.

## SYSTEM ARCHITECTURE

**Figure 1**



**Figure 2**

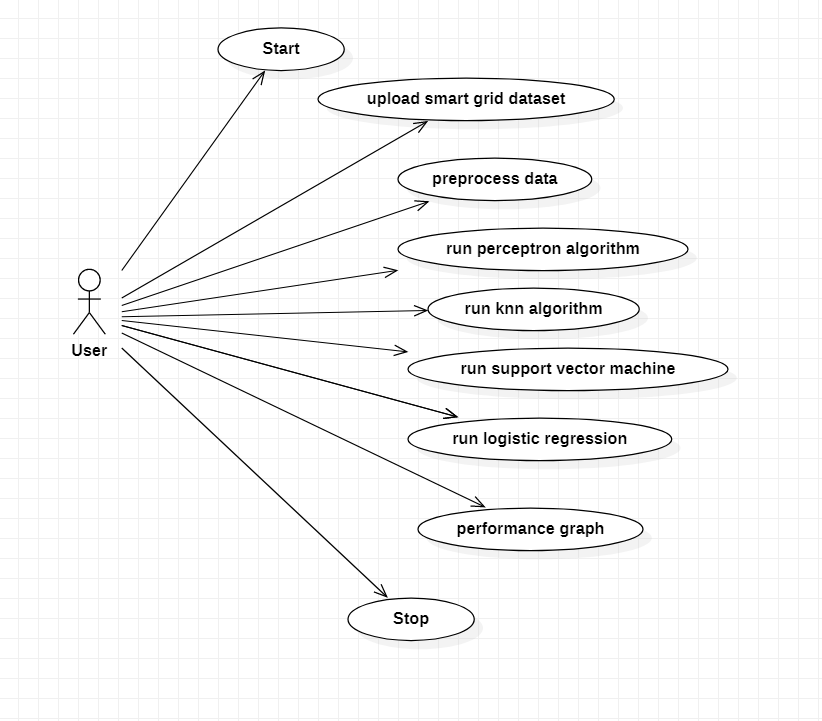


## UML Diagrams:

1. **USE CASE DIAGRAM**

Use case diagrams are used to depict and document a system's behavior and requirements. A system's high-level functionality and scope are depicted in use-case diagrams.

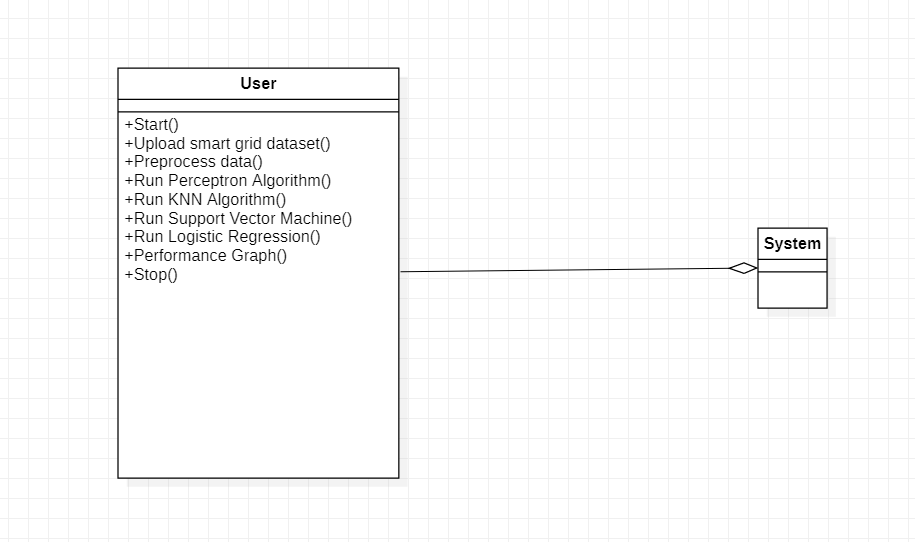
**Figure 3**

****

## 2. CLASS DIAGRAM:

The class diagram represents the attributes of class and its activities, as well as system constraints. Class diagrams are the UML diagrams that are translated directly into an object oriented language, they are employed frequently for modelling object oriented systems.

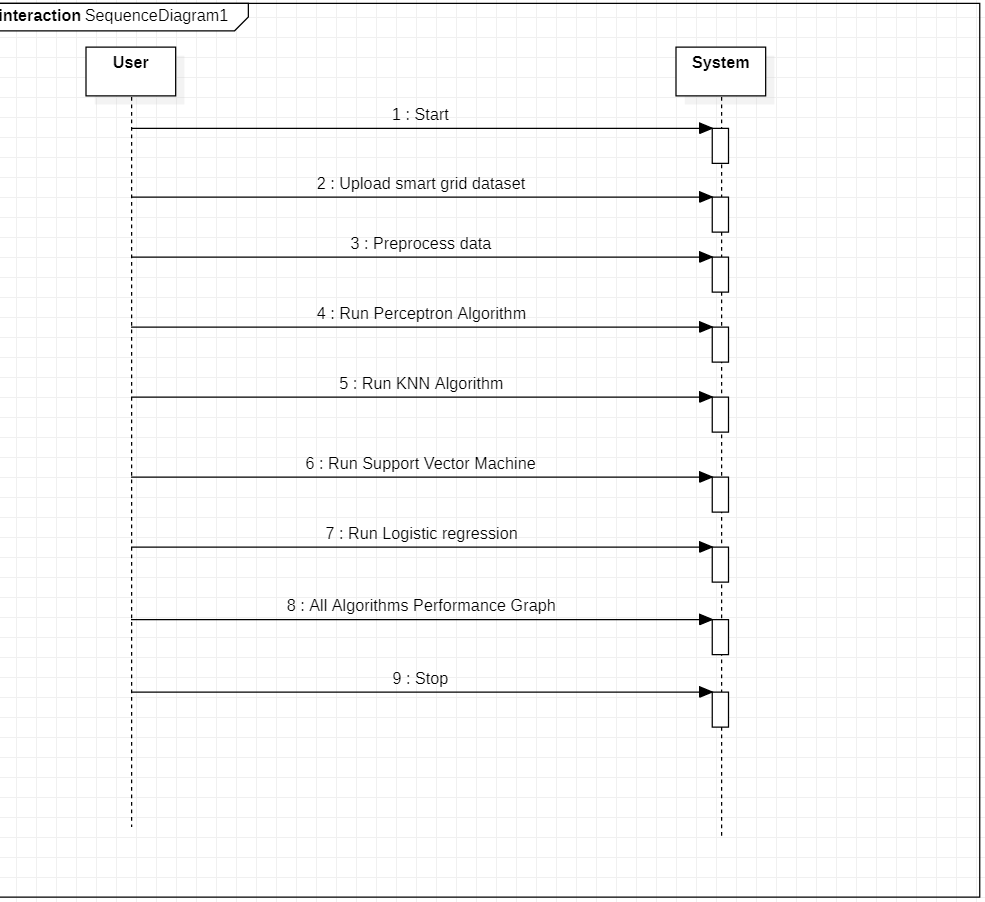
**Figure 4**

****

**3. SEQUENCE DIAGRAM:**

The sequence diagram depicts the messages flow through the system. It may be used to look at various dynamic settings. The communication between the lifelines is shown as events sequence that occurs in an order.

**Figure 5**



# 6. OVERVIEW OF TECHNOLOGIES

## SOFTWARE REQUIREMENTS

Operating system **:**Windows 7,8,10 Ultimate, Linux, Mac

Front-End **:**Python

Software Environment **:**Sublime Text and Anaconda (Jupyter)

Coding Language **:**Python

## HARDWARE REQUIREMENTS

System : Intel I-5, 3, 7 processor

Floppy Drive : 1.44 Mb

Hard Disk : 500 GB

Ram : 4Gb

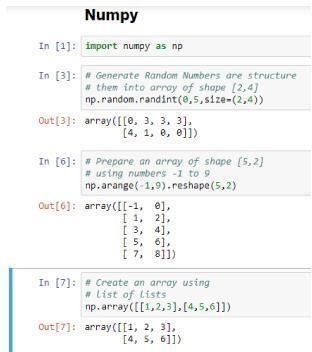
# 7. SYSTEM IMPLEMENTATION

**MODULES**

## 1. Numpy

Python provides a wide range of data types and data structures from which to pick. It wasn't, however, designed with Machine Learning in mind. Python is a computer language that uses NumPy (pronounced as num-pee). Numpy is a data management package that lets us work with multi-dimensional arrays and a wide range of mathematical functions.

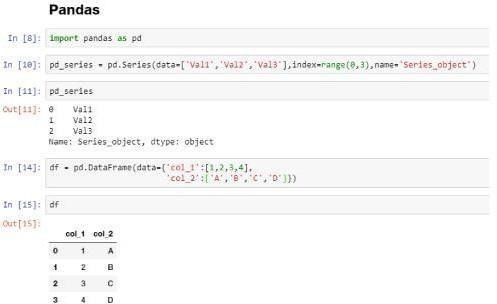
**Figure 6**



## 2. Pandas

## Pandas usually writes and reads the data from different formats including CSVs and other file systems. It includes features such as adding, updating, and deleting columns, merging or splitting data frames/series, working with datetime objects, imputation null/missing values, working with time series data, and converting to and from NumPy objects, among others.

## Figure 7



## 3. Matplotlib

## 

Matplotlib, another component of the SciPy stack, is a visualization package. It's compatible with NumPy objects (and its high-level derivatives like pandas). Matplotlib is a plotting environment similar to MATLAB that may be used to create high-quality figures/charts for papers, notebooks, web applications, and other applications.

**Figure 8**



## 4. Scikit-Learn

Scikit learn provides a suite of tools for machine learning and statistical learning that includes clustering, regression and dimensionality reduction using a python interface. The foundations of sklearn are mostly written in python.

**5. Tkinter**

Tkinter is used to create Graphical User Interfaces (GUI) and mostly commonly used.It is the fastest and easiest way to create a GUI .

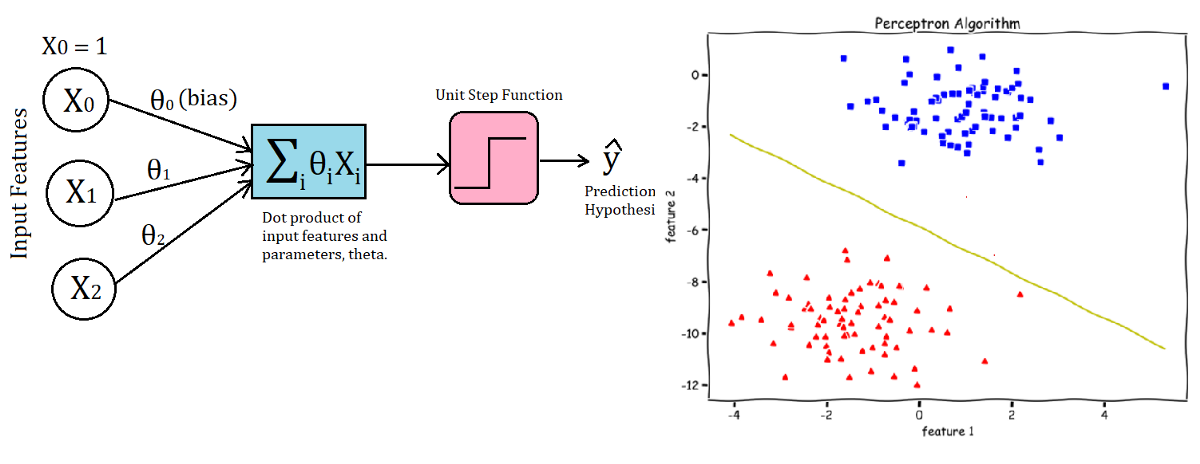
## PERCEPTRON LEARNING ALGORITHM

The Perceptron is a basic ANN unit inspired by the information processing of neurons. Before being transmitted to the activation function, the Perceptron, like the neuron, receives input signals from instances of training data that are weighted and blended in a linear equation (transfer function). It returns 1 if the result exceeds a particular threshold; else, it returns -1. In other words, if perceptron is given inputs ranging from x1 to xn, perceptron's output is



where the contribution of input xi to the perceptron output is determined by each wi, which is a weight.

**Figure 9**



## K- Nearest Neighbor (KNN) Algorithm

K nearest neighbor is a supervised learning approach that is used to resample datasets. It is used to predict the class of a new datapoint.

**Figure 10**



## Logistic Regression

Logistic regression comes under the supervised learning approach and is a popular algorithm. This algorithm is used to predict the dependent variable that is categorical from various independent variables.

The outcome can either be a discrete value or categorical value. The output may be no or yes, 1 or zero, false or true, etc.

**Figure 11**



**Support Vector Machine Algorithm:**

The main aim of the SVM algorithm is to decide the optimal line for classifying n dimensional space into the classes so that the subsequent data points may be placed easily in the relevant category. Most ideal decision boundary is known as a hyperplane.

**Figure 12**



# CODE:

# 

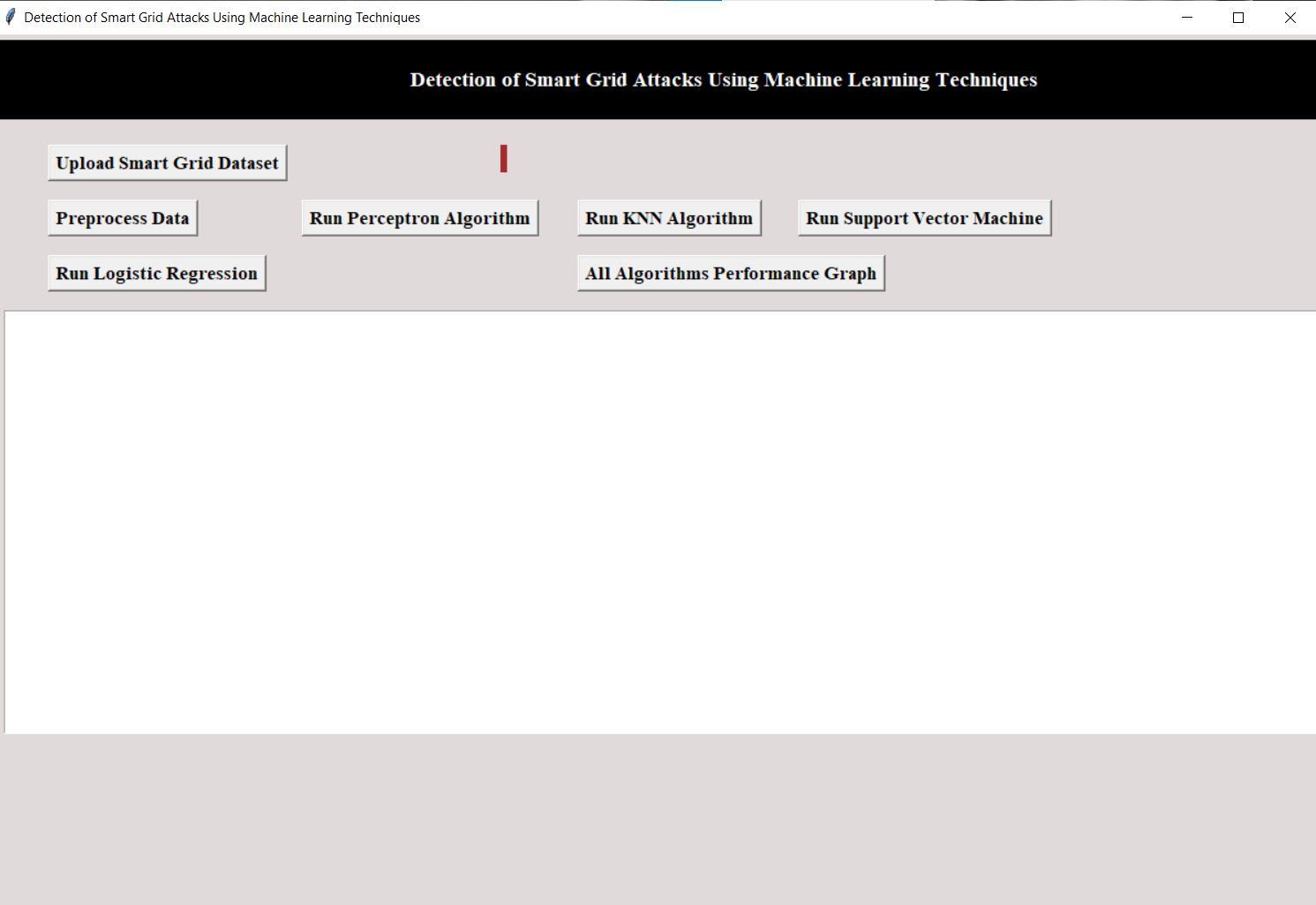
# 

# 

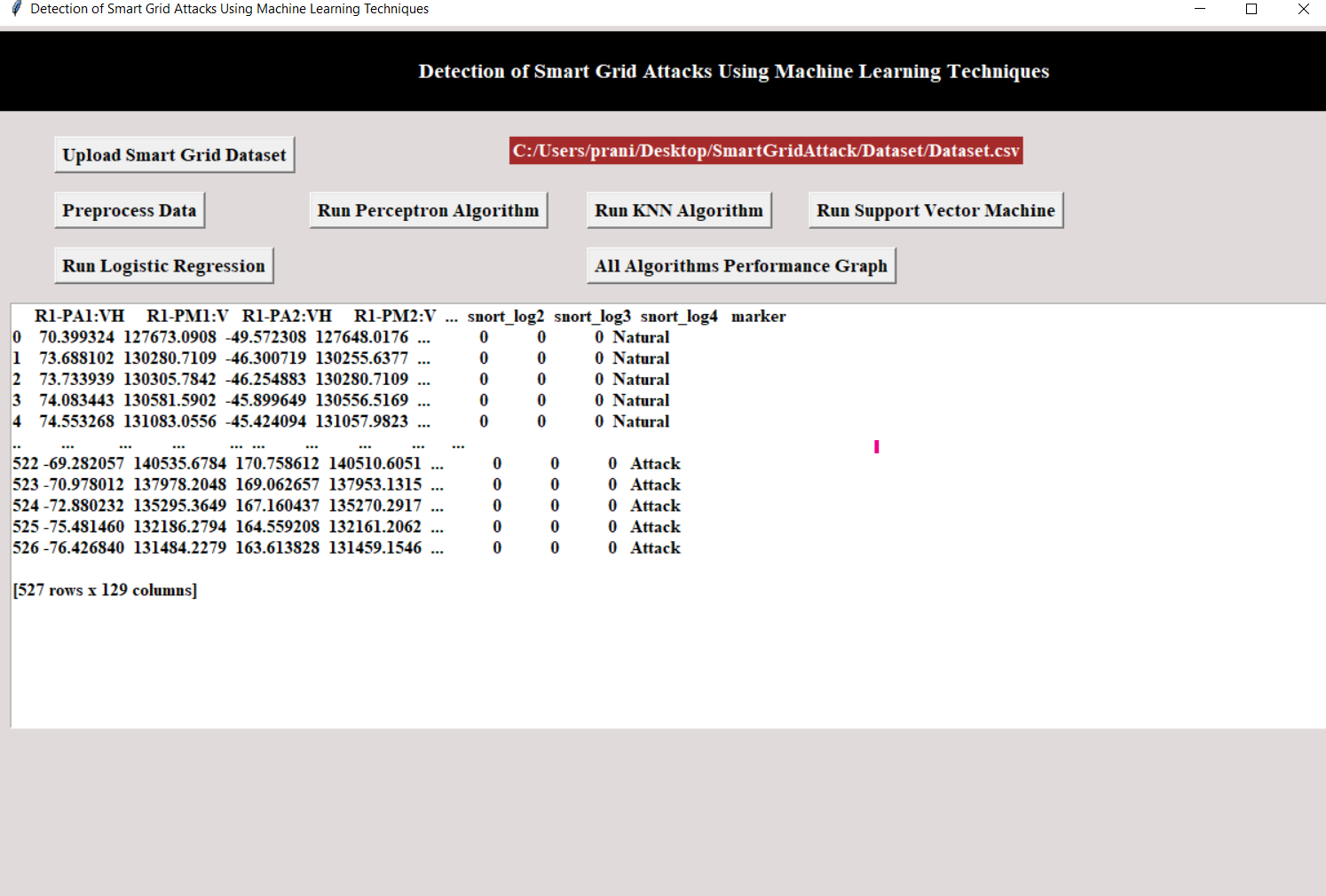
# 

# 

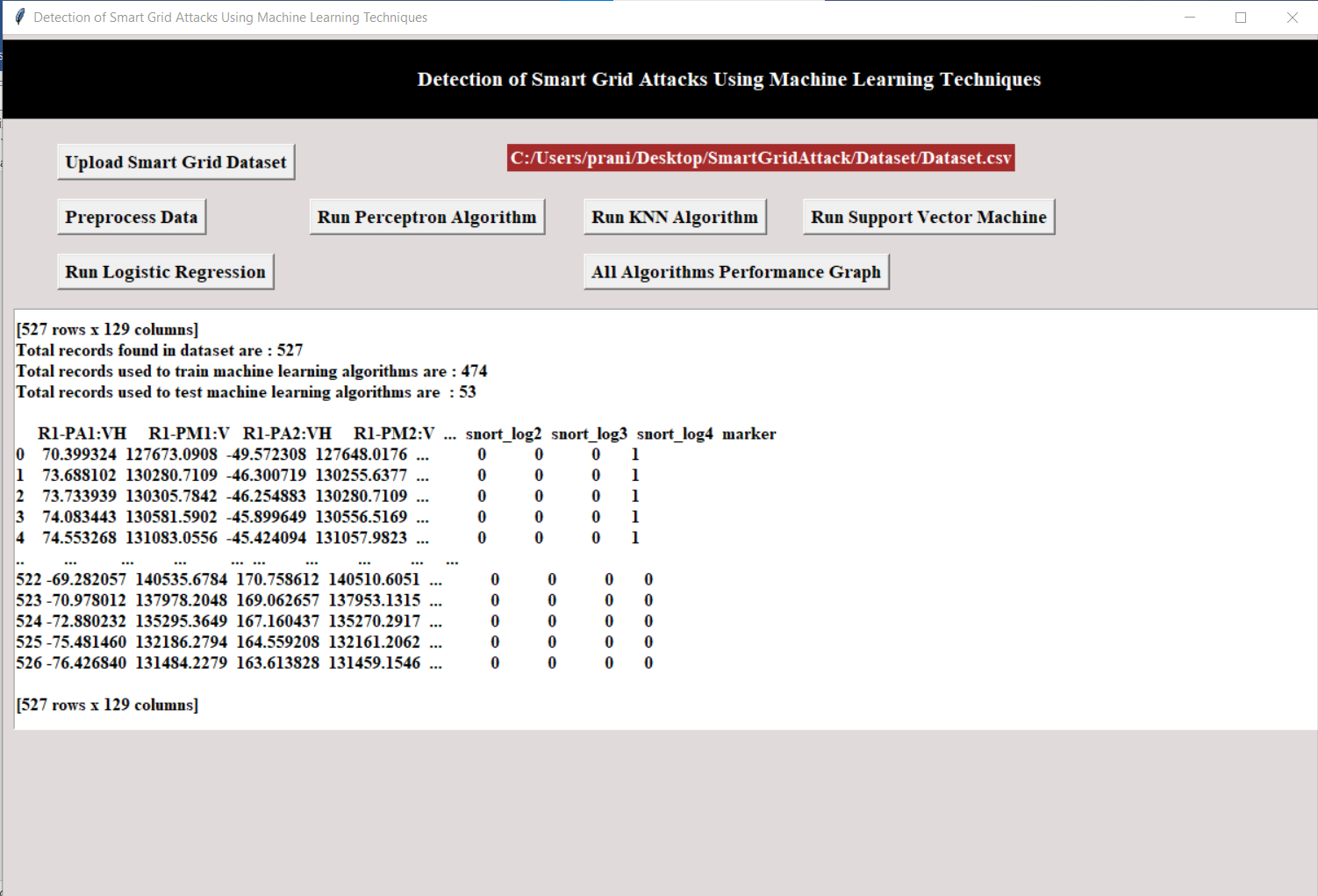
# 8. RESULTS



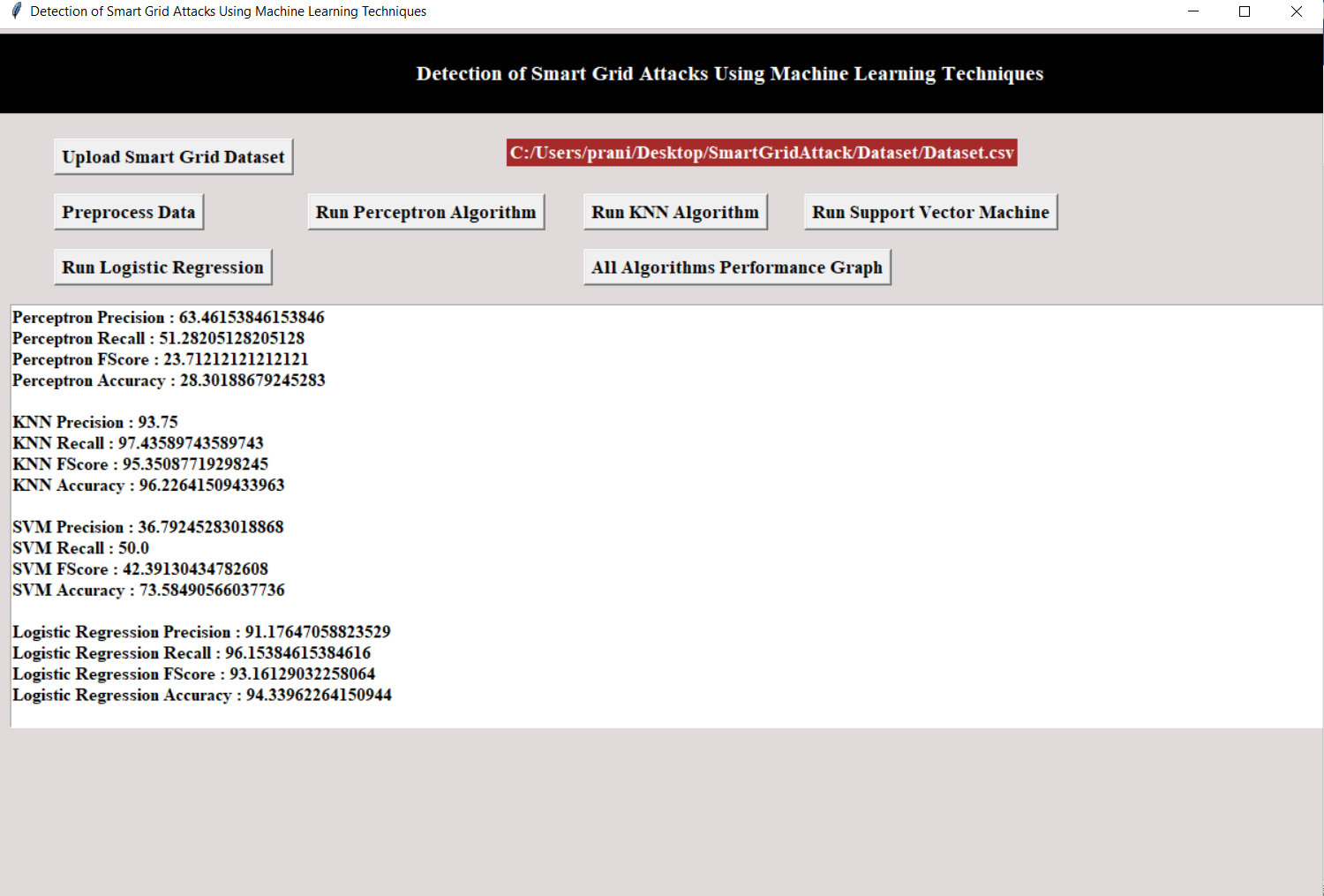
**After uploading Smart Grid dataset**



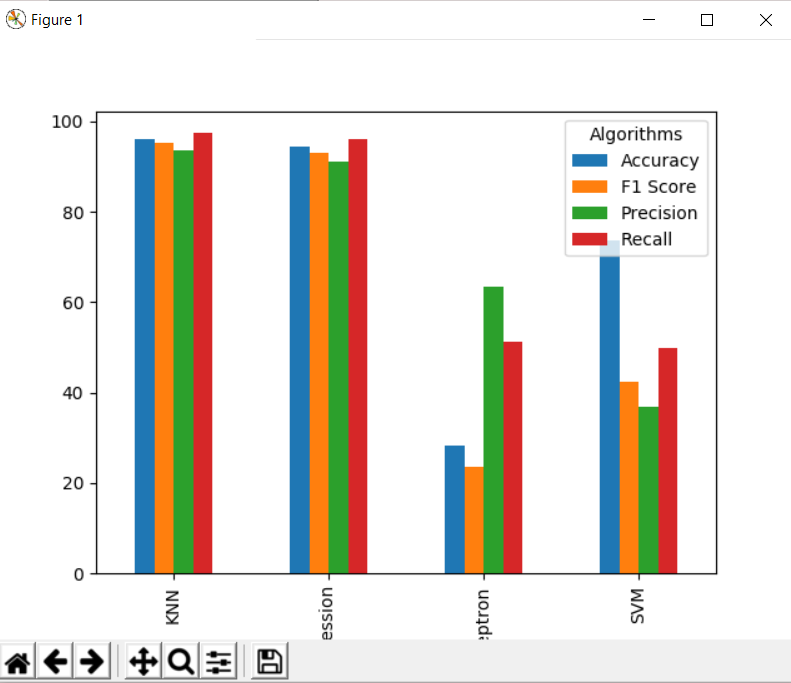
**Data Preprocessing:**



**Algorithms Output:**



**Performance Graph of All Algorithms:**



**9. CONCLUSION AND FUTURE SCOPE**

It is found that the perceptron is less susceptible to system size and the k-NN approach has a higher sensitivity than the other methods. The performance of K-NN's is harmed by the imbalanced data problem. As a result, when compared to other algorithms, in small systems, k-NN may perform better, but in big systems, it may perform worse. SVM outperforms the other techniques in the large scale systems. We've presented a system for predicting malicious assaults that could occur in the future of smart grids. A probabilistic distribution is used by the system to predict if an irregular style of operation would cause smart grid connections to be interrupted. The results of simulations reveal that the proposed system is capable of forewarning damaging threats. The suggested approach may also detect other hostile threats and anomalies allowing the control center to swiftly instruct smart meters to respond to such irregularities. We must first establish a baseline for calculating mistakes in smart grid networks with background traffic to identify actual aberrant behaviors. These are the topics that will be investigated more in the future.

# 