**Title (Font: Bookman Old Style, Size: 14)**

[Remaining text of the article is 10 size of same font and 1.5 lines as line spacing] Remove it

**Guide Name1**, Designation, Department of CSE,

Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

**Student Name2**, **Student Name3**, **Student Name4**, **Student Name5**

2,3,4,5 UG Students, Department of CSE,

Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

**1 E-Mail ID:** [**name@gmail.com**](mailto:name@gmail.com) **(Corresponding Author)**

**Abstract**

In order to locate the leakage fault location of low-voltage electrical lines, cut off the power supply of electrical lines, and ensure the safety and reliability of the power supply network. This paper proposes an electrical line leakage fault detection system and to detect the electric city theft in a particular area. In which the customer and government is facing the quality, increasing generation load, affecting the overall economy of the government etc. For lack of technology and data, insecurity, the complexity associated with traditional methods, untimeliness, and general human cost, electricity assets monitoring, and management have remained a big problem in many developing countries. By using proposed system, it can significantly reduce this loss by analysing the electric reduce this loss by analysing the electricity consumption. In this paper by using machine learning technique employing DBSCAN, K-MEANS Algorithm for detection. The electricity theft approach in a particular area by using the graphs and Traditional Grid (Pie table) for the prediction. The data collected is analysed and pre-processed before it is used for model training and testing. This study also provides the guide to various considerations when adopting this technology in terms of the choice of machine learning architecture, adequate training samples over multiple fault characteristics, effects of data augmentation, and balancing of intra-class heterogeneity.

**Keywords:** DBSCAN, Electricity Theft, Fault detection, K-Means, Machine Learning,

**Introduction**

Science and technology with all its wonderful progress has fascinated human life to a great extent that imagining a world without these innovations is hardly possible. While technology is on the rise, we need to: also note the increasing immoral activity. With a technical view, “Power Theft” is a crime not to be ignored that is highly common, and at the same time has a direct impact on the economy of a nation. This project is designed to discover such a power theft in the normal distribution lines. Even though there are certain practical difficulties in implementing this kind systems in the future, there is room for the development of these kind of systems. This project uses the principle of the differential protection scheme for identification of the power theft. Recently, Nigeria’s development agenda has been anchored in a vision that identifies energy as one of the vital infrastructural enablers for development. With the realisation that for them to successively make a significant positive transition in development, there must be an efficient, reliable, vast, and environment-friendly energy source, transmission, and distribution. This means that majority of the burden of energy demand is on power companies to provide and transmit quality energy services to consumers. Against this backdrop, investors in their transmission lines need accurate, cost and time-efficient methods to carry out existing asset inventory of the transmission lines for well-informed decisions and investment. In view of this, our hypothesis is that deep learning on high-resolution Aerial images should provide a cost and time-effective solution for power line assets inventory and studies, study fields.

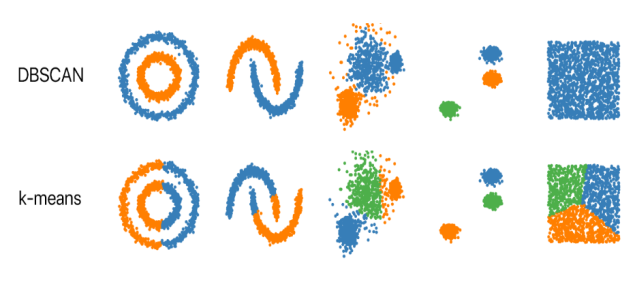


Fig.1. Description

**Fault Type Classifiers**

Conventional fault detection algorithms are designed based on current or voltage magnitude measurements [2,3]. Increase of current magnitude or decrease of voltage/impedance magnitude could be considered as a measure to detect a system fault. These algorithms are dependent on various factors such as fault resistance and power system short circuit capacity. Current based starters get confused when load current is significant compared to fault current. Conventional overcurrent-based starters may not be able to detect faults with high amount of fault resistance. For remote low current faults, no clear undervoltage condition arises at the relay location. In the case of a close-in fault on a weak system, all voltages deviate from the nominal value. Therefore, the voltage-based starters might not be able to perform correctly for different fault conditions. For the conventional based fault detectors, current and voltage magnitudes should be estimated correctly using appropriate filteringalgorithms. When a fault happens on a transmission line, the power system goes through a transient period. It might not be easy to determine current/voltage signal magnitude fast and precisely during the transient period after the occurrence of the fault. As power systems grow both in size and complexity, it becomes necessary to identify different system faults faster and more accurately using more powerful algorithms. It would be desirable to design a reliable and fast algorithm to classify different power system faults for various system parameters and fault states. ML algorithm in proposed in this paper as a transmission line fault detector and fault type classifier module.

### A. Power System

Poly phase power system is the common populated means to produce the electricity in the world. This system is proved high performance in power transmission. However, generators such as diesel generators is fuelled machines works to convert the mechanical energy into electrical energy.

**B. Transmission lines**

The conductors that carry the electric power from a known location into the other location is known as transmission lines. Commonly, two wires (conductors) are used for performing this service. The transmission of power in means of cables such as two cables are coming with another dielectric properties such as conductance and capacitance which need to be studied while selecting the cable, actually it varies and strongly dependent upon the length of conductor. The type of insulation is another concern for the power engineers, conductors and conductors guard materials can be used to isolate the conductors from the out world conditions and the guard is for the protection of conductors. The restrictions was very obvious in before years for fitting and making what is so called isolation materials, however, burden is mitigated on recent time where manufacturing of isolation and protection material is become so feasible. Studies found strong relations between the resistance of the conductors in transmission lines and the frequency of operation.

**C. Energy Consumer**

In here, three phase inductive load is used and rated of 80 MW and 60 MVar. Induction motors are the commonly used as loads with poly phases power system, they termed with their ability to provide high torque at the starting and simple design with improved efficiency. The study of [11] are listing the applications of industry where induction motors are applied successfully. The thee phase motors are grapping the attention for their designs, the performance lies their compact shape and lesser adaptation cost. The smooth operation is also one of the motivated point to adopt the induction motors, it is working smoothly with less vibration than the signal phase motors, this make it to last for longer time; it utilizes smaller place and heats slower than signal phase motors.

**System Implementation**

DBSCAN is a popular **density-based** data clustering algorithm. To cluster data points, this algorithm separates the high-density regions of the data from the low-density areas. Unlike the K-Means algorithm, the best thing with this algorithm is that we don’t need to provide the number of clusters required prior. DBSCAN algorithm group points based on distance measurement, usually the *Euclidean distance* and the *minimum number of points*. An essential property of this algorithm isthat it helps us track down the outliers as the points in low-density regions; hence it is not sensitive to outliers as is the case of K-Means clustering.

**Prerequisites**

To follow along the reader will need the following:

1. Python installed on your system or access to the Google Colab.
2. A Dataset available in the form of a CSV file.

DBSCAN algorithm works with two parameters.

These parameters are:

1. Epsilon (Eps): This is the least distance required for two points to be termed as a neighbor. This distance is known as Epsilon (Eps). Thus we consider *Eps* as a threshold for considering two points as neighbors, i.e., if the distance between two points is utmost *Eps*, then we consider the two points to be neighbors.
2. MinPoints: This refers to the minimum number of points needed to construct a cluster. We consider MinPoints as a threshold for considering a cluster as a cluster. A cluster is only recognized if the number of points is greater than or equal to the *MinPts*.

### DBSCAN Algorithm

The following are the DBSCAN clustering algorithmic steps:

* **Step 1:** Initially, the algorithms start by selecting a point (x) randomly from the data set and finding all the neighbor points within Eps from it. If the number of Eps-neighbours is greater than or equal to **MinPoints**, we consider x a core point. Then, with its Eps-neighbours, x forms the first cluster.

After creating the first cluster, we examine all its member points and find their respective Eps -neighbors. If a member has at least MinPoints Eps-neighbours, we expand the initial cluster by adding those Eps-neighbours to the cluster. This continues until there are no more points to add to this cluster.

1. **Core:** This is a point from which the two parameters above are fully defined, i.e., a point with at least Minpoints within the Eps distance from itself.
2. Border: This is any data point that is not a core point, but it has at least one *Core point* within *Eps* distance from itself.
3. Noise: This is a point with less than *Minpoints* within distance *Eps* from itself. Thus, it’s not a *Core* or a *Border*.
4. Step 2: For any other core point not assigned to cluster, create a new cluster.
5. Step 3: To the core point cluster, find and assign all points that are recursively connected to it.
6. Step 4: Iterate through all unattended points in the dataset and assign them to the nearest cluster at *Eps* distance from themselves. If a point does not fit any available clusters, locate it as a noise point.

### Python implementation of DBSCAN

As usual to any implementation, we get started with fetching the dataset and preparing it ready for our model implementation.

Step 1:- Importing Library Files

Step 2:- Input the data by Gitub Files

Step 3:-This plot will represents the data which we obtained by the bata sent by the data passed in the table

From the below plot, we note the maximum curvature of the curve is about eight, and thus we picked our *Eps* as 8. We now have our two parameters as:

MinPoints = 4

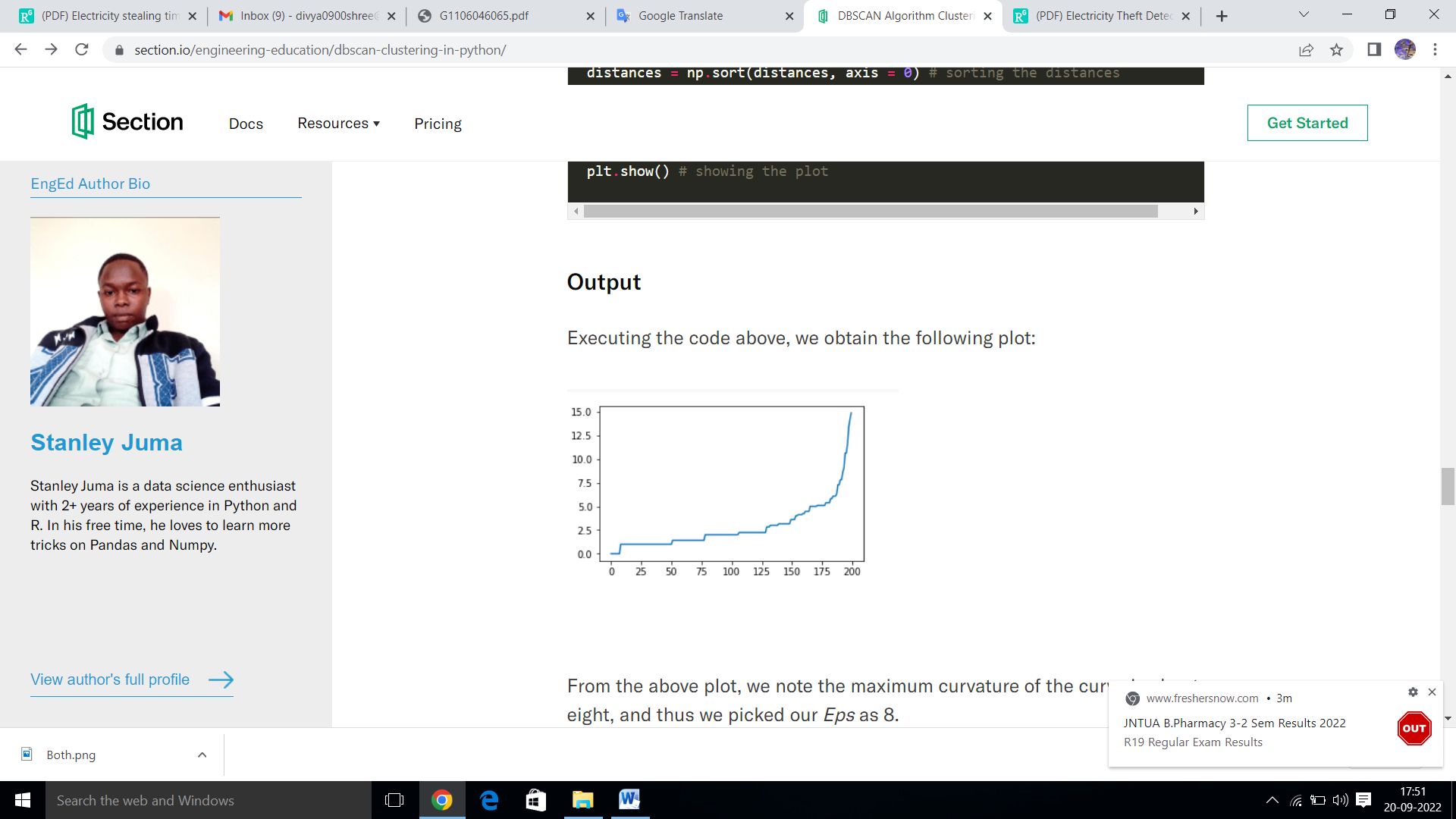


Fig.2. Description

Eps = 8 Clusters plots

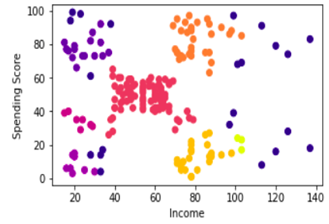


Fig.3. Description

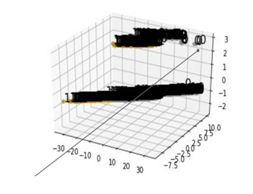
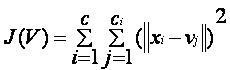


Fig.4. Description

**K- Means Algoritm**

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of theclusters resulting from the previous step. After we have these k new centroids, anew binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:



‘||xi - vj||’ is the Euclidean distance between xi and vj.

‘ci’ is the number of data points in ith cluster.

‘c’ is the number of cluster centers.

**Algorithmic steps for k-means clustering**

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select *‘c’* cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..

4) Recalculate the new cluster center using:

where, *‘ci’* represents the number of data points in *ith* cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

**Case Study**

Generator is supplying power to the load through transmission line of 400 km, in this three-phase system is used to provide 230 KV to an inductive load. Voltage is applied so that current can flow in each line of three phase system, the current in normal conditions is drawn according to the load capacity during the simulation as well as the losses introduced in the transmission lines. On other hand, at the time of fault, current with hug quantity is being drawn by the load. Fault is undesired event for the system and must be cleared upon its occurrence.

The fault should be cleared as quickly as potential. More and more equipment can be get collapsed if not quick clearance of fault begins. The negative effects of fault is commonly caused by short circuit of one or more phases with itself or with ground. More effects can be observed due to fault raisin in power system can be listed as below.

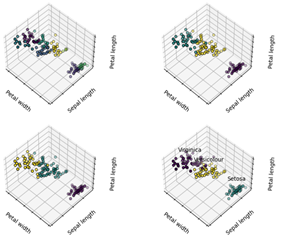


Fig.5. Description

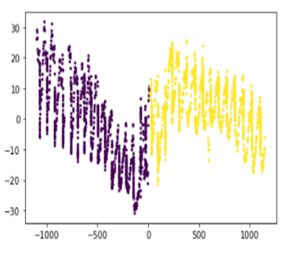


Fig.6. Description

Fault clearance is done by the control equipment lying on power system, relays and circuit breakers are the main devices used for that purpose, the new approaches to clear faults are designed to remove the faulty part quickly.The classifier will generate a message including the fault information to the controller and relay circuit which is in turn cutting of the connection between the generator and load till the fault is cleared. The control and monitoring algorithms that adopted using software and hardware integration are having great impact to recover the power system as compare to the traditional means of protection i.e. the mechanical switches. A smart fault detection system is being designed to detect the fault and classify it and ultimately separating the faulty part of power system until it is getting corrected.

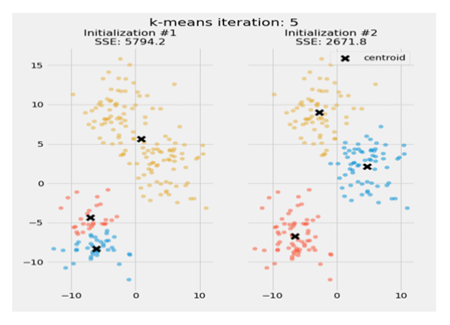


Fig.7. Description

**Conclusion**

In this article, we have developed a new method for finding the electricity theft in a perticular area by using the k-MEANS and DBSCAN Algorithum by this methods we can predect the loss of electricity in a perticular area. The protection devices may be chosen depending on the system specifications and work, so that short circuit level of each component may be calculated with care and accordingly such component can enter to service, such precautions are important to prevent any unwanted effects of error, the short circuit current can be defined as the maximum level of current amplitude in which the circuit breaker/breaking device can bare. Furthermore, if any error such as fault is took place within the system, it is mandatory to clear it immediately. A smart fault detection system is being designed to detect the fault and classify it and ultimately separating the faulty part of power system until it is getting corrected. Artificial neural network is proven means to perform fast and accurate solutions in this regard. By ausing DBSCAN algorithm we can view the flow of the current in a perticular area and the data can be stored in the big data in this projet we took the data from github where we can view the data and by using the K-means algorithm we combine the data and make a difference of noice data where we can see the loss of data and by this we can view different area at a same time we can view the difference between the electricity usage.

**References**

1. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
2. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.
3. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
4. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.
5. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
6. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.
7. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
8. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.
9. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
10. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.
11. Author Names, “Title of the article” like “Predictive Acknowledgement using TRE System to reduce cost and Bandwidth”, Journal Name, Vol. 11, Issue 3, March 2022.
12. Ashish B Deharkar “An Approach To Reducing Cloud Cost And Bandwidth Using Tre System”, Journal Name, Vol. 11, Issue 3, March 2022.