People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of 20 août 1955 - Skikda

Faculty of Sciences

Department of Computer Sciences

Specialty artificial intelligence

A project to obtain a master's degree + a start-up institution certificate within the framework of Ministerial Resolution 1275.

Project Title

Electricity theft detection model based on machine learning.

Logo

Trade name

DUANKOU

Prepared by the student:

Halais Yousra

Under the supervision of :

Dr. Bougamouza Fateh

THANKS AND GRATITUDE

With deep gratitude and praise to God Almighty, who has enabled us to complete this humble work as intended. Your countenance is majestic, and your authority is great.

الحمد والشكر الله عز وجل

I extend my heartfelt thanks to our supervising professor, Dr. Bougamouza Fateh, who has guided me throughout this project with valuable information and advice. I pray that God Almighty guides his steps and fulfills his aspirations. May God reward him abundantly.

I also express my sincere thanks to all the officials and workers of the Incubator and the Contracting House of the University of 20 août 1955 - Skikda, for their assistance and the facilities provided. May God reward them generously.

To all the professors of the Department of Mathematics and Computer Science, each in their respective positions, and to everyone who helped us, whether near or far, in completing this humble work, we extend our deepest gratitude.

Finally, we would like to thank the distinguished professors and members of the discussion committee for agreeing to discuss this memorandum.

DEDICATION

الى من ساندو خطاي المثعرة الى رمزي العطاءوالمحبة

إلى من أحمل اسمه بكل فخر أبي

إلى ملاذي الآمن مستقر روحي امي

الى ركائزي و فخري وسندي اخوتي

To my colleagues in my university career, each with his own name, led by Nabil Bourbia. You have been always a giver. Many thanks to you. May God reward you with all the best. May God bless you and yours and direct your steps toward what you desire.

To everyone who supported me with a kind word, prayer or advice To everyone who has an ounce of goodness for me in his heart May you be blessed

last but not least I wanna thank me

for beliving in me, I wanna thank me for doing all this hard work,

I wanna thank me for being always a giver.

ABSTRACT

Electricity theft poses not only considerable financial risks to utility providers but also threatens the stability and reliability of electrical grids. This growing issue necessitates innovative and effective solutions to ensure rapid and accurate anomaly detection.

In this work, we propose an application based on auto-encoders, an advanced type of artificial neural network, to detect electricity theft by identifying irregular consumption patterns. Auto-encoders are particularly well-suited to this task due to their ability to learn and represent complex data, thereby distinguishing normal consumption behaviors from anomalies.

Traditional detection methods, such as threshold-based analyses, often prove insufficient in the face of the increasing sophistication of electricity theft techniques. These approaches can lack precision and speed, leading to substantial financial losses and additional strain on electrical grid infrastructures. In response to these challenges, the use of advanced technologies like auto-encoders becomes crucial. These networks can analyze vast amounts of data in real-time, detecting subtle irregularities and abnormal consumption patterns that might indicate theft. By automating and enhancing detection, our application aims to provide utility providers with a powerful tool to protect their grids, reduce financial losses, and ensure a more secure and stable energy distribution.

Thus, our study highlights the importance of integrating machine learning technologies into the management of electrical grids and proposes a promising solution for the future of energy monitoring.

Résume

Le vol d’électricité représente non seulement des risques financiers considérables pour les fournisseurs de services publics, mais menace également la stabilité et la fiabilité des réseaux électriques. Cette problématique croissante nécessite des solutions innovantes et efficaces pour assurer une détection rapide et précise des anomalies.

Dans ce travail, nous proposons une application basée sur des auto-encodeurs, un type avancé de réseau neuronal artificiel, pour détecter le vol d'électricité en identifiant les schémas de consommation irréguliers. Les auto-encodeurs sont particulièrement adaptés à cette tâche en raison de leur capacité à apprendre et à représenter des données complexes, permettant ainsi de distinguer les comportements de consommation normaux des anomalies.

Les méthodes de détection traditionnelles, telles que les analyses basées sur des seuils simples, se révèlent souvent insuffisantes face à la sophistication croissante des techniques de vol d’électricité. Ces approches peuvent manquer de précision et de rapidité, entraînant des pertes financières substantielles et une charge supplémentaire pour les infrastructures des réseaux électriques. En réponse à ces défis, l'utilisation de technologies avancées comme les auto-encodeurs devient cruciale. Ces réseaux peuvent analyser de vastes quantités de données en temps réel, détectant des irrégularités subtiles et des modèles anormaux de consommation d'énergie qui pourraient indiquer un vol. En automatisant et en améliorant la détection, notre application vise à fournir aux fournisseurs de services publics un outil puissant pour protéger leurs réseaux, réduire les pertes financières et assurer une distribution d'énergie plus sécurisée et stable.

Ainsi, notre étude souligne l'importance de l'intégration des technologies de machine learning dans la gestion des réseaux électriques et propose une solution prometteuse pour l'avenir de la surveillance énergétique.

ملخص

لا تمثل سرقة الكهرباء فقط مخاطر مالية كبيرة لمزودي الخدمات العامة، بل تهدد كذلك استقرار وموثوقية الشبكات الكهربائية. تتطلب هذه المشكلة المتزايدة حلولا مبتكرة وفعالة لضمان الكشف السريع والدقيق عن الإختلالات.

في هذا العمل، نقترح تطبيقا قائما على أجهزة التشفير الذاتي، وهو نوع متقدم من الشبكات العصبية الاصطناعية، لاكتشاف سرقة الكهرباء عن طريق تحديد أنماط الاستهلاك غير المنتظمة. تعد أجهزة التشفير الذاتي مناسبة بشكل خاص لهذه المهمة نظرا لقدرتها على تعلم وتمثيل البيانات المعقدة، وبالتالي تمييز سلوكيات الاستهلاك الطبيعية عن التي تعتريها بعض الإختلالات.

ثبت أن الطرق التقليدية للكشف، مثل التحليلات المستندة إلى العتبات البسيطة، غالبا ما تكون غير كافية في مواجهة التعقيد المتزايد لتقنيات سرقة الكهرباء. قد تفتقر هذه الأساليب إلى الدقة والسرعة، مما يؤدي إلى خسائر مالية كبيرة وزيادة العبء على بنية الشبكات الكهربائية التحتية. استجابة لهذه التحديات، يصبح استخدام التقنيات المتقدمة مثل أجهزة التشفير الذاتي أمرا بالغ الأهمية. يمكن لهذه الشبكات تحليل كميات هائلة من البيانات في الوقت الفعلي، واكتشاف الإختلالات الطفيفة وأنماط الاستهلاك غير الطبيعية التي قد تشير إلى سرقة. من خلال تألية وتحسين عملية الكشف، يهدف تطبيقنا إلى تزويد مزودي الخدمات العامة بأداة قوية لحماية شبكاتهم، وتقليل الخسائر المالية، وضمان توزيع طاقة أكثر أمانا واستقرارا.

وبذلك، تسلط دراستنا الضوء على أهمية دمج تقنيات التعلم الآلي في إدارة الشبكات الكهربائية وتقدم حلا واعدا لمستقبل مراقبة الطاقة.

CONTENTS TABLE

Figure list

INTRODUCTION

Introduction

Electricity theft is a pervasive problem that affects utility companies and consumers worldwide. This illegal practice, which involves tampering with meters, bypassing them altogether, or illicitly reconnecting disconnected service, results in significant financial losses and operational challenges for utility providers. Additionally, it can pose serious safety risks to the public and the thieves themselves.

The economic impact of electricity theft is staggering. Utility companies face billions of dollars in losses annually due to unbilled consumption, which ultimately leads to higher costs for all customers as providers attempt to recoup their losses. This also hinders investments in infrastructure improvements and the maintenance of a reliable power supply. The ripple effects extend beyond financial losses, contributing to inefficiencies and unfair pricing in the energy market.

Traditional methods of detecting electricity theft, such as manual inspections and audits, are labor-intensive, time-consuming, and often ineffective in catching sophisticated theft techniques. However, advancements in technology, particularly in artificial intelligence (AI), offer new hope in tackling this age-old problem.

The application of deep learning autoencoders in electricity theft detection offers several advantages. By analyzing vast amounts of data from smart meters, billing systems, and customer records, these AI models can identify unusual consumption patterns that suggest theft. This automated process is not only faster but also more accurate than traditional methods, significantly enhancing the efficiency of theft detection.

For instance, our DUANKOU web application leverages the power of deep learning to combat electricity theft. It analyzes time-series data of electricity usage, demographic information, and environmental factors to detect anomalies indicative of theft. This sophisticated approach allows utility companies to quickly and accurately identify suspicious activities, reduce losses, and allocate resources more effectively.

CHAPTER1

ARTIFCIAL INTELLEGENCE

Introduction

Artificial intelligence (AI) encompasses the simulation of human intelligence by machines, including machine learning (ML) and deep learning (DL). ML involves algorithms that enable computers to learn from data, making predictions or decisions without being explicitly programmed. DL, a subset of ML, utilizes deep neural networks to learn complex patterns from large datasets, achieving remarkable success in various domains. Together, these fields advance the development of intelligent systems capable of perceiving, reasoning, and acting autonomously to solve diverse problems.

In this chapter, we will explore various artificial intelligence techniques commonly employed for pattern recognition. These techniques are categorized into two main types: machine learning and deep learning, with a special focus on Autoencoders within the deep learning category.

1. Machine learning

ML is a field of study that gives computers the ability to learn without being explicitly programmed [1]. These algorithms build models based on data training, also called "training data", in order to perform predictions or decision-making without the need for explicit programming this effect. Machine learning algorithms are widely used in various fields such as medicine and computer vision [2].

1.1 Learning types

Machine learning can be divided into several main categories: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and self-supervised learning.

1.1.1 Supervised learning

Supervised learning is a machine learning task that involves learning a function to map inputs to outputs based on example input-output pairs. It infers this function from labeled training data, which consist of training examples. Supervised learning algorithms require external guidance and typically involve dividing the input data set into a training set and a test set. The training set includes an output variable that needs to be predicted or classified. Algorithms learn patterns from the training set and apply these learned patterns to the test set for prediction or classification [3].

1.1.2 Unsupervised learning

Unsupervised learning occurs without the use of data labels. In this approach, the model learns internal representations or important features to discover unknown relationships or structures in the input data. Techniques such as clustering, dimensionality reduction, and generative models are often considered unsupervised learning methods [4].

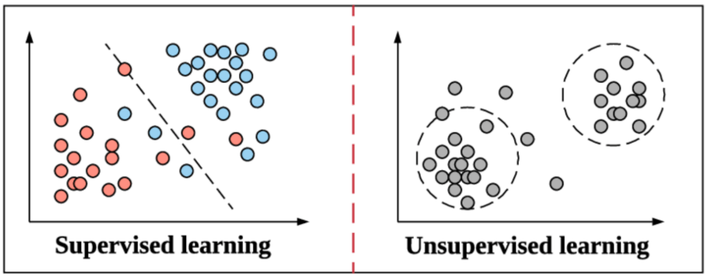


FIG 1.1 : Supervised and unsupervised learning.

1.1.3 Semi-Supervised Learning

Semi-supervised learning employs a combination of a small amount of labeled data and a large amount of unlabeled data to improve model performance while reducing the need for extensive labeled datasets. Initially, the model is trained using the available labeled data. It is then refined using the unlabeled data through methods such as pseudo-labeling or co-training. Finally, the refined model is retrained on the labeled data, now with improved generalization abilities due to the inclusion of information from the unlabeled data [5].

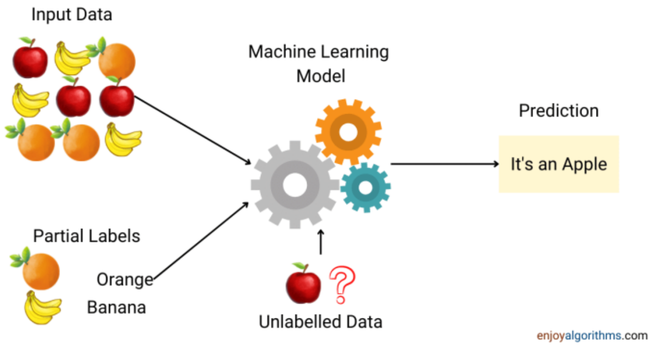


FIG 1.2 : semi-supervised learning.

1.1.4 Reinforcement learning

A key challenge unique to reinforcement learning is the trade-off between exploration and exploitation. To maximize rewards, a reinforcement learning agent must exploit actions that it has already tried and found effective. However, to identify these rewarding actions, the agent must also explore actions it has not previously selected. Therefore, the agent needs to balance exploiting its current knowledge to gain immediate rewards with exploring new actions to enhance its future decision-making. Successfully managing this balance is crucial for the agent's long-term success [6].

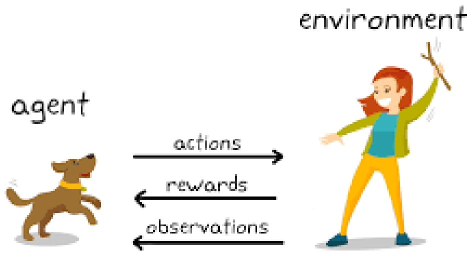


FIG 1.3: Reinforcement learning (web 01).

1.1.5 Self-Supervised Learning

Self-supervised learning primarily uses unlabeled data, creating its own labels from the data itself to learn useful representations or features without relying on manual labels. The process begins with a pretext task, where the model is trained on an auxiliary task such as predicting the next word in a sentence or filling in missing parts of an image. Through this auxiliary task, the model learns meaningful representations of the data. These learned representations are then fine-tuned on a specific downstream task using labeled data, if available. [5]

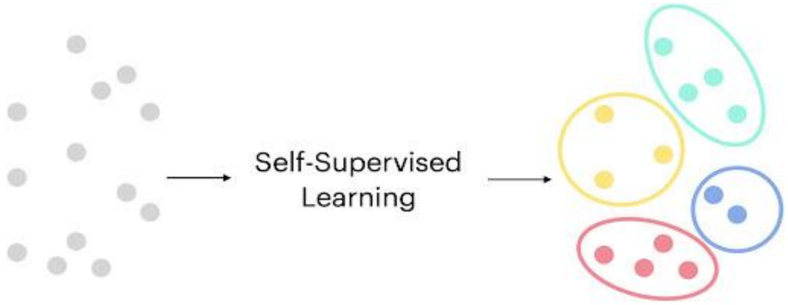


FIG 1.4: Self-supervised learning (web 02).

1.2 Machine learning methods

There are several machine learning methods. In this section, we concisely present the most commonly used and popular classification models Support vector machines (SVM),

1. Support Vector Machines (SVM): SVMs classify data by finding the optimal hyperplane that separates classes in a high-dimensional space. They are effective in high-dimensional settings and can handle both linear and non-linear classification through kernel functions.

2. K-Nearest Neighbors (KNN): KNN classifies data points based on the majority class of their 'k' nearest neighbors. It is simple and non-parametric but can be computationally intensive and sensitive to the choice of 'k' and distance metric.

3. Decision Trees: Decision trees use a flowchart-like structure to make decisions based on feature tests, leading to class labels. They are easy to interpret but can overfit if they become too complex.

4. Logistic Regression: Logistic regression is a linear model for binary classification that estimates the probability of an outcome using the logistic function. It is efficient, interpretable, and suitable for large datasets, despite its simplicity.

2. Deep learning

Deep learning, a subset of machine learning, utilizes artificial neural networks inspired by the human brain to create efficient algorithms for complex classification problems. It relies on complex neural networks and large, precise datasets to enable autonomous machine learning, similar to human learning. As a leading technology in artificial intelligence (AI), deep learning is widely used in various fields such as computer vision, natural language processing, healthcare, and finance [7].

2.1 Origin

The concept of biological neurons serves as the foundation for artificial neural networks, a key component of deep learning.

2.1.1 Biological neuron

The human brain consists of nearly 100 billion neurons, each intricately connected to many others, forming a complex network. Each neuron comprises several components: dendrites, which serve as the neuron's inputs and transmit signals to the cell body; the cell body, which processes the incoming information and determines the neuron's response; and the axon, through which the processed information is sent. The end of the axon forms synapses, the contact points with other neurons' dendrites or muscle fibers, ensuring communication and facilitating nervous responses [8].

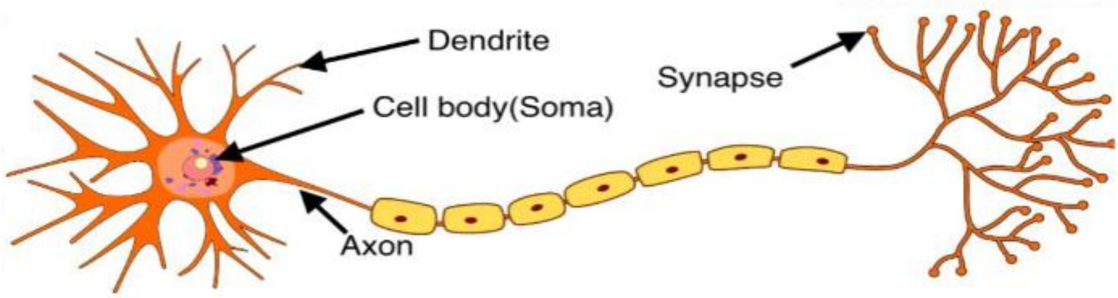


FIG 1.5: Biological neuron.

2.1.2 Artificial neural networks (ANN)

Artificial neural networks (ANNs) are computational models inspired by the brain's structure, consisting of interconnected "neurons" organized in layers. Each neuron's behavior is

determined by connections and associated parameters, with weights representing the strength of connections between neurons in consecutive layers. ANNs typically feature an input layer for receiving data, one or more hidden layers for processing information, and an output layer for producing predictions or classifications. The complexity of the problem being addressed determines the number of hidden layers and neurons within them. Figure 2.2 depicts a typical ANN architecture with two hidden layers, illustrating the network's interconnected structure that enables learning and task performance, such as classification, regression, or pattern recognition. [9].

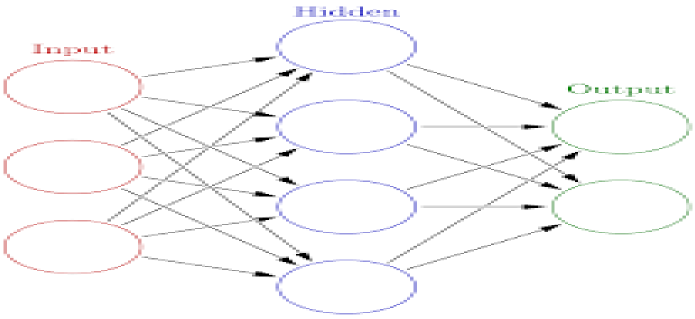


FIG 1.6: ANN architecture (web 03).

2.1.3 The multi-layer perceptron

A widely used model in artificial neural networks (ANNs) is the multi-layer perceptron (MLP), which is the most common type of ANN currently employed. The MLP architecture consists of an input layer, one or more hidden layers, and an output layer. Each layer contains multiple neurons, and neurons within one layer are connected to those in adjacent layers with different weights, determining the influence of one unit on another.

The primary task of an MLP is to generate the desired output based on specific input patterns. To achieve this, the perceptron learns by processing input-output pairs. Inputs and desired outputs are fed into the network, and the error between the actual and desired outputs is computed. During the training phase of the ANN, the system's parameters, namely the weights in the connections between neurons, are adjusted iteratively to minimize the discrepancy between the desired and actual outputs.

MLPs are particularly effective in addressing problems that involve non-linear separability, such as classification and approximation of continuous functions. Their layered structure and ability to learn complex relationships make them valuable tools in various domains of machine learning and artificial intelligence [10].

2.1.4 Dense network

A dense network, also known as a fully connected neural network, consists of layers of interconnected neurons, where each neuron in one layer is connected to every neuron in the subsequent layer. These connections are weighted, and during training, these weights are adjusted to minimize the error between predicted and actual outputs. Activation functions introduce non-linearities to the network, allowing it to capture complex relationships within the data. Trained through supervised learning, dense networks are widely utilized for tasks such as classification and regression, finding applications in areas like image recognition and natural language processing. However, they are susceptible to overfitting, a phenomenon where the model performs well on training data but fails to generalize to unseen data. Regularization techniques such as dropout and L2 regularization can help alleviate this issue by preventing the network from becoming overly complex, thereby enhancing its generalization capabilities [11].

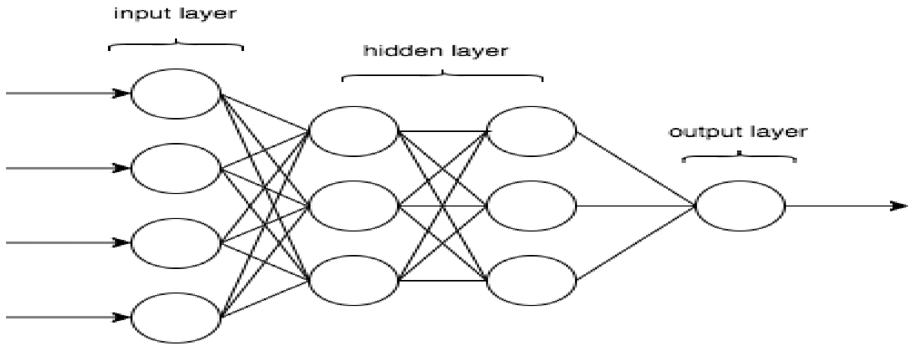


FIG 1.7: Dense network architecture (web 04)

Activation Functions

For neural networks to function properly, Activation Functions are essential, especially when there are multiple layers. The weights of each stratum are identical. This is accomplished by calculating the gradient of the global loss function concerning each weight vector.

For the gradient to be computed effectively, it must traverse each layer. The gradient's value is determined by how we convert the layer's input. To keep training with the proper gradient, it would also be preferable to provide non saturating activation functions. These are a few instances of frequently used activation functions

Logistics function (Sigmoid): This is one of the most frequently utilized functions. It is commonly referred to as the logistic function or the logistics sigmoid, and it is limited between 0 and 1, and it may be understood stochastically as the probability that the neuron activates. It is defined as:

$$\sigma(Z) = \frac{1}{1 + e^{-z}}$$

Hyperbolic Tangent (TanH): The TanH function is a trigonometric function whose waveform gives it the privilege of being selected among the activation functions. It is the same as the logistic sigmoid function, but better the value is between 1 and -1. After differentiation, the value of this function becomes less than 1. It is defined as:

$$\sigma(Z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

Rectified Linear Unit (ReLU): The ReLU function is likely the one that is most similar to its biological counterpart. Recently, many jobs (particularly those involving computer visions) have started to favor this function. As in the formula bellow, this function returns 0 if the entry z is less than 0 and returns z it self ,if it is greater than 0. It is defined as:

$$ReLU(Z) = \max(0, Z)$$

GAUSSIAN ERROR LINEAR UNIT (GELU): This function solves most of the previous activations function issues and more importantly avoids the vanishing gradients problem. It provides a well defined gradient in the neg ative area and prevents neurons from dying. The GELU is formula approximated by:

$$GELU(Z) = 0.5Z(1 + \tan[\sqrt{\frac{2}{\pi}}(Z + 0.044715Z^{3}])$$

Softmax: Softmax is a special activation function for classification networks with multi-class problems.The softmax function is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1. It is defined as :

$$softmax(\overrightarrow{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

The cost (loss) functions

In order to update the weights tying the neurons together, neural networks use a backpropagation technique based on the cost function. A cost function or the error function is how the global network did, a single value generally measured via the average difference between the output and the expected output of a training sample. We can conclude that cost function depends on the network weights, the biases of the network, and one training sample with the expected value of that training sample. Examples of commonly used cost functions :

- **Mean Squared Error (MSE):** Also known as the Quadratic cost function formula or maximum likelihood, it is the default choice for regression problems. Equation (8) explains how to calculate it.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

2.2 DL techniques

Deep learning encompasses various techniques that enable neural networks to learn complex patterns from data. Here are some key techniques commonly used in deep learning:

2.2.1Convolutional neural networks (CNN)

Convolutional neural networks (CNNs) derive their name from their utilization of the convolution operation in mathematics. They represent a specialized form of neural networks that employ convolution instead of traditional matrix multiplication, particularly in at least one of their layers. CNNs excel in learning from data with local correlations, as convolution allows for the extraction of relevant features. In CNNs, convolutional kernels generate outputs passed through a non-linear activation function, facilitating the learning of abstract features and introducing nonlinearity into the feature space. This nonlinearity enables CNNs to differentiate responses and learn semantic differences in images more effectively. The CNN architecture comprises three primary neural layers: Convolutional layer, Pooling layer, and Fully-Connected layer, each serving distinct roles in the network's processing of data [12].

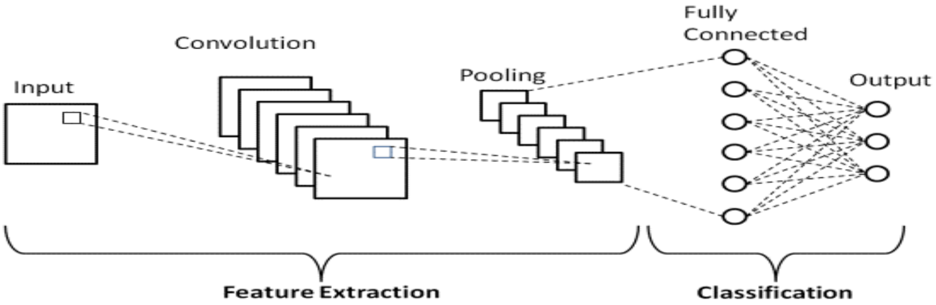


FIG 1.8: CNN architecture (web 07).

2.2.2 Recurrent Neural Networks (RNN)

Recurrent neural networks (RNNs) are specialized deep learning models primarily used for time series-related tasks. They excel in domains such as signal processing, natural language processing, and speech recognition. RNNs possess the unique capability to capture temporal dependencies by utilizing previous outputs as inputs for subsequent steps, facilitated by their internal memory. This temporal context consideration enables RNNs to achieve superior performance in tasks where timing is critical. However, traditional RNNs encounter challenges in learning long-term temporal dependencies, typically beyond a few discrete time steps (approximately 5 to 10) between relevant input events and target signals [13].**2.2.3 Transformer**

The Transformer, a revolutionary deep learning architecture introduced in the paper "Attention is All You Need" by Vaswani et al. (2017), revolutionized natural language processing tasks. Originally designed for sequential input data, Transformers leverage self-attention mechanisms to provide context to input sequences, allowing for non-linear processing and parallelization. Unlike recurrent neural network-based models, Transformers do not require sequential processing of input, leading to increased parallelization and reduced training time. The Transformer architecture consists of an encoder-decoder structure, comprising layers of Multi-Head Attention followed by MLP layers. Each layer incorporates residual connections to facilitate the flow of information. Input sequences, tokenized into words, are embedded into vectors of size model, enabling efficient processing and representation of textual data.

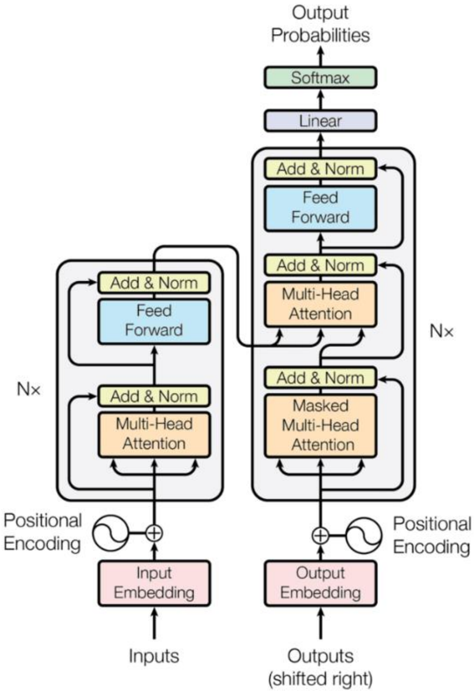


FIG 1.9: Transformer architecture (web 08).

3. Autoencoder [14]

3.1 Definition

The success of deep learning owes much to a method or group of methods known as unsupervised pre-training. This technique prepares deep neural networks to facilitate effective training via the backpropagation algorithm. While the concept of unsupervised pre-training has existed since the mid-1980s, it gained prominence when Hinton and his colleagues systematically explored its potential, demonstrating that deep neural networks (DNNs) could be effectively trained. Initially, they utilized a generative model called the restricted Boltzmann machine. However, it became evident that simpler feedforward networks, known as autoencoders, could achieve similar results. Autoencoders come in various forms and variations, but in this discussion, we'll focus on some of the most popular and fundamental types.

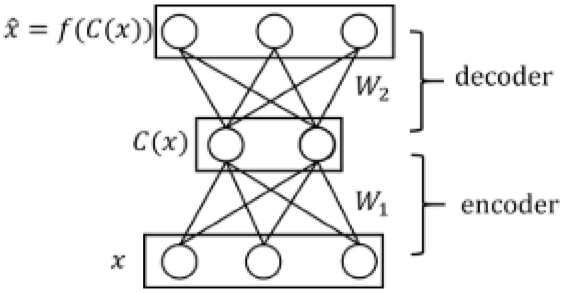


FIG 1.10: Simple autoencoder architecture.

3.2 Principle and functioning

Figure 3.1 shows a simple autoencoder with one hidden layer in the middle. Although the bottom and top layers show only three neurons and the middle only two, they may contain as many neurons as one designates. The bottom and top layers have the same number of neurons, and typically, but not always, the middle layer has fewer neurons. In such a case, the autoencoder is called an undercomplete autoencoder, and if the middle layer has more neurons, it is called an overcomplete autoencoder. The matrix W_1 is the collection of weights connecting the bottom and the middle layers and W_2 the middle and the top. They are usually, but not always, tied, i.e. $W_2 = W_1^T$. So now let W_1 = W and W_2 = W^T. The input x is fed into the bottom layer to produce

$$h = \sigma(Wx + b)$$

in the middle layer, and from this is gotten the output $\hat{x}$ in the top layer

$$\hat{x} = \sigma(W^T h + c)$$

The basic goal of an autoencoder is to make the output ˆx as close to the input x as possible. Now it cannot always be done if the number of neurons in the middle layer is much smaller than in the bottom or the top. In this case, the middle layer value h is sort of like compressed data of the input, and because of this reason, we write $h = C(x)$.

So in this sense, the neural network consisting of the bottom and the middle layers is called an encoder, and the network made up of the middle and top layers is called a decoder, and we write

$$\hat{x} = f\big(C(x)\big)$$

To see how to make this encoder-decoder combination work, let the dataset be

$$D = \{x^{(i)}\}_{i=1}^{T}$$

We write

$$\hat{x}^{(i)} = f(C\big(x^{(i)}\big)).$$

The error is usually the L 2 -error given by

$$E = \tfrac{1}{2}\textstyle\sum_i \left|\hat{x}^{(i)} - x^{(i)}\right|^2$$

$$E = \frac{1}{2}\sum_i \sum_k \left|\hat{x}^{(i)} - x^{(i)}\right|^2$$

and the training algorithm is the one utilizing the standard backpropagation algorithm of the feedforward network.

3.3 Autoencoder types

3.3.1 Stacked autoencoders

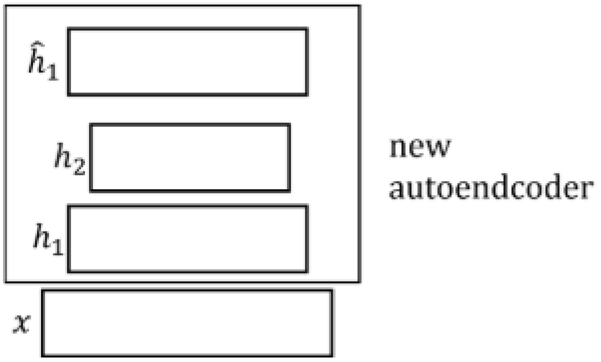


FIG 1.11: stacked autoencoder architecture.

The simple autoencoder described in the above section can be stacked to produce a deep autoencoder. We first describe the so-called layer wise pre-training. It goes as follows. First, as in the Figure above, the top layer of the simple autoencoder is removed while the bottom and the middle layers remain intact. The value of the middle layer is now renamed as $h_1$and by abuse of language we also denote this layer by h_1. In what follows, we also abuse the language to use the same symbol to represent the name of the layer and its value at the same time. Second, use layer h_1, the old middle layer, as an input and put two more layers h2 and $\hat{h}_1$. Then the three layers, h_1, $h_2$and $\hat{h}_1$, can be regarded as another simple autoencoder on its own right and is trained likewise. In particular, given an original input $x^{(i)}$, h1 is now regarded as the input for the newly created simple autoencoder $h_1 - h_2 - \hat{h}_1$

Now remove the top layer $\hat{h}_1$ and keep h_2. We now have a three-layer neural network consisting of x, h_1, h_2. The connection matrix between h1 and h2 is the one gotten from the simple autoencoder $h_1 - h_2 - \hat{h}_1$; and the connection matrix between x and h1 is the one gotten from the original simple autoencoder $x - h_1 - \hat{x}$ Now keep repeating the same process to add more layers one at a time to get a multi-layered autoencoder, hence the name deep autoencoder.

Note that up to this, everything was done using only the input x. Since we have no need for the label y, what we have done so far can be dubbed unsupervised learning.[14]

3.3.2Denoising autoencoders

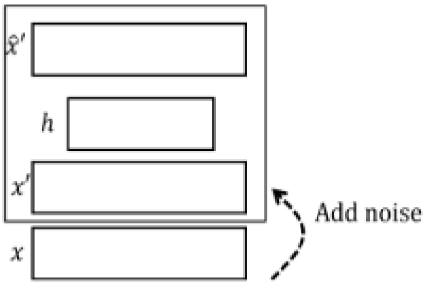


FIG 1.12: denoising autoencoder architecture.

In order to avoid overfitting autoencoders, various regularization techniques are applied. One of them is intentionally adding noise, in which case the resulting autoencoder is called a denoising autoencoder. The Figure above shows an example of a simple autoencoder to which noise is added. It works as follows. Instead of using the input x, one adds noise, typically Gaussian noise, and this corrupted input $\acute{x}$ is used as an input for the simple autoencoder enclosed in the box in Figure 1.12. However, when it comes to decoding, its target is not the corrupt $\acute{x}$ but the uncorrupted original input$x$. So the error is

$$E = \frac{1}{2}\sum_{i}\left|\hat{x}^{(i)} - x^{(i)}\right|^2$$

$$E = \frac{1}{2}\sum_{i}\sum_{k}\left|\hat{x}_k^{(i)} - x_k^{(i)}\right|^2$$

Where the value $\hat{x}_k^{(i)} = Cf(\ \acute{x}^{(i)})),$ i.e. the decoded value of $\acute{x}^{(i)}$ that is the corrupted $x^{(i)}$.

One can also apply the same denoising method for each layer in the layerwise pre-training.

The training we have described above is done layer-by-layer, hence the layer wise training. And we also call it pre-training if we want to emphasize the fact that the supervised learning involving the output (label) is the final goal and the construction of the autoencoder is only a preparatory step. There is another way to train autoencoders. In the construction of a deep autoencoder, instead of building up layer-by-layer, we put up the whole structure at once and train the whole network (find good connection weights) by minimizing the L^2-error between the input x at the bottom layer and the output $\hat{x}$ at the top layer. This is a feasible problem because nowadays training deep neural networks can be routinely done. Once the network is constructed, we remove the top half to get the network. This again is the deep autoencoder.

3.3.3Variational autoencoder

Variational autoencoders (VAEs) are powerful deep generative models widely used to represent high-dimensional complex data through a low-dimensional latent space learned in an unsupervised manner. In the original VAE model, the input data vectors are processed independently. Recently, a series of papers have presented different extensions of the VAE to process sequential data, which model not only the latent space but also the temporal dependencies within a sequence of data vectors and corresponding latent vectors, relying on recurrent neural networks or state-space models. In this paper, we perform a literature review of these models. We introduce and discuss a general class of models, called dynamical variational autoencoders (DVAEs), which encompasses a large subset of these temporal VAE extensions. Then, we present in detail seven recently proposed DVAE models, with an aim to homogenize the notations and presentation lines, as well as to relate these models with existing classical temporal models. We have reimplemented those seven DVAE models and present the results of an experimental benchmark conducted on the speech analysis-resynthesis task (the PyTorch code is made publicly available). The paper concludes with a discussion on important issues concerning the DVAE class of models and future research guidelines.[15]

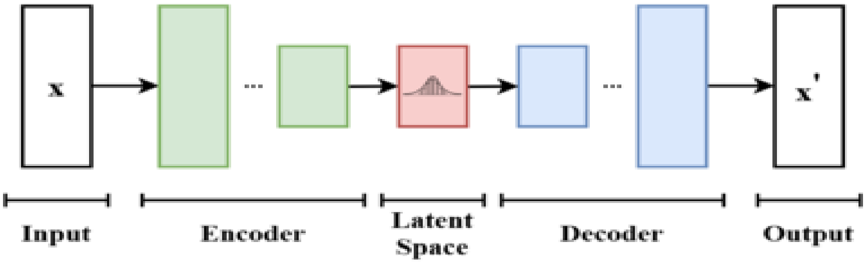


FIG 1.13: Variationalautoencoder architecture(web09)

3.3.4 Sparse autoencoder

Sparse Autoencoders are a type of artificial neural network that are used for unsupervised learning of efficient codings. The primary goal of a sparse autoencoder is to learn a representation (encoding) for a set of data, typically for the purpose of dimensionality reduction or feature extraction.

Sparse Autoencoders consist of an encoder, a decoder, and a loss function. The encoder is used to compress the input into a latent space representation, and the decoder is used to reconstruct the input from this representation. The sparsity constraint is typically enforced by adding a penalty term to the loss function that encourages the activations of the hidden units to be sparse.

The sparsity constraint can be implemented in various ways, such as by using a sparsity penalty, a sparsity regularizer, or a sparsity proportion. The sparsity penalty is a term added to the loss function that penalizes the network for having non-sparse activations. The sparsity regularizer is a function that encourages the network to have sparse activations. The sparsity proportion is a hyperparameter that determines the desired level of sparsity in the activations.[16]

3.4 Application domain of autoencoders

3.4.1Anomaly Detection: Autoencoders can learn to reconstruct normal instances of data accurately. When presented with anomalous or outlier data, the reconstruction error tends to be higher. By monitoring reconstruction errors, autoencoders can detect anomalies in various domains, including cybersecurity, fraud detection, and equipment maintenance.

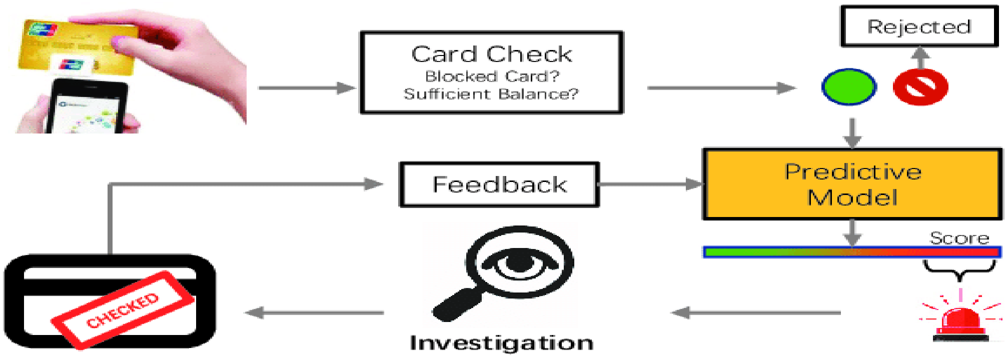


FIG. 1.14: Credit Card Fraud Detection (web10).

3.4.2Dimensionality Reduction: Autoencoders can compress high-dimensional data into a lower-dimensional representation while preserving essential features. This capability is useful in tasks like data visualization, where reducing the dimensionality of data facilitates easier analysis and interpretation.

3.4.3 Image Denoising and Reconstruction: Autoencoders trained on clean image data can be used to reconstruct noisy or corrupted images, effectively removing noise and restoring the original content. This application is valuable in image processing tasks where images are degraded by factors such as compression artifacts or sensor noise.

3.4.4 Feature Learning and Representation: Autoencoders learn compact and meaningful representations of data through unsupervised learning. These learned features can then be used as input to downstream supervised learning tasks, such as classification or regression, improving the performance of the overall system.

3.4.5 Generative Modeling: Variants of autoencoders, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), can generate new data samples similar to the training data. This capability is particularly useful in generative modeling tasks, including image generation, text generation, and music composition.

3.4.6 Recommendation Systems: Autoencoders can learn latent representations of user-item interactions in recommendation systems. By encoding user preferences and item characteristics into a lower-dimensional space, autoencoders can provide personalized recommendations, enhancing user experience and engagement in e-commerce platforms, streaming services, and content aggregators.

3.4.7 Time Series Analysis: Autoencoders can be applied to sequential data, such as time series or sequential sensor readings. By learning temporal dependencies and patterns in the data, autoencoders can perform tasks like anomaly detection, prediction, and sequence generation in domains like finance, healthcare, and manufacturing.

CHAPTER2

ELECTRICITY THEFT

Introduction

Electricity is crucial for our daily lives, powering everything from lights to transportation. Unfortunately, electricity theft is a widespread phenomenon throughout the world, whether in developed or underdeveloped countries. Still, it is common in undeveloped places due to the poverty experienced by the residents of these areas, which drives them to such immoral acts.

One of the reasons for the spread of such thefts is the lack of supervision and the difficulty of detecting them. Electricity theft remains a problem that affects the economy of all countries.

In this chapter, we will discuss electricity theft and the types of stealing, how can we figure them out, and the tools that allow it to be detected.

1. Electricity theft

Every year, energy theft globally causes enormous economic losses for electric utilities[1]. Energy theft is typically defined as the illegal usage of energy services with dishonest intentions, such as meter tampering, bypassing, or direct tapping from feeders[2]. Implementing smart meters and advanced metering infrastructure (AMI) can help electric utilities prevent traditional energy theft[3]. However, intelligent energy thieves may still manipulate smart meters using advanced techniques[4]. Electricity theft is considered a crime, offense, or even a minor mistake in different countries, depending on local parameters. In this section, various cases of electricity theft will be detailed.

2. General statistics

Several case studies exist covering energy theft, these studies include countries such as Russia, India, the U.S.A., and the U.K. According to the stay energy safe website, here are the following statistics:

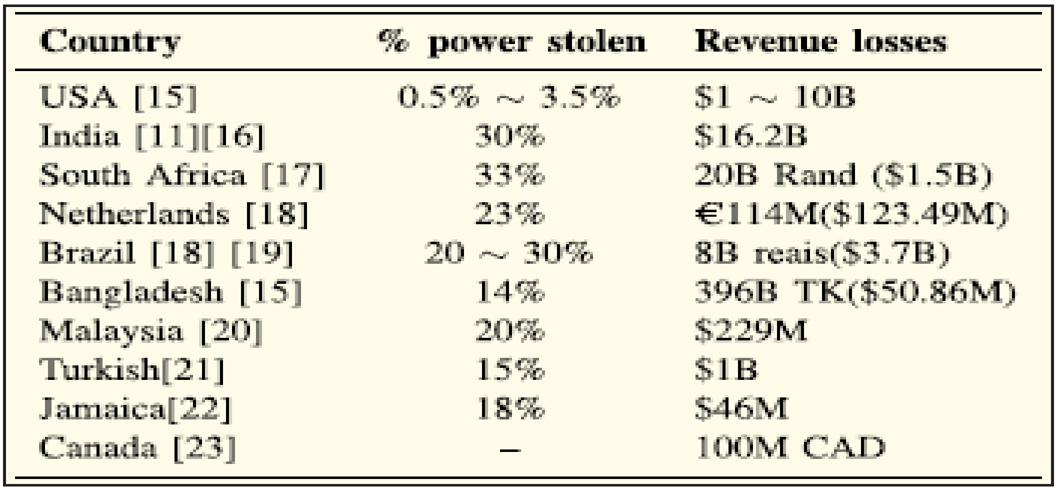
Here has been a 61% increase in reports since last year with over 1,000 reports a month received by Crimestoppers during the winter months (comparing Dec 21-Feb 22 to Dec 22- Feb 23)**.** Since 2017, it has been reported that there has been a 400% increase in people tampering with their energy supply. It is estimated that up to 250,000 cases of energy theft still go unreported every year. It is also estimated that energy theft is occurring in up to 1 in 150 homes every year.

A recent study by Direct Line business insurance found 43% of electricians and gas engineers have been asked by customers to tamper with their meters.

Also, nearly all (92%) of the electricians and gas engineers surveyed believe that increased energy costs have led to more people looking to make their meters run slowly to save money on energy. Crimestoppers' figures show that 12,076 reports about energy theft were received in the 12 months to April 2023, substantially up compared to the 7,776 reports received in the 12 months to April 2022.

Electricity theft has many negative effects, among which huge economic losses drive utility companies to figure out ways to reduce electricity theft greatly. Total worldwide revenue losses were total about $89.3 billion in 2014[5]. which rose to $96 billion in 2017[6]. In Table 1, we summarize recent statistics regarding ratios of stolen electricity to total power generation and revenue losses in several countries. The statistics in Table 1 indicate that electricity theft is much more serious in developing countries than in developed countries[7]. Particularly, India loses the most money among all countries in the world. Since utility companies cannot afford the enormous economic losses alone, they usually pass on these losses to all customers via higher tariffs. It is reported that each customer in the U.K. has to pay extra C30 for electricity theft[8].

In addition, electricity theft induces utility companies to underestimate customers' power demands. When electricity theft occurs, adversaries report lower electricity consumption than what is consumed. The incorrect information would mislead utility companies to generate less electricity than what is demanded, resulting in power quality degradation. As reported, in regions where electricity theft is pervasive (e.g., India), customers often experience voltage sags and intermittent power disruptions, especially during peak load periods[9].



Country	% power stolen	Revenue losses
USA [15]	0.5% ∼ 3.5%	$1 ∼ 10B
India [11][16]	30%	$16.2B
South Africa [17]	33%	20B Rand ($1.5B)
Netherlands [18]	23%	€114M($123.49M)
Brazil [18] [19]	20 ∼ 30%	8B reais($3.7B)
Bangladesh [15]	14%	396B TK($50.86M)
Malaysia [20]	20%	$229M
Turkish[21]	15%	$1B
Jamaica[22]	18%	$46M
Canada [23]	–	100M CAD

FIG. 2.1 : Sample Electricity Theft Statistics.

According to ELchourouk online electronic newspaper in 05/03/2017: Algerians Steal 13 Thousand Megawatts Of Electricity Annually.

The National Electricity and Gas Company Sonelgaz not secured 6 billion dinars from dues among public and private customers during 2016, to settle in the range of 58 billion dinars as it was estimated last year at 64 billion dinars.

According to The Sonelgaz CEO Mr Guitouni, these nefarious phenomena, especially electricity theft account for about 20 percent of the productive capabilities of the company with regard to electricity, as about 13 thousand megawatts per year are wasted because of these illegal practices deriving from wanton stealing, cheating and evasion of bills' payment.

Eldjoumhouria.dz website also published in 26/01/2024: Sidi Bel Abbes: 401 cases of electricity theft and 23 gigawatts recovered during the year 2023.

The Directorate of Electricity and Gas Distribution of Sidi Bel Abbes State has taken several measures to compensate for the damage resulting from recorded cases of electricity theft, including the amicable settlement, which enabled the recovery of 23 gigawatts during the year 2023, equivalent to about 9 billion and 300 million centimes. Sonelgaz's departments had recorded 401 cases of theft distributed across all parts of the state, which took two forms: either tampering with the real consumption of the consumed energy or stealing energy from the network without billing it. This theft also causes financial losses that harm the company's budget and has several consequences. Technical negativity, including pressure on the capacity of electrical and gas installations, and frequent fluctuations (Illegally exploited facilities, disturbances in the supply of energy by customers, a significant decrease in tension with repeated occurrences, short service life of the equipment, loading transformers and connections beyond their capacity, in addition to its social repercussions, which relate primarily to the growing phenomenon of attacks against agents The Directorate after recording cases of theft among citizens).

4. Electricity theft methods

As electric energy suppliers inform us, the ingenuity of so-called clients knows no boundaries. Based on their many years of experience, they assure us that there are about 300 different theft techniques. Collectors and controllers visiting the clients are alert to such ideas. There are cases of clients, in their attempt at tampering, breaking the meter in a way that makes it impossible to fix it or inflates the measured values. Signals indicating an act of illegal electric energy consumption are usually (although not always) scratch marks or damage of the plastic welds on the meter’s housing. Issues related to energy theft can be presented in the following categories:

1. classic energy theft from a line in front of the meter,
2. modification of the meter’s operation,
3. short circuiting orterminal modification and tampering with the physical security mechanisms of the meter,
4. modification of the meter’soperation - tampering with the memory and motherboard chip of the meter.

There is also another form of theft involving a collusion with an employee supervising the collecting and billing process. The most important and most popular theft techniques are:

4.1 Hidden connection in front of the meter

This technique is the most popular method of electric energy theft. It is most often encountered among clients of lower social standing, living in old apartment building whose old electrical installations facilitate performing illegal modifications. This method involves connecting an additional installation to the line in front of the metering system. The basic version of this kind of theft assumes total utilization of an illegal source, which is quite easy to detect. A more sophisticated method involves only partial use of an illegal source to make detection harder. The theft is often performed only on a seasonal basis, e.g. in winter, for heating.

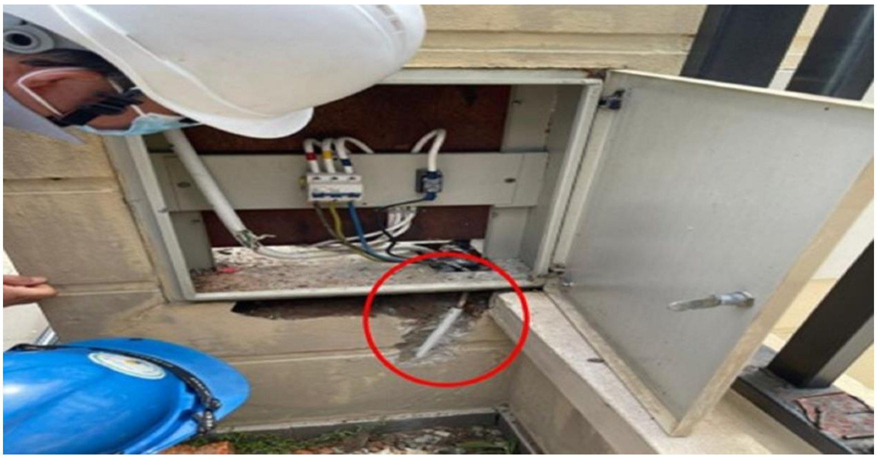


FIG. 2.2: Direct tapping cables(web01).

Physical tampering with the terminals or bridging

The consumer has a three-phase meter at their disposal. If they are knowledgeable in electrical engineering and disconnect the N neutral wire, taking care not to allow the installation's N point to be galvanically connected with the meter's N point, then, in case of asymmetrical energy consumption, the meter will understate the energy value, e.g. by 30% ,as potential equalization current will not flow through the N wire. Does the recipient consume energy in accordance with the contract? These days, some distribution companies purchase meters that facilitate detection of such actions.

4.3. Physical tampering with an analog meter's mechanism

4.3.1. Magnetic field

Influencing a meter with a strong electromagnetic field by means of neodymium magnets, which is one of the most popular methods of illegal energy consumption. That is because demagnetization of the inhibitory magnets initially slows down the counting wheel until it stops, but then it start moving again and speeds up, which in turn causes inflated readings and increased energy bills.

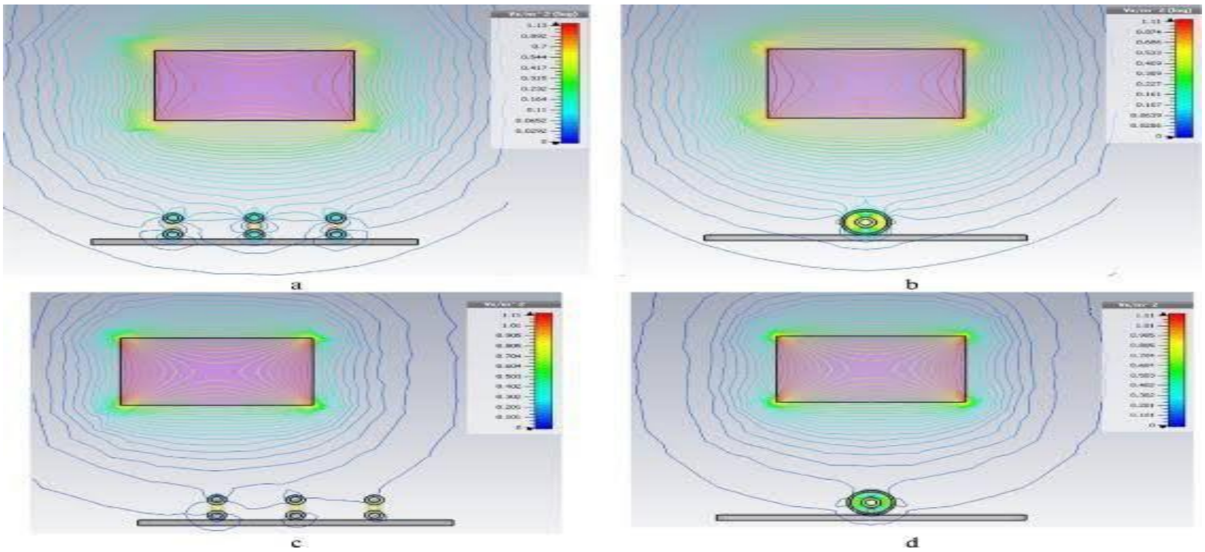


FIG. 2.3: Development of effective shielding against electricity (web02).

4.3.2. Photographic film

This fraud involves an attempt to stop the moving counting wheel of an analog meter by means of inserting photographic film between the rear housing of the meter and the glass front cover. This trick is popular because it does not necessitate opening the meter and it is easy to perform. The properties of film (it is narrow, elastic and durable) make it possible for it to be inserted into a meter (and moved if need be), leaving practically no traces. The curious thing is that photographic film passes well through narrow bends made by edges of the housing and cover. Inserted film rolls up inside the meter, hindering or completely blocking the movement of an analog meter's wheel.

4.3.3. Housing drilling

A technique very similar to the previous method, it involves making a discrete hole in the housing. The most common technique is drilling a hole with adiameter lesser than 1mm or burning such a hole in the plastic housing with a heated needle. The hole is small enough to allow for insertion of elastic print and, if need be (after resistance is gone), blanking it off. A hole with a minimal diameter made in the right spot can go unnoticed by a collector performing a reading.

4.4. Tampering with a digital meter's software

This technique involves making a logical change in the state of variables we can refer to by means of socalled OBIS (Object Identification System) codes.

These variables are responsible for storing values based on which the energy supplier bills a given client. Due to the process of breaking through many security measures being exceptionally complex and time-consuming, and the necessity to have highly specialized knowledge, this method of fraud is encountered extremely rarely. More information on security of smart electric energy meters can be found in the next point.

We find these methods almost all over the world. The difference is slight in the type of tools used, but the method is the same. For example in Algeria according to an article published in ELchourouk newspaper on 18/02/2016

Many citizens decided to resort to some methods and tricks to confront the increase in electricity bill prices, which was announced within the framework of the austerity policy pursued by the country, as they deliberately put some things in the meters to completely disrupt its movement for a short period of time or reduce its speed.

The infernal methods and tricks that many Algerians have used to confront the high prices of electricity and gas bills are considered plans that do not occur to anyone, as many of them have invented methods that lead to permanently disabling the movement of the meter for a short period of time, so that a large amount of consumption is not recorded, while others have adopted Methods that enable them to reduce the speed of the meters in order to reduce the value of the electricity consumption recorded by the meter.

According to responsible sources at the Sonelgaz Foundation, many citizens discovered some tricks through which they were able to stop the movement of the meter, especially with regard to old meters, as some of them resorted to using a small nail or needle, while others used a piece of magnet or tape. From "Al-Klishi," adding that placing a large block on top of the meter also stops the movement of the meter completely, such as placing a large stone or even "wages," because the weight of the block pressing on the meter greatly reduces the speed of the meter's rotation. As for the new meters, it is difficult to circumvent them, and despite this, some of them are wise to change the location of only two wires.

The same spokesman revealed that these methods were common before, but they increased significantly after the announcement of an increase in electricity prices due to the austerity policy, saying that they are removed a few days before the agents responsible for recording the amount of electricity consumed pass by.

5. Electricity theft detection

5.1 Meter Analysis

Abnormal Consumption Patterns: Utility companies analyze meter data to identify sudden spikes or irregularities in electricity consumption. A sudden increase in consumption outside of normal patterns can indicate potential theft.

Load Profiling: Load profiling involves creating profiles of typical energy usage for different types of customers. Deviations from these profiles may suggest theft.

5.2 Advanced Metering Infrastructure (AMI):

Smart Meters: These meters provide real-time data on electricity consumption. Utilities can remotely monitor usage, detect abnormalities, and receive alerts in case of potential theft.

5.3 Data Analytics

Using advanced analytics, utilities can process large volumes of data to identify patterns associated with theft. Machine learning algorithms can learn from historical data to recognize anomalies.

5.4 Remote Sensing Technologies

Infrared Imaging: Infrared cameras can detect heat signatures associated with unauthorized connections or hotspots on electrical equipment.

Drones: Drones equipped with infrared cameras can be used to survey large areas and identify irregularities.

5.5 Power Quality Analysis

Monitoring variations in voltage and current can help detect abnormalities associated with theft, such as the use of illegal connections or bypass circuits.

5.6 Tamper Detection Devices

Electronic seals or tamper detection devices can be installed on meters. They trigger alerts when tampering or unauthorized access is detected.

5.7 Field Inspections

Utility personnel conduct regular physical inspections to check for visible signs of tampering, illegal connections, or meter bypass. This may involve inspecting meters and the surrounding infrastructure.

5.8 Meter Seals and Security Measures

Utilizing secure meter seals and enclosures to prevent physical tampering. Tampering attempts can be identified if the seals are broken or compromised.

6. State of the art

6.1 Some AI software for electricity theft detection

6.1.1 C3 AI Energy

FIG. 2.5: Logo C3 AI (web03).

Energy management platform empowers sustainability, facilities, and operations managers to achieve targets for energy cost, GHG emissions, water consumption, and waste reduction. The application models fuel efficiency and emissions at every level from the individual equipment up to the facility as well as SKU-level product carbon footprints. Advanced AI models identify opportunities for fuel efficiency, prioritize emissions and cost reduction strategies with benchmarking and scenario analysis, alert operators to efficiency anomalies, and verify progress against sustainability goals across the enterprise (web03).

6.1.2 Itron analytics

FIG. 2.6: Logo Itron (web04).

Itron provides analytics solutions for utilities. Their software may include features for detecting anomalies and unusual patterns in energy consumption data(web04).

6.2 Data base

A database is a structured collection of data that is organized in a way that allows for efficient storage, retrieval, and management of information. It serves as a centralized repository where data is stored, and it typically employs a systematic method for organizing and accessing data. Databases are used in various applications, such as business systems, websites, and software applications, to store and manage large volumes of information. The data in a database is organized into tables, which consist of rows and columns, and relationships between different tables can be established to provide a comprehensive and interconnected view of the data. Database management systems (DBMS) are software applications or systems that facilitate the creation, maintenance, and manipulation of databases, ensuring data integrity, security, and efficient data handling.

6.2.1 SGCC Electricity Theft Detection

Electricity theft detection released by the State Grid Corporation of China (SGCC) dataset data set.csv contains 1037 columns and 42,372 rows for electric consumption from January first 2014 to 30 October 2016. SGCC data from column 1 to column 1035 is daily electricity consumption. Then column 1036 is consumer ID that is alphanumeric. The last column named flag is the labels in 0 and 1 values. 0 indicating no theft and 1 for theft.

6.2 Some related works

Haiqing Liu 2020

The new CSL power-stealing detection model proposed in this work deals with unbalanced data sets through CWGAN. The generated power-stealing data is mixed with the original data to form an enhanced data set for subsequent feature extractor training. Experiments show that the model not only makes the model converge quickly, but the MCC value is higher under the same epoch and the final MCC value of the model is increased by 0.1 to 0.8 compared to the case without data balancing operation.

In addition, in view of the interference noise phenomenon in the user's electricity data set, a comprehensive convolution and encoder idea is proposed to extract the power-stealing feature extractor SCDAE. On the one hand, the noise in the data set is filtered by the noise reduction auto-encoder to avoid the adverse impact of the noise data. On the other hand, the noise reduction autoencoders are stacked by convolution to extract more typical features in the theft of electricity, laying a good foundation for subsequent classification detection. Finally, through experiments comparing the training and test results obtained on different data-stealing detection models on the same data set, it is concluded that the CSL power-stealing detection model has improved the typical indicator accuracy from about 85% to more than 90% compared to the common power-stealing detection model, which has obvious advantages. The current work of this paper still has certain limitations, which involve the need to adjust a large number of parameters when using the LightGBM library. The LightGBM model parameters play an important role in the final effect of the power-stealing model. In this paper, only manual debugging is used to implement some parameters, and LightGBM's advantages in the classification of power-stealing features have not been fully utilized. Parameter adaptive adjustment methods can be added in the future to achieve the optimal approximation of model parameters.

Takiddin 2022

In this research, researchers developed Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids using time series data. they examine fully connected feed-forward and seq2seq structures in auto-encoders, focusing on RNN and LSTM architectures. Fully connected feed-forward structures are considered as a benchmark due to their simplicity, while LSTM-based RNN auto-encoders are designed to capture temporal correlations in time-series electricity consumption data. The study evaluates three auto-encoder structures: simple auto-encoder (SAE), variational auto-encoder (VAE), and auto-encoder with attention (AEA). VAE and AEA are found to improve detection performance by capturing variabilities and handling long sequences, respectively. The proposed auto-encoders are tested on two datasets and validated against six unseen electricity cyber-attacks. Their performance is compared to various detectors, and the deep AEA anomaly detector is identified as the most effective, achieving 94% in detection rate with low complexity and almost instant decision-making.

Pamir 2023

Researchers develop a novel approach efficiently detect electricity theft in smart grids. Using the salp swarm algorithm (SSA), gate convolutional autoencoder (GCAE), and cost-sensitive learning and long short-term memory (CSLSTM), an effective electricity theft detection (ETD) model named SSA–GCAE–CSLSTM is proposed in this work. Furthermore, a hybrid GCAE model is developed via the combination of gated recurrent unit and convolutional autoencoder. The proposed model comprises five submodules: (1) data preparation, (2) data balancing, (3) dimensionality reduction, (4) hyperparameters' optimization, and (5) electricity theft classification. The real-time EC data provided by the state grid corporation of China are used for performance evaluations via extensive simulations. The proposed model is compared with two basic models, CSLSTM and GCAE–CSLSTM, along with seven benchmarks, support vector machine, decision tree, extra trees, random forest, adaptive boosting, extreme gradient boosting, and convolutional neural network. The results exhibit that SSA–GCAE–CSLSTM yields 99.45% precision, 95.93% $F1$ score, 92.25% accuracy, and 71.13% area under the receiver operating characteristic curve score, and surpasses the other models.

Ruizhe 2023

This work proposes a semi-supervised ETD approach based on a hybrid replay strategy. From the data perspective, the reseqrchers design a hybrid replay strategy that includes a variational autoencoder (VAE) and sample scrambling ranking (SSR) methods, and uses a “rehearsal” method to obtain incremental ETD capability. From the detection method perspective, this work designs a semi-supervised ETD architecture that uses a temporal convolutional attention network (TCAN) as a feature extractor and uses contrastive learning to improve the utilization of unlabeled sensing samples, thus reducing the labeled sample size required for the fine-tuning process. Experimental results on the Irish smart energy trial (ISET) dataset show that the proposed scheme effectively solves the problem of incremental ETD in small sample size, and achieves 92.72%, 92.70%, and 92.57% on accuracy, precision, and f1-score, respectively.

CHAPTER3

CONCEPTION AND EXPERIMENTATION

Introduction

Detecting abnormal electricity consumption is a valuable approach for identifying theft by analyzing usage patterns. Classifying consumers as potential thieves relies on identifying non-technical losses. Our primary objective is to construct a model for categorizing consumption data using a stacked denoising autoencoder combined with various classification techniques. In this chapter, we will outline the steps for building the model and present the results.

1. Architecture of the system

The architecture of the electricity theft detection system consists of several essential steps illustrated in the block diagram depicted in Figure 1. Our approach starts with utilizing a learning database that contains the consumption data required to train the autoencoder model. Specifically the stacked denoising autoencoder. Then we take the extracted features or the constructed data (the output of the bottleneck layer of our autoencoder) and give it to another classification technique for further analysis.

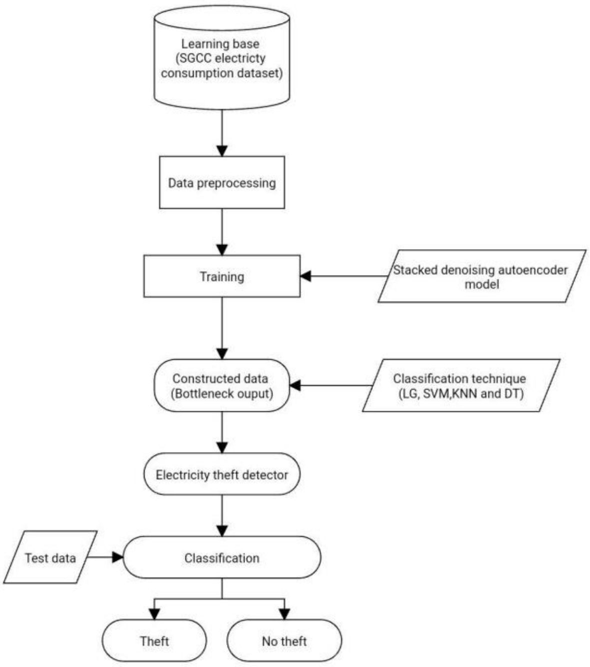

Learning base (SGCC electricty consumption dataset)
Data preprocessing
Training
Stacked denoising autoencoder model
Constructed data (Bottleneck ouput)
Classification technique (LG, SVM,KNN and DT)
Electricity theft detector
Test data
Classification
Theft
No theft

FIG. 3.1: Electricity theft detector model.

2 Presentation details of the system

2.1 Database

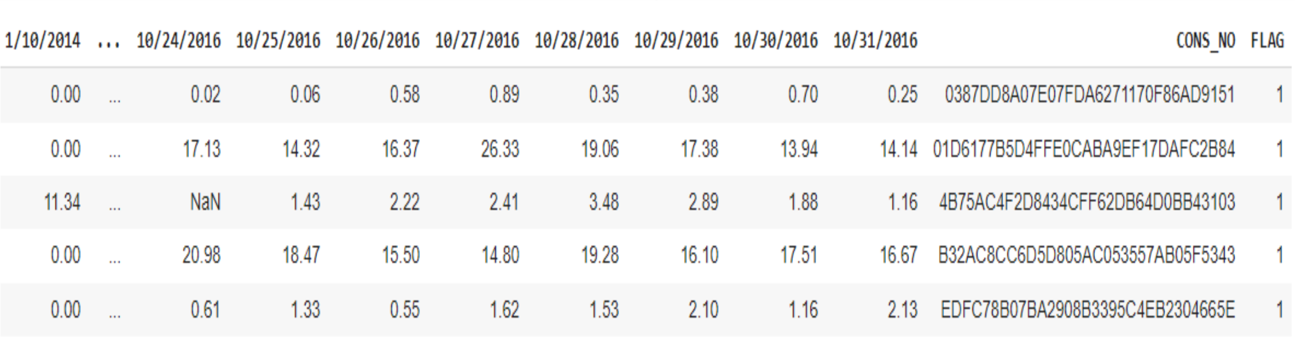
In order to realize our model, we must have databases representing the objects to be studied.

We used the SGCC electricity consumption database, which is composed of 43855020 values represent the daily consumption collected from 1035 China consumers.

2.1.1 Data mining

In the field of machine learning, Data quality is of paramount importance, as it directly influences the success of the project. Therefore, data preprocessing is generally considered the first step to be performed in a project.

In this section we will discover, visualize and draw meaningful information from our corpus, in order to make appropriate modifications and make informed decisions. Figures 1 represents the dataset used.

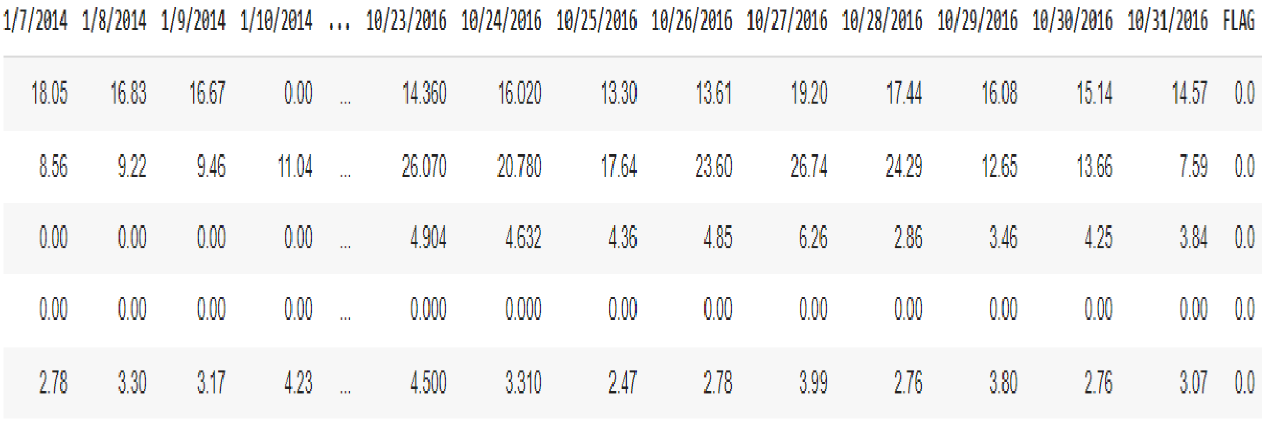


1/10/2014	...	10/24/2016	10/25/2016	10/26/2016	10/27/2016	10/28/2016	10/29/2016	10/30/2016	10/31/2016	CONS_NO	FLAG
0.00	...	0.02	0.06	0.58	0.89	0.35	0.38	0.70	0.25	0387DD8A07E07FDA6271170F86AD9151	1
0.00	...	17.13	14.32	16.37	26.33	19.06	17.38	13.94	14.14	01D6177B5D4FFE0CABA9EF17DAFC2B84	1
11.34	...	NaN	1.43	2.22	2.41	3.48	2.89	1.88	1.16	4B75AC4F2D8434CFF62DB64D0BB43103	1
0.00	...	20.98	18.47	15.50	14.80	19.28	16.10	17.51	16.67	B32AC8CC6D5D805AC053557AB05F5343	1
0.00	...	0.61	1.33	0.55	1.62	1.53	2.10	1.16	2.13	EDFC78B07BA2908B3395C4EB2304665E	1

FIG. 1: examples of the SGCC database

2.1.2 Data preprocessing

In our system, we perform preprocessing operations applied to numerical values to prepare them for training the model. These operations include dropping the unnecessary columns and rows and removing the missing and outlier values, applying a normalization, standardization methods for more cleaning. These operations aim to improve data diversity and prepare for better learning convergence.



1/7/2014	1/8/2014	1/9/2014	1/10/2014	...	10/23/2016	10/24/2016	10/25/2016	10/26/2016	10/27/2016	10/28/2016	10/29/2016	10/30/2016	10/31/2016	FLAG
18.05	16.83	16.67	0.00	...	14.360	16.020	13.30	13.61	19.20	17.44	16.08	15.14	14.57	0.0
8.56	9.22	9.46	11.04	...	26.070	20.780	17.64	23.60	26.74	24.29	12.65	13.66	7.59	0.0
0.00	0.00	0.00	0.00	...	4.904	4.632	4.36	4.85	6.26	2.86	3.46	4.25	3.84	0.0
0.00	0.00	0.00	0.00	...	0.000	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
2.78	3.30	3.17	4.23	...	4.500	3.310	2.47	2.78	3.99	2.76	3.80	2.76	3.07	0.0

FIG. 2: Final database

2.1.2 Data division

We have divided this database into three subsets: the training set, the validation set and the set test (train/val/test).

Training Set: This portion of the dataset is used for training both the autoencoder and the classifier. The autoencoder learns to reconstruct the input data, while the classifier learns to classify the encoded representations.

Validation Set: This set is used for hyperparameter tuning and model selection. During training, we will evaluate the performance of the autoencoder and the classifier on this dataset to choose the best hyperparameters and architecture configurations. It helps prevent overfitting and ensures that your models generalize well to unseen data.

Test Set: This set is kept separate from the training and validation data and is used to evaluate the final performance of our combined autoencoder-classifier model. Once we've selected the best hyperparameters and trained your models using the training and validation sets, we evaluate their performance on the test set to assess how well they generalize to new, unseen data.

2.1.3 Creation of data loaders

In this step, we create the data loaders to load data in batches during training and evaluation of the model. This facilitates batch processing of images and improves the effectiveness of the model.

2.1.4 Training

In a consumption data classification system using the combined autoencoder-classifier model, the main objective is to develop a system able to classify different consumers (if they are thieves or not) The main steps of training our combined autoencoder-classifier model are as follows:

Initialization of the combined autoencoder-classifier model

To initialize the combined autoencoder-classifier for my SGCC dataset, I first preprocessed the dataset, handling data cleaning, normalization, and feature scaling. Then, I split the dataset into training, validation, and test sets, ensuring each subset was representative of the overall data distribution. Next, I trained the autoencoder on the training set using unsupervised learning, defining its architecture and selecting appropriate hyperparameters. After training, I utilized the encoder component to extract features from all subsets. Subsequently, I initialized and trained a classifier on top of the extracted features, using the training set for training and the validation set for hyperparameter tuning. Finally, I evaluated the performance of the combined model on the test set, assessing its classification accuracy and generalization ability. Throughout the process, I monitored the training progress, adjusted hyperparameters as necessary, and validated the model's performance at each step to ensure effectiveness for classification tasks on the SGCC dataset.

Feature extraction

For feature extraction in the combined autoencoder-classifier setup on the SGCC dataset, I leveraged the encoder component of the trained autoencoder. This encoder, having learned to encode the input data into a lower-dimensional latent space, effectively captured essential features of the dataset. By passing instances from the training, validation, and test sets through the encoder, I obtained compressed representations that distilled the salient characteristics of the data. These extracted features served as input to the classifier, enabling it to learn discriminative patterns for classification tasks. Ensuring consistency across subsets, I employed the same encoder parameters for feature extraction. This approach facilitated efficient utilization of the learned representations, enabling the classifier to operate effectively on the encoded data. Through this feature extraction process, I aimed to enhance the classification performance of the combined model while maintaining computational efficiency and interpretability.

Adjusting model parameters

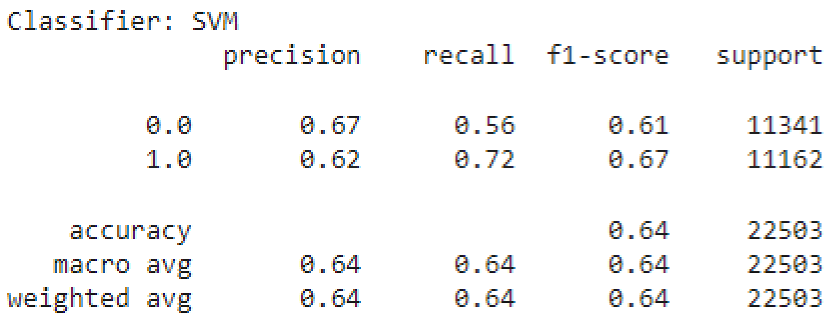
Optimizing model parameters stands as a pivotal stride within the realm of machine learning, pivotal for enhancing model performance. In the code snippet furnished, the process of parameter refinement unfolds through the prism of hyperparameter tuning via grid search. This method orchestrates a systematic exploration across a predefined space of hyperparameters, evaluating the model's efficacy across assorted parameter combinations. Each classifier, be it Support Vector Machine, KNN, Decision Tree, or Logistic Regression, comes accessorized with a distinct set of hyperparameters. These knobs and dials regulate the model's behavior during training, wielding a direct influence over its predictive prowess. Grid search, coupled with cross-validation, navigates through the hyperparameter maze by fragmenting the training data into multiple folds, scrutinizing performance metrics like accuracy, precision, recall, and F1-score. The crème de la crème model, crowned by grid search, is then put through its paces on the test set to furnish an impartial appraisal of its real-world performance. Furthermore, the creation of confusion matrices lends a visual scaffold, delineating the classifier's efficacy in terms of true positives, false positives, true negatives, and false negatives, thus elucidating the classification panorama. In essence, the art of tweaking model parameters, epitomized by techniques like grid search, ensures that machine learning models are finely calibrated to the idiosyncrasies of the dataset, culminating in heightened performance and resilience in practical applications.

2.2 Test phase

After training is complete, we evaluate the final model on the test set to get an unbiased estimate of its performance on unknown data. Each value in the test set is passed to the model to obtain classification predictions. The predictions are presented in the form of probabilities for each class. We calculate the evaluation metrics (accuracy) and analyze the results obtained to evaluate the performance and generalization ability of the model.

3. Experimental results

SVM

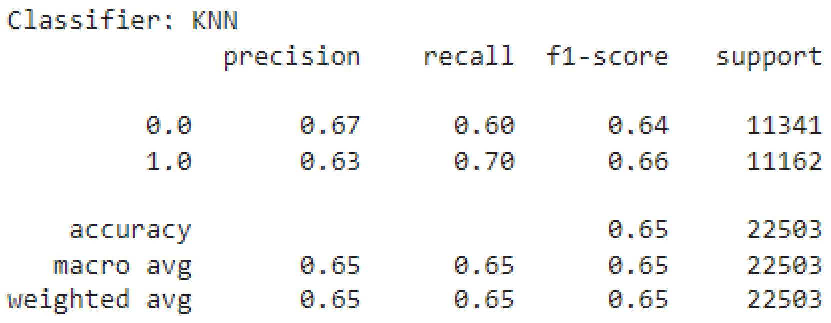


```
Classifier: SVM
              precision    recall  f1-score   support

         0.0       0.67      0.56      0.61     11341
         1.0       0.62      0.72      0.67     11162

    accuracy                           0.64     22503
   macro avg       0.64      0.64      0.64     22503
weighted avg       0.64      0.64      0.64     22503
```

KNN

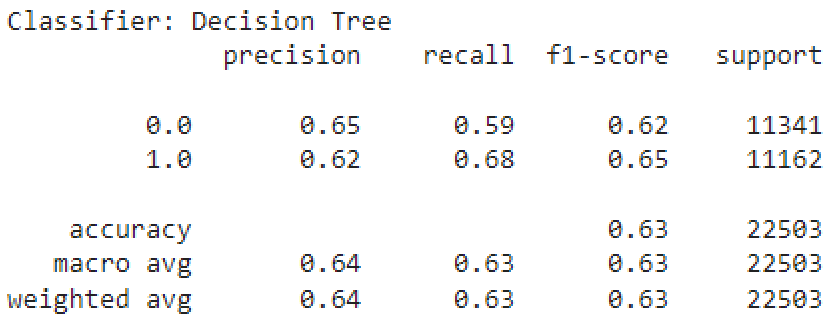


```
Classifier: KNN
              precision    recall  f1-score   support

         0.0       0.67      0.60      0.64     11341
         1.0       0.63      0.70      0.66     11162

    accuracy                           0.65     22503
   macro avg       0.65      0.65      0.65     22503
weighted avg       0.65      0.65      0.65     22503
```

DECISION TREE

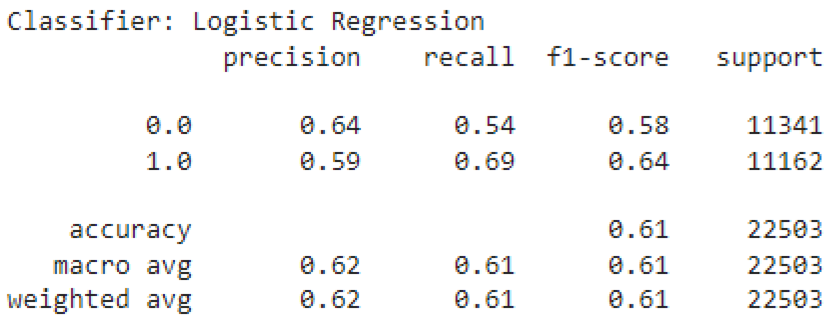


```
Classifier: Decision Tree
              precision    recall  f1-score   support

         0.0       0.65      0.59      0.62     11341
         1.0       0.62      0.68      0.65     11162

    accuracy                           0.63     22503
   macro avg       0.64      0.63      0.63     22503
weighted avg       0.64      0.63      0.63     22503
```

LOGISTIC REFRESSION



```
Classifier: Logistic Regression
              precision    recall  f1-score   support

         0.0       0.64      0.54      0.58     11341
         1.0       0.59      0.69      0.64     11162

    accuracy                           0.61     22503
   macro avg       0.62      0.61      0.61     22503
weighted avg       0.62      0.61      0.61     22503
```

CHAPTER4

IMPLEMENTATION

Introduction

In this annex, we are presenting our development tools, describeing the different steps for implementing the electricity theft detection systhem.

1 The platform and development environments

Google colaboratory

FIG.4.1: Google colaboratory logo.

In order to obtain concrete and scalable results in our study, we did heavy use of Google Colab, also known as Colab. Colab is a cloud service offered by Google which allows you to reproduce the environment Jupyter Notebook in the cloud. It provides many users with powerful graphics processors (GPUs), which is particularly advantageous for those who cannot afford to form learning projects remote automatic (Web 08). The table below presents the characteristics of the environment :

CPU Model	2-core Xeon 2.2GHz
RAM	25 GB
Disque dur	225 GB

Tab.1: Characteristics of the environment.

2 Python language

Python is an open source programming language belonging to the category interpreted languages. It allows developers to focus on actions that they carry out rather than on the different methods of carrying them out. Compared to compiled languages, Python offers programmers considerable time savings. It is widely used, appreciated and in high demand in the IT field. Python is a dynamically typed language, which means that the type of a variable can be modified. Annotations are used to specify types of arguments and return values of functions. Python is commonly used for web development, data analysis, artificial intelligence, neural networks, scientific computing and other areas of computing advanced. Programming is undoubtedly the main area of use of this language, which is suitable for both simple and complex projects(Web 9).

3 Libraries used

These libraries were used for the implementation and analysis of the models of classification of histopathological images of breast cancer based on Transformers, as well as for data manipulation and evaluation of results.

3.1 NumPy

NumPy, which is short for “Numerical Python”, is a library used in scientific programming in Python, mainly to manipulate digital data. It offers multidimensional data structures under form of tables, as well as a set of integrated tools to facilitate implementation in Python. NumPy essentially combines the advantages of the C language and Python, by processing numerical data in the form of tables to perform multidimensional operations and data manipulations (Web 10).

FIG. 4.2: Numpy logo.

3.2 Pandas

Pandas is a Python library used for data analysis. His development was motivated by the need to have a powerful and flexible tool to quantitative analysis, which allowed it to become one of the Python libraries the most popular. Pandas benefits from a very active community of contributors, which regularly contributes to its improvement and evolution (Web 11).

FIG. 4.3: Pandas logo.

3.3 TensorFlow

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML-powered applications.

TensorFlow was originally developed by researchers and engineers working within the Machine Intelligence team at Google Brain to conduct research in machine learning and neural networks. However, the framework is versatile enough to be used in other areas as well.

TensorFlow provides stable Python and C++ APIs, as well as a non-guaranteed backward compatible API for other languages(web 12).

FIG. 4.4: TensorFlow logo

3.4 Matplotlib

Matplotlib is widely recognized as the leading visualization and exploration library most popular data source. It offers a wide range of tools allowing create basic graphics, such as line graphs, clouds point charts, histograms, bar charts and pie charts.

Matplotlib forms the foundation for many other visualization libraries. It is a tracing library designed specifically for the language of Python programming and its numerical extension NumPy. Using Matplotlib, users can visualize patterns, trends and correlations which might not be detected by simply looking at text data(Web 13).

3.5 Seaborn

Seaborn is built on Matplotlib. Used to draw statistical graphs attractive and informative. Seaborn is also a specialized support for categorical variables to show observations. Seaborn has tools to select color palettes that reveal hidden patterns in the data (Web 14).

Key Features of Why We Use Seaborn:

– Feature: Uses less syntax and offers default themes simple and interesting,

– Flexibility: Provides the most used default themes,

– Management of several figures: Automates the creation of several figures that can cause low memory issues.

3.6 Sk-learn

Scikit-learn is an open-source machine learning library for Python that provides simple and efficient tools for data mining and data analysis. It is built on top of NumPy, SciPy, and matplotlib, and offers a wide range of algorithms for supervised learning (such as classification, regression, and clustering) and unsupervised learning (including various clustering algorithms). Scikit-learn also includes tools for model selection, like cross-validation and grid search, as well as metrics for model evaluation. Additionally, it provides data preprocessing, feature extraction, and feature selection functionalities, making it a comprehensive library for machine learning tasks.

3.7 Imbalanced-learn (imblearn)

Imbalanced-learn is an open-source Python library designed to handle imbalanced datasets, a common issue in machine learning where some classes are underrepresented. It offers various resampling techniques, such as oversampling (e.g., SMOTE), undersampling, and hybrid methods to balance class distribution. The library is compatible with scikit-learn's pipeline and other functionalities, providing a seamless integration for users. Imbalanced-learn's simple and intuitive API makes it easy to apply different resampling techniques, enhancing the performance of machine learning models on imbalanced data.

3.8 SciPy (spicy)

SciPy is an open-source Python library used for scientific and technical computing. Building on NumPy, it offers a vast array of higher-level functions for mathematical, scientific, and engineering computations. SciPy includes modules for optimization (such as function optimization, linear programming, and least squares), integration (including numerical integration and ordinary differential equation solvers), advanced linear algebra routines (like decompositions and eigenvalue problems), statistical functions and probability distributions, as well as signal processing tools (such as filtering and Fourier transforms). SciPy is a fundamental library for scientific research and technical applications in Python.

CHAPTER1

PROJECT SUBMISSION

Introduction

Presenting a successful project idea requires meticulous planning and thoughtful consideration. The idea must possess specific characteristics that contribute to its success. It should be innovative, offering a unique solution to existing problems, and meeting a clear societal need. The idea must stand out from existing projects by incorporating new elements and a unique methodology that adds real value to consumers. Moreover, the idea must be feasible and sustainable, with the necessary financial, human, and technological resources available for effective implementation. From a profitability standpoint, the idea should demonstrate the potential to achieve a significant financial return.

This chapter focuses on presenting our project idea, the team that will carry out its implementation, defining the project's desired goals, and establishing a timeline for achieving them.

1. The idea of the project

The idea of the project is a web application called DUANKOU, it detects electricity theft, which mean catches energy thieves. The idea of the project began after seeing the statistics of financial losses caused by non-technical energy losses.

Every year, Algeria loses approximately 300 million Algerian centimes, and this is according to what was published by Al-DJomhouria.dz newspaper about the Electricity and Gas Distribution Directorate of the state of Sidi Bel Abbes on January 26, 2024.

2. Suggested values

2.1. Explain how the application works

DUANKOU application is an easy-to-use application with a simple interface, which mean you do not have to employ people specialized in programming or computer science to use it.

Old clients in charge of computers can suffice. The application is designed to accept modifications according to the company's desires and, as their information is preserved and no one can access it.

2.2. Field of activity

An application to monitor electricity thefts falls within the scope of energy management or facilities management software. It is designed to help energy companies detect and prevent instances of electricity theft, thus improving revenue protection and ensuring equitable distribution of resources.

This application has never existed in Algeria, as electricity thefts are detected by sending patrols to check the meters to see if they have been modified, and therefore this application will continue with effort and time. Through it, these teams that cost a lot of money can be compensated with one free application used by the company.

Since it is an electronic application, it can be modified according to developments in thefts and the needs of energy companies.

3. The project owner

Halais Yousra Master2 Student specializes in computer science and artificial intelligence.

My role is:

- Programming the artificial intelligence model,
- Designing user-friendly interface.

4. Work team

Backend developer: handles server logic, databases, and APIs, ensuring security and performance.

Electrical specialist: Explain electricity distribution to create a well-labeled database for precise algorithms.

Marketing employee conducts market research, develops the brand, plans and executes marketing campaigns, engages with potential customers through promotions and live demonstrations, and implements sales strategies to drive subscriptions and licensing deals.

Lawyer ensures legal compliance, focusing on data privacy, security, and regulatory adherence, while also protecting intellectual property and drafting user agreements. Together, they support the project's financial health and legal integrity.

Accountant manages financial records, budgeting, and resource allocation, ensuring the project stays within budget and complies with financial regulations.

5. Project goals

DUANKOU application seeks to achieve the following strategic goals:

- Through this project, we seek to become the first application in Algeria to detect electricity theft using artificial intelligence and keeping pace with the times,
- We aim to become the first competitor for various applications or platforms that complement and serve our field,
- Striving to gain the trust and support of the state and investors through the credibility, transparency, procedures and transactions the platform provides, in order to make the DUANKOU application the only platform at the level Algeria as a whole, which specializes in the field of energy protection, and this calls us to be a reason for Eliminate all complaints that reach the Consumer Protection Authority as well, in order to reduce the rate Illegal behavior that violates official laws such as the Competition Law, the Consumer Protection Law, and all the laws and regulations that fall within the context of our project.
- use the AI model and backend architecture designed for anomaly detection in the application can be customized with UI adjustments to serve multiple industries. For instance, in healthcare, it could monitor patient vital signs for anomalies. In finance, it might detect fraudulent transactions. In manufacturing, it could identify equipment malfunctions. By tailoring the user interface, this technology can be effectively deployed across various sectors to enhance anomaly detection capabilities.

my application DUANKOU also seeks to achieve many other goals that we can highlight Below:

- It aims to detect electricity thefts and thus avoid material losses that affect the economy of the country as a whole,
- Reducing theft and thus fighting corruption,

- Reducing companies' burdens of material costs and human resources by replacing detection teams with just one application.

6. Timeline for project realization

The ultimate goal is to create a professional and high-quality application under the name DUANKOU.

To achieve this goal, we must rely on a set of steps and tasks, which are detailed as follows:

First task: Planning and requirements definition

We will begin with defining the project objectives starting with designing the application interface and specify all the need of the application (programming language, development, ...)

Second task: Data preparation

At this stage starting with collecting data to create our dataset or searching for an appropriate database already existed that can aguerd our needs.

Third task: Development and implementation of the application

At this stage, the task of developing the application, this includes building the user interface, the database preprocessing, and developing the ai model the electricity theft detector.

Firth task: Test and pilot the application

At this stage, the application is tested and piloted to ensure that it runs properly according to the requirements, and all functions and processes offered are also tested and reviewed. results to ensure its quality and smooth and effective operation.

Fifth task: Official introduction of the application into service

After going through all the previous stages and after making sure that the application is operating according to what was planned in advance, work can be started by collaborate with energy companies and do the necessary modification according to their desire.

Sixeth task: Training and employing employees

At this stage, the company's work team must be trained by us to use the application correctly.

Seventh task: Continuous monitoring and improvement

Once the application is operational, it must monitor the overall performance and constantly improve it according to guidelines, advice, or instructions. The efficiency of operations is also evaluated and data is analyzed periodically to identify areas that can be improved or modified.

Tasks				**Duration**				
Main tasks	**Scondery task**			**1Month**	**2Month**	**3Month**	**4Month**	**5Month**
First task: Planning and requirements definition				×				
Second task: Data preparation				×				
Third task: Development and implementation of the application	Ia mode building	App interfae building	imple mantat ion	×		×		
Firth task: Test and pilot the application							×	
Fifth task: Official introduction of the appplication into service								×
Sixeth task: Training and employing employees								×
Seventh task: Continuous monitoring and improvement								×

CHAPTER 2

INNOVATIVE ASPECTS

Introduction

In our project, we placed significant emphasis on innovation, recognizing it as the critical factor for our success. We diligently employed a variety of methods and technologies to foster innovation, maintain a competitive edge, and effectively compete in the market. This chapter focuses on the nature and scope of the innovations incorporated into our project.

1.The nature of innovation

Our project leverages multiple types of innovations to achieve its goals, ensuring we stay ahead of the market and compete effectively. These innovations are categorized into three main areas: radical innovations, market innovations, and technological innovations

1.2. Radical innovations

Radical innovations form the backbone of our project, transforming the way we approach theft detection and utility management:

- Automates theft detection using data analysis and machine learning, replacing manual inspections,
- Provides real-time monitoring and predictive capabilities based on usage patterns,
- Enhances scalability and efficiency, reducing operational costs,
- Offers user-friendly dashboards and real-time alerts for quick decision-making,
- Ensures secure data handling and supports regulatory compliance with detailed reporting.

1.2. Market innovations

- Introduces a new service for automating electricity theft detection, reducing reliance on manual inspections,
- Significantly lowers operational costs for utility companies, enhancing profitability,
- Increases customer trust by ensuring fair billing and reducing unauthorized usage,
- Differentiates utility companies as innovative and forward-thinking in the market,
- Offers scalable solutions adaptable to various regions and utility sizes,
- Creates partnership opportunities with smart meter manufacturers and data analytics firms,

- Improves revenue protection for utilities by minimizing losses due to theft,
- Assists utilities in meeting regulatory requirements and reducing compliance risks,
- Engages customers with tools to monitor their own usage, fostering trust and reducing disputes,
- Addresses a universal challenge, providing global market potential for the technology.

1.3. Technological innovations

Technological innovations are the driving force behind our project, enabling us to develop advanced solutions and maintain a competitive edge:

- Advanced data analysis algorithms for anomaly detection,
- Implementation of machine learning models like autoencoders,
- Real-time monitoring systems for data processing,
- Scalable infrastructure, such as cloud computing,
- Robust security measures for data privacy,
- Integration with IoT devices and sensors,
- Predictive analytics for forecasting theft activities,
- Automated reporting systems for alerts and notifications,
- Continuous learning algorithms for improved accuracy,
- Tools for regulatory compliance and standards adherence.

2. Areas of innovation

Our project encompasses various areas of innovation, each contributing to a more effective and comprehensive solution for electricity theft detection. These areas include detection technology, data analysis, automation, integration, security, scalability, user experience, regulatory compliance, predictive capabilities, and continuous improvement.

Innovating in:

- more accurate and efficient detection technologies,
- analyzing electricity usage data to identify anomalies,
- automating the detection process to improve efficiency,
- integrating various technologies for a comprehensive detection system,
- ensuring data security throughout the detection process,
- creating scalable solutions adaptable to different environments,

- Innovating in designing user-friendly interfaces for operators and customers,
- solutions to meet regulatory requirements,
- predictive analytics to forecast theft activities,
- mechanisms for ongoing algorithm enhancement.

CHAPTER3

STRATEGIC MARKET ANALYSIS

Introduction

To ensure the success of DUANKOU application in combating electricity theft, we need a thorough analysis of the electricity distribution market and a deep understanding of user needs. This process involves studying competitors' methods and performance to identify their strengths, weaknesses, and challenges.

Understanding user preferences for theft detection and reporting is essential, which we will achieve through comprehensive market research and surveys. These insights will inform a targeted marketing and development strategy for the "Catcher" application, focusing on attracting users and meeting their expectations.

Maintaining a competitive edge requires continuous monitoring and improvement of the platform. This approach ensures we understand the target market, assess competition effectively, and craft strategies that attract and retain users while delivering ongoing value.

1. The market sector

1.1 Potential market

Institutions Likely to Request an application to Detect Electricity Thefts:

- Utility Companies,
- Government Regulatory Bodies,
- Law Enforcement Agencies,
- Industrial and Commercial Entities,
- Smart Grid Technology Providers,
- Residential Communities and Property Managers.

1.2 Motivations:

Utility companies are motivated by financial losses due to electricity theft, impacting revenue and the integrity of power distribution networks. Government regulatory bodies aim to ensure fair pricing, efficient energy markets, and consumer protection by preventing illegal activities. Law enforcement agencies seek to curb electricity theft as part of broader efforts to combat criminal activities. Industrial and commercial entities want to avoid being billed for stolen electricity, protecting their financial interests and operational efficiency. Smart grid technology providers respond to the demand for solutions that enhance power grid efficiency and security, including theft detection. Residential communities and property managers aim to prevent inflated communal electricity costs and ensure fair charge distribution. Overall, these groups are driven by financial protection, legal compliance, operational efficiency, and technological advancement.

2. Measuring the degree of competition

Measuring the degree of competition in the market is essential to understand the unique advantages and challenges faced by our electricity theft detection application. This evaluation will help identify our strengths that can be leveraged and weaknesses that need addressing.

2.1 Strong Points

Our application has several strong points that position it favorably in the market:

- First-mover advantage in introducing electricity theft detection in Algeria,
- Addresses a significant issue affecting utilities and consumers,
- Utilizes modern technology such machine learning algorithms,
- Solution tailored to local issues and regulatory requirements.
- Leverages growing community of startup programmers for development and support.
- Potential for collaboration with local utility companies and government bodies.
- Scalable design for future expansion to other regions.
- Ability to add new features or integrate with other systems as needed.

2.2 Weak Points

Despite our strengths, there are several weak points we need to address:

- Infrastructure limitations that may hinder implementation or effectiveness.
- Possible reluctance from utilities or consumers to adopt new technology.
- Lack of awareness or understanding of the application’s benefits.
- Ensuring compliance with data protection regulations and maintaining user privacy.
- Ensuring the system can scale effectively with increased usage and data loads.

- Navigating potential regulatory and bureaucratic obstacles.
- Managing costs associated with meeting regulatory and compliance requirements.

3. Market strategies

To effectively promote our electricity theft detection application and ensure widespread adoption, we have developed a comprehensive set of market strategies. These strategies aim to raise awareness, build credibility, and foster trust among utility companies, government officials, and the public.

- Utilize social media to highlight the issue of electricity theft and promote the benefits of your application.

- Conduct sessions for utility companies, government officials, and the public to demonstrate the application's functionality and benefits.
- Partner with local utility companies to endorse the application as a trusted solution for reducing electricity theft.
- Participate in and sponsor local tech events and hackathons to raise visibility
- Organize live demonstrations to build trust and credibility.
- Create a professional website and maintain active social media profiles for information, updates, and customer support.
- Publish articles, blog posts, and whitepapers about electricity theft and the application.
- Collect and share testimonials from early adopters and successful pilot programs.
- Engage local community leaders to advocate for the application’s benefits.
- Tailor marketing materials to the local context and language.
- Actively seek feedback from users to improve the application.

CHAPTER4

PRODUCTION AND ORGANIZATION PLAN

Introduction

To successfully develop and deploy DUANKOU application for combating electricity theft, a comprehensive and well-structured approach is essential. This involves meticulously planning the production process, ensuring adequate supply and labor resources, and collaborating with key stakeholders. Below is a detailed outline of our strategy, including introductions for each section.

1. The production process

The production process encompasses all steps from initial planning to post-deployment, ensuring the application is developed efficiently and effectively meets user needs.

1.1 Requirement Analysis and Planning

- Identify Stakeholders: Engage with utility companies, regulatory bodies, and end-users to understand their needs.
- Define Objectives: Clarify the primary goals, such as reducing electricity theft, improving detection accuracy, and optimizing resource allocation.
- Scope and Specifications: Detail the project scope, including the types of data required, key features, performance metrics, and compliance requirements.

1.2 Data Collection and Preparation

- Data Sources: Gather data from smart meters, billing systems, customer records, and historical theft incidents.
- Data Types: Ensure a mix of time-series data (electricity usage patterns), customer demographic data, and environmental data.
- Data Cleaning: Handle missing values, outliers, and errors to ensure data quality.
- Data Labeling: Label data with theft and non-theft instances for supervised learning models.

1.3. Model Development.

- Model Selection: Choose our AI model, which is deep learning model (stacked denoising autoencoder).

- Training and Validation: Split the data into training and validation sets. Train the model on the training set and fine-tune hyperparameters.

1.4. Implementation and Integration

- System Architecture: Design the architecture for the application, including data pipelines, processing units, and storage.
- Backend Development: Implement the backend using appropriate technologies and frameworks (e.g., Python with TensorFlow/PyTorch for AI, Django/Flask for web backend).
- Frontend Development: Create a user-friendly interface for users to interact with the system (e.g., dashboards for utility companies).
- API Integration: Develop APIs to integrate with existing systems (e.g., utility billing systems, smart meter networks).

1.5. Testing

- Unit Testing: Test individual components of the application for functionality.
- Integration Testing: Ensure all components work together seamlessly.
- Performance Testing: Evaluate the system's performance under various conditions and loads.
- Security Testing: Ensure the application is secure against vulnerabilities and unauthorized access.

1.6. Deployment

- Environment Setup: Prepare the deployment environment (cloud-based or on-premises servers).
- Continuous Integration/Continuous Deployment (CI/CD): Implement CI/CD pipelines for automated testing and deployment.
- Monitoring and Maintenance: Set up monitoring tools to track system performance and detect issues in real-time.

1.7. Post-Deployment

- User Training: Provide training for end-users and stakeholders.
- Feedback Loop: Collect feedback from users for continuous improvement.
- Updates and Scaling: Regularly update the system with new features, improvements, and scalability measures.

2. Supply

Ensuring a reliable supply of resources is crucial for the smooth development and operation of the "Catcher" application.

- Human Resources: Domain experts, project managers.
- Hardware: Servers, workstations, storage solutions, network infrastructure, IoT devices (smart meters), GPUs.
- Data Sources: Historical data, real-time data from smart meters, external data sources.
- Licenses and Subscriptions**:** Software licenses, cloud services (AWS, Azure), data subscriptions.
- Infrastructure: Cloud infrastructure, on-premises infrastructure, security measures (firewalls, VPNs).
- Miscellaneous Supplies: Office space, collaboration tools (Slack, Teams), project management tools (Jira, Trello).

3. Labor

DevOps engineers are essential for managing the deployment pipeline, ensuring system scalability, and maintaining CI/CD processes. They handle the cloud infrastructure, container orchestration, and overall system reliability. Their responsibilities include.

3.1. DevOps Engineers

Manage the deployment pipeline, ensure system scalability, reliability, and maintain CI/CD processes. Handle cloud infrastructure and container orchestration.

Proficiency in CI/CD tools (Jenkins, GitHub Actions), containerization (Docker), orchestration (Kubernetes), cloud platforms (AWS, Azure, Google Cloud).

3.2. Domain Experts

Specialized expertise or short-term needs, such as cybersecurity experts to ensure data security.

Provide insights into electricity distribution and theft detection, help in understanding the problem domain, and validate the AI models.

Expertise in electricity distribution systems, knowledge of typical electricity usage patterns and theft indicators, and ability to interpret model outputs.

3.3. Project Managers

Coordinate the project, manage timelines, ensure communication between different teams, and keep the project on track.

Strong organizational and leadership skills, experience with project management tools (Jira, Trello, Asana), excellent communication skills.

4. Main companies

The parties that can help us in implementing the project and are considered intervening parties, carriers, Business incubators, banks, National Commercial Registry Center, Tax Directorate, investors, Exporters and importers…

CHAPTER5

FINANCIAL PLAN

I Introduction

In this chapter we will discuss the financial aspects of a project aimed at developing and implementing an application for a specific purpose. It covers various fees, costs, and revenue streams associated with the project, providing insights into its financial viability and sustainability over time

1. Costs and burdens

This section details the financial investments required for the project's development, maintenance, and promotion. It includes one-time expenses such as development costs, computer science equipment, promotional activities, and expert consultancy fees, as well as recurring costs like hosting, infrastructure, and ongoing maintenance and support.

The technical team will be engaged as partners for a one-time collaboration, and their compensation will be structured as consulting fees, which will encompass their salaries.

Development Costs.	**2 000 000.00 dz** (one-time)
Hosting and infrastructure.	**48 000.00dz** per year
Ongoing maintenance and support.	**600 000.00 dz** per year
computer science equipment	**250 000.00 dz** (one-time)
Promotional activities and direct sales efforts.	**200 000.00 dz** (one-time)
Expert computable and layer	**72 000.00 dz** (one-time)
Total	**3 170 000.00 dz**

2. Business Number

This section outlines the anticipated sources of revenue for the project, including one-time and recurring streams such as the sale of the application, subscription fees, consulting fees, and licensing. Each revenue stream contributes to the overall financial sustainability of the project.

Sale of the Application	**10 000 000.00 dz** (one-time)
Subscription Model	**360 000.00 dz** per year

Consulting Fees	**400 000.00 dz** per year
Licensing	**200 000.00 dz** per year

3. Expected Project revenue

Here, the projected revenue for both the first year and subsequent years is analyzed in detail. It includes calculations for total revenue, total costs, and taxes such as VAT and TAP. The section provides insights into the financial performance of the project over time, considering both one-time and recurring revenue streams.

3.1. Revenue Projections for the Optimistic Scenario

Annual Revenue for the First year

Revenue	**Total costs**
10 960 000.00 dz	**3 170 000.00 dz**
Total revenue 7 790 000.00 dz	

Annual Revenue Subsequent Years

Revenue	**Total costs**
960 000.00 dz	**648 000.00 dz**
Total revenue 315 000.00 dz	

3.1. Revenue Projections for the Pessimistic Scenario

The pessimistic scenario involves no sales and the project being rejected by the electricity company. After one year, we will reassess the project, develop new promotional and marketing methods, assemble a new work team, and update the application to make it suitable for anomaly detection. In this case, the costs, revenue, the new optimistic and pessimistic scenarios will be the same.

CHAPTER 6
PROTOTYPE

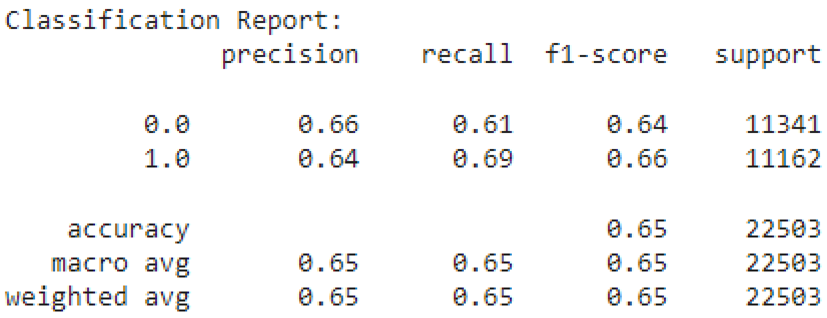
Introduction

In this chapter, we will explore the prototype of our application, including its user interface and functionality. As with any application, there is an admin who is responsible for managing the application. This application will have limited access permissions to ensure security and proper functionality.

1. Prototype

1.1 AI model

To build an AI model for anomaly detection in electricity theft detection, start by collecting historical electricity consumption data from smart meters, including normal consumption patterns and known theft cases. Preprocess the data by handling missing values, normalizing it, and engineering relevant features such as daily and seasonal consumption patterns. Split the data into training and testing sets, and handle class imbalance using SMOTE. Select an appropriate anomaly detection model like One-Class SVM, Autoencoders, or Isolation Forests, and train the model on the preprocessed data. Evaluate the model's performance using metrics such as accuracy, precision, recall, F1 score, and confusion matrix. Fine-tune the model by adjusting hyperparameters to improve performance. Finally, visualize the results using tools like Matplotlib and Seaborn to create plots such as ROC curves or confusion matrices.



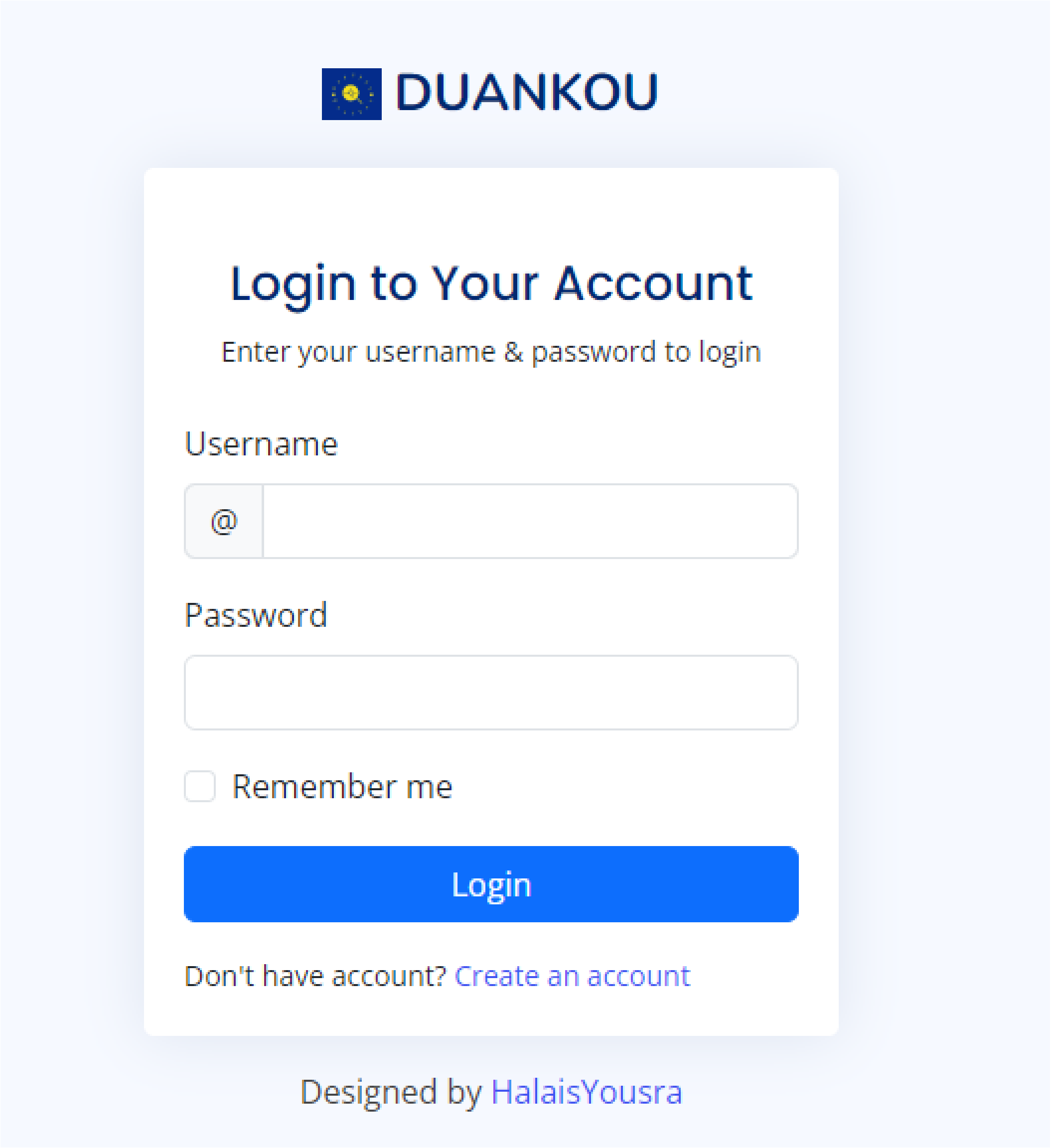
```
Classification Report:
              precision    recall  f1-score   support

         0.0       0.66      0.61      0.64     11341
         1.0       0.64      0.69      0.66     11162

    accuracy                           0.65     22503
   macro avg       0.65      0.65      0.65     22503
weighted avg       0.65      0.65      0.65     22503
```

1.2 User interface

This is the login page for the admin.



DUANKOU

Login to Your Account

Enter your username & password to login

Username

@

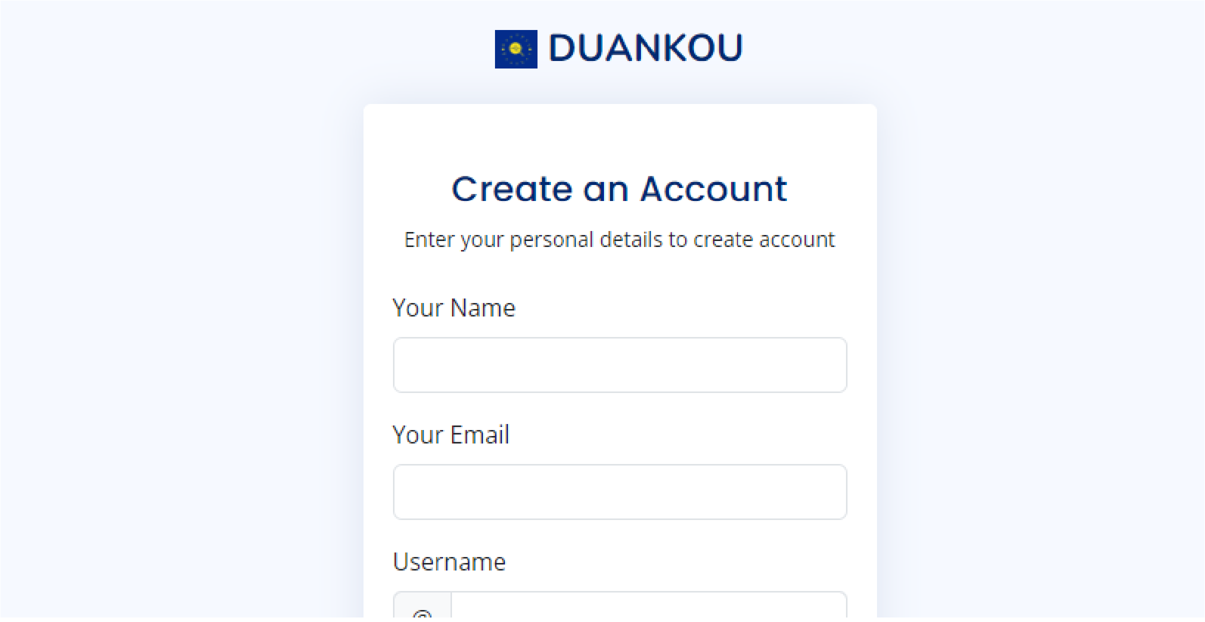
Password

□ Remember me

Don't have account? Create an account

Designed by HalaisYousra

Initially, the number of admin users will be limited to three. If a new admin does not have an account, they can click on "Create an account" to go to the account creation page.



DUANKOU

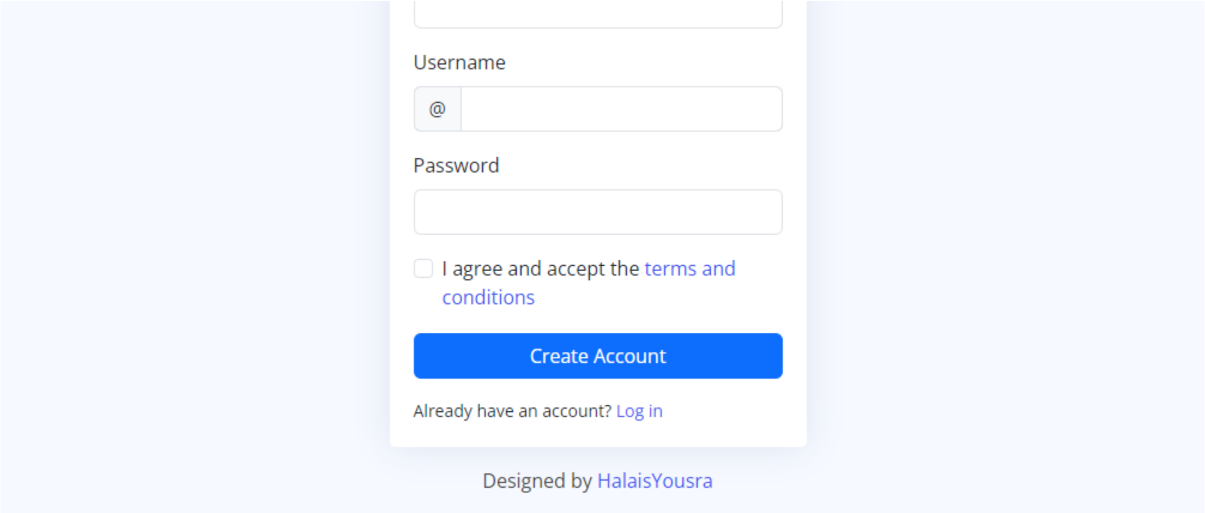
Create an Account

Enter your personal details to create account

Your Name

Your Email

Username



Username

@

Password

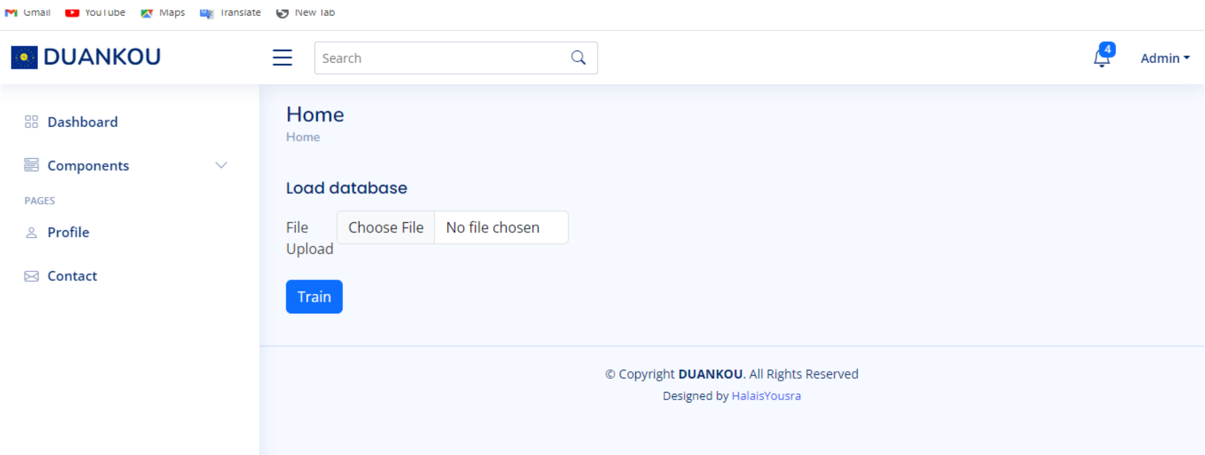
□ I agree and accept the terms and conditions

Create Account

Already have an account? Log in

Designed by HalaisYousra

Now when the admin log in he will find this page.



Home

Home

Load database

File Upload Choose File No file chosen

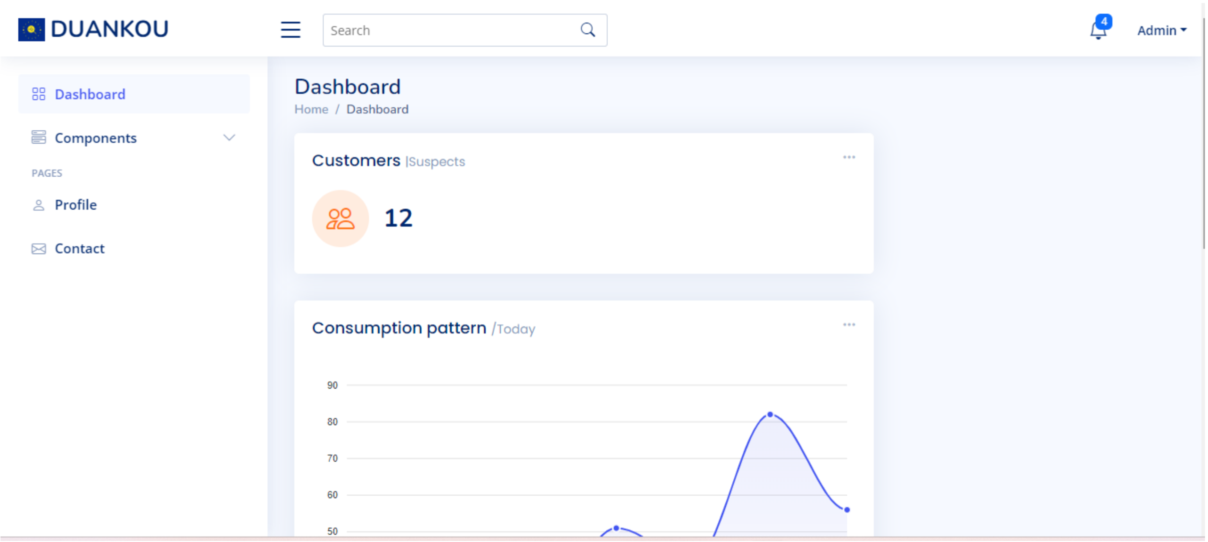
Train

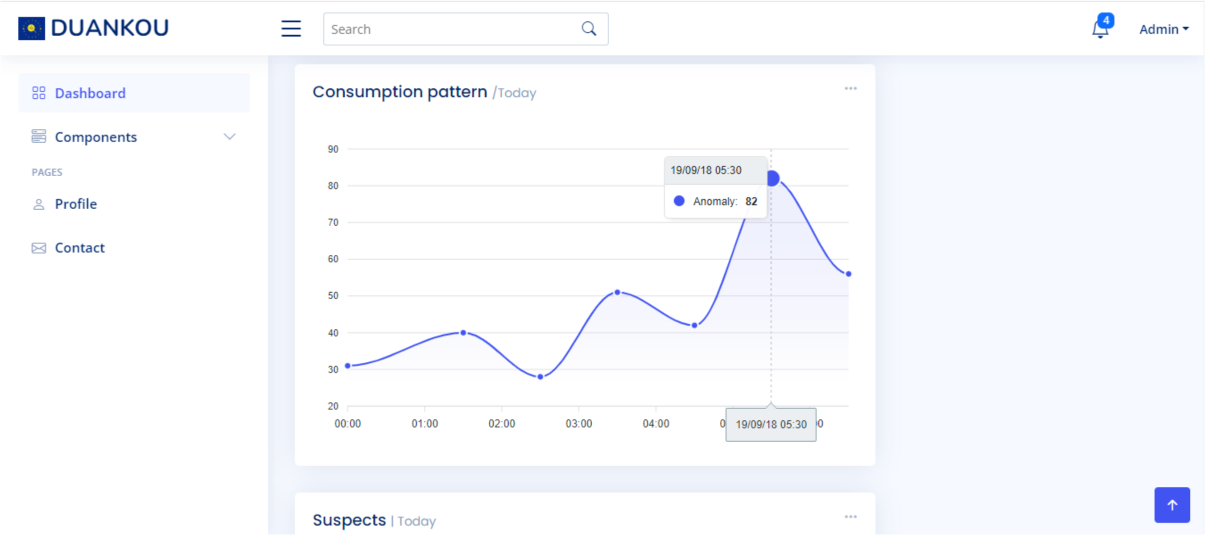
© Copyright DUANKOU. All Rights Reserved

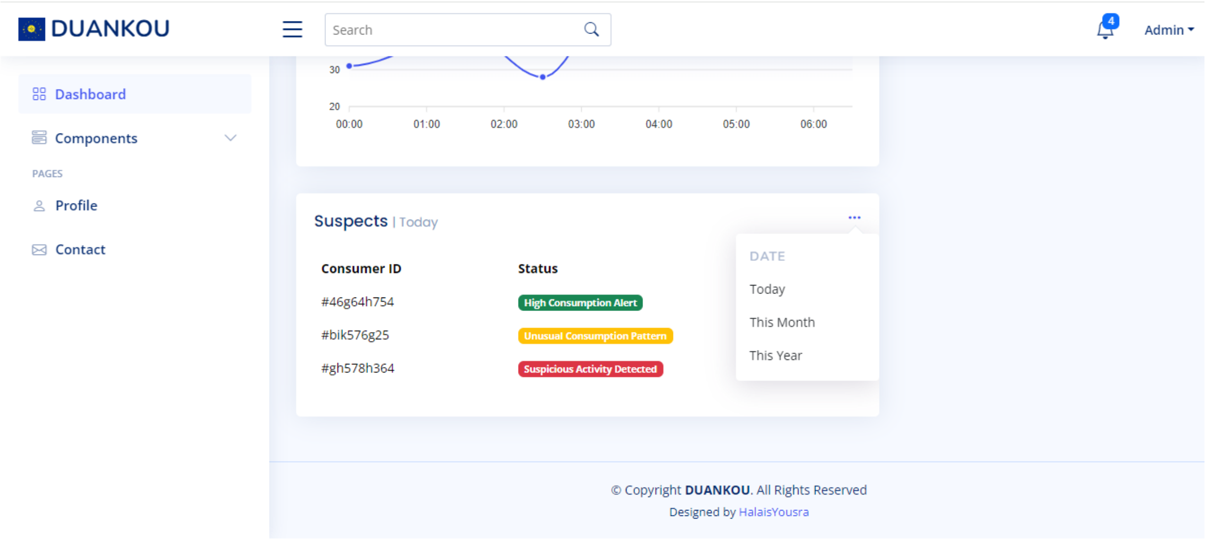
Designed by HalaisYousra

Here is where you will load your dataset, train it using the AI detector, and visualize the results on the dashboard. The dashboard consists of three boards: one displaying the number of suspected customers, another showing a usage pattern chart highlighting anomalies, and a third listing consumer names along with their status (green for normal, yellow for medium suspicion, and red for high suspicion). The status will be shown as notifications; for instance, clicking on the red notification will filter the suspects board to display only customers with high suspicion levels.

In the corner, there are three points displaying the date, allowing you to view the instances suspicion, this month’s suspicions, or the year’s suspicions, serving as a history filter.



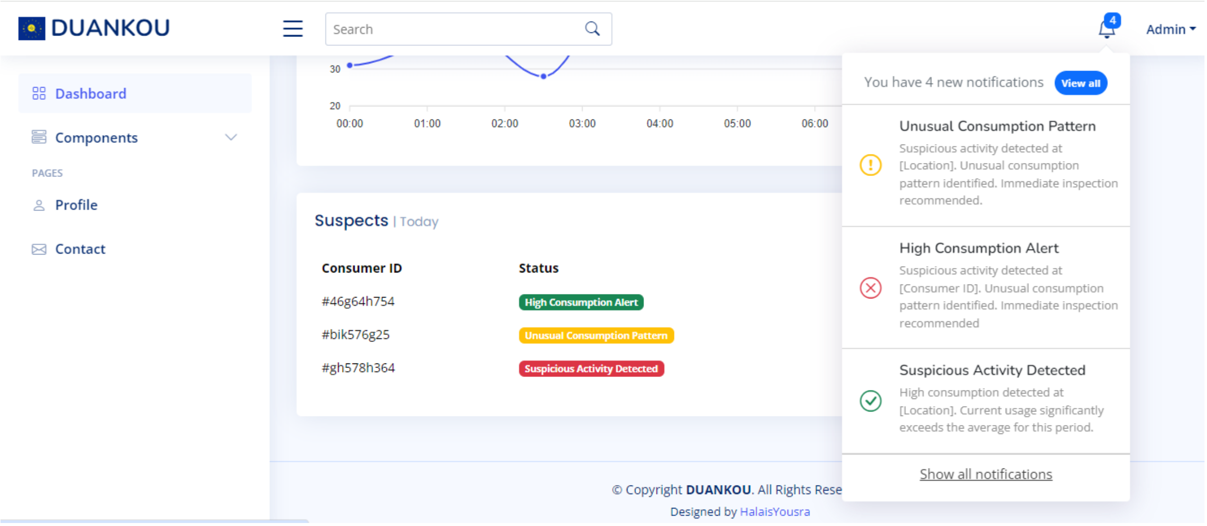




Suspects | Today

Consumer ID	Status
#46g64h754	High Consumption Alert
#bik576g25	Unusual Consumption Pattern
#gh578h364	Suspicious Activity Detected

© Copyright **DUANKOU**. All Rights Reserved
Designed by HalaisYousra



Suspects | Today

Consumer ID	Status
#46g64h754	High Consumption Alert
#bik576g25	Unusual Consumption Pattern
#gh578h364	Suspicious Activity Detected

You have 4 new notifications

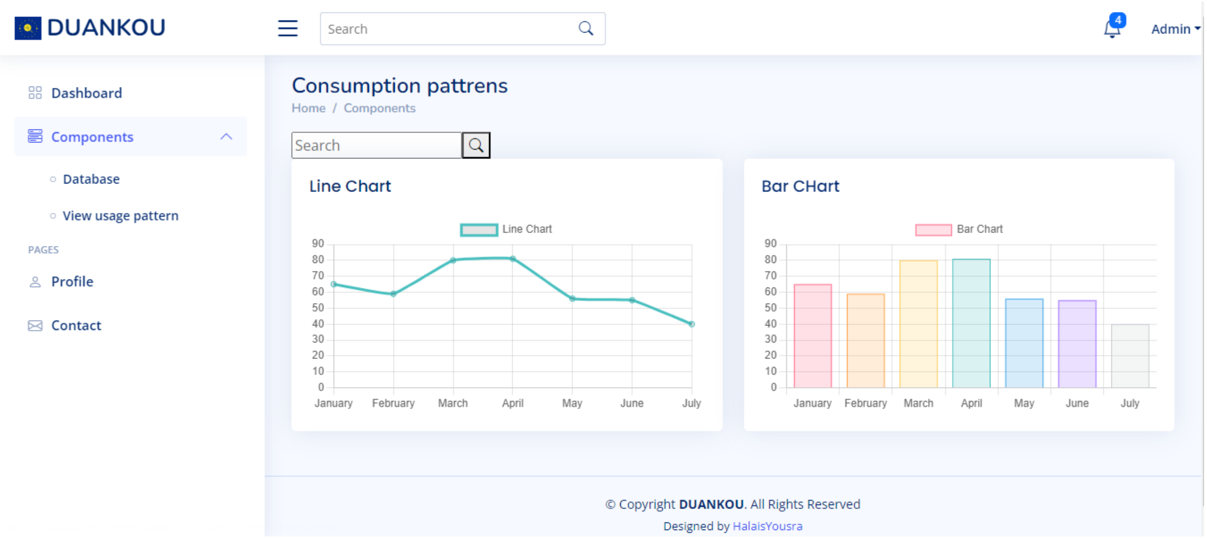
Unusual Consumption Pattern
Suspicious activity detected at [Location]. Unusual consumption pattern identified. Immediate inspection recommended.

High Consumption Alert
Suspicious activity detected at [Consumer ID]. Unusual consumption pattern identified. Immediate inspection recommended

Suspicious Activity Detected
High consumption detected at [Location]. Current usage significantly exceeds the average for this period.

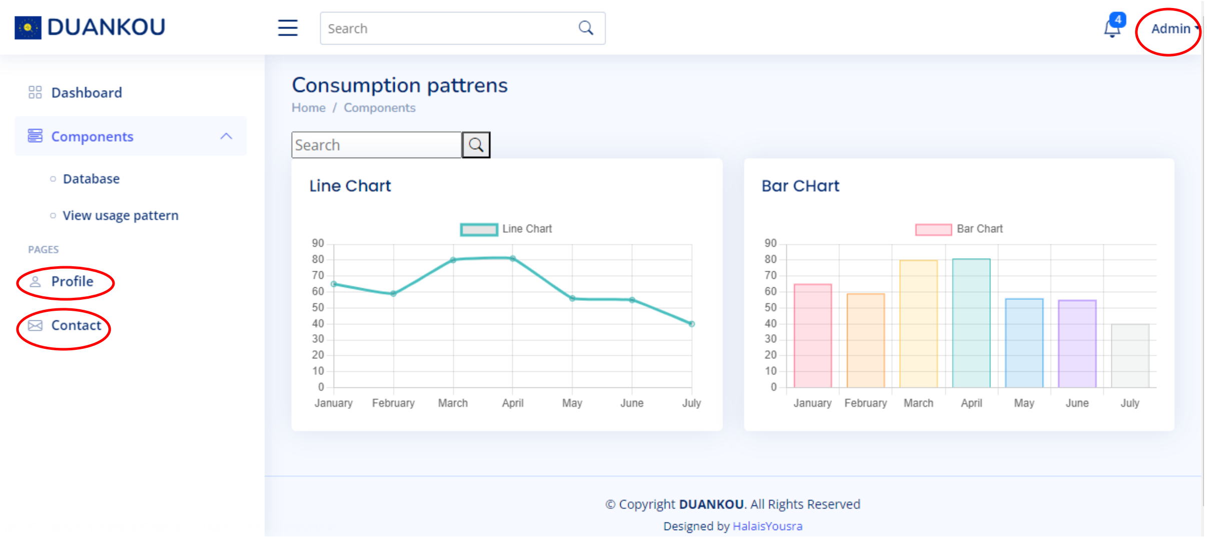
© Copyright **DUANKOU**. All Rights Rese
Designed by HalaisYousra

Now if you are looking for someone's usage pattern, you can click on "Components." Here, you will find "Database" (click on it if you want to change your database; it is the first page that appears when you log in) and "View Usage Pattern." The "View Usage Pattern" page allows you to check the consumption pattern of a customer by navigating to the search bar and entering the consumer's name.



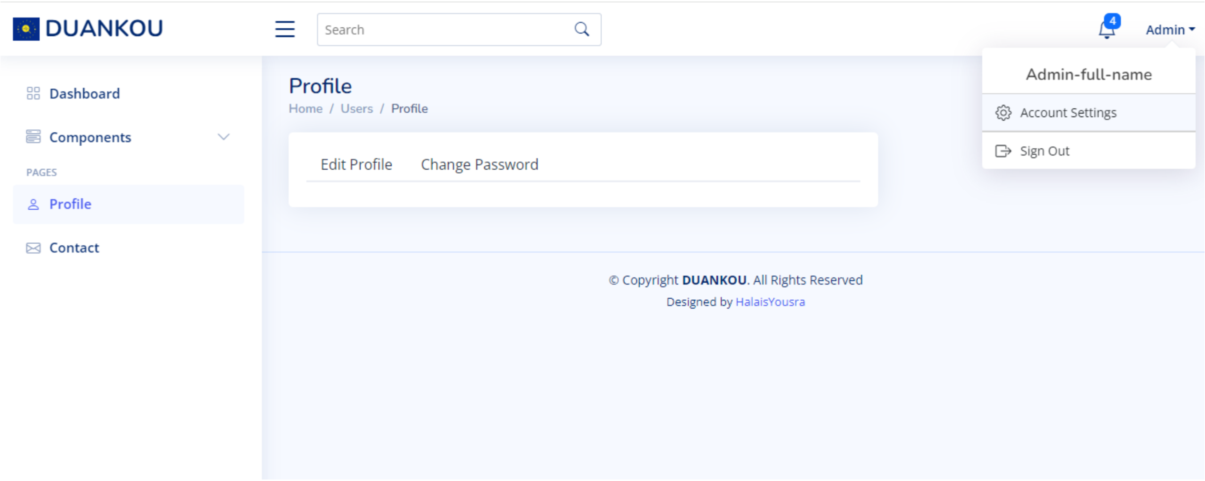
© Copyright **DUANKOU**. All Rights Reserved
Designed by HalaisYousra

Finally, there is three red-circled icons labeled:

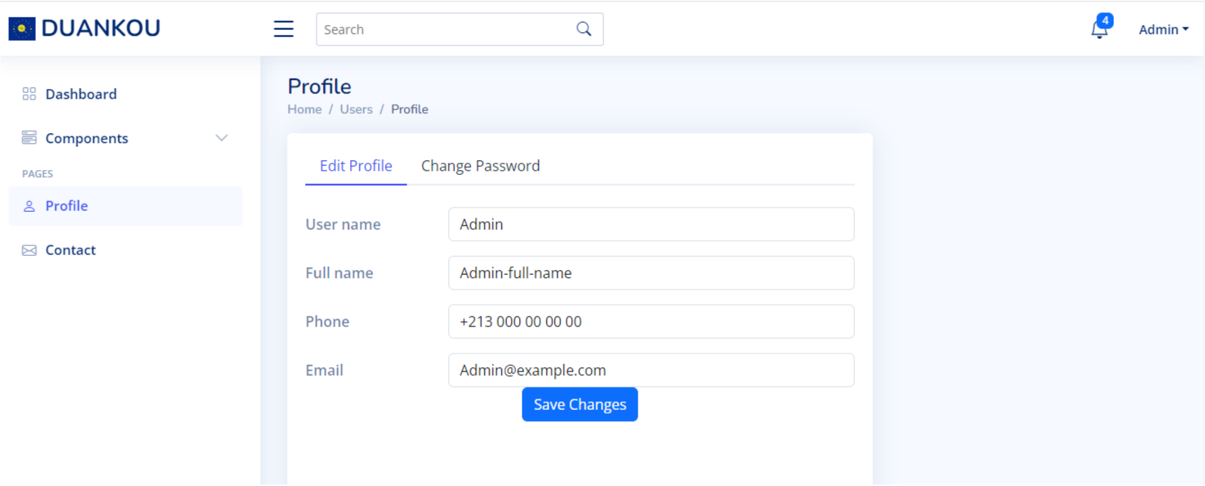


© Copyright **DUANKOU**. All Rights Reserved
Designed by HalaisYousra

"Admin." Here, you can log out or view and edit admin information by clicking on Account Settings, which takes you to the "profile" page where you can update these settings.



© Copyright **DUANKOU**. All Rights Reserved
Designed by HalaisYousra

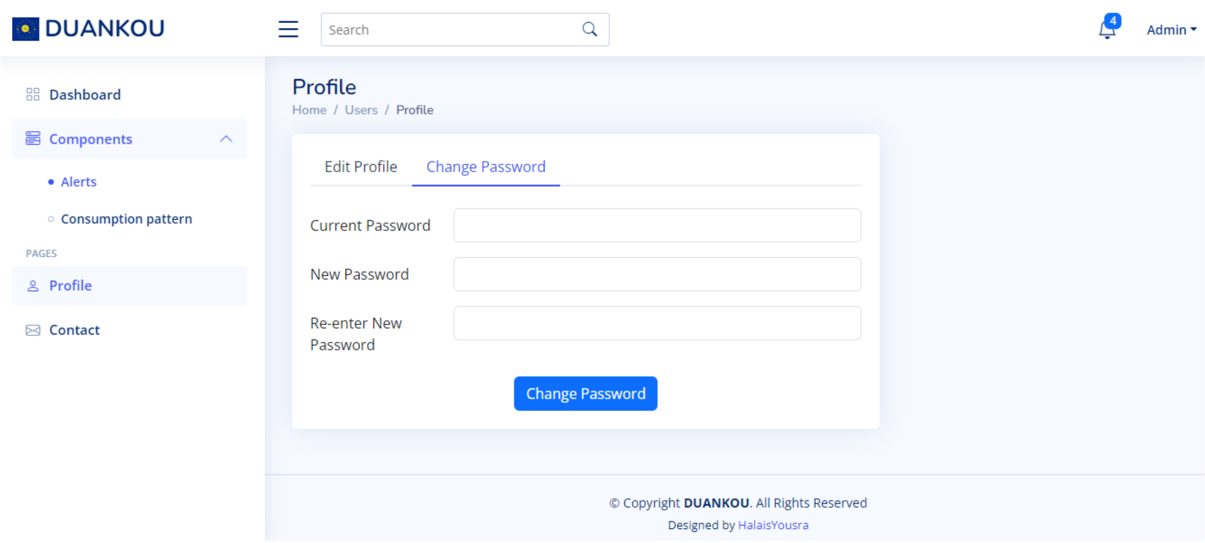


User name: Admin

Full name: Admin-full-name

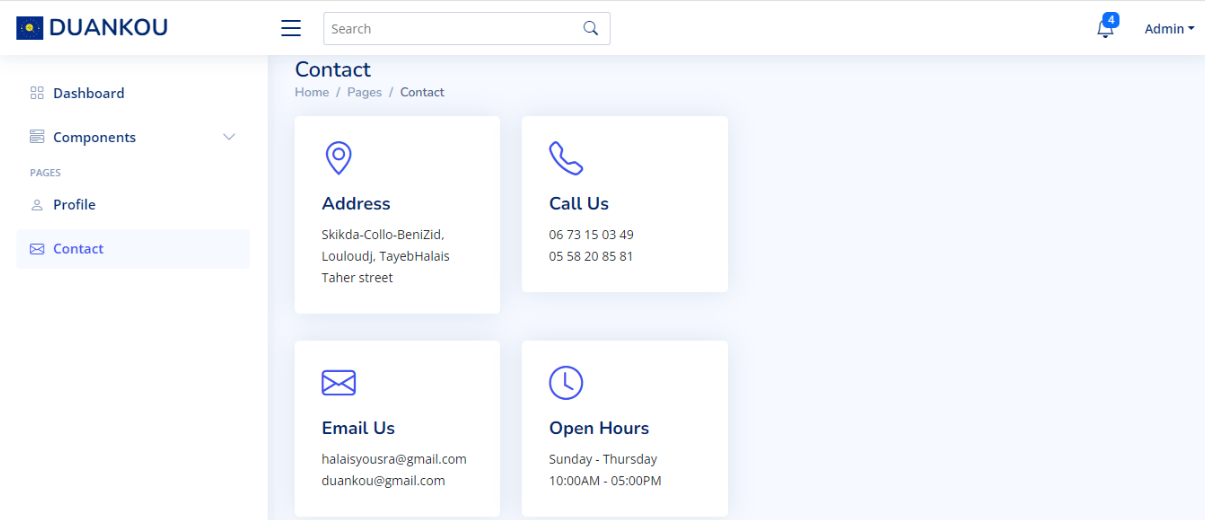
Phone: +213 000 00 00 00

Email: Admin@example.com



© Copyright **DUANKOU**. All Rights Reserved
Designed by HalaisYousra

For the "Contact" page containing the developer's (that's us) contact information.



Contact

Home / Pages / Contact

Address

Skikda-Collo-BeniZid,
Louloudj, TayebHalais
Taher street

Call Us

06 73 15 03 49
05 58 20 85 81

Email Us

halaisyousra@gmail.com
duankou@gmail.com

Open Hours

Sunday - Thursday
10:00AM - 05:00PM

BMC

This Business Model Canvas (BMC) outlines the key components of business strategy for developing and deploying an AI-based electricity theft detection system.

1.Key Partners

- **Government Company for Electricity and Gas Distribution (Sonelgaz):** The primary partner for deploying the system, providing access to electricity usage data.
- **Local Hosting Providers:** Ensure reliable hosting of the application within the country.
- **Cloud Providers:** Offer backup and scalability solutions for data storage and processing.
- **Backend Web Application Developers:** Responsible for developing and maintaining the web application.
- **Local Legal Consultants:** Ensure compliance with local laws and regulations.

2. Key Activities

- **Development of AI Model:** Creating and refining the AI algorithm to accurately detect electricity theft.
- **Designing User-Friendly Interface:** Developing an intuitive interface for users to interact with the application.
- **Data Collection and Integration:** Gathering and integrating data from electricity meters and sensors.
- **Testing and Validation:** Ensuring the accuracy and reliability of the AI model through rigorous testing.

3. Value Propositions

- **Accurate Detection of Electricity Theft:** Reduces financial losses by identifying and addressing theft.
- **Improved Monitoring and Management:** Enhances the efficiency of electricity distribution and management.
- **Regulatory Compliance:** Ensures that the application meets local regulatory requirements.

- **Transparency:** Provides clear and detailed information about electricity usage to consumers.
- **Data Security:** Guarantees the security and privacy of data in compliance with local laws.

The developers are not permitted to access the consumption database to ensure customer privacy.

4. Customer Relationships

- **Customer Support:** Available on Sunday - Thursday 10:00AM - 05:00PM:Offering continuous support to address user issues and queries.
- **Training Sessions and Resources:** Providing education to users on effectively utilizing the application.
- **Feedback Collection and Improvement:** Regularly gathering user feedback to enhance the application.

5. Customer Segments

- **Primary Segment:** Sonelgaz, the national utility company for electricity and gas distribution.
- **Secondary Segments:**
  - Algerian utility regulators and government bodies.
  - Other utility companies in Algeria and potentially in neighboring countries with similar needs.

6. Key Resources

- **Backend Developers:** Skilled professionals to build and maintain the application.
- **Computational Resources:** Necessary hardware and software for processing data and running the AI model.
- **High-Quality Data:** Reliable data from electricity meters and sensors.
- **Hosting Infrastructure:** Platforms for hosting the application, both locally and on the cloud.

7. Channels

- **Direct Engagement:** Meetings and presentations with Sonelgaz to demonstrate the application.
- **Email and Phone:** Communication channels for ongoing support and inquiries.

8. Cost Structure

- **Hosting and Infrastructure:** Costs for both local and cloud-based hosting.
- **Monitoring:** Continuous monitoring of the system for performance and security.
- **Maintenance and Support:** Ongoing technical support and updates.
- **Legal Consultations:** Ensuring compliance with local regulations through regular audits.
- **Salaries:** Compensation for the technical team.
- **Promotional Activities:** Marketing and sales efforts to promote the application.

9. Revenue Streams

- **Sale of the Application:** One-time payment for the application.
- **Subscription Model:** Recurring revenue from subscriptions for ongoing use, support, and updates.
- **Consulting Fees:** Additional revenue from consulting services related to data analysis, security, and compliance.
- **Licensing:** Licensing the technology to other utility companies or government bodies.

Key Partners	Key Activities	Value Propositions	Customer Relationships	Customer Segments
Government Company - for Electricity and Gas Distribution Sonelgaz. Local Hosting Providers. Cloud Providers for backup and scalability. Backend web application developers. Local Legal Consultants.	**Development of AI Model. Designing user-friendly interface. Data Collection and Integration. Testing and Validation. Deployment and Maintenance.**	**Accurate detection of electricity theft leading to reduced losses. Improved monitoring and management of electricity distribution. Ensuring compliance with**	**Available on Sunday - Thursday 10:00AM - 05:00PM. Providing training sessions and resources for users to effectively utilize the application. Regularly collecting and acting on**	**Primary Segment: Sonelgaz, the national utility company responsible for electricity and gas distribution. Secondary Segments: Algerian utility regulators and government bodies.**

<table>
<tr>
<td rowspan="2"></td>
<td></td>
<td rowspan="2">local regulations and improving data security.
Providing transparent and detailed electricity usage information.
Assurance of data security and privacy in accordance with local laws.
The developers are not permitted to access the consumption database to ensure customer privacy.</td>
<td>feedback to improve the application.</td>
<td rowspan="2">Other utility companies in Algeria and potentially in neighboring countries with similar needs.</td>
</tr>
<tr>
<td><b>Key Resources</b>
back-end developers.
Computational Resources.
High-quality data from electricity meters and sensors installed in Algeria.
Hosting Infrastructure.</td>
<td><b>Channels</b>
Direct engagement with Sonelgaz through meetings and presentations.
email.
Phone number.</td>
</tr>
<tr>
<td colspan="3"><b>Cost Structure</b>
Hosting and infrastructure (local and cloud).
Monitoring.
Ongoing maintenance and support.
Legal consultations and compliance audits with local regulations.
Salaries for a technical team.
Promotional activities and direct sales efforts.</td>
<td colspan="2"><b>Revenue Streams</b>
the sale of the application.
Subscription Model: Offering the application as a subscription service for ongoing use, support, and updates.
Consulting Fees: Charging for additional consulting services related to data analysis, security, and compliance.
Licensing: Potentially licensing the technology to other utility companies or government bodies within Algeria and the region.</td>
</tr>
</table>

CONCLUSION

Conclusion

The battle against electricity theft is critical for both utility companies and consumers, given its extensive economic and operational repercussions. Traditional detection methods have proven inadequate in addressing the sophisticated techniques employed by modern thieves. However, the advent of artificial intelligence, specifically deep learning and autoencoders, has revolutionized this fight. By harnessing the capabilities of autoencoders to analyze large datasets and identify anomalies, applications like DUANKOU represent a significant advancement in electricity theft detection. These AI-driven solutions not only enhance the accuracy and speed of identifying fraudulent activities but also help mitigate financial losses and ensure a fairer, more efficient energy market. As technology continues to evolve, the implementation of AI in electricity theft detection will be indispensable in securing a reliable and economically sound power supply for the future.

Bibliography

INTELLEGENCE ARTIFICIAL

1. Samuel , Arthur L (1959). “Machine learning”. In : *The Technology Review* 62.1,p. 42-45.

2. Hu , Junyan et al. (2020). “Voronoi-based multi-robot autonomous exploration in unknown environments via deep reinforcement learning”. In : *IEEE Transactions on Vehicular Technology* 69.12, p. 14413-14423.

3. Mahesh , Batta (2020). “Machine learning algorithms-a review”. In : *International Journal of Science and Research (IJSR).[Internet]* 9, p. 381-386.

4. Arulkumaran , Kai et al. (2017). “A brief survey of deep reinforcement learning”. In : *arXiv preprint arXiv :1708.05866*

5. https://www.statworx.com/en/content-hub/blog/5-types-of-machine-learning-algorithms-with-use-cases/

5. Sutton , Richard S, Andrew G Barto et al. (1998). *Introduction to reinforcement learning*. T. 135. MIT press Cambridge.

6. Tyagi , Amit Kumar et G Rekha (2020). “Challenges of applying deep learning in real-world applications”. In : *Challenges and applications for implementing machine learning in computer vision*. IGI Global, p. 92-118.

8. Moghaddamnia , A et al. (2009). “Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques”. In : *Advances in Water Resources* 32.1, p. 88-97.

9. Franco Ortellado , Blas Manuel et al. (2019). “Applications of artificial neural networks in three agro-environmental systems : microalgae production, nutritional characterization of soils and meteorological variables management”. In.

10. Brownlee , Jason (2017). “A gentle introduction to long short-term memory networks by the experts”. In : *Machine Learning Mastery* 19.

11. Macary , Manon (2022). “Analyse de données massives en temps réel pour l’extraction d’informations sémantiques et émotionnelles de la parole”. Thèse de doct. Le Mans Université.

12. Khan , Asifullah et al. (2020). “A survey of the recent architectures of deep convolutional neural networks”. In : *Artificial intelligence review* 53, p. 5455-5516.

13. Choi, Hyeong In. "Lecture 16: Autoencoders (Draft: version 0.7. 2)." (2019).

14. Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2020). Dynamical variational autoencoders: A comprehensive review. *arXiv preprint arXiv:2008.12595*.

15. https://saturncloud.io/glossary/sparse-autoencoders/

ELECTRICITY THEFT

1. Shah, A.L.; Mesbah, W.; Al-Awami, A.T. An Algorithm for Accurate Detection and Correction of Technical and Nontechnical Losses Using Smart Metering. IEEE Trans. Instrum. Meas. **2020**, 69, 8809–8820. [CrossRef]

2. Nagi, J.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K.; Mohamad, M. Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. IEEE Trans. Power Del. **2010**, 25, 1162–1171. [CrossRef]

3. Yan, Z.; Wen, H. Performance Analysis of Electricity Theft Detection for The Smart Grid: An Overview. IEEE Trans. Instrum. Meas. **2021**, 71, 1–28. [CrossRef]

4. Korba, A.A.; Karabadji, N.E.I. Smart Grid Energy Fraud Detection Using SVM. In Proceedings of the 2019 International Conference on Networking Advanced Systems, Annaba, Algeria, 26–27 June 2019.

5. Northeast Group, LLC. (2014). World Loses $89.3 Billion to Electricity Theft Annually, $58.7 Billion in Emerging Markets. [Online]. Available: http://www.prnewswire.com/news-releases/world-loses-893-billion-to-electricity-theft-annually-587-billion-in-emerging-markets-300006515.html

6. L. Northeast Group. (2017). 96 Billion Dollars Is Lost Every Year to Electricity Theft. [Online]. Available: https://www.prnewswire.com/newsreleases/96-billion-is-lost-every-year-to-electricitytheft-300453411.html

7. P. Antmann. (2009). Reducing Technical and Non-Technical Losses in the Power Sector. [Online]. Available: https://openknowledge.worldbank.org/handle/10986/20786

8. H. Arkell. (2014). How Middle-Class Families Are Turning to Crime by Getting Specialist Gangs to‘Hotwire’ Their Gas and Electricity Supplies to Beat Soaring Energy Bills. [Online]. Available: http://www.dailymail.co.U.K./news/article-2542487/Energy-theft.html

9. V. Gaur and E. Gupta, “The determinants of electricity theft: An empirical analysis of Indian states,” *Energy Policy*, vol. 93, pp. 127–136, Jun. 2016.

Takkidin 2022: Takiddin, A., Ismail, M., Zafar, U., & Serpedin, E. (2022). Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Systems Journal*, *16*(3), 4106-4117.

Guoying 2021: Lin, G., Feng, X., Guo, W., Cui, X., Liu, S., Jin, W., ... & Ding, Y. (2021). Electricity theft detection based on stacked autoencoder and the undersampling and resampling based random forest algorithm. *Ieee Access*, *9*, 124044-124058.

Younggyhu 2023: Sun, Y., Lee, J., Kim, S., Seon, J., Lee, S., Kyeong, C., & Kim, J. (2023). Energy Theft Detection Model Based on VAE-GAN for Imbalanced Dataset. *Energies*, *16*(3), 1109.

Pamir 2023: Pamir, Javaid, N., Javed, M. U., Houran, M. A., Almasoud, A. M., & Imran, M. (2023). Electricity theft detection for energy optimization using deep learning models. *Energy Science & Engineering*, *11*(10), 3575-3596.

Ruizhe 2023: Yao, R., Wang, N., Ke, W., Liu, Z., Yan, Z., & Sheng, X. (2023). Electricity Theft Detection in Incremental Scenario: A Novel Semi-supervised Approach based on Hybrid Replay Strategy. *IEEE Transactions on Instrumentation and Measurement*.

Webography

Web 01:

https://www.sarawakenergy.com/media-info/media-releases/2021/homeowner-caught-stealing-electricity-via-underground-direct-tapping-cables

Web 02:

https://www.google.com/imgres?imgurl=https%3A%2F%2Fars.els

Web 03: https://c3.ai/products/c3-ai-energy-management/

Web 04: https://www.itron.com/

Web 06:

https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FNeural_network_%2528machine_learning

Web 07:

https://www.google.com/url?sa=i&url=https%3A%2F%2Ftowardsdatascience.com%2Fa-comprehensive-guide-to-convolutional-neural-networks- E

Web 08 : Colaboratory. Retrieved June 8 ,2023.
website : https ://research.google.com/colaboratory/faq.html ?hl=fr

Web 9: https ://www.datarockstars.ai/glossary/python/
Web 10 : la bibliothèque Python la plus utilisée en Data Science. Retrieved June 8 ,2023. website : https ://datascientest.com/numpy/
Web 11 : Pandas. Retrieved June 8 ,2023. website : https ://datascientest.com/pandaspython-data-science
Web 12 : A SCIENTIFIC COMPUTING FRAMEWORK FOR LUAJIT. Retrieved June 8 ,2023. website : http ://torch.ch/
Web 13: Matplotlib. Retrieved June 8 ,2023. website : https://datascientest.com/matplotlibtout-savoir
Web 15 : Seaborn. Retrieved June 8 ,2023. website : https ://datascientest.com/seaborntout-Savoir