République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed El Bachir El Ibrahimi Bordj Bou Arreridj

Faculté des Mathématiques et informatique

Département d'Informatique

Mémoire de fin d'étude Master

Filière : Informatique

Option *: Réseaux et Multimédia*

Thème *:*

Prédiction de l'instabilité des réseaux électriques intelligents (smart grids) à l'aide des algorithmes d'apprentissage automatique

Présenté par :

- MANSOURI KHELIFA

Soutenu publiquement le23/06/2025 devant le jury composé de :

Président : Dr. NAILI MAKHLOUF

Examinateur : Dr. BADAOUI ATIKA

Encadreur : Dr. BOUMAZA FARID

2024/2025

Remerciements

Avant tout, je remercie ALLAH le tout puissant, le miséricordieux, qui m'a donné la force et le courage pour achever cet humble travail.

Je tiens à remercier vivement Monsieur **BOUMAZA FARID** *de m'avoir proposé le thème sur les smart grids qui a fait l'objet de mon travail. Qu'il trouve ici l'expression de ma gratitude pour ses conseils, son orientation et toute sa disponibilité*

Mes sincères remerciements s'adressent également, aux membres du jury
Dr. BADAOUI ATIKA et Dr. NAILI MAKHLOUF

Pour l'honneur qu'ils me feront en participant au jugement de ce travail.

Je tiens à remercier toutes les enseignantes et enseignants du faculté Maths et Informatique pour tous les efforts fournis durant tout mon cursus universitaire.

Enfin, Je tiens à remercier vivement toute personne qui m'a aidée, de près ou de loin, à finir mes études

Résumé

Les Smart Grids désignent des réseaux électriques intelligents utilisant les technologies de l'information et de la communication pour optimiser la production, la distribution et la consommation de l'électricité. La stabilité des réseaux électriques intelligents constitue un défi majeur en raison de leurs conditions d'exploitation dynamiques et incertaines. Cette étude propose une approche basée sur l'apprentissage automatique (Machine Learning, ML) pour prédire la stabilité de ces réseaux, en exploitant un ensemble de données contenant des paramètres dynamiques clés, tels que le taux de variation (τ), les variations de pression (p) et les gradients de distribution d'énergie (g). Ces données, disponibles sur Kaggle sous le nom « Electrical Grid Stability Simulated Dataset », sont traitées et analysées pour classer l'état du réseau en stable ou instable.

Une évaluation approfondie de plusieurs algorithmes d'apprentissage automatique, notamment la régression logistique (LR), les machines à vecteurs de support (SVM), Naïve Bayes (NB), les forêts aléatoires (RF) et Extreme Gradient Boosting (XGBoost), est réalisée pour évaluer leur efficacité à gérer des motifs non linéaires et des dépendances complexes. Les résultats expérimentaux mettent en évidence les forces et les limites de chaque méthode, en termes de précision prédictive et de capacité de généralisation. Les conclusions soulignent que les modèles avancés, en particulier les méthodes d'ensemble, améliorent de manière significative les prévisions de stabilité, optimisant ainsi l'efficacité opérationnelle et la résilience des réseaux. Cette étude contribue au développement de solutions basées sur les données pour la gestion des réseaux intelligents, et renforce le rôle de l'intelligence artificielle (IA) dans les systèmes énergétiques durables.

Mots-clés : Smart Grids, intelligence artificielle, apprentissage automatique, algorithmes de Machine Learning.

Abstract

Smart Grids refer to intelligent electrical networks that utilize information and communication technologies to optimize the production, distribution, and consumption of electricity. The stability of these smart grids presents a significant challenge due to their dynamic and uncertain operating conditions. This study proposes a machine learning (ML)-based approach to predict the stability of these networks by exploiting a dataset containing key dynamic parameters, such as the rate of variation (τ), pressure variations (p), and energy distribution gradients (g). This dataset, available on Kaggle under the name "Electrical Grid Stability Simulated Dataset," is processed and analyzed to classify the network state as stable or unstable.

An in-depth evaluation of several machine learning algorithms, including logistic regression (LR), support vector machines (SVM), Naïve Bayes (NB), random forests (RF), and Extreme Gradient Boosting (XGBoost), is conducted to assess their effectiveness in handling nonlinear patterns and complex dependencies. Experimental results highlight the strengths and limitations of each method in terms of predictive accuracy and generalization ability. The conclusions emphasize that advanced models, particularly ensemble methods, significantly improve stability predictions, thereby optimizing operational efficiency and network resilience. This study contributes to the development of data-driven solutions for managing smart grids and reinforces the role of artificial intelligence (AI) in sustainable energy systems.

Keywords: Smart Grids, artificial intelligence, machine learning, machine learning algorithms.

ملخص

تُشير الشبكات الذكية (Smart Grids) ,الى شبكات كهربائية ذكية تستفيد من تكنولوجيات المعلومات والاتصالات لتحسين ,انتاج وتوزيع واستهلاك الكهرباء. يعدّ استقرارية هذه الشبكات تحد ي رئيسي نظرا لظروف تشغيلها الديناميكية وغير المؤكدة. يقترح هذا البحث منهاجا يعتمد على التعلم الآلي (Machine Learning, ML) للتنبؤ باستقرارية هذه الشبكات، وذلك باستغلال مجموعة بيانات

تحتوي على متغيرات ديناميكية رئيسية، مثل معدل التغير(τ) ، تغيرات الضغط(p) ، وتدرجات توزيع الطاقة .(g) تمت معالجة هذه البيانات، المتوفرة على Kaggle تحت اسم"Electrical Grid Stability Simulated Dataset" ، وتحليلها لتصنيف حالة الشبكة ,الى مستقرة أو غير مستقرة.

أُجري تقييم شامل للعديد من خوارزميات التعلم الآلي، بما في ذلك الانحدار اللوجستي (LR) ، ألات المتجهات الداعمة(SVM) ، نايف بايز(NB) ، الغابات العشوائية(RF) ، وتعزيز التدرج المتطرف(XGBoost) ، لتقييم فعاليتها في التعامل مع انماط غير الخطية والتبعيات المعقدة. أبرزت النتائج التجريبية نقاط القوة والقيود لكل طريقة، من حيث دقة التنبؤ وقدرة التعميم. تُشير الاستنتاجات ,الى أن النماذج المتقدمة، لا سيما أساليب التجميع(ensemble methods) ، تُحسن بشكل كبير من توقعات الاستقرارية، وبالتالي تُحسن الكفاءة التشغيلية ومرونة الشبكات. تُسهم هذه الدراسة في تطوير حلول مبنية على البيانات ,لادارة الشبكات الذكية، وتُعزز دور الذكاء الاصطناعي (AI) في أنظمة الطاقة المستدامة.

الكلمات المفتاحية :الشبكات الذكية، الذكاء الاصطناعي، التعلم الآلي، خوارزميات التعلم الآلي.

Table des matières

Chapitre 2 : L'apprentissage automatique

Chapitre 3 : Expérimentation et discussion des résultats

Table des figures

Liste des tableaux

Liste des abréviations

SG : SMART GRIDS

DSGC :DECENTRALISED SMART GRIDS CONTROL

IA : INTELLIGENCE ARTIFICIELLE

ML : MACHINE LEARNING

RL : REGRESSION LOGISTIQUE

NB : NAIVE BAYES

SVM : SUPPORT VECTOR MACHINE

RF : RANDOM FOREST

XGBOOST : EXTREME GRADIANT BOOST

AUC- ROC : Area Under the Receiver Operating Characteristic Curve

CV : CROSS VALIDATION (VALIDATION CROISEE)

VP : VRAIS POSITIVES

VN : VRAIS NEGATIVES

FP : FAUX POSITIVES

FN : FAUX NEGATIVES

Introduction générale

Contexte

La durabilité environnementale, la fiabilité et l'efficacité Le réseau électrique intelligent, plus souvent désigné par son expression anglophone Smart Grid ont révolutionné la gestion et la distribution de l'énergie. Prévoir avec précision les besoins énergétiques facilite la gestion d'un réseau intelligent en optimisant l'allocation des ressources, en réagissant rapidement aux variations de la demande et en assurant la stabilité du système. Les techniques conventionnelles de prévision des charges énergétiques s'avèrent souvent inadéquates, car elles ne peuvent pas traiter efficacement les nombreuses connexions spatiales et temporelles inhérentes aux données de consommation d'énergie. Face à ce défaut inhérent, il est impératif d'explorer des modèles prédictifs plus sophistiqués, capables de relever efficacement ces défis et de générer des prévisions précises. L'apprentissage automatique plus souvent désigné par son expression anglophone Machine Learning prend une importance croissante dans tous les aspects de notre vie. Il a eu un impact significatif dans divers domaines, notamment la détection du cancer, la médecine de précision, les véhicules autonomes, la prévision prédictive et la reconnaissance vocale. Les extracteurs d'éléments fabriqués manuellement, utilisés dans les systèmes traditionnels d'apprentissage, de caractérisation et de reconnaissance de formes, ne sont pas adaptables aux ensembles de données à grande échelle. Selon le niveau de complexité, l'apprentissage automatique peut surmonter les limites des anciens réseaux superficiels qui entravaient le traitement et la représentation efficaces des structures hiérarchiques dans les données d'apprentissage multidimensionnelles. La prévision de la stabilité est essentielle dans les réseaux intelligents. Elle favorise la durabilité et permet aux administrations de mettre en œuvre des stratégies efficaces de planification et d'exploitation des systèmes électriques. L'électricité étant une ressource essentielle, sa consommation doit être synchronisée avec sa production afin d'éviter tout gaspillage inutile. Sa volatilité et son imprévisibilité caractérisent la prévision de la stabilité, et toute sous-estimation ou surestimation peut entraîner des problèmes importants. L'industrialisation et l'automatisation à l'échelle mondiale ont considérablement amélioré la production d'énergie électrique, améliorant ainsi sa fiabilité. Aujourd'hui, l'énergie verte est largement utilisée comme source d'énergie de substitution pour répondre aux besoins croissants en électricité. Néanmoins, de nombreux obstacles se dressent pour préserver la fiabilité et la sécurité du réseau électrique. Les systèmes électriques constituent l'un des principaux défis de la consommation énergétique mondiale. Certains ont proposé l'idée de réseaux intelligents comme solution à ces problèmes. Les réseaux intelligents sont un domaine multidisciplinaire qui associe les systèmes électriques aux avancées technologiques, aux technologies intelligentes, aux communications sans fil et à d'autres domaines connexes. Bien que les smart grids présentent de nombreux avantages, des améliorations restent encore nécessaires. Actuellement, la prévision de la charge est réalisée à l'aide de divers modèles et techniques. Néanmoins, face à l'augmentation des besoins énergétiques, il est impératif de créer des prototypes plus performants. La majorité des études ont tenté de relever ces défis [1].

Problématique

L'essor des énergies renouvelables et la transition vers une production décentralisée ont introduit de nouveaux défis dans la gestion de la stabilité des smart grids. Les méthodes traditionnelles, telles que le modèle de contrôle décentralisé des réseaux intelligents (DSGC), reposent sur des simplifications mathématiques qui ne parviennent pas à saisir la complexité dynamique des réseaux intelligents réels. Ces modèles ont du mal à prédire les instabilités du réseau causées par les fluctuations de la production et de la consommation d'énergie, en particulier dans les systèmes comptant plusieurs producteurs-consommateurs d'énergie. Pour combler cette lacune, nous avons besoin d'un modèle prédictif avancé capable de combiner plusieurs caractéristiques critiques, d'augmenter la capacité mémoire et d'améliorer la précision des prévisions. L'approche plus flexible et axée sur les données est nécessaire pour prédire la stabilité du réseau avec une plus grande précision et en temps réel.

Objectifs :

- Exploiter l'apprentissage automatique pour prédire la stabilité des réseaux intelligents avec une précision supérieure à celle des méthodes traditionnelles.
- Explorer diverses algorithmes de machine learning afin d'optimiser les performances dans la tâche de classification binaire (stable vs instable).
- Fournir des analyses pratiques sur la manière dont les modèles d'apprentissage automatique peuvent améliorer la gestion et la stabilité des réseaux dans des scénarios réels.

Ce projet de fin d'études permis d'apporter plusieurs contributions significatives à la compréhension et à la gestion proactive de l'instabilité dans les smart grids, en exploitant le potentiel des algorithmes de machine learning. Nos principaux apports se résument comme suit :

- Développement d'un Cadre de Prédiction Robuste.
- Évaluation Comparative d'Algorithmes de Machine Learning.
- Identification des Caractéristiques Clés de l'Instabilité.
- Démonstration de la Faisabilité et de l'Efficacité de l'Approche ML.
- Recommandations pour l'Opération des Smart Grids.

En somme, ce projet constitue une preuve de concept solide de l'application du machine learning à un défi industriel majeur, fournissant une base pour de futures recherches et un développement potentiel de systèmes d'aide à la décision en temps réel pour les opérateurs de réseaux intelligents.

Plan du mémoire

Ce mémoire est organisé en trois principaux chapitres comme suit :

1. Le 1er chapitre présente un aperçu général sur les smart grids, leurs définitions, fonctionnement, caractéristiques, avantages et leur type d'instabilité.
2. Le 2éme chapitre donne un aperçu sur l'apprentissage automatique, les algorithmes d'apprentissage supervisé qui peuvent nous aider à prédire l'instabilité des smart grids. En donnant une explication détaillée de chaque algorithme utilisé dans ce projet.
3. Le dernier chapitre présente d'abord une étude technique dans laquelle je définis l'environnement logiciel utilisé pour réaliser notre projet, puis une définition et une analyse détaillé de le dataset utilisés. Ensuite, les résultats sont présentés, comparés et interprétés.

   A la fin, ce travail est clôturé par une conclusion générale.

Chapitre 1 : Présentation des smart grids

1.1 Introduction

Les smart grids (SG) [2], émergent comme une solution prometteuse pour révolutionner notre manière de produire, de distribuer et de consommer l'électricité.

Ces systèmes avancés, qui intègrent les technologies de l'information et de la communication au cœur des infrastructures énergétiques, promettent d'optimiser l'efficacité du réseau électrique tout en favorisant l'intégration des énergies renouvelables.

Leur capacité à ajuster automatiquement la production à la demande, grâce à un réseau de capteurs et à l'analyse informatique des données en temps réel, ouvre la voie à une gestion plus durable et plus économique de l'énergie.

Historiquement, le développement des smart grids a été motivé par la nécessité de répondre aux nouveaux défis énergétiques, notamment l'augmentation de la demande en électricité, l'intégration des sources d'énergie renouvelable intermittentes et la volonté de réduire les émissions de gaz à effet de serre.

Les premiers pas vers des réseaux plus intelligents ont été marqués par l'amélioration du suivi et de la synchronisation des réseaux électriques, notamment aux États-Unis où le premier système de mesure utilisant des capteurs spécifiques est devenu opérationnel en 2000 [3].

Depuis lors, l'expression « smart grids » s'est généralisée, symbolisant une évolution majeure dans l'optimisation des réseaux électriques.

En mettant en lumière ces réseaux du futur, nous espérons contribuer à une meilleure compréhension de leur rôle dans la transition énergétique en explorant leurs avantages, leur fonctionnement, les défis à relever, ainsi que leur impact sur les consommateurs et l'environnement.

1.2 Architectures du réseaux électriques traditionnels

Le réseau électrique a été développé de manière progressive à partir de la fin du XIXe siècle [4]. Ce réseau a été développé avec une circulation à sens unique : les centrales électriques - à charbon, hydroélectriques, à gaz ou plus tard nucléaires - produisent l'électricité et cette dernière est acheminée aux consommateurs via le réseau de distribution.

Les systèmes électriques traditionnels se composent d'un ensemble d'infrastructures permettant d'acheminer l'énergie électrique produite vers les consommateurs. L'électricité transite donc depuis la centrale de production, par les réseaux de transport, de répartition, et de distribution pour arriver chez le consommateur [5].

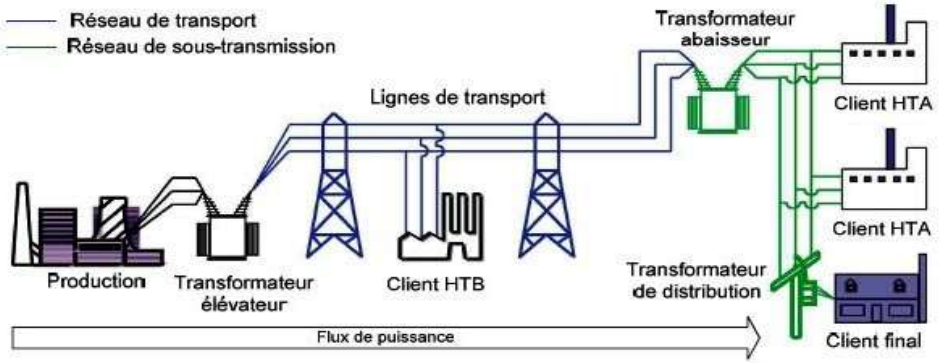


Figure 1.1 : Architecture du réseau électriques traditionnel [6]

1.3 Définition des Smart grids

les SG, est une approche nouvelle de la production et de la consommation électriques qui introduise dans le réseau électrique des capteurs, actionneurs et analyseurs afin d'optimiser la production et la consommation d'électricité. Ce concept permet entre autres d'intégrer les énergies renouvelables au dispositif existant et également de réduire l'impact écologique de l'électricité[7].

1.4 Description générale des Smart grids

Les réseaux intelligents sont une technologie qui permettrait d'affronter les changements actuels dans le paysage énergétique comme l'intégration des énergies renouvelables au réseau électrique, la gestion de l'augmentation de la consommation ou encore le développement de l'utilisation des voitures électriques. En effet ce réseau de distribution intelligent basé sur des Technologies informatiques augmenterait l'efficacité de la gestion entre l'offre et la demande d'électricité. Ce nouvel équilibre engendrerait une optimisation du réseau de distribution mais aussi de la production.

Les smart grids minimiseraient les pertes en ligne (comme produire inutilement de l'électricité lorsque l'offre est supérieure à la demande) ainsi que les problèmes causés par les énergies intermittentes (comme le solaire et l'éolien) sachant que l'énergie électrique est difficilement stockable. Ils seraient également l'alternative au remplacement et la construction de nouvelles lignes électriques.

La notion de smart grids combine deux idées : d'une part, rendre plus intelligents les réseaux électriques et, d'autre part, créer des mini-réseaux autonomes dans lesquels on pourra associer aisément différentes sources d'énergie. Le réseau intelligent possédera donc des caractéristiques différentes de celles du réseau électrique actuel qui nécessiteront des installations plus ou moins importantes. Ce réseau ne sera pas unidirectionnel mais bidirectionnel rendant l'ensemble des acteurs en interaction. Il se concentrera principalement sur le réseau de distribution car celui-ci n'est que faiblement doté de technologies de communication et gérera l'équilibre du système électrique par la demande (consommation) et non par l'offre. La figure I.7 ci-dessous illustre une configuration possible de ce type de réseau [8].

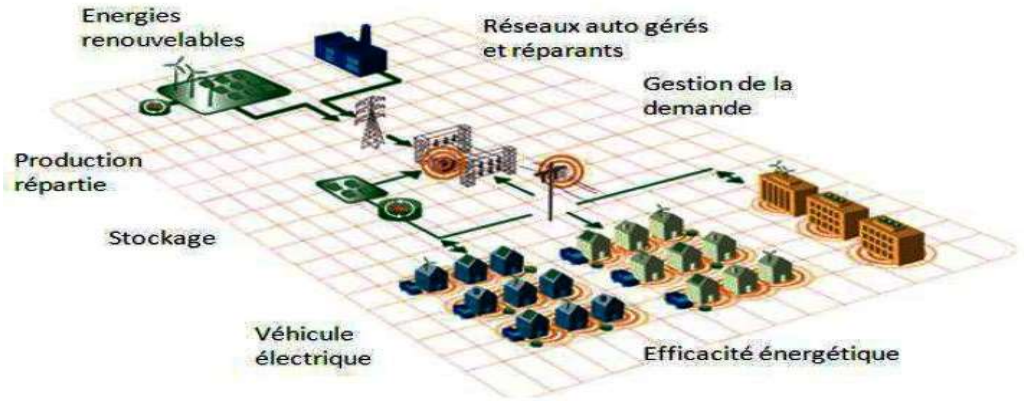


Figure 1.2 : Schéma d'un réseau intelligent [9]

1.5 Objectifs des smart grids

La modernisation des réseaux électriques actuels doit tendre vers le développement des smart grids afin d'améliorer leur fiabilité, faciliter l'intégration des énergies renouvelables et des voitures électriques, et mieux gérer la consommation d'électricité. Le développement des smart grids davantage de déploiement des ressources énergétiques telles que les batteries de stockage et la production décentralisée. La technologie associée aux smart grids vise entre autres les objectifs suivants :

- la libéralisation des marchés de l'énergie et la multiplicité des acteurs. Ceci a un impact sur la partition des responsabilités, la gestion des acteurs qui peuvent avoir des intérêts divergents ou l'obligation d'un traitement non discriminatoire de toute décision impactant ces acteurs. Par ailleurs, la gestion de l'information dans ce cadre devient primordiale pour le pilotage du système,
- le développement à grande échelle des énergies renouvelables qui se développent à grande échelle, en particulier en off-shore pour l'éolien pour certains pays par exemple. Cependant, le développement de ces énergies s'effectue également au niveau de plaques continentales impactant l'ensemble du réseau interconnecté des continents concernés,
- la remise en cause, au niveau local, le découplage traditionnel des réseaux de transport et de distribution par le développement de la production décentralisée ce qui conduit à une gestion plus complexe du réseau et un rôle accru des gestionnaires de réseaux. La figure I.8 illustre le rôle des gestionnaires de réseaux intelligents.

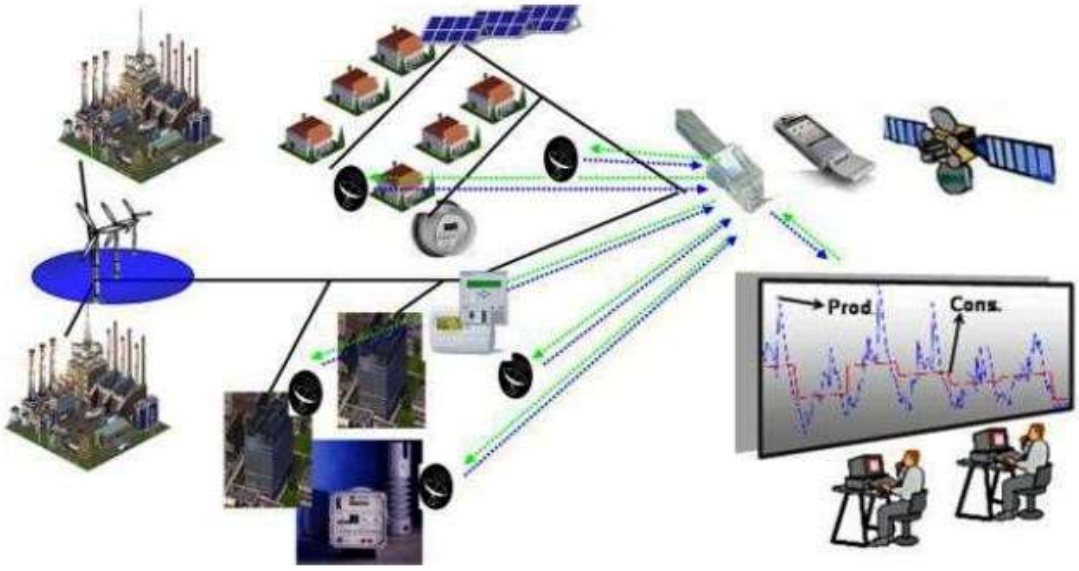


Figure 1.3 : Communication et intelligence embarquée en réseau [10].

1.6 Infrastructure du Smart Grids

La figure 1.4 montre un type d'architecture de réseau intelligent qui repose sur la combinaison de plusieurs couches infrastructures et logicielles permettant de communiquer, mesurer, contrôler et piloter. Sur cette base, il est possible d'imaginer la création de nombreux services en aval et en amont du compteur. Les équipements à mettre en œuvre diffèrent selon l'architecture énergétique. Un certain nombre de fonctions ou composants existent depuis de nombreuses années. Les grandes catégories de composants et de systèmes que l'on peut retrouver dans une architecture réseau intelligent sont [11] :

- L'infrastructure de communication qui comporte le réseau local LAN (Local Area Network), le réseau étendu WAN (Wide Area Network), le système de mesure avancé FAN (Field Area Network) et AMI (Advanced Metering Infrastructure), l'équipement CPE (Costumer Premise Equipement), le réseau domestique HAN (Home Area Network).
- l'infrastructure énergétique qui comporte les systèmes FACTS (Flexible AC Transmissions Systems), les SMES (Super conducting Magnetic Energy Storage), le système de transport en haute tension et à courant continu – HVDC (High Voltage Direct Current).
- les outils de mesure comportant le système PMU (Phasor Measurement Unit), les Capteurs, les compteurs intelligents.
- les systèmes de contrôle et de détection entrant dans le cadre de la télégestion tels que SCADA (Supervisory Control And Data Acquisition), WASA (Wide-Area Situational Awareness), WAMS (Wide Area Measurement System), WAAPCA (Wide Area Adaptive Protection, Control and Automation), MDMS (Meter Data Management System).
- les systèmes de pilotage comportant EMS (Energy Management System), GIS (Geographic Information System), DMS (Distribution Management System), OMS (Outage Management System), WMS (Workload Management System), DA (Distribution Automation) CEMS (Consumer Energy Management Systems).

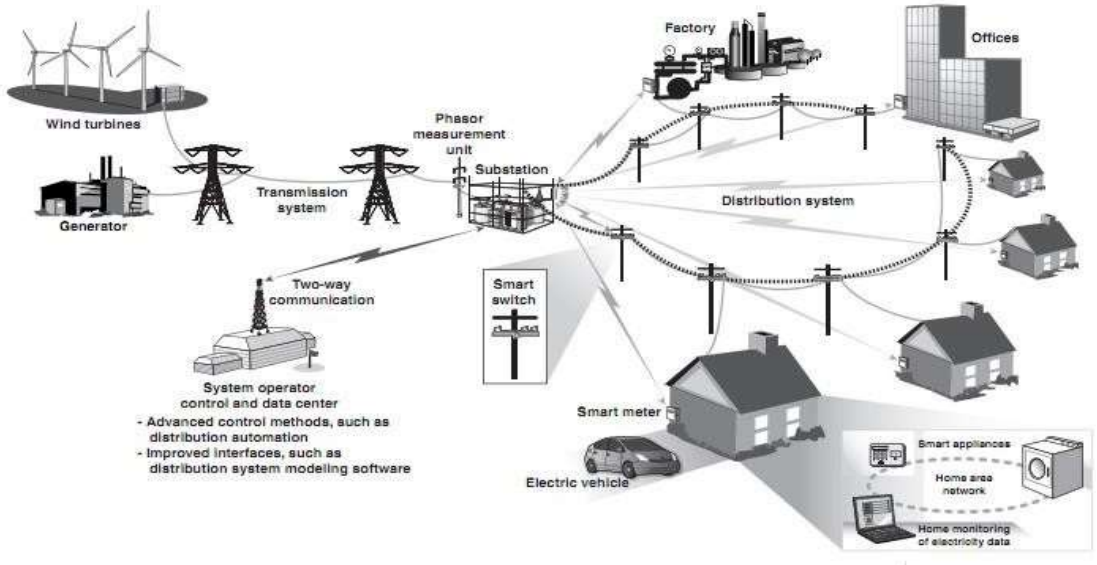


Figure 1.4 : Architecture du smart-Grid [12]

D’un point de vue technologique, le smart grid se compose de :

- D’un compteur analogique communiquant (smart meter) chez l’utilisateur final qui remplacera le compteur traditionnel,
- D’un logiciel de suivi et de gestion de la consommation client,
- D’une infrastructure de communication reliant le consommateur au producteur, plus ou moins dense suivant le mode de communication utilisé (par satellite, courant porteur en ligne, Wifi longue portée, radio fréquence, …),
- De serveurs informatiques et de logiciels de back-office permettant au producteur de stocker et d’analyser l’immense quantité d’informations générées par le smart grid [13].

L’exploitation du réseau électrique intelligent consiste à communiquer tout au long du réseau, jusqu’à l’intérieur des bâtiments. Le but est de réduire la consommation globale, mais aussi de l’adapter à la production. On commence donc à parler de « maisons intelligentes », qui pourraient gérer leur consommation d’énergie elles-mêmes.

L’idée d’une maison intelligente est notamment de réduire les pics de consommation : La maison pourra directement gérer l’utilisation du chauffage, des appareils électroménagers, etc. afin de réduire la consommation à toute heure, et spécialement aux heures de pointe. Du point

de vue des énergies renouvelables, cela peut aussi permettre d’utiliser les appareils quand l’énergie est disponible : la maison devra donc être connectée au réseau pour suivre la production d’énergie en temps réel. Les appareils électroménagers pourraient par exemple se mettre en marche le midi, lorsque la production des panneaux solaires est la plus forte, et ainsi créer des pics de consommation correspondant aux pics de production, évitant le stockage.

La maison intelligente va même plus loin en s’intégrant directement au réseau : les voitures électriques une fois branchées sur leur prise ne resteraient pas inertes. Après avoir emmagasiné de l’énergie, elles pourraient la restituer lors des pics de consommation en alimentant le réseau ou la maison directement.

En plus de la gestion de la consommation, les maisons intelligentes peuvent aussi gérer leur propre production d’énergie. Cette production peut venir d’énergie renouvelable (solaire, éolien,…). La décentralisation de la production d’énergie sera gérable à grande échelle grâce aux compteurs communicants [14].

1.7 Comment fonctionne les smart grids ?

Les smart grids, sont des systèmes énergétiques avancés qui intègrent des technologies de l’information et de la communication pour optimiser la gestion de l’électricité.

Dans ce type de situations, où la demande excède momentanément l’offre, les Smart Grids peuvent activer des programmes spécifiques de réponse à la demande. Ces programmes visent à réduire temporairement la consommation chez certains clients volontaires, libérant ainsi de la capacité sur le réseau et empêchant toute rupture d’approvisionnement

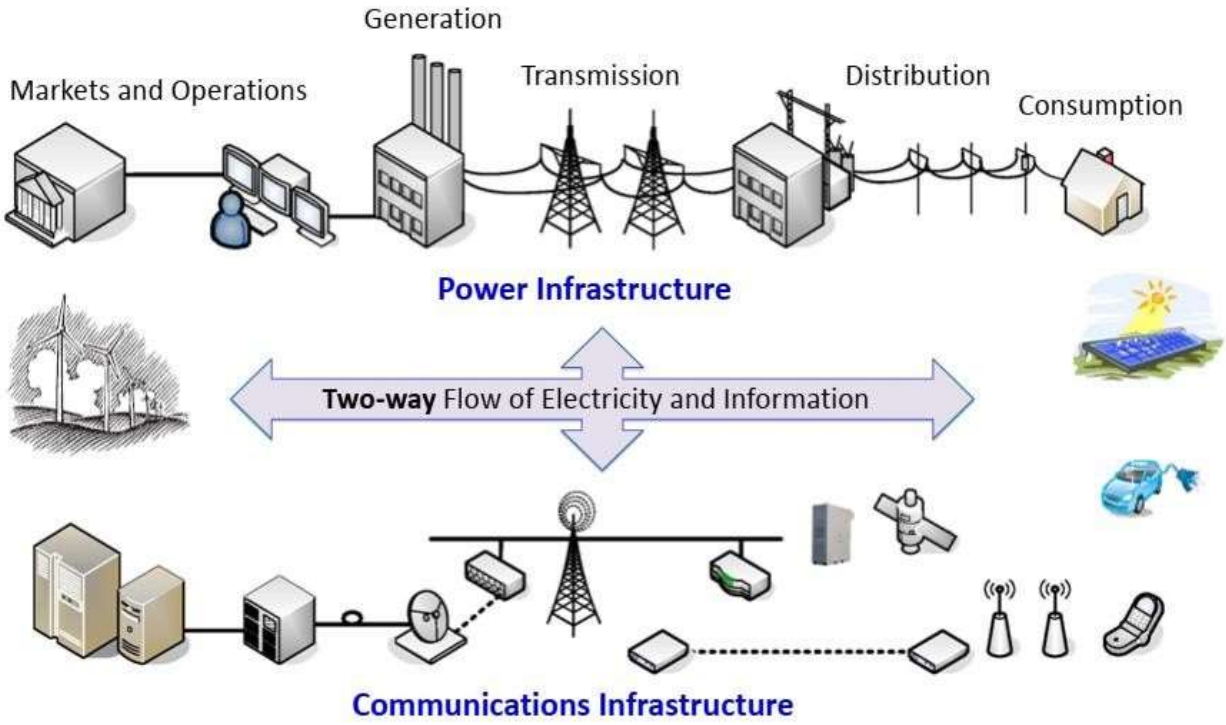


Figure 1.5 Infrastructure des Communications [15]

- **Utilisation de technologies numériques, de capteurs et de logiciels :**

Les smart grids utilisent des technologies numériques pour collecter des données en temps réel sur l'état du réseau électrique. Des capteurs et des compteurs Intelligents sont déployés à travers le réseau pour surveiller la consommation d'énergie, la production et les conditions du réseau.
Ces données sont ensuite analysées par des logiciels avancés pour mieux adapter l'offre et la demande d'électricité, améliorant ainsi l'efficacité et la fiabilité du réseau.

- **Échange bidirectionnel d'énergie et d'informations**

Contrairement aux réseaux traditionnels, les smart grids permettent un échange bidirectionnel d'énergie et d'informations.
Cela signifie que l'électricité peut être distribuée du fournisseur au consommateur et vice versa, permettant ainsi l'intégration de la production d'énergie décentralisée, comme celle issue des panneaux solaires domestiques.
Les informations peuvent circuler dans les deux sens, ce qui permet aux consommateurs et aux fournisseurs d'énergie de communiquer et de réagir aux changements de demande ou d'offre en temps réel.

- **Rôle des compteurs intelligents et des appareils connectés :**

Les compteurs intelligents jouent un rôle central dans les smart grids. Ils enregistrent la consommation d'énergie en détail et en temps réel, permettant aux consommateurs de mieux comprendre et gérer leur utilisation de l'énergie.

Les appareils connectés et les systèmes de gestion de l'énergie domestique peuvent utiliser ces informations pour optimiser l'utilisation de l'énergie, par exemple en éteignant automatiquement les appareils pendant les périodes de tarification élevée.

1.8 Caractéristiques des smart grids

- Adaptabilité : ces réseaux seront capables de se reconfigurer rapidement en fonction de la demande, des ressources disponibles, de la saison, du moment de la journée, …
- Self-healing: le réseau sera capable de détecter des anomalies sur le réseau et ainsi de reconfigurer le réseau de transport pour assurer la continuité de service.
- Réseau optimisé : utilisation optimale des ressources et du réseau.
- Prédictibilité : le réseau sera capable d'analyser et de prédire les besoins.
- Réseau distribué : mis en commun de toutes les sources d'énergie.
- Communication bidirectionnelle : les informations remontent du client final vers les lieux de production.
- Smart pricing: adapter le prix de l'électricité consommée en fonction de la manière dont l'électricité a été produite.

1.9 Les avantages des smart grids

Les smart grids, ou réseaux électriques intelligents, représentent une avancée majeure dans la gestion et la distribution de l'électricité, offrant de multiples avantages tant pour les consommateurs que pour les fournisseurs d'énergie.
Ces systèmes utilisent des technologies de l'information et de la communication pour optimiser la production, la distribution et la consommation d'électricité, en intégrant efficacement les sources d'énergie renouvelable et en améliorant la fiabilité du réseau électrique.

- **Transmission d'électricité plus efficace :**

Les smart grids permettent une gestion plus efficace de la transmission d'électricité, en adaptant la production aux besoins de consommation réels. Cette optimisation réduit les pertes d'énergie lors de la transmission et de la distribution, contribuant ainsi à une utilisation plus rationnelle des ressources énergétiques.

- **Restauration plus rapide après des perturbations électriques :**

Grâce à leur capacité à détecter et à réagir automatiquement aux pannes ou aux perturbations, les smart grids facilitent une restauration plus rapide du service électrique. Cette réactivité améliore la résilience du réseau face aux incidents, réduisant les désagréments pour les utilisateurs et les risques pour les infrastructures critiques.

- **Réduction des coûts d'exploitation et de gestion pour les services publics :**

Les smart grids offrent aux gestionnaires de réseaux une meilleure connaissance des flux d'énergies, permettant ainsi un dimensionnement plus précis des infrastructures et une gestion plus fine des ressources. Cette efficacité opérationnelle se traduit par une réduction des coûts d'exploitation et de gestion pour les services publics.

Une étude menée par l'Institut Montaigne en 2021 [**16**] a estimé que le déploiement des Smart Grids en France pouvait permettre d'économiser jusqu'à 7 milliards d'euros par an sur les coûts de production et de distribution d'électricité.

- **Diminution de la demande de pointe et des tarifs d'électricité :**

En permettant une gestion dynamique de la demande, notamment à travers la tarification incitative ou la réponse à la demande, les smart grids contribuent à lisser les pics de consommation.
Cette modulation de la demande aide à réduire les coûts énergétiques pour les consommateurs et à éviter les investissements coûteux dans des capacités de production supplémentaires.

- **Intégration accrue des systèmes d'énergie renouvelable à grande échelle :**

Les smart grids facilitent l'intégration des sources d'énergie renouvelable, comme l'éolien et le solaire, qui sont par nature intermittentes. En ajustant en temps réel la production à la consommation, ils permettent une exploitation plus large de ces énergies propres, contribuant ainsi à la transition énergétique.

- **Meilleure intégration des systèmes de production d'énergie appartenant aux consommateurs :**

Les smart grids transforment les consommateurs en acteurs de la production énergétique, en facilitant l'intégration de leur propre production d'énergie renouvelable au réseau. Cela encourage l'autoconsommation et le développement des énergies vertes à l'échelle locale.

- **Amélioration de la sécurité du réseau électrique :**

En intégrant des technologies avancées de surveillance et de contrôle, les smart grids améliorent la sécurité du réseau électrique. Ils permettent une détection précoce des anomalies et des menaces potentielles, renforçant ainsi la sûreté des infrastructures énergétiques.
Les smart grids jouent un rôle crucial dans l'amélioration de l'efficacité énergétique, la réduction des impacts environnementaux et la promotion des énergies renouvelables, tout en offrant une meilleure expérience aux consommateurs et en renforçant la sécurité et la résilience des réseaux électriques. Les smart grids et le consommateur

Les smart grids, transforment la relation entre les consommateurs et leur consommation d'énergie.
En offrant un contrôle accru, une facturation plus précise et la possibilité de participer activement à la gestion de la demande, ces technologies promettent de rendre la consommation d'énergie plus efficace et plus économique pour les utilisateurs finaux.

- **Contrôle accru des consommateurs sur leur utilisation de l'énergie :**

Les smart grids permettent aux consommateurs d'avoir un contrôle sans précédent sur leur consommation d'énergie. Grâce à l'utilisation de compteurs intelligents et d'applications de gestion de l'énergie, les utilisateurs peuvent surveiller en temps réel leur consommation d'électricité et ajuster leurs habitudes en conséquence.
Cette transparence accrue aide les consommateurs à identifier les appareils énergivores, à optimiser leur utilisation de l'énergie et à réduire leur empreinte carbone.

Facturation plus précise et économies potentielles grâce à la gestion de la consommation en temps réel Avec les smart grids, la facturation de l'électricité devient beaucoup plus précise, reflétant la consommation réelle plutôt que des estimations.

La tarification dynamique, qui ajuste les prix de l'électricité en fonction de la demande et de l'offre en temps réel, offre aux consommateurs la possibilité de réaliser des économies en déplaçant leur consommation vers des périodes où l'électricité est moins chère.
Cette gestion de la consommation en temps réel non seulement permet aux consommateurs de réduire leurs factures d'électricité, mais contribue également à l'équilibrage général du réseau.

- **Participation à la gestion de la demande et à la réponse à la demande :**

Les smart grids encouragent une participation active des consommateurs à la gestion de la demande et à la réponse à la demande.
En ajustant leur consommation en réponse aux signaux du réseau, tels que les prix de l'électricité ou les demandes de réduction de la consommation pendant les périodes de pointe, les consommateurs peuvent jouer un rôle clé dans la stabilisation du réseau et dans la réduction de la nécessité d'activer des centrales électriques coûteuses et polluantes.
Cette interaction bidirectionnelle entre les consommateurs et le réseau électrique favorise une utilisation plus durable de l'énergie et soutient l'intégration des sources renouvelables.
Les smart grids offrent aux consommateurs des outils puissants pour gérer leur consommation d'énergie de manière plus consciente et économique.
En participant activement à la gestion de la demande et en profitant d'une facturation plus précise, les consommateurs peuvent contribuer à la transition énergétique vers un système plus durable et moins dépendant des combustibles fossiles.

1.10 Technologies clés des Smart Grids

Plusieurs technologies novatrices sont au cœur du fonctionnement des Smart Grids et contribuent à améliorer leur efficacité, leur fiabilité et leur intelligence. Voici quelques-unes des principales technologies clés :

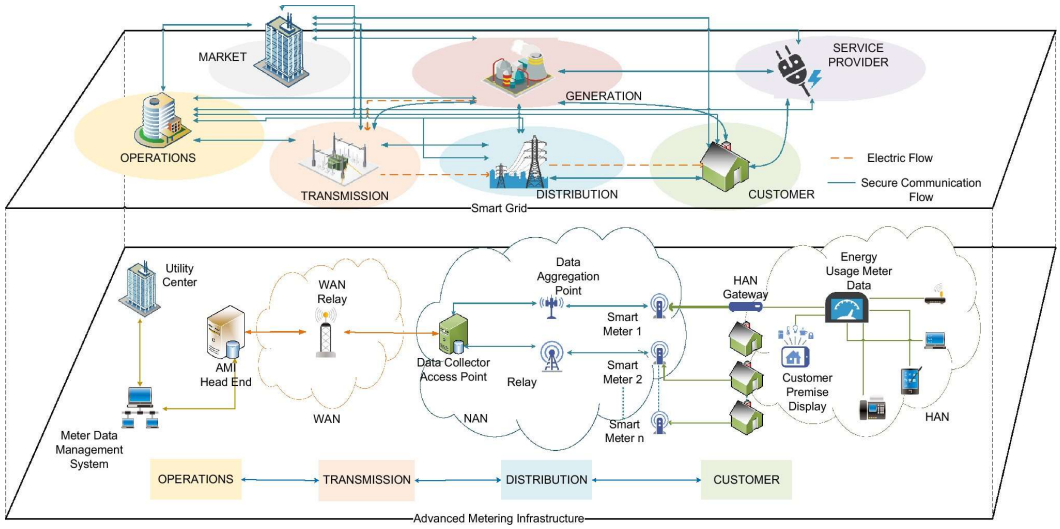


Figure 1.6 Technologies clés des Smart Grids [17]

- **Systèmes de surveillance et de contrôle avancés :**

Des capteurs IoT (Internet of Things – Internet des objets) et des dispositifs de mesure connectés collectent en permanence des données relatives à la production, la distribution et la consommation d'énergie.

Ces informations servent ensuite à piloter et à optimiser les performances du réseau en temps réel, minimisant ainsi les pertes d'énergie et améliorant la qualité de service.

Par exemple, Schneider Electric a annoncé en 2020 son intention d'investir 1,5 milliard d'euros dans la numérisation de ses activités, y compris le développement de solutions basées sur l'IoT pour les Smart Grids.

- **Stockage d'énergie distribué :**

Les batteries stationnaires et les systèmes de stockage thermique ou chimique distribués permettent de combler les lacunes entre l'offre et la demande d'électricité.

Ils accumulent l'excédent d'énergie pendant les périodes de faible consommation et le restituent lorsque la demande augmente, atténuant ainsi les pics de consommation et augmentant la résilience du réseau.

- **Gestion dynamique de la charge :**

Cette technique consiste à adapter la consommation d'électricité aux conditions du marché et aux ressources disponibles.

Elle permet notamment de reporter certaines tâches énergivores, telles que la recharge de véhicules électriques, aux moments où la demande est moindre et l'offre d'énergie plus importante.

Cela contribue à diminuer les pics de consommation et à mieux exploiter les ressources renouvelables. la gestion dynamique de la charge pourrait représenter une réduction potentielle des pointes de consommation allant jusqu'à 10 GW.

1.11 Smart Grids à l'échelle mondiale

Plusieurs exemples à travers le monde montrent l'importance des Smart Grids pour la bonne gestion des réseaux électriques :

- **Italie** : Pionnière en Europe, l'Italie a lancé dès 2001 le projet Telegestore, qui fut le premier exemple de smart grid. Ce projet a permis de connecter 27 millions de foyers au réseau, marquant une étape significative dans l'utilisation des compteurs intelligents pour surveiller la consommation électrique.
- **États-Unis** : Aux États-Unis, le Département de l'Énergie a estimé qu'une amélioration du réseau électrique grâce aux smart grids permettrait d'économiser entre 46 et 117 milliards de dollars de 2010 à 2023. Les smart grids ont été identifiés comme une solution clé pour moderniser un réseau souvent défaillant et obsolète, réduisant ainsi les pertes économiques annuelles dues aux coupures de courant, évaluées à 80 milliards de dollars.
- **Chine** : La Chine a atteint un taux de déploiement de compteurs intelligents de 100% dans certains domaines, démontrant une avance significative dans la digitalisation de son réseau de distribution électrique. Cette étape est cruciale pour l'intégration efficace des énergies

renouvelables et pour la gestion de la demande en temps réel.

- **Union Européenne** : La Commission européenne a identifié les smart grids comme une priorité pour intégrer l'énergie renouvelable, compléter le marché européen de l'énergie et permettre une régulation plus fine de la consommation énergétique.

1.12 Les smart grids et l'avenir

Les smart grids ouvrent la voie à un avenir énergétique plus durable et sécurisé.
Leur déploiement prépare le terrain pour l'intégration croissante des systèmes de recharge de véhicules électriques, offrant ainsi une réponse aux besoins croissants en mobilité électrique.
Ces réseaux facilitent le déploiement des systèmes d'énergie renouvelable, en favorisant l'autoconsommation, en améliorant l'efficacité des bâtiments et en réduisant les pannes électriques grâce à une gestion plus intelligente entre production d'électricité et consommation.
Enfin, les smart grids dessinent une vision d'un avenir plus durable et sécurisé, en contribuant à la réduction des émissions de gaz à effet de serre grâce à un mix énergétique plus efficace, et en favorisant l'augmentation de la rentabilité de l'énergie.

1.13 Les défis à relever :

Bien que les smart grids soient porteurs d'innovation et d'avantages significatifs, leur développement et leur déploiement ne sont pas exempts de défis techniques et réglementaires, La stabilité des smart grids, représente aujourd'hui un défi majeur.
D'une part, ces réseaux doivent intégrer de plus en plus d'énergies renouvelables, comme l'éolien et le solaire, dont la production est variable et difficile à prévoir.
D'autre part, ils font face à des variations imprévues de la demande, liées à l'évolution des habitudes de consommation ou à des événements ponctuels.
Cette double contrainte rend le réseau plus complexe à gérer, et augmente le risque d'instabilités.

1.13.1 Défi de la prédiction d'instabilité des Smart Grids

La prédiction de l'instabilité dans les smart grids est un domaine de recherche et de développement crucial, car les smart grids, malgré leurs nombreux avantages, sont confrontés à des défis uniques qui peuvent compromettre leur stabilité.
Contrairement aux réseaux électriques traditionnels, les smart grids intègrent les facteurs suivants :

- Une forte proportion d'énergies renouvelables intermittentes (solaire, éolien) : Leur production est variable et difficilement prévisible, ce qui rend l'équilibre entre l'offre et la demande plus complexe à maintenir.
- Une production décentralisée et bidirectionnelle **:** Les "prosumers" (producteurs-consommateurs) injectent de l'énergie dans le réseau, créant des flux de puissance plus complexes.
- Une interconnexion accrue et une dépendance aux TIC : Plus de capteurs, d'actionneurs et de systèmes de communication signifient plus de points de défaillance potentiels et une vulnérabilité aux cyberattaques.

- Une inertie réduite du système : Moins de grandes centrales thermiques avec des générateurs rotatifs diminuent l'inertie globale du réseau, le rendant plus susceptible aux fluctuations rapides de fréquence.

  Ces facteurs augmentent la probabilité d'événements d'instabilité tels que les fluctuations de tension et de fréquence, les surcharges, les pannes en cascade et, dans le pire des cas, les blackouts.

1.13.2. Objectifs de la prédiction d'instabilité

La prédiction d'instabilité vise à :

- Anticiper les scénarios critiques : Identifier les conditions qui pourraient mener à l'instabilité avant qu'elles ne se produisent.
- Permettre des actions préventives : Informer les opérateurs du réseau pour qu'ils puissent prendre des mesures correctives (reconfiguration du réseau, ajustement de la production/consommation, etc.).
- Améliorer la résilience du réseau : Rendre le smart grid plus robuste face aux perturbations.
- Optimiser l'exploitation **:** Assurer un fonctionnement sûr et efficace même avec l'intégration de nouvelles technologies.

1.13.3 Méthodes et Approches pour la Prédiction d'Instabilité

Les avancées en intelligence artificielle et en science des données jouent un rôle majeur dans la prédiction de l'instabilité des smart grids :

1. **Apprentissage Automatique:**
   - Classification : Des modèles sont entraînés pour classer l'état du réseau comme "stable" ou "instable" en fonction de diverses caractéristiques (tension, courant, fréquence, données de charge, production renouvelable, etc.). Des algorithmes comme les forêts aléatoires (Random Forest), les arbres de décision (Decision Trees), les machines à vecteurs de support (SVM), et les réseaux de neurones (Neural Networks) sont couramment utilisés.
   - Régression : Prédire des indices de stabilité numériques plutôt qu'une simple classification binaire.
   - Ensembles de modèles**:** Combiner les prédictions de plusieurs modèles ML pour améliorer la robustesse et la précision.
2. **Apprentissage Profond (Deep Learning - DL) :**
   - Les architectures de réseaux de neurones profonds, comme les réseaux neuronaux récurrents (RNN) pour les données séquentielles ou les réseaux génératifs antagonistes (GAN) pour générer des données d'instabilité synthétiques (ce qui est crucial étant donné la rareté des données d'événements instables réels), montrent des performances prometteuses.
   - L'intégration de mécanismes d'attention dans les architectures DL permet au modèle de se concentrer sur les caractéristiques les plus pertinentes pour la prédiction d'instabilité.
3. **Analyse de données temps réel et grands volumes (Big Data Analytics) :** Les smart grids génèrent d'énormes quantités de données à haute fréquence. L'analyse de ces données en temps

quasi réel est essentielle pour détecter les tendances et les anomalies qui pourraient précéder une instabilité.

4. **Techniques basées sur des modèles physiques :** Bien que l'apprentissage automatique soit très en vogue, les modèles physiques du réseau électrique continuent d'être importants. Ils peuvent être combinés avec des approches basées sur les données pour une meilleure robustesse (modèles hybrides).
5. **Détection d'anomalies :** Identifier des comportements inattendus ou des écarts par rapport aux schémas de fonctionnement normaux, qui peuvent être des signes précurseurs d'instabilité.

1.14. Intelligence artificielle et les smart grids :

L'IA et l'apprentissage automatique permettent aux Smart Grids de prendre des décisions autonomes et optimisées en temps réel, en anticipant les comportements des usagers, en adaptant la production et la distribution d'électricité aux conditions météorologiques locales et en tenant compte des tarifs dynamiques appliqués sur le marché.

En somme, les technologies clés mentionnées ci-dessus sont autant d'éléments cruciaux pour le bon fonctionnement des Smart Grids, assurant une gestion efficace et intelligente de l'énergie, tout en encourageant l'adoption massive des énergies renouvelables à travers le monde.

Pour permettrait d'éviter des pannes, réduire les pertes, et surtout renforcer la résilience du système électrique. L'objectif de ce travail est donc clair : prédire la stabilité du réseau à partir des paramètres dynamiques, afin d'anticiper les déséquilibres et d'aider à la prise de décision en temps réel, en intégrant les avantages qui offrent l'intelligence artificiel par les algorithmes de l'apprentissage automatique.

1.15 Conclusion

Dans ce chapitre on a donné une présentation générale sur les Smart Grids qui révolutionne la consommation et la production d'électricité en rendant intelligent l'ensemble du réseau. Cet apport est possible avec l'ajout de nouvelles technologies - capteurs, automates, analyseurs - dans toutes les mailles du réseau.

Le Smart Grid est la réponse retenue pour faire face à l'augmentation des besoins en électricité mais aussi à la nécessité d'intégrer des énergies renouvelables au réseau, afin par exemple de réduire les émissions de gaz à effet de serre. De plus, la maîtrise parfaite de la production/consommation d'électricité pourrait devenir demain un enjeu majeur avec le remplacement du parc automobile thermique par un parc électrique.

Les Smart Grids doivent maintenir plusieurs types de stabilité pour garantir un fonctionnement fiable et durable. Les approches basées sur l'apprentissage automatique, offrent des outils puissants pour prédire et gérer ces stabilités, en s'appuyant sur des paramètres dynamiques et des ensembles de données complexes.

Le chapitre suivant sera consacré pour la présentation d'apprentissage automatique et ces algorithmes.

Chapitre 2 :

Apprentissage Automatique

Chapitre 2 : Apprentissage Automatique

2.1 Introduction

L'apprentissage automatique (AA) est une branche de l'intelligence artificielle qui permet aux systèmes d'apprendre à partir de données, d'identifier des modèles et de prendre des décisions avec une intervention humaine minimale [17]. Au lieu d'être explicitement programmés pour chaque tâche, ces systèmes sont entraînés sur de vastes ensembles de données pour découvrir des relations et des structures cachées.

2.2 Les principaux types d'apprentissage automatique

Il existe trois grandes catégories d'apprentissage automatique, chacune adaptée à des types de problèmes différents :

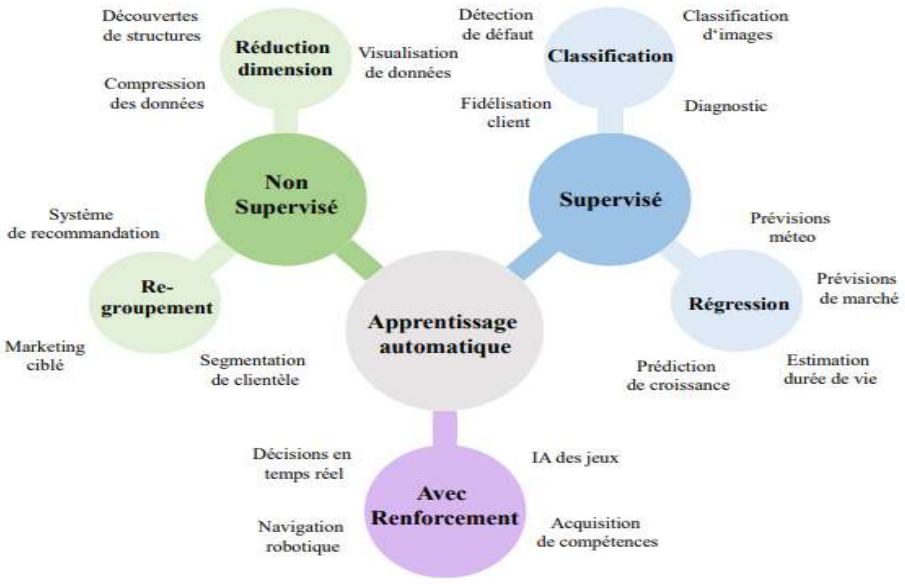


Figure 2.1 types d'apprentissage automatique [18].

- **Apprentissage supervisé (Supervised Learning):**
  - Principe: L'algorithme est entraîné sur un ensemble de données "étiquetées" (ou "labélisées"), c'est-à-dire des données pour lesquelles la sortie correcte est déjà connue. Le modèle apprend à mapper les entrées aux sorties souhaitées.
  - Exemples:
    - Classification: Prédire une catégorie discrète (ex: spam/non-spam, chat/chien, détection de fraude).
    - Régression: Prédire une valeur continue (ex: prix d'une maison, température, prévisions de vente).

- Algorithmes courants**:** Régression linéaire/logistique, arbres de décision, forêts aléatoires, Machines à Vecteurs de Support (SVM), k-plus proches voisins (k-NN), réseaux de neurones.

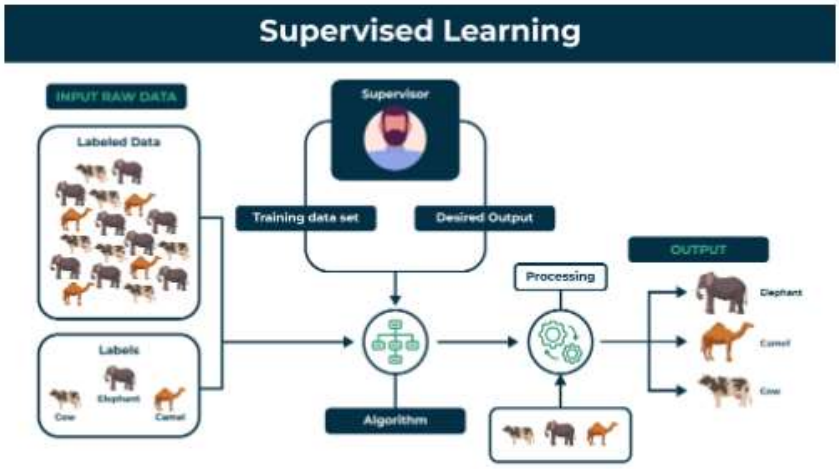


Figure 2.2 Apprentissage supervisé [19].

- **Apprentissage non supervisé (Unsupervised Learning):**
  - Principe: L'algorithme travaille avec des données "non étiquetées", c'est-à-dire sans sortie prédéfinie. L'objectif est de découvrir des structures cachées, des motifs ou des regroupements dans les données.
  - Exemples:
    - Clustering (regroupement): Diviser les données en groupes (clusters) en fonction de leurs similarités (ex: segmentation de clientèle, regroupement de documents similaires).
    - Réduction de dimensionnalité: Simplifier les données en réduisant le nombre de caractéristiques tout en conservant l'information essentielle (ex: analyse en composantes principales - PCA).
    - Règles d'association: Découvrir des relations entre des variables dans de grandes bases de données (ex: "les clients qui achètent du pain achètent aussi du lait").
  - Algorithmes courants**:** K-Means, Clustering hiérarchique, PCA, auto-encodeurs.

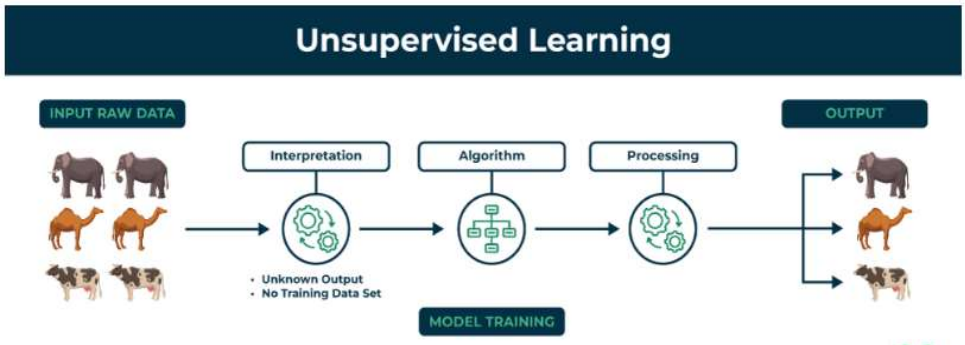


Figure 2.3 Apprentissage non supervisé [20].

- **Apprentissage par renforcement (Reinforcement Learning):**
    - Principe: Un agent (le programme d'IA) apprend à prendre des décisions en interagissant avec un environnement. Il reçoit des "récompenses" positives ou négatives en fonction de ses actions, et l'objectif est de maximiser la récompense cumulée au fil du temps. C'est un processus d'apprentissage par essai et erreur.
    - Exemples:
        - Jeux (AlphaGo, jeux vidéo).
        - Robotique (apprentissage de mouvements, navigation).
        - Systèmes de contrôle autonomes.
    - **Algorithmes courants:** Q-learning, SARSA, Deep Q-Networks (DQN).

D'autres catégories existent, comme l'apprentissage semi-supervisé (mélange de supervisé et non supervisé, utilisant un petit ensemble de données étiquetées et un grand ensemble non étiqueté) et l'apprentissage par transfert (utiliser un modèle pré-entraîné sur une tâche pour en résoudre une nouvelle).

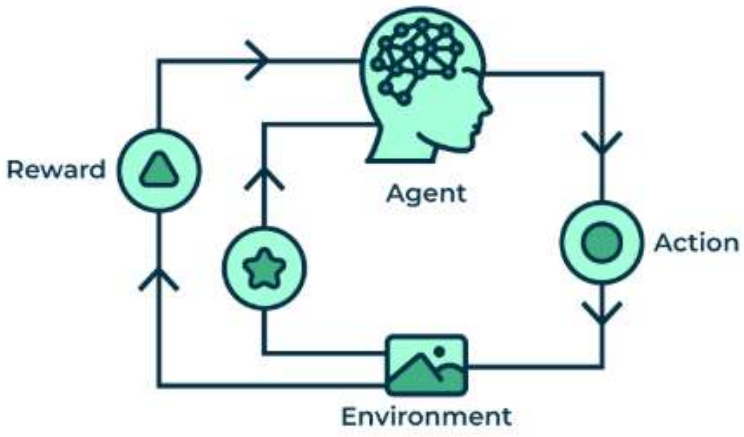


Figure 2.4 Apprentissage par renforcement [21].

2.3 Composants fondamentaux d'un modèle d'apprentissage automatique

Tout algorithme de Machine Learning repose sur trois éléments principaux :

- **Représentation :** Comment le modèle est-il structuré ? Comment la connaissance est-elle représentée ? (Ex: arbres de décision, réseaux de neurones, équations linéaires).
- **Évaluation :** Comment évalue-t-on la "qualité" d'un modèle ? C'est souvent mesuré par une "fonction de coût" (ou "fonction de perte") qui quantifie l'erreur entre les prédictions du modèle et les valeurs réelles. L'objectif est de minimiser cette erreur.
- **Optimisation :** Quel est le processus pour trouver le meilleur modèle ? Comment ajuster les paramètres du modèle pour minimiser la fonction de coût ? (Ex: descente de gradient, algorithmes génétiques).

2.4. Le processus d'apprentissage automatique

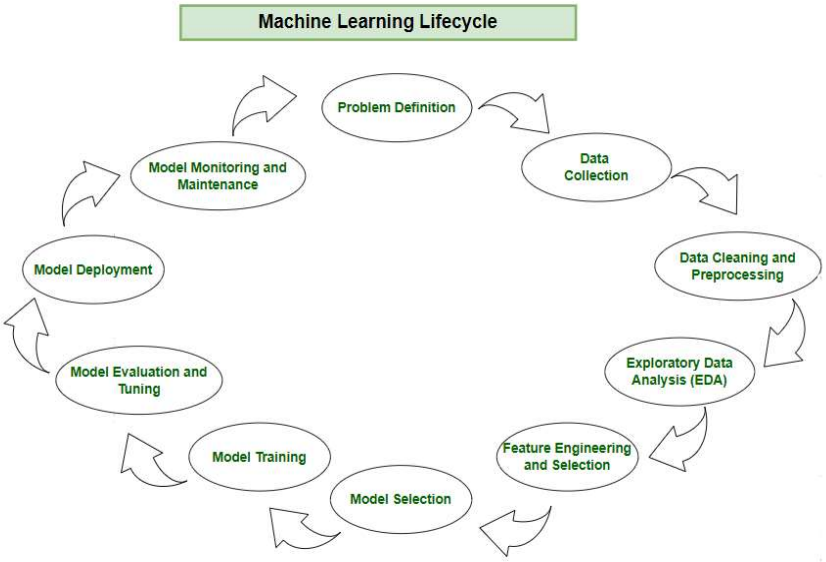


Figure 2.5 processus d'apprentissage automatique [22].

Le développement d'un modèle de Machine Learning suit généralement plusieurs étapes :

1. Collecte et préparation des données : Acquérir des données pertinentes et les nettoyer (gestion des valeurs manquantes, des erreurs, standardisation, etc.). C'est une étape cruciale qui représente souvent la majorité du travail.
2. Sélection des caractéristiques (Feature Engineering): Identifier les variables (caractéristiques) les plus pertinentes qui influenceront la sortie du modèle. Parfois, de nouvelles caractéristiques sont créées à partir des existantes.
3. Choix du modèle/algorithme : Sélectionner l'algorithme d'apprentissage automatique le plus approprié pour la tâche à accomplir (classification, régression, clustering, etc.).
4. Entraînement du modèle: Alimenter l'algorithme avec les données d'entraînement. Le modèle ajuste ses paramètres internes pour minimiser la fonction de coût.
5. Évaluation du modèle: Tester le modèle sur un ensemble de données "invisibles" (non utilisées pendant l'entraînement) pour évaluer ses performances et sa capacité à généraliser. Des métriques spécifiques (précision, rappel, F1-score pour la classification ; erreur quadratique moyenne pour la régression) sont utilisées.
6. Optimisation et réglage des hyperparamètres : Ajuster les hyperparamètres du modèle (paramètres qui ne sont pas appris directement par l'algorithme mais définis avant l'entraînement, comme le taux d'apprentissage ou le nombre de neurones dans une couche) pour améliorer les performances.
7. Déploiement : Une fois le modèle validé, il peut être déployé dans un environnement de production pour faire des prédictions sur de nouvelles données.

2.5. Fondements mathématiques

L'apprentissage automatique repose sur de solides bases mathématiques, notamment :

- Algèbre linéaire : Pour la manipulation des vecteurs et matrices de données, les transformations.
- Calcul différentiel : Pour l'optimisation des fonctions de coût (ex: descente de gradient).
- Probabilités et statistiques : Pour la modélisation de l'incertitude, l'inférence, la compréhension des distributions de données, les tests d'hypothèses.
- **Théorie de l'optimisation :** Pour trouver les meilleurs paramètres de modèle.

2.6. Applications

L'apprentissage automatique est omniprésent aujourd'hui et se retrouve dans de nombreuses applications :

- Reconnaissance d'image et de parole : Identification de visages, classification d'objets, assistants vocaux.
- Traitement du langage naturel (NLP): Traduction automatique, analyse de sentiment, chatbots, résumé de texte.
- Systèmes de recommandation : Films (Netflix), produits (Amazon), musique (Spotify).
- Détection de fraude : Transactions bancaires, assurance.
- Maintenance prédictive : Anticiper les pannes d'équipement.
- Véhicules autonomes : Prise de décision, reconnaissance d'obstacles.
- Santé : Diagnostic médical, découverte de médicaments.

En somme, l'apprentissage automatique est une discipline en constante évolution qui permet aux systèmes informatiques de devenir plus intelligents et autonomes en apprenant des données, ouvrant la voie à des innovations majeures dans presque tous les secteurs.

2.7. Les Algorithmes d'apprentissage automatique utilisés dans ce projet

Dans ce travail on utilise les Algorithmes de classification d'apprentissage supervisé suivants : Régression logistique (RL) - Naive Bayes (NB) - Support Vector Machine (SVM) - Random Forest (RF) Extrême gradient boost (XGBoost) pour concevoir un modèle de prédiction de l'instabilité des Smart grids

2.7.1. Régression Logistique

Bien que son nom contienne "régression", la régression logistique est principalement un algorithme de classification, très populaire pour les problèmes de classification binaire (deux classes). Elle peut être étendue pour la classification multi-classes.

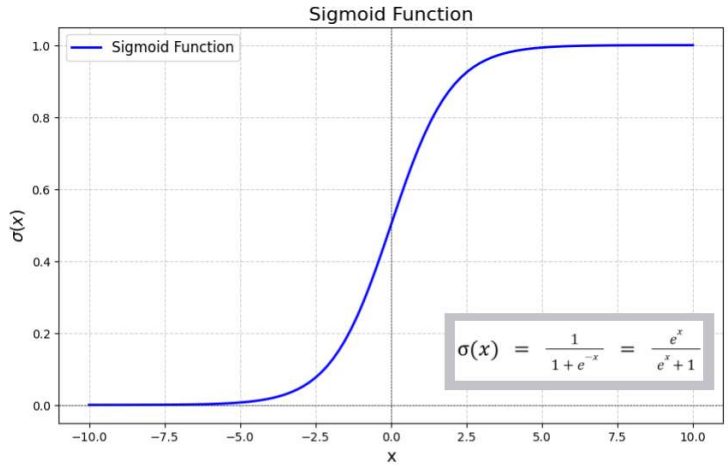


Figure 2.6 Régression Logistique [23].

- **Principe :**
  - La régression logistique modélise la probabilité qu'une instance appartienne à une certaine classe.
  - Elle utilise la fonction logistique (ou sigmoïde) pour transformer la combinaison linéaire des variables d'entrée en une probabilité qui est toujours comprise entre 0 et 1.
  - Un seuil de décision (généralement 0.5) est ensuite appliqué pour classer l'instance dans l'une des classes. Si la probabilité est supérieure à 0.5, elle est classée dans la classe positive (1) ; sinon, dans la classe négative (0).
  - Les coefficients (b_i) sont appris en maximisant la vraisemblance (ou en minimisant la fonction de coût, souvent la log-loss ou entropie croisée) sur les données d'entraînement.
- **Avantages :**
  - Simple et facile à comprendre et à interpréter : Les coefficients indiquent l'importance et la direction de l'influence de chaque variable sur la probabilité.
  - Rapide à entraîner et à prédire : Convient bien aux grands ensembles de données.
  - Donne des probabilités : Utile lorsque la confiance de la prédiction est importante.
  - Moins sujet au surapprentissage que des modèles plus complexes, surtout avec la régularisation.
  - Bon point de départ pour de nombreux problèmes de classification binaire.
- **Inconvenient:**
  - Hypothèse de linéarité : Suppose une relation linéaire entre les caractéristiques et le logit de la probabilité. Ne peut pas capturer des relations non linéaires complexes à moins d'ajouter des caractéristiques transformées manuellement.
  - Sensible aux valeurs aberrantes **:** Peut être influencée par des outliers extrêmes.
  - Nécessite que les caractéristiques soient indépendantes (ou du moins pas fortement corrélées) pour une interprétation facile.
  - Peut sous-performer si les données sont très complexes ou si les classes ne sont pas linéairement séparables.

- **Applications typiques :**
    - Prédiction du churn client (départ).
    - Diagnostic médical (maladie/pas de maladie).
    - Détection de fraude (transaction frauduleuse/non frauduleuse).
    - Filtrage de spams.
    - Scoring de crédit.

2.7.2. Machines à Vecteurs de Support

Les SVM sont des algorithmes puissants et polyvalents utilisés pour la classification et la régression, mais surtout connus pour la classification.

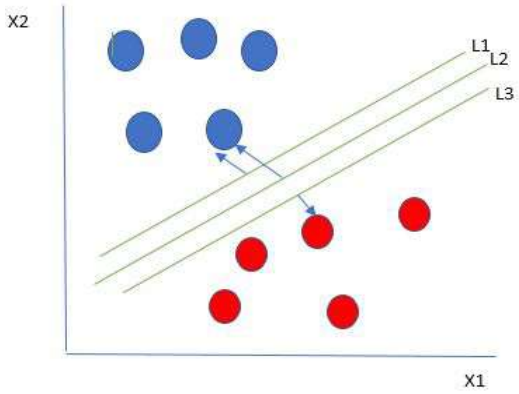


Figure 2.7 Algorithme SVM [24].

- **Principe :**
    - L'objectif principal des SVM est de trouver l'hyperplan de séparation optimal qui maximise la marge entre les différentes classes dans un espace de caractéristiques.
    - La marge est la distance entre l'hyperplan et les points de données les plus proches de chaque classe, appelés vecteurs de support.
    - Plus la marge est grande, plus le modèle est robuste et meilleure est sa capacité de généralisation.
    - Pour les données non linéairement séparables : Les SVM utilisent une technique appelée le "noyau" (ou Kernel Trick). Cela permet de projeter implicitement les données dans un espace de dimension supérieure où elles deviennent linéairement séparables. Les fonctions de noyau courantes incluent le noyau linéaire, polynomial, gaussien (RBF) et sigmoïde.
- **Avantages :**
    - Efficace dans les espaces de grande dimension : Fonctionne très bien même lorsque le nombre de caractéristiques est supérieur au nombre d'échantillons.
    - Efficace avec le "Kernel Trick" : Permet de modéliser des relations non linéaires complexes sans transformer explicitement les données dans un espace de grande dimension.

- Moins sujet au surapprentissage : Grâce à l'objectif de maximisation de la marge, qui favorise la généralisation.
- Robuste face aux valeurs aberrantes dans une certaine mesure (grâce à l'utilisation de variables de slack).

- **Inconvénients :**
  - Complexe à interpréter : Contrairement à la régression logistique ou aux arbres de décision, il est difficile de comprendre directement comment une prédiction est faite.
  - Lent pour les grands ensembles de données : Le temps d'entraînement peut être très long (complexité souvent en O(n^2) ou O(n^3) où n est le nombre d'échantillons).
  - Sensible au choix du noyau et à ses hyperparamètres : Un mauvais choix peut affecter considérablement les performances.
  - Ne fournit pas directement des probabilités, bien que certaines implémentations puissent estimer des probabilités.
- **Applications typiques :**
  - Reconnaissance faciale et de caractères.
  - Classification de texte et de documents.
  - Bioinformatique (classification de séquences de protéines).
  - Détection d'intrusion.

2.7.3. Naive Bayes (Bayes Naïf)

Le classifieur Naive Bayes est une famille d'algorithmes de classification basée sur le théorème de Bayes avec une "hypothèse de naïveté" forte.

- **Principe :**
  - Il est basé sur le théorème de Bayes, qui calcule la probabilité qu'un événement (A) se produise étant donné qu'un autre événement (B) s'est déjà produit :

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \qquad 2.1$$

  - L'hypothèse "naïve" est que toutes les caractéristiques sont indépendantes les unes des autres étant donné la classe. C'est une simplification forte qui est rarement vraie dans la réalité, mais l'algorithme fonctionne étonnamment bien malgré cela.
  - Pour la classification, il calcule la probabilité de chaque classe étant donné les caractéristiques d'une nouvelle instance, et attribue la classe ayant la probabilité la plus élevée.
  - Il existe différentes variantes de Naive Bayes selon la distribution supposée des données :
    - Gaussien Naive Bayes : Pour les caractéristiques continues (suppose une distribution gaussienne).
    - Multinomial Naive Bayes : Pour le comptage de mots (souvent utilisé en NLP).
    - Bernoulli Naive Bayes : Pour les caractéristiques binaires.

- **Avantages :**
    - Extrêmement rapide à entraîner et à prédire : Très efficace pour les grands ensembles de données.
    - Performances étonnamment bonnes : Souvent compétitif avec des algorithmes plus complexes, surtout pour les problèmes de classification de texte.
    - Facile à implémenter : Relativement simple à comprendre.
    - Peut gérer des données avec de nombreuses dimensions.
    - Fonctionne bien même avec un petit ensemble de données d'entraînement.
- **Inconvénients :**
    - L'hypothèse d'indépendance des caractéristiques est rarement vraie : Cela peut parfois limiter ses performances.
    - Peut être sensible aux problèmes de "probabilité nulle" : Si une caractéristique n'apparaît pas avec une certaine classe dans les données d'entraînement, la probabilité pour cette combinaison sera de zéro, ce qui peut fausser les prédictions (souvent géré par le lissage de Laplace).
    - Ne fournit pas des estimations de probabilité très fiables (souvent des estimations très proches de 0 ou 1).
- **Applications typiques :**
    - Filtrage de spams.
    - Classification de texte (analyse de sentiment, catégorisation de documents).
    - Systèmes de recommandation.
    - Diagnostic médical.

2.7.4. Forêts Aléatoires

Random Forest est un algorithme d'apprentissage ensembliste (ensemble learning) très populaire pour la classification et la régression. Il est basé sur la combinaison de plusieurs arbres de décision.

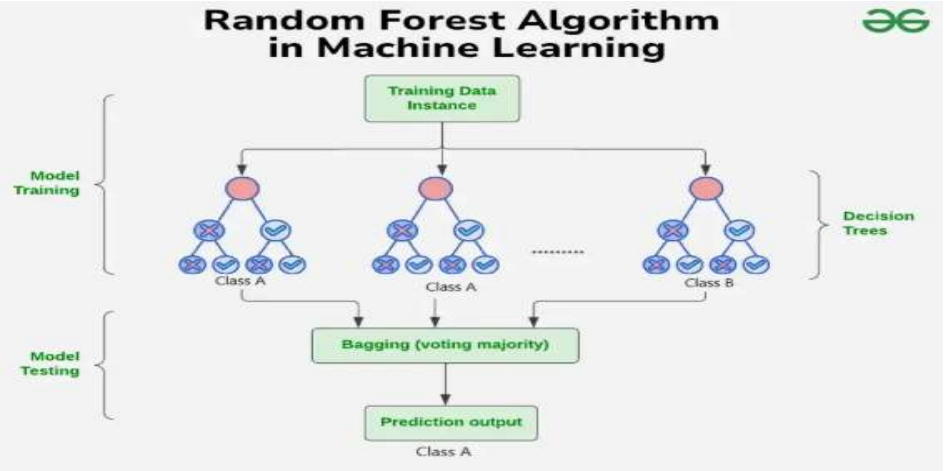


Figure 2.8 Random forest [25].

- **Principe :**
  - Une forêt aléatoire construit un grand nombre d'arbres de décision (d'où le nom "forêt") lors de l'entraînement.
  - Pour introduire de la diversité parmi les arbres :
    - Bagging (Bootstrap Aggregating) : Chaque arbre est entraîné sur un échantillon aléatoire (avec remplacement) des données d'entraînement (bootstrap).
    - Sélection aléatoire des caractéristiques : Lors de la construction de chaque nœud d'un arbre, seules un sous-ensemble aléatoire des caractéristiques est considéré pour la meilleure division.
  - Pour la prédiction :
    - Classification : Chaque arbre vote pour une classe, et la classe majoritaire est choisie (vote majoritaire).
    - Régression : La moyenne des prédictions de tous les arbres est calculée.
- **Avantages :**
  - Très précis et robuste : Généralement l'un des algorithmes les plus performants.
  - Moins sujet au surapprentissage : Grâce à l'agrégation de nombreux arbres et à la randomisation.
  - Gère bien les données non linéaires et les interactions complexes entre caractéristiques.
  - Peut gérer un grand nombre de caractéristiques et d'échantillons.
  - Peut gérer des valeurs manquantes.
  - Fournit l'importance des caractéristiques : Permet de comprendre quelles variables sont les plus pertinentes pour la prédiction.
  - Relativement facile à utiliser (moins de réglage d'hyperparamètres que les réseaux de neurones ou SVM).
- **Inconvénients :**
  - Manque d'interprétabilité : Bien qu'il fournisse l'importance des caractéristiques, il est difficile de comprendre le chemin exact qu'un seul arbre prend pour une prédiction donnée. C'est une "boîte noire" par rapport aux arbres de décision individuels.
  - Plus lent et plus gourmand en ressources que les arbres de décision individuels ou la régression logistique, car il entraîne de nombreux arbres.
  - Peut avoir des performances similaires à des arbres de décision élagués si les données sont très simples.
- **Applications typiques :**
  - Reconnaissance d'image.
  - Détection de fraude.
  - Modélisation prédictive dans de nombreux domaines (finance, santé, marketing).
  - Analyse de données génomiques.

2.7.5. XGBoost

XGBoost est une implémentation optimisée et hautement efficace de l'algorithme de Gradient Boosting, également un algorithme d'apprentissage ensembliste populaire pour la classification et la régression.

- **Principe du Gradient Boosting:**
  - Construit séquentiellement des arbres de décision.
  - Chaque nouvel arbre essaie de corriger les erreurs (résidus) commises par l'ensemble des arbres précédents.
  - L'apprentissage est "adaptatif" : les arbres sont construits de manière à se concentrer sur les exemples les plus difficiles.
  - Utilise la descente de gradient pour minimiser la fonction de perte.

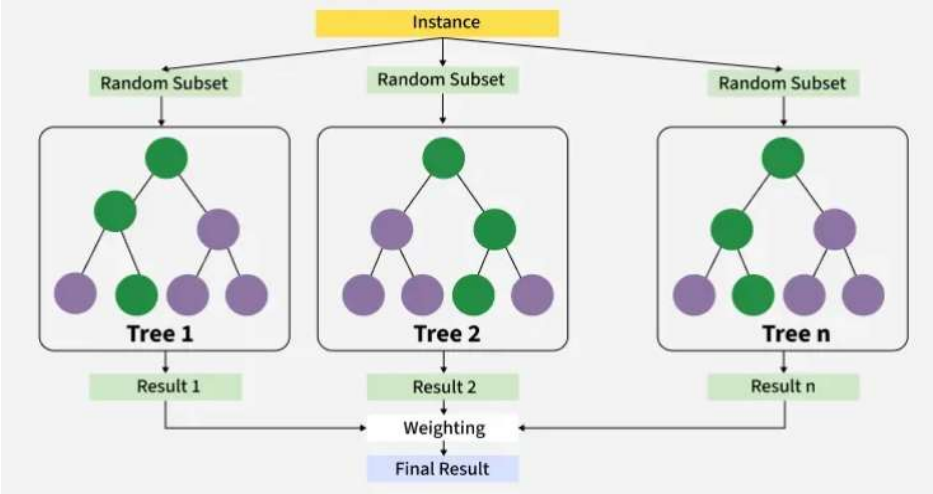


Figure 2.9 XGBoost [26].

- **Spécificités de XGBoost :**
  - Optimisation **:** Conçu pour être extrêmement rapide et performant, avec des optimisations au niveau du matériel et des algorithmes (parallélisation, gestion des valeurs manquantes, élagage des arbres, etc.).
  - Régularisation : Inclut des termes de régularisation (L1 et L2) pour prévenir le surapprentissage.
  - Gestion des valeurs manquantes : Peut apprendre à gérer les valeurs manquantes en les traitant comme des motifs distincts.
  - Optimisation des fonctions de perte personnalisées : Très flexible.
- **Avantages :**
  - Performance et précision exceptionnelles : Souvent vainqueur des compétitions de Machine Learning .
  - Rapidité d'exécution : Très optimisé, peut être beaucoup plus rapide que d'autres implémentations de boosting.
  - Flexibilité : Peut être utilisé pour des problèmes de classification, régression, et classement.
  - Gère bien les valeurs manquantes.
  - Gère un grand nombre de caractéristiques.
  - Robuste aux valeurs aberrantes.
  - Comprend des fonctionnalités pour la validation croisée intégrée.

- **Inconvénients :**
  - Moins interprétable : Comme Random Forest, c'est une "boîte noire" en raison de la complexité des arbres ensemblistes et du processus séquentiel.
  - Sensible aux hyperparamètres : Nécessite un réglage minutieux des hyperparamètres pour atteindre des performances optimales.
  - Peut prendre du temps d'entraînement si les données sont très grandes et les hyperparamètres ne sont pas optimisés.
  - Plus sujet au surapprentissage que Random Forest si les hyperparamètres ne sont pas correctement réglés.
- **Applications typiques :**
  - Pratiquement tous les problèmes de classification et de régression où la performance est primordiale.
  - Détection de fraude.
  - Scoring de crédit.
  - Prévisions de vente.
  - Compétitions de Machine Learning.

2.10. Métriques pour évaluer les performances d'un modèle

Pour comprendre comment évaluer un modèle de classification il est crucial, de connaitre les termes métriques suivants : Matrice de confusion, Accuracy (justesse), la précision, le rappel, le F1-score et l'AUC-ROC.

Matrice de confusion

La matrice de confusion est un outil fondamental en apprentissage automatique et en classification, utilisé pour évaluer la performance d'un modèle de classification.
Une matrice de confusion est un tableau carré qui présente le nombre de prédictions correctes et incorrectes, réparties selon les différentes classes. Elle est particulièrement utile pour les problèmes de classification binaire ou multi-classes.

	Prédit positif	Prédit negatif
Réel positif	Vrai positif (VP)	Faux négatif (FN)
Réel négatif	Faux positif (FP)	Vrai négatif (VN)

Tableau 2.1 Structure d'une matrice de confusion

- Vrai Positif (VP) : nombre d'exemples positifs correctement classés.
- Faux Positif (FP) : nombre d'exemples négatifs incorrectement classés comme positifs.
- Faux Négatif (FN) : nombre d'exemples positifs incorrectement classés comme négatifs.
- Vrai Négatif (VN) : nombre d'exemples négatifs correctement classés.

- **Accuracy**: est la métrique la plus simple et la plus intuitive. Elle mesure la proportion de prédictions correctes parmi toutes les prédictions faites par le modèle.

$$\text{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}} \qquad 2.2$$

Par exemple, si un modèle prédit correctement 90 cas sur 100, sa justesse est de 90 %. Bien que facile à comprendre, la justesse peut être trompeuse, surtout si les données sont déséquilibrées (par exemple, 95 % de cas négatifs et 5 % de cas positifs). Un modèle qui prédit toujours la classe majoritaire pourrait avoir une justesse élevée sans être réellement utile.

- **Précision** : La précision nous dit, parmi toutes les fois où le modèle a prédit la classe positive, combien de fois il a eu vraiment raison. Elle répond à la question : "Parmi les cas que j'ai identifiés comme positifs, quelle proportion est réellement positive ?"

$$\text{Précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}} \qquad 2.3$$

- Vrais Positifs : Le modèle a prédit positif, et c'était réellement positif.
- Faux Positifs : Le modèle a prédit positif, mais c'était réellement négatif (une "fausse alerte" ou erreur de Type I).

Si un modèle de détection de spams a une haute précision, cela signifie que peu de mails légitimes sont marqués à tort comme spams. C'est important quand le coût d'un faux positif est élevé.

- **Rappel (Recall)** : Le rappel (aussi appelé sensibilité ou True Positive Rate) mesure la capacité de votre modèle à trouver tous les cas positifs réels. Il répond à la question : "Parmi tous les cas qui étaient réellement positifs, quelle proportion mon modèle a-t-il réussi à identifier ?"

$$\text{Rappel} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Negatifs}} \qquad 2.4$$

- Vrais Positifs: Le modèle a prédit positif, et c'était réellement positif.
- Faux Négatifs: Le modèle a prédit négatif, mais c'était réellement positif (un "manque" ou erreur de Type II).

Un modèle de diagnostic médical avec un rappel élevé est crucial car il minimise le risque de passer à côté de patients malades.

- **F1-score** : Le F1-score est la moyenne harmonique de la précision et du rappel. C'est une métrique qui cherche à trouver un équilibre entre ces deux valeurs.

$$\text{F1-scor} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \qquad 2.5$$

Il est particulièrement utile lorsque vous avez un déséquilibre important entre les classes. Si un modèle a une très bonne précision mais un très mauvais rappel (ou vice-versa), son F1-score sera faible, signalant que le modèle ne performe pas bien de manière globale.

- **AUC-ROC (Area Under the Receiver Operating Characteristic Curve)**

L'AUC-ROC est une métrique qui évalue la capacité de votre modèle à distinguer entre les classes positives et négatives à travers tous les seuils de classification possibles.

- La courbe ROC est un graphique qui trace le Taux de Vrais Positifs (TVP) (le rappel) en fonction du Taux de Faux Positifs (TFP) à divers seuils de décision.

$$\text{TFP} = \frac{\text{Faux Positifs}}{\text{Faux Positifs} + \text{Vrais Negatifs}} \qquad 2.6$$

  - Vrais Négatifs : Le modèle a prédit négatif, et c'était réellement négatif.
- L'AUC est l'aire sous cette courbe ROC.
  - Une AUC de 1 représente un classifieur parfait (il distingue parfaitement les classes).
  - Une AUC de 0,5 indique un classifieur qui ne fait pas mieux que le hasard.
  - Plus l'AUC est élevée (proche de 1), meilleure est la capacité discriminatoire du modèle.

L'AUC-ROC est une métrique robuste pour les ensembles de données déséquilibrés car elle n'est pas influencée par la proportion de chaque classe. Elle nous donne une idée globale de la performance du modèle, indépendamment du seuil que nous choisissons pour décider si une prédiction est positive ou négative.

Ces métriques, prises ensemble, nous offrent une vision beaucoup plus complète et nuancée de la performance de notre modèle de classification, nous permettant de faire des choix éclairés selon le contexte spécifique de notre problème.

2.9 Conclusion

Dans ce chapitre, on a présenté les fondements d'apprentissage automatique et les algorithmes qui peuvent nous aider à prédire l'instabilité des smart grids. Dans l'étude qui suit, l'objectif principal est d'appliquer ces différents algorithmes, Régression logistique (RL) - Naive Bayes (NB) - Support Vector Machine (SVM) - Random Forest (RF) Extrême gradient boost (XGBoost) pour concevoir un modèle pertinent de prédiction de l'instabilité des Smart grids

Chapitre 3 :

Expérimentation et discussion des résultats

Chapitre 3 : Expérimentation et discussion des résultats

3.1 Introduction

Dans ce dernier chapitre, on présente d'abord une étude technique dans laquelle on défini l'environnement logiciel utilisée pour coder notre application, puis on définirai le dataset avec une description de ses caractéristiques et les étapes de pré-traitement des données (explorer, nettoyer, sélection de modèle et évaluation) pour analyser les résultats et choisir le meilleur modèle à suivre.

3.2 Outils et environnement de développement

3.2.1 Matérielles utilisés : Le tableau ci-dessous présente les spécifications du poste de travail utilisés pour la mise en œuvre du système développé.

PC	Caractéristiques
Marque	LENOVO IDAIPAD 700
System exploitation	Windows 10 Professionnel 64 bits
Processor	Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz 2.60 GHz
RAM	16 Go

Table 3.1 caractéristique du PC

3.2.2 Langage de programmation Python

Python [27] C'est un langage de programmation multiparadigme et le langage de programmation dominant dans la data science avec de nombreuses implémentations ce qui le rend encore plus intéressant.

Pourquoi Python ?

- Écosystème ML Robuste : Python dispose de l'écosystème le plus riche et le plus mature pour le machine learning et la science des données.
- Facilité d'Utilisation : Sa syntaxe claire et intuitive réduit le temps de développement, permettant de se concentrer davantage sur la modélisation et l'analyse.
- Grandes Communautés : Une vaste communauté signifie de nombreuses ressources, tutoriels, et un support pour résoudre les problèmes.
- Intégration Facile : Python s'intègre bien avec d'autres systèmes et langages si nécessaire.
- Visualisation de Données **:** D'excellentes bibliothèques pour créer des graphiques et visualiser les résultats, ce qui est crucial pour l'analyse et la présentation.

3.2.3 Bibliothèques de Machine Learning (ML) Essentielles

Voici les bibliothèques Python incontournables pour ce projet, spécifiquement pour les algorithmes nécessaires pour le travail.

- **NumPy :**
  - **Rôle :** C'est la bibliothèque fondamentale pour le calcul numérique en Python. Elle est indispensable pour manipuler des tableaux (arrays) et des matrices de grandes dimensions de manière efficace. Presque toutes les autres bibliothèques scientifiques s'appuient sur NumPy.
  - **Utilisation :** Pour la manipulation de vos données numériques (tensions, courants, fréquences, etc.) avant et pendant le traitement ML.
- **Pandas :**
  - **Rôle :** La bibliothèque de référence pour la manipulation et l'analyse de données structurées. Elle offre des structures de données comme les DataFrames, idéales pour gérer vos datasets (données de capteurs, états du réseau, labels d'instabilité).
  - **Utilisation :** Pour le chargement, le nettoyage, la transformation, le filtrage et l'exploration des données. Essentielle pour la phase de prétraitement.
- **Scikit-learn (sklearn) :**
  - **Rôle :** La bibliothèque la plus populaire et complète pour le machine learning classique en Python. Elle fournit des implémentations efficaces et faciles à utiliser pour la plupart des algorithmes d'apprentissage supervisé et non supervisé.
  - **Utilisation :** Vous y trouverez les implémentations de :
    - LogisticRegression (Régression Logistique)
    - SVC (Support Vector Classifier pour SVM)
    - GaussianNB, MultinomialNB (Naïf Bayes)
    - RandomForestClassifier (Random Forest)
    - Ainsi que des outils pour la sélection de caractéristiques, le prétraitement des données (mise à l'échelle, encodage), la division des jeux de données (train_test_split), la validation croisée (cross_val_score, GridSearchCV), et l'évaluation des modèles (metrics : accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, confusion_matrix, classification_report).
- **XGBoost (Extreme Gradient Boosting) :**
  - **Rôle :** Une bibliothèque optimisée et hautement performante qui implémente l'algorithme Gradient Boosting. Elle est célèbre pour sa vitesse et sa précision, souvent gagnante dans les compétitions de Kaggle.
  - **Utilisation :** Pour l'implémentation de l'algorithme XGBoostClassifier, qui sera probablement un des modèles les plus performants de votre projet.

- **Matplotlib et Seaborn :**
    - **Rôle :** Les bibliothèques de référence pour la création de visualisations statiques et interactives de haute qualité.
    - **Utilisation :** Pour visualiser la distribution de vos données, les corrélations, les courbes ROC, les matrices de confusion, et présenter les performances de vos modèles de manière claire et percutante dans votre rapport.
- **SciPy (Scientific Python) :**
    - **Rôle :** Offre des modules pour l'optimisation, l'algèbre linéaire, l'intégration, les statistiques, le traitement du signal et de l'image.
    - **Utilisation :** Moins centrale que NumPy ou Pandas pour ce type de projet, mais elle peut être utile pour des opérations statistiques avancées ou certaines optimisations spécifiques. Scikit-learn s'appuie souvent sur SciPy.

3.2.4 Environnent de Développent

Le choix de votre environnement de développement est important pour la productivité et la gestion de notre code et de nos dépendances.

- **Anaconda :**
    - **Rôle :** Fortement recommandé. C'est une distribution Python qui inclut un gestionnaire de paquets (Conda) et un gestionnaire d'environnements virtuels. Anaconda est très complet et vient avec de nombreuses bibliothèques préinstallées.
    - **Avantages :** Facilite l'installation et la gestion de toutes les bibliothèques nécessaires, résout les problèmes de compatibilité et permet de créer des **environnements virtuels isolés** pour chaque projet. C'est essentiel pour éviter les conflits de versions entre les différents projets Python.
- **Jupyter Notebook / JupyterLab :**
    - **Rôle :** Des environnements de calcul interactifs basés sur le web. Ils permettent de combiner du code, des visualisations, du texte explicatif (en Markdown) et des équations dans un seul document. Idéal pour l'exploration de données, le prototypage, l'expérimentation rapide avec les modèles et la documentation de démarche pas à pas.
    - **Avantages :** Très visuel, parfait pour la phase d'exploration des données et la présentation incrémentale de vos résultats.

3.3 Définition de l'ensemble des données et des variables utilisés

Le dataset utilisé smart_grid_stability_augmented comprend 60 000 enregistrements et est destiné à la prévision de la stabilité des smart grids, il a été téléchargé depuis le lien suivant : https://www.kaggle.com/datasets/pcbreviglieri/smart-grid-stability

Il contient quatorze attributs tels que le taux de variation (t1,t2,t3,t4), la pression (p1,p2,p3,p4) et les gradients (g1,g2,g3,g4), ainsi que des étiquettes de stabilité (stab, stabf) classant l'état du réseau en "stable" ou "instable", servant de variable cible pour l'apprentissage.

Les variables dynamiques (t1–t4), les pressions (p1–p4) et les gradients (g1–g4) fournissent une vue complète du comportement du réseau.

Ce jeu de données est une version améliorée du jeu de données simulé sur la stabilité du réseau électrique, développé par Vadim Arzamasov (Institut de Technologie de Karlsruhe, Allemagne) et mis à disposition sur le répertoire UCI (Université de Californie à Irvine) Machine Learning, une plateforme de référence pour les chercheurs et étudiants en apprentissage automatique.

3.3.1 Description du dataset

Le dataset contient des variables liées aux caractéristiques d'un réseau électrique intelligent, notamment :

- tau1, tau2, tau3, tau4 : Temps de réaction du participant, tau1 correspond au producteur d'électricité, et tau2, tau3, tau4 aux consommateurs.
- p1, p2, p3, p4 : Mesures ou demandes de puissance en quatre points du réseau, Puissance nominale consommée (négative) ou produite (positive). P1 est une valeur composite (abs(p2 + p3 + p4)). Pour les consommateurs les valeurs sont généralement dans la plage [-0.5, -2].
- g1, g2, g3, g4 : Coefficient (gamma) proportionnel à l'élasticité prix correspond a la Capacités de production ou paramètres d'efficacité des générateurs à quatre emplacements.
- stab : Indicateur numérique de stabilité.
- stabf : Indicateur catégoriel de stabilité (stable ou unstable).

Variable	Type	Description	Intervale
tau1,tau2,tau3, tau4	flaot	Constantes de temps associées à différents composants du réseau.	Minimum: 0.5 Maximum: 10
P1(fourisseur) P2,P3,P4(consommateur)	flaot	Mesures ou demandes de puissance en quatre points du réseau.	P1: Minimum: 1.6 Maximum: 5.9
	flaot		P2,P3,P4 Minimum : -2 Maximum: -0.5
G1 ,G2 ,G3,G4	flaot	Capacités de production ou paramètres d'efficacité des générateurs à quatre emplacements.	Minimum: 0.05 Maximum: 1
STAB	flaot	Indicateur numérique de stabilité.	Minimum: -0.08 Maximum: 0.1
STABF	string	Indicateur catégoriel de stabilité (stable ou unstable).	stable ou unstable

Table 3.2- Définition des variables

3.3.2 Prétraitement des données du dataset

C'est le processus de nettoyage, de transformation et de préparation des données brutes pour les rendre adaptées et performantes pour l'entraînement d'un modèle d'apprentissage automatique. Voici les étapes clés à considérer :

- **Gestion des valeurs manquantes** : Le dataset « smart_grid_stability_augmented « est généralement propre et ne contient pas de valeurs manquantes**.**

```
data.isnull().sum()
```
✓ 0.0s

```
tau1     0
tau2     0
tau3     0
tau4     0
p1       0
p2       0
p3       0
p4       0
g1       0
g2       0
g3       0
g4       0
stab     0
stabf    0
dtype: int64
```

Figure 3.2 Gestion des valeurs manquantes

- **Encodage des variables catégorielles :** Le dataset contient une variable cible catégorielle : stabf (stable/unstable). Pour que les modèles d'apprentissage automatique puissent l'utiliser, elle doit être convertie en format numérique**.**

```
data['stabf'] = data['stabf'].map({'unstable': 0, 'stable': 1})
data.head()
```
✓ 0.0s

	tau1	tau2	tau3	tau4	p1	p2	p3	p4	g1	g2	g3	g4	stab	stabf
0	0.869615	5.390493	9.483204	7.521103	4.081518	-1.234632	-1.661369	-1.185516	0.772575	0.758536	0.193251	0.879593	0.021209	0
1	7.529498	8.333825	1.489076	9.640714	2.066343	-0.772024	-0.568704	-0.725614	0.571514	0.265756	0.902673	0.468411	0.004152	0
2	3.132813	2.712341	6.994670	6.887259	4.234673	-1.930731	-1.594576	-0.709367	0.188647	0.055684	0.301834	0.428633	-0.025340	1
3	2.700884	2.815071	3.798056	3.463444	3.143668	-0.739843	-1.334164	-1.069661	0.468997	0.594340	0.290011	0.557437	0.002921	0
4	7.561637	4.650711	8.933701	3.343445	4.231189	-1.674766	-1.152435	-1.403988	0.214005	0.119668	0.870198	0.521203	0.021623	0

Figure 3.3 Encodage des variables catégorielles

- **Normalisation :** La normalisation est une technique de mise à l'échelle des caractéristiques (feature scaling) utilisée dans le prétraitement des données. Son objectif principal est de redimensionner les valeurs des caractéristiques numériques dans une plage spécifique, généralement entre **0 et 1**.

  la méthode de normalisation utilisée est Z-score qui est une technique de prétraitement des données essentielle en science des données et en apprentissage automatique. Son objectif est de transformer les données numériques de manière à ce qu'elles aient une moyenne de 0 et un écart-type de 1.

```
# Normalization
X_norm = X.copy()
for column in X_norm.columns:
    X_norm[column] = (X_norm[column] - X_norm[column].mean()) / X_norm[column].std()
X_norm.head()
```

	tau1	tau2	tau3	tau4	p1	p2	p3	p4	g1	g2	g3	g4
0	0.756004	1.576772	0.288758	-1.714840	1.298135	-1.385842	-0.400816	-0.468137	1.567079	1.066100	1.451349	0.644285
1	1.220197	1.558265	1.406330	1.598513	-1.392774	0.055967	1.306706	1.056506	0.815391	0.898159	-0.860945	1.545457
2	0.099197	-1.134462	-1.320093	-1.688119	1.188741	-1.447238	-0.259396	-0.358150	0.746571	-0.330345	-1.087999	0.134403
3	1.029053	1.669101	0.596029	-1.300808	-2.312732	1.036117	1.263693	1.717289	1.593635	1.224469	0.517604	-1.608699
4	0.083502	1.155927	-0.952976	-1.080151	-1.188634	0.991917	0.752080	0.320599	0.259402	-0.051547	-0.468656	-1.642606

Figure 3.4 normalisation des données

- **Division des Données :** la division des données est une étape absolument essentielle dans le processus de Machine Learning. Elle consiste à séparer l'ensemble de données initial en deux sous-ensembles distincts : l'ensemble d'entraînement, l'ensemble de test.

	X	Y
Train(80%)	**(48000 , 12)**	**48000**
Test(20%)	**(12000 , 12)**	**12000**

Table 3.3 division des données train/test

3.4 Analyse exploratoire des données

L'Analyse Exploratoire des Données (AED), ou Exploratory Data Analysis (EDA) en anglais, est une étape fondamentale et itérative dans tout projet de science des données. C'est le processus d'investigation initiale des données pour découvrir des modèles, détecter des anomalies, tester des hypothèses et résumer leurs principales caractéristiques, souvent avec des méthodes de visualisation.

3.4.1 Analyse de la distribution des classes

La Figure 3.5 Illustre la distribution de l'ensemble de données entre les deux classes cibles, la stabilité et l'instabilité. Cette visualisation met en évidence l'équilibre entre les classes et fournit des informations sur la structure des données, ce qui est crucial pour l'évaluation des modèles d'apprentissage automatique

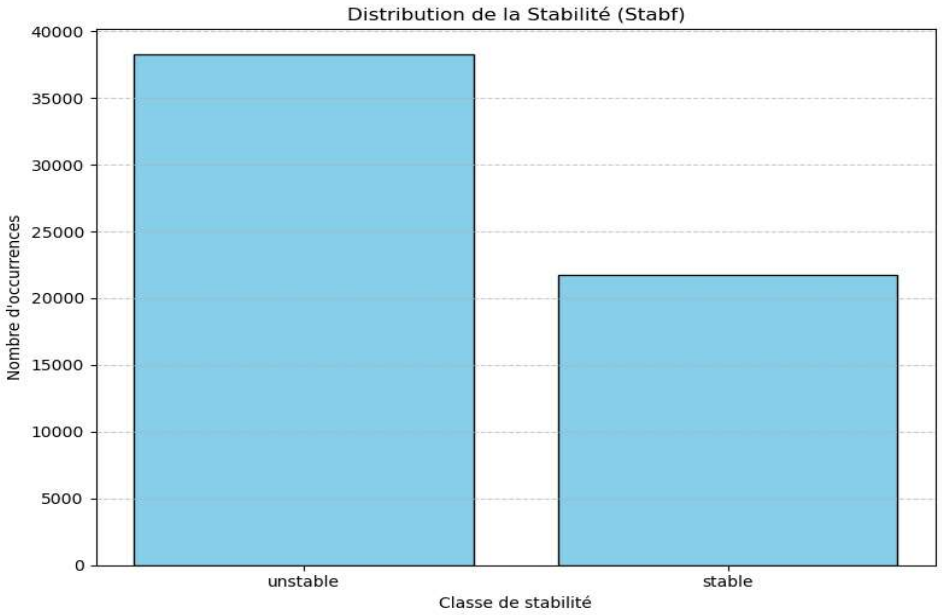


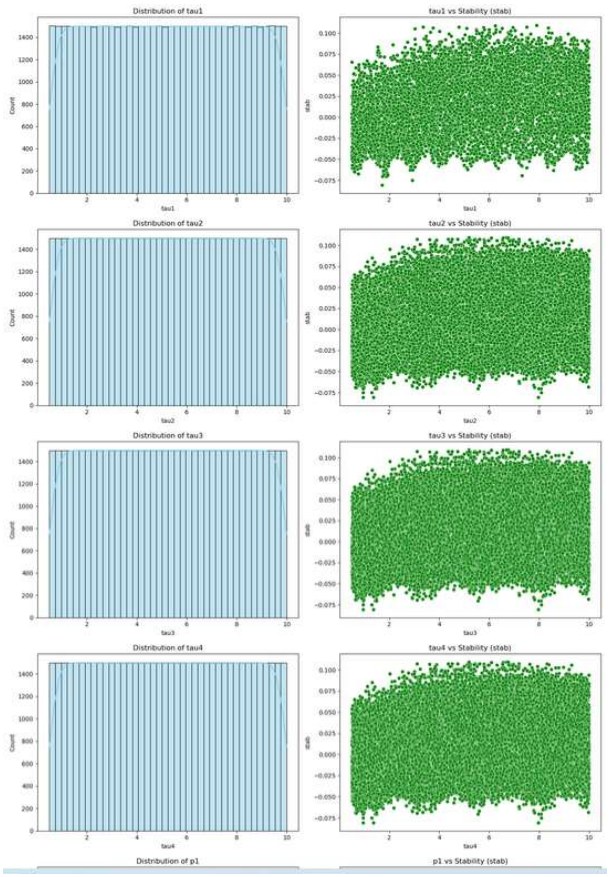
Figure 3.5 Distribution de la Stabilité (Stabf)

Classe	Nombre d'enregistrements	Pourcentage (%)
unstable (0)	38280	63.8 %
stable (1)	21720	36.2 %

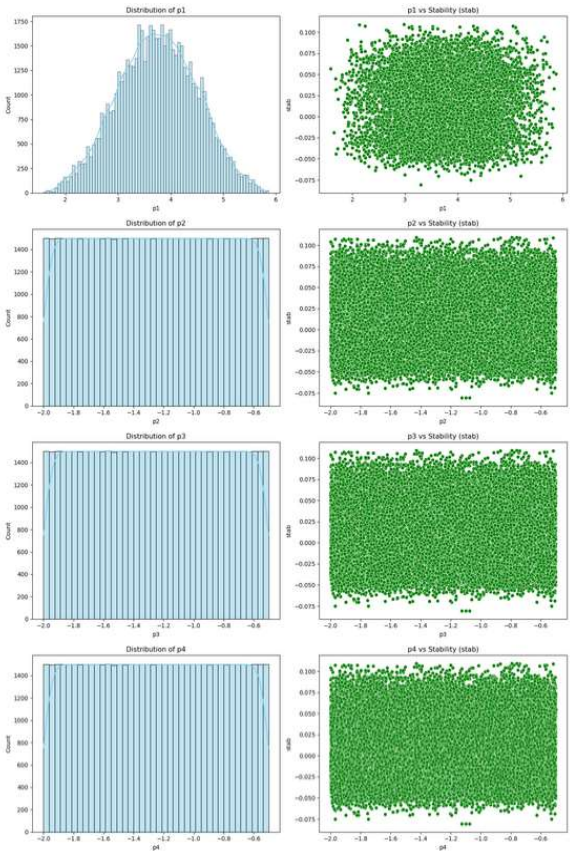
Table 3.4 distribution des classes

3.4.2 Analyse des distributions et des corrélations

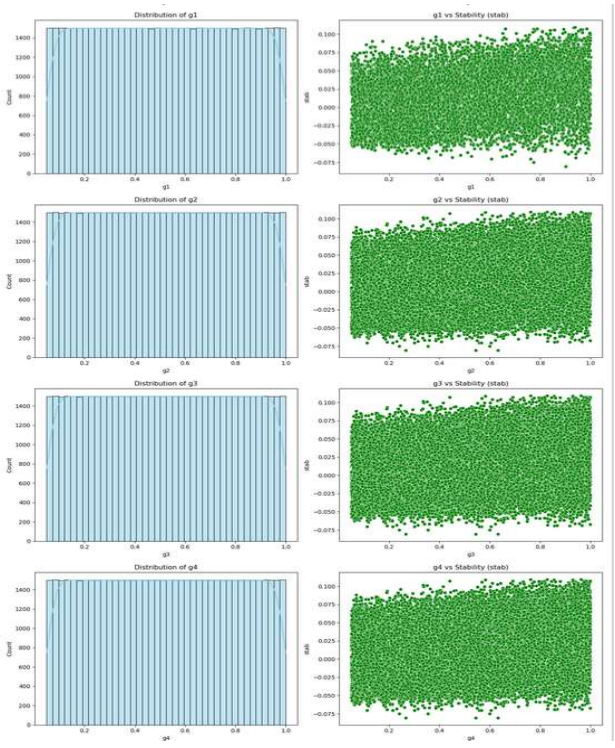
Une analyse approfondie des distributions et des corrélations est un pas essentiel et fondamental dans la démarche exploratoire des données (EDA – Exploratory Data Analysis) qui va nous permettre d'appréhender la structure de nos données, de découvrir des tendances cachées, d'identifier les relations entre les variables, de repérer des difficultés, afin de mieux construire un modèle.



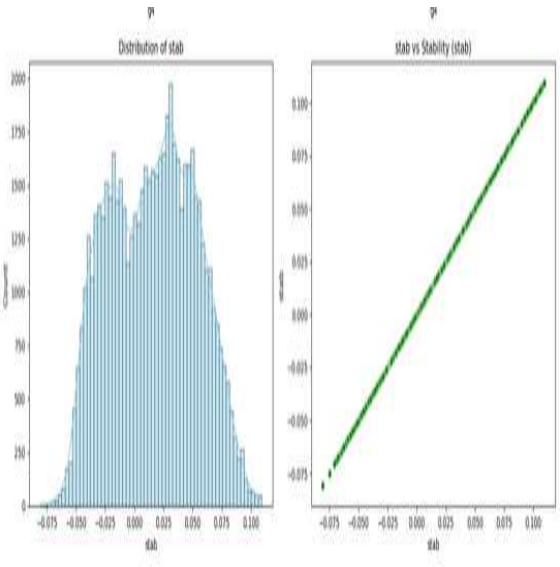
Correlation tau1..tau4 avec stab



Correlation p1..p4 avec stab



Correlation g1..g4 avec stab



Correlation stab avec stab

Figure 3.6 Distribution et corrélations des caractéristiques (t,p,g,stab)

3.4.3 Calcule de la matrice de corrélation en excluant « stabf »

La matrice de corrélation est un outil statistique et visuel fondamental en analyse de données. C'est un tableau carré qui affiche les coefficients de corrélation entre toutes les paires de variables numériques au sein du dataset. Elle nous donne un aperçu rapide et compréhensible de la force et de la direction des relations linéaires entre nos différentes caractéristiques.

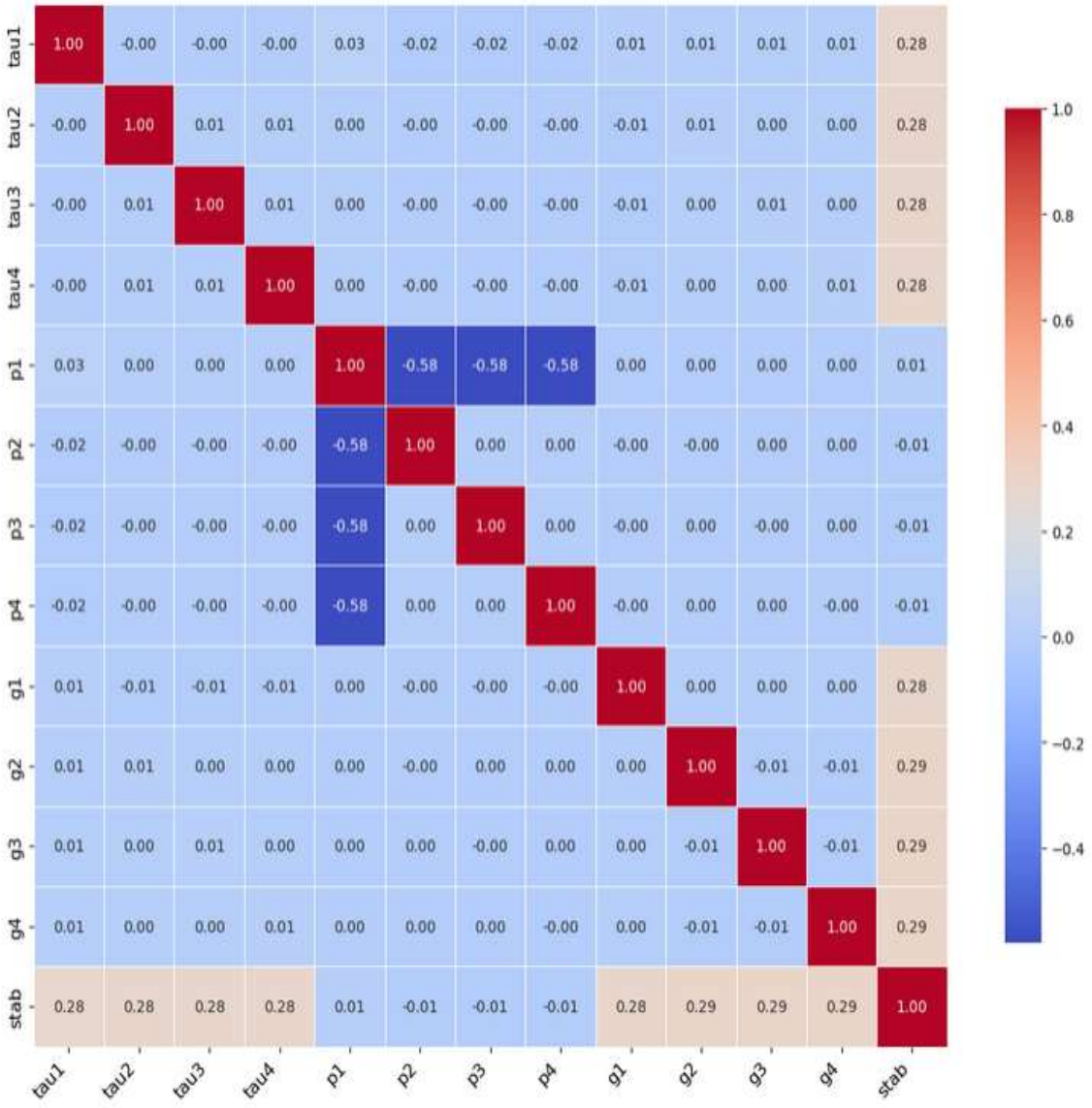


Figure 3.7 Matrice de corrélation des caractéristiques

➢ **Corrélation avec stab :**

- Les variables tau1, tau2, tau3 et tau4 présentent une corrélation positive modérée avec stab (environ 0,28). Cela suggère que le temps de réaction (tau) est quelque peu lié à la stabilité du système.

- Caractéristiques liées à la puissance (p1 à p4) :Il existe une forte corrélation négative (-0,58) entre la puissance du fournisseur (p1) et les puissances des consommateurs (p2, p3 et p4). Cela est logique car p1 représente la puissance fournie, qui doit être égale en valeur absolue mais de signe opposé à la somme des puissances consommées.
- Les caractéristiques liées à la puissance (p1 à p4) ne montrent pas de forte corrélation avec stab, ce qui suggère que les valeurs nominales de puissance n'ont pas une influence significative sur la stabilité du système.

Corrélation des élasticités-prix (g1 à g4) :

- Les coefficients d'élasticité-prix (g1, g2, g3, g4) présentent de faibles corrélations entre eux, ce qui indique que l'élasticité de chaque consommateur se comporte de manière relativement indépendante des autres.
- Ces caractéristiques montrent aussi une corrélation positive modérée (~0,29) avec stab, ce qui suggère que l'élasticité-prix pourrait jouer un certain rôle dans la stabilité du système.

Tendance générale :

- La plupart des caractéristiques ne présentent pas de forte corrélation entre elles (à l'exception des relations de puissance attendues), ce qui indique que le jeu de données contient une diversité de variables avec des comportements relativement indépendants.
- La variable stab a sa plus forte corrélation avec le temps de réaction (tau) et l'élasticité-prix (g), ce qui désigne ces deux groupes de variables comme des facteurs clés influençant la stabilité du système.

3.5. Entraînement et validation des Modèles Machine Learning ML

L'entraînement des modèles de Machine Learning est l'étape cruciale où un algorithme apprend à identifier des motifs et des relations dans les données. Le but est de lui permettre de faire des prédictions ou de prendre des décisions sur de nouvelles données jamais vues auparavant. Il s'agit de l'une des sections les plus importantes. On va testé de nombreux modèles d'apprentissage automatique, notamment des modèles traditionnels comme la régression logistique et naïve Bayes, ainsi que des techniques avancées comme SVM, Random Forest et XGBoost. ensuite on présente leurs résultats afin de comparer leurs performances ultérieurement.

3.5.1 Régression Logistique :

Ce modèle de régression logistique a été choisi pour sa simplicité et sa facilité d'interprétation ; nous souhaitions donc en faire notre point de départ. Afin d'améliorer les performances du modèle dans le traitement de notre ensemble de données, ces hyperparamètres ont été ajustés comme suit :(random_state: 42, max_iter: 1000).

La figure 3.8 montre la matrice de confusion et les autres métriques obtenues après avoir entrainer le modèle régression logistique, (1 représente stable, 0 représente unstable)

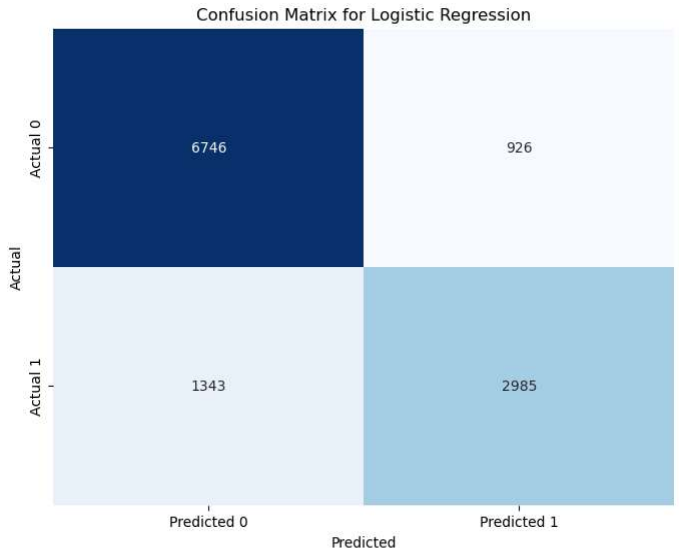


Figure 3.8 Matrice de confusion du model régression logistique

Accuracy= 81.09%

Class	Precision	Recall	F1-Score	Support
0	0.83	0.88	0.86	7588
1	0.76	0.69	0.72	4412

Table 3.5– Classification report de régression logistique

- régression logistique présente le plus grand nombre de faux positifs (901) pour la Classe 0 et le plus grand nombre de faux négatifs (1275) pour la Classe 1, ce qui impacte fortement ses scores de précision et de rappel pour les deux classes.

3.5.2 Naïf Bayes

Le classifieur Naive Bayes est une famille d'algorithmes de classification basée sur le théorème de Bayes avec une "hypothèse de naïveté" forte. La figure 3.9 montre la matrice de confusion et les autres métriques obtenues après avoir entrainer le modèle Naive Bayes, on a implémenté l'algorithme Naive Bayes avec distribution gaussienne (GaussianNB).

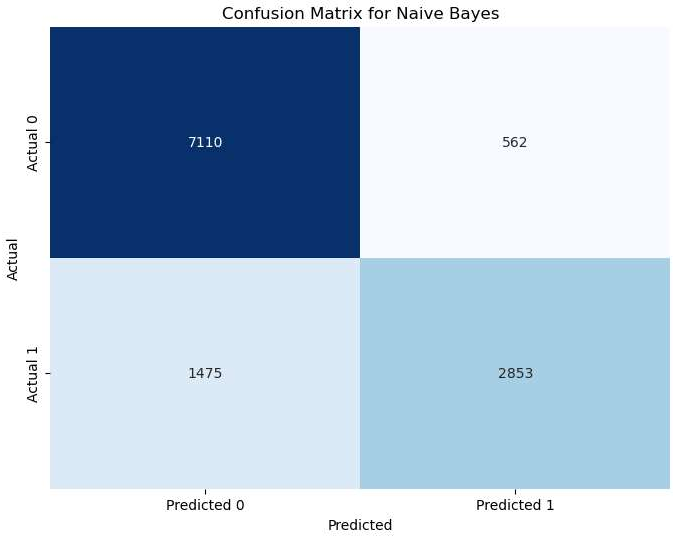


Figure 3.9 Matrice de confusion du model Naive Bayes

Accuracy :82.98

Class	Precision	Recall	F1-Score	Support
0	0.83	0.93	0.87	7588
1	0.84	0.66	0.74	4412

Table 3.6– Classification report de Naive Bayes

- le model Naive Bayes Le faible rappel pour la Classe 1 (0,66) est un problème majeur, signifiant qu'il manque un grand nombre d'instances réelles de la Classe 1 (1475 faux négatifs). Sa capacité à discriminer les classes est donc moins bonne.

3.5.3 Support Vector Machine (SVM)

Le model SVM est un algorithme puissant et polyvalent utilisés pour la classification et la régression, mais surtout connus pour la classification, ces hyperparamètres ont été ajustés comme suit :(kernel='rbf',random_state: 42), Création du modèle SVM avec un noyau RBF (Radial Basis Function). La figure 3.10 montre la matrice de confusion et les autres métriques obtenues après avoir entrainer le modèle SVM .

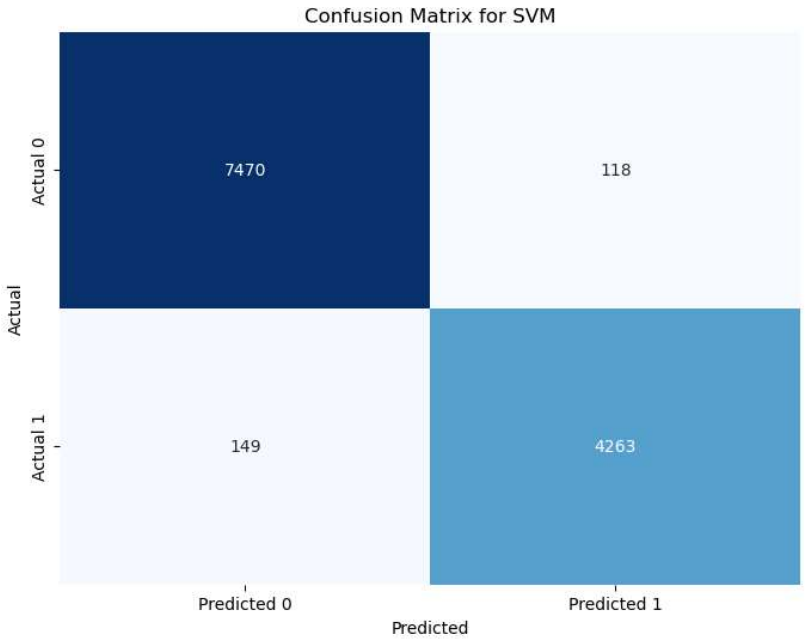


Figure 3.10 Matrice de confusion du model SVM

Accuracy :97.78

Class	Precision	Recall	F1-Score	Support
0	0.97	0.98	0.98	7588
1	0.97	0.97	0.97	4412

Table 3.7– Classification report SVM

- SVM montre une très bonne performance globale Sa précision (0.97)est supérieur à celle des Naive Bayes et RL, Sa capacité à discriminer les classes est donc très bonne avec (118) faux positifs et(149) faux négatifs seulement .

3.5.4 Random Forest

Random Forest est un algorithme d'apprentissage ensembliste (ensemble learning) très populaire pour la classification et la régression. Il est basé sur la combinaison de plusieurs arbres de décision, ces hyperparamètres ont été ajustés comme suit :(random_state= 42, n_estimators=100), La figure 3.11 montre la matrice de confusion et les autres métriques obtenues après avoir entrainer le modèle Random Forest.

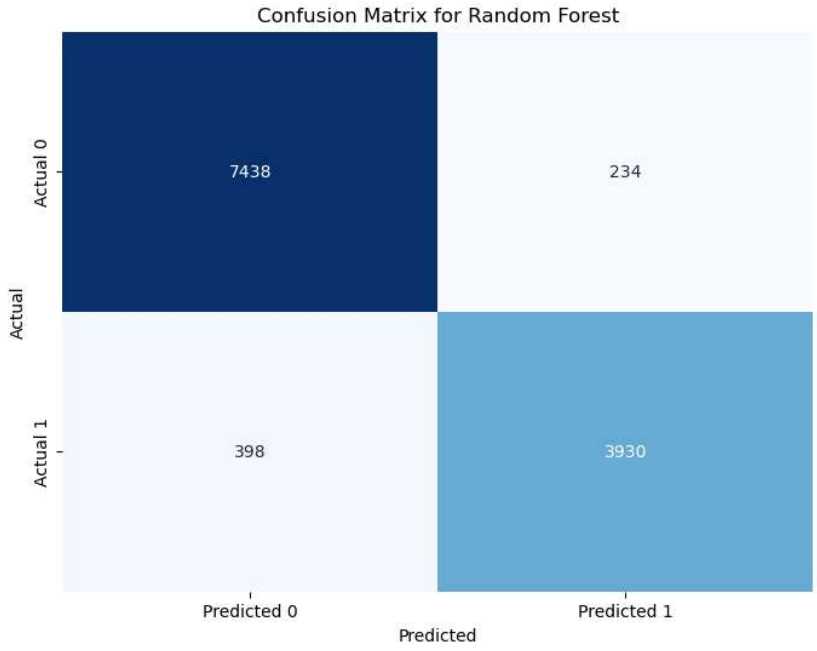


Figure 3.11 Matrice de confusion du model Random Forest

Accuracy : 94.73

Class	Precision	Recall	F1-Score	Support
0	0.95	0.97	0.96	7672
1	0.94	0.91	0.93	4328

Table 3.8– Classification report Random Forest

- Random Forest présente plus de classifications erronées (234 faux positifs et 398 faux négatifs), ce qui se reflète dans ses scores de précision et de rappel légèrement inférieurs, en particulier pour la Classe 1.

3.5.5 XGBoost :

- XGBoost est une implémentation optimisée et hautement efficace de l'algorithme de Gradient Boosting, également un algorithme d'apprentissage ensembliste populaire pour la classification et la régression, Nous avons choisi ce modèle car il est capable de gérer des données déséquilibrées puisque nous avons un déséquilibre entre les classe stable et ceux qui ne le sont pas. Pour plus de détails sur ce modèle, veuillez vous référer au chapitre 2. Et afin d'améliorer les performances du modèle dans la gestion de notre ensemble de données, ces hyperparamètres ont été ajustés après optimisation par l'algorithme RandomizedSearchCV on a trouvé que les meilleurs paramètres sont :
- **n_estimators** (nombres d'arbre): **500**
- **learning_rate** (taux d'apprentissage) : **0,2**
- **max_depth** (profondeur maximale des arbres):**8**
- **subsample** (fraction des données d'entraînement utilisée pour chaque arbre) : **0,7**
- **colsample_bytree** (fraction des colonnes utilisées pour chaque arbre) :**1,0**
- **gamma** (seuil de réduction de la perte pour permettre un split): **0,1**

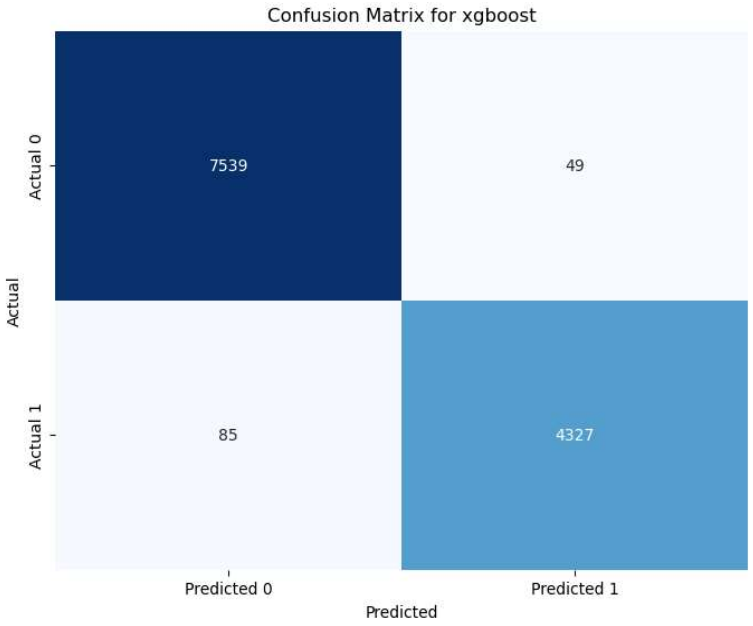


Figure 3.12 Matrice de confusion du model XGBoost

Accuracy : 98.88

Class	Precision	Recall	F1-Score	Support
0	0.99	0.99	0.99	7588
1	0.99	0.98	0.98	4412

Table 3.9– Classification report XGBoost

- XGBoost démontre une performance exceptionnelle et équilibrée sur les deux classes, confirmée par la précision la plus élevée, Les très faibles nombres de faux positifs (49) et de faux négatifs (85) pour la Classe 0 et la Classe 1 signifie que le modèle est capable de distinguer les classes positives et négatives .

3.6 Analyse et discussion des résultats:

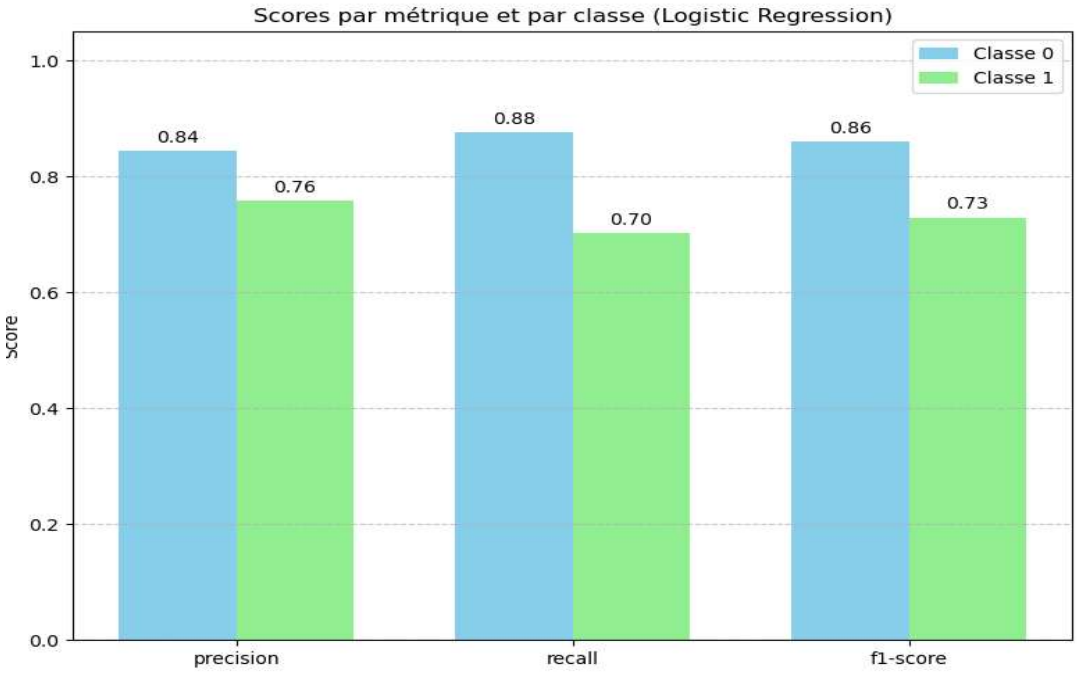
Cette étude sur la prédiction de la stabilité des smart grids s'appuie sur cinq algorithmes de machine learning :Regression Logistique , Support Vector Machine (SVM), Naive Bayes , Random Forest et extrême gradient boosting (XGBOOST) .L'évaluation a été réalisée à l'aide d'indicateurs classiques de classification, notamment l'Accuracy, précision, Recallv,F1-score et l'aire sous la courbe ROC (AUC), ainsi que les matrices de confusion. En général, pour la détection d'instabilité dans les Smart Grids, on observe souvent la hiérarchie de performance suivante de la plus simple à la plus complexe/performante.

	Accuracy (%)	Precision (Classe 0)	Rappel (Classe 0)	F1-score (Classe 0)	Precision (Classe 1)	Rappel (Classe 1)	F1-score (Classe 1)
Regression Logistique	**81.27**	**0.84**	**0.87**	**0.85**	**0.75**	**0.69**	**0.72**
Naive Bayes	**83.16**	**0.83**	**0.91**	**0.87**	**0.82**	**0.67**	**0.73**
SVM	**97.59**	**0.98**	**0.98**	**0.96**	**0.96**	**0.96**	**0.97**
Random Forest	**95**	**0.94**	**0.97**	**0.96**	**0.95**	**0.90**	**0.92**
XGBOOST	**98.65**	**0.99**	**0.99**	**0.99**	**0.98**	**0.97**	**0.98**

Table 3.10 métriques de performance des cinq Algorithmes

1 **Régression Logistique** (accuracy=82 %)

- Précision (Classe 0) : 0,84
- Précision (Classe 1) : 0,76
- Rappel (Classe 0) : 0,88
- Rappel (Classe 1) : 0,70
- Score F1 (Classe 0) : 0,86
- Score F1 (Classe 1) : 0,73



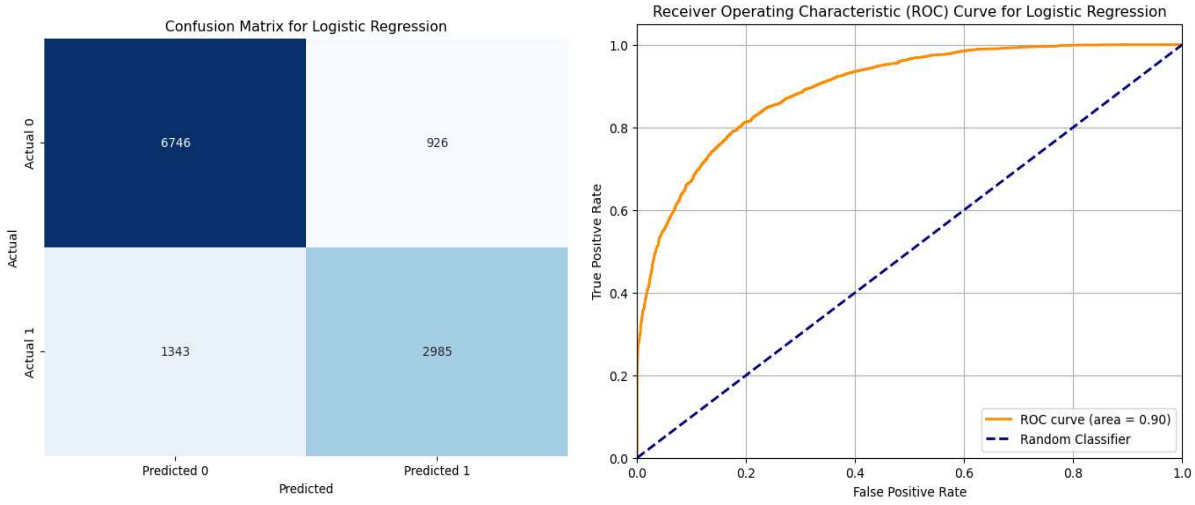


Figure 3.13 Rrésultat des performances Régression Logistique

- **Interprétation :** La Régression Logistique est le modèle le moins performant de cette comparaison, avec la précision la plus basse et un AUC ROC de 0,90. présente le plus grand nombre de faux positifs (901) pour la Classe 0 et le plus grand nombre de faux négatifs (1275) pour la Classe 1, ce qui impacte fortement ses scores de précision et de rappel pour les deux classes.
- **Conclusion :** La Régression Logistique est la moins efficace pour cette tâche, offrant une capacité de discrimination avec plus d'erreurs de classification.

2. **Naive Bayes (accuracy=84 %)**

- Précision (Classe 0) : 0,84
- Précision (Classe 1) : 0,82
- Rappel (Classe 0) : 0,92
- Rappel (Classe 1) : 0,67
- Score F1 (Classe 0) : 0,88
- Score F1 (Classe 1) : 0,74

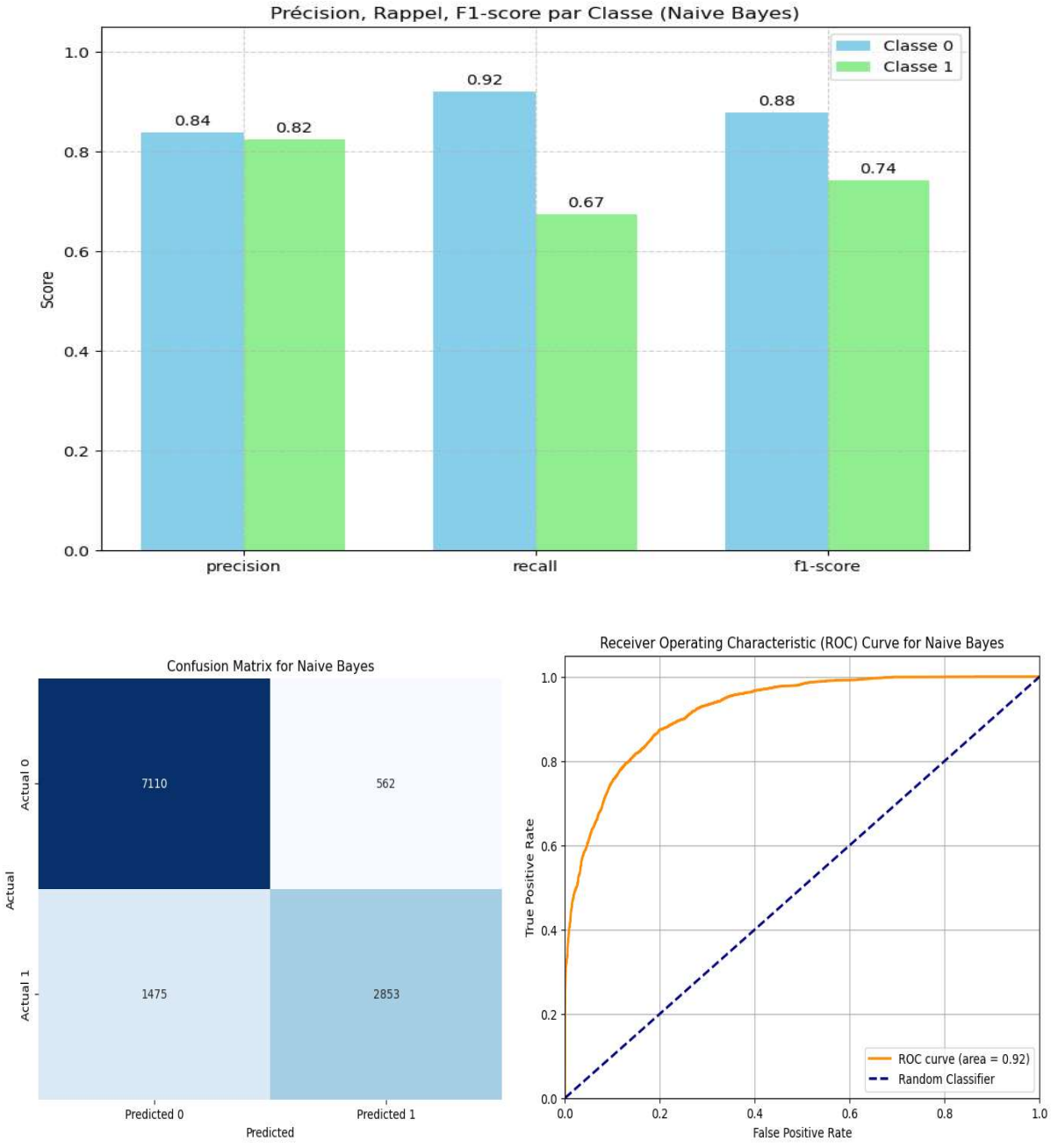


Figure 3.14 Résultat des performances Naïve Bayes

- **Interprétation :** Naive Bayes est nettement moins performant que les méthodes d'ensemble et le SVM, comme l'indiquent sa précision plus faible et son AUC ROC de 0,92. Le faible rappel pour la Classe 1 (0,673646) est un problème majeur, signifiant qu'il manque un grand nombre d'instances réelles de la Classe 1 (1433 faux négatifs). Sa capacité à discriminer les classes est donc moins bonne que celle des meilleurs modèles.
- **Conclusion :** Naive Bayes est moins adapté si une identification précise de la Classe 1 est critique, et sa capacité de discrimination est modérée par rapport aux autres.

3. **Random Forest** (accuracy=95 %)

- Précision (Classe 0) : 0,95
- Précision (Classe 1) : 0,94
- Rappel (Classe 0) : 0,97
- Rappel (Classe 1) : 0,91
- Score F1 (Classe 0) : 0,96
- Score F1 (Classe 1) **:** 0,92

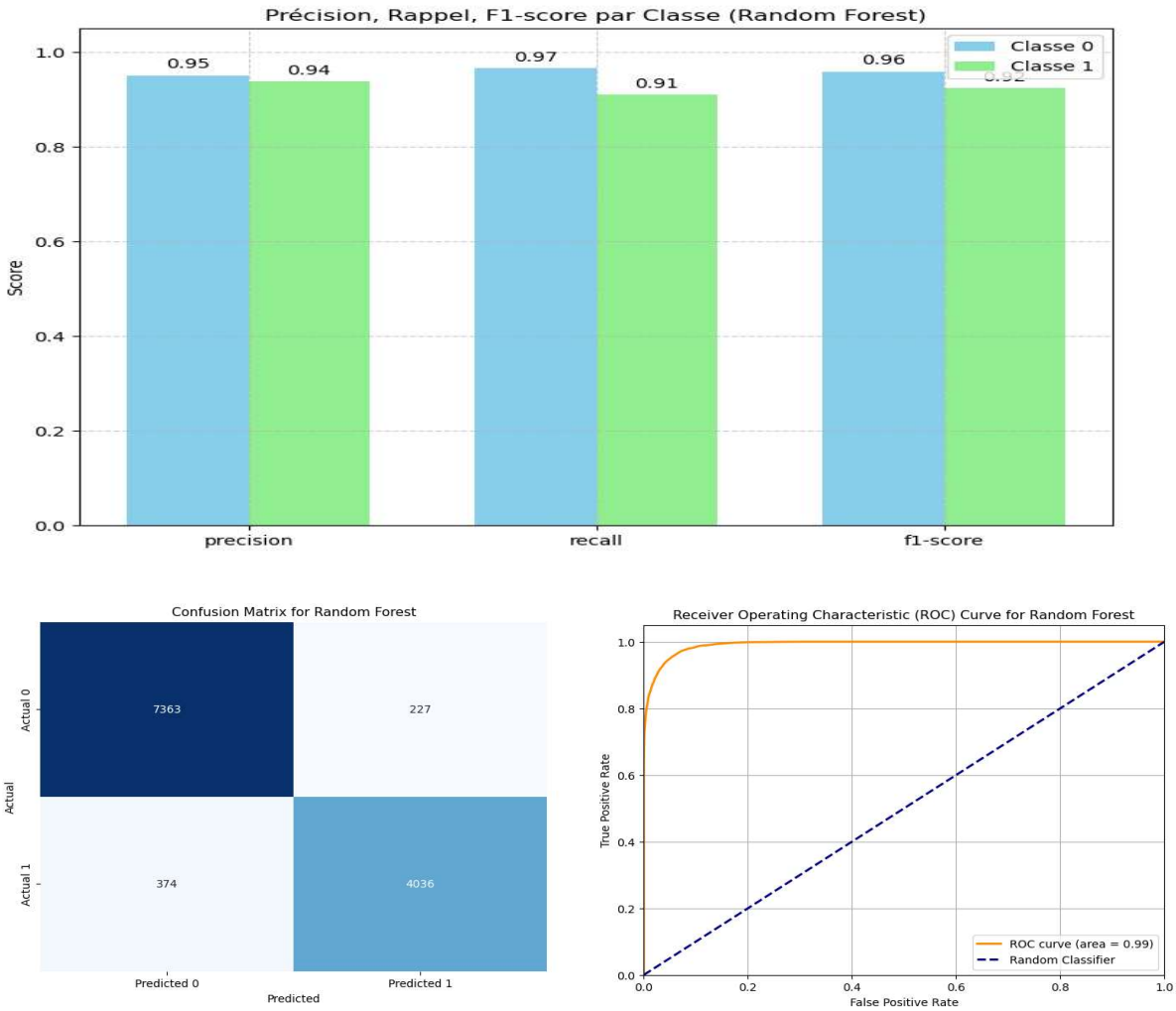


Figure 3.15 Rrésultat des performances Random Forest

- **Interprétation :** Random Forest montre une très bonne performance globale avec un AUC ROC de 0,99, ce qui est excellent et indique une forte capacité à distinguer les classes. Cependant, il présente plus de classifications erronées (227 faux positifs et 374 faux négatifs) que SVM et XGBoost, ce qui se reflète dans ses scores de précision et de rappel légèrement inférieurs, en particulier pour la Classe 1.
- **Conclusion :** Random Forest est un modèle robuste avec une excellente capacité de classification, bien qu'un peu moins précis que les deux modèles de tête.

4. SVM (accuracy=97.98%)

- Précision (Classe 0) : 0,98
- Précision (Classe 1) : 0,98
- Rappel (Classe 0) : 0,99
- Rappel (Classe 1) : 0,97
- Score F1 (Classe 0) : 0,98
- Score F1 (Classe 1) : 0,97

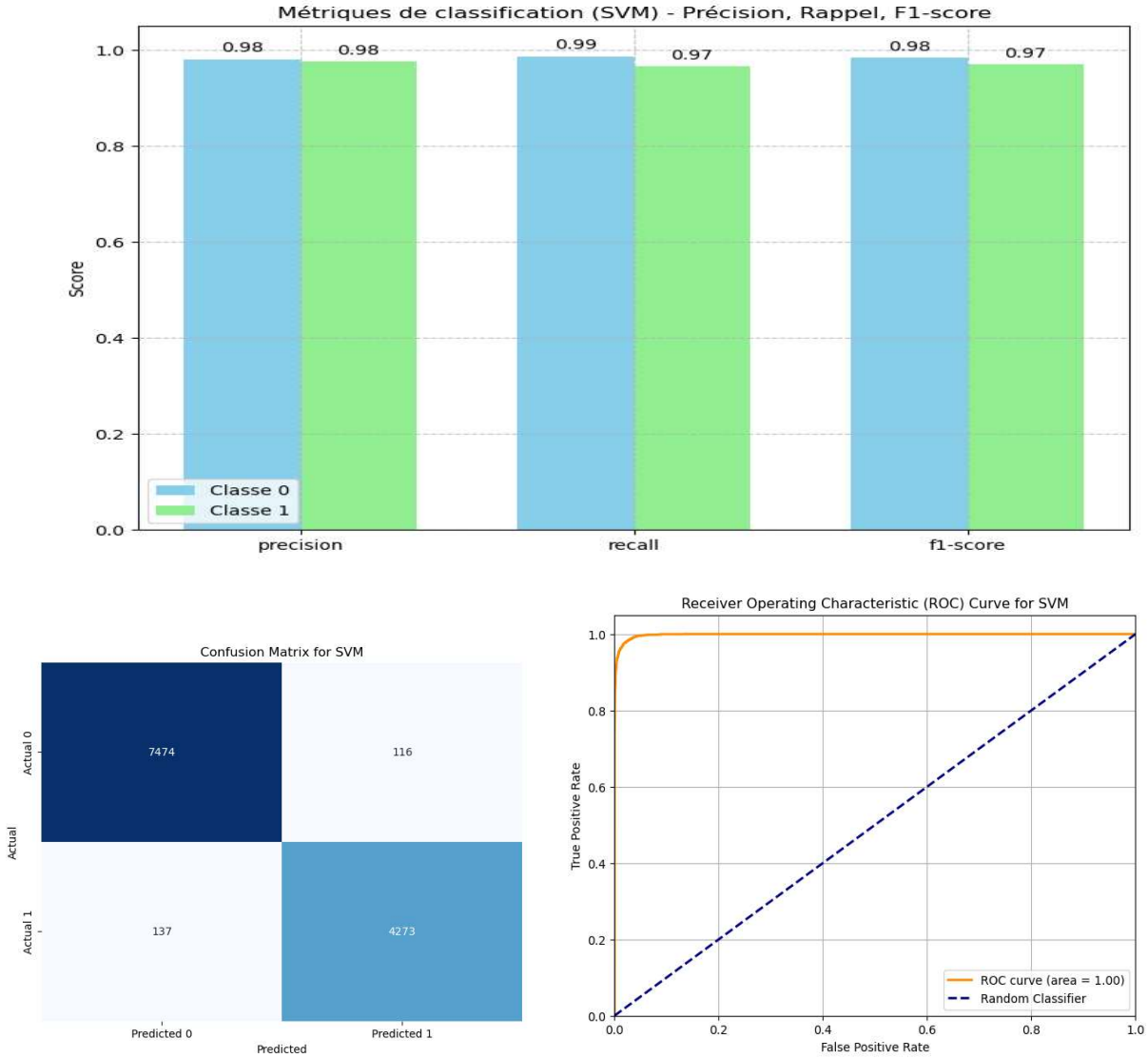


Figure 3.16 Résultat des performances SVM

- **Interprétation :** Le SVM à un AUC ROC parfait de 1,00, indiquant également une excellente capacité de discrimination. Sa précision est légèrement inférieure à celle de XGBoost, principalement due à un nombre légèrement plus élevé de faux positifs (116) et faux négatifs (137) par rapport à XGBoost. Le rappel pour la Classe 1 (0,97) est légèrement inférieur à celui de la Classe 0.
- **Conclusion :** Le SVM est un concurrent extrêmement fort, offrant une capacité de distinction des classes aussi bonne que XGBoost.

5. **XGBoost (accuracy=98.76%)**
   - Précision (Classe 0) : 0,99
   - Précision (Classe 1) : 0,98
   - Rappel (Classe 0) : 0,99
   - Rappel (Classe 1) : 0,98
   - Score F1 (Classe 0) : 0,99
   - Score F1 (Classe 1) : 0,98

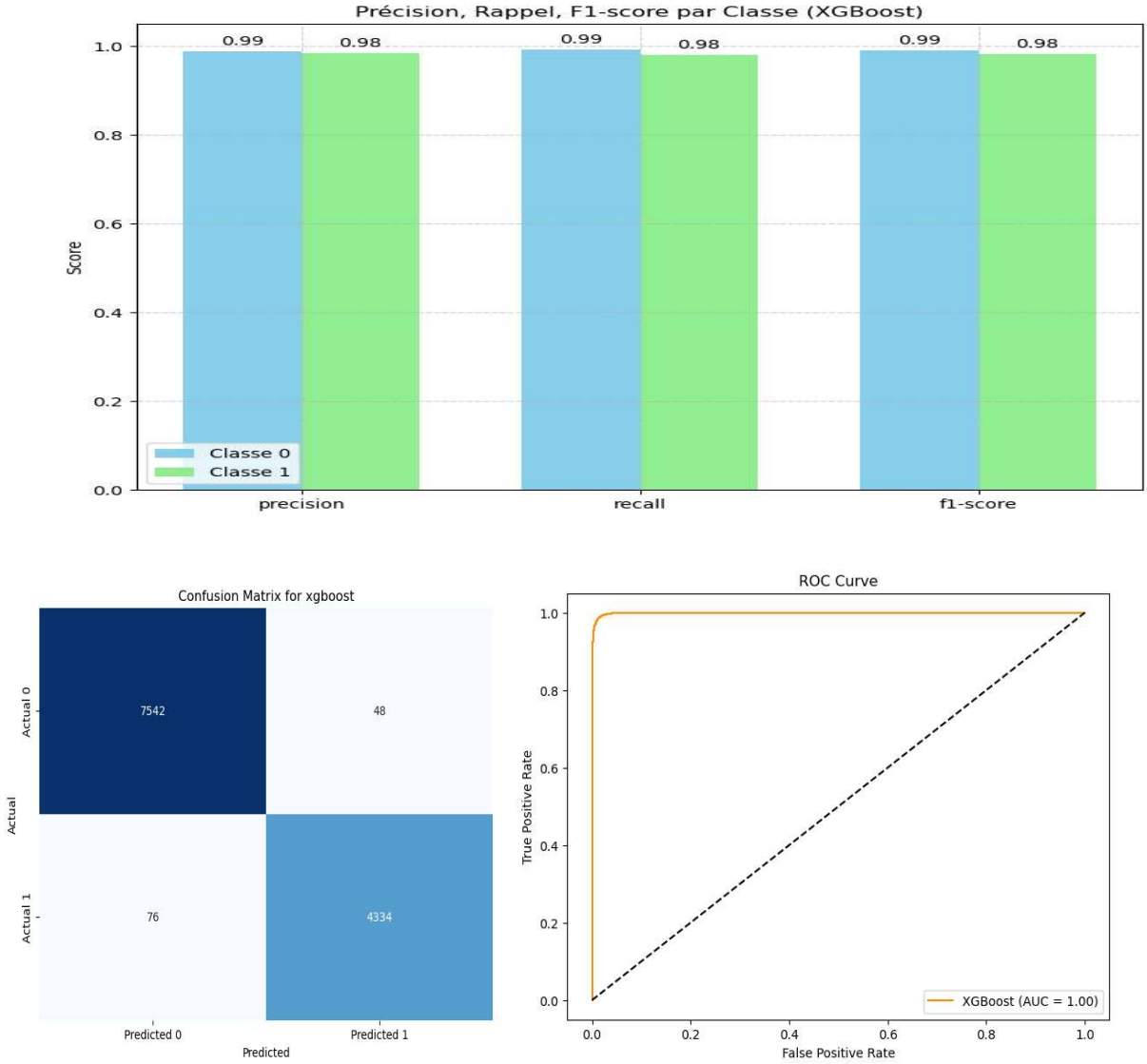


Figure 3.17 Rrésultat des performances XGBOOST

- **Interprétation :** XGBoost démontre une performance exceptionnelle et équilibrée sur les deux classes, confirmée par la précision la plus élevée et un AUC ROC parfait de 1,00. Cela signifie que le modèle est capable de distinguer les classes positives et négatives avec une très grande fiabilité. Les très faibles nombres de faux positifs (48) et de faux négatifs (76) pour la Classe 0 et la Classe 1 soutiennent cette conclusion.
- **Conclusion :** XGBoost est le modèle le plus performant, montrant une généralisation excellente et un pouvoir discriminant maximal.

3.7 Résultat final du model prédictif :

- **XGBoost et SVM sont les leaders incontestés :** Avec une AUC ROC de 1,00 et des précisions très élevées, ces deux modèles sont les meilleurs choix si l'objectif est d'obtenir la performance prédictive la plus élevée et la meilleure capacité de discrimination entre les classes. Leurs matrices de confusion montrent des erreurs de classification minimes.
- **Random Forest est une excellente alternative :** Son AUC ROC de 0,99 confirme qu'il est également très performant en termes de discrimination, bien que ses autres métriques (précision, rappel, F1-score) soient légèrement inférieures à celles de XGBoost et SVM.
- **Naive Bayes et Régression Logistique sont nettement moins efficaces :** Leurs AUC ROC de 0,89, bien que respectables, indiquent une capacité de discrimination inférieure par rapport aux modèles plus complexes. Ils sont moins adaptés si une haute précision pour les deux classes ou une capacité de distinction parfaite est requise. Ils semblent également plus sensibles à un potentiel déséquilibre des classes, sous-performant en particulier sur la Classe 1 (comme le montre le rappel plus faible pour cette classe).

En résumé : XGBoost est le modèle le plus adapté à cette tâche, en raison de ses excellentes performances en termes de précision, rappel, exactitude et robustesse globale. Il gère efficacement le compromis entre sous-ajustement et surajustement, ce qui le rend particulièrement adapté aux jeux de données complexes

3.8 Applications aux Smart Grids

- **Renforcement de la résilience** : Les modèles prédictifs basés sur l'apprentissage automatique peuvent rapidement identifier les conditions menant à l'instabilité. Par exemple, des prédictions précises d'instabilité peuvent alerter les opérateurs pour ajuster les paramètres dynamiques (τ, p, g) en temps réel.
- **Optimisation de la gestion** : La mise en œuvre d'algorithmes, en particulier XGBoost et SVM, peut améliorer les systèmes de contrôle automatisés, réduisant ainsi le risque de perturbations imprévues.
- **Scalabilité**: Les algorithmes comme XGBoost et SVM,sont bien adaptés pour inclure des sources de données supplémentaires (comme les conditions météorologiques et la consommation énergétique), renforçant leur utilité pratique pour les réseaux complexes.

3.9. Conclusion

Dans ce chapitre on a présenté les différents étapes de prétraitement des données tels que l'exploration et la visualisation des données, et on a montré l'efficacité des techniques de machine learning pour prédire la stabilité des smart grids à partir de paramètres dynamiques temps (t), pression(p) et gradient (g) et de labels de stabilité du dataset « smart_grid_stability_augmented ». En exploitant des modèles comme XGBOOST et SVM, nous pouvons non seulement prédire avec précision les instabilités, mais aussi optimiser la gestion en temps réel des réseaux électriques.

Cependant, pour maximiser l'impact, il est essentiel d'intégrer des facteurs externes et d'explorer des architectures plus avancées.

Les résultats obtenus ouvrent la voie à des applications concrètes dans la gestion dynamique des réseaux et l'intégration des énergies renouvelables, renforçant ainsi la résilience et l'efficacité des infrastructures électriques du futur.

Conclusion générale

Cette étude a démontré l'efficacité des techniques d'apprentissage automatique pour prédire l'instabilité des réseaux intelligents en utilisant un ensemble de données comprenant des paramètres dynamiques clés tels que τ (taux de variation), p (pression), g (gradients) et des étiquettes de stabilité (stab, stabf). Les résultats mettent en évidence une précision prédictive élevée, en particulier pour les XGBOOST (98.66%) et SVM (97.59%), confirmant leur fiabilité pour évaluer la stabilité des réseaux électriques. Ces performances sont renforcées par des métriques d'évaluation avancées, avec des valeurs AUC atteignant 1,00 pour XGBOOST et les SVM, soulignant leur robustesse. Random Forest a aussi une performance pas mal avec une précision de (94.64 %), les algorithmes plus simples, tels que logistique régression et Naïve Bayes, bien qu'efficaces, affichent des performances légèrement inférieures, en particulier en termes de précision et de rappel, soulignant l'importance de choisir des modèles adaptés à la complexité des données. Au-delà de la classification précise, ces modèles offrent une base solide pour la gestion proactive des perturbations dans les réseaux intelligents, facilitant l'anticipation et l'atténuation des risques d'instabilité.

Les travaux futurs se concentreront sur l'enrichissement de l'ensemble de données en intégrant d'autres facteurs tels que les conditions météorologiques, les modèles de consommation et les variables socio-économiques, ce qui pourrait améliorer la précision des prédictions et élargir l'applicabilité des modèles. De plus, l'intégration de ces modèles dans des systèmes de contrôle en temps réel permettrait un ajustement dynamique des paramètres du réseau, contribuant à la prévention des instabilités. Une exploration plus approfondie des architectures d'apprentissage profond, telles que les réseaux neuronaux récurrents (RNN) et les transformateurs, pourrait améliorer les performances, en particulier pour gérer les dépendances temporelles complexes.

Pour améliorer le modèle dans le futur, divers éléments supplémentaires pour aider au calcul seraient intégrés comme des données liées aux tendances météo, aux consommations des clients ou des statistiques réseaux en temps réels. Plusieurs architectures de deep learning plus sophistiquées, et méthodes d'optimisations, pourraient donne de nouveaux modèles plus précis et plus efficaces à la prédiction de la consommation d'énergie. L'ajout d'outils techniques émergents comme l'Internet des objets (IoT) et la blockchain pourrait encore améliorer la gestion et l'optimisation de l'énergie en temps réel du modèle

Bibliographie

[1] Q. Biyu, Y. Lin, M. Jiayi, L. Tianqi, Z. Xiaotian, and W. Youyin, “Analysis of Influence with Connected Wind Farm Power Changing and Improvement Strategies on Grid Voltage Stability,” *Proc. China Int. Conf. Electr. Distrib.*, Tianjin, pp. 2029–2033, 2018.

[2] G. Pellerin, “Initiation au Machine Learning avec Python,” *Makina Corpus*, 2017. [En ligne]. Disponible : https://makina-corpus.com/blog/metier/2017/initiation-au-machine-learning-avec-python-pratique

[3] H. Bahnhofstrasse, *Smart Grid, document connaissances de base*, Association des entreprises électriques, Suisse, févr. 2014.

[4] D. J. Thorp, A. G. Phadke, and S. H. Horowitz, “The Role of Phasor Measurement Units in Wide Area Monitoring Systems,” *IEEE Power Eng. Soc. Winter Meeting*, vol. 2, pp. 872–876, 2001. doi: 10.1109/PESW.2001.917239.

[5] L. Gyugyi, “Solid State Control of Electric Power in Flexible AC Transmission Systems,” *Int. Symp. Electr. Energy Conversion in Power Systems*, Italy, 1989.

[6] A. N. Auteur, *Architecture des Réseaux de Distribution du Futur en Présence de Production Décentralisée*, Thèse de doctorat, Institut National Polytechnique de Grenoble, déc. 2009.

[7] S. M. Amin and P. F. Schewe, “What is the Smart Grid?,” *IEEE Smart Grid News*. [En ligne]. Disponible : https://smartgrid.ieee.org/news/what-is-the-smart-grid

[8] G. Leboyer, A. Girodet, and J.-C. V. Bounezou, *Systèmes d’énergie électrique, guide de référence, les postes THT/HT*, ELEC Int. Symp., 1998.

[9] J. D. Glover and E. Morel, *Charte Smart Grid Côte d’Azur : Solutions pour l’Aménagement d’un Écoquartier Innovant*, Chambre de Commerce et d’Industrie Nice Côte d’Azur, déc. 2012.

[10] S. Bouvier and P. Strubel, *Déployer un réseau plus intelligent grâce à des solutions et services de câblage*, Livre blanc, Réseaux d’énergie intelligents (Smart Grids).

[11] Auteur anonyme, *Mémoire de fin d’études en vue de l’obtention du diplôme de Master*, Université Ibn Khaldoun Biskra, Thème : Conduite des réseaux électriques dans les smart grids.

[12] Étude d’Alcimed sur les Smart Grids, *Enerzine, Batirama, Le Monde de l’Énergie*, avr. 2011.

[13] A. Friconneau, “Réseaux électriques intelligents,” *DSpace, Université de Lorraine*, [En ligne]. Disponible : https://dumas.ccsd.cnrs.fr/dumas-02175125/document. [Consulté le 18 juin 2025].

[14] A. Friconneau, “Réseaux électriques intelligents,” *DSpace, Université de Lorraine*, [En ligne]. Disponible : https://dumas.ccsd.cnrs.fr/dumas-02175125/document. [Consulté le 18 juin 2025].

[15] A. Friconneau, “Réseaux électriques intelligents,” *DSpace, Université de Lorraine*, [En ligne]. Disponible : https://dumas.ccsd.cnrs.fr/dumas-02175125/document. [Consulté le 18 juin 2025].

[16] A. Friconneau, “Réseaux électriques intelligents,” *DSpace, Université de Lorraine*, [En ligne]. Disponible : https://dumas.ccsd.cnrs.fr/dumas-02175125/document. [Consulté le 18 juin 2025].

[17] A. Friconneau, “Réseaux électriques intelligents,” *DSpace, Université de Lorraine*, [En ligne]. Disponible : https://dumas.ccsd.cnrs.fr/dumas-02175125/document. [Consulté le 18 juin 2025].
[18] “Machine Learning – Projects,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/machine-learning-projects/
[19] “Supervised Machine Learning,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/supervised-machine-learning/
[20] “Unsupervised Learning,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/unsupervised-learning/
[21] “What is Reinforcement Learning,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/what-is-reinforcement-learning/
[22] “Machine Learning Lifecycle,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/machine-learning-lifecycle/
[23] “Understanding Logistic Regression,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/understanding-logistic-regression/
[24] “Support Vector Machine Algorithm,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/
[25] “Random Forest Algorithm in Machine Learning,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/
[26] “XGBoost Algorithm in Machine Learning,” GeeksforGeeks. [En ligne]. Disponible : https://www.geeksforgeeks.org/machine-learning/xgboost -algorithm-in-machine-learning/
[27] *Python Notes for Professionals*, Goalkicker.com, 1re éd., 2018. [En ligne]. Disponible : https://books.goalkicker.com/PythonBook/

Annex

Technologies de l'information et de la communication : Les Technologies de l'Information et de la Communication (TIC) désignent l'ensemble des outils et des services utilisés pour la production, le traitement, le stockage, la communication et la diffusion de l'information.
Elles regroupent un vaste éventail de technologies, incluant l'informatique (ordinateurs, logiciels), les télécommunications (internet, téléphonie mobile, réseaux), l'audiovisuel (télévision, radio) et les médias numériques.

Capteurs : Ce sont des dispositifs qui collectent des données en temps réel sur l'état du réseau électrique. Ils mesurent des paramètres comme la tension, le courant, la fréquence, ou la température. Ces informations sont cruciales pour surveiller et comprendre ce qui se passe sur le réseau.

Actionneurs : Ce sont des composants qui agissent sur le réseau en réponse aux informations des capteurs et aux commandes du système de gestion. Ils peuvent, par exemple, ouvrir ou fermer des disjoncteurs, ajuster la puissance des équipements ou reconfigurer le réseau pour optimiser la distribution.

Analyseurs : Il s'agit de systèmes logiciels ou matériels qui traitent et interprètent les vastes quantités de données collectées par les capteurs. Ils utilisent des algorithmes complexes pour identifier les problèmes (pannes), prévoir la demande, optimiser les flux d'énergie et prendre des décisions pour améliorer l'efficacité et la fiabilité du réseau.

Compteur intelligent : C'est un compteur d'électricité avancé qui mesure la consommation d'énergie de manière détaillée (par exemple, toutes les 15 minutes) et communique ces données bidirectionnellement avec le fournisseur d'énergie. Il permet aux consommateurs de mieux suivre leur consommation et aux gestionnaires de réseau de collecter des informations précises pour optimiser la production et la distribution.

Énergies renouvelables : Ce sont des sources d'énergie inépuisables ou qui se reconstituent naturellement, comme l'énergie solaire (photovoltaïque), éolienne, hydraulique ou géothermique. Dans les smart grids, leur intégration est clé pour réduire la dépendance aux énergies fossiles et rendre le réseau plus durable, même si leur nature intermittente représente un défi.