

M2 SIIA

Artificial intelligence applied to the predictive
maintenance of intelligent solar inverters

Ibtissem

26 août 2019

UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Faculté des Sciences et Techniques

Département INFORMATIQUE



Rapport de stage M2

Systemes Informatiques Intelligents et Autonomes :

Intelligence Artificielle appliquée à la
maintenance prédictive d'onduleurs

- Réaliser par :
Ibtissem TAZTAZ

- Tuteur :
Romain Le FORSTIER
- Enseignant référent :
Vincent RODIN

Rapport Confidentiel

2018-2019

REMERCIEMENTS

Je tiens à adresser en premier lieu mes remerciements et ma profonde gratitude, respect et admiration à mon tuteur Monsieur Romain LE FORSTIER et à Monsieur Cristophe Gouasguen de par leurs disponibilités, et leurs feedbacks précis.

Je tiens aussi à remercier Monsieur Vincent RODIN, pour d'abord la qualité de formation en tant qu'enseignant du module Méthodes computationnelles. Ensuite, pour ses précieux conseils et recadrages lors de mon stage en tant qu'enseignant référent.

Mes remerciements s'étendent aussi aux membres du jury, Monsieur Jérémy Rivière et Monsieur Thierry Duval, qui m'ont fait l'honneur d'examiner et de juger mon travail.

Je tiens enfin à remercier toute personne ayant contribué de près ou de loin à l'aboutissement de ce projet.

Ibtissem TAZTAZ

Ce présent mémoire est le fruit d'un dur labeur qui vient s'inscrire comme la fin d'un cycle d'apprentissage : le Master.

Dans l'espérance de nouveaux projets et défis encore plus passionnants et constructifs, je dédie ce travail :

A mes parents

Toujours présents, disponibles et attentifs, ce travail est le fruit de vos efforts et de vos sacrifices. Merci de m'avoir accompagnée et aidée dans mes études. Que Dieu vous accorde longue vie et santé.

TABLE DES MATIÈRES

Liste des tableaux	vii
Table des figures	viii
Introduction générale	viii
1 Organisme d'accueil : IMEON ENERGY	1
1.1 IMEON ENERGY	1
1.1.1 IMEON : Onduleur connecté intelligent	2
1.1.2 Mode de fonctionnement	2
2 La maintenance prédictive aux aurores de l'Intelligence Artificielle	5
2.1 Industrie 4.0	5
2.2 Types de maintenance	7
2.2.1 Corrective	7
2.2.2 Préventive	8
2.2.3 Prédictive	8
2.3 Surveillance en continu des données	9
2.4 Schémas des pannes	10
2.5 Systèmes de maintenance prédictive	10
2.5.1 Technique de maintenance prédictive	11
2.5.1.1 Approches statistiques :	11
2.5.1.2 Approche basée Apprentissage Automatique :	12
2.6 Intelligence Artificielle appliquée à la maintenance prédictive : Etat de L'art . .	17
3 Conception	21
3.1 Vue d'ensemble de la conception	21
3.2 Visualisation des données	22
3.3 Extraction des données	22
3.4 Prétraitement des données	23
3.4.1 Nature des données	23
3.4.1.1 Données remontées par les onduleurs	23
3.4.1.2 Données internes : Maintenance/ ERP	24
3.4.2 Techniques de prétraitement des données	24
3.4.2.1 Normalisation	24
3.4.2.2 Filtrage des données	24
3.4.2.3 Extraction d'information	25

3.4.2.4	Data mining	26
3.5	Apprentissage Automatique	27
3.5.1	Apprentissage des données onduleurs	27
3.5.2	Apprentissage des données batteries	28
3.5.3	Apprentissage des données PV	28
4	Implémentation et résultats	30
4.1	Implémentation	30
4.1.1	Environnement de développement	30
4.1.1.1	Psycopg – PostgreSQL database adapter for Python	31
4.1.1.2	Outil de Data visualisation	31
4.1.1.3	Scikit-learn, librairie d'apprentissage automatique	34
4.1.2	Choix d'implémentation	35
4.1.3	Architectures modèles prédictifs	35
4.1.3.1	SVM	35
4.1.3.2	MLP	36
4.1.3.3	RandomForest Classifier	36
4.1.4	Mesures d'évaluation	36
4.1.4.1	Mesure de précision	36
4.1.4.2	Mesure d'erreur	37
4.2	Résultats	37
4.2.1	Performances des algorithmes d'apprentissage	37
4.2.1.1	Performances des prédictions des erreurs onduleurs	37
4.2.1.2	Performances des prédictions des erreurs batteries	37
4.2.1.3	Performances des prédictions des erreurs PV	38
4.2.2	Evaluation de notre travail par rapport aux travaux antérieurs	38
4.2.3	Module Prédictif	39
	Conclusion générale et perspectives	41
	Bibliographie	48

LISTE DES TABLEAUX

4.1	Tableau comparatif des choix d'implémentation [LOUKIL, 2018]	35
4.2	Tableau comparatif des résultats de l'apprentissage sur des données des erreurs onduleurs	37
4.3	Tableau comparatif des résultats de l'apprentissage sur des données des erreurs batteries	38
4.4	Tableau comparatif des résultats de l'apprentissage sur des données des erreurs PV	38
4.5	Comparaison des précisions de nos modèles prédictifs avec les résultats des tra- vaux antérieurs	39

TABLE DES FIGURES

1.1	Organigramme de l'entreprise IMEON ENERGY.	1
1.2	Schémas d'installation solaire. [FORESTIER, 2016]	2
1.3	Les différents modes de fonctionnement de l'onduleur IMEON [www.imeon energy.com, 2019]	3
1.4	Exemple d'utilisation d'IMEON [www.imeon energy.com, 2019]	4
2.1	Composants clés d'industrie 4.0	7
2.2	Moment d'intervention des différents types de maintenance [www.mps maintenance.fr, 2019]	9
2.3	Les six étapes-clés de l'analyse prédictive [ZedNet, 2015].	10
2.4	Architecture d'un système expert.	11
2.5	Corrélation des données. [Guillaume, 2010]	12
2.6	Neurone artificiel [Chrislb, 2016]	15
3.1	Schéma récapitulatif des étapes de conception l'outil de prédiction de panne	21
3.2	Application de data visualisation	22
3.3	Base de données monitoring	23
3.4	Champs vides	25
3.5	Boîte à moustache	26
3.6	Les codes erreur onduleur	27
3.7	Timeline	28
3.8	Les codes erreurs batteries	28
3.9	Les codes erreur PV	29
4.1	Tableau de bord de l'application de visualisation des données.	32
4.2	graphique et tableau résumant l'évolution des installation des onduleurs.	32
4.3	Partie statistique	33
4.4	Tableau récapitulatif	33
4.5	Analyse des taux de connections des onduleurs	33
4.6	Model viewer controller	34
4.7	Consommation de mémoire par les deux API d'apprentissage automatique.	35
4.8	L'ensemble des modules / applications internes d'onduleur.	39
4.9	Le Module prédictif.	40

INTRODUCTION GÉNÉRALE

L'ère industrielle est loin derrière : nous faisons face à une révolution du numérique ; nous vivons une époque de digitalisation extrême. Les TIC font désormais partie intégrante de notre quotidien : pour toute action nous consultons internet, nos smartphones et objets connectés. L'industrie n'est pas en reste de ce bouleversement ; le domaine industriel connaît une expansion fulgurante, ne cessant d'évoluer et étant au cœur de toutes sortes d'innovation.

Les objets connectés (IoT) ont submergé le marché industriel sous un amant de données phénoménal. Avec de plus en plus de machines connectées, un grand nombre de données utiles sont générées. On parle du big Data.

Le Big Data nous pousse à proposer aux clients des solutions customisées en ciblant les besoins de chacun de façon à y répondre instantanément. L'un des plus grands besoins des clients, dans le domaine industriel, est d'assurer la maintenance des machines, ainsi que de pouvoir estimer la durée de vie d'un composant et/ou prédire des défaillances potentiels ; ceci passe inévitablement par la maintenance prédictive.

La maintenance prédictive permet aux clients de mieux utiliser leurs machines en détectant l'apparition d'anomalies sur ces machines avant qu'elles ne deviennent trop graves, et aux fabricants d'anticiper les pannes et d'estimer les coûts de la maintenance par la suite. Beaucoup de chercheurs se sont penchés sur le problème de la maintenance prédictive en proposant diverses solutions avec des approches différentes. L'intelligence artificielle et notamment l'apprentissage automatique ont révolutionné la prédiction en général et la maintenance en particulier. Certains algorithmes d'apprentissage permettent des prédictions de très grande précision et à long terme grâce à leur capacité à inférer des modèles prédictifs qui se basent sur l'exploitation des historiques des données collectées.

Dans ce contexte, notre travail consiste en l'exploitation du contenu des données industrielles générées par la gamme des onduleurs solaires connectés IMEON afin de mettre en place un module prédictif qui permet détecter la date et la nature de la prochaine panne prévisible.

Contribution

Ce projet nous a permis d'explorer le domaine d'exploitation des données industrielles et les techniques de la maintenance prédictive. Notre contribution expérimente les algorithmes d'apprentissage automatique classés parmi plus performants dans le domaine de prédiction selon

les travaux de recherche récents.

Par le biais de ce mémoire nous présenterons toutes les étapes de notre approche de réalisation d'un outil d'aide à la décision aux futurs utilisateurs des onduleurs solaires IMEON. Nous aurons ainsi réalisé la première solution de maintenance prédictive de la gamme IMEON, avec une technique de prédiction basée apprentissage automatique.

Organisation du rapport de stage

Nous avons organisé notre rapport de stage en quatre chapitres :

Dans le premier chapitre « **Organisme d'accueil : IMEON ENERGY** » nous présenterons l'entreprise accueillante IMEON ENERGY, ses secteurs d'activités. Nous présenterons aussi ses technologies innovantes, ainsi que leurs produits, notamment la gamme d'onduleur solaire IMEON et son mode de fonctionnement.

Dans le second chapitre « **La maintenance prédictive aux aurores de l'Intelligence Artificielle** » nous aborderons les systèmes de maintenance prédictive. Nous présenterons d'abord le contexte qui est Industrie 4.0 et les différents sources de données de ces systèmes, ainsi que les techniques de surveillance des données pour leur exploitation. Nous présenterons aussi les différentes techniques de prédiction. Enfin nous clôturerons ce chapitre avec un état de l'art sur les travaux antérieurs les plus récents et pertinents qui ont abordé le problème de la maintenance prédictive.

Le troisième chapitre « **CONCEPTION** » détaillera toute la phase conceptuelle de notre travail : nous y présenterons la source exploitée pour la collecte de données, les étapes de collecte et de prétraitement ainsi que notre approche d'apprentissage automatique. Enfin nous y détaillerons la conception de notre outil de prédiction.

Le quatrième chapitre « **IMPLÉMENTATION ET RÉSULTATS** » nous y présenterons les aspects techniques de notre travail, ainsi que les résultats obtenus avec notre approche d'apprentissage automatique. Nous y effectuerons également une étude comparative et critique de nos résultats. Enfin, nous finirons avec une présentation de notre outil de prédiction de pannes.

La conclusion générale portera sur l'ensemble de notre travail, ainsi que les perspectives d'amélioration et d'extension de notre projet.

CHAPITRE 1

ORGANISME D'ACCUEIL : IMEON ENERGY

1.1 IMEON ENERGY

Imeon Energy est une entreprise fondée par Christophe Goasguen en 2013, elle est située à Brest - France. Il s'agit d'une entreprise spécialisée dans la fabrication d'onduleurs « intelligent » dédiés aux installations photovoltaïque en autoconsommation. L'onduleur IMEON est le fruit de nombreuses années de recherche et développement, cette technologie est protégée par un brevet international.

La start-up Bretonne est constituée de deux commerciaux, quatre ingénieurs, un directeur des opérations, une responsable logistique et une responsable qualité :

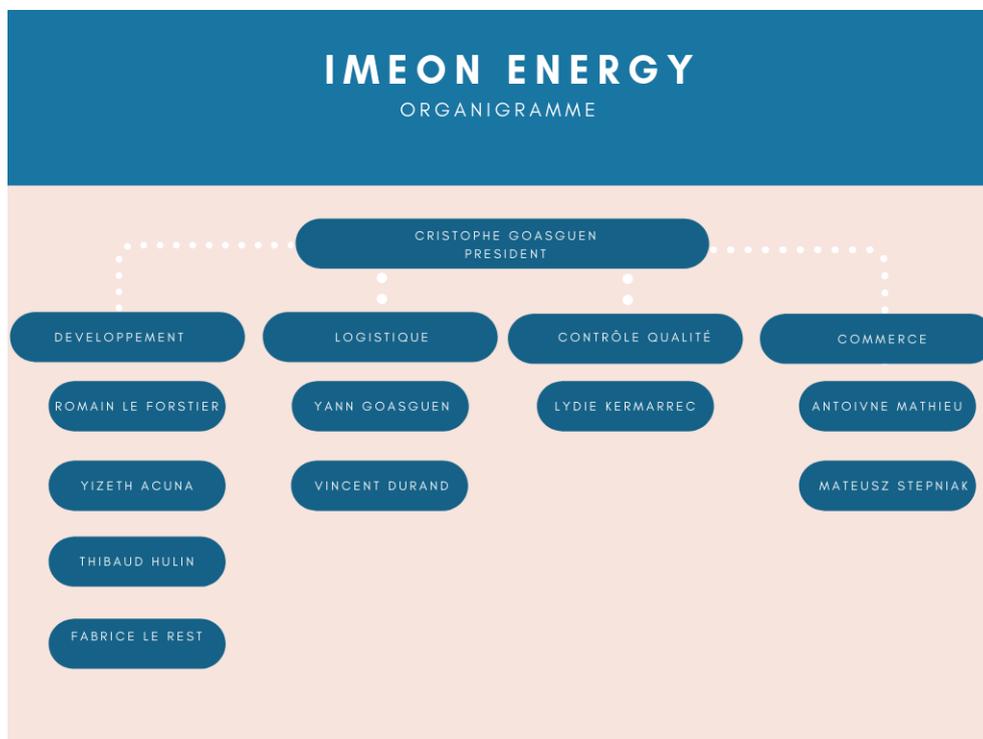


FIGURE 1.1 – Organigramme de l'entreprise IMEON ENERGY.

Le bureau d'étude de Brest s'occupe de la conception des onduleurs IMEON et une partie d'assemblage des pièces. La fabrication des pièces est réalisée en Chine.

La société réalise 95% de son chiffre d'affaires à l'export, commercialisé via un réseau international de 30 distributeurs et disponible dans plus de 70 pays.

1.1.1 IMEON : Onduleur connecté intelligent

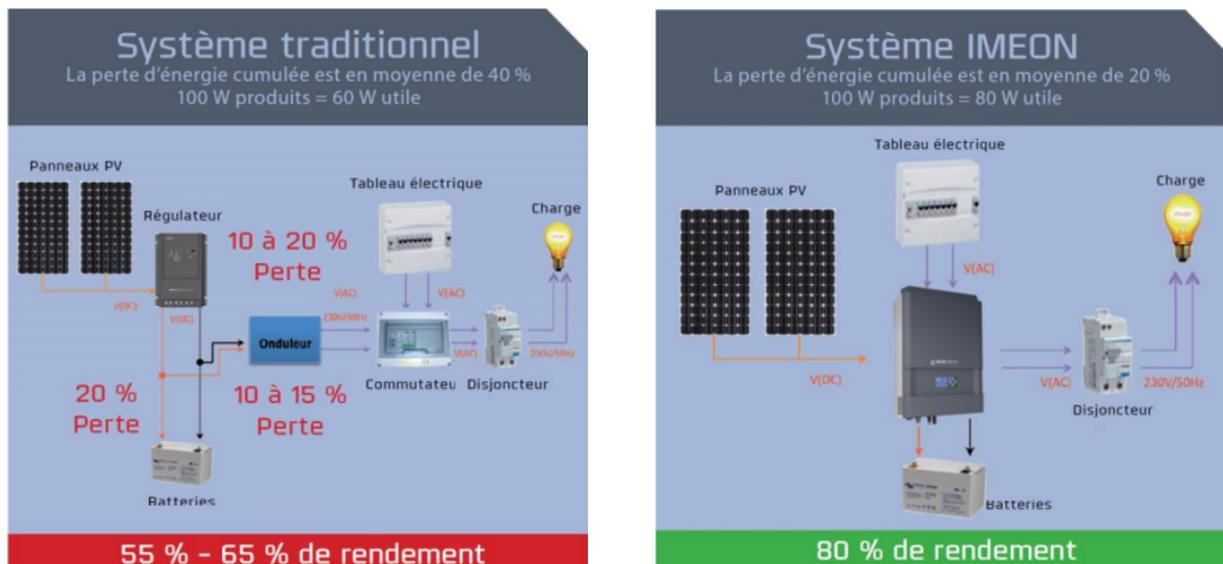
IMEON intègre au sein d'un processeur ARM CORTEX, une Intelligence Artificielle lui permettant d'apprendre, de comprendre et d'anticiper, dans le but d'atteindre les objectifs de l'utilisateur (Autonomie, économies, sécurisation, rendement...).

Cette nouvelle génération d'onduleurs pour l'autoconsommation crée une rupture radicale avec les technologies actuellement disponibles sur le marché avec 5ans d'avance [www.imeonenergy.com, 2019].

1.1.2 Mode de fonctionnement

IMEON gère plusieurs sources d'énergie (PV/Batteries/ Réseau) et oriente l'électricité disponible en fonction des conditions de production et des besoins en consommation. Chaque KWh générée est consommé directement, stocké dans des batteries où vendu à un opérateur.

IMEON est conçu pour s'adapter sur tous les types d'installation solaires en autoconsommation : site isolé (off-Grid) raccordé en réseau On Grid, ou hybride (off-grid et on). Il remplace : onduleurs, régulateur de charge, inverseur de sources, routeur, etc. IMEON simplifie la mise en œuvre d'un système solaire photovoltaïque et réduit donc fortement le temps d'installation :



(a) Installation solaire traditionnelle

(b) Installation solaire avec un IMEON

FIGURE 1.2 – Schémas d'installation solaire. [FORESTIER, 2016]

Imeon gère le stockage de manière intelligente pour limiter l'utilisation des batteries. Le système charge uniquement l'excédent de production dans les batteries et les décharge seulement pour assurer le complément nécessaire à l'énergie solaire.

IMEON permet donc de réduire la capacité du stockage et de prolonger sa durée de vie en évitant un maximum de solliciter les batteries (nombre de cycles réduit). Les systèmes Off-Grid gèrent le stockage par seuils fixes alors qu'IMEON, via une compensation dynamique, ajuste les seuils en permanence en fonction des courants de charge (production) et de décharge (consommation) :

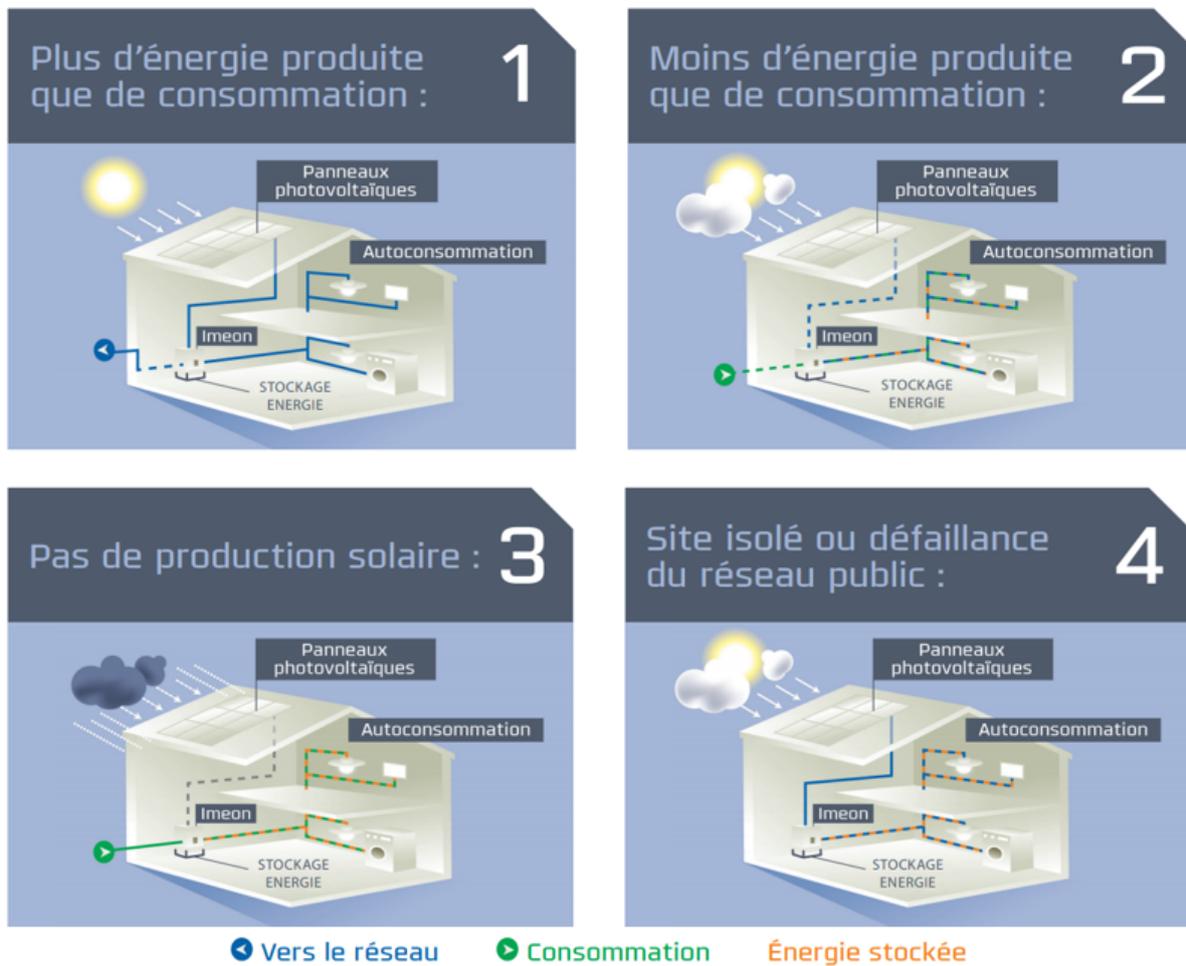


FIGURE 1.3 – Les différents modes de fonctionnement de l'onduleur IMEON [www.imeon energy.com, 2019]

IMEON réduit le coût de l'électricité produite par une installation solaire photovoltaïque jusqu'à 30%.

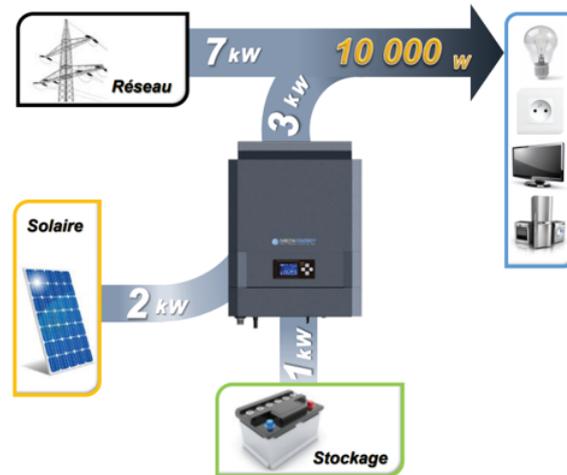


FIGURE 1.4 – Exemple d'utilisation d'IMEON [www.imeon energy.com, 2019]

CHAPITRE 2

LA MAINTENANCE PRÉDICTIVE AUX AURORES DE L'INTELLIGENCE ARTIFICIELLE

Sans prévision il n'est pas possible de planifier. Chacun d'entre nous se fie aux prévisions météorologiques pour s'habiller le matin. Si nous prenons le parapluie pour sortir, c'est qu'il pleut ou que nous prévoyons qu'il va pleuvoir.

[Seiersen, 2006]

Les économistes et les mathématiciens se sont penchés sur la question de prédiction depuis très longtemps. L'économiste Adam Smith vers la fin du 17ème siècle expliquait que le prévisionniste ne donnait pas seulement le rythme du marché mais il cherche aussi à l'anticiper. Des siècles après, ils se penchent toujours sur le problème de prédiction pour pouvoir anticiper des crises économiques et expliciter des fluctuations non prévues des marchés.

Avec les avancées technologiques notamment l'apparition de l'intelligence artificielle, de nouvelles méthodes de prévision ont vu le jour, à l'image de l'Apprentissage Automatique et l'Apprentissage Profond. Nous nous concentrons dans cette partie sur le domaine de la maintenance, et plus précisément ce qui est appelé aujourd'hui les systèmes de maintenance prédictive dans le domaine de l'industrie.

Dans ce chapitre, nous définirons brièvement la maintenance, ses différentes formes, la surveillance des données et les systèmes de prédiction des pannes, mais tout d'abord nous expliciterons la source de données de ces systèmes, ainsi que les différents traitements nécessaires afin d'extraire et d'exploiter efficacement ces données dans le processus de prédiction.

Nous aborderons ensuite le deuxième plier de ce chapitre qui est l'intelligence artificielle. Nous clôturons ce chapitre avec un état de l'art sur l'expansion de la maintenance prédictive aux aurores de l'intelligence artificielle.

2.1 Industrie 4.0

La production industrielle continue de progresser depuis le tout début. Parfois, les changements étaient si puissants et si importants que nous devons les décrire sous le terme de "Révolutions industrielles". Le terme Industrie 4.0 est utilisé pour reconnaître les trois autres

révolutions précédentes de l'industrie.

Aujourd'hui, nous assistons à la quatrième révolution industrielle, également connue sous le nom Industrie 4.0. Il combine les atouts d'une fabrication industrielle optimisée avec les technologies Internet : Big Data (BD) Data Mining (DM), les systèmes cyber-physiques (CPS), l'Internet des objets (IoT) et les services Internet (IoS), qui modifient considérablement le processus de fabrication, la gestion de la maintenance et les stratégies de maintenance. Les technologies de l'information et de la communication se développent ensemble et touchent tous les domaines de la vie. Les appareils et les systèmes de notre environnement réel contrôlés par un logiciel intégré sont désormais intégrés au réseau de communication mondial, où Internet est le terme clé.

Le monde réel et le monde virtuel grandissent clairement ensemble. Industrie 4.0 est une sorte de mot à la mode de nos jours, largement discuté entre praticiens et théoriciens, et facilite la version de « Smart Factory » [Weyer et al., 2015]. Il a été introduit au salon de Hanovre en 2011 en Allemagne pour présenter une nouvelle tendance vers la mise en réseau des industries traditionnelles. À partir de ce moment, il a été démontré que de nombreux projets et programmes similaires reflétaient le concept d'industrie 4.0, tels que « Intelligent Manufacturing system » système de fabrication intelligente et "Smart Manufacturing" fabrication intelligente.

Sur la base de l'analyse de la littérature, la définition générale d'Industrie 4.0 peut être résumée comme suit : Industrie 4.0 est un terme collectif désignant les technologies et concepts d'organisation de la chaîne de production [Thuemmler and Bai, 2017]. Le principe de base de l'Industrie 4.0 est l'application d'IoT et de la fabrication intelligente, qui permet aux composants et aux machines de production de collecter et de partager des données en temps réel, et de tirer les meilleurs avantages de ces ressources coûteuses. Industrie 4.0 est la superposition de plusieurs développements technologiques liés aux systèmes CPS, IOT, IoS et DM. CPS fait référence à une nouvelle génération de systèmes dotés de capacités informatiques et physiques intégrées pouvant interagir avec l'humain via de nombreuses nouvelles modalités. La clé est la capacité d'interagir avec, et d'élargir les capacités du monde physique, par le biais du calcul, de la communication et du contrôle [Baheti and Gill, 2011]. En outre, Industrie 4.0 facilite le développement de systèmes de contrôle de production intelligents et flexibles, qui appliquent les technologies de l'information et de la communication pour que les machines puissent communiquer et interagir. Typiquement, Industrie 4.0 comprend principalement les composants clés suivants :

- Systèmes cyber-physiques
- Internet des objets
- Big Data et Data Mining
- Internet des services

La figure 2.1 montre les rôles de ces composants clés :

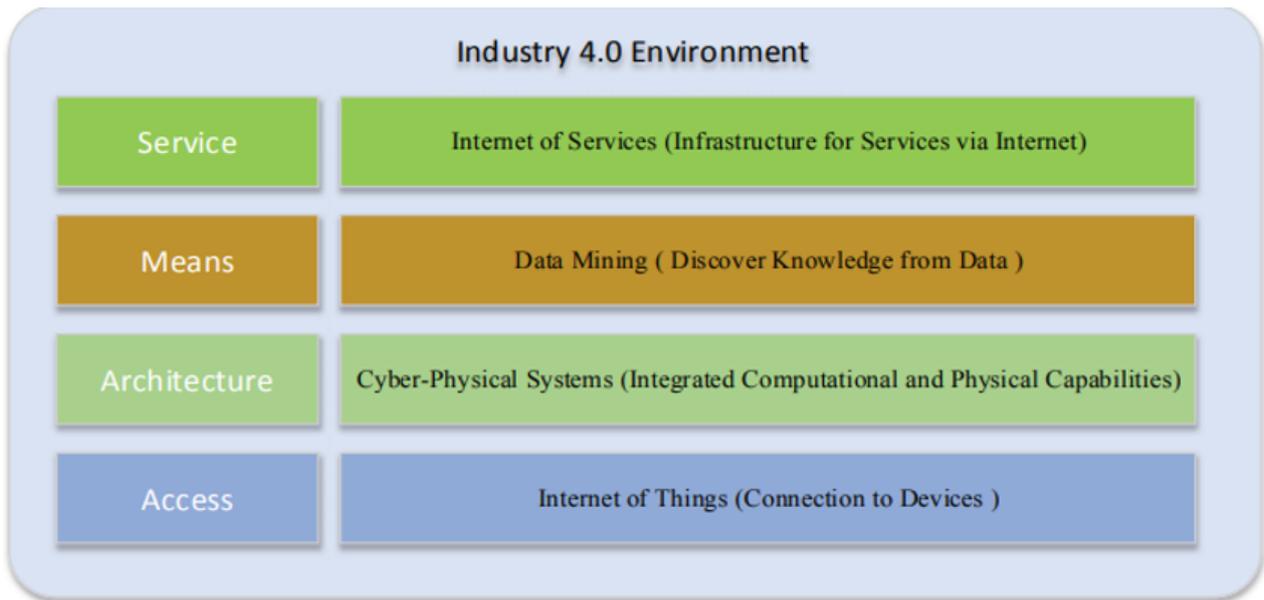


FIGURE 2.1 – Composants clés d'industrie 4.0

Nous sommes aux aurores de l'industrie 4.0 où les industriels doivent s'adapter à toutes les tendances naissantes pour éviter à leurs business une fin tragique. Dans ce contexte, la maintenance constitue le moyen d'assurance le plus répandu entre les éventuels industriels et clients. Cette dernière a connu un grand nombre d'évolutions au sein du projet 4.0 :

2.2 Types de maintenance

La maintenance est définie par la combinaison de toutes les actions techniques de dépannage, réparation, vérification, contrôle, déclassement, réforme et gestion d'équipements afin d'assurer une activité de production. Il existe essentiellement deux types de stratégies de maintenance : la maintenance corrective et la maintenance préventive. Ils peuvent également être divisés en sous-catégories [www.iri lyon.com, 2019] :

2.2.1 Corrective

L'intervention est effectuée après la détection d'une panne entraînant des arrêts de production. Il s'agit d'une réparation (maintenance curative) ou d'un dépannage (maintenance palliative) :

a Curative :

Elle consiste en la réparation complète de la machine suite à une identification du problème sans perturbation de la chaîne de production ou de l'utilisation du système rattaché à cette machine. En effet, une maintenance corrective curative visera à remplacer un / plusieurs composant(s) défectueux par un / plusieurs composant(s) neufs afin de résoudre définitivement le problème.

b Palliative :

La maintenance corrective palliative intervient lorsqu'un problème important ou bloquant, a été identifié. Ici, l'objectif est de réparer provisoirement le défaut afin de réta-

blir l'utilisation du matériel. Afin d'éviter tout risque futur d'une nouvelle panne, une opération de maintenance curative est nécessaire. Cette dernière permettra de corriger définitivement le problème par une réparation ou même un changement complet du matérielle.

2.2.2 Préventive

la maintenance préventive est effectué dans l'intention de réduire la probabilité de défaillance ou la dégradation du fonctionnement d'un bien (machine). Il peut s'agir d'inspection, de contrôle ou de visites sur équipements :

- a Maintenance programmée :
Maintenance préventive exécutée selon un calendrier préétabli ou selon un nombre défini d'unités d'usage.

- b Maintenance systématique : Maintenance préventive exécutée à des intervalles de temps préétablis ou selon un nombre défini d'unités d'usage mais sans contrôle préalable de l'état du bien.

2.2.3 Prédictive

La maintenance prédictive permet de détecter des anomalies sur des machines avant qu'elles ne deviennent trop graves. La force de la maintenance prédictive est donc d'anticiper les pannes. Ce qui évite tout arrêt de la chaîne de production ; qui peut engendrer des réparations coûteuses.

Fondée sur l'analyse de données collectées grâce à des capteurs intelligents positionnés sur l'équipement, elle permet de pousser à bout la logique du « juste à temps » et ainsi d'optimiser l'usage d'une machine avant de procéder à sa réparation ou à son remplacement, juste avant qu'une panne ne se produise.

La maintenance prédictive permet de prévoir les incidents techniques plutôt que de les subir. Grâce à cette stratégie, il est ainsi possible de programmer des interventions que lorsque des signaux émis par une machine reflètent une panne probable à court terme. C'est donc l'état réel de l'actif, et non un calendrier théorique, qui déclenche une intervention.

En comparant avec les stratégies de maintenance corrective et préventive, la maintenance prédictive présente les avantages suivants :

1. L'équipement nécessitant une maintenance est arrêté avant une panne imminente.
2. Réduire le temps total consacré à la maintenance des équipements.
3. Réduire les coûts de maintenance en évitant les dégâts catastrophiques.
4. Allonger la durée de vie des équipements et des processus.

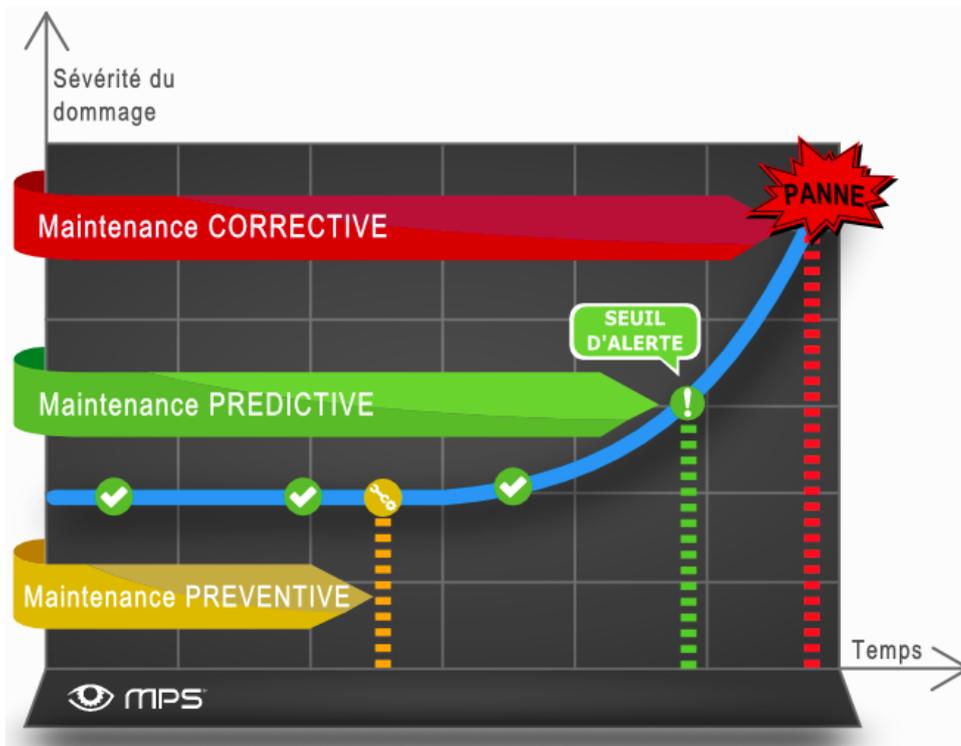


FIGURE 2.2 – Moment d'intervention des différents types de maintenance [www.mps maintenance.fr, 2019]

En tant que politique de maintenance la plus répandue et la plus moderne, la maintenance prédictive réduit les coûts de maintenance préventive en prévoyant au mieux le moment du changement des pièces seulement quand cela est réellement nécessaire : ni trop tôt (maintenance préventive), ni trop tard (maintenance corrective). Pour cela il est nécessaire de mettre en place une surveillance intelligente périodique de plusieurs paramètres sur les machines en fonctionnement.

2.3 Surveillance en continu des données

L'acquisition des données en contenu à partir de sources multiples est le fondement physique de la maintenance prédictive. L'objectif principale de cette étape, consiste à sélectionner les capteurs, les sources de données et la stratégie de collecte des données appropriés afin de refléter l'image réelle d'un actif (composant/ machine) [Baheti and Gill, 2011]. Ces données vont nourrir des algorithmes prédictifs par la suite.

La méthodologie traditionnelle de surveillance des données repose sur le principe suivant : lorsque l'équipement commence à tomber en panne, divers types de signes, tels que les fluctuations de température, les vibrations ou le bruit, peuvent être détectés, mais ceci nécessite une intervention sur place pour détecter les précurseurs de la défaillance [Li, 2018]. De nos jours, à l'ère des progrès technologiques rapides, la variété des capteurs disponible est suffisamment avancée pour jouer ce rôle sans une intervention humaine. En revanche, le coût de ces équipements de surveillance des conditions est souvent élevé. Par conséquent, la maintenance prédictive génère un coût d'investissement relativement élevé et de nombreuses entreprises ne voient pas facilement d'économies futures.

2.4 Schémas des pannes

La modélisation d'un schéma de panne est une étape cruciale pour mettre en place une stratégie de maintenance prédictive, là où chaque événement entrant est analysé afin de repérer les marqueurs d'une défaillance. En se basant sur l'historique de fonctionnement des machines, il est possible d'identifier les causes racines de la panne.

Un schéma de panne est souvent la combinaison de plusieurs facteurs difficiles à visualiser et à repérer avec une inspection humaine seule. Mais parfois, un seul événement peut signaler l'imminence d'une panne, ce qui nécessite aussi de mettre en place des modèles de détection des anomalies, afin d'émettre une alerte dès qu'un tel événement survient [Li, 2018].

2.5 Systèmes de maintenance prédictive

L'analyse prédictive sert à effectuer des prédictions dans divers domaines : environnement, justice, économie, etc. Il s'agit de deviner un fait, une donnée précise, à partir d'un historique. Un système de prédiction s'appuie exclusivement sur cette analyse descriptive ; sa construction est déterminée par les étapes suivantes [ZedNet, 2015] :

- Identification et extraction des données utiles en utilisant diverses sources.
- Prétraitement des données : normalisation, correction, etc.
- Construction d'un modèle prédictif à partir d'algorithmes statistiques et d'apprentissage automatique.
- Evaluation de l'efficacité et de la précision du modèle prédictif.
- Utilisation du modèle pour les tâches de l'aide à la décision.
- Maintien de l'efficacité du modèle prédictif avec une évaluation continue.

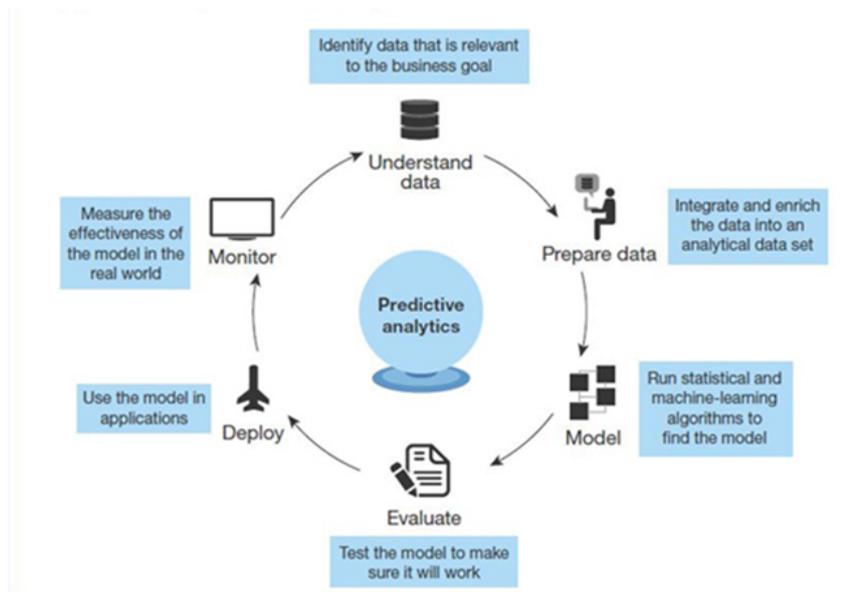


FIGURE 2.3 – Les six étapes-clés de l'analyse prédictive [ZedNet, 2015].

Un système de maintenance prédictive retourne l'image la plus probable de ce que l'état actuels/futurs sont/seront. Bien évidemment, les chiffres prévisionnels sont fortement dépendants des techniques utilisées. Comme précisé, pour construire un modèle de prédiction deux approches sont possibles, l'approche basée statistique et l'approche basée Apprentissage Automatique. Dans ce qui suit nous essayerons de présenter les différentes méthodes de chaque

approche.

2.5.1 Technique de maintenance prédictive

La maintenance prédictive est l'un des grands services modernes qu'offre divers secteurs de l'industrie. Elle utilise différentes méthodes statistiques ou d'Apprentissage Automatique pour déduire des informations présentes, une vision future.

2.5.1.1 Approches statistiques :

1. **Système expert** : Un système expert (SE) est un logiciel qui résout les problèmes d'un utilisateur d'une manière proche de celle d'un expert [Laforet et al., 1990].

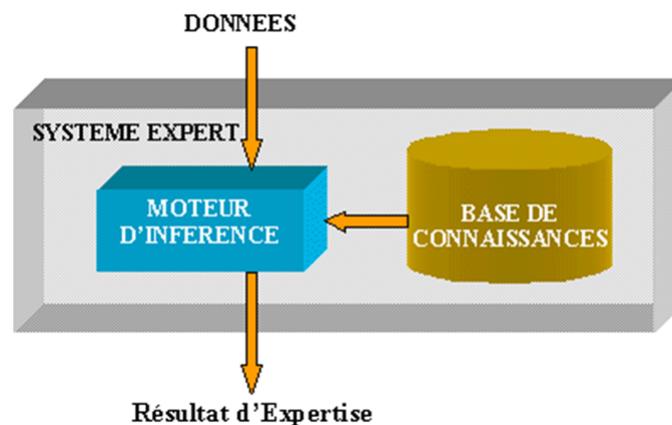


FIGURE 2.4 – Architecture d'un système expert.

Le processus de la prédiction des pannes à l'aide d'un système expert repose sur des données statistiques dans le but de prédire une panne d'un actif (composant / machine).

Comme défini dans la figure 2.4 le fonctionnement d'un système expert se base sur des « données » qu'on appelle base des faits et qui seront l'entrée d'un moteur d'inférence, qui a pour rôle de sélectionner les règles applicables à l'aide d'une base de connaissances. Le système expert retourne à l'utilisateur le résultat d'inférence (par exemple la panne estimée).

2. **Corrélation** :

La corrélation linéaire est un outil mathématique d'aide à la prévision. On parle de corrélation linéaire entre deux variables lorsqu'on constate un lien entre elles (si l'une augmente l'autre aussi et inversement). Le coefficient de corrélation linéaire entre deux attributs A et B se calcul de la manière suivante :

$$r(A, B) = \frac{\sum_{i=1}^n (A - \bar{A})(B - \bar{B})}{(n - 1)\delta A \delta B}$$

Tel que :

- n : le nombre de tuples.
- A et B : Moyenne de fréquence de l'attribut A et B , respectivement.
- δA et δB : déviation standard de A et B (la racine carré de la variance), respectivement :

$$\delta A = \sqrt{\frac{\sum(A - \bar{A})^2}{n}}$$

Le coefficient de corrélation peut être :

- Positif : les deux variables A et B sont Positivement corrélées.
- Négatif : A et B sont négativement corrélées.
- Nul : A et B sont indépendantes.

A l'aide du coefficient de corrélation, un nuage de point peut être représenté :

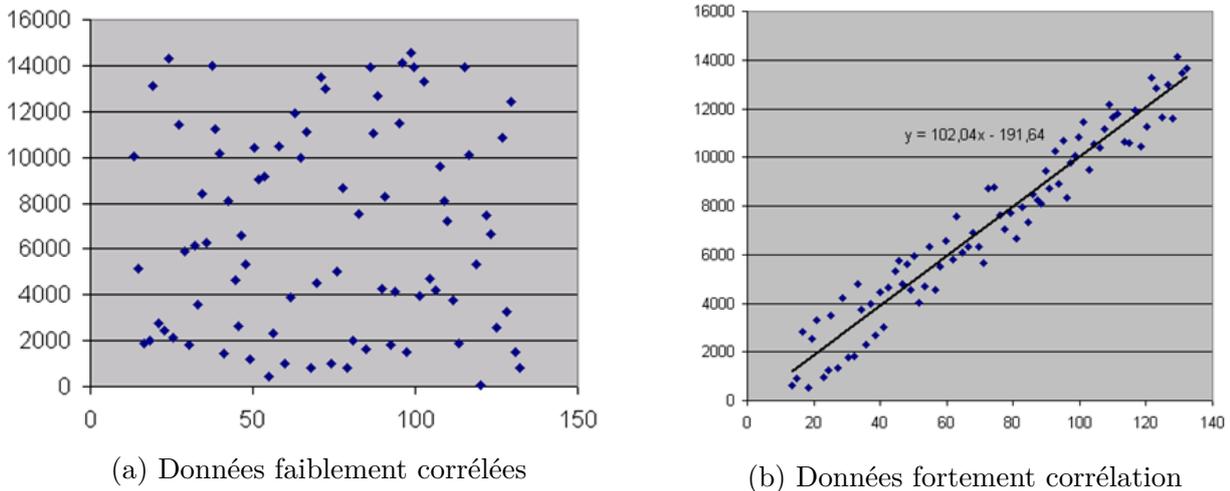


FIGURE 2.5 – Corrélation des données. [Guillaume, 2010]

On constate que dans la 2.5a les points sont éparpillés : il n'y a aucune relation évidente entre les deux attributs. Contrairement à la 2.5b qui représente un nuage de point dense qui permet, en effet, de dresser une droite regroupant un maximum de points : dans ce cas les prévisions seront plus fiables.

2.5.1.2 Approche basée Apprentissage Automatique :

Comme nous l'avons expliqué précédemment les systèmes de prédiction peuvent utiliser des techniques d'Apprentissage Automatique.

Nous leur consacrerons donc cette section. Nous aborderons des généralités sur cette discipline et exposerons quelques techniques liées à la prédiction citées dans ce rapport.

1. Apprentissage automatique :

" Apprendre aux machines à développer des processus d'apprentissage afin qu'elles deviennent intelligentes ".

[Lecun, 2010]

L'Apprentissage Automatique est une discipline de l'Intelligence Artificielle. Il consiste à utiliser des machines pour l'inférence d'un modèle de traitement de l'information selon certains critères de performances à partir d'observations. Ce modèle de traitement est impossible à déterminer à l'aide d'algorithmes classiques pour la simple raison que leurs résultats sont généralement médiocres et coûteux en temps.

L'Apprentissage Automatique est applicable à tous les domaines, pour des tâches simples ou complexes, afin d'améliorer les résultats grâce à l'expérience. On peut citer le domaine industriel (routage de paquets, contrôle et maintenance de machines, etc.), les jeux (Backgammon, Échecs, etc.) ou même le domaine d'imagerie médicale (reconnaissance d'objets) [Mitchell, 1997]. Le but même de l'Apprentissage Automatique est la prédiction de nouvelles données à partir d'un historique appris.

Le but de l'Apprentissage Automatique (AA) est de produire des modèles d'Apprentissage traitant un ensemble de caractéristiques (attributs). Par exemple, les caractéristiques d'un bilan médical ce sont généralement les valeurs qui ont été mesurées (âge, pression artérielle, niveau de cholestérol, etc.). Les tâches de l'AA peuvent être des tâches de classification, de Régression ou de détection d'anomalies.

2. Types d'Apprentissage Automatiques :

Les algorithmes du Machine Learning (ou Apprentissage Automatique) sont catégorisés selon le mode d'apprentissage qu'ils emploient [Mitchell, 1997] :

- Supervisé : Quand le système apprend à classer les données selon un modèle ; on parle alors d'Apprentissage Supervisé. Un algorithme d'Apprentissage Supervisé analyse la base de données et apprend à classer les données en se basant sur leurs caractéristiques. Il consiste donc à prédire l'étiquette (classe) d'une nouvelle donnée.
- Non supervisé : Contrairement à l'apprentissage supervisé, le système ne dispose que d'exemples (pas de classe/ étiquette). L'apprentissage non supervisé ou le clustering consiste en la division de l'ensemble de données en des sous-ensembles (clusters) similaires. La similarité est généralement calculée selon une fonction de distance entre paires d'exemples.
- Semi-supervisé : Lorsque les données contiennent des données étiquetées et non étiquetées. Une combinaison des deux algorithmes (supervisé et non supervisé) est effectuée. Il est mis en œuvre dans le cas de manque de données ou d'étiquettes.

Durant la phase d'apprentissage, l'algorithme peut rencontrer certains problèmes :

- a. *Le sous-apprentissage* : C'est le cas lorsque le modèle n'arrive pas à obtenir un taux d'erreur suffisamment réduit à partir des données d'apprentissage. En d'autres termes, l'algorithme n'arrive pas à induire un modèle qui satisfait les données d'apprentissage.
- b. *Le sur-apprentissage* : C'est la situation où la distance entre l'erreur d'apprentissage

et l'erreur de test est trop grande. Autrement dit, le programme retourne de bons résultats sur les données d'apprentissage mais n'est malheureusement pas aussi performant (voir pas du tout performant) sur les données qu'il n'a pas encore rencontrées (ensemble de test). Les paramètres appris ne permettent pas une généralisation efficace.

3. **Algorithmes d'Apprentissage Automatique** : Dans ce qui suit, nous nous intéressons aux méthodes qui sont souvent utilisées en Apprentissage Automatique [Mitchell, 1997] :

a. **Régression Linéaire**

Le modèle de régression linéaire désigne un modèle qui cherche à établir une relation linéaire entre une variable, dite expliquée, et une ou plusieurs variables, dites explicatives. Autrement dit, il consiste à rechercher la droite permettant d'expliquer le comportement d'une variable statistique Y comme étant une fonction affine d'une autre variable statistique X :

$$Y = F(x)$$

On considère le modèle pour l'individu i . Pour chaque individu, la variable expliquée s'écrit comme une fonction linéaire des variables explicatives, tel que :

$$y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i} + \varepsilon_i$$

Où :

* Y_i, X_i sont fixes et ε_i représente l'erreur.

* β le vecteur des paramètres du modèle $(\beta[0], \dots, \beta[k])$

En Apprentissage Automatique, la méthode de régression linéaire est considérée comme une méthode d'apprentissage supervisé utilisée pour prédire une variable quantitative. Le modèle de régression linéaire est aussi bien utilisé pour chercher à prédire un phénomène que pour chercher à l'expliquer. Après avoir estimé un modèle de régression linéaire, on peut prédire quel serait le niveau de Y pour des valeurs particulières de x .

Le modèle de régression linéaire est souvent estimé par la méthode des moindres carrés, mais il existe aussi de nombreuses autres méthodes pour estimer ce modèle.

b. **Les réseaux de neurones**

Le réseau neuronal s'inspire du fonctionnement du système nerveux humain. C'est une modélisation mathématique et algorithmique qui simule les unités de "calcul" que nous avons par milliards. Le modèle se compose de trois couches principales : couche de données d'entrée (exemple, les caractéristiques d'une voiture), couche (s) cachée(s) (souvent appelée «boîte noire») et couche de sortie (exemple, prix estimé de la voiture).

Le réseau neuronal est un réseau interconnecté de neurones artificiels avec une règle pour ajuster la force ou le poids des connexions entre les unités en réponse à des

données fournies extérieurement (Stanley et al., 1998).

Un neurone (un perceptron) est une unité de calcul. Elle comporte :

- * Un vecteur comme entrée (noté X)
- * Un vecteur de poids (noté W) avec un biais W_0
- * Une valeur de sortie (noté Y)

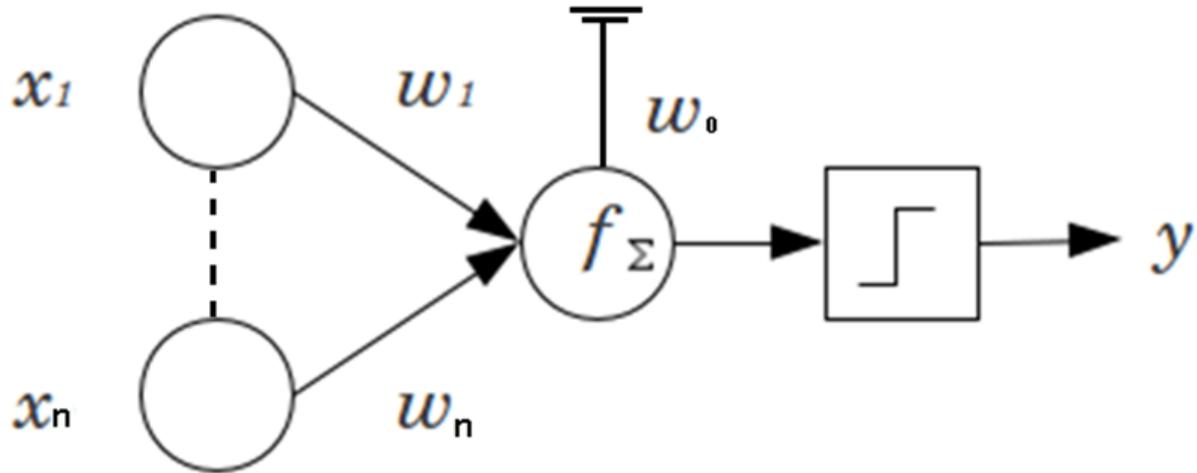


FIGURE 2.6 – Neurone artificiel [Chrislb, 2016]

Le principe de fonctionnement est assez simple, chaque neurone ayant comme entrée un vecteur X le multiplie par un vecteur de poids W. Le résultat de cette opération est, par la suite, injecté comme entrée dans une fonction d'activation F. On pourra noter le résultat comme ci :

$$y = F_{\text{activation}} \left(\sum x_n \times w_n \right)$$

La fonction $f(x)$ d'activation peut être une fonction :

- * seuil avec un résultat binaire : exemple si $f > 0.5$ alors output=1, sinon output=0,
- * à Sigmoid
- * à tangente hyperbolique

Il existe divers types de réseaux de neurones : les réseaux de neurones multicouches (MLP) utilisant la rétro propagation du gradient de l'erreur, mais aussi d'autres types de réseaux encore plus efficaces tels que les réseaux de neurones récurrents LSTM (Long Short-Term Memory) et les réseaux de neurones convolutifs.

c. Méthodes des K Plus Proches Voisins

La méthode des k plus proches voisins (KPPV) de l'anglais k-Nearest Neighbors (K-NN), est une méthode d'apprentissage supervisé. Dans ce cadre, on dispose d'une base de données d'apprentissage constituée de N couples « entrée-sortie ». Pour estimer la sortie associée à une nouvelle entrée x, la méthode des k Plus Proches Voisins consiste à prendre en compte (de façon identique) les k échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée x, selon une distance à définir.

Par exemple, dans un problème de classification, on retiendra la classe la plus représentée parmi les k sorties associées aux k entrées les plus proches de la nouvelle entrée x .

d. Classification Naive de Bayes

C'est un classifieur linéaire probabiliste basé sur le théorème de Bayes avec une indépendance naïve, c.-à-d qu'il se base sur le fait qu'une caractéristique pour une classe est indépendante des autres caractéristiques. Il s'agit d'un apprentissage supervisé très simple à implémenter mais qui donne de très bons résultats de classification [Chrislb, 2016].

e. Les arbres de décision

Un arbre de décision est la représentation graphique d'une procédure de classification ou d'association. Cette association est définie en commençant à la racine de l'arbre et en descendant dans l'arbre selon les réponses aux tests qui étiquettent les nœuds internes. La classe associée est alors la classe par défaut associée à la feuille qui correspond à la description. La procédure de classification obtenue a une traduction immédiate en termes de règles de décision. Les systèmes de règles obtenus sont particuliers car l'ordre dans lequel on examine les attributs est fixé et les règles de décision sont mutuellement exclusives. C'est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les feuilles) et sont atteintes en fonction de décisions prises à chaque étape.

Il existe plusieurs algorithmes qui implémentent les arbres de décision à l'image d'ID3 ou M5 ou J48.

f. Random Forest

Random Forest ou forêts d'arbres décisionnels aléatoires est une technique d'Apprentissage Automatique. L'algorithme des forêts d'arbres décisionnels effectue un apprentissage (régression ou classification) sur de multiples arbres de décision ; entraînés sur des sous-ensembles de données légèrement différents.

L'algorithme se repose sur l'apprentissage par arbre de décision, il se procède comme suit :

1. Créer B nouveaux ensembles d'apprentissage par un double processus d'échantillonnage :
 - a. sur les observations, en utilisant un tirage avec remise d'un nombre N d'observations identique à celui des données d'origine (technique connue sous le nom de bootstrap),
 - b. sur les p prédicteurs, en n'en retenant qu'un échantillon de cardinal $m < p$ (la limite n'étant qu'indicative).
2. Sur chaque échantillon, on entraîne un arbre de décision selon une des techniques connues, en limitant sa croissance par validation croisée.
3. On stocke les B prédictions de la variable d'intérêt pour chaque observation d'origine.
4. La prédiction de la forêt aléatoire est alors un simple vote majoritaire (Ensemble learning).

Le principal revers de cette méthode est que l'on perd l'aspect visuel des arbres de décision uniques. Elle effectue cependant des prédictions très précises par rapport aux arbres décisionnels ou divers autres modèles de régression puisqu'elle capture les

variances de plusieurs variables en entrée en parallèle et les inclue dans la prédiction [Chrislb, 2016].

g. **Machines à Vecteurs de Support**

Les SVM, de l'anglais Support Vectors Machines, représente un des séparateurs à vaste marge, une technique d'apprentissage supervisé adaptée aux données à grande dimensionnalité et pouvant être utilisée pour la classification, la distribution et la régression.

Les SVM reposent sur deux concepts : la marge maximale et la fonction noyau permettant ainsi de traiter les problèmes linéaires et non linéaires. La marge maximale permet de trouver la frontière séparatrice optimale des données à partir d'un ensemble d'apprentissage. Quant aux données non séparables linéairement, on utilise la fonction noyau qui utilise l'hypothèse qu'une probable séparation linéaire existe dans un plan dimensionnel plus grand et donc transforme les données en des données de plus grande dimensionnalité (peut arriver jusqu'à une dimensionnalité infinie) [chang Chih-Chung and Chih-Jen, 2013].

Il existe plusieurs variations d'optimisation pour les algorithmes SVM selon le problème, pour la classification, C-SVM et V-SVM sont les plus utilisés, on peut y ajouter le one-class SVM. Pour la régression, nous retrouvons SVR (epsilon SVR) et V-SVR (Simple VectorRegression). Ces divers algorithmes diffèrent principalement de par leurs fonctions de calcul d'erreur.

Cette section nous a permis d'introduire l'Intelligence Artificielle et ses différentes méthodes et algorithmes qui sont utilisés pour l'exploitation des données dans l'industrie et notamment la maintenance prédictive. Dans ce qui suit, nous présenterons quelques travaux récents que nous avons jugés pertinents par rapport à leurs approches et leurs résultats.

2.6 Intelligence Artificielle appliquée à la maintenance prédictive : Etat de L'art

Beaucoup de travaux de recherche autour de la maintenance prédictive dans le domaine de l'industrie ont été effectués durant cette dernière décennie. Nous essayerons de couvrir par notre état de l'art les travaux les plus importants afin de surligner les différentes approches déjà abordées, ainsi que les problèmes et les obstacles rencontrés, dans le but d'essayer d'améliorer par notre contribution ce domaine de recherche.

Bien avant l'engouement autour de l'Intelligence Artificielle(IA), de l'Apprentissage Automatiques (AA) et l'Apprentissage profond (AP), nombreux de travaux de maintenance prédictive ont utilisé l'approche statistique avec des modèles qui reposent la plupart du temps sur des hypothèses relativement restrictives quant à la description du système (par exemple sur les lois de dégradation et la représentation du contexte) ou encore des systèmes basés règles (schéma de panne), en se focalisant sur les connaissances des experts du domaine, afin de déterminer les facteurs influant sur les dysfonctionnements et les pannes.

Le système expert convient aux problèmes généralement résolus par des spécialistes humains. Cependant, il est difficile d'obtenir des connaissances sur le domaine et de les convertir

en règles. Dès lors, un système expert n'a pas la capacité de gérer de nouvelles situations qui ne peuvent pas être converties explicitement dans les bases de connaissances. En outre, lorsque le nombre de règles augmente considérablement, la situation peut provoquer une "explosion combinatoire", qui implique des problèmes de calcul.

Citant par exemple [Courtois and Delsarte, 2003] qui ont fait une analyse basée sur un modèle constitué de scénarios réalistes, avec des statistiques faibles obtenus en général par des simulations. C'est donc une approche qui vise à prédire le comportement réel d'une installation, erreurs humaines comprises. Cette analyse présente en contrepartie des limites vis-à-vis des incertitudes qui sont toujours associées aux statistiques manipulées notamment.

Dans un contexte probabiliste, [Celeux et al., 2006] proposent l'utilisation des Modèles Graphiques Probabilistes Markoviens MGPM qui sont des outils plus génériques, basés notamment sur des modèles graphiques permettant la représentation des systèmes plus complexes à l'aide des chaînes de Markov. Ce cadre de travail offre en effet un formalisme puissant et intuitif en matière de modélisation ainsi qu'un ensemble d'outils d'analyse adaptés à l'évaluation de politiques de maintenance prédictive :

Soient X l'ensemble des états du système et A l'ensemble des actions de maintenance. L'ensemble A est partitionné tel que : $A = A_x \cup A_{\bar{x}}$ où A_x et $A_{\bar{x}}$ représentent respectivement l'ensemble des actions laissant le système invariant et celui des actions affectant le système.

Ce modèle se repose sur plusieurs formules mathématiques représentant le fonctionnement dynamique d'un système multi-composant. Dans ce cas, le processus de dégradation de chaque composant est lui-même modélisé par une chaîne de Markov.

Naturellement, lorsque le nombre de composants augmente, les calculs deviennent de plus en plus complexes.

Plus récemment, [Ayadi, 2013] dans le cadre de sa thèse, intitulée « Optimisation des politiques de maintenance préventive dans un cadre de modélisation par modèles graphiques probabilistes. », propose un modèle probabiliste graphique appliqué à un problème réel de maintenance de la voie ferroviaire. Cette approche se concentre sur l'analyse de dégradation de la voie ferroviaire avec une temporisation des données à l'aide des chaînes de Markov. [Ayadi, 2013] a conclu que malgré la limitation imposée par les chaînes de Markov, qui exigent obligatoirement une loi de temps de séjour de type exponentielle mais la solution s'adapte tout à fait à la formalisation des relations de dépendance entre les différents composants d'un système complexe évoluant au cours du temps.

En revanche, ces travaux nécessitent toujours une grande période d'analyse et d'interprétation des résultats et ne fournissent aucune solution algorithmique générale aux problèmes d'inférence soulevés, et de s'auto-adaptation avec de nouvelles situations.

L'IA et l'AA ont démontré un succès fulgurant concernant la prédiction en général et la maintenance prédictive en particulier. Certains chercheurs ont comparé les méthodes statistiques et les méthodes de l'AA : c'est le cas de [Paolanti and Romeo, 2018] qui ont étudié la maintenance prédictive des moteurs électriques et autres équipements utilisés par l'industrie. Ils ont comparé l'approche statistique avec corrélation et l'approche Apprentissage Automatique avec réseau de neurones.

Le système a été testé sur un exemple réel du secteur, en développant la collecte de données et l'analyse de système de données, en appliquant l'approche Machine Learning et en la comparant à l'analyse d'outil de simulation. Les données ont été collectées par divers capteurs, automates machine et protocoles de communication et mises à la disposition de Data Analysis Tool sur l'architecture Azure Cloud pour la phase de visualisation. Ils ont conclu que malgré un taux de précision de 65% sur la prévision des pannes les modèles statistiques étaient moins performants que les réseaux de neurones qui ont plafonné une précision de 95%. Ils ont expliqué que cela était principalement dû au fait que les relations entre les données requises des capteurs et les pannes sont non linéaires, ainsi que des erreurs contenues dans les données, sachant que les réseaux de neurones sont plus robustes aux erreurs que les méthodes statistiques.

[RABENORO et al., 2015] dans le cadre de sa thèse de doctorat, a implémenté un « Outils statistiques de traitement d'indicateurs pour le diagnostic et le pronostic des moteurs d'avions ». La thèse propose une méthodologie permettant de combiner le savoir expert à des méthodes d'apprentissage automatique tout en respectant les exigences liées à l'interprétabilité des résultats et aux traitements des données : qu'il s'agisse des difficultés concernant le stockage des données ou bien des difficultés liées à la définition des algorithmes de traitement eux-mêmes.

La question d'interprétabilité est très importante car les opérateurs ne veulent pas travailler avec des outils de type boîte noire ; ils doivent pouvoir interpréter facilement les informations fournies par les algorithmes de surveillance, notamment parce qu'il leur appartient de valider le résultat.

[RABENORO et al., 2015] s'est basé dans un premier temps sur l'exploitation du savoir expert à plusieurs niveaux : la détermination des variables d'intérêt, la suppression des effets de contexte, l'identification des signatures ou des axes de projection des données, la prise de décision. Ensuite, une phase apprentissage automatique a été mise en œuvre avec l'algorithme Random Forest (les forêts d'arbres aléatoires) avec une précision de 97%. Les résultats fournis par la méthodologie peuvent être très bons, mais malgré cela, les opérateurs métier ont besoin de comprendre comment la décision a été prise, afin d'être en mesure de confirmer ou non une décision difficile à prendre. C'est d'ailleurs pour cette raison que, le choix de la visualisation ainsi que la sélection d'indicateurs, avec un classifieur bayésien naïf, sont des étapes nécessaires pour que l'opérateur prenne confiance en la méthodologie et surtout remédier au problème d'haute interprétabilité.

[WANG et al., 2017] dans cet article, les auteurs proposent la méthode de prédiction DES-SVM une solution hybride basé sur SVM et DES (lissage exponentiel double). Les résultats expérimentaux ont démontré que le DES-SVM avait une précision élevée (95%) dans la prévision des défaillances des équipements optiques. Par conséquent, il est possible de sauvegarder et de basculer les services sur des liaisons sûres. Il peut ainsi protéger le réseau contre la perte de données.

Pour des soucis vis-à-vis l'environnement de travail (CPU : ARM Cortex Texas instruments 32 bits avec 8 Go de stockage) nous avons étudié quelques articles qui tournent autour l'implémentation des techniques d'apprentissage automatique (Machine learning) et d'apprentissage profond (Deep learning) dans un environnement peu performant / embarqué : [Vanhoucke et al., 2011] ont conçu un papier qui sert comme tutorial pour les étudiants et les chercheurs, montrant des techniques qui peuvent être utilisées pour optimiser le coût de calcul considérable dans l'absence d'un GPU.

Pour cela, ils ont pris la reconnaissance de la parole comme exemple de leurs études (avec un benchmark de 440 inputs, Intel Xeon DP Quad Core E5640 et Ubuntu OS). Cet article montre que des techniques simples peuvent améliorer considérablement les performances des algorithmes d'apprentissage (HMM/réseaux de neurones). Les opérations SIMD à point fixe récemment introduites sont particulièrement intéressantes dans les processeurs x86 qui font pencher la balance des performances en faveur de l'arithmétique en virgule fixe.

Ils ont montré qu'en utilisant ces instructions plus rapides, il était possible de construire un système de reconnaissance vocale en temps réel à l'aide d'un très grand réseau hybride sans coût en précision.

Ces articles présentés ne sont certainement pas les seuls qui abordent la maintenance prédictive. Néanmoins ils nous permettent de tirer quelques conclusions :

- L'approche classique est limitée aux accidents engendrés par des défaillances aléatoires. Ceci suppose le bienfondé de la conception basée sur l'approche IA.
- L'analyse des actions humaines est très difficile et aucune approche pertinente n'a encore été définie pour les erreurs de réalisation (commission error).
- La relation entre les attributs décrivant les caractéristiques des machines et les dysfonctionnements probables ne sont pas linéaires. De ce fait, les algorithmes non-linéaires donnent de meilleurs résultats que les algorithmes linéaires.
- Nous n'avons pas trouver de travaux liés à la maintenance prédictive des onduleurs.
- Les résultats sont fortement dépendants du niveau de connaissances et limités par ce dernier. Ils changent donc lorsque les connaissances évoluent.
- Les algorithmes de prédiction qui ont donné les meilleurs taux de précision sont SVM, réseau de neurones (MLR) et Random Forest.
- Les réseaux de neurones et SVM sont plus interprétables par rapport aux forêts d'arbres aléatoires (Random Forest).
- Les modèles en vogue aujourd'hui sont les SVM et random forests. Ils s'accommodent des variables discrètes ou continues et des valeurs manquantes. Ce n'est pas le cas des réseaux de neurones qui sont un peu passés de mode même si ceux-ci reviennent sur le devant de la scène avec les Deep Neural Network.
- La fourniture d'une stratégie de maintenance prédictive est entraînée de devenir indispensable dans les usines futur

Conclusion

Ce chapitre nous aura donné l'occasion de faire une présentation des données industrielles : sources des données, types des données et les différentes techniques de prétraitement et d'interprétations des données. Nous avons ainsi, présenté des notions de base de la maintenance prédictive et plus particulièrement l'analyse prédictive qui est utilisées pour ce faire. Nous avons, par la même occasion, introduit les différents technique de maintenance prédictive à savoir (l'approche statistique et l'approches basée machine Learning).

Enfin, nous avons illustré notre chapitre par des travaux antérieurs qui utilisent ces techniques dans le domaine de la maintenance prédictive.

Dans ce chapitre nous aborderons la partie conceptuelle de notre travail : nous présenterons la source exploitée pour la collecte des données, les étapes de collecte et de prétraitements effectués, les méthodologies d'apprentissage et d'évaluation de la précision de notre solution, ainsi que le processus de détection de panne.

3.1 Vue d'ensemble de la conception

Notre travail consiste en l'exploitation du contenu des données industrielles générées par la gamme des onduleurs solaires connectés IMEON afin de mettre en place un module prédictif qui permet détecter la date et la nature de la prochaine panne prévisible. Les étapes de conception de notre outil sont résumées dans la figure 3.1.

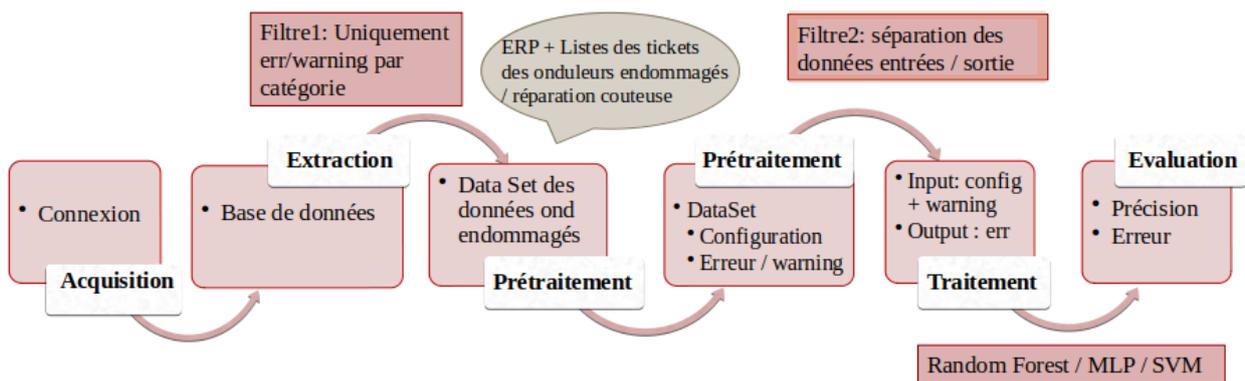


FIGURE 3.1 – Schéma récapitulatif des étapes de conception l'outil de prédiction de panne

Ainsi, nous avons en premier recensé les bases de données de l'entreprise pour choisir les données les plus pertinentes. Nous avons ensuite établi un schéma d'extraction des données remontées par les onduleurs. Une fois les données acquises, nous sommes passés à la phase de prétraitement, étape très importante dans le but d'améliorer au maximum la qualité des données, pour une qualité d'apprentissage optimale. L'étape suivante est la phase d'apprentissage, nous avons choisi un apprentissage à deux phases, la première consiste en l'utilisation des données structurées brutes uniquement ; la seconde est un raffinement avec l'exploitation

d'autre source de données (tickets maintenance/ERP). Une fois le modèle prédictif inféré, la dernière étape était de concevoir l'outil prédictif.

3.2 Visualisation des données

Nous avons choisi d'implémenter un outil qui nous permet de visualiser nos données. Il s'agit d'une application Web pour la visualisation de données interactif de manière très intuitif. Dans un premiers temps, il nous a permis de mettre en évidence plusieurs bruits et erreurs contenus dans les données que nous avons extraites et extraire les facteurs clés provoquant certaines pannes.



FIGURE 3.2 – Application de data visualisation

Dans un future proche, cette application servira comme outil de visualisations innovantes à destination des métiers (opérateurs, gestionnaires de parc, spécialistes systèmes).

3.3 Extraction des données

Dans le cadre du processus de collecte des données, il est essentiel d'avoir une connectivité de bout en bout entre le capteur et le serveur backend. Il s'agit là d'une étape très importante dans un projet de maintenance prédictive ; puisque la fiabilité de cette connexion déterminera le niveau de confiance à accorder aux systèmes de prédiction. La gamme d'onduleur IMEON possède de nombreux types de capteurs. La totalité des données obtenues par ces capteurs sont récupérées et stockées en local (onduleur) et centralisées par la suite dans un serveur distant afin de permettre aux experts métiers leur analyse. Ceci a rendu possible la mise en place d'un projet de maintenance prédictive.

Afin d'extraire ces données nous avons construit un aspirateur de données. La structure globale des données étant quasiment identiques, il s'agit de collecter les données structurées/semi-structurées stockées dans les bases de données de l'entreprise, ERP et les tickets de maintenance. Nous avons donc suivi le même cheminement pour construire l'aspirateur de données pour les différentes catégories de pannes. Le pseudo code suivant explique le procédé :

1. Création d'un nouveau projet aspirateur de données.
2. Définition des objets à extraire : les différentes informations relatives à l'onduleur.
3. Définir la source de données :
 - a. DB_HOST : l'adresse du serveur ou est hébergé la bases des données
 - b. DB_PORT : le port de réception et d'émission des données
 - c. DB_USER : le nom d'utilisateur
 - d. DB_NAME : Le nom de la base des données
 - e. DB_PWD : le mot de passe pour accéder à la base des données
4. Préciser le format et le chemin de stockage des données ;
5. Lancer le processus d'aspiration.

3.4 Prétraitement des données

La phase de prétraitement des données est souvent la plus laborieuse et qui demande le plus de temps. Ceci dû en particulier à l'absence de structuration et à la grande quantité de bruit existant dans les données brutes d'usage. Le prétraitement des données industrielles consiste à nettoyer et structurer ces données afin de les préparer à une future analyse.

Avant la phase de prétraitement des données, il est important d'étudier la structure et la nature des données afin de déterminer quels processus de prétraitement devons nous appliquer.

3.4.1 Nature des données

Lors de la phase d'aspiration, nous avons nettoyé les données afin d'obtenir des données brutes (sans balisage). Les données obtenues se subdivisent en deux grande partie :

3.4.1.1 Données remontées par les onduleurs

Les données remontées par l'onduleur révèlent des informations sur son état global, ses configurations, et l'historique des panne :

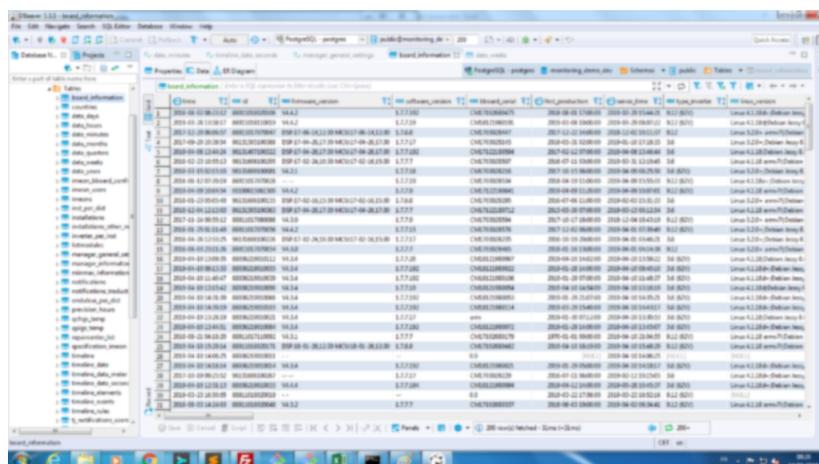


FIGURE 3.3 – Base de données monitoring

NB : La prise écran étant floue pour des soucis de RGPD (Règlement général sur la protection des données).

3.4.1.2 Données internes : Maintenance/ ERP

Nous avons exploité d'autres sources de données autre que l'onduleur, qui sont la base des données ERP et les ticket de maintenance :

- i. Enterprise Ressource Planning (ERP) : Il s'agit d'un logiciel de gestion qui permet à l'entreprise d'intégrer différentes fonctionnalités nécessaires pour le bon fonctionnement de cette dernière, comme la gestion des stocks, de paie et de comptabilité.
à IMEON Energy, l'ERP regroupe toutes les informations :
 - Qui touchent aux différentes transactions avec les fournisseurs/clients.
 - Des informations sur l'avancement des différentes tâches et projets.
 - Des informations logistiques et suivie qualité
- ii. Maintenance : Une opération de maintenance à IMEON ENERGY, commence par l'ouverture d'un ticket de maintenance. Ces tickets regroupe toutes les informations qui concerne le client et l'onduleur en question : la date, numéro de ticket, numéro de série de l'onduleur et son type, coordonnées du client, description du problème, pièce(s) à réparer ou changer si nécessaire, avis du technicien ... etc.

3.4.2 Techniques de prétraitement des données

Le prétraitement se fait à l'aide de méthodes spécifiques au traitement des amonts de données. Dans cette section nous présenterons les techniques utilisées pour le prétraitement des données :

3.4.2.1 Normalisation

La normalisation consiste à projeter des données dans un intervalle bien précis. Dans le cas des données onduleurs, nous avons choisi de convertir les attributs qui peuvent atteindre des grandes valeurs comme par exemple : les tensions des panneaux solaires, les puissances des composants onduleurs, etc afin d'éliminer les grandes écarts de ces valeurs.

3.4.2.2 Filtrage des données

- **Segmentation des attributs :**

Une seule base de données monitoring est utilisée pour stockées toutes les données peu importe le type d'onduleur. En effet, des standard (règles) d'enregistrement des données sont mises en place pour créer automatiser ce processus, on peut citer l'exemple de : l'attribut « voltage des panneaux solaires » : englobe 2 valeurs séparées par virgule dans le cas d'un IMEON 9.12, mais qu'une seule valeur pour la gamme IMEON 3.6. Nous avons séparé les deux attributs pour pouvoir les exploiter chacun à part dans la phase d'apprentissage. Ils existent d'autres attributs qui sont enregistré au format JSON, nous avons procédé de la même manière pour segmenter ces données.

- **Cohérence des données :**

Le nettoyage des données vise à éliminer toutes les données considérées comme inutiles de l'ensemble de données de départ. En effet, une quantité non négligeable des données

sont générées par des onduleurs du test interne à l'entreprise, ce qui rajoute du bruit à notre ensemble de données vu que la majorités des pannes sont provoquées dans le but de faire des contrôle qualité, des tests ou même pour des raison de maintenance. Nous avons donc éliminer toutes ces données incohérentes.

3.4.2.3 Extraction d'information

Étant donné qu'une partie des informations relative à l'onduleur et aux clients est saisie par les clients eu même via une interface de monitoring propre à chaque onduleur, nous consacrons cette section pour montrer les différents types d'anomalies existantes dans les données saisies.

La saisie d'un formulaire est probablement l'étape la plus délicate sur un site web. C'est à ce stade que de nombreux clients hésitent, se trompent et parfois abandonnent. L'ergonomie de l'interface joue un rôle clé dans l'accompagnement du client dans sa saisie et l'encouragement à poursuivre son chemin. Il est envisageable de proposer une validation automatique des formulaires.

Cette proposition sera importante dès que le concepteur définit comme « Obligatoire » le remplissage d'un certain nombre de champs. La validation automatique des champs de saisie se justifie également quand le contenu d'un champ doit respecter une mise en forme particulière : type, valeurs, etc. Enfin, elle est nécessaire lorsque le site doit vérifier la disponibilité d'une information pertinente, par exemple le type de l'onduleur ou certaines configuration, pour valider une entrée.

- Champs vide

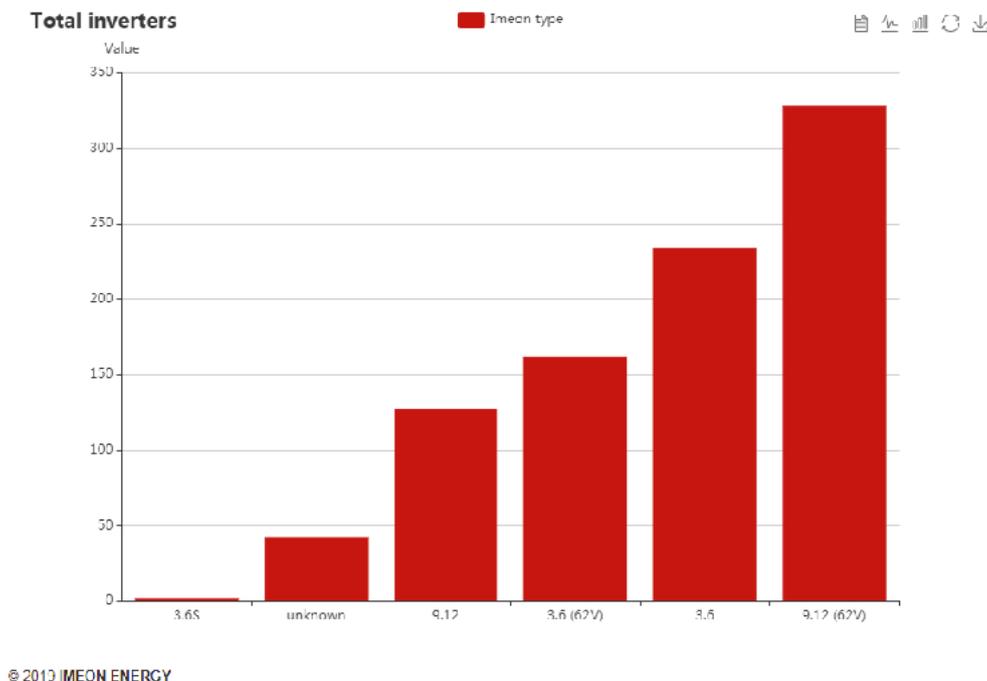


FIGURE 3.4 – Champs vides

- Bruit

Lors de la phase de visualisation des données nous avons constaté que nos données contiennent beaucoup de bruit. Une donnée aberrante (bruit) est une valeur ou une observation qui est "distante" des autres observations effectuées sur le même phénomène.

Dans ce qui suit, nous présenterons un exemple de certaines données aberrantes qui la date de la première production de l'installation solaire. Cette date est renseignée par le client :

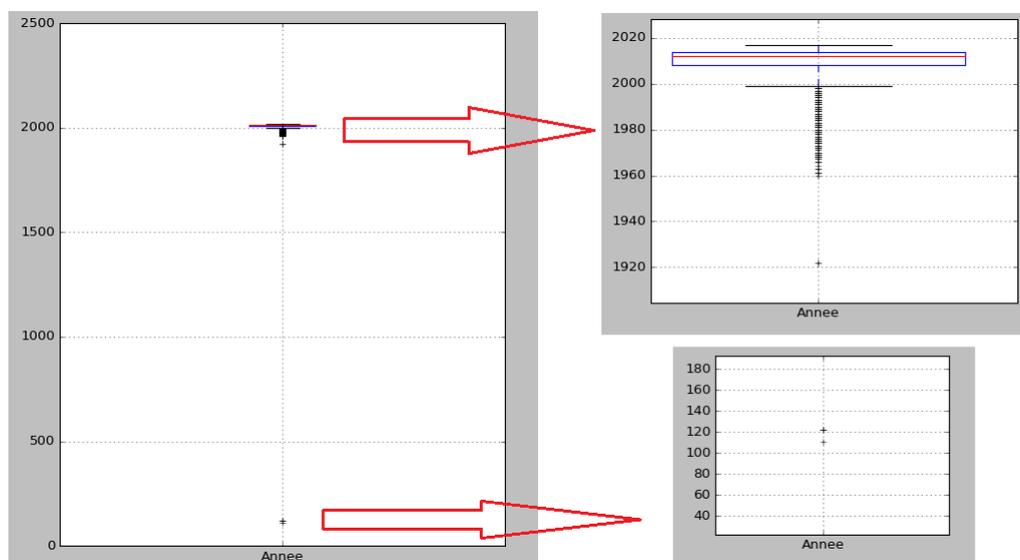


FIGURE 3.5 – Boite à moustache

La boite à moustache de l'attribut « first_production » d'une installation solaire 3.5 montre que certaines valeurs sont inférieures à 2015, le début des installations des onduleurs connectés, ce qui représente certainement un bruit par rapport aux années des première production qui sont logiquement comprises entre 2015 et 2019.

Pour remédier à ce problème, nous avons exploiter d'autre source de données pour extraire de l'information ; comme les données disponible sur l'ERP pour extraire le type d'onduleur et les données remontées par l'onduleur pour corriger la date de la première production.

3.4.2.4 Data mining

a. Réduction de la dimensionnalité

La sélection d'attributs consiste à choisir parmi un ensemble d'attribut, un sous-ensemble d'attributs intéressants pour le problème étudié.

Dans ce contexte, l'objectif de la sélection est de trouver un sous-ensemble optimal d'attributs qui a les propriétés suivantes : il doit être composé d'attributs pertinents et à éviter les attributs redondants. De plus, cet ensemble doit permettre de satisfaire au mieux la prédiction de panne, à savoir la précision de l'apprentissage, la rapidité ou bien encore l'applicabilité de l'algorithme d'apprentissage. Autrement dit, la réduction de la dimensionnalité permet de réduire le nombre d'attributs pour accélérer les calculs et représenter les données sous un format optimal pour l'exploration.

Notre approche consiste à étudier la corrélation des attributs avec le code erreur afin de mieux sélectionner le sous ensemble d'apprentissage.

3.5 Apprentissage Automatique

Une fois que nous avons effectué la phase de prétraitement, l'apprentissage automatique sur les données pouvait être lancé.

L'état de l'art nous a permis de dégager les trois algorithmes d'apprentissage automatique qui enregistrent les taux de précision les plus élevés dans le domaine de l'analyse prédictive : Random Forest, MLR et SVM. Dans notre approche d'analyse prédictive des erreurs, nous comparerons l'efficacité de ces trois algorithmes avec trois différentes catégories d'erreur : Onduleurs, batteries et panneaux photo voltaïque.

Nous avons choisi de faire un apprentissage à deux phases dans une première étape nous inférons un modèle d'apprentissage qui est basé sur l'exploitation des données qui reflètent des informations générales sur l'onduleur. La seconde étape concerne l'exploitation des données enregistré avant et au moment de l'erreur, nous rajoutons ainsi à nos données des informations extraites de l'ERP et les ticket de maintenance dans le but d'améliorer la précision de notre modèle inféré lors de la première étape.

Les attributs structurés sur lesquels va porter la première phase d'apprentissage sont : type d'onduleur, version linux, version software, version firmware, type batterie, le mode utilisé, des spécifications relatives à l'utilisation d'onduleur (la tension maximale autorisé, l'autorisation d'injection dans le réseau,...) et d'autres informations plus générales comme le pays d'installation, le mode de connectivité...,etc. Avec le code erreur comme valeur de sortie à apprendre. La deuxième phase concerne l'ajout de nouveaux attributs constitués après l'exploitation d'autre source comme expliqué précédemment :

3.5.1 Apprentissage des données onduleurs

la figure 3.6 représente les codes qui correspondes aux erreurs onduleurs avec leurs description :

A	B	C	D
ID	Warning / Error	Category	Title level A
54	Erreur	Error Ond	Over leakage current.
55	Erreur	Error Ond	Inverter DC <u>overcurrent</u> due to grid fluctuation.
56	Erreur	Error Ond	Grid measurement error on processor1.
57	Erreur	Error Ond	<u>GFCI</u> sensor failure.
58	Erreur	Error Ond	AC input ground loss.
59	Erreur	Error Ond	DC Bus discharge failure.
60	Erreur	Error Ond	DC Bus discharge start timeout.
61	Erreur	Error Ond	AC Output overload.
62	Erreur	Error Ond	<u>Long period overcurrent</u> detected on AC Output.
63	Erreur	Error Ond	AC Output short-circuit.
64	Erreur	Error Ond	<u>Fan failure</u> .
65	Erreur	Error Ond	AC Output sensor failure.
66	Erreur	Error Ond	Hardware version error.
67	Erreur	Error Ond	Input / Output reversal.
68	Erreur	Error Ond	AC Output <u>undervoltage</u> .
69	Erreur	Error Ond	AC Output <u>overvoltage</u> .
74	Erreur	Error Ond	Auxiliary internal supply voltage failure.
112	Erreur	Error Ond	Neutral / Ground wires are reversed.

FIGURE 3.6 – Les codes erreur onduleur

Dans cette catégorie, nous avons établi une liste d'attributs à rajouter. Il s'agit de l'ensemble des signes transmises par l'onduleur, avant et au moment de la panne : tension de sortie

alternative, tension d'entrée des panneaux solaires, le mode d'injection au moment de la panne, la disponibilité de batterie ou pas, ...etc.

Nous avons exploiter ainsi, une liste des codes d'avertissements (warnings) qui est générée par un module interne à l'onduleur "Timeline". Ce dernier a pour mission de contrôler la bonne utilisation de l'onduleur, en respectant les spécifications et les consignes d'usage ; si une anomalie est détecté, ce même module informe l'utilisateur en lui indiquant le code d'avertissement et des recommandations pour corriger la situation.

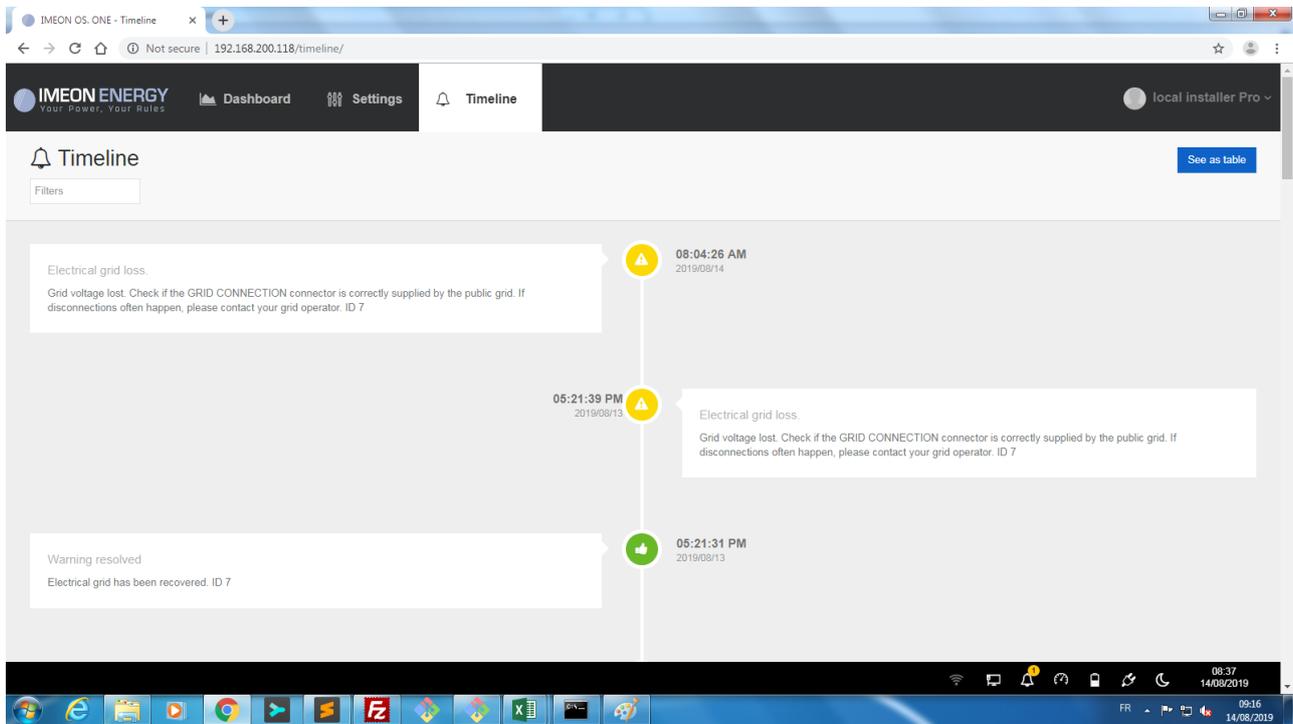


FIGURE 3.7 – Timeline

3.5.2 Apprentissage des données batteries

la figure 3.8 représente les codes qui correspondes aux erreurs batterie avec leurs descriptions :

ID	Warning / Error	Category	Title level A
70	Erreur	Error BAT	Battery <u>overvoltage</u> .
72	Erreur	Error BAT	Charger current failure.
73	Erreur	Error BAT	Battery voltage differs from communicated value.
76	Erreur	Error BAT	The battery reported an error.
87	Erreur	Error BAT	Battery current exceeds max battery current limit.

FIGURE 3.8 – Les codes erreurs batteries

3.5.3 Apprentissage des données PV

la figure 3.9 représente les codes qui correspondes aux erreurs PV avec leurs descriptions :

A	B	C	D
ID	Warning / Error	Category	Title level A
42	Erreur	Error PV	Solar input <u>overvoltage</u> .
43	Erreur	Error PV	Solar input <u>overcurrent</u> .
44	Erreur	Error PV	Solar isolation resistance too low.

FIGURE 3.9 – Les codes erreur PV

Conclusion

Dans ce chapitre, nous avons exposé toute la phase de conception de notre travail. Nous avons expliqué tous nos choix et détaillé toute la phase de prétraitement de données. Enfin, nous avons présenté la conception générale de notre outil d'analyse prédictive.

Le prochain chapitre concernera la réalisation de notre travail, les détails de l'implémentation ainsi que les résultats.

CHAPITRE 4

IMPLÉMENTATION ET RÉSULTATS

Dans le chapitre précédent nous avons abordé toute la phase de conception de notre travail, ainsi que la partie préparation des données.

Dans ce chapitre, nous détaillerons l'implémentation de notre solution : environnement de travail, détails des données, librairies utilisées, etc. Nous discuterons aussi les résultats que nous avons obtenus que nous comparerons aux résultats des travaux antérieurs. Pour finir nous présenterons notre outil de prédiction des pannes.

4.1 Implémentation

Dans ce qui suit nous présenterons tous les outils que nous avons utilisés afin de mettre en œuvre notre application de prédiction des pannes.

4.1.1 Environnement de développement

Tout notre travail a été développé grâce au langage de programmation Python dans sa version 3.6.8. Pour l'écriture et le débogage de nos scripts nous avons utilisé l'environnement de développement intégré Anaconda. Travailler sous Python était une évidence ; le langage a en effet fait ses preuves pour tout ce qui est pré traitement des données, apprentissage automatique de par la simplicité de sa syntaxe et les différentes librairies qu'il permet d'intégrer.

Notre apprentissage a été exécuté sur un serveur interne possédant un processeur Intel(R) Xeon(R) CPU 4 cores, 8 logical Processors de 8Go de RAM, sous Windows 7 dans une version de 64 bits.

Pour la réalisation de chaque étape (extraction des données, visualisation, apprentissage, réalisation d'interface, etc.) nous avons utilisé plusieurs bibliothèques intégrées à python. Nous avons aussi utilisé des outils de Data visualisation.

4.1.1.1 Psycopg – PostgreSQL database adapter for Python

Psycopg est l'adaptateur de base de données PostgreSQL le plus populaire pour le langage de programmation Python. Ses principales caractéristiques sont l'implémentation complète de la spécification Python DB API 2.0 et la sécurité des threads (plusieurs threads peuvent partager la même connexion). Il a été conçu pour les applications fortement multithreads qui créent et détruisent de nombreux curseurs et génèrent un grand nombre d'insertions ou de mises à jour simultanées [www.Psycopg.org, 2019].

Psycopg 2 est principalement implémenté en C en tant que wrapper libpq, ce qui permet d'être à la fois efficace et sécurisé. Il comporte des curseurs côté client et côté serveur, une communication et des notifications asynchrones et une prise en charge de COPY. De nombreux types Python sont pris en charge immédiatement et adaptés aux types de données PostgreSQL correspondants ; l'adaptation peut être étendue et personnalisée grâce à un système d'adaptation d'objets flexibles.

Nous avons créé notre extracteur de données grâce à ce Framework. Son automatisation permet de parcourir la base de données de production (monitoring) et d'en extraire le contenu disponibles.

4.1.1.2 Outil de Data visualisation

Nous avons choisi d'implémenter un outil qui nous permet de visualiser nos données. Il s'agit d'une application Web pour la visualisation de données interactif de manière très intuitif. Il nous a permis de mettre en évidence plusieurs bruits et erreurs contenus dans les données que nous avons extraites et extraire les facteurs clés provoquant certaines pannes.

pour se faire, nous avons utilisé la bibliothèque matplotlib afin d'explorer plus en détail certains attributs des données extraites. Matplotlib est une bibliothèque Python qui permet de créer des figures principalement statistiques en 2D. Pour manipuler les données nous avons également utilisé la bibliothèque Pandas qui nous a permis de manipuler les données plus facilement, grâce à sa structure DataFrame très puissante et légère pour la représentation des données.

Pour des graphiques innovantes et dynamiques, nous avons utilisé L'API ECharts. C'est un outil de visualisation JavaScript Open source, qui peut s'exécuter facilement sur un ordinateur et des appareils mobiles. Il est compatible avec la plupart des navigateurs Web modernes, tels que IE8, Chrome, Firefox, Safari, etc. ECharts s'appuie sur ZRender, un moteur de rendu graphique, pour créer des graphiques intuitifs, interactifs et hautement personnalisables [Echarts, 2019].

- a. Dashboard : Nous avons choisi d'implémenter quelques graphiques et tableaux concernant l'évolution du stock et les intégrer dans l'interface principale de l'application, pour donner des aider les experts métiers dans leurs analyses :

Evolution of inverter's installation per month

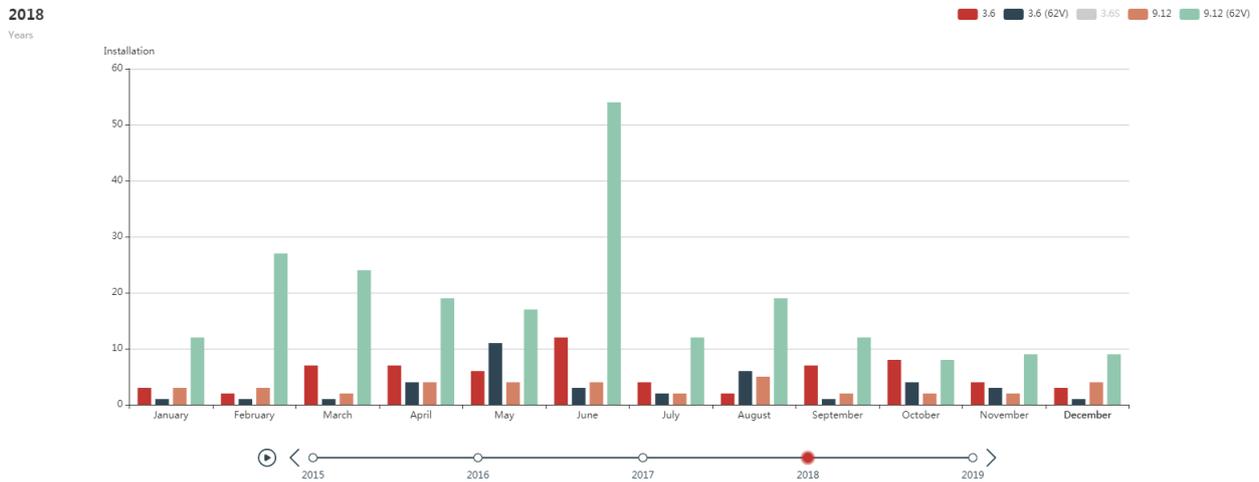


FIGURE 4.1 – Tableau de bord de l'application de visualisation des données.

The total number of inverters installed

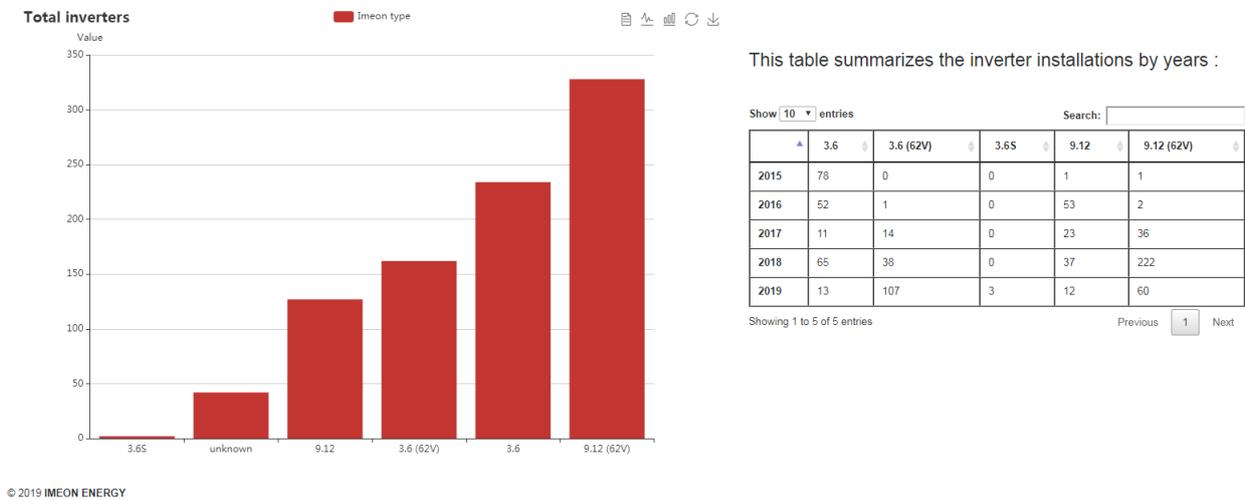


FIGURE 4.2 – graphique et tableau résumant l'évolution des installation des onduleurs.

- b. Statistiques :

Nous avons consacré un partie pour faire des statistiques dédié aux équipes de développements et de maintenances pour visualiser et analyser par la suite les résultats : s'il y a par exemple une erreur plus fréquente avec un certain type d'onduleur ou certaines version de software.

Error distribution by type of inverter

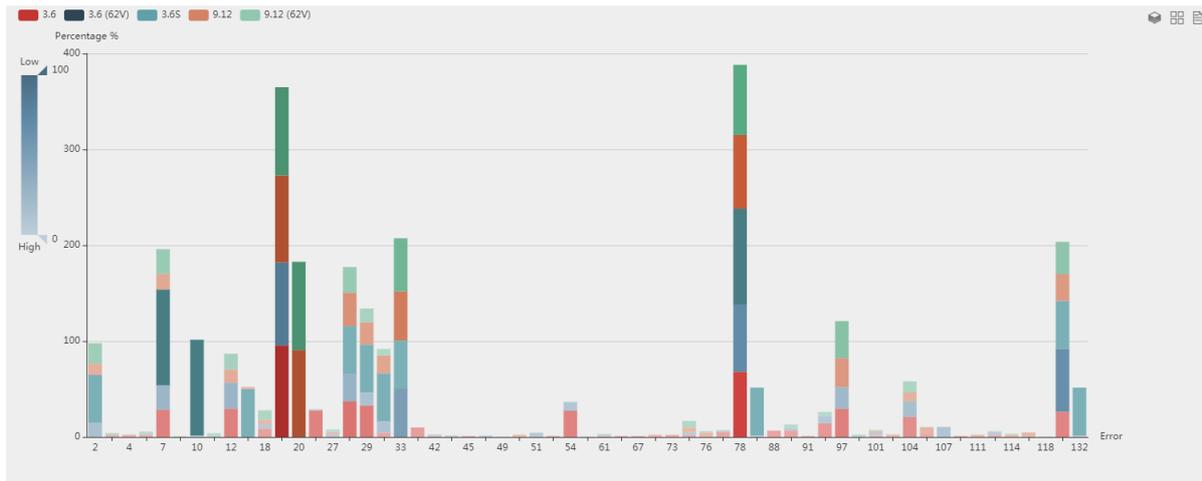


FIGURE 4.3 – Partie statistique

Summary Table

Percentage :

Show 10 entries

	3.6	3.6 (62V)	3.6S	9.12	9.12 (62V)
2	0.00	14.93	50.0	11.63	21.29
3	2.20	0.00	0.0	0.00	1.98
4	2.20	0.00	0.0	0.00	0.50
5	2.20	1.49	0.0	0.00	1.98
7	29.67	25.37	100.0	16.28	25.74
9	0.00	0.00	0.0	0.00	0.99
10	0.00	1.49	100.0	0.00	0.50
11	0.00	1.49	0.0	0.00	2.48
12	29.67	26.87	0.0	13.95	16.34
13	0.00	0.00	50.0	2.33	0.00

Showing 1 to 10 of 59 entries

FIGURE 4.4 – Tableau récapitulatif

c. Indicateurs de performance la figure 4.5 représente un indicateur de performance implémenter permettant le suivi des taux de connections des onduleurs :

Connection rate analysis

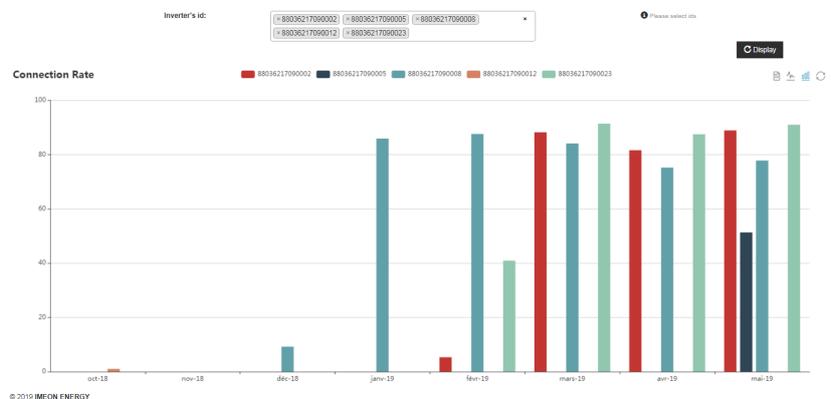


FIGURE 4.5 – Analyse des taux de connections des onduleurs

- Modèle Vue Contrôleur

Nous avons utilisé l'architecture MVC pour notre applications. Le motif est composé de trois types de modules ayant trois responsabilités différentes : les modèles, les vues et les contrôleurs :

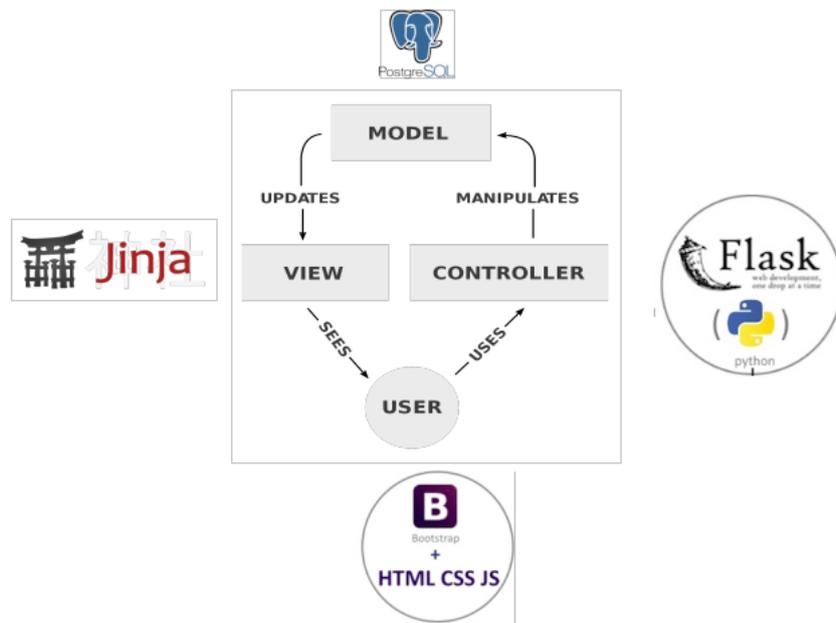


FIGURE 4.6 – Model viewer controller

Nous avons utilisé Jinja pour faciliter l’affichage des informations dans plusieurs pages, c’est un moteur de templates écrit en Python. Il est inspiré du moteur Django tout en restant plus simple dans son utilisation. Jinja charge ce patron et réalise la substitution des champs en fonction du contexte d’exécution.

4.1.1.3 Scikit-learn, librairie d’apprentissage automatique

Scikit-learn est une librairie python open source et collaborative, permettant de construire divers méthodes d’apprentissage automatique : classification, régression et clustering. Elle offre un large choix d’algorithmes [Sckitlearn, 2019b]

Nous avons choisi Scikit-learn, tout d’abord pour la qualité de documentation disponible sur internet, une communauté très active qui permet le maintien du projet et son développement. cette Api offre la facilité d’implémentation et de test des divers algorithmes. En effet Scikit-learn permet avec quelques lignes de code d’utiliser les algorithmes d’apprentissage, ainsi que de les paramétrer. Scikit-learn est la librairie à utiliser pour des projets sur courte période et qui mettent l’accent sur les données et le choix des algorithmes.

La raison principale de notre choix de cette api, est la contrainte lié à l’environnement de test embarqué (onduleur) et la capacité limité de stockage. En effet, nous avons comparer scikit learn et tensorflow, qui est une bibliothèque open-source de deep learning créée par Google.

Comme le montre la figure 4.7, contrairement à tensorflow, sickit learn ne consomme pas autant de mémoire.

Discussion : Tensorflow est principalement une bibliothèque dédiée aux deep learning. Elle nécessite la transformation des données en tensor (sorte de tableau) qui servent comme entrées des algorithmes d’apprentissage profond ; afin de pouvoir les utiliser avec des accélérateur graphique.

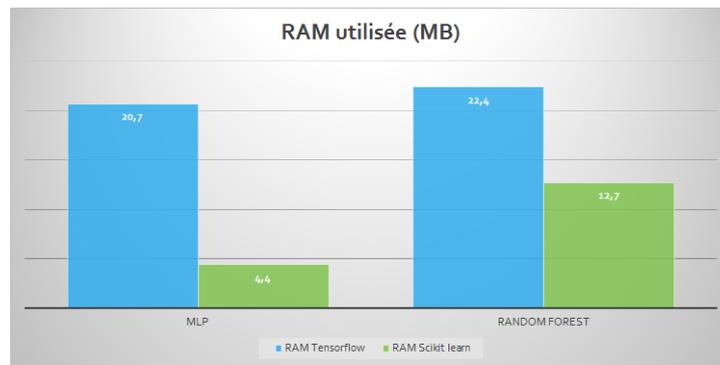


FIGURE 4.7 – Consommation de mémoire par les deux API d’apprentissage automatique.

4.1.2 Choix d’implémentation

Pour un projet de maintenance prédictive, il existe trois type d’implémentation :

Implémentation	Explication	Avantage	Inconvénient
Cloud ou data center	Apprentissage du modèle et exécution dans le cloud ou dans les serveurs du data center.	Grandes capacités de calcul, de stockage et de data.	Dépendance du réseau
Système hybride	Apprentissage du modèle dans le cloud /data center et exécution dans le système embarqué.	Désengorgement des réseaux.	Capacités de calcul et de stockage limitées en local.
Local	Système embarqué auto-apprenant qui effectue lui-même l’apprentissage de son modèle avec ses propres données, puis son exécution.	Système 100% autonome.	Coûteux - Difficultés de mise à jour.

Tableau 4.1 – Tableau comparatif des choix d’implémentation [LOUKIL, 2018]

Après études et discussion avec les équipes métiers, nous avons opté pour une solution hybride.

4.1.3 Architectures modèles prédictifs

Afin d’obtenir le meilleur système prédictif, nous avons exploré plusieurs combinaisons de paramètres relatifs à chaque algorithme d’apprentissage. Ces paramètres expérimentaux représentent l’architecture caractéristique de chaque modèle prédictif. Dans cette section nous présenterons les paramètres sur lesquels notre apprentissage a été testé et qui correspondent à l’implémentation propre à Scikit-learn de ces algorithmes.

4.1.3.1 SVM

Nous avons implémenté Support Vector Machine, avec la fonction RBF. Cette version de l’algorithme nous donne la possibilité de paramétrer deux caractéristiques C et gamma¹, qui

1. Pour pouvoir obtenir les valeurs idéales de C et de Gamma nous avons utilisé la méthode GridSearchCV.

nous permettent de régler l'hyper plan de la ligne prédictive des données. La caractéristique C joue sur la taille de l'hyper plan où se situent les droites de prédiction, pendant que γ joue sur la pente de ces droites et améliore la précision. Un trop grand γ peut mener au sur-apprentissage, pendant qu'un trop petit peut mener au sous-apprentissage et inversement pour le C [Sckitlearn, 2019c].

4.1.3.2 MLP

Afin de construire le modèle de réseau de neurones le plus optimal, nous avons au début calculé les coefficients de régression de chaque attribut par rapport au code erreur. Ensuite nous avons essayé d'inférer plusieurs modèles avec diverses combinaisons de sous-ensembles d'attributs selon leurs importances, en faisant varier le nombre de neurone et la fonction d'activation. Nous avons fixé le nombre de couches cachées à 1 ; Vu le nombre insuffisant des données pour passer au deep learning.

4.1.3.3 RandomForest Classifier

RandomForest classifieur est un algorithme d'apprentissage qui entraîne plusieurs sous-ensembles d'arbres de décision. Les paramètres expérimentaux de cet algorithme sont : [Sckitlearn, 2019a].

- *Max-features* : le nombre maximum d'attributs à tester dans un arbre.
- *N-estimators* : le nombre maximal d'arbres à construire avant de calculer la moyenne des prédictions des arbres.
- *Min-simple-leaf* : le nombre minimal de feuilles dans chaque arbre.
- *n-jobs* : le nombre de processus à exécuter en parallèle.
- *random-state* : le nombre de validations à exécuter aléatoirement avant de calculer la précision.

4.1.4 Mesures d'évaluation

Afin d'évaluer les modèles d'apprentissage inférés, nous avons choisi deux mesures : le calcul de la précision, ainsi que le taux d'erreur.

4.1.4.1 Mesure de précision

Afin de pouvoir comparer la précision de notre prédiction aux résultats des travaux antérieurs, nous avons choisi de ne pas customiser la mesure de précision. En effet, nous avons utilisé la méthode de Scikit-Learn « score » où chaque algorithme d'apprentissage implémente sa propre méthode de calcul de précision. Néanmoins, le RandomForest classifieur, le SVM et le MLP ont la même mesure de précision [Sckitlearn, 0199] qui est une maximisation de R^2 , telle que

$$R^2 = (1 - U/V) \text{ où,}$$

$$U : ((y_true - y_pred)^2).sum :$$

la somme des régressions de la valeur réelle moins la valeur prédite.

$$V : ((y_true - Y_true.mean())^2).sum() :$$

la somme de la valeur réelle moins la moyenne des valeurs réelles au carré.

4.1.4.2 Mesure d'erreur

Pour calculer l'erreur, nous avons implémenté notre propre méthode afin de prendre en considération les code erreurs que nous essayons de prédire. Ainsi, le taux d'erreur est calculé comme suit :

$$\text{Erreur} = (y_{\text{pred}} - y_{\text{test}}) - (y_{\text{pred}} - y_{\text{test}}) * 10\% / y_{\text{test}} \text{ Où,}$$

Y_{pred} : les valeurs prédites par le modèle.

Y_{test} : les valeurs de sortie réelles.

4.2 Résultats

4.2.1 Performances des algorithmes d'apprentissage

Nous avons expliqué dans la partie implémentation, les mesures d'évaluation de nos modèles inférés. Nous utiliserons ainsi deux mesures, la précision et le taux d'erreurs. Nous présenterons les performances des trois catégories de pannes séparément, étant donné que ce sont trois problèmes distincts. Afin de déterminer l'algorithme d'apprentissage le plus efficace pour le problème de prédiction de panne de chaque catégorie, nous avons implémenter les 3 algorithmes pour chaque catégorie.

4.2.1.1 Performances des prédictions des erreurs onduleurs

Nous présentons dans le tableau 4.2, les meilleurs résultats obtenus après la phase d'apprentissage et de validation par chaque algorithme de prédiction des pannes onduleurs.

Algorithme d'apprentissage	Précision	Taux d'erreur	Mémoire utilisée
MLP	0.60	1.53	79.8
SVM	0.96	0.36	14.1
RandomForest Classifier	0.79	1.07	16.6

Tableau 4.2 – Tableau comparatif des résultats de l'apprentissage sur des données des erreurs onduleurs

Analyse :

Dans le tableau 4.2 sont en effet représentés les précisions et taux d'erreurs moyens des données de test des meilleurs modèles inférés. Il n'est pas à préciser que l'algorithme SVM est celui qui enregistre la précision la plus haute et le taux d'erreur le plus bas, avec une moyenne de précision de 96% et un taux d'erreur moyen d'environ 36%. Les deux algorithmes MLP et RamndomForest ne dépassent pas les 79% de précision, avec une erreur moyen assez élevé environ 1.53 enregistré par MLP.

4.2.1.2 Performances des prédictions des erreurs batteries

Nous présentons dans le tableau 4.3, les meilleurs résultats obtenus après la phase d'apprentissage et de validation par chaque algorithme de prédiction des pannes batteries.

Algorithme d'apprentissage	Précision	Taux d'erreur	Mémoire utilisée
MLP	0.88	0.51	8.9
SVM	0.90	0.41	2.0
RandomForest Classifier	0.92	0.39	13.9

Tableau 4.3 – Tableau comparatif des résultats de l'apprentissage sur des données des erreurs batteries

Analyse :

Dans le tableau 4.3 sont en effet représentés les précisions et taux d'erreurs moyen sur les données de test des meilleurs modèles inférés concernant les pannes batteries. Dans cette catégorie l'algorithme RandomForest Classifier est celui qui enregistre la précision la plus haute et le taux d'erreur le plus bas, avec une moyenne de précision de 92% et un taux d'erreur moyen d'environ 39%. SVM vient en deuxième position avec un petit écart de 2%.

4.2.1.3 Performances des prédictions des erreurs PV

Nous présentons dans le tableau 4.4, les meilleurs résultats obtenus après la phase d'apprentissage et de validation par chaque algorithme de prédiction des pannes PV.

Algorithme d'apprentissage	Précision	Taux d'erreur	Mémoire utilisée
MLP	0.88	0.31	2.6
SVM	0.90	0.28	4.2
RandomForest Classifier	0.86	0.31	0.3

Tableau 4.4 – Tableau comparatif des résultats de l'apprentissage sur des données des erreurs PV

Analyse :

Selon le tableau 4.4, SVM a obtenue le meilleur taux de précision avec 90% et un taux d'erreur bas avec 28%.

4.2.2 Evaluation de notre travail par rapport aux travaux antérieurs

Comme mentionné dans le chapitre 2, beaucoup de travaux de recherche se sont penchés sur la maintenance prédictive grâce aux techniques d'apprentissage automatique. Dans cette section nous comparerons nos résultats en général aux résultats déjà existants.

Analyse

Nous remarquons que notre modèle prédictif des pannes PV est un peu en dessous de la moyenne de prédiction des autres modèles, alors que notre modèle prédictif des pannes onduleurs dépasse quelques résultats (2/3).

Nous précisons que nous avons exploité un volume de données plus petit que tous les travaux déjà existants. En effet, nous avons environs les 150 000 entrées pour la catégorie onduleurs, 24 000 pour les pannes batteries et uniquement 4600 pour les PV ; ce qui explique les résultats.

Auteurs	Meilleure précision (%)
[Paolanti and Romeo, 2018]	95
[RABENORO et al., 2015]	97
[WANG et al., 2017]	95
Notre modèle de prédiction des pannes Onduleurs	96
Notre modèle de prédiction des pannes Batterie	92
Notre modèle de prédiction des pannes PV	90

Tableau 4.5 – Comparaison des précisions de nos modèles prédictifs avec les résultats des travaux antérieurs

A noter aussi que, pour notre approche, nous avons choisi d'essayer d'inférer un modèle spécifique pour chaque catégorie (pannes onduleurs, batterie ou pv).

4.2.3 Module Prédictif

au sein de l'onduleur, on trouve un module principal appelé "Le coremod" qui regroupe tous les modules assurant le bon fonctionnement de l'onduleur. Grâce à ces modules qui sont conçus avec les toutes dernières technologies, l'onduleur est de plus en plus autonome et intelligent. Le coremod gère la communication avec des composantes externe comme par exemple le module BMS "Battery Manager System" qui s'occupe de la gestion des batteries.

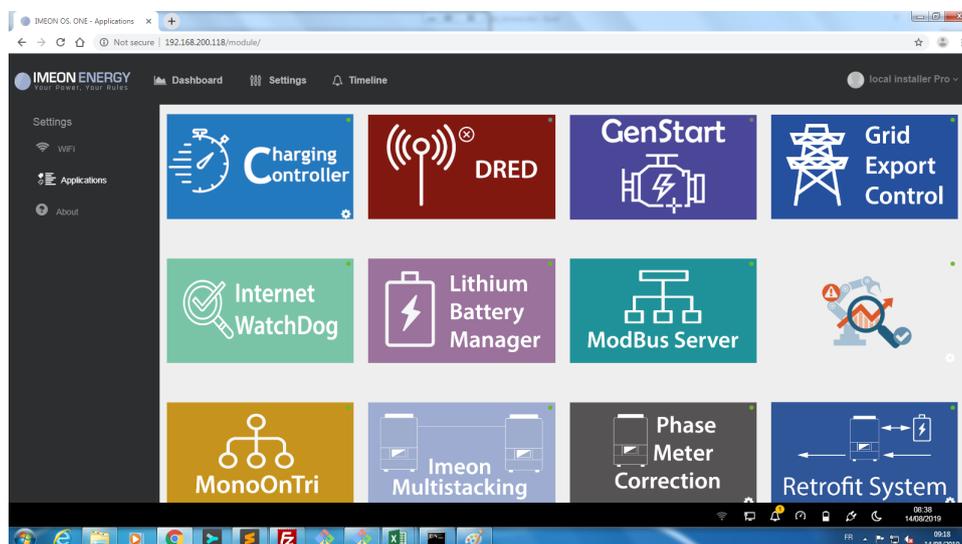


FIGURE 4.8 – L'ensemble des modules / applications internes d'onduleur.

Chaque module possède sa propre base des données, permettant un stockage permanent des paramètres et des données.

Nous avons implémenté une version Beta du module prédictif, qui se lance chaque semaine avec une tâche Cron.

Cette version sera tester au fur est à mesure dans les onduleurs des test, avant d'établir une version finale du module.

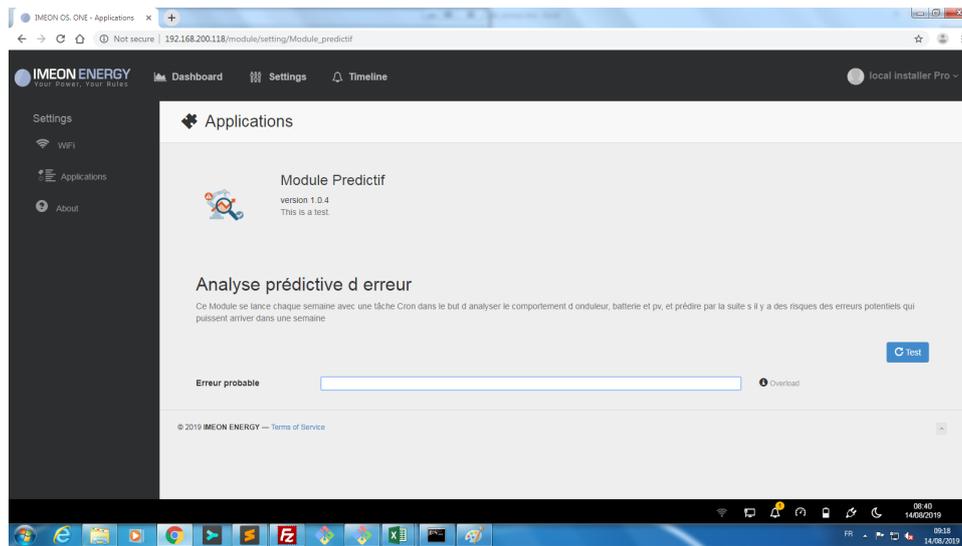


FIGURE 4.9 – Le Module prédictif.

Conclusion

Dans ce chapitre nous avons présenté tous les détails relatifs à notre implémentation. Nous avons présenté nos résultats, et avons fait une étude comparative de ces derniers entre eux et avec les résultats des travaux antérieurs. Enfin, nous avons présenté en général nos modèles prédictifs, ce qui nous permet ainsi de clôturer notre travail.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

De nombreuses entreprises maintiennent leurs équipements et leurs machines en mode curatif. On répare quand la panne survient. Ce qui génère des surcoûts, des retards de production liés aux arrêts de chaîne et, par conséquent, une productivité en berne [Thuemmler and Bai, 2017]. La maintenance industrielle devient partie prenante de l'usine du futur et non plus une simple fonction annexe. Une solution pour contourner cette problématique est d'adopter une démarche de maintenance prédictive.

Dans ce contexte, IMEON Energy, le leader français de la fabrication des onduleurs intelligents connectés, opte pour un projet de maintenance prédictive de la gamme onduleur IMEON.

Le fruit de notre travail se positionne dans cet écosystème en expansion. Nous nous sommes penchées sur l'exploitation des données des industrielles dans le but de réaliser un outil de prédiction des pannes basée sur des algorithmes d'apprentissage automatique. Grâce à ces algorithmes et dans le cadre de la maintenance prédictive, les coûts de maintenance et arrêts non programmés perturbant la production devraient être significativement réduits.

Ce mémoire a été réalisé dans le but de présenter toutes les étapes effectuées pour la réalisation de notre projet. Dans un premier temps, nous avons étudié l'environnement actuel et les tendances du marché industrielle afin de comprendre les besoins actuels.

Ensuite, nous avons effectué un long travail de recherche sur le domaine de la maintenance prédictive, répertoriant ainsi les différents travaux significatifs et performants. Nous avons présenté également les techniques de prétraitement et d'exploration des données nécessaires.

La phase la plus fastidieuse et la plus longue de notre projet était le prétraitement des données. En effet, une fois que nous avons collecté les données, nous devons les analyser, normaliser, trier, etc, afin de créer un dataset exploitable qui nous permettrait d'inférer le système de prédiction le plus précis qui soit. Nous avons ensuite présenté notre approche pour la prédiction des pannes, qui nous permet d'exploiter efficacement les données :

Dans un premier temps, nous avons choisi d'effectuer notre apprentissage automatique sur les données avec trois algorithmes : SVM, MLR et RandomForest. Dans notre approche, nous avons choisi d'inférer des modèles qui essaient de prédire le code erreur exact, et ainsi, nous avons choisi d'implémenter une variante de ces algorithmes qui permet d'effectuer des classification plutôt qu'une régression.

L'algorithme qui a donné le meilleur taux de précision et le taux d'erreur minimal est le SVM : 96% de précision et 36% d'erreur pour les erreurs onduleurs.

L'écart des taux précision et d'erreur entre les prédictions des pannes onduleurs, batterie et PV est lié à la qualité des données exploitées. En effet, bien que d'apparence le problème de prédiction une panne onduleur (surcharge, courant de fuite, ...) se veut plus complexe que la prédiction des autres pannes, néanmoins lors de la phase d'exploration des données, nous avons remarqué que les données des onduleurs étaient qualitativement bien plus diversifiées que les autres données.

La dernière étape de notre projet était la réalisation de l'outil prédictif en local, notre outil offre la possibilité de prédire les pannes avant une semaine..

Nos résultats de prédiction des pannes sont satisfaisants comparés aux résultats des travaux qui se penchent sur le même sujet de maintenance prédictive.

Nous tenons à rajouter que les taux de précision de nos systèmes prédictifs sont plus que satisfaisants étant donné que nous essayons de prédire une classe d'erreur réelle avant une semaine de son arrivé, ce qui peut laisser une marge d'erreur assez souple par rapport à d'autres problèmes de prédiction.

Comme tout travail le nôtre peut connaître des améliorations et des extensions. La suite de ce projet pourrait se faire dans différentes directions :

- L'extension des prédictions à d'autre catégorie des pannes.
- Mettre en place un système d'identification des causes des pannes.
- La création d'un outil de conseil pour mieux conseiller le client et l'aider à prendre des décisions vis à vis la maintenance de l'onduleur, sans avoir recours aux service de maintenance.

[ref,] " " .

- [Abdallah, 2015] Abdallah, S. (2015). *"Using Text Mining To Analyze Real Estate Classifieds"*. Library of Congress Cataloging-in-Publication Data , British University in Dubai, United Arab Emirates and University of Edinburgh, United Kingdom.
- [Absolunet's, 2017] Absolunet's (2017). "10 ecommerce trends for 2017" livre digital en ligne. <http://10ecommercetrends.com/> en ligne ; accédé 28 Février 2017.
- [Adams et al., 2017] Adams, L., Burkholder, E., and Hamilton, K. (2017). *"Micro-Moments : Your Guide to Winning the Shift to Mobile"*. Google.
- [Alexa, 2017] Alexa (2017). "find, reach, and convert your audience with marketing that works". <http://www.alexa.com/> ; en ligne accédé 25 Mai 2017.
- [analyticsvidhya, 2015] analyticsvidhya (2015). "data visualization resource". <https://www.analyticsvidhya.com/blog/2015/05/data-visualization-resource/> ; accédé 02 Mars 2017.
- [Annie Laforet, 2015] Annie Laforet, G. M. e. J.-M. P. (2015). *"Statistique"*. Institut de Statistique - Université Libre de Bruxelles - Belgique.
- [Aubay, 2016] Aubay (2016). "text mining". <http://www.aubay.com/wp-content/uploads/2016/04/Regard-Aubay-Text-Mining-V0.1.pdf> ; accédé 02 Mars 2017.
- [Ayadi, 2013] Ayadi, I. (2013). « *Optimisation des politiques de maintenance préventive dans un cadre de modélisation par modèles graphiques probabilistes.* ». PhD thesis, l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux. Paris, France.
- [Baheti and Gill, 2011] Baheti, R. and Gill, H. (2011). Cyber-physical systems.
- [BusinessInsider, 2016] BusinessInsider (2016). "e-commerce online payment technologies overview trends" publié en 26 décembre 2016 par andrew meola. <http://www.businessinsider.com/ecommerce-payment-technologies-overview-trends-2016-10?IR=T> en ligne ; accédé 28 Février 2017.
- [Caciandone and Chiru, 2016] Caciandone and Chiru, C.-G. (2016). "using machine learning to generate predictions based on the information extracted from automobile ads ". *University Politehnica Bucharest*.
- [Celeux et al., 2006] Celeux, G., Corset, F., Lannoy, A., and Ricard, B. (2006). *" Designing a bayesian network for preventive maintenance from expert opinions in a rapid and reliable way. 849–856 "*. Reliability Engineering System Safety.
- [chang Chih-Chung and Chih-Jen, 2013] chang Chih-Chung and Chih-Jen, L. (2013). *"LIBSVM : A Library for Support Vector Machines"*. Department of Computer Science National Taiwan University, Taipei, Taiwan Last updated : March 4, 2013.

- [Chrislb, 2016] Chrislb, D. (2016). "neurone artificiel" , accédé le 25 février 2017. https://commons.wikimedia.org/wiki/File%3AArtificialNeuronModel_francais.png.
- [Continuations, 2011] Continuations (2011). "a brief history of disruption classified ads" publié en 29 juillet 2011 par albert wenger. <http://continuations.com/post/8214667310/a-brief-history-of-disruption-classified-ads> en ligne ; accédé 28 Février 2017.
- [Cotton, 2010] Cotton, N. J. (2010). "a neural network implementation on embedded systems". *Auburn, Alabama*.
- [Courtois and Delsarte, 2003] Courtois, P. and Delsarte, P. (2003). "*On the optimal scheduling of periodic tests and maintenance for reliable redundant components.66–72*". Reliability Engineering System Safety 91.
- [CultureMedia, 2012] CultureMedia (2012). "social commerce : Est-ce l'avenir du e-commerce?" publié en 8 novembre 2012. <http://www.culturecrossmedia.com/social-commerce-est-ce-lavenir-du-e-commerce-2/> en ligne ; accédé 28 Février 2017.
- [Delcacho, 2010] Delcacho, M. (2010). *SVM , KPPV, génétique*. PhD thesis, "University".
- [DUNCOMBE, 2010] DUNCOMBE, C. (2010). fautes orthographe commerce internet. rapport bbc news. <http://www.slate.fr/lien/41129/fautes-orthographe-commerce-internet> en ligne accédé 28 Mai 2017.
- [Echarts, 2019] Echarts (2019). <https://echarts.apache.org/en/feature.html> en ligne ; accédé 29 Juillet 2019.
- [EcommerceFoundation, 2016] EcommerceFoundation (2016). *Ecommerce Benchmark Retail Report 2016*. Ecommerce Foundation Raadhuisstraat 22 1016 DE AMSTERDAM The Netherlands.
- [Econsultancy, 2013] Econsultancy (2013). "*The Realities of Online Personalisation*". Econsultancy digital report.
- [edutechwiki, 2016a] edutechwiki (2016a). "half the webs traffic comes from bots" accédé 05 février 2017. <http://www.computerworld.com/article/3070058/big-data/half-the-webs-traffic-comes-from-bots.html>.
- [edutechwiki, 2016b] edutechwiki (2016b). "web scraping" en ligne ; accédé le 05 février 2017. http://edutechwiki.unige.ch/fr/Web_scrapingD.C3.A9finitio/.
- [Fiwiclassifieds, 2013] Fiwiclassifieds (2013). "the history of classified ads" publié en 10 aout 2013 par michelle. <https://www.fiwiclassifieds.com/blog/the-history-of-classified-ads/> en ligne ; accédé 28 Février 2017.
- [FORESTIER, 2016] FORESTIER, R. L. (2016). *Développement API communication inter-objet (Smart-Grid/ M2M)*. PhD thesis, Université de Bretagne Occidentale.
- [Frew and G., 2003] Frew and G., J. (2003). "estimating the value of apartment buildings.". "*Journal of Real Estate Research* ".
- [future science, 2016] future science (2016). "informatique-comprendre-facebook" publié en 22 janvier 2016. <http://www.futura-sciences.com/tech/dossiers/informatique-comprendre-facebook-1360/page/20/> en ligne ; accédé 20 Février 2017.
- [Gaffiot, 1934] Gaffiot, F. (1934). *Dictionnaire latin-français*. Hachette.
- [GoogleTrends, 2017] GoogleTrends (2017). "google trends for algeria". <https://trends.google.fr/trends/explore?cat=18geo=DZ> ; en ligne accédé 25 Mai 2017.
- [Guillaume, 2010] Guillaume, S.-J. (2010). "méthodes de prévision et formules excel". <https://www.lokad.com/fr/methodes-previsions-et-formules-excel> ; accédé 15 Mars 2017.
- [Gérard, 1989] Gérard, S. (1989). "*L'intelligence artificielle et le langage*". Hermès, Paris.

- [HuffpostAlgeria, 2016] HuffpostAlgeria (2016). "l'algérie pas prête pour le passage au commerce en ligne, selon les nations unies" publié en 03 mai 2016. http://www.huffpostmaghreb.com/2016/05/03/algerie-e-commerce-pas-pr_n_829126.html en ligne; accédé 28 Février 2017.
- [Immo2.pro, 2016] Immo2.pro (2016). "les 8 technologies et tendances qui transforment l'immobilier" publié en 22 janvier 2016. <https://immo2.pro/les-8-technologies-et-tendances-qui-transforment-limmobilier-2016/> en ligne; accédé 28 Février 2017.
- [KeyDifferences, 2015] KeyDifferences (2015). "difference between e-commerce and m-commerce" publié le 16 octobre 2015 par surbhi s. <http://keydifferences.com/difference-between-e-commerce-and-m-commerce.html> en ligne; accédé 28 Février 2017.
- [KUKICH, 1992] KUKICH, K. (1992). *"Techniques for automatically correcting words in text. ACM Computing Surveys"*. Bellcore, 44.5 South Street, Morristown, NJ 07962-1910.
- [Kumar et al., 2015] Kumar, A., Anil, B., Anand, C., Aniruddha, S., and Kumar, R. (2015). *"Machine Learning Approach to Predict Real Estate Prices"*. Discovery, 2015, 44(205), 173-178.
- [Laforet et al., 1990] Laforet, A., Mélard, G., and Pasteels, J.-M. (1990). *"VERS UN SYSTÈME EXPERT DE PRÉVISION ET DE STATISTIQUE ÉCONOMIQUE"*. Institut de Statistique - Université Libre de Bruxelles - Belgique.
- [Lecun, 2010] Lecun, Y. (2010). "directeur du laboratoire d'ia de facebook, professeur d'informatique à l'université de new york.". <https://www.youtube.com/watch?v=QxRdWIIyDG8>; accédé 18 Février 2017.
- [LeMonde, 2016] LeMonde (2016). "une photo de mark zuckerberg ravive la peur de la réalité virtuelle" publié en 22 février 2016 par leloup, d. et tual, m. http://www.lemonde.fr/pixels/article/2016/02/22/une-photo-de-mark-zuckerberg-ravive-la-peur-de-la-realite-virtuelle_4869605_4408996.html en ligne; accédé 28 Février 2017.
- [Levenshtein, 1965] Levenshtein, V. (1965). *Distance d'édition minimale*. PhD thesis, "Université d'État de Moscou".
- [Li, 2018] Li, Z. (April 2018). *"Deep Learning Driven Approaches for Predictive Maintenance. A Framework of Intelligent Fault Diagnosis and Prognosis in the Industry 4.0 Era"*. PhD thesis, Norwegian University of Science and Technology, Trondheim.
- [LibertéAlgérie, 2016] LibertéAlgérie (2016). "e-commerce c'est quoi et ou en est l'algérie" publié en 16 décembre 2016 par mohamed aniss amrah. <http://www.liberte-algerie.com/soleil-hec-red-num-liberte/e-commerce-cest-quoi-et-ou-en-est-lalgerie-260520> en ligne; accédé 28 Février 2017.
- [LISTIANI, 2009] LISTIANI, M. (2009). *Support Vector Regression Analysis for Price Prediction in a Car Leasing Application*. PhD thesis, "Hamburg University of Technology".
- [Lorentz, 1998] Lorentz, F. (1998). *Rapport sur le commerce électronique -Addendum*. Ministère de l'économie, des Finances et de l'industrie français.
- [LOUKIL, 2018] LOUKIL, R. (2018). "le défi de l'intégration de l'ia dans les systèmes embarqués". *L'usine Nouvelle*.
- [McKinseyCompany, 2016] McKinseyCompany (2016). *"Automotive revolution" -perspective towards 2030 How the convergence of disruptive technology-driven trends could transform the auto industry*. Advanced Industries.
- [MillwardBrown, 2017] MillwardBrown (2017). "the definitive guide to how connectivity is reshaping consumers, connections, content and commerce" publié en 2017 par millward brown. http://connectedlife.tnsglobal.com/?utm_source=kmbutm_medium=webutm_campaign=Connected-Life-2016 en ligne; accédé 28 Février 2017.

- [Seiersen, 2006] Seiersen, N. (2006). *"Systèmes d'information logistique"*. Revue technique de l'ingénieur.
- [SimilarWeb, 2017] SimilarWeb (2017). "market intelligence solutions to understand, monitor and win your market". <https://www.similarweb.com/>; en ligne accédé 25 Mai 2017.
- [Statista, 2017] Statista (2017). "e-commerce worldwilde" publié par statista the statistics portal. <https://www.statista.com/outlook/243/100/e-commerce/worldwidemarket-revenue> en ligne; accédé 28 Février 2017.
- [ThinkWithGoogle, 2015] ThinkWithGoogle (2015). "holiday shopping trends 2015 : Three predictions for retailers" publié en 18 novembre 2015 par matt lawson. <https://www.thinkwithgoogle.com/qs/documents/958/holiday-shopping-trends-2015-interactive.pdf> en ligne; accédé 28 Février 2017.
- [ThinkWithGoogle, 2016] ThinkWithGoogle (2016). "marketing insights mobile first media plans" publié en décembre 2016 par joy, j. et zmuda, n. <https://www.thinkwithgoogle.com/marketing-resources/data-measurement/marketing-insights-mobile-first-media-plans-2017/> en ligne; accédé 28 Février 2017.
- [Thuemmler and Bai, 2017] Thuemmler, C. and Bai, C. (2017). *"Health 4.0 : How Virtualization and Big Data are Revolutionizing Healthcare"*. Springer.
- [UE, 2013] UE (2013). "handbook on residential property prices indices (rppis)". *Luxembourg : Publications Office of the European Union*, page 179.
- [UN, 1998] UN (1998). *"Loi type de la CNUDCI sur le commerce électronique Guide pour son incorporation et 1996 tel qu'adopté en 1998"*. United Nations Publication.
- [Vanhoucke et al., 2011] Vanhoucke, V., Senior, A., and Mao, M. Z. (2011). *Improving the speed of neural networks on CPUs*. PhD thesis, Deep Learning and Unsupervised Feature Learning Workshop. Google, Mountain View, CANADA.
- [WANG et al., 2017] WANG, Z., ZHANG, M., WANG, D., SONG, C., LIU, M., LI, J., LOU, L., and LIU, Z. (2017). *"Failure prediction using machine learning and time series in optical network"*. PhD thesis, State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China. China Mobile Communications Corporation, Beijing 100033, China 2017 Optical Society of America.
- [Wearesocial and Hootsuite, 2017] Wearesocial and Hootsuite (2017). "digital, social, mobile : Les chiffres 2017" publié le 25 janvier 2017 par nathan blaison. <https://wearesocial.com/fr/blog/2017/01/digital-social-mobile-les-chiffres-2017> en ligne; accédé 28 Février 2017.
- [Weyer et al., 2015] Weyer, S., Schmitt, M., Ohmer, M., and Gorecky, D. (2015). "towards industry 4.0 - standardization as the crucial challenge for highly modular, multi-vendor production systems". *IFAC-PapersOnLine*.
- [Wood, 2005] Wood, R. (2005). "a comparison of uk residential house price". *Indices in Real Estate Indicators and Financial Stability Bank for International Settlements, Washington DC : The International Monetary Fund*, pages 212–227.
- [WU et al., 2009] WU, D., J., HSU, C., C., CHEN, and C., H. (2009). *An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications*. PhD thesis, "University".
- [www.imeon energy.com, 2019] www.imeon energy.com (2019). <http://www.imeon-energy.com> en ligne; consulté Mai 2019.
- [www.iri lyon.com, 2019] www.iri lyon.com (2019). <https://www.iri-lyon.com/nos-domaines/maintenance-industrielle> en ligne; consulté Mai 2019.

- [www.mps maintenance.fr, 2019] www.mps maintenance.fr (2019). [http ://www.mps-maintenance.fr/maintenance-predictive/](http://www.mps-maintenance.fr/maintenance-predictive/) en ligne ; consulté Juin 2019.
- [www.Psycopg.org, 2019] www.Psycopg.org (2019). [http ://initd.org/psycopg/docs/](http://initd.org/psycopg/docs/) en ligne ; accédé 29 juillet 2019.
- [ZedNet, 2015] ZedNet (2015). " big data : quel intérêt pour l'analyse prédictive?". [http ://www.zdnet.fr/actualites/big-data-quel-interet-pour-l-analyse-predictive-39824666.htm](http://www.zdnet.fr/actualites/big-data-quel-interet-pour-l-analyse-predictive-39824666.htm) ; accédé 01 Mars 2017.
- [Zhang and LeCun, 2016] Zhang, X. and LeCun, Y. (2016). "text understanding from scratch". *"Facebook Research "*.
- [ZILIO and LUGIEZ, 2002] ZILIO, S. D. and LUGIEZ, D. (2002). *Fondements de l'Interrogation des Données Semi Structurées*. PhD thesis, Laboratoire d'Informatique Fondamentale de Marseille.

