



الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية العليا للتكنولوجيا والهندسة – عنابة

ECOLE NATIONALE SUPERIEURE DE TECHNOLOGIE ET D'INGENIERIE – ANNABA

Département Génie Industriel

MEMOIRE

En vue d'obtention du diplôme de MASTER

Domaine : Science et Technologie

Filière : Génie Industriel

Spécialité : Maintenance et Fiabilité des Systèmes Industriels

Présenté par

HAFSA DAKKICHE

FATTOUME SEMMOUN

CONCEPTION D'UN MODÈLE PRÉDICTIF POUR LA DÉTECTION ET LA PRÉVENTION DE QUALITÉ DE SOUDAGE : VERS UNE MAINTENANCE INTELLIGENTE.

Encadré par

Dr. Abdallah BOUZITOUNA

ENSTI Annaba

Co-encadré par

Dr. Ammar AYAD

ENSTI Annaba

Membres du jury :

Dr. Nour Elislem KARABADJI

Président

ENSTI-Annaba

Dr. Youcef ABIDI

Examineur

ENSTI-Annaba

Dr. Tarek KEBABSA

Examineur

ENSTI-Annaba

Année 2025

Remerciement

Avant toute chose, nous tenons à remercier Allah, le Tout-Puissant, de nous avoir donné la force, la patience et la volonté nécessaires pour mener ce travail à son terme. Sa grâce nous a guidés à travers les moments de doute et de difficulté.

Nos remerciements s'adressent ensuite à nos encadrants de mémoire, Monsieur BOUZITOUNA Abdallah et Monsieur AYAD Ammar. Nous les remercions pour leurs directions scientifiques, leur disponibilité constante et leurs conseils avisés qui nous ont guidés tout au long de ce projet.

Nous souhaitons également remercier les membres du jury, d'avoir accepté d'évaluer ce mémoire. Leurs relectures attentives et leurs remarques pertinentes ont grandement contribué à enrichir notre réflexion.

Nous exprimons notre profonde gratitude à l'entreprise EMB-Azzaba de nous avoir accueillis durant notre stage. Nous remercions tout particulièrement notre tuteur de stage, Monsieur BOUGUELAA Marouane, pour son encadrement, sa confiance et le temps précieux qu'il nous a consacré. Nos remerciements vont également à l'ensemble du personnel pour son accueil chaleureux et sa collaboration.

Notre reconnaissance va aussi à l'ensemble du département de Génie Industriel pour leur soutien durant ces trois années.

Nous n'oublions pas nos collègues et amis de promotion, pour leur soutien moral, les discussions stimulantes et l'excellente ambiance de travail qui a régné entre nous.

Dédicace

Du plus profond de mon cœur, je dédie ce mémoire à ma famille. À mes parents, mon frères MOHAMED et ma sœur SONDESE, pour leur patience infinie, leurs sacrifices silencieux et leur amour inconditionnel. Vous êtes mon ancre et ma fierté.

Une gratitude immense et toute particulière se tourne vers ma chère tante (3amti) :D.R. Plus qu'une tante, tu as été une seconde mère, un pilier et un refuge dans les moments de doute. Tes encouragements ont éclairé mon chemin. Ce succès est aussi le tien.

Un grand remerciement à ma sœur et ma collègue SEMMOUNE FATOUME

Une grande dédicace à l'ingénieur responsable de FAB LAB ZAHRA ZARDAZI

Merci d'être qui vous êtes.

Hafsa Dakkiche

Dédicace

الحمد لله الذي بنعمته تتم الصالحات، والصلاة والسلام على سيدنا محمد، خير هادٍ ومعلم.

أهدي هذا العمل إلى:

الله تعالى، الذي بفضلِهِ وتوفيقِهِ أكملت هذا المشوار، فله الحمد أولاً وآخرًا.

لنفسي الصبورة التي تحملت التعب والسهر، وواصلت رغم كل التحديات، فكان هذا الإنجاز ثمرة صبرك وإصرارك

والديّ الغاليين، اللذين كانا نور دربي وسندي في كل خطوة، فلهما كل التقدير والحب.

إخوتي وأختي (حسني، عبد الجبار، هيثم، سيفو، وبثينة)، شكرًا لوجودكم الدائم بجانبني، ولحبكم الذي كان وقودًا

لتحملي وتقديمي.

صديقتي العزيزات (فريال صالح، فريال بوشليق، زينب، منى، إيمان، أميمة، يسرى، آية، ميساء)، لأنكن جعلتن

رحلتي أكثر جمالًا بصادقتكن ونكرياتنا التي لا تنسى.

زميلتي حفصة دقيش، شريكة الجهد والنجاح، فبدون تعاونك وإخلاصك ما كنا لنصل إلى هذه النتيجة.

شكرًا لكل من ساهم في رحلتي، ولو بكلمة طيبة أو دعم معنوي، فأنتم جزء من هذا الإنجاز. وَمَا تَوْفِيقِي إِلَّا بِالْأ

لَّهِ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ

Semmoun Fattoume

Résumé

Ce mémoire de Master est consacré à la conception et à l'évaluation d'un modèle prédictif de classification de la qualité des soudures des boîtes métalliques. Face à la nécessité d'une détection fiable et automatique des défauts, la méthodologie CRISP-DM est adoptée pour ce constat. Des données des capteurs ainsi des autres simulées, intégrant des bruits pour refléter des conditions réelles de production, ont servi de base à l'entraînement des modèles. Plusieurs algorithmes d'apprentissage automatique de la bibliothèque Scikit-learn, incluant des approches reconnues pour des applications similaires ainsi que d'autres explorées spécifiquement pour ce travail, ont été évalués pour une tâche de classification multiple (défectueuse, médiocre, bonne, excellente). La mode d'évaluation retenue dans le cadre du présent travail repose fondamentalement sur le score F1 macro, ce qui est tout à fait pertinent pour le type de classification mis en œuvre. Les résultats de cette étude révèlent que le modèle Naïve Bayes a donné la meilleure performance, apparaissant comme une véritable solution d'amélioration du contrôle qualité concernant la fabrication des boîtes métalliques.

Mots-clés : Apprentissage Automatique, Qualité de Soudure, Boîtes Métalliques, Modélisation Prédictive, Naïve Bayes, Classification multiple, CRISP-DM, Contrôle Qualité.

Abstract

This Master's thesis focuses on the design and evaluation of a predictive model for classifying the weld quality of metallic cans. Addressing the need for reliable and automated defect detection, this study adopted the CRISP-DM methodology. Simulated sensor data, incorporating noise to reflect real-world production conditions, served as the basis for model training. Several machine learning algorithms from the Scikit-learn library, including approaches known for similar applications and others specifically explored for this work, were evaluated for a multi-class classification task (defective, good, excellent, mediocre). The evaluation primarily relied on the macro F1-score, relevant for this type of classification. The results demonstrate that the Naive Bayes model provided the best performance among the tested classifiers, proving to be a promising solution for enhancing quality control in metallic can manufacturing.

Keywords: Machine Learning, Weld Quality, Metallic Cans, Predictive Modeling, Naive Bayes, Multi-class Classification, CRISP-DM, Quality Control.

ملخص:

هذا البحث في درجة الماجستير مخصص لتصميم وتقييم نموذج تنبؤي لتصنيف جودة لحامات العلب المعدنية. يبرز البحث منهجية CRISP-DM لضمان كشف تلقائي موثوق للعيوب. تم استخدام بيانات من المستشعرات وأخرى محاكاة، مع إدخال ضوضاء لتعكس ظروف الإنتاج الحقيقية، كأساس لتدريب النماذج. تم تقييم عدة خوارزميات تعلم آلي من مكتبة Scikit-learn، تتضمن تقنيات معروفة لتطبيقات مشابهة وأخرى تم استكشافها خصيصًا لهذا البحث، لتصنيف متعدد المستويات (معيب، متوسط، جيد، ممتاز). تم الاعتماد على معيار F1 macro لتقييم النماذج، وهو معيار مناسب لهذا النوع من التصنيف. أظهرت النتائج أن نموذج Naïve Bayes قدم أفضل أداء، مما يجعله حلًا فعالًا لتحسين مراقبة الجودة في تصنيع العلب المعدنية.

كلمات مفتاحية : تعلم الآلة، جودة اللحام، العلب المعدنية، النمذجة التنبؤية، بايز الساذج، تصنيف متعدد الفئات، CRISP-DM، مراقبة الجودة.

Table des matières

INTRODUCTION GENERALE.....	1
CHAPITRE 1 : PRESENTATION D'ENTREPRISE ET CONTEXTE INDUSTRIELLE	3
1.1 PRESENTATION D'ENTREPRISE	3
1.2 MOTIVATION	3
1.3 PROBLEMATIQUE	4
CONCLUSION.....	4
CHAPITRE 2 : ÉTAT DE L'ART ET REVUE BIOGRAPHIQUE.....	5
2.1 LE SOUDAGE PAR RESISTANCE A LA MOLETTE : PROCESSUS, CONTROLE QUALITE ET DEFAULTS.....	6
2.2 METHODES TRADITIONNELLES DE CONTROLE QUALITE ET LEURS LIMITES	6
2.3 L'INTELLIGENCE ARTIFICIELLE ET SON IMPACT SUR L'AMELIORATION DE LA QUALITE DU SOUDAGE PAR RESISTANCE A LA MOLETTE	7
2.3.1. SURVEILLANCE DU PROCESSUS DE SOUDAGE A L'AIDE DE CAPTEURS	7
2.3.2. MODELES DE MACHINE LEARNING POUR LA PREDICTION DE LA QUALITE DE SOUDURE.....	7
2.3.3. DETECTION D'ANOMALIES ET DE DEFAULTS EN TEMPS REEL	8
2.4 INDUSTRIE 4.0 DANS LES SYSTEMES DE SOUDAGE PAR RESISTANCE A LA MOLETTE.....	8
2.4.1. APPLICATIONS DE L'IA POUR LA MAINTENANCE DES EQUIPEMENTS DE SOUDAGE PAR RESISTANCE A LA MOLETTE	8
CONCLUSION.....	8
CHAPITRE 3 : CONCEPTION D'UN MODELE PREDICTIF DE QUALITE	10
3.1 METHODE CRISP-DM POUR UNE APPROCHE STRUCTUREE	11
3.1.1. ADOPTION DE LA METHODOLOGIE CRISP-DM	11

3.2 CONCEPTION ET ELABORATION D'UN CORPUS DE DONNEES SIMULE POUR L'ANALYSE MULTI-CLASSE DE QUALITE	11
3.2.1. SYSTEME DE SOUDAGE DE REFERENCE ET GRANDEURS SIMULEES.....	12
3.2.2. PROTOCOLE DE CREATION DES SIGNAUX CAPTEURS BRUTES	12
3.2.3. ETABLISSEMENT DES CRITERES D'ETIQUETAGE POUR LA QUALITE MULTI-CLASSE	13
3.2.4. DESCRIPTION DU JEU DE DONNEES FINAL.....	13
3.3 CONDITIONNEMENT DES DONNEES POUR LA MODELISATION PREDICTIVE.....	15
3.3.1. SELECTION ET JUSTIFICATION DES VARIABLES EXPLICATIVES	15
3.3.2. TRANSFORMATION DES DONNEES : MISE A L'ECHELLE PAR STANDARDISATION	16
3.4 ARCHITECTURE DES MODELES DE CLASSIFICATION MULTIPLE ET PARAMETRAGE DES EXPERIMENTATIONS	16
3.4.1. ALGORITHMES DE CLASSIFICATION	16
3.4.2. SCHEMA D'APPRENTISSAGE ET DE VALIDATION.....	17
3.4.3. CHOIX DES INDICATEURS D'EVALUATION POUR LA CLASSIFICATION MULTIPLE	18
CONCLUSION.....	19
CHAPITRE 4 : RESULTATS ET DISCUSSION.....	20
4.1 RESULTATS PAR MODELE	21
4.1.1. MODELE DE REGRESSION LOGISTIQUE	21
4.1.2. MODELE DE LES K PLUS PROCHES VOISINS.....	22
4.1.3. MODELE ARBRE DE DECISION	23
4.1.4. MODELE RESEAU DE NEURONES – MLP	24
4.1.5. MODELE BAYES NAÏF GAUSSIEN	25
4.2 COMPARAISON DES MODELES	26

4.2.1.	RESULTATS COMPARATIFS	26
4.2.2.	ANALYSE COMPARATIVE	27
4.3	CHOIX FINAL DE MODELE	28
	CONCLUSION	28
	REFERENCES	30

Liste des tableaux

Table 3.1. Aperçu des statistiques descriptives du jeu de données	14
Table 3.2. Aperçu d'un extrait des données générées	15

Liste des Figures

Figure 3-1. Approche CRISP-DM.....	11
Figure 3-2. Distribution de la qualité en graphique.....	15
Figure 4-1. Matrice de confusion du Régression Logistique.....	21
Figure 4-2 Matrice de confusion du KNN.....	22
Figure 4-3 Matrice de confusion de l'Arbre de Décision.....	23
Figure 4-4 Matrice de confusion du MLP	24
Figure 4-5 Matrice de confusion du Naïve Bayes Gaussien.....	25
Figure 4-6 Comparaison des performances des modèles	26
Figure 4-7 Courbe ROC des modèles utilisés	27

Introduction Générale

En raison de l'importance de la qualité pour garantir des structures fiables, le soudage est une technique d'assemblage cruciale dans de nombreux secteurs comme l'automobile ou l'aéronautique. Les méthodes de contrôle traditionnelles ont longtemps prévalu mais génèrent des coûts, reposent sur un long temps d'isolement, etc. et pour n'être en tout cas essentiellement qu'en post-production. La montée de l'Industrie 4.0 et l'émergence de méthodes en apprentissage automatique (Machine Learning) ouvrent désormais des perspectives pour la gestion prédictive de la qualité.

Ce mémoire fait écho à ces enjeux, et propose d'exploiter les algorithmes d'apprentissage automatique pour prédire la qualité des liaisons soudées. Grâce aux données possibles sur les paramètres du procédé de soudage, il s'agit de développer des modèles permettant d'identifier, avant même la production, les configurations pouvant aboutir à des défauts ou bien garantir la conformité. L'objectif est de fournir un outil d'aide à la décision, pour optimiser les procédés, réduire les rebuts et améliorer la productivité.

Le développement de notre solution repose dans un premier temps sur une revue approfondie de la littérature ; dans un second, sur la description de notre méthodologie pour la collecte et le traitement de nos données et pour le développement de nos modèles prédictifs ; puis, dans un troisième temps, sur l'analyse et la discussion des résultats obtenus; *enfin, sur l'analyse et la discussion des résultats obtenus, concluant sur les implications pratiques et les limites de cette recherche pour un contrôle qualité intelligent dans le domaine du soudage..*

Chapitre 1 : Présentation d'entreprise et contexte industrielle

Chapitre1 : Présentation d'entreprise et contexte industrielle

Ce chapitre présente brièvement l'entreprise EPE SPA EMB AZZABA , justifie les raisons de choix de ce projet et détermine la problématique présente dans cette entreprise.

1.1 Présentation d'entreprise

L'entreprise EPE SPA EMB AZZABA est un société par action qui fabrique les emballage métallique de conserve pour l'industrie agroalimentaire créé le 28 /11/19 située dans la région d' Azzaba dans la wilaya de Skikda pour satisfaire le besoin national , elle compte environ 132 salariés entre l'administration et la production ,Le produit fini de l'entreprise sont les boite métallique pour les conserves de différent type (boites de 1Kg ,boites de 500 g et boites de 250 g) ,l'usine contient 4 ligne de production de différent atelier comme atelier presse ,imprimerie ,train ,et bouchons couronnes . Ces lignes intègrent des processus tels que la découpe, l'emboutissage, l'assemblage, et le soudage par molette, qui est une étape critique dans l'obtention d'un emballage étanche et fiable. EMB Azzaba joue un rôle clé dans la chaîne d'approvisionnement en emballages métalliques à l'échelle nationale. Grâce à son expérience et à son savoir-faire, EPE SPA EMB AZZABA s'est positionnée comme un acteur crucial dans le secteur de l'emballage métallique en Algérie.

1.2 Motivation

Pour l'entreprise EPE SPA EMB Azzaba la qualité de produit fini est très important spécialement pour le soudage des boites car une soudure de mal qualité entraîner des pertes pour l'entreprise au niveau de matière première, le temps et la satisfaction des clients, actuellement la qualité de soudage est contrôlée visuellement ou avec des tests manuels mais cette méthode ne donne pas des résultats en temp réel ce qui expose le risque de non-conformité et de gaspillage de matières premières. La motivation de ce projet est de proposer à l'entreprise une solution moderne qui marche avec le développement de le monde et l'utilisation de l'intelligence artificiel. L'objectif est de renforcer la fiabilité du processus de soudage, de réduire les rebuts, et de garantir une qualité constante des produits finis.

Chapitre1 : Présentation d'entreprise et contexte industrielle

1.3 Problématique

Le soudage par résistance à la molette est le procédé utilisé dans le soudage des boîtes métallique à l'entreprise EMB Azzaba, ce procédé est très sensible à des paramètres spécifiques qui sont la température de soudage, le courant de soudage, et la force appliqué par la molette sur les flancs. Tous ces paramètres influencent sur la qualité du soudage, ou il faut les contrôler et les suivis en temp réel, cependant le suivi et le contrôle de ces paramètres seulement n'est pas efficace sans l'analyse et l'interprétation de ces résultat, pour cela la problématique qui se pose est: comment développer le meilleur modèle de Machine Learning à utiliser pour anticiper la qualité de cordons de soudage selon la variabilité des paramètres prédéfinis.

Conclusion

Le présent chapitre a permis d'identifier la structure EPE SPA EMB AZZABA et de justifier le choix du projet en lien avec ses nécessités spécifiques. La problématique rencontrée au sein de la structure de l'entreprise est le point de départ pour les analyses et les réponses qui seront abordées dans les chapitres suivants.

Chapitre 2 : État de l'art et revue biographique

Chapitre 02 : Etat de l'art et revue bibliographique.

La qualité du soudage est un critère essentiel dans la fabrication des produits métalliques, garantissant leur fiabilité et leur durabilité. Ce chapitre a pour objectif de passer en revue les études récentes qui explorent l'utilisation de l'IA pour prévoir la qualité des soudures et pour la maintenance prédictive des installations de soudage. Nous allons concentrer plus particulièrement sur le soudage par résistance à la molette.

2.1 Le Soudage par Résistance à la Molette : Processus, Contrôle Qualité et Défauts

Le soudage à la molette se distingue du soudage par points en remplaçant des électrodes par des disques cuivre, nommées molettes. Grâce à leur rotation, les molettes permettent de réaliser des soudures étanches et continues par recouvrement. Sur l'intégralité du cordon à souder, la soudure s'effectue de manière progressive et continue combinant le serrage des tôles et l'application de courant. Les facteurs qui influencent la qualité de soudure sont plusieurs notamment le courant, la pression appliquée par les molettes, la température, la vitesse de déplacement des boîtes et l'état des pièces à souder. Des soudures de mauvaise qualité peuvent conduire à des défauts tels que des collages imparfaits, des fuites, des fissures et une résistance mécanique réduite. (Technique, n.d.).

2.2 Méthodes traditionnelles de contrôle qualité et leurs limites

Couramment, le contrôle de qualité des soudures des boîtes s'effectue par différentes méthodes (SOUDRONIC, Manuel SOUDRONIC – qualité):

- **Contrôle visuel, aspect extérieur** : Cette méthode permet de détecter des défauts tels que criques, manques, points de soudure trop faibles, points de soudure brisés, trous forts, points de soudure irréguliers et l'absence de points de soudure.
- **Contrôle mécanique, directement sur la machine** : Ceci permet une première appréciation du cordon de soudure et peut utiliser de grands moyens techniques auxiliaires comme : l'essai au déchirement, l'essai de pliage, l'essai par bille, et l'essai de torsion.
- **Contrôle de séries sur la ligne de production** : Ces essais de séries sont nécessaires en vue d'assurer la fabrication et incluent le moniteur, le "can tester" sous vide ou sous pression, l'essai de pression dans bain d'eau (périodique) et l'essai sous vide (périodique).
- **Essais de laboratoire** : Ces essais permettent d'obtenir un rapport exact sur la qualité de soudure par l'appréciation de la structure resolidifiée à l'aide de micrographies

Chapitre 02 : Etat de l'art et revue bibliographique.

réalisées perpendiculairement, longitudinalement et obliquement au cordon, et l'essai d'arrachement.

Bien que ces méthodes soient utilisées depuis longtemps, elles contiennent des limitations. La plupart de ces contrôles sont effectués après la soudure, ce qui complique la correction des anomalies en temps réel. Il est donc essentiel de mettre en place des techniques de contrôle et de prévision constante de la qualité pour améliorer le processus du soudage à la molette et minimiser les défauts.

2.3 L'Intelligence Artificielle et son Impact sur l'Amélioration de la Qualité du Soudage par Résistance à la Molette

2.3.1. Surveillance du processus de soudage à l'aide de capteurs

Pour assurer un bon fonctionnement de l'IA, il est nécessaire de fournir des données. En ce qui concerne le soudage par résistance à la molette, cela implique la mise en place de capteurs qui vont contrôler en direct les opérations. Les capteurs qui on peut installer sont : de courant, de tension, de pression exercée par les molettes, de déplacement et bien aussi de température. L'idée, c'est de capturer le plus d'informations possible sur le processus. Le défi, c'est ensuite de synchroniser toutes ces données et de les organiser de manière à ce qu'elles soient exploitables.

2.3.2. Modèles de Machine Learning pour la prédiction de la qualité de soudure

Il existe plusieurs algorithmes de machine Learning (ML) qui peuvent être appliqués à la prédiction de la qualité du soudage par résistance à la molette. Ces derniers peuvent être utilisés soit pour la classification (prédire si une soudure est bonne ou mauvaise) ou la régression (prédire la résistance mécanique de la soudure). Des algorithmes comme la régression linéaire, les machines à vecteurs de support (SVM), les forêts aléatoires, le Gradient Boosting, le CatBoost, les K-plus proches voisins (KNN), ou le Deep Learning y sont employés. Cependant, l'efficacité de ces modèles dépend largement sur la pertinence des paramètres d'entrée (paramètres de soudage et données capteurs), soulignant l'importance d'une sélection méthodique pour garantir des prédictions fiables en milieu industriel. (MEZHER, 2024).

Chapitre 02 : Etat de l’art et revue bibliographique.

2.3.3. Détection d'anomalies et de défauts en temps réel

L’intégration des capteurs IoT dans les systèmes SRM¹ permet de recueillir en temps réel des données essentielles tels que : courant, tension, force, vitesse, température et même aussi les émissions acoustiques. Ces données, recueillies avec une fréquence élevée, sont essentielles pour une analyse approfondie du procédé, la détection des anomalies et la facilitation d'interventions rapides, ce qui aide à réduire les déchets et les interruptions. (CUI, 2025).

2.4 Industrie 4.0 dans les Systèmes de Soudage par Résistance à la Molette

La maintenance prédictive est une approche de maintenance qui implique d’évaluer constamment l’état de la machine et d’analyser les informations en temps réel pour anticiper les pannes éventuelles avant qu’elles n’interviennent. À l’opposé de la maintenance corrective, qui intervient en cas de défaillance, ou de la maintenance préventive, qui s’effectue selon un planning établi, la maintenance prédictive autorise uniquement l’action si elle est indispensable, ce qui optimise l’exploitation des ressources et réduit au minimum les temps d’arrêt. (MARCOPACK, 2023).

2.4.1. Applications de l'IA pour la maintenance des équipements de soudage par résistance à la molette

L’intelligence artificielle (IA) révolutionne aujourd’hui les stratégies de maintenance industrielle, notamment dans le domaine du soudage par résistance à la molette. Des diverses applications sont en forte et indissociable corrélation avec de réels enjeux qualité et productivité. En effet, les producteurs exploitant les divers outils d’intelligence artificielle qu’ils sont en mesure de mettre en œuvre une surveillance renforcée de leurs processus de soudage, projettent à toute fin utile leurs besoins de maintenance ultérieure et participent dès lors à l’optimisation de la qualité de leur produit garantissant la meilleure qualité de soudure. [(BRUNO, 2024)]

Conclusion

Pour conclure ce chapitre, notre travail a cherché à démontrer que la qualité est primordiale dans le domaine du soudage par résistance à la molette tout en pointant les limites des méthodes de contrôle classiques. Puis, nous avons vu que l’IA offre un potentiel disruptif, tant dans

¹ Soudage par résistance à la molette.



Chapitre 02 : Etat de l'art et revue bibliographique.

l'utilisation de capteurs et de modèles de Machine Learning pour prédire la qualité des soudures et détecter les anomalies en temps réel, que dans la perspective d'une maintenance prédictive sur les équipements de soudage. Ces jalons théoriques sont à la source de l'application pratique que nous allons parcourir dans les chapitres à venir.

Chapitre 3 : Conception d'un modèle prédictif de qualité

Chapitre 3 : Conception d'un modèle prédictif de qualité

La qualité est le facteur principal qui influence la réputation de l'entreprise, et tandis que le monde est en évolution constante, la détection de la qualité des produits connaît elle aussi un grand développement. Parmi les dernières avancées, on trouve l'utilisation de programmes d'apprentissage automatique (ML) pour la détection, ce dont nous allons parler dans ce chapitre.

3.1 Méthode CRISP-DM pour une Approche Structurée

3.1.1. Adoption de la méthodologie CRISP-DM

Depuis plusieurs années, et dans la conduite des projets de données, la méthode CRISP-DM est la plus utilisée et connue dans ce domaine. CRISP-DM définit un cycle de vie prévisible des projets de la compréhension du métier (Business Understanding), Compréhension des données (Data Understanding), Préparation des données (Data Preparation), Modélisation (Modeling), Évaluation (Evaluation), et le Déploiement (Deployment) d'une première version d'un modèle. (SHEARER, 2000).

Le schéma présenté ci-après (Figure 1) détaille la reconfiguration de la démarche CRISP-DM à notre problématique singulière. Il souligne les jalons essentiels, depuis la formulation des objectifs de simulation jusqu'à la mise en place des protocoles de modélisation.

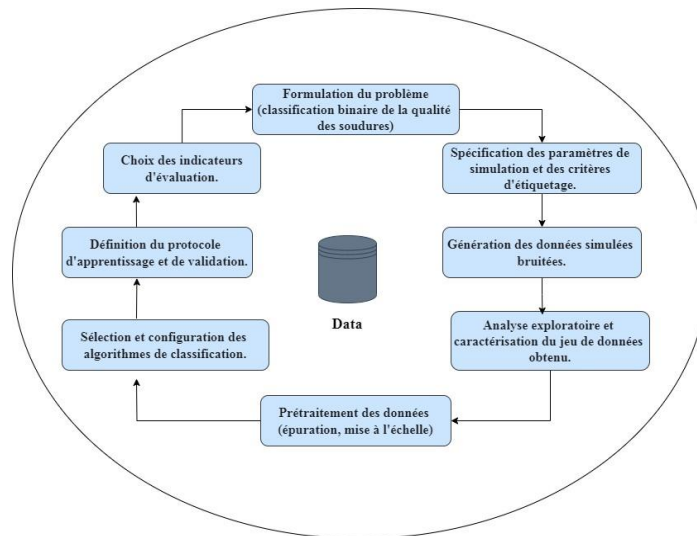


Figure 3-1. Approche CRISP-DM

3.2 Conception et élaboration d'un Corpus de données simulé pour l'analyse multi-classe de qualité

L'objectif pour créer un corpus de données synthétiques s'est imposée en raison de le manque de données industrielles. Ainsi, cette démarche permet d'avoir la maîtrise des attributs des données et d'évaluer tout aussi bien la pertinence de la modélisation prédictive construite dans un environnement sous contrôle.

Chapitre 3 : Conception d'un modèle prédictif de qualité

3.2.1. Système de soudage de référence et grandeurs simulées

Ces données, bien que synthétiques, s'inspirent de la technique de soudage industriel qu'est le soudage par résistance à la molette, pour lequel la qualité finale est étroitement tributaire de la gestion précise de ses paramètres opératoires. Dans cette étude, nous avons choisi quatre variables instrumentales dont nous savons qu'elles conditionnent fortement la qualité du joint (SOUDRONIC, TECHNICAL REPORT) :

- **Courant de soudage (Ampères)** : il agit directement sur l'apport thermique ;
- **Température (°C)** : elle constitue un indicateur important de l'énergie apportée et des transformations métallurgiques ;
- **Force appliquée (DaN)** : elle est indispensable à la garantie de l'homogénéité du contact et la cohésion du matériau ;
- **Vitesse d'avance (m/min)** : elle module le temps de contact et l'apport d'énergie linéaire.

Ces grandeurs sont le plus souvent disponibles en temps réel au sein des systèmes de soudage automatisés, ce qui en fait des candidats privilégiés pour une démarche de prédiction.

3.2.2. Protocole de création des signaux capteurs bruités

Afin de simuler des données de capteurs en rapport avec le réel, il a été décidé de procéder de la manière suivante :

1. **Valeurs de Consigne Optimales** : Plages de valeurs "optimales" pour chaque capteur : température, courant, pression, vitesse, correspondant à une "bonne" soudure.
2. **Ajout de Bruit Réaliste** : Bruit gaussien ajouté aux données de consigne, écart-type proportionnel à la valeur nominale du signal (5-10 % par exemple), cela pour simuler les variations de mesure, à savoir : $Valeur_perturbée = Valeur_optimale + Bruit_Gaussien$.
3. **Génération de Non-Conformités** : Pour avoir des instances "non conformes", certaines valeurs de consigne ont dû être décalées en dehors de leur zone de l'idéal avant d'être bruitées ou le bruit a pu avoir une amplitude augmentée.
4. **Écrêtage (Clipping)** : Les valeurs simulées ont dû être écrêtées pour être réalistes vis-à-vis de la physique et les plages de fonctionnement des capteurs.
5. **Échantillonnage** : Les données ont dû être produites avec un temps fixe entre chaque échantillon, chaque "soudure" étant représentée par un jeu de capteurs uniques (moyennés ou stabilisés).

Ce protocole a comme but de proposer un jeu de données certes synthétique mais représentatif des conditions réelles pour évaluer la robustesse des analyses qui suivront.

Chapitre 3 : Conception d'un modèle prédictif de qualité

3.2.3. Etablissement des critères d'étiquetage pour la qualité multi-classe

La procédure consistant à générer les valeurs remaniées du capteur est suivie de la catégorisation de chaque instance de soudure simulée sur quatre niveaux de qualité : «Excellente», «Bonne», «Médiocre», et «Défectueuse».

Cette classification est construite sur la conformité finale aux plages de valeurs des quatre paramètres (température, courant, pression, vitesse) indiquées pour chaque niveau :

- « **Excellente** » et « **Bonne** » : Les paramètres sont inclus dans des fenêtres de procédé optimales, adaptées avec des rigidités plus fortes pour « Excellente ».
- « **Médiocre** » : Les paramètres s'éloignent significativement des plages optimales, et se signalent comme dégradés sans être cependant autant que fautifs.
- « **Défectueuse** » : Au moins un des paramètres dépasse les seuils acceptables des classes supérieures, ou de manière biaisée est contraint à déviation extrême inductivement conçue précédemment.

Ce choix permet d'évaluer la qualité, d'une façon nuancée et tend à mettre à l'échelle du jeu de données final une distribution des classes non plus uniformément réparties.

3.2.4. Description du jeu de données final

La démarche de simulation a abouti à la constitution d'un corpus de données final se caractérisant par :

- **Volume total d'instances** : 2400 opérations de soudage simulées.
- **Nombre de variables explicatives (features)** : 4 variables numériques (Courant, Température, Pression, Vitesse).
- **Variable cible** : Qualité ("Bonne" :0, "Défectueuse" : 1, "Excellente" : 2, "Médiocre" : 3)

Chapitre 3 : Conception d'un modèle prédictif de qualité

Aperçu des statistiques descriptives des variables (sur les valeurs perturbées) :

Statistique	Température (°C)	Courant (A)	Pression (DaN)	Vitesse(Boites/min)
count	2400	2400	2400	2400
mean	314.96	3552.33	47.61	259.4
std	26.63	741.61	2.26	14.26
min	250.71	1042	40	200
25%	308.47	3349.75	46.76	254
50%	312.43	3552	47.53	260
75%	316.2	3693	48.38	265
max	498.62	7553	55	300

Table 3.1. Aperçu des statistiques descriptives du jeu de données

Discussion

Le tableau présente des différents statistiques pour chaque paramètre de soudage, on peut conclure des choses :

- **Température** : La moyenne est de **314.96°C**, avec un écart-type de **26.63**, il y a donc une variation modérée. Les valeurs extrêmes vont de **250.71** à **498.62**, mais la médiane (**312.43**) et les quartiles sont proches de la moyenne, ce qui montre que la distribution est assez stable.
- **Courant** : la moyenne est de 3552.33 A avec un écart-type de 741.61 A, on peut suggérer qu'il existe une dispersion notable. En effet, l'écart entre le min et le max, ce qui pourrait indiquer des valeurs extrêmes ou des sous-groupes dans les données.
- **Pression** : La moyenne observée est de 47.61 DaN, avec un écart-type de 2.26 DaN, ce qui indique une faible dispersion. Les valeurs sont bien réparties autour de la moyenne avec peu de variation.
- **Vitesse** : La moyenne est de 259.4 boites/min, avec un écart-type de 14.26 boites/min, il y a donc une certaine variabilité. Les quartiles (25%, 75%) sont centres autour la médiane (50%) ce qui renforce que la distribution est normale.

Chapitre 3 : Conception d'un modèle prédictif de qualité

Horodatage	Température (°C)	Courant (A)	Pression (DaN)	Vitesse(Boites/min)	Qualite_soudure
4/26/2025 8:15	313.54	3657	47.72	264	Excellente
4/26/2025 15:26	308.35	3610	48.45	256	Bonne
4/29/2025 8:20	330.05	2908	43.98	247	Médiocre
4/29/2025 14:17	305.91	6637	51.21	218	Défectueuse

Illustration d'un extrait du jeu de données (5 premières observations) :

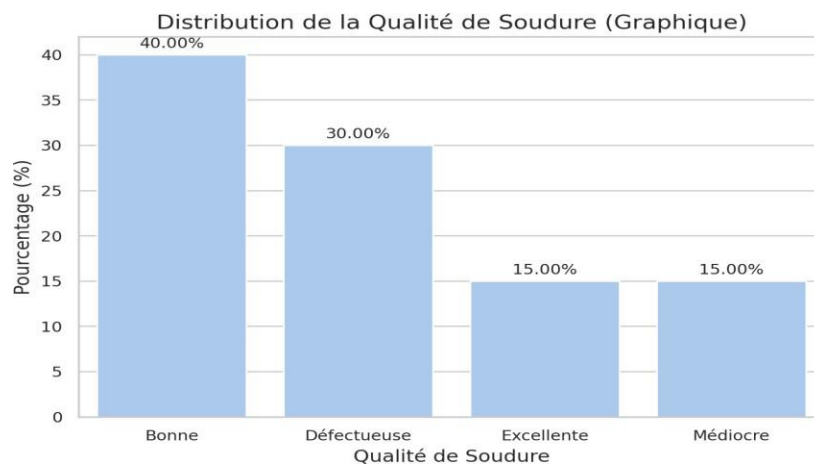


Figure 3-2. Distribution de la qualité en graphique

Discussion :

D'après la figure 2, nous voyons que la classe 'Bonne' a le pourcentage supérieur (40%) avec un effectif de 960, suivie par la classe 'Défectueuse' par 30% avec un effectif de 720, et à la fin on trouve l'égalité entre les 2 classe 'Excellente' et 'Médiocre' par 15% avec un effectif de 360 pour chacune des classes.

3.3 Conditionnement des données pour la modélisation prédictive

3.3.1. Sélection et justification des variables explicatives

Dans cette phase initiale de développement des modèles, les variables explicatives retenues correspondent directement aux quatre grandeurs synthétisées et bruitées : Température, Courant, Pression et Vitesse. Leur sélection repose sur leur influence avérée sur la qualité du

Chapitre 3 : Conception d'un modèle prédictif de qualité

soudage, ainsi que sur leur fiabilité en tant que mesures fournies et validées par l'entreprise. Aucun travail avancé d'ingénierie des caractéristiques (feature engineering) n'a été réalisé à ce stade, l'objectif étant d'évaluer la performance des modèles sur la base des données les plus directes et représentatives du procédé industriel.

3.3.2. Transformation des données : Mise à l'échelle par standardisation

Certains algorithmes d'apprentissage automatique, tels que ceux basés sur les mesures de distance (ex : SVM) et ceux reposant sur la descente de gradient (ex : Régression Logistique), semblent être plus sensibles que d'autres à l'échelle des variables d'entrée (Shaibu, 2024). En effet, les variables dont la plage de valeurs est étendue pourraient dominer les variables dont les valeurs sont mieux contrôlées sur des plages plus étroites et vont ainsi fausser les performances du modèle et la vitesse de convergence.

Pour atténuer ce problème, une standardisation (normalisation Z-score) des variables explicatives (données capteurs) est appliquée. Chaque valeur x est associée x' selon la formule (Shaibu, 2024):

$$x' = \frac{(x - \mu)}{\sigma}$$

où : μ et σ sont respectivement la moyenne et l'écart-type de la variable (calculés uniquement sur l'ensemble d'apprentissage). Cette opération permet de centrer les données autour d'une moyenne de zéro avec un écart-type unitaire.

Cette transformation est cruciale et sera effectuée après la séparation des données en ensembles d'apprentissage et de test, afin d'éviter toute fuite d'information de l'ensemble de test vers celui d'apprentissage, garantissant ainsi une évaluation impartiale du modèle.

3.4 Architecture des modèles de classification multiple et paramétrage des expérimentations

3.4.1. Algorithmes de classification

Dans le cadre de prédiction de la qualité de soudure, nous avons choisis des modèles que nous avons déjà étudié, ainsi qu'à d'autres modèles fréquemment utilisés dans la littérature pour des problèmes similaires de prédiction de qualité :

- **Régression Logistique** : Cet algorithme sert principalement à évaluer la probabilité qu'un élément fasse partie d'une catégorie spécifique. Lorsque cette probabilité dépasse 50 %, le modèle considère que l'élément appartient à la classe positive (marquée "1") ; dans le cas contraire, il est classé comme négatif (étiqueté "0"). Il s'agit donc d'un outil de classification binaire. (GÉRON, 2019).

Chapitre 3 : Conception d'un modèle prédictif de qualité

- **K-Plus Proches Voisins (KNN)** : est un type d'apprentissage par instances (ou non généralisant) : elle ne tente pas de construire un modèle interne général, mais se contente de stocker les instances des données d'entraînement. La classification est calculée par un vote majoritaire simple des plus proches voisins de chaque point : un point de requête est assigné à la classe qui a le plus de représentants parmi ses voisins les plus proches. (SCIKIT-LEARN, n.d.).
- **Arbre de Décision** : C'est un algorithme polyvalent d'apprentissage automatique capable d'effectuer des tâches de classification et de régression, voire des tâches à sorties multiples. C'est un algorithme puissant, apte à s'adapter à des jeux de données complexes. (GÉRON, 2019).
- **Bayes Naïf Gaussien** : Cet algorithme supervisé repose sur le théorème de Bayes, en supposant (de façon simplificatrice) que les caractéristiques sont indépendantes entre elles conditionnellement à la classe. La version Gaussienne (GaussianNB) modélise la distribution des données par une loi normale. Les paramètres sont calculés en maximisant la vraisemblance. (SCIKIT-LEARN., n.d.).
- **Perceptron Multicouche (Multi-Layer Perceptron - MLP) / Réseau de Neurones Artificiels (ANN)** : Il est composé d'une couche d'entrée directe, d'une ou plusieurs couches de TLUs (appelées couches cachées), et d'une dernière couche de TLUs nommée couche de sortie. Les couches proches de l'entrée sont généralement appelées couches inférieures, tandis que celles proches des sorties sont appelées couches supérieures. Toutes les couches, sauf celle de sortie, incluent un neurone de biais et sont entièrement connectées à la couche suivante. (GÉRON, 2019)

Ces choix des modèles permettent de faire dialoguer des philosophies très différentes et éventuellement de pointer celles qui sont les plus à même de rendre compte de la nature de nos données simulées.

3.4.2. Schéma d'apprentissage et de validation

Pour une vision d'estimation robuste des performances des modèles mais également d'évitement du surapprentissage, l'ensemble de données sera divisé en ensembles d'apprentissage (80 %) et de test (20 %). Cette séparation sera réalisée de manière stratifiée, dans le but de garantir la proportion des classes dans les deux sous-ensembles, tout en prévenant du surapprentissage.

Chapitre 3 : Conception d'un modèle prédictif de qualité

3.4.3. Choix des indicateurs d'évaluation pour la classification multiple

Pour juger du bon fonctionnement de nos modèles de classification multiple, les métriques suivantes seront calculées sur l'ensemble de test :

- **Matrice de Confusion** : pour visualiser les Vrais Positifs (VP), Faux Positifs (FP), Vrais Négatifs (VN) et Faux Négatifs (FN), en considérant la classe "Défectueuse" comme la classe positive.

- **Métriques par Classe** :

Pour évaluer la performance sur chacune des quatre classes de qualité ("Excellente", "Bonne", "Médiocre", "Défectueuse"), les indicateurs suivants seront calculés pour chaque classe :

- **Précision** : Proportion d'instances correctement classées pour une classe donnée parmi toutes celles prédites comme appartenant à cette classe.
- **Rappel** : Proportion d'instances d'une classe donnée que le modèle a correctement identifiées.
- **Score F1** : Moyenne harmonique de la Précision et du Rappel pour chaque classe.

- **Métriques Moyennées (Agrégées)**

Afin d'obtenir une mesure synthétique de la performance globale tout en tenant compte de la nature multi-classe et d'un éventuel déséquilibre, les moyennes suivantes seront rapportées :

- **Exactitude(précision global) (Accuracy)** : mesure globale du pourcentage de prédictions correctes.
 - **Macro-Average F1-score** : Calcule la métrique pour chaque classe et trouve leur moyenne non pondérée. Traite toutes les classes de manière égale.
 - **Weighted-Average F1-score (ou Précision/Rappel)** : Calcule la métrique pour chaque classe et trouve leur moyenne, pondérée par le nombre d'instances réelles dans chaque classe. Tient compte du déséquilibre.
- **L'Aire sous la Courbe ROC (AUC-ROC)** : qualité de discrimination du modèle entre les classes « Bonne » et « Défectueuse » sur l'ensemble des seuils de classification. (FAWCETT, 2006).

Chapitre 3 : Conception d'un modèle prédictif de qualité

Conclusion

Ce chapitre a détaillé l'architecture méthodologique pour la prédiction multi-classe de la qualité de soudure selon la démarche CRISP-DM, depuis la constitution d'un jeu de données synthétiques bruitées jusqu'au conditionnement des données par standardisation des variables. Notre dispositif expérimental rigoureux reposant sur la sélection d'algorithmes de classification variés, d'un protocole d'apprentissage, ainsi le choix d'indicateurs d'évaluation pertinents. Ces bases expérimentales solides, permettent une mise en œuvre fiable des modèles et préparent le terrain pour le Chapitre 4 qui présentera une analyse approfondie des résultats, discutera leurs implications pratiques, complétant ainsi l'approche systématique initiée dans ce chapitre méthodologique.

Chapitre 4 : Résultats et Discussion

Chapitre 4 : Résultats et Discussion

Dans ce chapitre tous les modèles de classification de la qualité de soudage seront évalués et le meilleur modèle sera sélectionné.

4.1 Résultats par modèle

4.1.1. Modèle de Régression Logistique

- Présentation des performances globale

Modèle	Classe	Précision	Rappel	F1-score	Support
Régression logistique	Excellente	0.37	0.92	0.52	108
Régression logistique	Bonne	0.62	0.26	0.37	288
Régression logistique	Médiocre	0.17	0.28	0.21	108
Régression logistique	Défectueuse	0.45	0.32	0.38	216

Tableau 4-1. Présentation des performances du modèle Régression Logistique

Description

Ce tableau détaille les performances du modèle de régression logistique pour la classification de la qualité de soudures. Il présente les résultats pour chacune des quatre classes prédites : « Excellente », « Bonne », « Médiocre » et « Défectueuse ». Pour chaque classe, les métriques de précision, de rappel et le F1-score sont spécifiés. La colonne "Support" indique le nombre d'échantillons réels appartenant à chaque catégorie dans les données.

- Matrice de confusion

Cette matrice de confusion illustre les résultats d'un modèle de régression logistique. Le modèle identifie avec une grande précision la classe « Excellente » (99 prédictions justes). Cependant, il commet une erreur majeure en classant 90 soudures de la classe « Bonne » comme « Médiocres ». On observe aussi une confusion notable entre les soudures de la classe « Défectueuse » et « Excellents » (64 cas).

- Analyse des résultats

L'analyse du modèle de régression logistique révèle une performance globale faible et déséquilibrée. Le modèle se distingue uniquement par sa capacité à identifier la classe "Excellente" (Rappel de 0.92), mais échoue sur toutes les autres catégories, comme en témoignent leurs très faibles scores F1. Il confond massivement les classes, le rendant ainsi inutilisable en pratique pour cette tâche de classification.

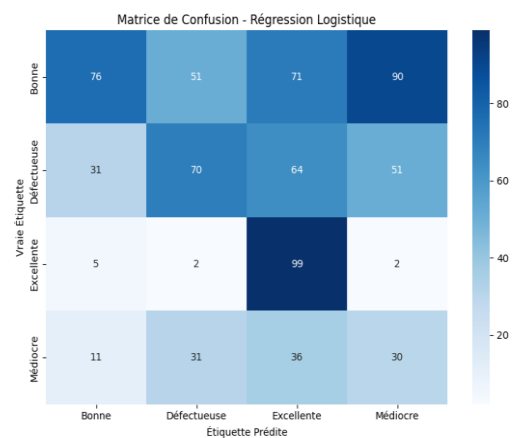


Figure 4-1. Matrice de confusion du Régression Logistique

Chapitre 4 : Résultats et Discussion

4.1.2. Modèle de Les K plus proches Voisins

- **Présentation des performances globale**

Modèle	Classe	Précision	Rappel	F1-score	Support
KNN	Excellente	0.83	0.99	0.90	108
KNN	Bonne	0.87	0.92	0.89	288
KNN	Médiocre	0.55	0.44	0.49	108
KNN	Défectueuse	0.82	0.75	0.79	216

Tableau 4-2. Performances globales du KNN

Description

Ce tableau présente les résultats de performance du modèle k-plus proches voisins (KNN) sur une tâche de classification. Il fournit une évaluation détaillée pour chacune des quatre classes : « Excellente », « Bonne », « Médiocre » et « Défectueuse ». Les métriques de précision, de rappel et le F1-score sont affichées pour chaque catégorie. La colonne "Support" indique le nombre d'instances réelles par classe dans le jeu de données.

- **Matrice de confusion**

Cette matrice de confusion présente les résultats de modèle K-NN (k=5). Le modèle identifie avec une grande efficacité les classes « Bonne » (265 cas) et « Excellente » (107 cas). Sa principale faiblesse réside dans la prédiction de la classe « Médiocre », la moins bien reconnue. L'erreur la plus significative est la confusion entre les soudures « Défectueuse » et « Médiocre ».

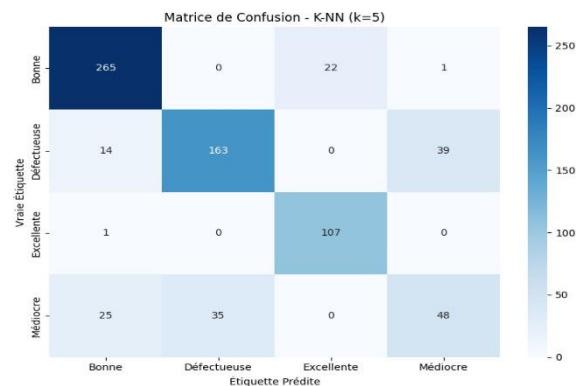


Figure 4-2 Matrice de confusion du KNN

- **Analyse des résultats**

L'évaluation du modèle K-NN (k=5) démontre une performance globale robuste et bien équilibrée, surpassant de manière significative le modèle de régression logistique.

Le modèle atteint des scores F1 élevés pour les classes "Excellente" (0.90), "Bonne" (0.89) et "Défectueuse" (0.79), indiquant une forte capacité à les identifier avec précision et sensibilité. La matrice de confusion confirme une nette distinction entre ces catégories, avec un nombre d'erreurs de classification très limité.

La seule faiblesse relative concerne la classe "Médiocre" (F1-score de 0.49), pour laquelle le modèle montre encore une certaine confusion avec les classes adjacentes.

Chapitre 4 : Résultats et Discussion

4.1.3. Modèle Arbre de Décision

- Présentation des performances globale

Modèle	Classe	Précision	Rappel	F1-score	Support
Arbre de décision	Excellente	0.89	0.86	0.87	108
Arbre de décision	Bonne	0.87	0.87	0.87	288
Arbre de décision	Médiocre	0.44	0.41	0.42	108
Arbre de décision	Défectueuse	0.74	0.77	0.75	216

Tableau 4-3 Performances globales de l'Arbre de Décision

Description

Ce tableau résume les performances d'un modèle d'arbre de décision pour une tâche de classification. Il détaille les résultats obtenus pour chacune des quatre classes : "Excellente", "Bonne", "Médiocre" et "Défectueuse". Pour chaque classe, les indicateurs de précision, de rappel et le F1-score sont présentés. La colonne "Support" spécifie le nombre d'instances réelles par catégorie dans le jeu de données.

- Matrice de confusion

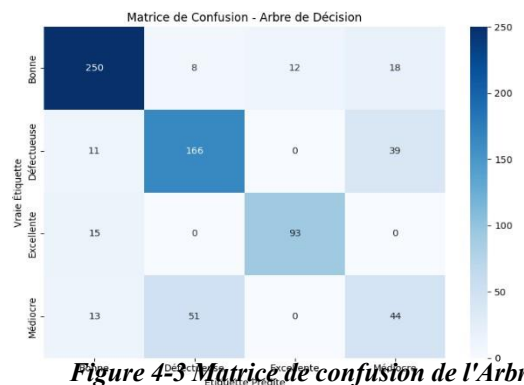
Cette matrice de confusion détaille les prédictions de modèle d'Arbre de Décision. Le modèle est très performant pour identifier les classes « Bonne » (250) et « Défectueuse » (166). L'erreur la plus notable est la classification de 51 soudures de la classe « Médiocre » dans la classe « Défectueuse ». La classe « Médiocre » est la moins bien distinguée, avec le plus faible nombre de prédictions correctes (44).

- Analyse des résultats

L'analyse de l'Arbre de Décision révèle des performances globales élevées, le positionnant comme un modèle très compétitif, avec des résultats comparables à ceux du modèle KNN.

Le modèle démontre une grande robustesse, attestée par des scores F1 identiques de 0.87 pour les classes "Bonne" et "Excellente", et un score solide de 0.75 pour la classe "Défectueuse". Ces métriques, confirmées par une diagonale fortement marquée dans la matrice de confusion, confirment sa capacité à distinguer efficacement ces catégories.

La principale limitation du modèle, à l'instar du KNN, réside dans la classification de la classe "Médiocre", qui n'atteint qu'un F1-score de 0.42. La confusion se produit principalement avec la classe "Défectueuse".



Chapitre 4 : Résultats et Discussion

4.1.4. Modèle Réseau de Neurones – MLP

- Présentation des performances globale

Modèle	Classe	Précision	Rappel	F1-score	Support
MLP	Bonne	0.94	0.93	0.94	288
MLP	Médiocre	0.51	0.33	0.40	108
MLP	Défectueuse	0.75	0.84	0.79	216
MLP	Excellente	0.88	0.99	0.93	108

Tableau 4-4 Performances globales du MLP

Description

Ce tableau présente les performances d'un modèle de réseau de neurones appliqué à une tâche de classification. Il fournit les résultats détaillés pour les quatre classes prédites : "Excellente ", "Bonne", "Médiocre", et "Défectueuse". Pour chaque classe, les métriques de précision, de rappel et le F1-score sont affichés. La colonne "Support" indique le nombre d'instances réelles par catégorie dans l'ensemble de données.

- Matrice de confusion

Cette matrice de confusion évalue le modèle de réseau de neurones (MLP) créé avec Keras. Le modèle identifie avec une grande précision les classes « Excellente », « Bonne », et « Défectueuse ». Son erreur la plus significative est de classer 59 soudures de la classe « Médiocre » comme étant des soudures de la classe

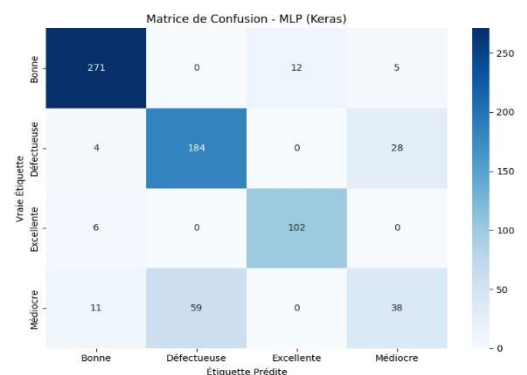


Figure 4-4 Matrice de confusion du MLP

« Défectueuse ». La classe « Médiocre » est ainsi la moins bien reconnue, avec seulement 38 prédictions correctes.

- Analyse des résultats

L'évaluation du Perceptron Multicouches (MLP) révèle les performances les plus élevées parmi tous les modèles testés, le positionnant comme la solution optimale pour ce problème de classification. Le modèle se distingue par des scores F1 exceptionnels pour les classes "Bonne" (0.94) et "Excellente" (0.93), et maintient une performance très solide sur la classe "Défectueuse" (0.79). La matrice de confusion valide cette supériorité, avec une diagonale très prononcée et un nombre d'erreurs de classification minimal entre les catégories. La faiblesse persistante, commune à tous les modèles, concerne la classe "Médiocre" (F1-score de 0.40), qui est encore largement confondue avec la classe "Défectueuse".

Chapitre 4 : Résultats et Discussion

4.1.5. Modèle Bayes Naïf Gaussien

- **Présentation des performances globale**

Modèle	Classe	Précision	Rappel	F1-score	Support
Gaussien bayes naïve	Excellente	0.91	0.92	0.93	108
Gaussien bayes naïve	Bonne	0.94	0.95	0.96	288
Gaussien bayes naïve	Médiocre	0.48	0.49	0.50	108
Gaussien bayes naïve	Défectueuse	0.81	0.82	0.83	216

Tableau 4-5 Performances globales du Bayes Naïf Gaussien

Description

Ce tableau résume les performances du modèle de classification Bayes Naïf Gaussien. Il présente les résultats détaillés pour chacune des quatre classes : "Excellente", "Bonne", "Médiocre" et "Défectueuse". Pour chaque classe, les métriques de précision, de rappel et le F1-score sont affichés. La colonne "Support" spécifie le nombre d'échantillons réels par catégorie dans les données.

- **Matrice de confusion**

Cette matrice de confusion illustre les résultats de modèle Gaussien Naïf Bayes. Le modèle identifie très bien les classes « Bonne » (274) et « Excellente » (104). Son erreur la plus importante est de classer 65 soudures « Défectueux » comme des soudures de la classe « Médiocre ». La classe « Médiocre » est également souvent confondue, notamment avec la classe « Défectueuse ».

- **Analyse des résultats**

L'évaluation du classifieur Bayes Naïf Gaussien révèle des performances globales exceptionnelles, le positionnant comme le modèle le plus efficace et le mieux équilibré parmi toutes les approches testées. Ce modèle se distingue par des

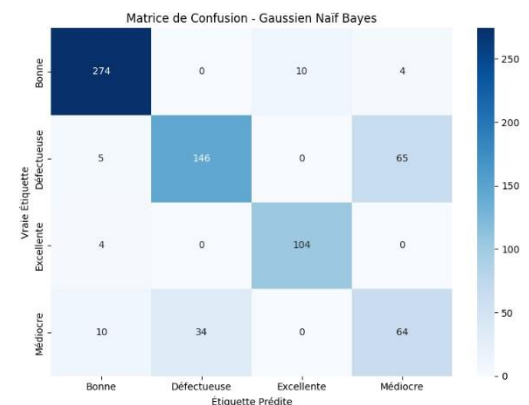


Figure 4-5 Matrice de confusion du Naïve Bayes Gaussien

scores F1 remarquables, atteignant des sommets pour chaque catégorie : 0.96 pour "Bonne", 0.93 pour "Excellente" et 0.83 pour "Défectueuse". De manière notable, il est également le plus performant sur la classe la plus ardue, "Médiocre", avec un F1-score de 0.50. La matrice de confusion confirme cette supériorité, avec une discrimination très nette entre les classes et une minimisation des erreurs critiques. La seule confusion significative persiste entre les catégories "Médiocre" et "Défectueuse", bien qu'elle soit mieux gérée que par les autres modèles.

Chapitre 4 : Résultats et Discussion

4.2 Comparaison des modèles

4.2.1. Résultats comparatifs

Modèle	Précision globale	F1-score macro	F1-score pondérée	AUC ROC	Temps d'entrainement	Temps de prédiction
Régression logistique	0.38	0.37	0.37	0.5496	0.0095	0.0011
Arbre de décision	0.77	0.73	0.77	0.8231	0.0117	0.0008
KNN	0.81	0.77	0.80	0.9322	0.0017	0.0431
MLP	0.83	0.77	0.81	0.9596	28.1701	0.4889
Bayes Naïf Gaussien	0.52	0.79	0.82	0.9534	0.0015	0.0011

Tableau 4-6 Résultats Comparatifs des modèles

Description

Ce tableau comparatif évalue les performances globales de cinq modèles d'apprentissage automatique. Il présente pour chaque modèle des métriques d'évaluation comme la précision, les F1-scores et l'AUC ROC. En complément, il mesure l'efficacité en indiquant les temps d'entraînement et de prédiction. L'ensemble de ces informations permet de confronter les modèles sur la base de leur performance et de leur rapidité.

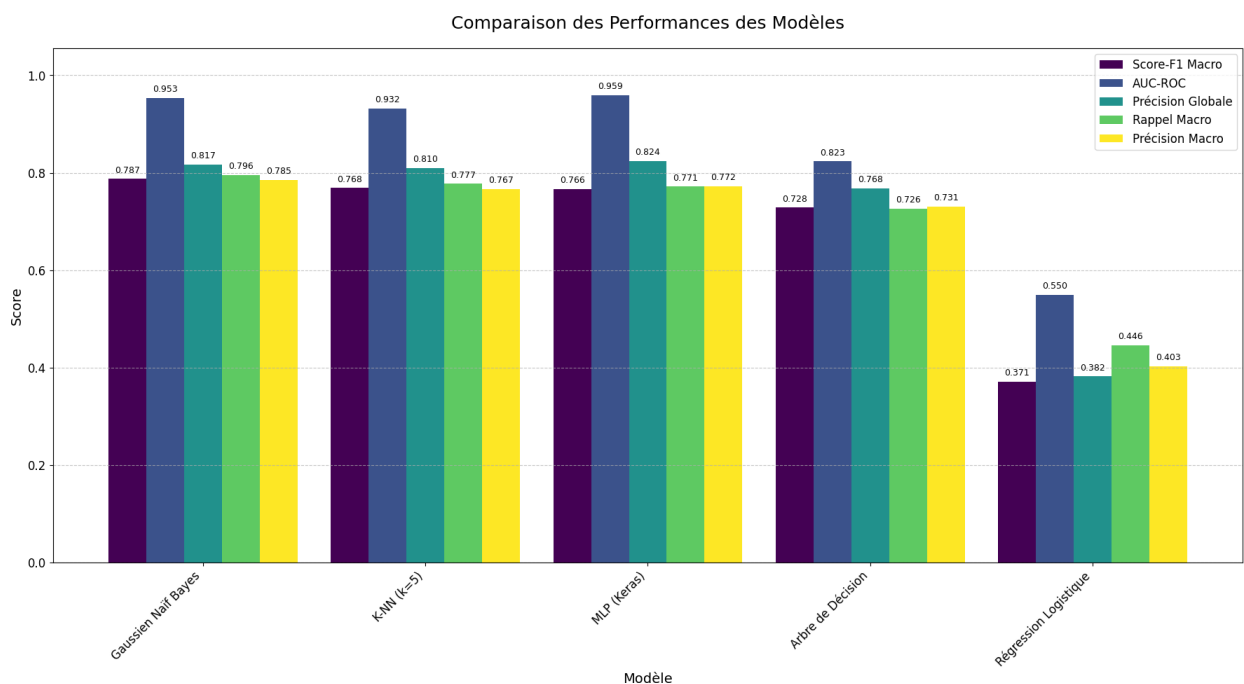


Figure 4-6 Comparaison des performances des modèles

Chapitre 4 : Résultats et Discussion

Description

Ce diagramme en barres présente une comparaison visuelle des performances de cinq modèles d'apprentissage automatique. Pour chaque modèle, le graphique affiche un groupe de barres représentant cinq métriques d'évaluation différentes. Ces métriques incluent le Score-F1 Macro, l'AUC-ROC, la Précision Globale, le Rappel Macro et la Précision Macro. La hauteur de chaque barre correspond au score obtenu, permettant d'évaluer facilement l'efficacité de chaque modèle.

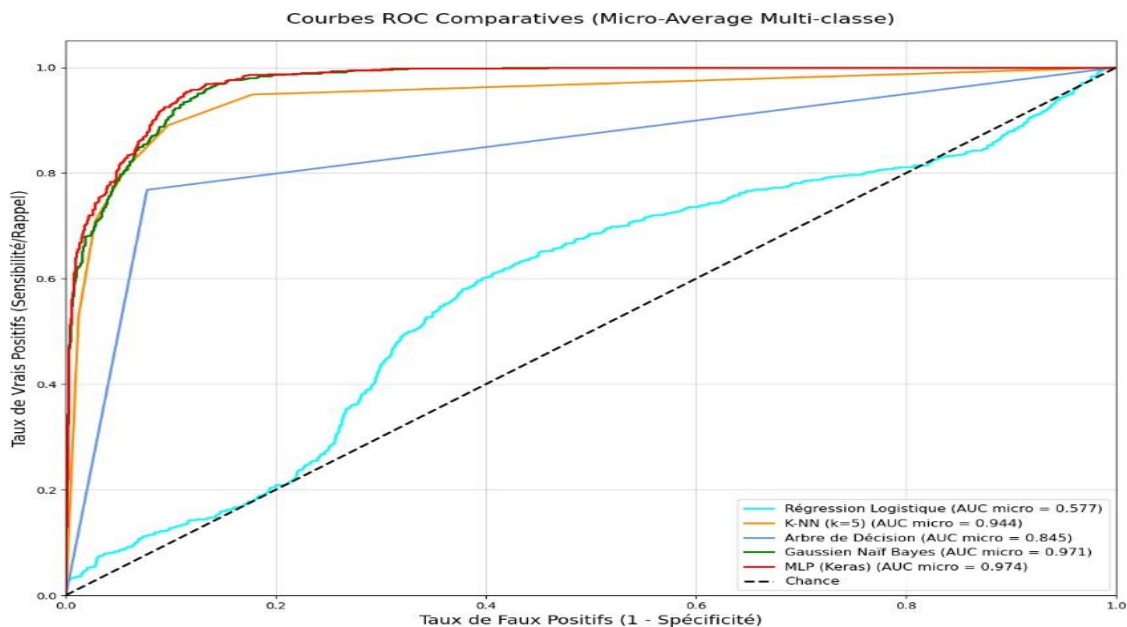


Figure 4-7 Courbe ROC des modèles utilisés

Description

Ce graphique présente les courbes ROC (Receiver Operating Characteristic) comparatives pour cinq modèles de classification. Il met en relation le taux de vrais positifs (sensibilité) avec le taux de faux positifs pour visualiser la performance de chaque modèle. Chaque courbe colorée représente un algorithme spécifique, dont la valeur de l'aire sous la courbe (AUC micro) est précisée dans la légende. La ligne diagonale en pointillés sert de référence, correspondant à la performance d'un classificateur aléatoire (chance).

4.2.2. Analyse comparative

L'analyse comparative des cinq modèles révèle une hiérarchie claire de performance et d'efficacité. Les modèles Bayes Naïf Gaussien, MLP et KNN se distinguent par leurs excellentes capacités prédictives, avec des scores AUC-ROC supérieurs à 0.93, surpassant nettement l'Arbre de Décision et la Régression Logistique. Cependant, l'introduction des coûts

Chapitre 4 : Résultats et Discussion

de calcul est déterminante. Le modèle MLP, bien que très performant (AUC de 0.96), présente un temps d'entraînement et de prédiction excessif, le rendant peu pratique. À l'inverse, le modèle Bayes Naïf Gaussien se démarque de manière exceptionnelle. Il combine des métriques de performance de premier plan (AUC de 0.95, F1-score pondéré de 0.82) avec une efficacité de calcul optimale (temps d'exécution quasi-instantané).

4.3 Choix final de modèle

Selon l'analyse des résultats de l'application des différents modèles de Machine Learning, le modèle de Bayes Naïf Gaussien atteint des bonnes performances avec temps de prédiction et d'entraînement rapide et précision globale excellente.

Ce modèle sera adopté lors des prochaines phases du projet, notamment lors de la phase de déploiement au sein du système de maintenance prédictive.

Conclusion

Ce chapitre compare plusieurs algorithmes de classification de la qualité des soudures à partir de données de capteurs. Les résultats montrent que le modèle Bayes Naïf Gaussien est le plus performant, avec une précision élevée (81.67%), ainsi qu'une bonne performance globale (AUC 95.34%) ,avec un temps d'entraînement et prédiction rapide.

En comparaison, des algorithmes tels que le modèle de MLP et KNN qui donne bons résultats, sachant que le modèle de régression logistique affichait de faibles performances.

Sur la base de ces résultats, le modèle Bayes Naïf Gaussien a été adopté comme modèle final pour l'étape suivante, notamment pour les applications de maintenance prédictive en environnement industriel.

Conclusion Générale

L'objectif de ce travail est d'améliorer la stratégie de contrôle qualité pour le soudage des boîtes métalliques, sur la base des données collectées par des capteurs industriels qui mesurent les paramètres influençant la qualité, à savoir le courant électrique, la force, la température et la vitesse d'avancement des boîtes.

Dans un premier temps, une base de données a été créée, comprenant la qualité des soudures divisée en quatre classes : Excellente, Bonne, Médiocre et Défectueuse. Après cela, différents modèles d'auto-apprentissage ont été appliqués et comparés les uns aux autres, tels que : le modèle de régression logistique, Modèle des K plus proche voisins, Arbre de décision et Réseau de Neurones.

Diverses métriques tels que la précision globale, le F1-score macro, AUC-ROC, le temps d'entraînement et le temps de prédiction, compte tenu de leur importance pour l'assurance qualité. Les résultats ont montré que le modèle Bayes Naïf Gaussien était le meilleur modèle en raison de ses performances globales élevées avec une précision de 81.67% de et un F1 score macro 78.72 % ; en revanche, nous considérons que le modèle de régression logistique est le plus faible en termes de performance car il obtient de mauvais résultats par rapport aux résultats des autres modèles.

Ces résultats confirment la puissance et l'efficacité du modèle Bayes Naïf Gaussien dans le traitement de données industrielles de nature variable et complexe, car le modèle a démontré une grande capacité efficace dans la prédiction de la qualité des soudures.

En conclusion, le modèle Bayes Naïf Gaussien est le meilleur modèle d'apprentissage automatique qui peut être appliqué pour prédire la qualité des soudures dans un environnement industriel complexe.

Références Bibliographie

Références

- [1] AG, S. (2023). *High-speed welding systems for metal packaging*.
- [2] AG, S. (s.d.). *SPEETEC 1D – Non-contact motion sensors*. Récupéré sur SICK: <https://www.sick.com/ch/en/catalog/products/motion-control-sensors/non-contact-motion-sensors/speetec-1d/c/g550876>
- [3] BRUNO, G. A. (2024). *Convolutional Neural Network for Quality Monitoring and Predictive Maintenance in Resistance Spot Welding*. Advances in Computational Intelligence and Robotics Book Series.
- [4] C9C - compression et miniature. (s.d.). Récupéré sur HBM Test and Measurement: https://www.hbm.com/fr/3928/c9c-capteur-de-force/?product_type_no=C9C
- [5] Corporation, K. (2025). *Model DSB™ Miniature Displacement Sensor*. Récupéré sur Keyence (France): https://www.keyence.fr/mykeyence/?ptn=001&dlSiteId=8&dlLanguageId=fr-FR&dlLangType=fr-FR&ad_local=modeldsbtlm
- [6] CUI, S. Z. (2025). Research on an Online Intelligent Monitoring System for Resistance Spot Welding Based on Wireless Communication. 25.
- [7] Drouart, C. &. (2016). *Le soudage par résistance*. Récupéré sur formation-soudure.com: <https://www.formation-soudure.com/Livret%20Soudage%20par%20Resistance%20-%20SDs%202016.pdf>
- [8] FAWCETT, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters. 27.
- [9] GÉRON, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- [10] Gutiérrez, R. e. (2023). AI-based process control in resistance seam welding. *IEEE Transactions on Industrial Informatics*,
- [11] Inc., A. (s.d.). *AcuCTFlex Series Rogowski Coil CT Datasheet*. Récupéré sur Accuenergy: <https://accucdn.accuenergy.com/wp-content/uploads/acuct-flex-series-rogowski-coil-ct-datasheet.pdf>
- [12] Jean-Marie Auberville. (2004). *Maintenance industrielle : de l'entretien de base à l'optimisation de la sûreté*. Paris: Ellipses (Collection Technosup).
- [13] Kleniewski, A. (1995). Emballages métalliques. *Les Techniques de l'Ingénieur*, 1–24.
- [14] MARCOPACK. (2023). *Évitez les temps d'arrêt grâce à la maintenance prédictive avec l'IA*.
- [15] Récupéré sur MarcoPack : <https://marcopack.com/fr/evitez-les-temps-darret-grace-a-la-maintenance-predictive-avec-lia/>
- [16] MEZHER, M. P. (2024). Predicting the effect of RSW parameters on the shear force and nugget diameter of similar and dissimilar joints using machine learning algorithms and multilayer perceptron. 17.
- [17] Novak, M. e. (2022). Inline monitoring for seam weld quality in tinsplate cans. *Welding Journal*, 45–52.

Références Bibliographie

- [18] ROUX, B. L. (2016). *management avancé pour l'ingénieur de production* . ellipses.
- [19] SCIKIT-LEARN. (s.d.). *Nearest Neighbors Classification*In. *Scikit-learn User Guide [online]*.
Récupéré sur SCIKIT-LEARN: <https://scikit-learn.org/stable/modules/neighbors.html#classification>
- [20] SCIKIT-LEARN. (s.d.). *Naive Bayes*. In: *Scikit-learn User Guide [online]*. . Récupéré sur SCIKIT-LEARN.: https://scikit-learn.org/stable/modules/naive_bayes.html#naive-bayes
- [21] Shaibu, S. (2024, 10 15). *Normalization vs. Standardization: How to Know the Difference*. Récupéré sur DataCamp: <https://www.datacamp.com/tutorial/normalization-vs-standardization>
- [22] SHEARER, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. 5.
- [23] SOUDRONIC. (s.d.). *Manuel SOUDRONIC – qualité*. EMB-AZZABA.
- [24] SOUDRONIC. (s.d.). *TECHNICAL REPORT*.
- [25] SOUDRONIC. (s.d.). *Technique de soudage* .
- [26] Systems, E. (s.d.). *ESP32 Datasheet*. Récupéré sur Espressif Systems :
https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf
- [27] Technique, d. (s.d.). *Soudage à la molette*. Récupéré sur Technique de l'Ingénieur – Base documentaire Mécatronique.
- [28] Weman, K. (2020). *Procédés de soudage*. Paris : Dunod.