ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

RAPPORT DE PROJET PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE
LA MAÎTRISE EN GÉNIE LOGICIEL

INDUSTRIE 4.0 : PRÉDICTION DE DÉFAUTS DE PIÈCES MANUFACTURÉS À L'AIDE DES DONNÉES D'ASSURANCE QUALITÉ

PAR

Nicolas LEBRUN

MONTRÉAL, LE 29 AVRIL 2019

PRÉSENTATION DU JURY

CE RAPPORT DE PROJET A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

Professeur Alain April, directeur de projet
Département de Génie Logiciel et TI à l'École de technologie supérieure

Professeur Abdelaoued Gherbi, président du jury
Département de Génie Logiciel et TI à l'École de technologie supérieure

REMERCIEMENTS

Je tiens à remercier le professeur Alain April pour m'avoir permis de travailler sur ce sujet lors de mon projet de maîtrise. Je le remercie également pour sa disponibilité et ses conseils durant toute la période du projet.

J'aimerais de même remercier M. Iannick Gagnon pour m'avoir apporté ces conseils et son expérience afin d'améliorer mon travail.

INDUSTRIE 4.0 : PRÉDICTION DE DÉFAUTS DE PIÈCES MANUFACTURÉS À L'AIDE DES DONNÉES D'ASSURANCE QUALITÉ

Nicolas Lebrun

RÉSUMÉ

L'objectif de ce projet de maîtrise appliqué est d'utiliser l'apprentissage machine ainsi que les données d'assurance qualité provenant d'une machine à mesurer tridimensionnelle pour prédire si une pièce automobile manufacturé sera défectueuse avant sa production.

Dans un premier temps, il s'agit d'étudier la littérature pour avoir une vision d'ensemble sur ce qui se fait dans le domaine de cette industrie avec l'apprentissage machine.

Ensuite, les données expérimentales provenant d'une entreprise Québécoise de production de pièces automobile sont analysées pour déterminer des attributs pertinents pour l'apprentissage machine. Il s'agit ici d'une étape de *feature engineering.* Deux catégories d'attributs ont été choisis, l'une sur les aspects géométriques des pièces et l'autre sur la notion de temps et de tendance qualité des dernières pièces manufacturées.

La troisième étape de cette recherche a été le prétraitement des données pour faciliter le processus d'apprentissage. Les données sont ici normalisées sur un intervalle simple [0; 1].

L'étape suivante est l'apprentissage des modèles et l'analyse des résultats pour identifier les performances prédictives obtenues. Pour améliorer ces résultats, les deux étapes précédentes sont itérées. Le choix a été fait d'optimiser les modèles pour obtenir le moins de pièces défectueuses prédites comme conformes.

Le projet se termine par une comparaison de la performance des 4 modèles expérimentés.

Mots clés : Prédiction de défauts des pièces manufacturés, machine à mesurer tridimensionnelle, apprentissage machine, réseau de neurones, forêt aléatoire, machine à vecteur de support, régression logistique.

INDUSTRY 4.0: PREDICTING DEFECTS OF MANUFACTURED PARTS USING QUALITY ASSURANCE DATA

Nicolas Lebrun

ABSTRACT

The goal of this master's project is to use machine learning techniques as well as quality assurance data issued by a coordinate-measuring machine to predict if a manufactured auto part will be defective before its production.

First, studying the literature we can see that it is possible to have an overview of what is done in the manufacturing field concerning machine learning for predicting defects.

Then, the experimental data is analysed to choose relevant features by using feature engineering. Two categories of features were identified for the prediction model training, one on the geometric aspects of the manufactured component and the other using the notion of time and trend of the last manufactured component produced.

The third step is to pre-process this data to facilitate the learning process. The data is normalised on a simple interval [0; 1].

The next step is to train the models and analyse the prediction results to identify the performance achieved. To improve the prediction results, these two steps are iterated. Finally, to end the optimisation, a decision was made to optimise the models to obtain the least defective parts predicted as compliant.

This report ends with a comparison of the models performance.

Key words: manufactured parts defect prediction, coordinate-measuring machines, machine learning, neural network, random forest, support vector machine, logistic regression.

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AM : Apprentissage machine
MVS : Machines à vecteurs de support
AD : Arbre de décisions
RNA : Réseau de neurones artificiels
AMLS : Azure Machine Learning Studio
CSV : Comma separated values

INTRODUCTION

L'industrie 4.0 est l'avenir du secteur automobile. C'est une promesse d'amélioration et d'optimisation des processus pour arriver à réduire les coûts et augmenter la productivité le tout en réduisant l'empreinte carbone de la fabrication.

L'apprentissage machine est l'un des volets de l'industrie 4.0. Elle a pour vocation d'intervenir dans tous les secteurs de l'industrie automobile, aussi bien dans la planification et la maintenance que les relations clients.

Dans ce contexte, ce projet de recherche appliquée, à l'École de Technologie Supérieure, a pour objectif de se pencher sur l'utilisation de l'apprentissage machine pour prédire, avant sa production, si une pièce manufacturée sera conforme ou non à partir de données historiques d'assurance qualité provenant des machines à mesurer tridimensionelles. Cet enjeu permettrait de grandement améliorer la l'efficacité et la productivité des industriels du secteur de la production de pièces automobile Québécois. Il deviendrait possible d'intervenir durant le processus de fabrication pour apporter les modifications nécessaires selon les prédictions.

Pour atteindre cet objectif, plusieurs étapes sont nécessaires. Dans un premier temps, une grande quantité de données de production de pièces doivent être collectées. Pour ce projet, il s'agit des données historiques générées par des machines à mesures tridimensionnelles. Une fois collectées, ces données sont traitées pour faciliter les prochaines étapes et améliorer les résultats prédictifs. L'étape suivante est la phase d'apprentissage du modèle. Elle consiste à utiliser ces données pour entrainer les modèles. Enfin, chaque modèle expérimenté est évalué sur des données de test pour pouvoir l'améliorer si possible par la suite.

Le projet est divisé en trois parties. Dans un premier temps, une étude de la littérature est effectuée pour observer ce qui est déjà publié dans ce domaine et les pratiques utilisées par

cette industrie spécifique. Les différentes utilisations de l'apprentissage machine y sont résumé, les contextes spécifiques et les principales techniques utilisées. La deuxième section du rapport traite des étapes pour créer un modèle prédictif. Les données utilisées sont détaillées ainsi que les attributs choisis et l'outil d'apprentissage machine expérimenté. Quatre algorithmes d'apprentissage sont expérimentés : 1) les réseaux de neurones; 2) les forêts aléatoires 3) les machines à vecteurs de support; et 4) la régression logistique. La troisième partie du rapport présente les résultats de chaque modèle et discute des résultats obtenus.

CHAPITRE 1

Revue de littérature

1.1 Introduction

Grâce à l'Internet des Objets, l'industrie entre dans une nouvelle ère. Les composants autrefois indépendants présents dans une industrie sont maintenant surveillés et liés. Il s'agit de l'industrie 4.0 et ses usines intelligentes. Cette revue de littérature a pour objectif de montrer qu'il est possible de faire des prédictions, dans cette industrie grâce, à l'apprentissage machine (AM) et d'identifier les différentes techniques utilisées. Dans un premier temps, les différentes utilisations de l'AM dans l'industrie seront synthétisées. Ensuite, les industries qui utilisent ce procédé pour prédire la qualité d'un produit seront détaillées. Les techniques utilisées pour prédire la qualité seront détaillées. Enfin, les résultats obtenus pour chacune des techniques seront résumés.

1.2 Les différentes utilisations de l'apprentissage machine dans l'industrie

La prédiction dans l'industrie 4.0 est en pleine évolution. Elle est utilisée pour anticiper des états réels en fonction de données d'entrées. Cette section présente une vision d'ensemble des applications publiées de l'AM pertinentes à cette recherche.

1.2.1 Les systèmes de support à la décision

Il s'agit de donner des outils d'aide à la prise de décision dans le domaine industriel. Par exemple, la logique floue a été utilisée pour modifier les planifications dans les travaux de Grabot et coll. (7). Aussi, des outils ont été développés en utilisant les arbres de décisions (AD, en anglais : *decision trees*), entre autres, pour trouver des motifs dans les chaînes de production. Toutes les techniques développées se basent sur l'exploration de données (« Data mining »).

1.2.2 Les systèmes de gestion des relations clients

Cette facette de la prédiction cherche à anticiper le comportement des clients. Là aussi, l'exploration de données est une technique très utilisée. Les machines à vecteurs de support (MVS, en anglais : *support vector machines*, *SVM*) sont une méthode qui a été utilisée dans les travaux de Tsang et coll. [8] pour prédire le fournisseur préféré ce qui revient à un problème de classification multiple.

1.2.3 Les systèmes de gestion d'atelier de fabrication

L'AM est utilisé dans ce domaine pour estimer les délais de productions et la répartition optimale du travail pour optimiser les performances de production en utilisant au mieux les ressources disponibles. Les AD ont été utilisés pour prédire un délai de production. Des travaux ont été réalisés par Wang et coll. (9) en utilisant un système hybride d'AD avec un réseau de neurones artificiels (RNA, en anglais : *artificial neural network, ANN*) pour déterminer la règle de répartition adéquate à partir des données de production.

1.2.4 Les systèmes de production

Certains projets se sont penchés sur les systèmes de production pour, par exemple, déterminer l'influence de certains paramètres lors de la production. Ainsi, l'exploration de données a été utilisée pour identifier l'impact de paramètres dans l'industrie des semi-conducteurs par Backus et coll. (10). Le temps de cycle des produits a été étudié en utilisant les AD, et la méthode des k plus proches voisins.

1.2.5 Les systèmes de maintenance

Cette branche de la prédiction permet d'anticiper des pannes et ainsi améliorer les performances de la maintenance préventive. Elle utilise des techniques comme la régression,

les réseaux de neurones ou encore les arbres de décisions. Lin et coll. (11) ont ainsi utilisé une méthode de RNA (en anglais : *cerebellar model articulation controller, CMAC*) pour synthétiser des informations de bas niveau comme des signaux de vibrations avec des informations de haut niveau comme des statistiques de fiabilité pour ainsi améliorer la maintenance.

1.2.6 Les systèmes de gestion/prédiction de la qualité

La prédiction de la conformité des pièces avant leur production est le domaine qui nous intéresse. Il permet de déterminer, en cours de production, la qualité finale du produit pour pouvoir ainsi ajuster des paramètres. Des méthodes comme les MVS, les AD, les RNA et de manière générale, les techniques de régression sont les plus courantes dans la prédiction de la qualité. La section suivante détail des utilisations de ces méthodes.

1.3 Les industries qui utilisent l'apprentissage machine pour prédire la qualité

La prédiction de la qualité d'un produit est utilisée dans plusieurs types d'industries bien précises et spatialisées où le processus de fabrication est maîtrisé et la pièce contrôlée. L'objectif de cette section est de présenter une synthèse des publications et faire un parallèle avec l'industrie de la fabrication de pièces pour l'automobile.

1.3.1 L'industrie de la fabrication de semi-conducteurs

Chatterjee et coll. (6) ont travaillé sur la prédiction de la résistance de plaques de semi-conducteur (en anglais : *wafers*). Leurs travaux couvrent toutes les étapes nécessaires à la mise en place de l'AM pour prédire la résistance. Les auteurs ont cherché à collecter les données utiles. Ainsi il a fallu réduire la quantité de données en sélectionnant les meilleurs paramètres et en moyennant les valeurs de certains paramètres sur le temps. En effet, il s'agit d'une étape importante, car la qualité des données conditionne le résultat. Les données sont divisées en trois catégories :

- Les données qui caractérisent l'état de la machine durant la gravure ;
- Les données de métrologie pour les mesures de dimensions critiques ;
- Les mesures de résistance électrique.

Une fois ces données récoltées, il a fallu faire un prétraitement pour ne garder que les données utiles et ainsi améliorer les performances de l'algorithme d'AM. Les auteurs ont décidé de comparer plusieurs algorithmes d'AM :

- Les RNA ;
- Le raisonnement par cas (en anglais : *case-based reasoning*) ;
- Les AD

En plus de prédire la valeur de résistance électrique, les auteurs ont utilisé l'AD pour observer les facteurs ayant les plus d'impact. Ainsi l'AD aide à la compréhension des résultats obtenus.

Les méthodes d'AM utilisées ont démontrés des résultats concluants arrivant à prédire la résistance électrique du semi-conducteur à +/- 1 ohm dans plus de 90% des cas.

1.3.2 L'industrie de la fabrication des microprocesseurs

Dans leurs travaux, Weiss et coll. (4) cherchent à prédire la vitesse d'un microprocesseur avant sa fabrication. Le modèle de prédiction idéal doit pouvoir anticiper si la vitesse du microprocesseur est trop lente ou trop rapide pour pouvoir appliquer des mesures correctives durant sa fabrication. Un microprocesseur trop lent n'a pas les performances attendues et un microprocesseur trop rapide consomme davantage. Les actions correctives doivent être appliquées avant la moitié de la fabrication. L'une des difficultés rencontrées par ces chercheurs est la quantité de données manquante, car elles sont échantillonnées avec un taux de 10%.

Les étapes suivies pour prédire les valeurs de performance des microprocesseurs sont les mêmes que pour les semi-conducteurs.

Les méthodes d'apprentissages utilisées sont les suivantes :

- La régression linéaire (en anglais : *linear regression*) ;
- Les forêts aléatoires (en anglais : *random forest*).

Ces deux méthodes sont combinées en effectuant une moyenne des résultats obtenus. Ils précisent que cette stratégie est efficace dans le cadre d'une production temps réel.

Les auteurs ont réalisé deux scénarios de test. L'un en faisant une simulation en utilisant des données historiques complètes. Le second scénario est un test réel en production à plus petite échelle. La précision obtenue lors du premier scénario est très bonne. De l'ordre de 90% des microprocesseurs déterminés comme hors limite sont effectivement à corriger. Cependant, le second test montre des prédictions plus décevantes, de l'ordre de 65%. D'après les auteurs, des expériences durant le second test sur le processus ont influencé les données, expliquant la différence.

1.3.3 La fabrication par machine à commandes numériques

Dans le cadre des machines à commande numérique, des travaux de recherche ont déjà été effectués, la plupart sur les opérations de tournage. Les travaux de Han et coll. (1) ont pour objectif de prédire l'usure de l'outil pour ajuster le décalage de l'outil. Les auteurs commencent par trier les données en trois catégories :

- Les caractéristiques du matériau ;
- Les données de la machine ;
- Les données d'inspection du produit.

Ces trois catégories coïncident avec celles présentées dans les travaux de Chatterjee et coll (6).

Le problème de prédiction est identifié comme étant continu et non linéaire, ainsi les auteurs ont décidé d'utiliser les MVS. Pour considérer la non-linéarité, le noyau utilisé est une fonction à base radiale gaussienne.

1.4 Les techniques utilisées pour prédire la qualité et leurs résultats

Cette section résume les méthodes d'AM utilisées dans la littérature. L'objectif est d'identifier les méthodes utilisables dans le cadre du projet.

1.4.1 La régression linéaire et logistique

La technique de prédiction à l'aide d'une régression linéaire, souvent rencontrée dans la littérature, ne s'applique pas au projet, car il est demandé de prédire des valeurs discrètes. Ainsi une pièce fabriquée est considérée défectueuse ou non. Il s'agit d'un problème de classification binaire.

En revanche, la régression logistique peut être utilisée pour résoudre des problèmes de classification binaire. Elle n'a pas été développée dans les articles retenus, mais s'applique totalement dans le cadre du projet.

1.4.2 Les machines à vecteurs de support

Les techniques de MVS sont un ensemble de méthodes d'AP utilisées dans beaucoup de publications de prédiction. Par exemple dans l'article de Çaydaş et coll. (2), cette méthode est utilisée pour déterminer la rugosité d'une surface après une opération de tournage. Les auteurs comparent ainsi différentes boîtes à outils en langage MATLAB. Dans ces travaux de recherche, les auteurs ont également comparé les résultats obtenus par MVS et RNA. Ils

observent que les SVM obtiennent de meilleurs résultats, dans leur cas précis. De plus, le temps de calcul est de moins d'une seconde pour la MVS contre 123 secondes pour le RNA.

L'utilisation de MVS est très rependue dans la littérature. Han et coll. (1) l'utilisent pour prédire l'usure d'un outil d'usinage. Cependant, comme beaucoup d'autres articles, la méthode MVS est utilisée pour des problèmes de régression qui ne sont pas le sujet de ce projet.

Dans leurs travaux, Mohammadi et coll. (3), utilisent aussi une MVS pour des problèmes de classification afin de déterminer la qualité du produit en séparant les observations en trois classes (A, B et C). Pour se faire, ils divisent le problème en trois classifieurs binaires spécialisés (A vs. B, A vs. C et B vs. C). Ils comparent le cas d'un problème linéaire et non linéaire. Leurs résultats démontrent que le modèle utilisant une fonction à base radiale gaussienne obtient de meilleurs résultats que le modèle linéaire. De plus, ils proposent que les classifieurs sont plus performant lorsque la différence entre les deux classes est grande. Ces travaux sont intéressants, car dans le cas de ce projet de recherche appliquée, il s'agit de réaliser un classifieur pour prédire si la pièce sera défectueuse ou non.

1.4.3 Les réseaux de neurones

Les techniques de prédiction basées sur les réseaux de neurones sont aussi populaires dans les travaux réalisés dans le domaine de la prédiction de la qualité. Bai et coll. (5) utilisent une architecture d'apprentissage de réseaux de neurones profonds (en anglais : *deep neural network, DNN*). Cette architecture est composée d'une couche d'entrée, une couche cachée et une couche de sortie. La couche cachée est un réseau de conviction profonde (en anglais : *deep belief network, DBN*) qui n'est autre que plusieurs machines de Boltzmann restreinte. Dans leur expérience, cette architecture a permis d'obtenir des résultats plus précis, ainsi l'erreur absolue moyenne en pourcentage de la prédiction ne dépasse pas 4,5%.

Chatterjee et coll. (6) ont également utilisé des RNA pour déterminer la valeur des résistances produites. Le réseau de neurones résultant possède 16 nœuds d'entrées avec une couche cachée de 3 à 15 nœuds puis un nœud de sortie. Ce réseau de neurones a été capable de déterminer la résistance à plus ou moins un ohm dans plus de 90 % des cas.

1.4.4 Les forêts aléatoires

Dans leurs travaux, Weiss et coll. (4) utilisent la méthode des forêts aléatoire pour déterminer la vitesse des microprocesseurs durant leur fabrication. À l'aide des données historiques, s'étalant sur deux mois, soit 24 lots de microprocesseurs produits sur cette période, le modèle a déterminé que 3 lots étaient inférieurs à la limite et 3 étaient supérieurs à la limite. À la suite de vérifications, ces 6 lots étaient effectivement hors-normes.

Cette technique est intéressante, car une fois les résultats obtenus, il est possible d'analyser les arbres pour observer les facteurs qui influencent plus le résultat final (Chatterjee et coll. [6]).

1.5 Le prétraitement des données

Avant l'apprentissage, les étapes importantes sont la collecte et la préparation des données. Il s'agit souvent des étapes les plus longues. Il est important, dans beaucoup de travaux, de bien sélectionner les données pour faciliter l'apprentissage comme le soulignent les travaux de Chatterjee et coll. (6). Ils ont diminué les variables venant des capteurs de 60 à 15. Une première étape a consisté à supprimer les variables constantes, aléatoires ou redondantes pour ainsi descendre à 30. Ensuite une analyse en composantes principales (en anglais : *principal component analysis*) des 30 variables a été réalisée pour sélectionner les 15 premiers. Les auteurs ont également fait une moyenne à travers le temps de plusieurs valeurs.

Enfin, les données doivent être mises à l'échelle. Pour leur prétraitement, Han et coll. (1) utilisent deux méthodes différentes :

- La standardisation ;
- La normalisation ([0, 1] et [-1, 1]).

Les méthodes qui obtiennent les meilleurs résultats dans leurs travaux sont la normalisation de [0,1] et la standardisation. La normalisation de [-1, 1] donne des résultats nettement moins bons. Une conclusion qui peut être tirée de cette étude est que la méthode de prétraitement choisie est très importante, car elle impacte significativement les résultats obtenus. De plus, des données qui ne sont pas préparées détérioreront grandement les prédictions, car le modèle doit travailler sur des données étalées sur une grande plage.

Pour le projet, il est donc nécessaire, dans un premier temps, de bien identifier les attributs principaux. Ensuite il sera sans doute utile de traiter les données, en privilégiant la normalisation de [0, 1], car les données n'auront surement pas une distribution gaussienne.

1.6 Conclusion

Cette revue littéraire a présenté qu'il existe de nombreuses méthodes de prédiction dans les différentes industries de fabrication et que les techniques sont variées et dépendent de chaque situation. Le prétraitement des données est nécessaire pour réduire la nuisance des données mesurées sur différentes échelles et des données manquantes pour obtenir des résultats plus précis. De plus, la plupart des études classent les données de la même manière, en classant les informations intrinsèques au matériau d'usinage, les données de la machine en cours d'usinage puis les mesures du produit une fois usiné. Toutes les techniques d'AM ne seront pas adaptées au projet, car il s'agit d'un problème de classification binaire. Il a été démontré que les RNA et les AD semblent être des techniques fiables. Les travaux montrent que ces techniques permettent d'obtenir des résultats de prédiction de l'ordre de 90%. Il est également possible de combiner plusieurs méthodes pour diminuer l'influence de certains

attributs. La littérature a permis d'avoir une idée générale du domaine pour pouvoir créer un modèle qui puisse déterminer, à partir des données d'assurance qualité, si une pièce sera défectueuse ou non.

CHAPITRE 2

Création d'un modèle

2.1 Les données d'assurance qualité

Les données d'assurance qualité utilisées pour la prédiction des défauts sont constituées de 9363 fichiers CSV. Un fichier regroupe les différentes mesures effectuées sur une pièce tel que présenté en Annexe I.

Ces fichiers sont constitués d'un entête pour identifier la pièce. L'échantillon de fichiers comprend 101 pièces différentes présentées en Annexe II.

Les fichiers comprennent différents champs afin d'obtenir des informations concernant les mesures.

Champ	Valeur
Feat. Type	Flatness, Circle, Pos. Tol., Distance, Point, Profile of Line, Profile of Surface, Perpend., Angle, Sphere, Text/Value, Slot, Paral., Coaxiality, Plane, Section, Cone, Surf. Pnt
Feat. Name	Nom de la mesure
Value	Valeur associée à la mesure
Actual	Valeur réelle mesurée
Nominal	Valeur de référence
Dev.	L'écart entre la valeur réelle et la valeur nominale
Tol-	Tolérance inférieure
Tol+	Tolérance supérieure
Out of Tol.	Précise de combien la mesure est en hors tolérance si c'est la cas
Comment	Donne des commentaires sur la mesure

Une fois les données détaillées, il est important d'être capable de savoir si une pièce est conforme ou non. Pour cela il suffit d'observer la colonne « Out of Tol ». Si la colonne est vide, cela signifie qu'aucune mesure sur la pièce n'est hors tolérance. Ainsi on considère la pièce comme conforme. Inversement, une ou plusieurs valeurs différentes de zéro dans cette colonne prouvent que la pièce est non-conforme comme le décrit l'Annexe I.

Maintenant que la non-conformité est définie, il est possible d'évaluer l'échantillon des pièces mesurées. Sur les 9363 pièces évaluées, 6247 comportent au moins une côte hors tolérance ce qui donne environ 67% de pièces défectueuses dans le lot étudié.

2.2 Choix des attributs d'entrée du modèle

Après avoir détaillé les données, la prochaine étape consiste à sélectionner les attributs d'entrée pour le modèle prédictif. Le choix va se porter sur les données disponibles avant la production de la pièce qui pourraient influencer sa qualité.

2.2.1 Nombre de mesures avec une tolérance

Un premier attribut qu'il est possible d'obtenir à partir d'un fichier de mesures d'une pièce est le nombre de mesures avec tolérances. La figure 1 présente la proportion de pièces défectueuses en fonction du nombre de mesures avec tolérance.

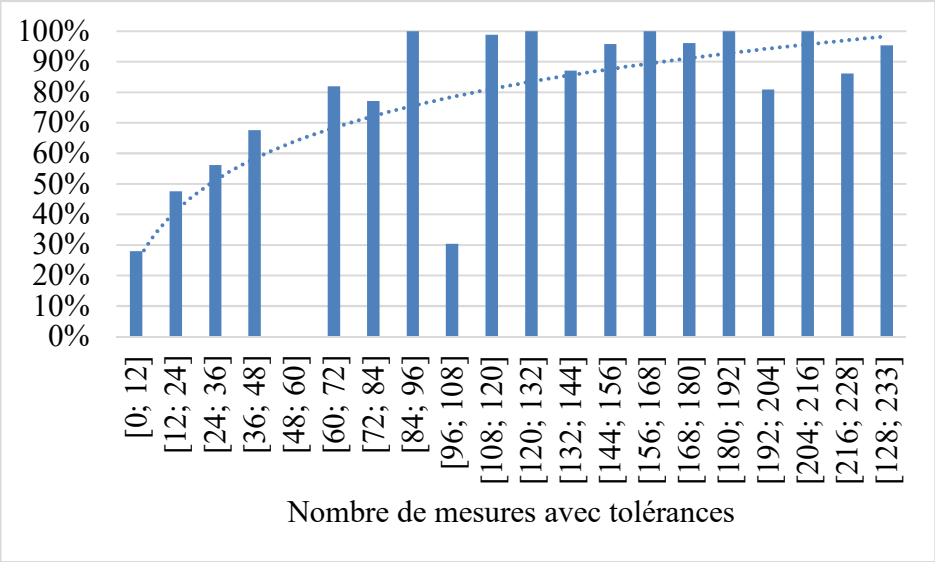

Figure 1 Relation entre la proportion de pièces défectueuses et le nombre de mesures avec tolérances

À la figure 1, la tendance démontre que le nombre de défauts augmente avec le nombre de mesures avec tolérances. Il est possible de conclure un lien entre le nombre de tolérances et la qualité de la pièce.

2.2.2 La déviation maximale

Un autre attribut qu'il est possible d'obtenir, pour chaque pièce, est la déviation maximale (colonne « Dev. » des fichiers CSV). Il s'agit de la valeur maximale de l'écart entre la valeur réelle mesurée et la valeur nominale. La figure 2 présente le pourcentage de pièces non conforme en fonction de la déviation maximale.

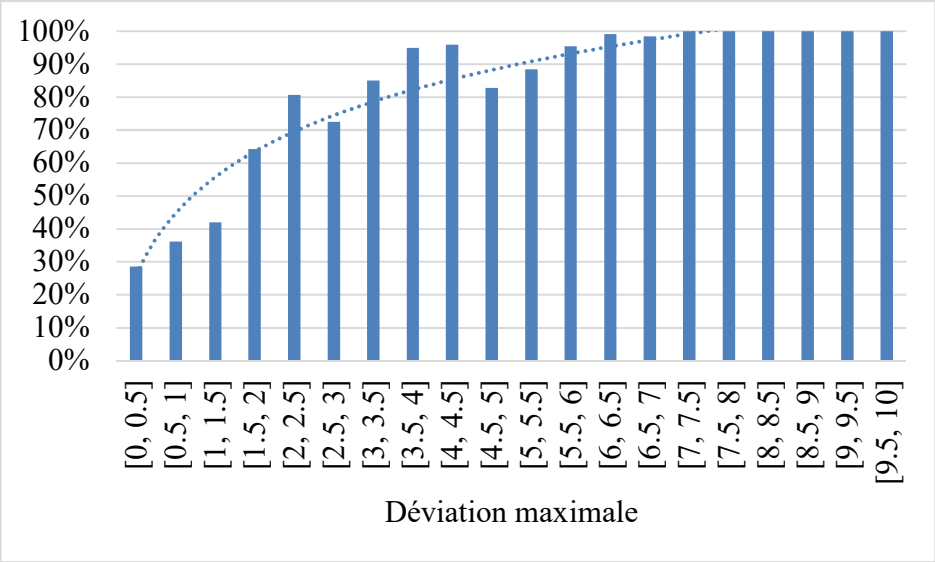

Figure 2 Relation entre la proportion de pièces défectueuses et la déviation maximale d'une pièce

Encore une fois, il est possible d'observer que le pourcentage de défauts augmente pour de déviation de plus en plus grande. Il existe donc un lien entre la déviation maximale observée sur une pièce et sa qualité. Cependant cette valeur étant obtenue après la production de la pièce, elle ne sera pas utilisée par la suite.

2.2.3 Tolérance maximale

De la même manière, il est possible d'imaginer un attribut qui prenne en compte la tolérance maximale de la pièce. Voici en figure 3 la mise en relation entre les pièces défectueuses et la tolérance maximale de chacune d'elle.

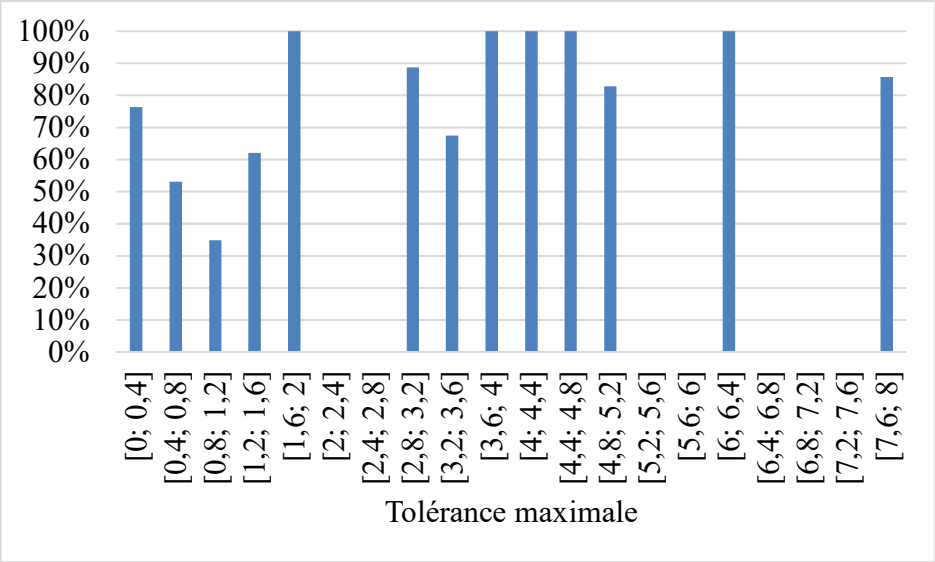

Figure 3 Relation entre la proportion de pièces défectueuses et la tolérance maximale d'une pièce

Il est difficile de conclure un lien entre la tolérance maximale et le nombre de défauts trouvés. Ce paramètre ne sera donc pas retenu pour la suite.

2.2.4 Types de mesures

Un attribut supplémentaire pourrait être de comptabiliser le nombre des types de mesures par pièce. Il est possible d'imaginer que certains types de mesures découvrent plus de défauts sur les pièces. Le tableau 1 présente le pourcentage de mesures hors tolérance par type de mesure.

Tableau 1 Pourcentage de défauts par type de mesures

Type de mesures	Nb de mesures avec tolérance	Nb de mesures hors tolérance	Pourcentage [%]
Distance	516675	22823	4
Pos. Tol.	188645	6080	3
Circle	92099	7927	9
Profile of Line	35366	1044	3

Flatness	11793	874	7
Text/Value	8371	374	4
Perpend.	5065	21	0
Paral.	3236	239	7
Slot	2926	2	0
Point	2406	1	0
Section	1884	157	8
Angle	809	0	0
Coaxiality	406	36	9
Sphere	78	3	4
Cone	15	15	100
Profile of Surface	12	0	0
Surf. Pnt	8	8	100
Plane	0	0	0
Total	**869794**	**39604**	**5**

Il est possible d'observer que certains types de mesures contiennent plus de défauts. Par exemple, l'attribut « circle » a 9% de non-conformité, ce qui est quasiment deux fois plus que la moyenne de 5%. Les types faiblement représentés obtiennent rapidement des valeurs extrêmes, ils ne seront pas compris par la suite.

2.2.5 Le nom de la pièce

L'attribut suivant est le nom de la pièce. En effet, certaines pièces sont plus affectées par les défauts que d'autres. L'Annexe III présente le pourcentage de défauts par type de pièce. On observe clairement qu'une trentaine de types de pièces dans notre échantillon ont un taux de non-conformité de 100%. Cependant, prendre le nom de la pièce comme attribut ne permet pas de différencier les pièces du même nom entre elles. Cela reviendrait à admettre qu'une pièce peut être défectueuse simplement parce que son nom est associé à de nombreuses non-conformités. Même si dans les données actuelles, certains types de pièces ont 100% de pièces défectueuses, cela ne veut pas dire que par la suite les pièces de ce type soient conformes, ce

qui est très probable. De plus, le modèle ne fonctionnerait pas en cas de nouvelle pièce ou de faute de frappe d'un opérateur.

Cependant, il est possible d'obtenir d'autres informations du nom de la pièce. En effet, certaines expressions sont récurrentes dans les noms :

- « Complet » ou « Complete ».
- « Free State »

De la même manière, un numéro de module est souvent utilisé dans le nom : module 44, module 51, module 16, module 18, module 64, module 67, module 50, module 01

Le tableau 2 présente le pourcentage de pièces défectueuses associé à une expression.

Tableau 2 Pourcentage de pièces défectueuses par expression

Formule	Total	Nombre de pièces défectueuses	Proportion [%]	Écart [%]
Complet	6238	5276	85	18
Free State	1577	366	23	44
Module 44	3070	1571	51	16
Module 51	4091	3117	76	9
Module 16	15	14	93	26
Module 18	450	362	80	13
Module 64	2296	1571	68	1
Module 67	56	46	82	15
Module 50	463	351	76	9
Module 01	232	179	77	10

La colonne « Écart » permet d'obtenir la différence du pourcentage de l'expression par rapport au pourcentage de pièces défectueuses global de l'échantillon qui est de 67%. Les expressions avec un grand écart peuvent permettre de différentier plus facilement la qualité de ces pièces. Pour la suite, seulement les modules ayant un écart conséquent avec la moyenne et étant représenté souvent. Cela va permettre de ne pas trop augmenter le nombre d'attributs, car le module est un attribut catégorique nominal. Seulement le module 44 et 51 seront utilisés par la suite.

2.2.6 Notion de temps

Maintenant que les aspects géométriques sont couverts, la notion de temps et d'ordre de fabrication des pièces va être ajoutée aux attributs. Pour ce faire, trois attributs vont être créés pour indiquer si dans les 10, 5 ou 2 dernières pièces produites, une pièce du même nom est défectueuse. Leur valeur sera mise à 1 lorsqu'une pièce défectueuse du même nom sera présente.

Un attribut sera également utilisé pour savoir si la dernière pièce produite ayant le même nom était conforme ou non.

Enfin deux tendances seront évaluées. La première tendance concerne la déviation moyenne des deux dernières pièces produites du même nom. La seconde se penche sur la déviation moyenne des deux dernières pièces produites.

2.2.7 Récapitulatif des attributs

Le tableau x récapitule les attributs utilisés ainsi que les valeurs observées sur les données actuelles avant normalisation.

Tableau 3 Récapitulatif des attributs

No	Attributs	Valeurs
1	Nombre de mesures avec une tolérance	[0; 233]
2	Nombre de mesures de type "Distance" avec tolérance	[0; 176]
3	Nombre de mesures de type "Pos. Tol." avec tolérance	[0; 69]
4	Nombre de mesures de type "Circle" avec tolérance	[0; 204]
5	Nombre de mesures de type "Profile of Line" avec tolérance	[0; 19]
6	Nombre de mesures de type "Flatness" avec tolérance	[0; 5]
7	Expression "complet" présent dans le nom	0 ou 1
8	Expression "free state" présent dans le nom	0 ou 1
9	Expression "module 44" présent dans le nom	0 ou 1
10	Expression "module 51" présent dans le nom	0 ou 1
11	Présence d'une pièce défectueuse du même nom dans les 10	0 ou 1

	dernières pièces produites	
12	Présence d'une pièce défectueuse du même nom dans les 5 dernières pièces produites	0 ou 1
13	Présence d'une pièce défectueuse du même nom dans les 2 dernières pièces produites	0 ou 1
14	Etat de la dernière pièce produite du même nom	0 ou 1
15	Tendance de la déviation moyenne des 2 dernières pièces du même nom	0 ou 1
16	Tendance de la déviation moyenne des 2 dernières pièces	0 ou 1

2.3 Mise en forme des données

Il est nécessaire maintenant de récupérer les valeurs des attributs pour chaque pièce. Pour ce faire, un programme codé en Python le module « csv » boucle sur les 9363 fichiers. Pour chacun des fichiers, le programme boucle sur toutes les lignes des fichiers. Il récupère ainsi le nom de la pièce à la ligne 2. En passant à travers toutes les lignes, il incrémente les différents compteurs. Le programme prétraite le nom de la pièce en éliminant les caractères inutiles. De plus, lors de la boucle, le programme vérifie si les mesures sont hors tolérance. Un attribut est ainsi ajouté à la pièce qui spécifie si elle est défectueuse ou non. Il est traduit par un 0 si la pièce est conforme est 1 si elle est non conforme. Ainsi, à la fin de chaque fichier, une ligne est écrite dans le nouveau fichier csv qui récapitule les valeurs des paramètres pour chaque fichier. Ce fichier sera nommé « Dataset » par la suite. L'Annexe IV présente quelques lignes de ce fichier.

2.4 Outil utilisé pour l'apprentissage

Maintenant que les données sont générées pour l'apprentissage, il faut commencer à générer un modèle. Pour ce faire, Microsoft Azure Machine Learning Studio sera utilisé. Il s'agit d'un environnement simple et puissant basé sur une interface visuelle sur navigateur. De plus, une documentation claire et des exemples simples et complexes sont disponibles pour facilement prendre en main le système.

Grâce à cette solution, il sera possible de mettre en place plusieurs modèles et de les comparer tout en agissant sur les données pour le prétraitement.

2.5 Prétraitement des données

Présentation de la normalisation : explication + exemple sur un des paramètres
Une fois les données générées et ajoutées à Azure Machine Learning Studio, il est nécessaire de les prétraiter pour avoir des données de qualité et améliorer les résultats.

Les données étant relativement simples, le prétraitement aura peu d'étapes. Dans un premier temps, il est nécessaire de sélectionner les colonnes intéressantes pour l'apprentissage. Ici, il suffit de prendre en compte toutes les colonnes sauf la colonne « ID ».

Ensuite, il ne peut pas y avoir de données manquantes, car toutes les valeurs sont initialisées à 0 avant d'être incrémentées dans le programme. Donc il n'est pas nécessaire de gérer les données manquantes.

La prochaine étape est de réduire la plage des valeurs. Tous les paramètres numériques sont ainsi normalisés entre [0; 1] suivant la formule suivante :

$$x_i := \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

2.6 Modèles comparés

Plusieurs modèles seront comparés pour observer quel algorithme sera le plus performant. Quatre algorithmes ont été sélectionnés en fonction du travail de recherche effectué précédemment.

2.6.1 Réseaux de neurones

Le réseau de neurones est l'un des algorithmes les plus utilisés dans la littérature (5, 6, 9, 11) pour la prédiction. Un réseau de neurones est composé d'une couche d'entrée et d'une couche de sortie liées par différentes couches constituées de nœuds liés d'une couche à l'autre par des branches synaptiques avec des poids différents. Le nombre de couches au milieu dépend de la complexité du problème. Les poids sont modifiés à chaque itération de l'apprentissage.

Le réseau de neurones présente quelques avantages. Il est polyvalent et peut ainsi traiter une grande variété de problèmes. Il a été démontré expérimentalement que les réseaux de neurones sont capables d'atteindre des précisions élevées pour une variété de différents problèmes.

2.6.2 Forêts aléatoires

Pour créer une forêt, il faut premièrement échantillonner les données pour obtenir plusieurs échantillons (en anglais : *bootstraps*). Ensuite, pour chaque échantillon de données, un AD est créé. Cependant lors de sa création, au moment de couper un nœud, un sous ensemble des attributs est tiré au hasard parmi tous les attributs pour ensuite avoir le meilleur découpage. Ainsi tous les AD sont différents. Pour terminer, le résultat est obtenu par vote majoritaire de tous les AD. Il s'agit de la méthode du *bagging*.

Les forêts aléatoires sont adaptées à la classification ainsi qu'à la régression. Un des avantages d'utiliser des AD est qu'ils sont faciles à visualiser et interpréter. Enfin elle permet de gérer de larges quantités de données avec une grande dimension.

2.6.3 Machines à vecteurs de support

Cet algorithme permet, lors de l'apprentissage, de cartographier les données d'entrées dans un espace mathématique multidimensionnel en tant que points. Ainsi il est possible d'identifier des catégories en les séparant par des hyperplans. Pour prédire, l'algorithme place les nouvelles entrées dans ce même espace.

Cette méthode est efficace même pour un grand nombre d'attributs. De par son fonctionnement, elle fonctionne lorsque le ratio du nombre de données sur le nombre d'attributs est faible.

2.6.4 Régression logistique

La régression logistique est un modèle linéaire qui permet de classifier en prédisant 0 ou 1. La fonction linéaire peut s'écrire ainsi :

$$S\left(X^{(i)}\right) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_n x_n \quad (2)$$

Avec

- $X^{(i)}$: une variable à prédire sous forme de vecteur ;
- x_i : les attributs de $X^{(i)}$;
- θ_i : les poids des attributs ;
- θ_0 : le biais.

Le but est de déterminer les poids des attributs pour que la fonction soit positive lorsque la classe vaut 1 et négative lorsque la classe vaut 0. Ensuite la fonction sigmoïde (3) est utilisée pour obtenir des valeurs comprises entre 0 et 1.

$$Sigmoid(x) = \frac{1}{1+e^x} \quad (3)$$

Le résultat peut être interprété comme la probabilité d'obtenir la classe 1.

2.7 Paramètres des modèles

La prochaine étape avant l'apprentissage est de paramétrer les algorithmes. Azure Machine Learning Studio propose un outil qui permet d'obtenir les meilleurs paramètres de chaque modèle. Une plage de paramètres est spécifiée par l'utilisateur puis l'outil teste les différentes combinaisons de paramètres pour obtenir le meilleur résultat en comparant les exactitudes obtenues par chaque modèle.

2.8 Évaluation des modèles

Pour évaluer les modèles lors de la classification, plusieurs métriques existent.

- **Matrice de confusion** : Cette matrice donne le nombre de vrais positifs ainsi que les faux positifs ce qui représente respectivement les prédictions positives bonnes et les prédictions positives fausses.

Tableau 4 Matrice de confusion

		Réel	
		Oui	Non
Prédit	Oui	Vrai positif	Faux positif
	Non	Faux négatif	Vrai négatif

Donc dans le cas du projet, vrai positif représente une prédiction d'une pièce défectueuse réussie.

- **Faux négatifs** : Ils vont être intéressants à observer, car ils représentent les pièces défectueuses prédites comme non défectueuses. Si, par exemple, les pièces coûtent cher à produire, il s'agira de réduire au maximum cette métrique.

- **La précision** : il s'agit d'évaluer le nombre de bonnes prédictions positives par rapport au total de valeurs positives. Cette métrique donne la qualité du modèle pour prédire ce pour quoi il a été entrainé. Pour le projet, il s'agit d'une bonne indication sur la capacité du modèle à prédire de manière efficace un défaut en gardant un œil sur les faux positifs. Voici la formule :

$$Précision = \frac{Vrais\ positifs}{Vra\quad positif + Fau\quad positif} \tag{4}$$

- **Le taux de faux négatif** : il s'agit d'une métrique qui représente le pourcentage de pièces défectueuses prédites comme non défectueuses sur le total de pièces défectueuses :

$$Taux\ de\ faux\ négatifs = \frac{Faux\ négatifs}{Faux\ négatifs + Vrais\ positifs} \tag{5}$$

Ces métriques vont permettre d'évaluer la capacité du modèle à prédire les pièces non conformes.

CHAPITRE 3

Résultats

3.1 Réseau de neurones

3.1.1 Paramètres

Les seuls paramètres sur lesquels il est possible d'agir sur AMLS sont les suivants :

- Nombre de nœuds de la couche cachés
- Le taux d'apprentissage
- Le nombre maximum d'itérations d'apprentissage
- Le poids initial
- Le momentum

Il est impossible pour l'utilisateur de gérer les algorithmes ni même de savoir lesquels sont utilisés.

Pour obtenir les paramètres, AMLS propose un module qui permet de faire varier les paramètres sur une certaine plage pour ensuite indiquer la meilleure combinaison en fonction des précisions obtenues. J'ai commencé par indiquer des plages relativement importantes pour ensuite ne faire varier manuellement qu'un paramètre à la fois.

Le premier paramètre évalué est le nombre de nœuds de l'unique couche cachée. Il peut être obtenu par la formule empirique suivante :

$$Nombre\ de\ n\text{œ}uds = \sqrt{n\text{œ}uds\ d'entrée + \ n\text{œ}uds\ de\ sortie} + A \text{ avec } A \in [0; 10]$$

Le nombre de nœuds varie donc entre 4 et 14. La figure 4 présente le nombre de faux négatifs en fonction du nombre de nœuds. Le nombre de faux négatif est utilisé ici pour déterminer les meilleurs paramètres. Il aurait été possible de les évaluer en fonction de la

précision, cela peut dépendre de la volonté de l'industriel, car souvent un faible nombre de faux négatifs entraine un nombre de faux négatif plus important. L'échantillon de test est constitué de 2809 pièces.

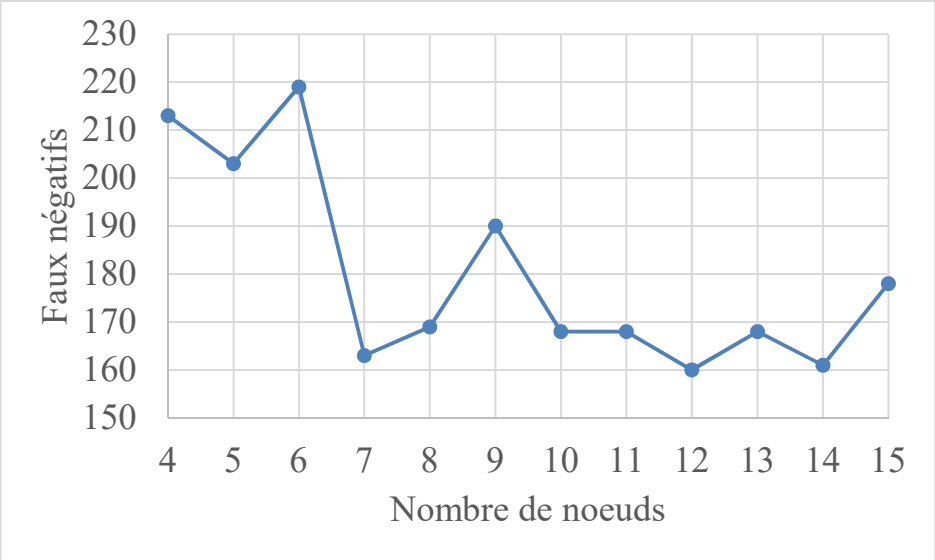

Figure 4 Nombre de faux négatifs en fonction du nombre de nœuds de la couche cachée

Le plus faible nombre de faux négatifs est obtenu avec 12 nœuds.

Le momentum est une valeur comprise entre 0 et 1. Elle permet d'intégrer, lors du calcul des nouveaux poids, la valeur de l'itération précédente. Voici l'impact du momentum sur les faux négatifs :

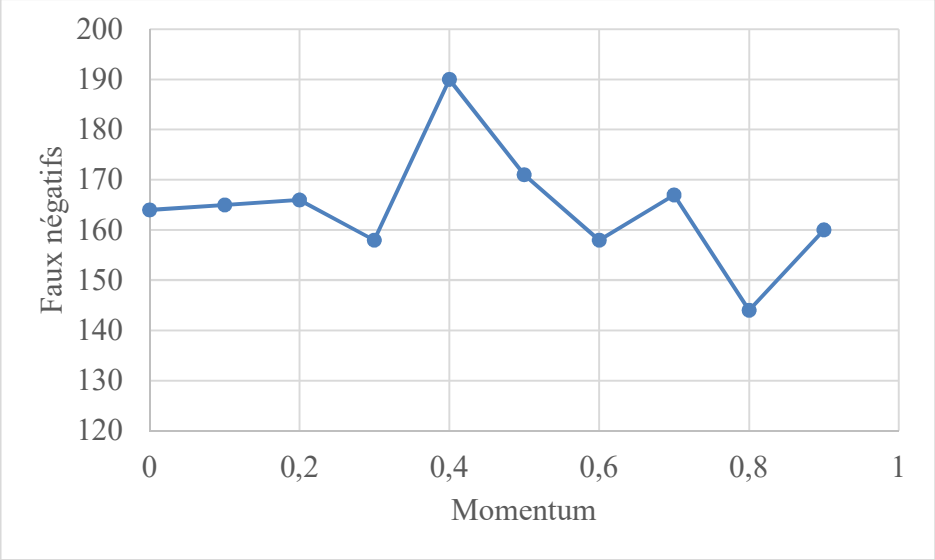

Figure 5 Nombre de faux négatifs en fonction du momentum

Pour un momentum de 0,8 le nombre de faux négatifs est de 144.

Le taux d'apprentissage permet de moduler l'avancement de chaque itération. Il a été modifié sur une plage de 0.005 à 0.5 pour savoir dans quel cas l'algorithme obtenait le moins de faux négatifs.

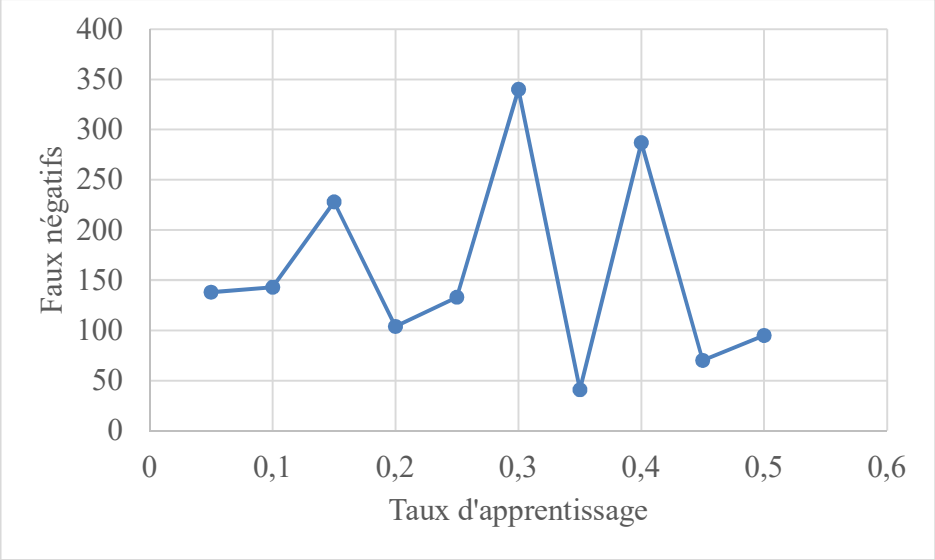

Figure 6 Nombre de faux négatif en fonction du taux d'apprentissage

Le taux d'apprentissage est donc fixé à 0,35 pour obtenir le minimum de faux négatifs, soit 41.

Le tableau 5 présente les meilleurs paramètres trouvés pour un faible nombre de faux négatifs.

Tableau 5 Paramètres du réseau de neurones

Paramètre	**Valeur**
Nombre de nœuds de la couche cachés	12
Taux d'apprentissage (en anglais : *learning rate*)	0.35
Nombre maximum d'itérations d'apprentissage	60
Poids initial	0.1
Momentum	0.8

3.1.2 Résultats

Les résultats présentés sont ceux obtenus avec les paramètres présentés, la normalisation des données ainsi que tous les attributs présentés. L'échantillon test est constitué de 2809 pièces.

Tableau 6 Matrice de confusion du réseau de neurones

		Réel	
		Non-conforme	Conforme
Prédit	Non-conforme	Vrais positifs 1851	Faux positifs 432
	Conforme	**Faux négatifs 41**	Vrais négatifs 485

Tableau 7 Résultats du réseau de neurones

Métrique	Valeur
Précision	**81,1%**
Taux de faux négatifs	**2,2%**

La précision du modèle n'est pas très bonne. En effet, les paramètres sont optimisés pour obtenir le plus faible nombre de faux négatifs, peu importe le nombre de faux positifs. Donc certes le nombre de faux négatifs est faible, mais le modèle obtient 432 faux positifs.

3.2 Forêt aléatoire

3.2.1 Paramètres

Les seuls paramètres sur lesquels il est possible d'agir sur AMLS sont les suivants :

- Le nombre maximal d'AD ;

- La profondeur maximale d'un AD ;
- Le nombre d'échantillons aléatoires d'attributs utilisés pour créer un nœud ;
- Le nombre minimum de nœuds avant une feuille.

La forêt aléatoire utilise le Bagging (*bootstrap aggregating*) pour le rééchantillonnage.

Le nombre d'arbres permet de définir combien d'AD sont créés pour la forêt.

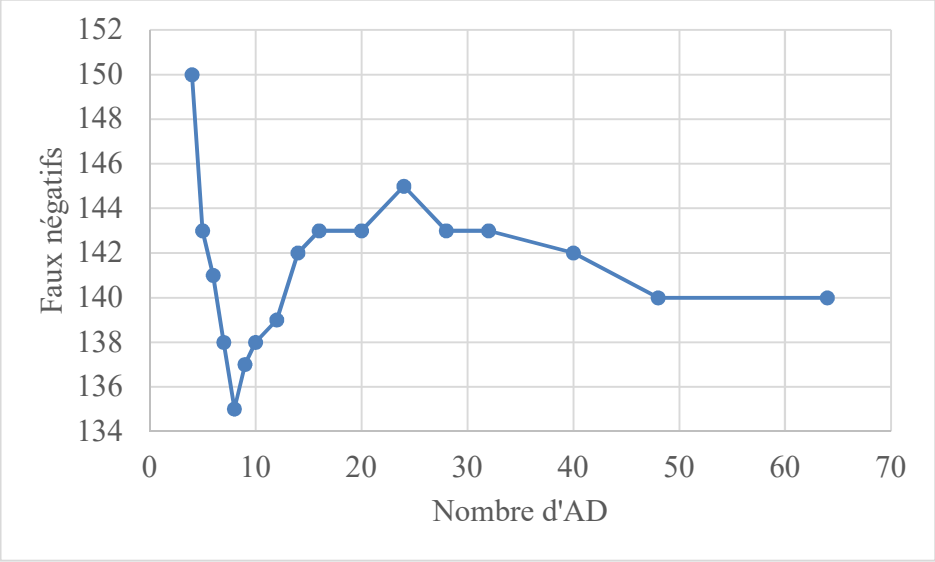

Figure 7 Nombre de faux négatifs en fonction du nombre d'AD

Le plus faible nombre de faux négatifs a été obtenu avec 8 AD.

Le second paramètre étudié est la profondeur maximale d'un AD. Elle représente la longueur maximale entre le nœud racine et une feuille. La figure 8 présente son influence sur les faux négatifs.

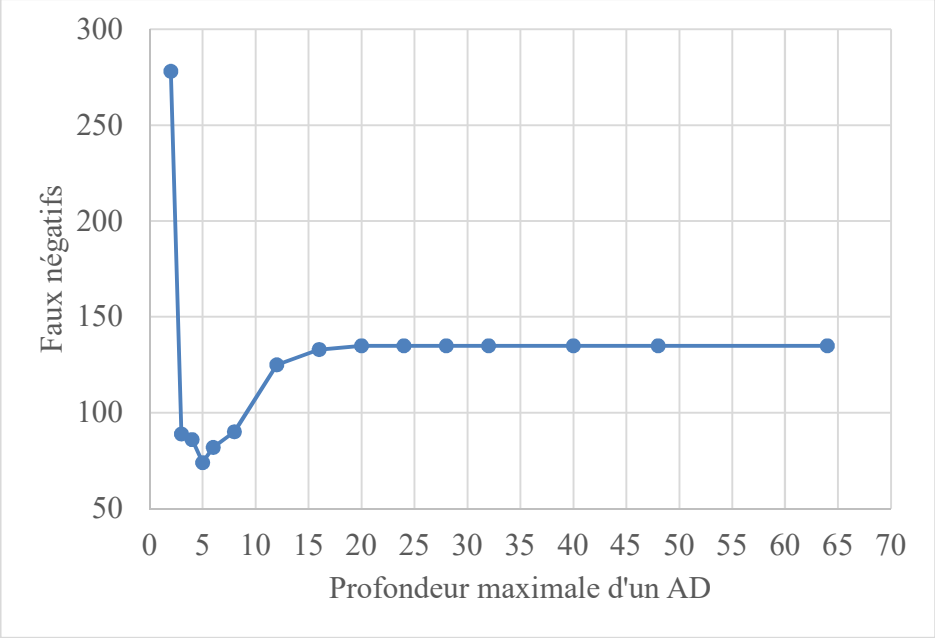

Figure 8 Nombre de faux négatifs en fonction de la profondeur maximale d'un AD

Le plus faible nombre de faux négatifs est obtenu pour une profondeur maximale de 5.

La figure 9 présente l'impact du nombre d'échantillons d'attributs aléatoires utilisé pour créer un nœud sur les faux négatifs. À chaque nœud, un certain nombre d'attributs sont sélectionnés au hasard parmi ces échantillons pour obtenir la meilleure division possible. Cela permet d'obtenir des AD différents les uns des autres.

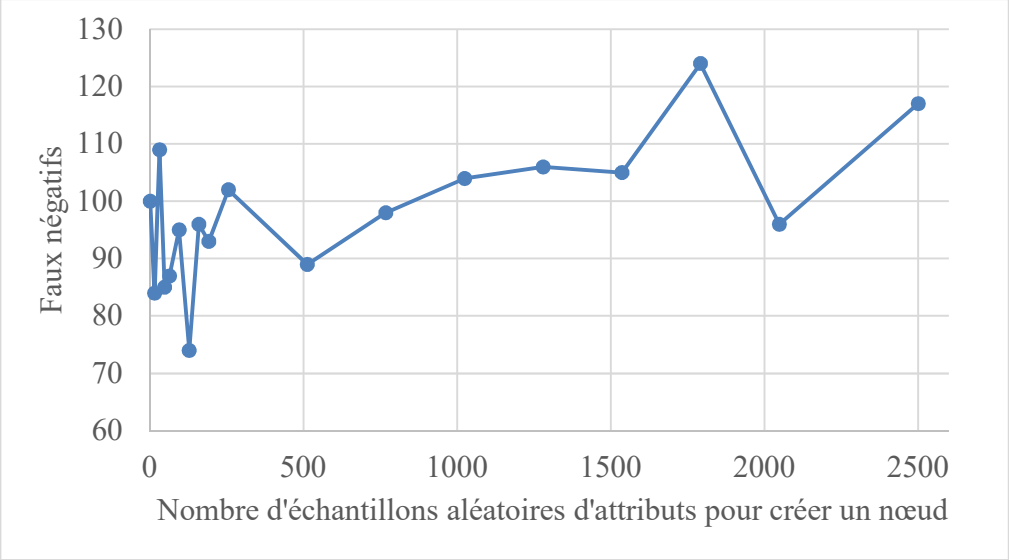

Figure 9 Faux négatifs en fonction du nombre d'échantillons aléatoires utilisés pour créer un nœud

Le nombre de faux négatifs le plus faible est obtenu pour un nombre d'échantillons de 128.

Le tableau 8 présente les meilleurs paramètres trouvés pour un faible nombre de faux négatifs.

Tableau 8 Paramètres de la forêt aléatoire

Paramètre	**Valeur**
Nombre d'arbres	8
La profondeur maximale d'un arbre	5
Nombre maximum de divisions aléatoires par nœud	128
Nombre minimum de tests avant une feuille	1

3.2.2 Résultats

Les résultats présentés sont ceux obtenus avec les paramètres présentés, la normalisation des données ainsi que tous les attributs présentés. L'échantillon test identique à celui utilisé pour le réseau de neurones.

Tableau 9 Matrice de confusion de la forêt aléatoire

		Réel	
		Non-conforme	Conforme
Prédit	Non-conforme	Vrais positifs 1818	Faux positifs 275
	Conforme	**Faux négatifs 74**	Vrais négatifs 642

Tableau 10 Résultats de la forêt aléatoire

Métrique	Valeur
Précision	**86,9 %**
Taux de faux négatifs	**3,9 %**

En optimisant le modèle pour obtenir le moins de faux négatifs possible, la précision est tout de même restée assez haute. Cela montre que le nombre de faux positifs n'est pas

3.3 Machine à vecteur de support

3.3.1 Paramètres

Les seuls paramètres sur lesquels il est possible d'agir sur AMLS sont les suivants :

- Le nombre d'itérations
- Lambda

Il est impossible pour l'utilisateur de gérer les algorithmes ni même de savoir lesquels sont utilisés. De plus, ces paramètres ne sont pas courants pour paramétrer une MVS. Le plus important est le noyau utilisé et gamma. La documentation ne permet pas d'avoir des renseignements suffisant pour comprendre l'implémentation de la MVS selon AMLS.

Cependant, les paramètres ont été étudiés comme précédemment. La figure 10 présente l'influence du nombre d'itérations sur les faux négatifs.

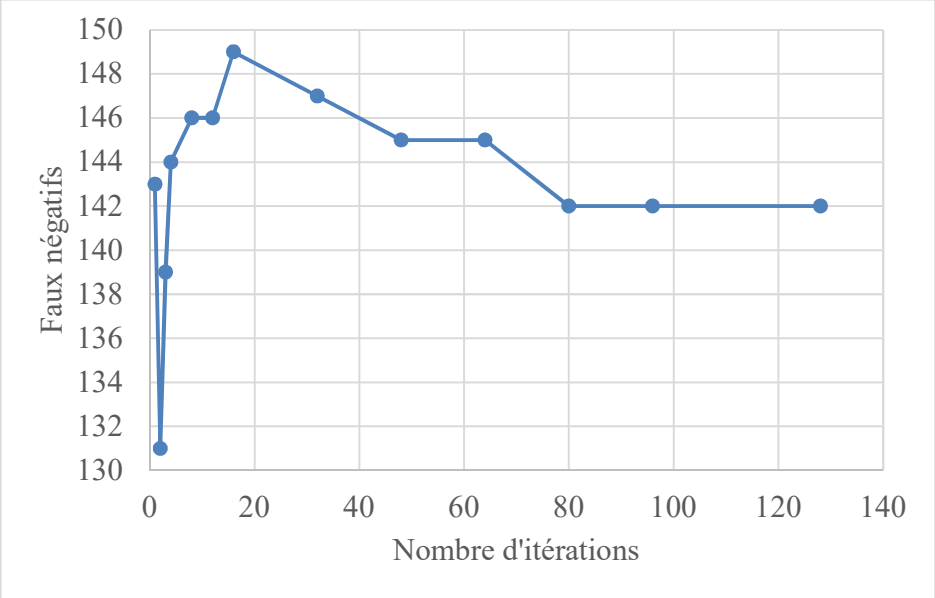

Figure 10 Faux négatifs en fonction du nombre d'itérations

Seulement 131 faux négatifs sont obtenus pour 2 itérations.

La figure 11 permet d'observer l'impact de lambda sur le nombre de faux négatifs. Lambda est le coefficient de régularisation L1

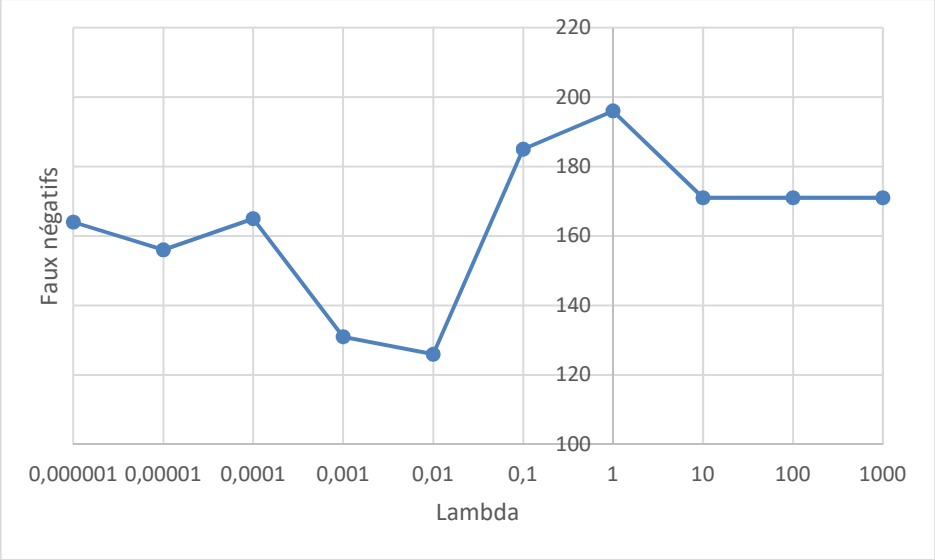

Figure 11 Faux négatifs en fonction de lambda

Un lambda de 0,01 permet d'obtenir le nombre le plus faible de faux négatifs.

Tableau 11 Paramètres pour la MVS

Paramètre	**Valeur**
Nombre d'itérations	2
Lambda	0,01

3.3.2 Résultats

Les résultats présentés sont ceux obtenus avec les paramètres présentés, la normalisation des données ainsi que tous les attributs présentés. L'échantillon test identique à celui utilisé par les deux modèles précédents.

Tableau 12 Matrice de confusion de la MVS

		Réel	
		Non-conforme	Conforme
Prédit	Non-conforme	Vrais positifs 1766	Faux positifs 280
	Conforme	**Faux négatifs 126**	Vrais négatifs 637

Tableau 13 Résultats de la MVS

Métrique	Valeur
Précision	**86,3 %**
Taux de faux négatifs	**6,7 %**

La précision du modèle est acceptable cependant le taux de faux négatif est élevé. Sur 100 pièces défectueuses, le modèle en prédit plus de 6 comme conformes.

3.4 Régression logistique

3.4.1 Paramètres

Les seuls paramètres sur lesquels il est possible d'agir sur AMLS sont les suivants :

- La tolérance d'optimisation ;
- Le coefficient de régularisation L1 ;
- Le coefficient de régularisation L2 ;
- La taille de la mémoire pour la méthode L-BFGS.

L'algorithme utilisé est L-BFGS (en anglais : *Limited-memory Broyden Fletcher Goldfarb Shanno*).

La tolérance d'optimisation permet de définir un seuil pour arrêter la convergence du modèle. La figure 12 présente les faux négatifs et la précision en fonction de la tolérance d'optimisation.

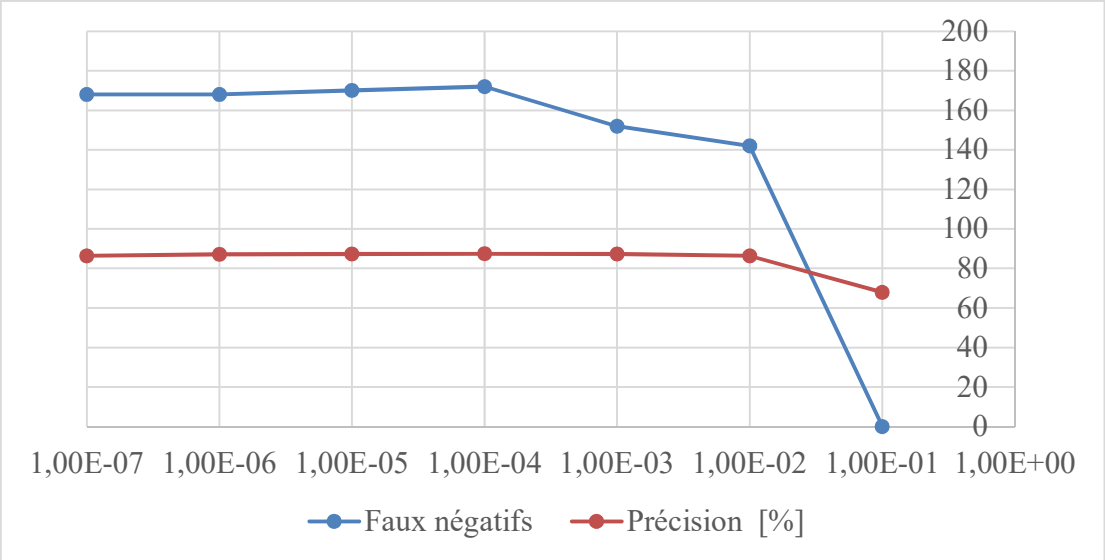

Figure 12 Faux négatifs et précision en fonction de la tolérance d'optimisation

Augmenter le seuil de tolérance d'optimisation fait réduire le nombre de faux négatifs jusqu'à 0. Cependant ce n'est pas bénéfique, car la précision du modèle diminue significativement. Il s'agit d'utiliser le seuil précèdent pour garder une bonne précision et optimiser le nombre d'itérations.

Le coefficient de régularisation L1 permet de pénaliser le modèle pour éviter le surapprentissage. Le coefficient L1 agit sur les poids des attributs de l'équation (2).

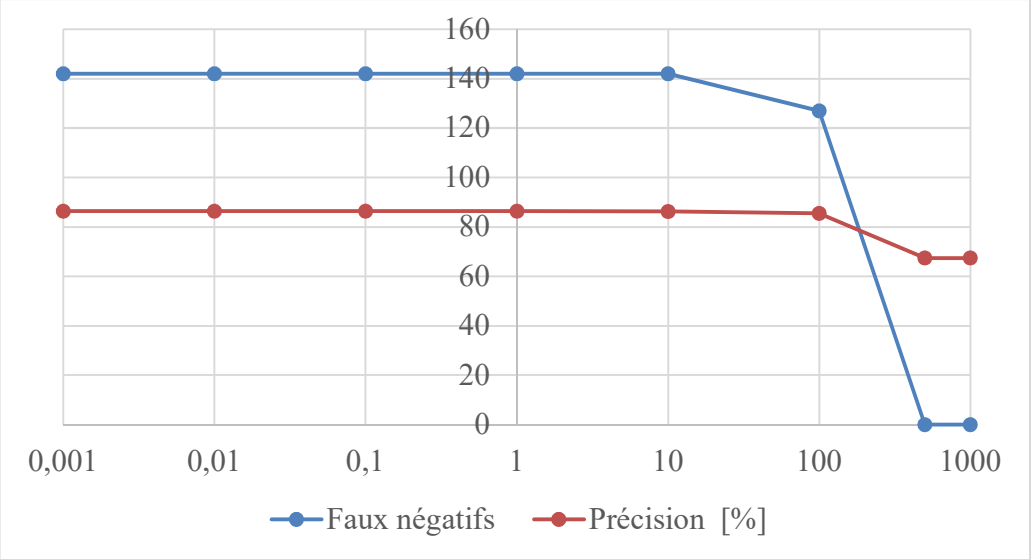

Figure 13 Faux négatifs et précision en fonction du coefficient de régularisation L1

Pour un coefficient supérieur à 100, la précision chute. Un coefficient de 100 permet de minimiser les faux négatifs en gardant une bonne précision.

Le coefficient de régularisation L2, comme le coefficient L1, permet de pénaliser le modèle pour éviter le surapprentissage. Il agit également sur les poids des attributs de l'équation (2).

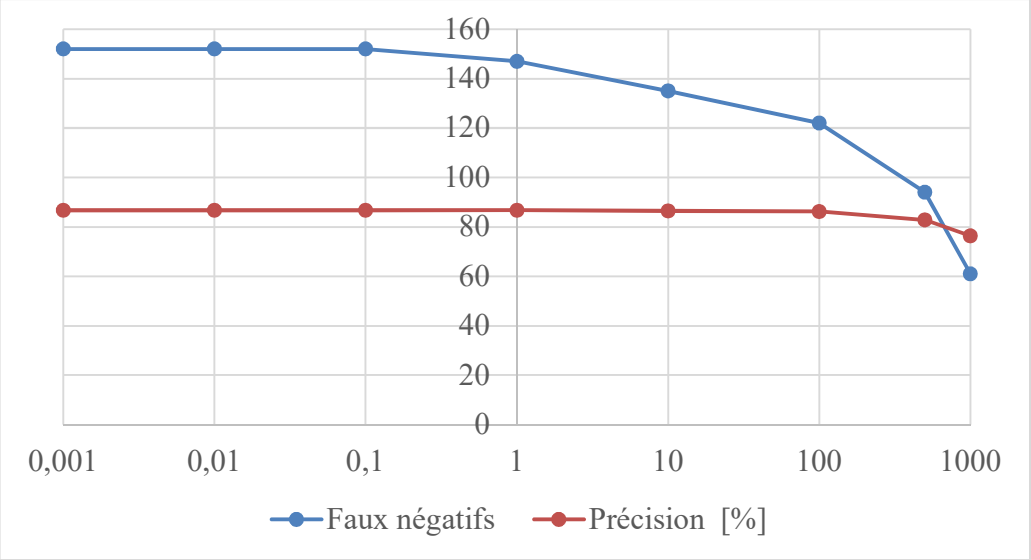

Figure 14 Faux négatifs et précision en fonction du coefficient de régularisation L2

Le meilleur résultat obtenu en tenant compte à la fois des faux négatifs et de la précision du modèle est pour un coefficient de 500.

Les coefficients de régularisation permettent d'agir sur les poids pour éviter le surapprentissage. Le tableau montre les différences entre les poids avec et sans la régularisation.

Tableau 14 Poids des attributs avec et sans la régularisation

Avec régularisation		**Sans régularisation**	
Attributs	**Poids**	**Attributs**	**Poids**
Etat dernière du même nom	0.414426	Etat dernière du même nom	2.20884
Complet	0.362211	Bias	-1.85468
Nb Tolérances	0.291644	Nb Tolérances	1.69366
Distance	0.237379	Circle	1.52307
Free State	-0.23302	Distance	1.38148
Circle	0.165497	Pos. Tol.	1.30012
Pos. Tol.	0.163665	Free State	-0.522687
Module-44	-0.118416	Dernières 10 pièces	0.513781

Dernières 10 pièces	0.0720544	Tendance croissante dev moy 2 dernières mêmes pièces	-0.445992
Profile of Line	0.0593703	Profile of Line	-0.347797
Dernières 5 pièces	0.0385888	Complet	-0.294093
Dernières 2 pièces	0.0248662	Dernières 2 pièces	0.176882
Bias	-0.0199423	Flatness	0.166608
Module-51	0.00825505	Module-44	0.165122
Tendance croissante dev moy 2 dernières mêmes pièces	-0.00128933	Dernières 5 pièces	0.145153
Flatness	0	Module-51	-0.0975435
Tendance croissante dev moy 2 dernières pièces	0	Tendance croissante dev moy 2 dernières pièces	-0.080486

La norme L1 a la propriété de sélectionner des attributs en leur accordant un poids nul. La régularisation L2 force les poids à être très petits pour simplifier l'équation (2).

La taille de la mémoire n'a pas d'incidence sur les résultats du modèle tant qu'elle est de l'ordre du Mb. La valeur utilisée sera 10Mb.

Le tableau 12 présente les meilleurs paramètres trouvés pour un faible nombre de faux négatifs et une précision correcte.

Tableau 15 Paramètres de la régression logistique

Paramètre	**Valeur**
Tolérance d'optimisation	0.01
Coefficient de régularisation L1	100
Coefficient de régularisation L2	500
Taille de la mémoire	10

3.4.2 Résultats

Les résultats présentés sont ceux obtenus avec les paramètres présentés, la normalisation des données ainsi que tous les attributs présentés. L'échantillon test identique à celui utilisé par les modèles précédents.

Tableau 16 Matrice de confusion pour la régression logistique

		Réel	
		Non-conforme	Conforme
Prédit	Non-conforme	Vrais positifs 1831	Faux positifs 435
	Conforme	**Faux négatifs 61**	Vrais négatifs 482

Tableau 17 Résultats de la régression logistique

Métrique	Valeur
Précision	**80,8 %**
Taux de faux négatifs	**3,2 %**

La précision du modèle n'est pas parfaite, mais il a été optimisé pour réduire le nombre de faux négatifs. Donc sur 100 pièces défectueuses, le modèle ne classe que 3 d'entre elles comme conformes.

3.5 Comparaison

Tous les modèles ont été entrainés est testé sur les mêmes échantillons de données. Il est donc possible de comparer la capacité des modèles à déterminer si une pièce est défectueuse ou non.

Le tableau 18 présente les résultats obtenus.

Tableau 18 Résultats des modèles

	Nombre de faux négatifs	**Précision [%]**	**Taux de faux négatifs [%]**
Réseau de neurones	41	81,1	2,2
Forêt aléatoire	74	86,9	3,9
MVS	126	86,3	6,7
Régression logistique	61	80,8	3,2

Tous les modèles ont été paramétrés pour essayer de réduire au maximum le nombre de faux négatifs et non d'augmenter la précision. Cela dépend du coût de la production de la pièce en question. Si une pièce coûte très cher à produire, il sera préférable d'avoir plus de faux positifs que de faux négatifs en améliorant la précision. Ici l'hypothèse a été faite que les pièces coûtent cher à produire. Dans cette optique, le meilleur modèle est celui du réseau de neurones avec un taux de faux négatif de seulement 2,2%. La forêt aléatoire est un bon compromis entre précision et taux de faux négatifs. Lorsque les modèles ont été entrainés pour avoir la plus haute précision, la forêt aléatoire a obtenu le meilleur résultat avec 90%.

CONCLUSION

L'objectif de ce projet de recherche appliqué était d'expérimenter avec l'apprentissage machine pour faire de la prédiction de défaut en utilisant des données d'assurance qualité. Pour ce faire, une première étape d'analyse de la littérature sur le domaine a permis de s'approprier les notions. Ensuite les données d'assurance qualité ont été analysées pour déterminer des attributs géométriques et temporels pour l'apprentissage machine. Quatre modèles prédictifs ont été entrainés en utilisant des algorithmes différents : 1) le réseau de neurones; 2) la forêt aléatoire; 3) la machine à vecteurs de support; et 4) la régression logistique. L'apprentissage des modèles a été optimisé pour réduire le nombre de pièces défectueuses prédites comme conformes. L'outil utilisé, AMLS, permet de mettre en place des modèles facilement et simplement. Il peut cependant devenir limité si l'on cherche à maitriser tous les paramètres d'un algorithme.

Le réseau de neurones a montré les meilleurs résultats et a obtenu un faible nombre de faux négatifs. Ce modèle obtient une précision de 81,1% et un taux de faux négatif de 2,2%. En optimisant les modèles pour avoir une meilleure précision, la forêt aléatoire obtient le meilleur résultat avec une précision de 90%. Ces résultats démontrent qu'il est possible de prédire l'état d'une pièce avant sa production en utilisant l'apprentissage machine et des données d'assurance qualité provenant d'une machine à mesurer tridimensionnelle.

ANNEXE I

Exemple de fichier CSV

Feat. Type	Feat. Name	Value	Actual	Nominal	Dev.	Tol-	Tol+	Out of Tol.	Comment
Text/Value	VAR_PROGRAM	VAL							N:\Metrolog XG\Programmes\Ford 3,5L GTDI\Ford 3.5L Free State (Module 44).gm2'
Text/Value	VAR_EVENEMENT	VAL							'OTS-2'
Text/Value	VAR_COMMENTAIRES	VAL							'P1'
Circle	CIR_FLANGE_1	X	-16.066	-15.944	-0.122				Criterion : Least Square (From 8 Pts On PL_FLANGE_1)
Circle	CIR_FLANGE_1	Z	-48.709	-48.650	-0.059				
Circle	CIR_FLANGE_2	X	-43.673	-43.444	-0.229				Criterion : Least Square (From 8 Pts On PL_FLANGE_2)
Circle	CIR_FLANGE_2	Z	52.804	52.850	-0.046				
Circle	CIR_FLANGE_3	X	-16.273	-15.944	-0.329				Criterion : Least Square (From 8 Pts On PL_FLANGE_3)
Circle	CIR_FLANGE_3	Z	158.711	158.849	-0.138				
Circle	CIR_FLANGE_4	X	-43.438	-43.444	0.006				Criterion : Least Square (From 8 Pts On PL_FLANGE_4)
Circle	CIR_FLANGE_4	Z	264.765	264.849	-0.084				
Circle	CIR_FLANGE_5	X	-221.821	-221.700	-0.121				Criterion : Least Square (From 8 Pts On PL_FLANGE_5)
Circle	CIR_FLANGE_5	Z	296.820	296.849	-0.029				
Circle	CIR_FLANGE_6	X	-195.030	-195.208	0.178				Criterion : Least Square (From 8 Pts On PL_FLANGE_6)
Circle	CIR_FLANGE_6	Z	195.414	195.349	0.065				
Circle	CIR_FLANGE_7	X	-222.571	-222.700	0.129				Criterion : Least Square (From 8 Pts On PL_FLANGE_7)
Circle	CIR_FLANGE_7	Z	89.453	89.349	0.104				
Circle	CIR_FLANGE_8	X	-195.131	-195.208	0.077				Criterion : Least Square (From 8 Pts On PL_FLANGE_8)
Circle	CIR_FLANGE_8	Z	-16.556	-16.651	0.095				
Distance	BOSS_HEIGHT_1	D1	26.098	25.900	0.198	-0.200	0.100	0.098	CIR_TOP_1 - CIR_FLANGE_1 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_2	D1	25.916	25.900	0.016	-0.200	0.100		CIR_TOP_2 - CIR_FLANGE_2 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_3	D1	26.004	25.900	0.104	-0.200	0.100	0.004	CIR_TOP_3 - CIR_FLANGE_3 / CS_BF_MNTG_POINTS

Distance	BOSS_HEIGHT_4	D1	26.191	25.900	0.291	-0.200	0.100	0.191	CIR_TOP_4 - CIR_FLANGE_4 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_5	D1	26.031	25.900	0.131	-0.200	0.100	0.031	CIR_TOP_5 - CIR_FLANGE_5 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_6	D1	25.989	25.900	0.089	-0.200	0.100		CIR_TOP_6 - CIR_FLANGE_6 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_7	D1	26.036	25.900	0.136	-0.200	0.100	0.036	CIR_TOP_7 - CIR_FLANGE_7 / CS_BF_MNTG_POINTS
Distance	BOSS_HEIGHT_8	D1	26.071	25.900	0.171	-0.200	0.100	0.071	CIR_TOP_8 - CIR_FLANGE_8 / CS_BF_MNTG_POINTS
Pos. Tol.	POS_1	PTOL	0.271	0.000	0.271	0.000	1.500		CIR_FLANGE_1/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_2	PTOL	0.467	0.000	0.467	0.000	1.500		CIR_FLANGE_2/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_3	PTOL	0.714	0.000	0.714	0.000	1.500		CIR_FLANGE_3/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_4	PTOL	0.168	0.000	0.168	0.000	1.500		CIR_FLANGE_4/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_5	PTOL	0.248	0.000	0.248	0.000	1.500		CIR_FLANGE_5/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_6	PTOL	0.378	0.000	0.378	0.000	1.500		CIR_FLANGE_6/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_7	PTOL	0.332	0.000	0.332	0.000	1.500		CIR_FLANGE_7/NOMINAL - CS_ABC - C-Zone - XZ - Nominal
Pos. Tol.	POS_8	PTOL	0.244	0.000	0.244	0.000	1.500		CIR_FLANGE_8/NOMINAL - CS_ABC - C-Zone - XZ - Nominal

ANNEXE II

Les différentes pièces et leur quantité

Pièce	Nombre
GM HF FWD Gen II C1xx Complete (Module 64).gm2	569
Denso 640F Complet (Module 51).gm2	444
GM HF FWD Gen II High Hood 31xx Complete (Module 64).gm2	420
Corvette LT1 Complete (Module 51).gm2	389
GM HF RWD Gen II Complete (Module 64).gm2	362
Corvette LT1 Free State Boss Height (Module 44).gm2	327
V10 6.8L Triton Complete Module 51.gm2	313
D33 RWD Complete (Module 44).gm2	262
Ford Transmission Pan 10R Complete (Module_51).gm2	242
Ford 3.5L GTDI Complete (Module 44).gm2	238
GM HF FWD Gen II E2xx Complete (Module 64).gm2	238
Mercury Marine TigerShark Complete (Module 51).gm2	226
V10 6.8L Triton Stage-1 Module 51.gm2	216
Denso 507F Complet (Module 51).gm2	196
D33 RWD Crush Rib (Module 44).gm2	187
GM HF RWD Gen II Free State Flatness Flange (Module 64).gm2	185
D-35 Hauteur Plastic Crush Rib Moule 2 (Module 44).gm2	171
D-37 with TMAP Complete (Module 44).gm2	163
GDTI Gap Entretoises Module_50.gm2	163
D-37 Hauteur Plastic Crush Rib Module 44.gm2	162
D-35 Moule 2 Complete with TMAP (Module 44).gm2	157
GDTI Complete Module_18.gm2	154
Mercury Marine Complete (Module 51).gm2	152
Ford 2L GTDI Complete (Module 51).gm2	142
Ford Nano MY2018 Complete (Module_51).gm2	138
Ford Nano 2.7L Complete (Module_51).gm2	134
Ford Transmission Pan 10R80 Free State Cavité 1 (Module_44).gm2	133
Ford Transmission Pan 10R80 Complete Cavité 1 (Module_51).gm2	129
Ford Transmission Pan 10R60 Free State (Module_44).gm2	115
Ford 3.5L Free State (Module 44).gm2	109
Denso 507F Stage-1 Module 51.gm2	104
GMI 700 Moule 1 Complete Module 51.gm2	97

Ford Transmission Pan 10R60 Complete Cavité 1 (Module_51).gm2	96

Ford Transmission Pan 10R60 Complete Cavité 2 (Module_51).gm2	92
D-35 Flatness Free State Moule 2 (Module 44).gm2	90
D-35 Hauteur Plastic Crush Rib Moule 1 (Module 44).gm2	90
Ford Transmission Pan 10R80 Free State Cavité 2 (Module_44).gm2	89
Ford Transmission Pan 10R80 Complete Cavité 2 (Module_51).gm2	83
Ford 2L GTDI Complete Moule 2 (Module 51).gm2	82
D-35 Moule 1 Complete with TMAP (Module 44).gm2	82
Denso 640F Stage-1(Module 51).gm2	78
I4 Complete Moule 2 (Module 44).gm2	75
Ford Transmission Pan 10R60 Free State Cavité 1 (Module_44).gm2	75
Ford Nano 2.7L Free State (Module 51).gm2	73
Ford Nano MY2018 Free State (Module 51).gm2	71
D37 Mustang Hauteur Plastique Crush Rib (Module 44).gm2	67
D37 Mustang Complete (Module 44).gm2	66
Ford Transmission Pan 10R80 Complete Cavité 4 (Module_51).gm2	63
GMI 700 Moule 3 Complete (Module 51).gm2	61
GM Twin Turbo Complete (Module 51).gm2	60
GM Twin Turbo Free State Retaining Ring Groove (Module 67 & 44).gm2	56
GM Twin Turbo Boss Height Free State (Module 51).gm2	54
I4 Complete Moule 1 (Module 44).gm2	52
D-35 Flatness Free State Moule 1 (Module 44).gm2	43
GM Impala Validation Stage-1(Module 51).gm2	38
Ford Transmission Pan 10R80 Complete Cavité 3 (Module_51).gm2	36
Ford Transmission Pan 10R Free State (Module_44).gm2	33
D-37 Validation Stage-1 (Module 44).gm2	31
Ford 2L Maverick FWD Complete (Module 44).gm2	31
D37 Mustang Validation Stage-1 (Module 44).gm2	29
TBA 2V Complete Module 51.gm2	25
Ford Transmission Pan 10R80 Free State Cavité 3 (Module_44).gm2	25
Ford Transmission Pan 10R80 Free State Cavité 4 (Module_44).gm2	25
TBA 3V Platine (Module 51).gm2	24
TBA 3V Complete (Module 51).gm2	24
GM Impala Complete Module 18.gm2	23
Ford Transmission Pan 10R60 Free State Cavité 2 (Module_44).gm2	23
Ford Transmission Pan 10R80 MHT Free State (Module_44).gm2	22
GMI 700 Moule 3 Hauteur Crush Rib (Module 51).gm2	22
GM Impala Boss Height Module 18.gm2	21
Mercury Marine Great White V6 Plenum (Module 51).gm2	21
Corvette LS7 Complete (Module 18).gm2	20

GM HF RWD Continental Complete Module 51.gm2	18
Road Runner Complete Module 51.gm2	17
Mercury Marine Great White V8 Plenum (Module 51).gm2	17
Ford Transmission Pan 10R80 Transit Free State (Module_44).gm2	17
Ford 2.3L Maverick RWD Complete (Module 51).gm2	16
P-415 Complete (Module_16).gm2	15
Mercury Marine Great White V8 Runner Pack LH (Module 51).gm2	14
Mercury Marine Great White V6 Runner Pack RH (Module 51).gm2	13
Dual Plenum Complete (Module 18).gm2	12
Mercury Marine Great White V8 Runner Pack RH (Module 51).gm2	12
Ford Nano MY2018 Program EGR (Module_51) .gm2	12
Mercury Marine Great White V6 Runner Pack LH (Module 51).gm2	12
D33 MY2020 Crush Rib (Module 44).gm2	12
GMX Complete Module 01.gm2	11
GMI 700 Moule 2 Complete Module 51.gm2	9
GMI 700 Moule 1 Hauteur Crush Rib (Module 51).gm2	8
Dual Plenum	6
Ford Transmission Pan 10R80 Free State Cavité 3 (Module_44) .gm2	6
Ford Transmission Pan 10R80 Free State Cavité 4 (Module_44) .gm2	6
GM HF FWD Gen II E2xx Cover ASM (Module 51).gm2	4
Mercury Marine Great White V8 Runner Pack LH--TAB(Module 51).gm2	4
GMI 700 Moule 4 Hauteur Crush Rib (Module 51).gm2	3
Mercury Marine Great White V8 Runner Pack RH--NO TAB(Module 51).gm2	3
GM Impala	2
Viper Cover Air Cleaner Complet (Module 51).gm2	2
Original Program with LSRC	2
Toyota 930N Complet (Module 44).gm2	1
Ford Transmission Pan 10R80 Complete Cavité 3 (Module_51) .gm2	1
GMI 700 Moule 4 Complete Module 51.gm2	1

ANNEXE III

Pourcentage de défaut par type de pièce

Nom pièce	Nombre de pièces	Nombre de pièces défectueuses	Pourcentage
Mercury Marine TigerShark Complete (Module 51).gm2	226	226	100%
Denso 507F Complet (Module 51).gm2	196	196	100%
I4 Complete Moule 1 (Module 44).gm2	52	52	100%
D37 Mustang Complete (Module 44).gm2	66	66	100%
D-35 Moule 1 Complete with TMAP (Module 44).gm2	82	82	100%
GMI 700 Moule 2 Complete Module 51.gm2	9	9	100%
GM Twin Turbo Complete (Module 51).gm2	60	60	100%
TBA 3V Complete (Module 51).gm2	24	24	100%
GM HF RWD Continental Complete Module 51.gm2	18	18	100%
GMI 700 Moule 3 Complete (Module 51).gm2	61	61	100%
Corvette LS7 Complete (Module 18).gm2	20	20	100%
I4 Complete Moule 2 (Module 44).gm2	75	75	100%
GM Impala Validation Stage-1(Module 51).gm2	38	38	100%
D-37 Validation Stage-1 (Module 44).gm2	31	31	100%
Road Runner Complete Module 51.gm2	17	17	100%
Dual Plenum	6	6	100%
GM HF FWD Gen II E2xx Cover ASM (Module 51).gm2	4	4	100%
GMI 700 Moule 1 Hauteur Crush Rib (Module 51).gm2	8	8	100%
Mercury Marine Great White V8 Plenum (Module 51).gm2	17	17	100%
Ford Nano MY2018 Program EGR (Module_51) .gm2	12	12	100%
Mercury Marine Great White V6 Plenum (Module 51).gm2	21	21	100%
GMI 700 Moule 4 Hauteur Crush Rib (Module 51).gm2	3	3	100%
GMI 700 Moule 3 Hauteur Crush Rib (Module 51).gm2	22	22	100%
Ford Transmission Pan 10R80 Free State Cavité 3 (Module_44) .gm2	6	6	100%
Ford Transmission Pan 10R80 Free State Cavité 4 (Module_44) .gm2	6	6	100%
Ford Transmission Pan 10R80 Complete Cavité 3 (Module_51) .gm2	1	1	100%
GMI 700 Moule 4 Complete Module 51.gm2	1	1	100%
Ford Transmission Pan 10R60 Complete Cavité 1 (Module_51).gm2	96	96	100%
Ford Transmission Pan 10R60 Complete Cavité 2 (Module_51).gm2	92	92	100%
D-37 with TMAP Complete (Module 44).gm2	163	162	99%
D-35 Moule 2 Complete with TMAP (Module 44).gm2	157	156	99%

Mercury Marine Complete (Module 51).gm2	152	151	99%
GDTI Complete Module_18.gm2	154	152	99%
Ford 2L GTDI Complete (Module 51).gm2	142	140	99%
Ford Nano MY2018 Complete (Module_51).gm2	138	130	94%
P-415 Complete (Module_16).gm2	15	14	93%
Ford Transmission Pan 10R Complete (Module_51).gm2	242	221	91%
GMI 700 Moule 1 Complete Module 51.gm2	97	88	91%
Ford Transmission Pan 10R80 Transit Free State (Module_44).gm2	17	15	88%
GM HF RWD Gen II Complete (Module 64).gm2	362	317	88%
Ford 2L Maverick FWD Complete (Module 44).gm2	31	27	87%
D33 RWD Complete (Module 44).gm2	262	225	86%
Ford Transmission Pan 10R80 Complete Cavité 2 (Module_51).gm2	83	70	84%
Ford Transmission Pan 10R80 Complete Cavité 3 (Module_51).gm2	36	30	83%
Ford Nano 2.7L Complete (Module_51).gm2	134	111	83%
GM Twin Turbo Free State Retaining Ring Groove (Module 67 & 44).gm2	56	46	82%
Ford Transmission Pan 10R80 Complete Cavité 1 (Module_51).gm2	129	105	81%
GDTI Gap Entretoises Module_50.gm2	163	132	81%
Ford 3.5L GTDI Complete (Module 44).gm2	238	192	81%
GM HF FWD Gen II High Hood 31xx Complete (Module 64).gm2	420	333	79%
TBA 2V Complete Module 51.gm2	25	19	76%
Corvette LT1 Complete (Module 51).gm2	389	293	75%
Denso 640F Complet (Module 51).gm2	444	334	75%
GM HF FWD Gen II E2xx Complete (Module 64).gm2	238	179	75%
Dual Plenum Complete (Module 18).gm2	12	9	75%
Mercury Marine Great White V8 Runner Pack LH--TAB(Module 51).gm2	4	3	75%
V10 6.8L Triton Complete Module 51.gm2	313	234	75%
Ford 2L GTDI Complete Moule 2 (Module 51).gm2	82	61	74%
GMX Complete Module 01.gm2	11	8	73%
D-35 Hauteur Plastic Crush Rib Moule 1 (Module 44).gm2	90	63	70%
Ford Transmission Pan 10R Free State (Module_44).gm2	33	23	70%
Mercury Marine Great White V8 Runner Pack RH--NO TAB(Module 51).gm2	3	2	67%
GM HF FWD Gen II C1xx Complete (Module 64).gm2	569	376	66%
Ford Transmission Pan 10R80 MHT Free State (Module_44).gm2	22	14	64%
Ford Transmission Pan 10R60 Free State (Module_44).gm2	115	72	63%
Ford Transmission Pan 10R80 Free State Cavité 3 (Module_44).gm2	25	14	56%
Mercury Marine Great White V6 Runner Pack RH (Module 51).gm2	13	7	54%
Ford Transmission Pan 10R80 Complete Cavité 4 (Module_51).gm2	63	33	52%

D-35 Hauteur Plastic Crush Rib Moule 2 (Module 44).gm2	171	89	52%

GM Twin Turbo Boss Height Free State (Module 51).gm2	54	28	52%
Original Program with LSRC	2	1	50%
Mercury Marine Great White V6 Runner Pack LH (Module 51).gm2	12	5	42%
Ford Nano MY2018 Free State (Module 51).gm2	71	29	41%
Ford Transmission Pan 10R60 Free State Cavité 2 (Module_44).gm2	23	9	39%
Denso 640F Stage-1(Module 51).gm2	78	30	38%
Ford Nano 2.7L Free State (Module 51).gm2	73	27	37%
Ford Transmission Pan 10R80 Free State Cavité 4 (Module_44).gm2	25	9	36%
Mercury Marine Great White V8 Runner Pack LH (Module 51).gm2	14	5	36%
Ford 3.5L Free State (Module 44).gm2	109	38	35%
GM Impala Complete Module 18.gm2	23	8	35%
Mercury Marine Great White V8 Runner Pack RH (Module 51).gm2	12	4	33%
D37 Mustang Hauteur Plastique Crush Rib (Module 44).gm2	67	21	31%
Denso 507F Stage-1 Module 51.gm2	104	23	22%
D-37 Hauteur Plastic Crush Rib Module 44.gm2	162	34	21%
TBA 3V Platine (Module 51).gm2	24	5	21%
Ford Transmission Pan 10R80 Free State Cavité 2 (Module_44).gm2	89	12	13%
Ford 2.3L Maverick RWD Complete (Module 51).gm2	16	2	13%
Ford Transmission Pan 10R60 Free State Cavité 1 (Module_44).gm2	75	9	12%
GM Impala Boss Height Module 18.gm2	21	2	10%
D33 RWD Crush Rib (Module 44).gm2	187	16	9%
Ford Transmission Pan 10R80 Free State Cavité 1 (Module_44).gm2	133	7	5%
GM HF RWD Gen II Free State Flatness Flange (Module 64).gm2	185	2	1%
V10 6.8L Triton Stage-1 Module 51.gm2	216	1	0%
Corvette LT1 Free State Boss Height (Module 44).gm2	327	0	0%
D-35 Flatness Free State Moule 2 (Module 44).gm2	90	0	0%
D-35 Flatness Free State Moule 1 (Module 44).gm2	43	0	0%
GM Impala	2	0	0%
Viper Cover Air Cleaner Complet (Module 51).gm2	2	0	0%
D37 Mustang Validation Stage-1 (Module 44).gm2	29	0	0%
Toyota 930N Complet (Module 44).gm2	1	0	0%
D33 MY2020 Crush Rib (Module 44).gm2	12	0	0%

ANNEXE IV

Premières lignes du fichier Dataset

ID	Nom	Nb Tolérances	Distance	Pos. Tol.	Circle	Profile of Line	Flatness	Complet	Free State	Module 44	Module 51	Dernières 10 pièces	Dernières 5 pièces	Dernières 2 pièces	Etat dernière du même nom	Tendance dev moy 2 dernières avec mêmes noms	Tendance dev moy 2 dernières pièces	Défectueuse
1	denso 507f stage 1 module 51	103	102	0	12	0	1	0	0	0	1	0	0	0	0	0	0	1
2	corvette lt1 free state boss height module 44	12	0	0	0	0	2	0	1	1	0	0	0	0	0	0	0	0
3	corvette lt1 complet module 51	220	149	33	92	16	3	1	0	0	1	0	0	0	0	0	1	1
4	denso 507f complet module 51	154	106	31	78	0	3	1	0	0	1	0	0	0	0	0	1	1
5	v10 6.8l triton stage 1 module 51	47	12	24	48	10	1	0	0	0	1	0	0	0	0	0	1	0
6	gm impala boss height module 18	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	ford 2l gtdi complet moule 2 module 51	134	87	30	70	0	1	1	0	0	1	0	0	0	0	0	0	1
8	ford 2l gtdi complet moule 2 module 51	134	87	30	70	0	1	1	0	0	1	1	1	1	1	0	1	1
9	gm impala complet module 18	172	114	30	72	19	2	1	0	0	0	0	0	0	0	0	1	0
10	mercury marine tigershark complet module 51	162	84	46	122	4	1	1	0	0	1	0	0	0	0	0	0	1

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Han, J. H., & Chi, S. Y. (2016). Consideration of manufacturing data to apply machine learning methods for predictive manufacturing. *International Conference on Ubiquitous and Future Networks, ICUFN*, *2016–Augus*, 109–113. https://doi.org/10.1109/ICUFN.2016.7536995

[2] Çaydaş, U., & Ekici, S. (2012). Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. *Journal of Intelligent Manufacturing*, *23*(3), 639–650. https://doi.org/10.1007/s10845-010-0415-2

[3] Mohammadi, P., & Wang, Z. J. (2016). Machine Learning for Quality Prediction in Abrasion- Resistant Material Manufacturing Process. *Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 4–7. https://doi.org/10.1109/CCECE.2016.7726783

[4] Weiss, S. M., & Baseman, R. J. (n.d.). Improving Quality Control by Early Prediction of Manufacturing Outcomes Categories and Subject Descriptors, 1258–1266. https://doi.org/10.1145/2487575.2488192

[5] Bai, Y., Li, C., Sun, Z., & Chen, H. (2017). Deep neural network for manufacturing quality prediction. *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, 1–5. https://doi.org/10.1109/PHM.2017.8079165

[6] Chatterjee, A., Croley, D., Ramamurti, V., & Kui-Yu Chang. (1996). Application of machine learning to manufacturing: results from metal etch. *Nineteenth IEEE/CPMT International Electronics Manufacturing Technology Symposium*, 372–377. https://doi.org/10.1109/IEMT.1996.559762

[7] Grabot, B., Blanc, J. C., and Binda, C., (1996). A Decision Support System for Production Activity Control. *NineteenthDecision Support Sys., 16, pp.87*–101. https://www.sciencedirect.com/science/article/pii/0167923695000038

[8] Tseng, T. L., Huang, C. C., Jiang, F., & Ho, J. C. (2006). Applying a hybrid data mining approach to prediction problems: A Case of preferred supplier prediction. *International Journal of Production Research 44(14), 2935*–2954. https://www.tandfonline.com/doi/abs/10.1080/00207540600654525

[9] Wang, K. J., Chen, J. C., & Lin, Y. S. (2005). A hybrid knowledge discovery model using decision tree and neural network for selecting dispatching rules of semiconductor final testing factory. *Production Planning and Control, 16(6), 665–680.* https://www.tandfonline.com/doi/full/10.1080/09537280500213757

[10] Backus, P., Janakiram, M.,Mowzoon, S., Runger, G. C.,&Bhargava. A. (2006). Factory cycle time prediction with data-mining approach. *IEEE Transactions on Semiconductor Manufacturing, 12(2).* https://ieeexplore.ieee.org/document/1628987

[11] Lin, C. C., & Tseng, Y. H. (2005). A neural network application for reliability modelling and condition-based predictive maintenance. *International Journal of Advance Manufacturing Technology, 25, 174–179.* https://link.springer.com/article/10.1007/s00170-003-1835-3

BIBLIOGRAPHIE

[1] Machine Learning Studio Documentation (2019). Consulté sur https://docs.microsoft.com/en-us/azure/machine-learning/studio/