

MÉTHODOLOGIE D'ÉVALUATION DE LA FIABILITÉ ET DE RÉDUCTION DES COÛTS DE GARANTIE POUR UN PARC À AUTOS UTILISANT DES ALGORITHMES D'INTELLIGENCE ARTIFICIELLE

Gabriel CRISTEA¹, Cristian ANDREESCU², Daniel Mihail IOZSA³,
Dan Mihai CONSTANTINESCU⁴

Commentary texts in warranty claims contain a lot of useful information that is often overlooked or used too little. In order to bring this information to light, effective text analysis is needed, which invariably begins with transforming the text into a structured format that can be processed with artificial intelligence algorithms.

In the next step, the data is capitalized through graphical representations and analytical reports that help to better understand the issues behind this information.

Les textes des commentaires dans les demandes de garantie contiennent de nombreuses informations utiles qui sont souvent négligées ou utilisées trop peu. Afin de mettre en lumière ces informations, une analyse de texte efficace est nécessaire, qui commence invariablement par la transformation du texte en un format structuré pouvant être traité avec des algorithmes d'intelligence artificielle.

Dans l'étape suivante, les données sont capitalisées à travers des représentations graphiques et des rapports analytiques qui aident à mieux comprendre les problèmes derrière ces informations.

Keywords: Reliability, Artificial intelligence, Machine learning, Natural language processing, Data mining.

Mots-clés: Fiabilité, Intelligence artificielle, Apprentissage automatique, Traitement du langage naturel, Extraction de données.

1. Introduction

Les constructeurs automobiles du monde entier dépensent des dizaines de milliards d'euros par an en réclamations de garantie de produits. Rien qu'en 2019, plus de 46 milliards de dollars ont été dépensés à cette fin [1]. Pour cette raison,

¹ Ing., Université POLITEHNICA de Bucarest, Roumanie, e-mail: gabriel.cristea@upb.ro

² Prof., Dép. d'ingénierie automobile, Université POLITEHNICA de Bucarest, Roumanie, e-mail: andreescu.cristian@gmail.com

³ Conf., Dép. d'ingénierie automobile, Université POLITEHNICA de Bucarest, Roumanie, e-mail: daniel_iozsa@yahoo.com

⁴ Prof., Dép. de la résistance des matériaux, Université POLITEHNICA de Bucarest, Roumanie, e-mail: dan.constantinescu@upb.ro

ils investissent massivement dans les systèmes analytiques et le personnel, dans un effort pour réduire les coûts, améliorer la fiabilité des produits et augmenter la satisfaction des clients. Afin de poursuivre leur objectif, les constructeurs automobiles recueillent les commentaires des réclamations de garantie, mais sont confrontés à un problème difficile lorsqu'ils doivent décider quoi en faire.

Par exemple, identifier les problèmes de qualité, réduire les réparations excessives et détecter les réparations frauduleuses ne sont que quelques-unes des principales opportunités d'économiser de l'argent en analysant les commentaires dans les demandes de garantie.

Cet article vise à présenter une méthode d'analyse de la fiabilité des composants et de réduction des coûts pendant la période de garantie en utilisant une méthode de base d'apprentissage automatique, à savoir le traitement du langage naturel [2] effectué sur les textes dans les commentaires des réclamations de garantie.

Dans une étude sur les solutions d'intelligence artificielle dans l'industrie automobile, Gusikhin [3] discute de la prévision des pannes, des services après-vente et des demandes de garantie. Un travail représentatif dans ce domaine est celui de Buddhakulsomsiri [4]. Il présente un algorithme d'extraction de données qui extrait des modèles associatifs et séquentiels d'une grande base de données de garanties automobiles, capturant les relations entre l'apparition de demandes de garantie au fil du temps. En utilisant une représentation simple de la procédure algorithmique *si - alors*, l'algorithme filtre les motifs insignifiants en utilisant le critère de puissance des paramètres. Cependant, dans leur article, aucune information n'est disponible sur l'utilisation du véhicule et les connaissances découvertes sont de nature statistique en référence aux relations entre les défaillances courantes.

Un autre travail représentatif est celui de Rajpathak [5] dans lequel il présente un système d'extraction de texte basé sur l'ontologie. L'ontologie [6] peut être vue comme un modèle informationnel qui décrit explicitement des entités et des concepts qui existent dans un champ de discours avec leurs propriétés. Selon la distinction proposée par la communauté de l'ingénierie ontologique, l'ontologie diagnostique est une ontologie légère [7], dans laquelle chaque étiquette de nœud peut être traduite en un codage sans équivoque de sa signification. Ainsi il classe les réparations afin d'identifier les meilleures réparations.

2. Traitement du langage naturel - notions de base

Le traitement du langage naturel (TLN) est une discipline à la frontière entre l'informatique, la linguistique et l'intelligence artificielle.

L'une des motivations de cette discipline est la compréhension du langage naturel, c'est-à-dire sa signification dans le but déclaré de pouvoir interagir avec

les êtres humains. Les domaines d'application possible étant bien entendu orientés autour de la langue on peut citer les applications de:

- Traduction automatique
- Correction orthographique
- Communication personne-machine
- Synthèse et reconnaissance de la parole
- Analyse des opinions / sentiments

Pour y parvenir, on peut utiliser une première approche appelée méthode linguistique, dont le but est de décrire les informations qui seront extraites à l'aide de règles linguistiques (grammaires, formulations) ou on peut utiliser l'apprentissage automatique. Cette deuxième méthode est celle que nous utiliserons.

Le traitement du langage naturel (TLN) est un ensemble de processus, d'algorithmes et d'outils utilisés par des systèmes intelligents pour interpréter des données textuelles écrites en langage humain afin d'obtenir des informations qui peuvent être utilisées. Le TLN fait référence à l'interprétation de données non structurées. Il organise des données textuelles non structurées et utilise des méthodes sophistiquées pour résoudre de nombreux problèmes, tels que l'analyse d'opinion, la classification de documents et la synthèse de texte.

Le diagramme suivant (Fig. 1) représente schématiquement les étapes hiérarchiques de base, impliquées dans le traitement du langage naturel:

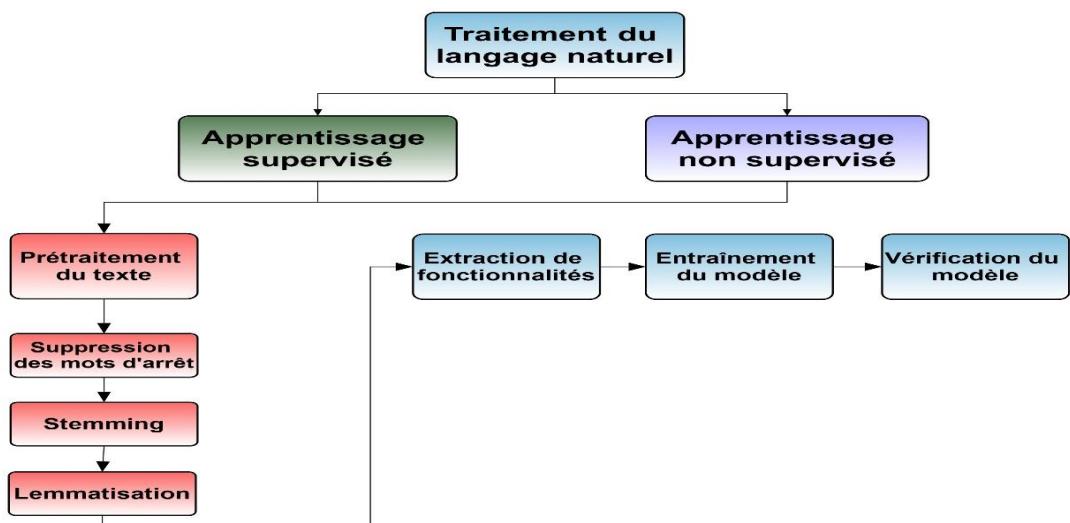


Fig. 1. Les étapes hiérarchiques de base impliquées dans le traitement du langage naturel.

- **Type d'apprentissage automatique:** Le TLN peut être réalisé soit à l'aide d'algorithmes d'apprentissage supervisé, soit d'algorithmes d'apprentissage non supervisé [2]. Les algorithmes d'apprentissage supervisé incluent Naïve Bayes,

Support Vector Machine et Random Forest. Les algorithmes d'apprentissage non supervisé comprennent les réseaux de neurones alimentés directement (Perceptron multicouche) et les réseaux de neurones récurrents (RNR). Une chose importante à noter ici est que les étapes de prétraitement et d'extraction de caractéristiques sont les mêmes pour les deux classes d'algorithmes. Ce qui est différent, c'est la façon dont le modèle sera entraîné. L'apprentissage supervisé nécessite une sortie et une entrée étiquetées, et l'apprentissage non supervisé pourrait prédire le résultat sans aucune sortie étiquetée.

• **Prétraitement du texte:** Cette étape est nécessaire, car le texte brut ne peut pas être utilisé dans les systèmes TLN. Cela conduira à des résultats incorrects ou pas très précis. Certaines des étapes courantes du prétraitement du texte consistent à supprimer les mots d'arrêt, à convertir le texte en minuscules et à supprimer les caractères spéciaux. D'avantage, certaines des techniques de base et génériques qui améliorent la précision impliquent la suppression des nombres (selon le contexte), la suppression de la ponctuation, la suppression des espaces blancs (parfois, cela ajoute du bruit au signal d'entrée) et la suppression des termes rares du document.

- ✓ **Suppression des mots d'arrêt:** Les mots d'arrêt sont des mots qui apparaissent plus souvent dans la phrase et rendent le texte plus difficile à analyser, et pour cela, ils doivent être exclus du texte. Ces mots d'arrêt n'ont aucune signification contextuelle et augmentent la taille des vecteurs de termes.
- ✓ **Stemming** est la conversion d'un mot d'un texte dans son format brut. Différentes formes d'un mot communiquent souvent essentiellement la même signification. Le but de cette méthode est de supprimer certains suffixes pour réduire le nombre de mots. Cela permet également d'éliminer la confusion pendant la période d'entraînement du modèle.
- ✓ **Lemmatisation** est légèrement différente de *stemming* (cette dernière supprime parfois les suffixes qui ajoutent du sens à un mot). *Lemmatisation* essaie de surmonter cette limitation de *stemming*. Elle essaie de trouver la forme de base du mot, appelé lemme, à partir d'un vocabulaire des mots qu'elle possède et d'une analyse morphologique des mots.

• **Extraction de fonctionnalités:** Pour que tout algorithme d'apprentissage automatique (ML) fonctionne sur des textes, ils doivent être convertis en une forme numérique. L'extraction de fonctionnalités utilise des techniques courantes pour convertir du texte en nombres sous forme de vecteurs.

• **Entraînement du modèle:** L'entraînement du modèle est le processus d'établissement ou de recherche d'une fonction mathématique pouvant être utilisé pour prédire le résultat en fonction de l'entrée donnée. Le processus de recherche d'une fonction implique plusieurs itérations et l'ajustement des paramètres.

• **Vérification du modèle:** Cette étape est le processus de vérification des modèles résultant du processus d'entraînement du modèle. En général, l'ensemble de données d'apprentissage est divisé en un rapport 80:20. 80% des données sont

utilisées pour entraîner le modèle et 20% des données sont utilisées pour valider l'exactitude du modèle. En cas de divergence, les étapes de création du modèle sont ajustées et les validations sont reprises.

3. Présentation des données - commentaires des demandes de garantie

Les prestataires de services de garantie collectent généralement deux types principaux de données textuelles: les commentaires des clients et les commentaires des techniciens. Les commentaires du client expliquent la raison de sa présence à la station-service, comme un dysfonctionnement du système électrique, une fuite de liquide sous la voiture, un bruit de moteur plus spécial, etc.

Les commentaires des clients sont enregistrés par le propre représentant du réparateur. Les commentaires du technicien décrivent le travail effectué ou la résolution et sont documentés, plus ou moins en détail, par les réparateurs. Les deux types de commentaires sont finalement introduits dans un système de demande de garantie, auquel le fabricant peut accéder pour une analyse plus approfondie.

Contrairement à d'autres sources traditionnelles de données textuelles, les commentaires techniques sous la garantie sont généralement documentés dans un langage informel et contiennent souvent du jargon technique, des acronymes, des mots mal orthographiés, des transcriptions et des incohérences dans l'utilisation des noms techniques classiques.

La base de données utilisée contient des informations sur les réparations effectuées sur les véhicules d'une entreprise de messagerie pendant deux ans. Chaque visite d'un véhicule à la station-service est enregistrée sous une forme structurée, étiquetée avec la date et le kilométrage, qui détaille les pièces changées et les opérations effectuées. Les pièces et les opérations sont marquées avec des codes d'identification normalisés.

Malheureusement, cependant, il n'y a pas de codes des raisons des opérations. Dans certains cas, ceux-ci peuvent être déduits de commentaires libres rédigés par le personnel de l'atelier, mais pas toujours. La qualité et le niveau de détail de ces commentaires varient considérablement. Il s'agit d'une limitation sérieuse, car elle introduit beaucoup de bruit dans les étiquettes de classification des données d'apprentissage. Dans le pire des cas, il peut arriver qu'un bon composant soit remplacé lors du processus de résolution d'un autre problème.

4. La relation entre constructeurs, concessionnaires et réparateurs dans l'industrie automobile

Dans le cas des concessionnaires, les constructeurs automobiles font un exercice d'équilibre - ils dépendent du réseau de concessionnaires pour une part importante de leurs ventes - mais ils ont du mal à gérer le côté service. Il est essentiel que les constructeurs s'assurent que les réparateurs fournissent une

prestation de service client équitable, car, d'après l'expérience [8], près de 70% des clients ne reviendront pas chez le concessionnaire après l'expiration de la garantie. Les constructeurs doivent travailler en permanence avec les concessionnaires franchisés pour assurer leur fidélité et maximiser la satisfaction des clients.

Dans le secteur des services, des priorités commerciales concurrentes peuvent entraîner des conflits entre constructeurs, concessionnaires et réparateurs. Par exemple :

- Les concessionnaires ou réparateurs souhaitent maximiser le coût par réparation
- Les distributeurs et réparateurs souhaitent augmenter le nombre de réparations dans leurs ateliers
- Pour les travaux sous garantie, les constructeurs veulent minimiser les deux chiffres.

La clé pour les constructeurs dans la gestion de ce conflit, tout en garantissant une grande satisfaction des clients et des concessionnaires, est d'avoir des rapports et des analyses de réclamations automatisées de haute qualité. L'utilisation de cette technologie résoudra des milliers de plaintes. Il identifiera les coûts de réparation élevés, les réparations à haute fréquence et les concessionnaires et réparateurs essayant de «tromper le système».

5. Méthodologie utilisée

La plupart des informations de la base de données ont été saisies manuellement. Cela conduit à diverses erreurs humaines, telles que des fautes de frappe et des valeurs manquantes. Mais un problème plus profond concerne les données et les kilométrages incorrects. Ceci est en partie dû à un manque de compréhension de la part des techniciens de l'atelier. Leur objectif principal est de garder une trace des factures et des composants défectueux et moins d'enregistrer avec précision les données du calendrier.

Table 1

Un extrait avec les premières lignes de la base de données étudiée.

Date	Facture	Km	Véhicule	Nom	Unité	Qté	Valeur
110417	16508	10000	B52TAZ	FILTRE D'HUILE	pce	1	7,29
240617	17071	21500	B52TAZ	TAPIS DE PLANCHE	jeu	1	36,00
200417	16735	17900	B53TAZ	AMPOULE PHARE	pce	1	6,40
060417	16439	8400	B52TAZ	ACUMULATEUR	pce	1	98,50
101017	50157	60000	B53TAZ	FILTRE À AIR	pce	1	17,73
101017	50157	60000	B53TAZ	GALET TENDEUP	pce	1	125,56
251017	50461	63000	B52TAZ	BOÎTIER DIRECTION	pce	1	920,17
011117	50572	64000	B53TAZ	BRAS TRAVERSAL	pce	1	140,37
310517	17595	30750	B53TAZ	BOUGIES	jeu	4	12,86
101017	50157	60000	B53TAZ	COURROIE DISTRIB.	pce	1	42,75
271217	55315	86000	B52TAZ	BOÎTIER PAPILLON	pce	1	630,25
240617	17071	21500	B52TAZ	COUVRE-SIÈGE	jeu	1	75,00

Dans certains cas, la date peut être considérée comme la date du diagnostic, c'est-à-dire quand le problème a été découvert et peut ne pas être la même que la date de réparation.

Dans la table 1, on peut voir, marqués en rouge, certains mots mal orthographiés; en vert ont été mis en évidence des exemples de mots d'arrêt qui doivent être supprimés.

La préparation des données, qui représente plus de 80% du temps total alloué à l'analyse, comporte plusieurs étapes.

Étape 1 - normalisation des données

La normalisation du texte consistera à supprimer les signes diacritiques, les caractères spéciaux et la transformation de toutes les majuscules en minuscules.

Étape 2 - suppression des mots d'arrêt

Les mots d'arrêt sont des mots qui sont très souvent utilisés dans une langue mais qui n'apportent pas de contribution supplémentaire au sens de la phrase, à titre d'exemple, ceux marqués en vert (Fig. 2).

Étape 3 - stemming

L'étape de stemming consiste à supprimer les suffixes et préfixes des mots et à ne conserver que les formes singulières.

Étape 4 - lemmatisation

L'étape de lemmatisation prend en compte une analyse morphologique des mots afin de ne trouver que des noms.

Étape 5 - la détermination du nombre d'occurrences de chaque mot

Pour cela, un tableau est créé: sur les colonnes sont tous les mots utilisés et sur les lignes sont listés les commentaires individuels de notre base de données.

La première action sera de créer la matrice d'apparence des mots dans les différents commentaires en utilisant la fonction *CountVectorizer ()* [9] du module Scikit-Learn en Python.

La matrice de mots est utilisée pour déterminer combien de fois un mot apparaît dans un texte (table 2). Cependant, cela ne signifie pas qu'un mot avec un grand nombre d'occurrences soit déterminant dans le classement du commentaire. En effet, le même mot peut être répété sept fois dans un seul commentaire.

Table 2
Fréquence des mots dans la base de données étudiée.

pièce	courroie distribution	ensemble biellette	accumulateur	courroie alternateur	tête de biellette	disque embrayage	bascule	bras transversal	pompe à eau	galet tendeur	barre anti-évers	boîtier papillon	boîtier direction	assistance freinage
fréquence des mots	119	105	86	81	53	36	25	22	21	18	14	14	13	4

Nous préférons donner un poids aux mots en fonction de leur fréquence d'occurrence dans un commentaire TF (*Term Frequency*) et de leur nombre d'occurrences dans l'ensemble de commentaires IDF (*Inverse Document Frequency*). Ce poids appelé TF-IDF qui est calculé avec la fonction *TdifTransformer()* [9] est également utilisé pour exprimer la rareté du mot dans tous les commentaires.

Nous pouvons créer un diagramme de fréquence des termes les plus courants (Fig. 2):

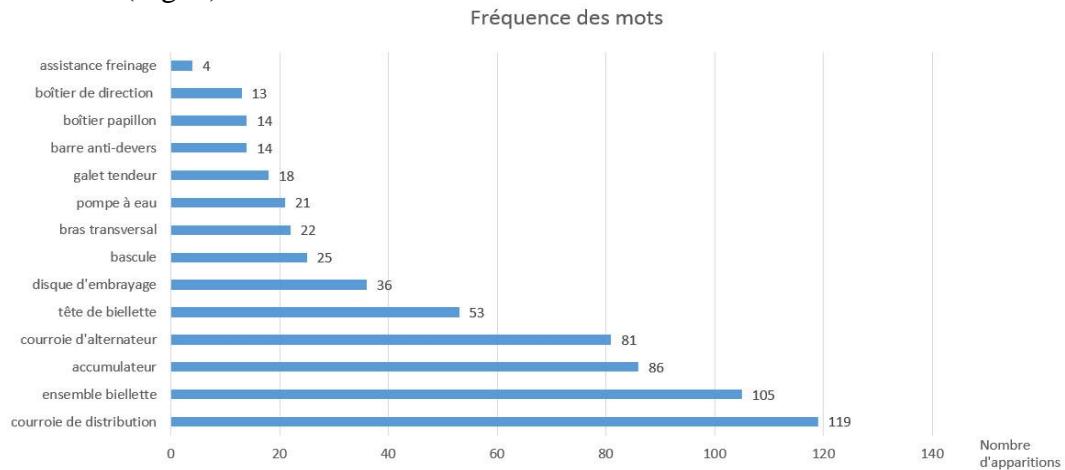


Fig. 2. Diagramme de fréquence pour les mots faisant référence aux pièces couvertes par la garantie.

On obtiendra ainsi le diagramme de Pareto (Fig. 3) avec l'importance des dépenses sur chaque type de pièce remplacée pendant la période de garantie:

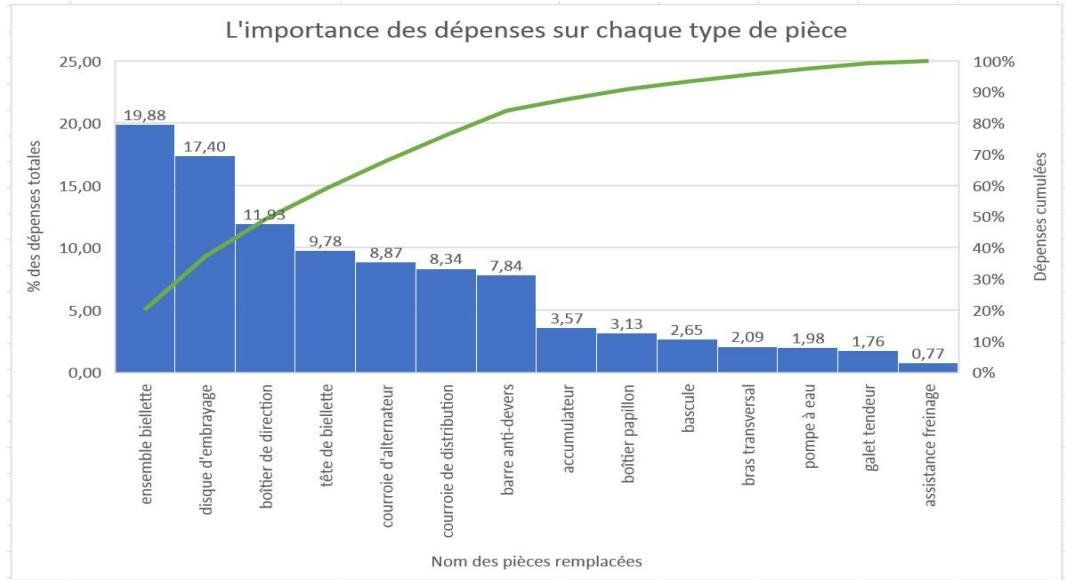


Fig. 3. Diagramme de Pareto - l'importance des dépenses sur chaque type de pièce remplacée pendant la période de garantie, en pourcentage des dépenses totales.

6. Résultats

Lors d'une première analyse du diagramme de Pareto (Fig. 4), on observe que l'importance des dépenses pour les quatre premières pièces représente 58,99% des dépenses totales sous garantie. Si nous étendons la zone d'analyse aux quatre pièces suivantes, nous atteignons un pourcentage de 87,61% du coût total de la garantie.

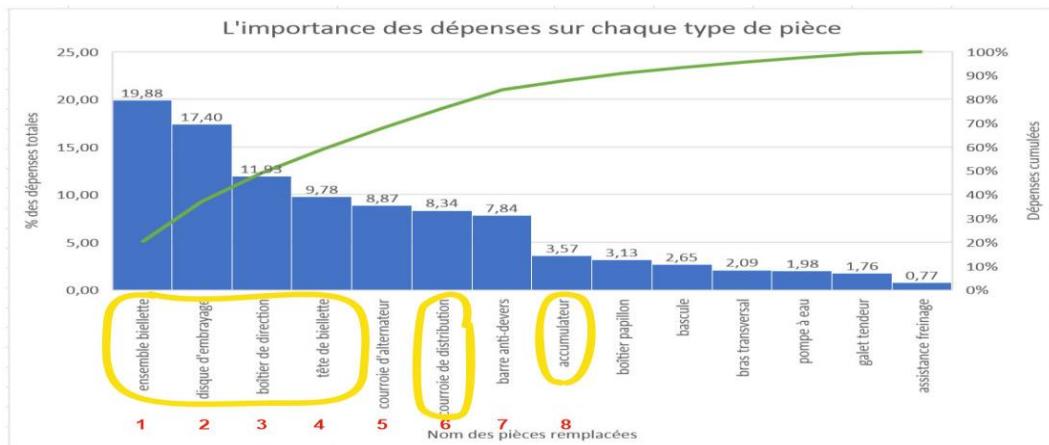


Fig. 4. Diagramme de Pareto - l'importance des dépenses sur les pièces marquées en jaune représente 70,90% des dépenses totales sous garantie.

La prochaine étape est une analyse plus approfondie des pièces marquées en jaune. La pièce 1 a un très grand nombre de plaintes, 105, pour un parc à 100 véhicules sur une période de 24 mois. À la suite de l'analyse, le constructeur a demandé au fournisseur d'équipement d'origine de remplacer la pièce non performante aux frais de ce dernier.

La pièce 2, bien qu'elle ait un volume moyen de plaintes – 36, implique des dépenses très élevées, 17,40% des dépenses totales et dispose d'une procédure de diagnostic et de réparation très bien standardisée. Il faut vérifier si les réparateurs respectent les procédures de diagnostic et de réparation spécifiées.

Après une enquête plus approfondie, il a été constaté que les techniciens n'avaient pas suivi les étapes de diagnostic et de réparation appropriées avant de remplacer la pièce 2. Les techniciens ont détourné des étapes importantes de la procédure de service et ont sauté à la mauvaise conclusion pour remplacer la pièce 2. En conséquence, le volume des demandes de réparation a été gonflé, ce qui a entraîné des coûts très élevés pour la pièce 2.

Si les techniciens avaient suivi la procédure de réparation, une réparation beaucoup plus simple et moins coûteuse aurait pu être effectuée.

La pièce 8, quoiqu'elle ne représente que 3,57% du coût total de la garantie, peut affecter l'image et la réputation de la marque en raison du très grand nombre de réclamations - 86. Le fournisseur a reconnu qu'il avait des problèmes de qualité des produits et a décidé de prendre en charge la garantie complète.

7. Conclusions

Partant de notions théoriques sur l'apprentissage automatique, le traitement du langage naturel et l'extraction de données, l'article présente une méthodologie moderne pour évaluer la fiabilité des composants mécaniques et réduire les coûts de garantie d'une flotte de véhicules. La base de données réelle contenant l'historique des interventions effectuées sur les véhicules en deux ans, fournie par une société de transport automobile, a été utilisée. L'originalité de l'article réside dans le fait qu'il construit un lien concret et utile entre les aspects théoriques modernes et les cas rencontrés dans la réalité quotidienne.

En guise de nouveauté, de nombreuses informations utiles, généralement ignorées, ont été mises au jour, à l'aide de cette méthodologie d'analyse efficace.

Les résultats obtenus à partir de cette analyse ont conduit à une réduction des coûts de réparation de 12%, à une reconsideration du remplacement des pièces problématiques, à une diminution du nombre d'événements réclamés de 17% et par conséquent à une augmentation de la confiance des clients dans la marque.

Enfin, on a été mis en évidence, le fait que l'analyse de texte peut réduire les coûts des constructeurs en découvrant plus rapidement les problèmes de fiabilité et en détectant les problèmes de maintenance.

RÉFÉRENCES

- [1]. Warranty Week, August, 20, 2019, <https://www.warrantyweek.com/archive/ww20190822.htm>
- [2]. A. Deshpande, M. Kumar, Artificial Intelligence for Big Data, Packt Publisher, 2018.
- [3]. O. Gusikhin, N. Rychtyckyj, D. Filev, 2007. Intelligent systems in the automotive industry: applications and trends. *Knowl. Inf. Syst.* **12**, 147–168.
- [4]. J. Buddhakulsomsiri, A. Zakarian, Sequential pattern mining algorithm for automotive warranty data. *Comput. Ind. Eng.* **57**, 137–147, 2009.
- [5]. D. G. Rajpathak, An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain. *Comput. Ind.* **64**, 565–580, 2013.
- [6]. T. R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition* **5** (2) (1993) 199-220.
- [7]. R. Studer, V. R. Benjamins, D. Fensel, Knowledge engineering: principles and methods, *IEEE Transactions on Knowledge and Data Engineering* **25** (1-2) (1998) 161-197.
- [8]. S. Masson, New Car Warranty – Do I have to have my car serviced by a dealership?, January, 30, 2019, <https://www.thecarexpert.co.uk/new-car-warranty-servicing/>
- [9]. J. Brownlee, Deep Learning for Natural Language Processing – Develop Models for Natural Language in Python, Machine Learning Mastery, 2017.