* **Critère de fraude**
* Les incohérences simples telle qu’un écart entre le total et la somme des prix, entre le total et le total payé, ou entre la quantité, le prix unitaire et le prix du produit.
* Détection de copier-coller … suivre ce lien page 4 : chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://coria-taln-2023.sciencesconf.org/461873/document
* Modification du prix d’un article , son nom, le moyen de paiement. La falsification peut également viser une extension indue de la garantie en modifiant la date.
* CPI : copier-coller à l’intérieur du document (ou copy-move forgery)
* CPO : copier-coller à l’extérieur du document (ou splicing)
* IMI : pour imitation, boite textuelle imitant la police de caractères du ticket
* CUT : suppression d’un ou plusieurs caractères
* Vérification de variation des prix : recherche des prix aberrants
* Vérification du total à payer : examen des incohérences dans les prix des articles et le montant à payer
* Vérification de texte manquant : recherche des mots-clés qui impliquent un élément d’information spécifique, mais cette information est manquante
* Vérification des remises : recherche des incohérences dans les promotions
* Vérification des quantités : recherche des incohérences dans la formule quantité × prix de l’article = somme
* Vérification des dates : recherche des dates non valides
* Modules basés sur l’image (utilisant OpenCV) :
* Contrôle des couleurs : recherche de saturation artificielle, de noirceur, ou de « bruits de poivre » (pixels noirs)
* Vérification des parties effacées : recherche des zones blanches non naturelles ou de grandes zones homogènes (qui n’ont pas de bruit)
* Vérification des copier-coller : recherche des composantes connexes identiques dans les images binarisées

**Somme des prix = total** Nous calculons d’abord la somme des prix de chaque produit du ticket et nous la comparons au montant total que nous avons extrait. En effet, il est fréquent dans notre corpus que les fraudeurs aient oublié de modifier les totaux en conséquence du changement des prix des produits.

**Somme d’articles = Nombre articles** Le deuxième indice porte sur le nombre extrait d’articles, comparé au nombre calculé de produits extraits. Il est possible que les fraudeurs ne fassent pas attention à ce genre de détails quand ils suppriment ou ajoutent un produit. La suppression de produit peut par exemple avoir lieu dans le cas de remboursement de frais de mission, où il n’est pas forcément bien vu de prendre certaines boissons alcoolisées. Cette information n’est cependant pas relevée sur tous les tickets de caisse, ce qui implique que cet indice ne pourra être pertinent que pour les tickets Carrefour.

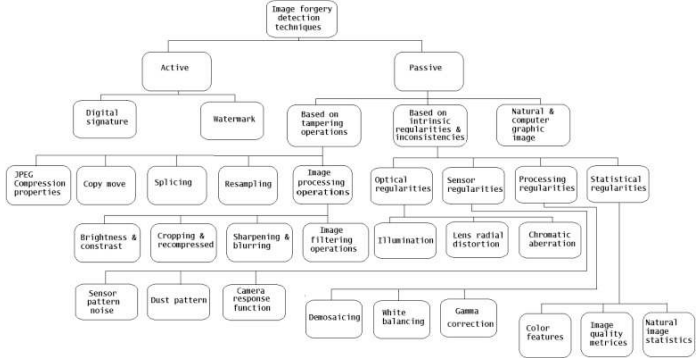
**Total = paiement** Le troisième indice concerne le montant payé extrait et le montant total extrait. En effet, nous avons remarqué dans notre corpus qu’il est fréquent que les fraudeurs oublient de reporter le montant total fraudé sur le montant payé, probablement parce que les abréviations utilisées rendent difficile l’identification de cette information pour les profanes, qui ne voient pas dans des « CB EMV », « TR » et autres « ESP » la signification que notre outil, lui, peut relever.

**Quantité × prix unitaire = prix** Sur les tickets de Carrefour, nous relevons la quantité et le prix unitaire des produits. Certaines fraudes portent sur ces informations et nous cherchons donc à vérifier que les fraudeurs n’ont pas fait d’erreurs dans la multiplication de ces deux informations : le prix total du produit doit en effet correspondre au résultat de cette multiplication.

**Poids × prix au kilogramme = prix** Nous vérifions également, dans le cinquième indice, l’égalité du résultat de la multiplication du poids d’un produit par son prix au kilogramme et du prix total, quand ces informations existent et sont relevées.

**Date et heure** Le sixième indice porte sur la vérification du format de la date et de l’heure

**Loi de Benford** Le dernier indice cherche à vérifier la « loi de Benford » sur les prix des produits. La loi généralisée de Benford établit que la distribution de certains chiffres dans de nombreuses séquences de chiffres de la vie réelle ne suit pas une distribution uniforme. Cela signifie que, dans un ensemble de données sur n’importe quel domaine (comptabilité, démographie, presse, articles scientifiques...), il est fréquent qu’un chiffre soit sur-représenté par rapport aux autres (Nigrini 2012). D’après Durtschi et al. (2004), cette loi est souvent utilisée pour détecter les fraudes, avec plus ou moins de résultats, lors d’audits comptables. Le constat est que la répartition du premier chiffre significatif d’un nombre est logarithmique : il y a ainsi plus de nombres commençant par 1 que par 2, plus par 2 que par 3... Nous avons donc testé cette loi sur chacun des tickets de notre corpus comme septième indice, ainsi que sur l’ensemble de notre corpus (voir l’annexe C).



Sept manipulations de fraude sont identifiées par Poisel & Tjoa (2011) :

— le Copy-Move Forgery (CMF) : la duplication d’une partie de l’image dans la

Même image ;

— la retouche d’image ;

— le filtrage d’une partie non-désirée de l’image ;

— la suppression partielle d’un objet de l’image ;

— le splicing : la combinaison de plusieurs images ;

— la manipulation sur la luminance, la couleur et le contraste ;

— la manipulation sur la géométrie de l’image.- le Copy-Move Forgery Detection (CMFD) : duplication d’une ou plusieurs parties de l’image.

Viennent s’ajouter a cela les bruits locaux, les aberrations chromatiques, les flous et contrastes, identification de la source.

* Détecter si un caractère n’est pas aligné dans le sens du document puis comparer aux autres ou détecter si une ligne n’est pas alignée aux autres sur le document.