Etat de l'art - Outils de détection de fuites de données sensibles
Groupe 29

Nom de l'outil	Types de données détectées	Types de fichiers analysés	Approche	Précision	Testé ?
Symantec DLP	Numéro de téléphone, de documents d'identité, de comptes bancaires / cartes bancaires, dates de naissance, adresses postales / email, code source, chiffres stratégiques / financiers, numéros d'assurance maladie, informations médicales, etc.	330 types de fichiers pris en charge, dont : - Fichiers texte - Office (.docx, xlsx, etc) - Images - Code source - Bases de données (.sql, etc..)	Scan automatique des fichiers sur le système d'information de l'entreprise, qui détecte les fuites de données sensibles contenues dans les documents. Prévention des fuites en bloquant le partage des fichiers, en alertant les utilisateurs, etc. Utilisation d'OCR	Détection de la validité des numéros de carte bancaire avec l'algorithme de Luhn. Utilisation d'expressions régulières pour déterminer la provenance des numéros de téléphone.	❌
Varonis DLP	Adresses email, IBAN, adresses physiques, numéros de documents d'identité, données bancaires, numéros d'assurance maladie et de sécurité sociale, mots de passe, chiffres stratégiques (statistiques, financiers, …), dates de naissance , etc	60 types de fichiers pris en charge, dont : - Documents (ex : .docx, .odt, …) - Présentations (.pptx par exemple) - PDF - Images - Feuilles de calcul (ex : .xlsx) - Archives ZIP - Code source	Système de classification automatique via des tags des documents en fonction des données sensibles retrouvées dans leurs contenus. Détecte également automatiquement quand des fichiers avec des données sensibles sont partagés. Score de risque lié à chaque employé afin de pouvoir détecter de potentielles menaces internes. Utilisation de l'IA pour la classification	Détection des émetteurs de carte bancaire, des pays de provenance des adresses physiques, …	❌

			automatique.		
Forcepoint DLP	Mots clés particuliers, expressions régulières, numéro de cartes bancaires, de téléphone,	De nombreux types de fichiers pris en charge, dont : - Documents Microsoft Office - PDF - Archives ZIP - Fichiers texte - Images - Code source - Fichiers FTP - Fichiers Flash	Système d'analyse en temps réel relié à tous les services utilisés par l'entreprise (Microsoft Office, AWS, …) qui émet des alertes si des données sensibles sont retrouvées dans les fichiers. Les fichiers sont ensuite classifiés automatiquement selon les données qu'ils contiennent, et certains blocages peuvent être mis dessus (ex : interdire leur impression ou leur envoi par mail.) Cela peut être personnalisé via des règles de sécurité. Visualisation du niveau de risque à chaque employé via un score. Utilisation de LLMs pour la détection, ainsi que des expressions régulières.	Vérification de la validité des cartes bancaires via l'algorithme de Luhn.	❌(mais démo disponible sur Youtube
Digital Guardian (Fortra)	Adresses, numéros de téléphone, de sécurité sociale, de documents d'identité, IBAN, mots de passe, dates de naissance, mots de passe, identifiants internes, clé d'API et tokens, coordonnées GPS, etc.	Beaucoup de types de fichiers pris en charge, dont : - Documents Microsoft Office - PDF - Images - Fichiers texte - Code source	Système de classification des fichiers, qui précise quel type de données sensibles le fichier comporte via des tags dans ses métadonnées. Ensuite, selon cette classification, certaines restrictions peuvent être mises	Reconnaissance des émetteurs de carte bancaire (ex : Visa, Mastercard, …) Reconnaissance du pays de provenance d'un numéro d'identification, pareil pour les numéros de téléphone et les adresses.	❌(mais démo disponible sur Youtube)

		- Bases de données (ex : .sql) - Fichiers CAO	sur le fichier (ex : pas possible de l'envoyer par mail à un certain service). Mais si le fichier a déjà été partagé par exemple, le système permet de supprimer les différents accès au fichiers, voire même de le supprimer à certains endroits.		
Teramind	Numéros de carte bancaire, de sécurité sociale, de téléphone, mots-clés particuliers, adresses, adresses IP, numéro d'immatriculation, URLs privés, clé cryptographiques privés, etc.	Beaucoup de types de fichiers pris en charge, dont : - Documents Microsoft Office - PDF - Images - Fichiers texte - Code source - Bases de données (ex : .sql) - Fichiers CAO - Archives .zip	Système de règles qui permet de définir un contexte de détection, les types de données à détecter et permet de définir des conditions particulières, et un niveau de gravité. Chaque employé à un score de "risque" calculé à partir du nombre de règles déclenchées sur des données le concernant. Teramind permet aussi de déclencher des actions, comme avertir l'utilisateur, faire une commande particulière sur l'ordinateur, etc.	Possibilité de cibler uniquement les numéros téléphones / mails etc.. d'un certain pays, d'un certain opérateur ou d'une certaine banque.	✅