Rapport sur la détection de l'URL malveillante à l'aide de machines à vecteurs de support

Filière : Cycle d'ingénieur en DATA SCIENCE
Niveau : 4ème Année

Elaboré par:
Eya Bouhouch
Hajer Zitouni

Sommaire

Liste des figures

1 Introduction

Les sites Web malveillants représentent un risque important en matière de cybersécurité. Ils présentent souvent un contenu non sollicité tel que du spam, des fraudes, des tentatives de phishing, des exploits et bien plus encore. Ces sites attirent les utilisateurs d'Internet sans méfiance pour devenir victimes de fraudes et d'escroqueries, causant ainsi des dommages de plusieurs milliards de dollars chaque année. Il est crucial de découvrir et d'agir rapidement contre de telles menaces.

2 Objectif et Solution

Notre objectif principal est de concevoir un système efficace pour détecter les URL malveillantes, réduisant ainsi les risques potentiels pour les utilisateurs en ligne. Pour atteindre cet objectif, nous avons choisi d'utiliser un modèle de machine à vecteurs de support (SVM) en raison de sa capacité à traiter efficacement des ensembles de données complexes et à généraliser les modèles appris.

Pour résoudre ces problèmes, nous avons travaillé sur un système qui détectera les URL malveillantes en utilisant des techniques d'apprentissage automatique. L'aspect important de la détection d'URL uniformes malveillantes est l'extraction des caractéristiques. Nous avons collecté un ensemble de données d'entraînement qui est ensuite entraîné pour les fonctionnalités extraites. Les principaux attributs que nous avons extraits sont basés sur l'hôte, basés sur le lexique et basés sur la popularité.

Nous avons choisi le modèle SVM en raison de sa capacité à traiter efficacement les données à plusieurs dimensions et à séparer linéairement les classes dans l'espace des caractéristiques. Ce modèle offre également une flexibilité pour explorer différents noyaux, nous permettant d'expérimenter avec différents noyaux pour obtenir les meilleurs résultats de classification.

Nous allons utiliser ce modèle SVM pour construire notre système de détection d'URL malveillantes et évaluer sa performance sur des ensembles de données de test et de validation. En extrayant judicieusement les caractéristiques des URL et en utilisant un modèle SVM bien ajusté, nous visons à fournir une solution efficace pour protéger les utilisateurs contre les menaces en ligne.

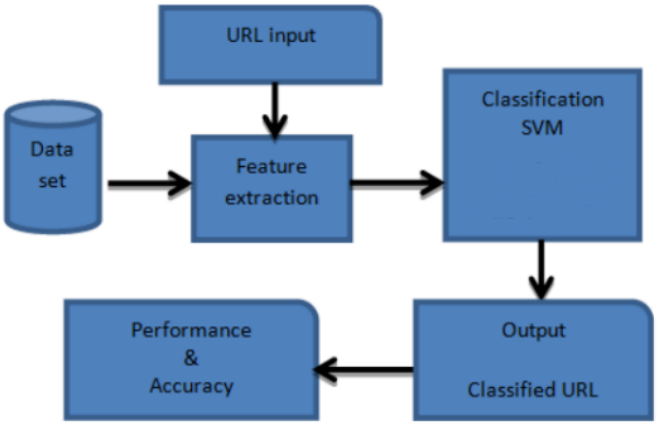

FIGURE 1 – Schéma de la solution

3 Exploration du jeu de données

Nous avons collecté un ensemble de données énorme de 651 191 URL, parmi lesquelles 428 103 sont des URL bénignes ou sécurisées, 96 457 sont des URL de défiguration, 94 111 sont des URL de phishing, et 32 520 sont des URL de logiciels malveillants. Pour collecter des URL bénignes, de phishing, de logiciels malveillants et de défiguration, nous avons utilisé le jeu de données URL (ISCX-URL-2016). Pour augmenter les URL de phishing et de logiciels malveillants, nous avons utilisé le jeu de données de la liste noire de domaines malveillants. Nous avons augmenté les URL bénignes en utilisant le dépôt Git de Faizan. Enfin, nous avons augmenté un plus grand nombre d'URL de phishing en utilisant le jeu de données Phishtank et le jeu de données PhishStorm. Comme nous vous l'avons dit, l'ensemble de données est collecté auprès de différentes sources. Nous avons d'abord collecté les URL auprès de différentes sources dans un DataFrame séparé, puis les avons fusionnées pour ne conserver que les URL et leur type de classe.

	url	type
0	br-icloud.com.br	phishing
1	mp3raid.com/music/krizz_kaliko.html	benign
2	bopsecrets.org/rexroth/cr/1.htm	benign
3	http://www.garage-pirenne.be/index.php?option=...	defacement
4	http://adventure-nicaragua.net/index.php?optio...	defacement
...	...	...
651186	xbox360.ign.com/objects/850/850402.html	phishing
651187	games.teamxbox.com/xbox-360/1860/Dead-Space/	phishing
651188	www.gamespot.com/xbox360/action/deadspace/	phishing
651189	en.wikipedia.org/wiki/Dead_Space_(video_game)	phishing
651190	www.angelfire.com/goth/devilmaycrytonite/	phishing

651191 rows × 2 columns

FIGURE 2 – Extrail de la dataset

4 Prétraitement des données et extraction des caractéristiques

Nous avons extrait plusieurs caractéristiques des URL pour les utiliser dans notre modèle de détection. Cela comprend des caractéristiques lexicales telles que la longueur de l'URL, la longueur de chaque composant de l'URL, le nombre de caractères spéciaux, etc. Nous avons également utilisé des caractéristiques basées sur l'hôte, telles que la présence d'une adresse IP dans l'URL, la normalité de l'URL basée sur le nom d'hôte, et d'autres caractéristiques telles que la présence de services de raccourcissement d'URL, le nombre de mots suspects, etc.

5 Description du modèle

Le modèle que nous avons choisi pour détecter les URL malveillantes est le Support Vector Machine (SVM). Les SVM sont des modèles d'apprentissage supervisé qui sont efficaces pour la classification binaire et peuvent également être étendus à des tâches de classification multiclasse. Ils fonctionnent en trouvant l'hyperplan qui sépare au mieux les exemples des différentes classes dans l'espace des caractéristiques.

Dans notre cas, nous avons utilisé un SVM avec un noyau linéaire pour séparer les URL malveillantes des URL bénignes dans un espace de caractéristiques de grande dimension. Ce choix a été fait en raison de la nature linéairement séparable des données après extraction des caractéristiques. Cependant, les SVM offrent également la possibilité d'utiliser d'autres noyaux, tels que le noyau gaussien (RBF), pour des données plus complexes qui ne sont pas linéairement séparables.

6 Construction et formation du modèle

Nous avons divisé nos données en ensembles d'entraînement, de validation et de test. En utilisant un modèle de machine à vecteurs de support avec un noyau linéaire, nous avons formé notre modèle sur l'ensemble d'entraînement. Ensuite, nous avons évalué les performances du modèle sur l'ensemble de validation pour ajuster les hyperparamètres si nécessaire. Enfin, nous avons testé notre modèle sur l'ensemble de test final pour évaluer sa performance.

7 Résultats

Sur l'ensemble de validation, notre modèle a montré une précision de 90,27 %, une précision de 90,67 %, un rappel de 79,70 %, et un score F1 de 84,83 %. Sur l'ensemble de test final, les performances étaient similaires avec une précision de 90,10 %, une précision de 90,66 %, un rappel de 79,15 %, et un score F1 de 84,52 %.

8 Conclusion

Dans ce projet, nous avons démontré comment les techniques d'apprentissage automatique peuvent aider à la détection d'URL malveillantes en fonction de l'ensemble de fonctionnalités fourni. En extrayant et en utilisant judicieusement les caractéristiques des URL, nous pouvons construire des modèles efficaces pour protéger les utilisateurs contre les menaces en ligne.