

***Gabriela Aparecida Alcântara Silva**

Discente do curso tecnólogo em Análise e Desenvolvimento de Sistemas da Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Brasil.
ORCID: 0009-0006-3467-6182
E-mail: gabrielaalcantara18@gmail.com

****Wesley Dias Maciel**

Docente da Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Brasil.
ORCID: 0000-0003-0879-0943
E-mail: wesleydiasmaciel@gmail.com

SMISHING E INTELIGÊNCIA ARTIFICIAL: uma Análise baseada em Revisão Sistemática de Literatura sobre as Abordagens e Desafios em Pesquisas Recentes

DOI: doi.org/10.5281/zenodo.22713185

RESUMO

Este trabalho é baseado em uma Revisão Sistemática de Literatura com o objetivo de analisar o estado da arte do smishing, um golpe digital que utiliza mensagens SMS para enganar usuários e obter dados confidenciais. A análise evidenciou a escassez de coleções de mensagens públicas e centralizadas, o que compromete o desenvolvimento e o treinamento de modelos de aprendizado de máquina. Diante desse cenário, o estudo destaca a necessidade de criação e disponibilização de coleções de mensagens mais acessíveis. Essa iniciativa pode impulsionar pesquisas futuras e contribuir para o fortalecimento da segurança digital frente a ataques de engenharia social cada vez mais sofisticados.

Palavras-chave: Smishing; Short Message Service; Machine Learning; Segurança Digital.

SMISHING AND ARTIFICIAL INTELLIGENCE: An Analysis Based on a Systematic Literature Review of Approaches and Challenges in Recent Research

ABSTRACT

This work is based on a Systematic Literature Review aimed at analyzing the state of the art of smishing, a digital scam that uses SMS messages to deceive users and obtain confidential data. The analysis revealed a scarcity of public and centralized message collections, which compromises the development and training of machine learning models. Given this scenario, the study highlights the need to create and make available more accessible message collections. This initiative can boost future research and contribute to strengthening digital security against increasingly sophisticated social engineering attacks.

Keywords: Smishing. Short Message Service. Machine Learning. Digital Security.

SMISHING E INTELIGENCIA ARTIFICIAL: Un análisis basado en una revisión sistemática de la literatura sobre enfoques y desafíos en la investigación reciente

RESUMEN

Este trabajo se basa en una revisión sistemática de la literatura cuyo objetivo es analizar el estado del arte del smishing, una estafa digital que utiliza mensajes SMS para engañar a los usuarios y obtener datos confidenciales. El análisis reveló la escasez de colecciones de mensajes públicas y centralizadas, lo que dificulta el desarrollo y el entrenamiento de modelos de aprendizaje automático. Ante esta

situación, el estudio subraya la necesidad de crear y poner a disposición colecciones de mensajes más accesibles. Esta iniciativa puede impulsar la investigación futura y contribuir a fortalecer la seguridad digital frente a ataques de ingeniería social cada vez más sofisticados.

Palabras clave: Smishing. Servicio de Mensajes Cortos. Aprendizaje Automático. Seguridad Digital.

INTRODUÇÃO

Com o avanço da comunicação móvel, o Short Message Service (SMS) consolidou-se como uma forma de comunicação rápida, sendo amplamente utilizado em diversas aplicações empresariais, como cobranças, transações, campanhas promocionais, lembretes e envio de códigos de segurança (Rai, 2025). Entretanto, a popularização do SMS também trouxe novos desafios, especialmente o aumento no número de mensagens fraudulentas, um golpe conhecido como smishing. O smishing tem o objetivo de enganar os destinatários, induzindo-os a fornecer informações confidenciais ou a realizar ações prejudiciais. Entretanto, existe uma escassez de coleções de documentos sobre mensagens de smishing que estejam publicamente disponíveis. Tais coleções são importantes para subsidiar pesquisas em segurança digital, especialmente no treinamento, teste e validação de algoritmos de aprendizado de máquina voltados à detecção automática de fraudes (Mishra; Soni, 2023).

A ausência de bases abrangentes e atualizadas compromete o desempenho e a generalização de modelos de detecção de ataques de smishing, sobretudo diante da constante evolução das estratégias utilizadas por golpistas (Yerima; Bashar, 2022). Neste trabalho, foram analisados estudos sobre detecção de smishing por meio de inteligência artificial. O objetivo foi analisar as abordagens para classificação de mensagens fraudulentas e compreender os desafios no combate ao smishing. A justificativa do trabalho foi avaliar as coleções de mensagens utilizadas nos algoritmos de aprendizado de máquina, buscando identificar características, limitações e contribuições para o avanço das pesquisas na área. O trabalho também examinou os algoritmos empregados, considerando critérios de pré-processamento das mensagens e a acurácia alcançada nos resultados.

A metodologia adotada foi baseada em uma revisão sistemática de literatura (Kitchenham, 2004; Kitchenham et al., 2008; Kitchenham et al. 2010; Kitchenham et al. 2011). Os resultados mostraram o uso combinado de bases públicas e privadas, juntamente

com mensagens próprias e específicas de cada estudo avaliado. Também mostraram a utilização de algoritmos como Random Forest, Support Vector Machine (SVM) e Naive Bayes. A avaliação e tratamento de mensagens de smishing para treinamento, teste e validação de algoritmos de aprendizado de máquina ajuda desenvolver pesquisas aplicadas à detecção de mensagens maliciosas, impactando no desempenho dos algoritmos e na reprodutibilidade dos estudos. Assim, como contribuição, o trabalho apresentou um panorama do combate ao smishing auxiliado por inteligência artificial, destacando a necessidade de criação de um repositório de mensagens que possa ser avaliado, atualizado e utilizado pela comunidade científica publicamente.

1. FUNDAMENTAÇÃO TEÓRICA

A difusão da Internet impulsionou significativamente a cooperação e o compartilhamento de informação em tempo real, transformando a forma como indivíduos e organizações se comunicam (Mehmood et al, 2024). Com o avanço das tecnologias móveis, o SMS consolidou como um dos meios de comunicação mais utilizados, devido à sua praticidade, ampla acessibilidade e baixo custo (Patra et al, 2023). Inicialmente restrito a interações interpessoais, o SMS passou a ser amplamente incorporado por diversos setores empresariais, principalmente para estratégias de marketing digital, autenticação de usuários e relacionamento com consumidores (Franco, 2024) (Quadro 1).

Quadro 1: principais usos do SMS por empresas públicas e privadas e os meios como esse tipo de mensagem chega aos usuários.

Principais Usos do SMS	Forma de Recebimento do SMS
Marketing e campanhas promocionais	Alertas de vendas, cupons, promoções e atualizações de eventos, com foco em aumentar o engajamento e impulsionar vendas. Exemplo: ações de cashback para compras futuras
Autenticação de dois fatores (2FA)	Confirmação de identidade por meio do envio de códigos temporários, oferecendo uma camada adicional de segurança no acesso a sistemas e aplicativos

Alertas urgentes e notificações transacionais	Mensagens sobre tentativas de fraude, lembretes de pagamento e comprovantes de transações financeiras, como compras com cartão
Atendimento ao cliente	Suporte automatizado, envio de atualizações de pedidos, lembretes de compromissos e posição em filas de espera, muito comum em restaurantes, bares e prestadores de serviços
Pesquisas e coleta de feedback	Questionários rápidos, como os de Net Promoter Score (NPS), que permitem identificar melhorias em produtos e serviços
Comunicações internas	Envio de alertas e atualizações críticas, com alta taxa de leitura e resposta, especialmente úteis para equipes de tecnologia da informação e de operações remotas
Confirmações de entrega e status de pedidos	Acompanhamento logístico, por meio de notificações com links de rastreamento, utilizado por empresas de e-commerce e serviços postais como os dos Correios

Fonte: elaborado pelos autores.

A expansão do uso do SMS também acarretou desafios no campo da segurança digital. Entre as ameaças, destaca-se o smishing, termo derivado da combinação de SMS e phishing, que se refere ao envio de mensagens fraudulentas contendo links maliciosos, com o intuito de induzir o usuário a fornecer informações sensíveis ou executar ações prejudiciais (Patra, 2023). Tais mensagens frequentemente simulam comunicações legítimas de instituições conhecidas e empregam técnicas de engenharia social (Kumarasinghe et al, 2023), levando as vítimas a clicar em links, responder a mensagens ou informar dados pessoais sigilosos. Como consequência, informações confidenciais, como números de cartões de crédito, Card Verification Value (CVV) e credenciais bancárias, podem ser expostas ilegalmente (Mehmood et al, 2024). Em alguns casos, o ataque resulta ainda na instalação de malwares que comprometem a integridade de dispositivos (Kumarasiri et al, 2023).

Estudos recentes apontam o aprendizado de máquina como uma alternativa promissora para a detecção de mensagens fraudulentas (Boukari et al, 2021). A eficácia dos algoritmos de aprendizado de máquina depende diretamente da qualidade de coleções de

treinamento, teste e validação, para classificação e detecção automática (Mishra, Soni, 2023). Nesse contexto, o smishing é um problema clássico de classificação binária, em que mensagens SMS são categorizadas como maliciosas (smishing) ou legítimas (ham) (Chen et al., 2023).

2. TRABALHOS RELACIONADOS

Abdillah et al. (2022) realizaram uma revisão sistemática de literatura relacionada a ataques de phishing em que as análises foram organizadas em três categorias: coleções de mensagens, avaliação de desempenho e tipos de phishing. As coleções de mensagens foram classificadas em três grupos: públicas, privadas e combinadas (públicas e privadas). As principais bases utilizadas foram PhishTank que forneceu publicamente informações sobre URLs suspeitas de phishing (34 autores), Aburrous et al. (2010 apud Abdillah et al., 2022), enquanto a Alexa (15 autores), Chen et al. (2014 apud Abdillah et al., 2022) é uma ferramenta de análise da Web que obtém dados de clientes por meio de uma barra de ferramenta instalada e UCI Machine learning foi utilizado principalmente para realizar benchmarks em estudos sobre phishing (12 autores) foi utilizado principalmente para realizar benchmarks em estudos sobre phishing (12 autores), Chiew et al. (apud Abdillah et al., 2022).

Dou et al. (2017) realizaram um estudo sobre phishing abordando quatro aspectos: taxonomia, coleções de mensagens, recursos e técnicas de detecção. Os autores destacaram que a disponibilidade pública de conjuntos de mensagens é necessária para padronizar a avaliação e a comparação entre diferentes abordagens. No entanto, a ausência de bases de referência consolidadas tornam complexas a reprodução de experimentos e a validação dos resultados. Os autores apontaram diversas coleções de mensagens: Anti-Phishing Working Group (APWG), Aburrous et al. (2010 apud Dou et al., 2017), Thomas et al. (2011 apud Dou et al., 2017) e Wenyin et al. (2005 apud Dou et al., 2017), PhishTank, Zhang et al. (2007 apud Abdillah et al., 2022) e Ma et al. (2009 apud Abdillah et al., 2022), MillerSmiles, Pan e Ding (2006 apud Dou et al., 2017) e Ramesh et al. (2014 apud Dou et al., 2017), SpamScatter, Prakash et al. (2010 apud Dou et al., 2017) e Ma et al. (2009 apud Dou et al., 2017), MalwareProl, Le et al. (2011 apud Dou et al., 2017), Spam URI Realtime Block List (SURBL), Thomas et al. (2011 apud Dou et al., 2017), e CLEAN

MX, Sanglerdsinlapachai e Rungsawang (2010 apud Dou et al., 2017). Eles também relataram que a ausência de informação como tamanho dos conjuntos de mensagens, redundância e proporção entre os dados de treinamento e teste dificultam pesquisas sobre phishing.

Wangchuck e Gonsalves (2025) conduziram um estudo sobre phishing multimodal em redes sociais. Em redes sociais, ataques de phishing ocorrem por meio de textos, imagens, voz e vídeos. Os autores destacaram que a qualidade e a disponibilidade dos dados representam desafios para o treinamento de modelos. Além disso, os pesquisadores apontaram que sistemas multimodais geram modelos mais complexos, de difícil escalabilidade e com alta demanda por recursos computacionais.

Kyam et al. (2024) conduziram um estudo sobre o uso de técnicas de deep learning (DL) na detecção de e-mails de phishing. Os autores identificaram como limitação a escassez de conjuntos de mensagens, ressaltando que a eficácia dos modelos de DL depende diretamente do tamanho e da diversidade da coleção de mensagens utilizadas no treinamento. No estudo, os pesquisadores demonstraram que conjuntos pequenos, homogêneos ou com características irrelevantes reduzem a capacidade de generalização do modelo para novos padrões de phishing. Embora não tenham definido um tamanho mínimo ideal, eles recomendam pelo menos 3.000 amostras de treinamento para classificadores tradicionais, enquanto modelos DL exigem volumes significativamente maiores. Além disso, os autores propuseram o uso de amostras sintéticas geradas por sistemas de IA generativa como uma alternativa para ampliar a coleção de mensagens, aumentando a diversidade dos dados e a capacidade do modelo de reconhecer novos cenários.

Cossa et al. (2021) conduziram um estudo sobre a detecção de phishing bancário que pode ocorrer por meio de SMS (smishing) ou ligações telefônicas (vishing). A coleta da coleção de mensagens foi conduzida pelos autores através de um aplicativo Android e entrevistas com indivíduos que já sofreram esses ataques. Para classificar as mensagens fraudulentas, os autores analisaram padrões recorrentes em fraudes de loterias e de recibos falsos.

Kaddoura et al. (2022) destacaram que a detecção de spam em mídias sociais ocorre em três etapas principais: coleta, extração de características e aplicação de métodos de

detecção. A coleta foi realizada por meio de Application Programming Interface (APIs), como as do Facebook e do Twitter, que permitem extrair textos, hashtags e palavras-chave. Em seguida, foi realizado um pré-processamento da coleção de mensagens e a extração das características para a aplicação de métodos de detecção. Os autores disponibilizaram livremente as coleções de mensagens para serem utilizadas em estudos de detecção de spam. Além disso, ressaltaram que há um número limitado de conjuntos de mensagens rotulados disponíveis para textos de spam.

Gaber et al. (2024) destacaram que criar um conjunto de mensagens públicas é um desafio. Os autores apontam que existem restrições legais e de direitos autorais associadas à disseminação de dados e informação. Os autores identificaram que há potenciais riscos de segurança quando arquivos maliciosos são liberados para um público que pode não adotar precauções adequadas ao manuseá-los. Por esses motivos, é mais comum utilizar conjuntos de mensagens pequenos e privados em pesquisas. Entretanto, essa prática dificulta a mensuração do progresso e dos resultados, pois os estudos podem não ser replicáveis, já que os pesquisadores trabalham com diferentes conjuntos de mensagens. Nesse contexto, os autores analisaram cinco conjuntos disponíveis: EMBER (Anderson e Roth, 2018), BODMAS (Yang et al., 2022), SOREL-20M (Harang e Rudd, 2022), VirusShare (2023), que é um repositório que contém arquivos com malwares ativos e Malware Bazaar (Abuse.ch, 2023).

Xu et al. (2025) afirmaram que as coleções de mensagens utilizadas no treinamento e teste de ameaças em segurança digital ainda são limitadas, o que evidencia a necessidade de coleções mais abrangentes. Os autores destacaram que a escassez e a baixa qualidade dessas coleções representam desafios significativos, uma vez que as bases públicas não apresentam robustez, consistência e diversidade adequadas. Além disso, ressaltaram que as mensagens disponíveis frequentemente são desbalanceadas ou enviesadas. Como forma de mitigar essas limitações, os pesquisadores propuseram o uso de Large Language Models (LLMs) para ampliar e enriquecer as coleções de mensagens existentes.

Os estudos analisados mostraram que ainda são escassas as coleções de mensagens para treinar, validar e testar algoritmos de aprendizado de máquina. Essa limitação ocorre por fatores como a falta de repositórios que centralizem publicamente mensagens de smishing. Tais restrições comprometem a reprodutibilidade e dificultam o avanço de novas

pesquisas. Diante desse cenário, neste estudo, foi apresentado um levantamento do estado da arte na construção de coleções de mensagens sobre ataque smishing, destacando lacunas que podem orientar o desenvolvimento de futuras abordagens em segurança digital. Também foram levantados os algoritmos de aprendizado de máquina mais utilizados na detecção de smishing.

3. METODOLOGIA

A metodologia adotada no trabalho foi baseada em revisão sistemática de literatura (Kitchenham, 2004; Kitchenham et al., 2008; Kitchenham et al. 2010; Kitchenham et al. 2011). O primeiro passo foi definir as questões de pesquisa (Research Questions, RQ). Depois, foram definidos os termos de busca. Os termos de busca foram empregados na formação do string de busca. O string de busca foi submetido nas bases de artigos selecionadas para a pesquisa. Os artigos retornados foram inicialmente analisados por título e resumo. Então, critérios de inclusão e exclusão foram aplicados para determinar os trabalhos selecionados para responder às questões de pesquisa. O trabalho foi realizado entre agosto e dezembro de 2025. A escrita do artigo ocorreu de agosto de 2025 a março de 2026.

3.1 Questões de Pesquisa

As questões de pesquisa (Research Questions, RQs) foram elaboradas para avaliar os principais desafios e limitações relacionados à detecção de smishing por meio do aprendizado de máquina (Quadro 2).

Quadro 2: questões (Research Questions, RQs) elaboradas para investigar a detecção de smishing por aprendizado de máquina.

ID	Questão	Justificativa
RQ1	Por que o smishing é considerado um tema relevante?	Compreender a relevância do smishing no contexto atual, destacando sua ameaça para a sociedade e para a área de segurança digital.

RQ2	As coleções de documentos utilizadas nas pesquisas de smishing são públicas ou construídas pelos próprios autores?	Avaliar a reprodutibilidade dos estudos de coleções de mensagens de smishing.
RQ3	Quais critérios de pré-processamento de mensagem são utilizados para os algoritmos de aprendizado de máquina?	Analisar os critérios de pré-processamento aplicados às coleções de mensagens de smishing.
RQ4	Quais algoritmos de aprendizado de máquina são mais utilizados na detecção de smishing?	Identificar quais algoritmos de aprendizado de máquina são empregados na detecção de smishing.

Fonte: elaborado pelos autores.

3.2 Busca

As buscas foram conduzidas manualmente em artigos científicos e anais de conferências publicados de janeiro de 2020 a junho de 2025, nas bases de periódicos IEEE Xplore e ACM Digital Library (Quadro 3).

Quadro 3: strings de busca (query strings).

ID	Query String
1	"smishing" AND "machine learning"
2	"smishing" AND "naive bayes"
3	"smishing" AND "artificial intelligence classification algorithms"
4	"smishing" AND "neural networks"
5	"datasets smishing classification"

Fonte: elaborado pelos autores.

3.3 Critérios de Inclusão

Com o objetivo de orientar o processo de seleção das pesquisas, foram estabelecidos critérios de inclusão que delimitam o escopo do estudo. Esses critérios contemplam o período de publicação, a metodologia de construção da coleção de mensagens, o idioma dos artigos e a pertinência temática (Quadro 4).

Quadro 4: critérios de inclusão.

ID	Critério de Inclusão
01	Artigos publicados entre janeiro de 2020 e julho de 2025
02	Demonstrar como foi realizada a construção da coleção de mensagens para a pesquisa
03	Idioma inglês ou português
04	Pesquisas sobre smishing

Fonte: elaborado pelos autores.

3.4 Critérios de Exclusão

Após a seleção dos estudos, foram aplicados os critérios de exclusão, com o objetivo de verificar se os artigos mantinham relação com smishing, se foram publicados após janeiro de 2020 e se apresentavam a descrição do processo de construção da coleção de mensagens. Os estudos que não atenderam a esses parâmetros foram excluídos da análise (Quadro 5).

Quadro 5: critérios de exclusão.

ID	Critérios de exclusão
01	Artigos que, após leitura do título e resumo, não apresentaram relação direta com o tema da pesquisa
02	Artigos publicados antes de 2020
03	Estudos que não demonstram a construção da coleção de mensagens

Fonte: elaborado pelos autores.

4. RESULTADOS

Os resultados evidenciaram que o smishing é um tema relevante no contexto da segurança digital, devido ao crescente uso de dispositivos móveis e à sofisticação das técnicas de engenharia social que visam ao roubo de dados e prejuízos financeiros. As pesquisas demonstram uma predominância no uso combinado de bases de dados públicas,

como Kaggle, Mendeley e UCI, juntamente com coleções próprias construídas pelos autores, a fim de atender necessidades específicas dos experimentos. Foi observada também uma padronização nas etapas de pré-processamento, com destaque para normalização textual (conversão para caracteres minúsculos e remoção de termos irrelevantes), exclusão de stop words, tokenização, stemming e lematização. Quanto aos algoritmos de aprendizado de máquina, os modelos mais utilizados foram Random Forest, Support Vector Machine (SVM) e Naive Bayes, alcançando elevados níveis de acurácia.

Os estudos selecionados contemplam tanto investigações empíricas quanto levantamentos bibliográficos, além de serem frequentemente referenciados em revisões sistemáticas de literatura relacionadas à engenharia de software, com ênfase em segurança digital e inteligência artificial (Quadro 6).

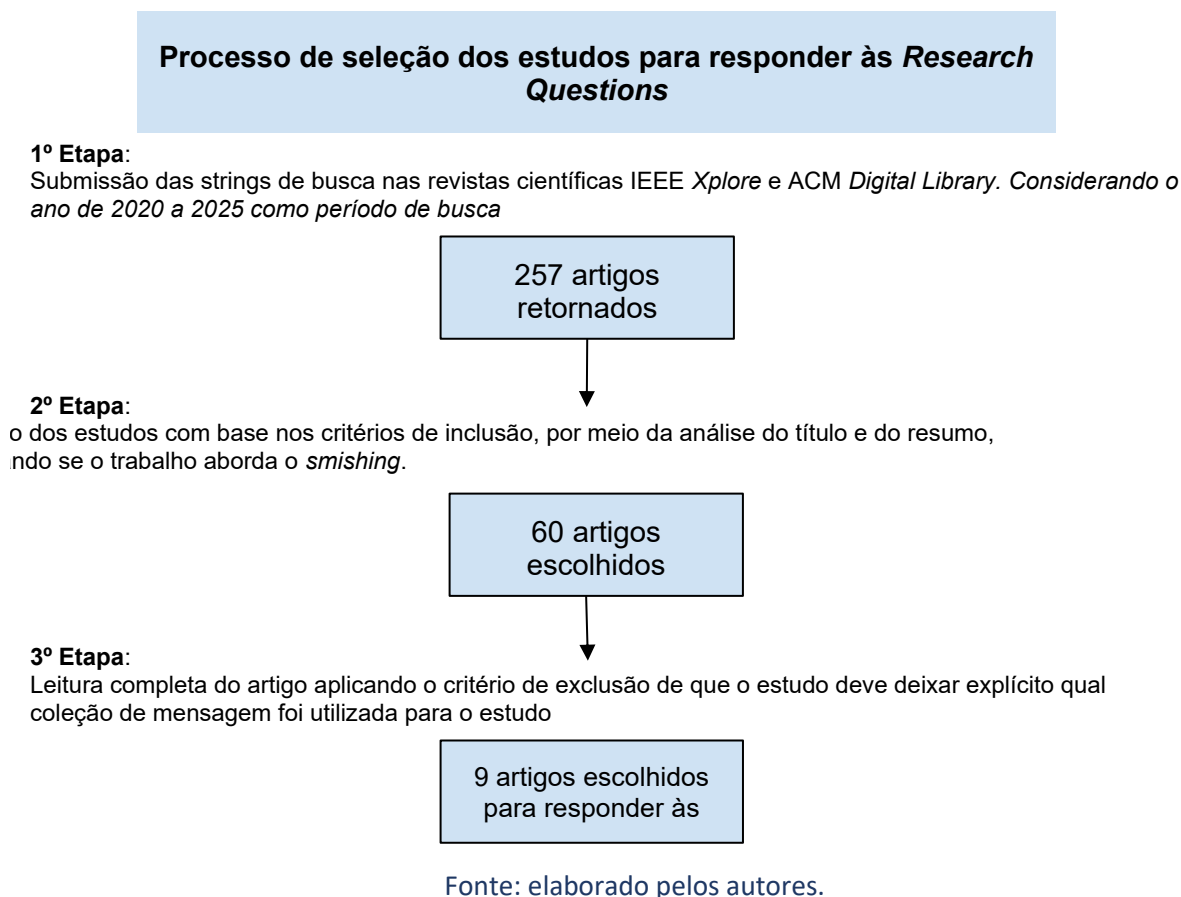
Quadro 6: artigos retornados para os strings de busca.

String de Busca	Revista/Base	Artigos
smishing and machine learning	IEEE Xplore	9
	ACM	65
smishing and naive bayes	IEEE Xplore	5
	ACM	27
smishing and artificial intelligence classification algorithms	IEEE Xplore	1
	ACM	35
smishing and neural networks	IEEE Xplore	15
	ACM	31
datasets smishing classification	IEEE Xplore	14
	ACM	55
Total		257

Fonte: elaborado pelos autores.

Após aplicar os critérios de inclusão e exclusão, foram escolhidos os estudos que responderiam às RQs. Dos 322 artigos encontrados, apenas 9 foram selecionados para responderem às RQs (Figura 1).

Figura 1: resultado da aplicação dos critérios de inclusão e exclusão utilizados para a construção das RQs.



RQ1 - Por que o *smishing* é considerado um tema relevante?

O *smishing* foi retratado como um tipo de golpe virtual no campo da segurança digital (Mishra; Soni; 2019). Esse golpe utiliza técnicas de engenharia social para induzir as vítimas a fornecer informações confidenciais, acessar links maliciosos ou realizar download de arquivos com malware (Samad et. al; 2024). Com a popularização do uso de dispositivos móveis, o *smishing* tem se tornado comum sobretudo pela falta de conhecimento sobre práticas de segurança digital e pela ausência de ferramentas que ajudem o usuário a analisar as mensagens que recebe no celular (Patra et. al; 2023). As consequências do *smishing* são o roubo de dados pessoais, perdas financeiras e comprometimento de

credenciais corporativas, afetando tanto indivíduos quanto organizações (Rai; 2025). Nesse contexto, o uso de algoritmos de aprendizado de máquina aplicados a coleções de mensagens SMS tem se destacado como uma forma de aprimorar a análise e a detecção de ameaças (Chen et. al; 2023). Essa aplicação não apenas impulsiona o avanço científico e tecnológico, como também viabiliza o desenvolvimento de soluções que aumentam a segurança do usuário.

RQ2 - As coleções de documentos utilizadas nas pesquisas de smishing são públicas ou construídas pelos próprios autores?

As pesquisas demonstram uma predominância no uso combinado de bases de dados públicas, como Kaggle, Mendeley e UCI. Entretanto, alguns trabalhos fundamentam-se em coleções próprias, construídas pelos autores a fim de atender necessidades específicas dos experimentos (Quadro 7).

Quadro 7: relação entre os artigos analisados e as coleções de mensagens utilizadas em suas pesquisas sobre smishing.

Artigo	Coleção de mensagens utilizada
Cybersmish: A Proactive Approach for Smishing Detection and Prevention using Machine Learning (Kumarasiri et. al; 2023)	Kaggle (2016), Kaggle (2022)
Using Machine Learning for Efficient Smishing Detection (Chen et al.; 2023)	Coleção de mensagens criada pelos autores
Ensemble Learning Approach in Tackling Smishing Messages (Febryan et al.; 2025)	Kaggle (2016) e coleção de mensagens criada pelos autores
SMS-Shield: A Lightweight Approach for Smishing Detection using Machine Learning (Samad et al.; 2024)	Kaggle (2022)
Enhancing Smishing Detection (Mehmood et al.; 2024)	Kaggle (2021) e Mendeley (2022)
SMSDect: A Prediction Model for Smishing Attack Detection using Machine Learning and Text Analysis	Kaggle e UCI, Jain e Gupta (2018 apud Patra et al., 2023)

Fonte: elaborado pelos autores.

O uso combinado de bases públicas, como as disponíveis no Kaggle, Mendeley e UCI, juntamente com conjuntos de mensagens próprios, constitui uma prática recorrente nas pesquisas sobre smishing. Em geral, os autores não se limitam a uma única base de dados, complementando fontes públicas ou privadas com conjuntos rotulados especificamente para atender aos objetivos e contextos de seus estudos.

RQ3 - Quais critérios de pré-processamento de mensagem são utilizados para os algoritmos de aprendizado de máquina?

Os critérios de pré-processamento nos estudos analisados seguem etapas que visam preparar as coleções de mensagem para o treinamento, validação e teste de algoritmos de aprendizado de máquina. A primeira etapa é de limpeza e normalização do texto, buscando padronizar o conteúdo e remover elementos que não contribuem para o processo de classificação (Febryan et al; 2025). Entre as ações recorrentes estão: (i) a conversão de todas as palavras para letras minúsculas e as remoções de (ii) pontuação, (iii) números, (iv) símbolos, (v) stop words e (vi) tags HTML (Chen et al., 2023; Kumarasiri et al., 2023; Mehmood et al., 2024). Em seguida, são aplicadas técnicas de (i) tokenização, (ii) lematização e (iii) stemming, que têm como finalidade segmentar o texto em unidades linguísticas menores (tokens) e reduzir as palavras às suas formas básicas, seja por meio da lematização ou do stemming (Mehmood et al., 2024) (Quadro 8).

Quadro 8: etapas de pré-processamento utilizadas nos trabalhos analisados.

Etapas de Pré-Processamento	Estudos encontrados
(i) Normalização de texto (ii) Palavras-chave suspeitas (iii) Identificação de links maliciosos ou encurtadores de links ou domínios incomuns (iv) Personagens de evasão - Golpistas burlam a detecção substituindo caracteres na mensagem por símbolos de aparência semelhante para burlar filtros de spam (v) Contagem de palavras em caixa alta (vi) Contagem de exclamações	Ensemble Learning Approach in Tackling Smishing Messages (Febryan et al; 2025)

(vii) Contagem de palavras de urgência (viii) Menção ao dinheiro - Golpistas induzem leitores a doar seu dinheiro a troca de algo como um sorteio	
Pré-processamento de mensagens: (i) Remoção de URLs, números e tags HTML (ii) Rotulação das mensagens (iii) Carregamento da biblioteca Transformer no Hugging Face (iv) Tokenização Pré-processamento de URL: (i) Remoção de nulos (Remover ou tratar registros onde a URL não existe ou está vazia) (ii) Mapeamento de rótulos (iii) Balanceamento de classes (iv) Tokenização (v) Definição de comprimento Analisador de Conteúdo HTML: (i) Remoção de valores nulos (Remover ou tratar registros onde o HTML não existe ou está vazio) (ii) Filtragem de idioma (iii) Conversão para letras minúsculas (iv) Tokenização (v) Filtragem de entradas apenas alfanuméricas (vi) Remoção de stopwords e pontuação (vii) Stemming	Cybersmish: A Proactive Approach for Smishing Detection and Prevention using Machine Learning (Kumarasiri et. al; 2023)
(i) Remoção de pontuação. (ii) Remoção de espaços extras e stopwords (iii) Aplicação da Frequência do Termo e a Frequência Inversa do Documento (TF-IDF) para calcular a importância dos termos no texto (iv) Uso de expressões regulares para extrair características de URLs seguindo o padrão World Wide Web Consortium (W3C)	Using Machine Learning for Efficient Smishing Detection (Chen et al; 2023)

(i) Eliminação de dados brutos para reduzir a complexidade do texto (ii) Uso do Natural Language Toolkit (NLTK) para o pré-processamento (iii) Remoção de pontuação e números por meio de expressões regulares (iv) Conversão do texto para minúsculas (v) Remoção de stopwords (vi) Stemming (vii) Aplicação de lematização com WordNet Lemmatizer	Enhancing Smishing Detection (Mehmood et. al; 2024)
(i) Conversão do texto para minúsculas (Padronizar o texto, reduzindo variações lexicais e melhorando a qualidade do processamento) (ii) Tokenização (iii) Remoção de stopwords (iv) Tratar caracteres especiais	SMS-Shield: A Lightweight Approach for Smishing Detection using Machine Learning (Samad et. al; 2024)
(i) Contração e expansão de palavras (ii) Limpeza de caracteres (iii) Conversão do texto para minúsculas (iv) Tokenização (v) Remoção de stopwords (vi) Lematização	A Machine Learning-based Approach for Detecting Smishing Attacks at End-user Level (Kohilan et al; 2023)

Fonte: elaborado pelos autores.

RQ4 - Quais algoritmos de aprendizado de máquina são mais utilizados na detecção de smishing?

Os algoritmos de aprendizado de máquina utilizados na detecção de smishing são modelos de classificação valorizados por eficácia e desempenho. Estudos demonstram que algoritmos como Random Forest, Support Vector Machine (SVM) e Naive Bayes são consistentemente aplicados e geralmente alcançam as melhores acurácias ao analisar as coleções de mensagens (Quadro 9).

Quadro 9: algoritmos empregados na detecção de smishing e a maior acurácia obtida nos trabalhos considerados.

Algoritmos utilizados no estudo	Acurácia	Estudo
Random Forest Regressão Logística Naive Bayes K-Nearest Neighbors (KNN) Support Vector Machine (SVM) XGBoost	O algoritmo Random Forest obteve a acurácia de 99%	SMSDect: A Prediction Model for Smishing Attack Detection using Machine Learning and Text Analysis (Patra et. al; 2023)
Support Vector Machine (SVM) Kernel Linear Support Vector Machine (SVM) Kernel Polinomial (SVM Poly) Support Vector Machine (SVM) Kernel Radial Basis Function (RBF) Regressão Logística	O algoritmo SVM com kernel linear obteve a acurácia de 97.30%	Smishing Detection Using Support Vector Machines (Rai; 2025)
Random Forest Support Vector Machine (SVM) XGBoost Regressão Logística Rede Neural Artificial (RNA)	O algoritmo Regressão Logística obteve a acurácia de 95.95%	Ensemble Learning Approach in Tackling Smishing Messages (Febryan et al; 2025)

Fonte: elaborado pelos autores.

5. DISCUSSÃO

Os resultados desta revisão permitiram identificar apenas nove estudos que atenderam aos critérios de inclusão estabelecidos. Embora esse número possa parecer reduzido, ele representa uma lacuna nas pesquisas sobre detecção de smishing. Um dos critérios adotados nesta revisão foi de que os estudos descrevessem como foi selecionada a coleção de mensagens utilizada no estudo. A baixa quantidade de trabalhos que atenderam a esse requisito demonstra a limitada disponibilidade de conjuntos de dados centralizados para pesquisas nessa área. Esse cenário é preocupante, pois a qualidade e a transparência dos dados utilizados influenciam diretamente a confiabilidade dos modelos de detecção desenvolvidos. A ausência de informações detalhadas sobre as coleções de mensagens dificulta a reprodução dos experimentos, a comparação entre abordagens

propostas e a validação dos resultados obtidos por diferentes pesquisadores. Consequentemente, o avanço científico da área torna-se mais lento e menos consistente. A relevância desse problema é ampliada pelo crescimento contínuo dos ataques de smishing, que exploram mensagens SMS para enganar vítimas e obter informações pessoais, financeiras e credenciais de acesso. O smishing utiliza técnicas de engenharia social que exploram vulnerabilidades humanas, tornando a conscientização dos usuários e o desenvolvimento de mecanismos automatizados de detecção elementos fundamentais para a mitigação desse tipo de ameaça. Nesse contexto, os resultados encontrados reforçam a necessidade de iniciativas voltadas à criação e manutenção de repositórios públicos de mensagens de smishing. A existência de bases padronizadas permitiria maior reprodutibilidade das pesquisas, facilitaria a comparação entre métodos e contribuiria para o desenvolvimento de soluções. Portanto, a quantidade de estudos incluídos nesta revisão deve ser interpretada como uma lacuna existente na literatura. Ao identificar essa deficiência e sistematizar as informações disponíveis sobre as coleções de mensagens utilizadas nas pesquisas, este estudo contribui para o entendimento do estado atual da área.

6. CONCLUSÃO

Esta revisão sistemática da literatura analisa as abordagens e os desafios em pesquisas sobre detecção de mensagens de smishing, considerando aspectos relacionados à engenharia social, à construção das bases de mensagens e à aplicação de algoritmos de aprendizado de máquina. A relevância do tema se justifica pelo aumento de golpes disseminados por SMS, os quais exploram fragilidades humanas, como a falta de conhecimento em segurança digital evidenciando a necessidade de estratégias na área de segurança digital. Os resultados indicam um uso combinado de bases públicas, como as disponíveis no Kaggle, Mendeley e UCI, juntamente com bases de mensagens privadas rotuladas especificamente para atender aos objetivos e contextos das pesquisas consideradas. Os resultados também demonstram um predomínio de algoritmos como Random Forest, SVM e Naive Bayes. Como contribuição, esta revisão apresenta um panorama das pesquisas sobre o combate ao smishing, destacando a necessidade de criação de um repositório de mensagens que possa ser avaliado, atualizado e utilizado pela

comunidade científica gratuitamente. Como trabalhos futuros, será criada uma coleção de documentos sobre mensagens de smishing para treinamento de algoritmos de inteligência artificial com o intuito de disponibilizá-la livremente para a comunidade científica. Também serão comparados algoritmos tradicionais de aprendizado de máquina, como Random Forest, Support Vector Machine (SVM) e Naive Bayes com algoritmos baseados em Large Language Model (LLM).

REFERÊNCIAS

ABDILLAH, R. et al. Phishing classification techniques: a systematic literature review. **IEEE Access**, v. 10, p. 41574-41591, 2022. Disponível em: <https://doi.org/10.1109/ACCESS.2022.3166474>. Acesso em: 01 dez. 2025.

ANDERSON, Hyrum S.; ROTH, Phil. **EMBER: Elastic Malware Benchmark for Empowering Researchers**. 2018. Disponível em: <https://arxiv.org/abs/1804.04637>. Acesso em: 18 mar. 2026.

ABUSE.ch. **MalwareBazaar**. 2023. Disponível em: <https://bazaar.abuse.ch>. Acesso em: 18 mar. 2026.

ALMOUSA, M. **Hyperlinks**. 2022. Disponível em: <https://www.kaggle.com/datasets/mayalmousa/afahyperlinks>. Acesso em: 18 mar. 2026.

BOUKARI, B. E et al, M. Machine learning detection for SMiShing frauds. In: **IEEE annual consumer communications & networking conference (CCNC)**, 18., 2021, Las Vegas, NV, USA. Anais... Las Vegas: IEEE, 2021. p. 1–2. Disponível em: <https://doi.org/10.1109/CCNC49032.2021.9369640>. Acesso em: 01 dez. 2025.

CHEN, C.-C. et al. Using large language models to detect smishing messages in telecom networks. In: **international symposium on computer, consumer and control (IS3C)**, 7., 2025, Taichung, Taiwan. Anais... Taichung: IEEE, 2025. p. 1–4. Disponível em: <https://doi.org/10.1109/IS3C65361.2025.11131109>. Acesso em: 01 dez. 2025.

CHEN, S.-S. et al. Using machine learning for efficient smishing detection. In: **international conference on consumer electronics - taiwan (icce-taiwan)**, 2023, PingTung, Taiwan. Anais... PingTung: IEEE, 2023. p. 207–208. Disponível em: <https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226901>. Acesso em: 01 dez. 2025.

CHITHAMBARAMANI, R. et al. SMS Guard: Android application to prevent smishing attacks. In: **international conference on intelligent cyber physical systems and internet of things (icoici)**, 2., 2024, Coimbatore, India. Anais Coimbatore: IEEE, 2024. p. 69–74. Disponível em: <https://doi.org/10.1109/ICoICI62503.2024.10696674>. Acesso em: 01 dez. 2025.

COSSA, O. F. et al. Prediction of bank frauds by SMS or voice, from cell phone data analysis: a systematic literature review. in: **iberian conference on information systems and technologies (cisti)**, 16., 2021, Chaves, Portugal. Anais [...]. Chaves, Portugal: CISTI, 2021. p. 1-8. Disponível em: <https://doi.org/10.23919/CISTI52073.2021.9476380>. Acesso em: 01 dez. 2025.

DOU, Z. et al. Systematization of knowledge (SoK): a systematic review of software-based web phishing detection. **IEEE Communications Surveys & Tutorials**, v. 19, n. 4, p. 2797-2819, 2017. Disponível em: <https://doi.org/10.1109/COMST.2017.2752087>. Acesso em: 01 dez. 2025.

FEBRYAN, J. et al. Ensemble learning approach in tackling smishing messages. In: **IEEE international conference on industry 4.0, artificial intelligence, and communications technology (iaict)**, 2025, Bali, Indonesia. Anais [...]. Bali, Indonesia: IEEE, 2025. p. 477-482. Disponível em: <https://doi.org/10.1109/IAICT65714.2025.11100814>. Acesso em: 01 dez. 2025.

FRANCO, Cristiano Mendes. Filtros de alta precisão e performance para mensagens de SMS fraudulentas, com base em redes neurais. 2024. **Trabalho de Conclusão de Curso (MBA)** – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024. Disponível em: <https://doi.org/10.11606/003225083>. Acesso em: 18 jul. 2026.

GABER, Matthew G. et al. Malware detection with artificial intelligence: a systematic literature review. **ACM Computing Surveys**, v. 56, n. 6, art. 148, jun. 2024. Disponível em: <https://doi.org/10.1145/3638552>. Acesso em: 18 jul. 2026.

HARANG, Richard E.; RUDD, Ethan M. **SoReL-20M**. 2022. Disponível em: <https://github.com/sophos/SOREL-20M>. Acesso em: 18 mar. 2026.

KADDOURA, Sanaa et al. A systematic literature review on spam content detection and classification. **PeerJ Computer Science**, v. 8, e830, 2022. Disponível em: <https://doi.org/10.7717/peerj-cs.830>. Acesso em: 18 jul. 2026.

KITCHENHAM, Barbara. **Procedures for performing systematic reviews**. Keele, UK: Keele University, 2004. Disponível em: <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>. Acesso em: 18 jul. 2026.

KITCHENHAM, Barbara et al. Systematic literature reviews in software engineering – a systematic literature review. **Information and Software Technology**, v. 51, n. 1, p. 7-15, 2009. Disponível em: <https://doi.org/10.1016/j.infsof.2008.09.009>. Acesso em: 18 jul. 2026.

KITCHENHAM, Barbara et al. Systematic literature reviews in software engineering – a tertiary study. **Information and Software Technology**, v. 52, n. 8, p. 792-805, 2010. Disponível em: KITCHENHAM, Barbara et al. Systematic literature reviews in software engineering – a tertiary study. **Information and Software Technology**, v. 52, n. 8, p. 792-805, 2010. Disponível em: <https://doi.org/10.1016/j.infsof.2010.03.006>. Acesso em: 18 jul. 2026.

KITCHENHAM, B. et al. Repeatability of systematic literature reviews. In: **annual conference on evaluation & assessment in software engineering (EASE)**, 15., 2011, Durham. Anais. Durham: IEEE, 2011. p. 46-55. Disponível em: <https://doi.org/10.1049/ic.2011.0006>. Acesso em: 18 jul. 2026.

KOHILAN, R. et al. A Machine Learning-based Approach for Detecting Smishing Attacks at End-user Level. In: **ieee international conference on e-business engineering (ICEBE)**, 2023,

Sydney, Australia. Anais [...]. Sydney, Australia: **IEEE**, 2023. p. 149-154. Disponível em: <https://doi.org/10.1109/ICEBE59045.2023.00042>. Acesso em: 18 jul. 2026.

KUMARASIRI, W. L. T. T. N. et al. Cybersmish: A proactive approach for smishing detection and prevention using machine learning. In: **international conference on i-smac (iot in social, mobile, analytics and cloud)**, 7., 2023, Kirtipur, Nepal. Anais [...]. Kirtipur: IEEE, 2023. p. 210-217. Disponível em: <https://doi.org/10.1109/I-SMAC58438.2023.10290228>. Acesso em: 18 jul. 2026.

KUMARASINGHE, P. et al. user behavior analysis in determining the vulnerable category of vishing and smishing. in: **international conference on advancements in computing (ICAC)**, 5., 2023, Colombo, Sri Lanka. Anais [...]. Colombo: IEEE, 2023. p. 35-40. Disponível em: <https://doi.org/10.1109/ICAC60630.2023.10417682>. Acesso em: 18 jul. 2026.

KYAW, Phyto Htet. et al. A Systematic Review of Deep Learning Techniques for Phishing Email Detection. **Electronics**, v. 13, n. 19, p. 3823, 2024. Disponível em: <https://doi.org/10.3390/electronics13193823>. Acesso em: 18 jul. 2026.

MEHMOOD, M. K. et al. Enhancing smishing detection: A deep learning approach for improved accuracy and reduced false positives. **IEEE Access**, v. 12, p. 137176-137193, 2024. Disponível em: <https://doi.org/10.1109/ACCESS.2024.3463871>. Acesso em: 18 jul. 2026.

MISHRA, Sandhya; SONI, Devpriya. **SMS phishing dataset for machine learning and pattern recognition**. 2022. Disponível em: <https://doi.org/10.17632/f45bkkt8pr.1>. Acesso em: 18 jul. 2026.

MISHRA, Sandhya; SONI, Devpriya. SMS phishing dataset for machine learning and pattern recognition. In: ABRAHAM, A. et al. (org.). **Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)**. Cham: Springer, 2023. v. 648. Disponível em: https://doi.org/10.1007/978-3-031-27524-1_57. Acesso em: 18 jul. 2026.

MISHRA, Sandhya; SONI, Devpriya. SMS phishing and mitigation approaches. In: **international conference on contemporary computing (IC3)**, 12., 2019, Noida, Índia. Anais [...]. Noida: IEEE, 2019. p. 1-5. Disponível em: <https://doi.org/10.1109/IC3.2019.8844920>. Acesso em: 18 jul. 2026.

PATRA, C. et al. SMSDelect: A prediction model for smishing attack detection using machine learning and text analysis. in: **ieee global communications conference (globecom)**, 2023, Kuala Lumpur, Malásia. Anais [...]. Kuala Lumpur: IEEE, 2023. p. 3837-3842. Disponível em: <https://doi.org/10.1109/GLOBECOM54140.2023.10437712>. Acesso em: 1 dez. 2025.

RAI, P. Smishing detection using support vector machines. in: **international conference on intelligent control, computing and communications (IC3)**, 2025, Mathura, Índia. Anais [...]. Mathura: IEEE, 2025. p. 253-258. Disponível em: <https://doi.org/10.1109/IC363308.2025.10956699>. Acesso em: 1 dez. 2025.

SAMAD, S. R. et al. SMS-Shield: a lightweight approach for smishing detection using machine learning. In: **international conference on innovative engineering sciences and technological research (iciestr)**, 1., 2024, Muscat, Oman. Anais [...]. Muscat, Oman:

ICIESTR, 2024. p. 1-6. Disponível em: <https://doi.org/10.1109/ICIESTR60916.2024.10798213>. Acesso em: 15 nov. 2026.

UCI MACHINE LEARNING. **SMS Spam Collection Dataset**. 2016. Disponível em: <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>. Acesso em: 18 mar. 2026.

VirusShare. **Because sharing is caring**. 2023. Disponível em: <https://virusshare.com/> . Acesso em: 18 mar. 2026.

WANGCHUK, T.; GONSALVES, T. Multimodal Phishing Detection on Social Networking Sites: A Systematic Review. **IEEE Access**, v. 13, p. 103405-103416, 2025. Disponível em: <https://doi.org/10.1109/ACCESS.2025.3579584>. Acesso em: 1 nov. 2025.

XU, Hanxiang et al. Large Language Models for Cyber Security: A Systematic Literature Review. [S. l.]: **Huazhong University of Science and Technology**; Hamad Bin Khalifa University; Nanyang Technological University, 2025. Disponível em: <https://dl.acm.org/doi/10.1145/3769676>. Acesso em: 1 dez. 2025.

YANG, Limin et al. **BODMAS Malware Dataset**. 2022. Disponível em: <https://whyisyoung.github.io/BODMAS/>. Acesso em: 18 mar. 2026.

YERIMA, S. Y.; BASHAR, A. Semi-supervised novelty detection with one class SVM for SMS spam detection. In: **international conference on systems, signals and image processing (IWSSIP)**, 29., 2022, Sofia, Bulgária. Anais [...]. Sofia: IEEE, 2022. p. 1-4. Disponível em: <https://doi.org/10.1109/IWSSIP55020.2022.9854496>. Acesso em: 1 dez. 2025.