GolpeBR: Construction and Validation of an Annotated Dataset on Banking Scams and Fraud

Tamyres Vial de Souza[1], Jhonata Tirloni[1], Felipe Belo[1], Nelcileno Virgilio Araujo[1], Thiago M. Ventura[1], Allan Gonçalves de Oliveira[1]

[1]Instituto de Computação – UFMT – CEP 78.060-719 – Cuiabá – MT – Brasil

{tamyres.souza1, jhonata.tirloni, felipe.belo}@sou.ufmt.br,
{nelcileno, thiago, allan}@ic.ufmt.br

Abstract. *This article details the construction and validation of the GolpeBR dataset, which was created from news articles and Reddit posts. Automated Python methods were used to extract and process the data, which was then annotated using the Deepseek-R1 LLM and the 5W1H methodology. A cybersecurity expert classified and validated the records, distinguishing between banking cybercrimes and unrelated crimes. For the dataset's validation, supervised learning algorithms were applied. It was found that models trained with data structured by the 5W1H methodology demonstrated better accuracy, reaching 0.83 for the Logistic Regression and Random Forest algorithms.*

1. Introduction

Cybercrime refers to illicit activities executed via the digital environment, where digital devices or information systems may function as the tool, the target, or both in the commission of these crimes [Sabillon et al. 2016]. Digital crime encompasses different types of illicit activities for profit, ranging from downloading illegal records to stealing money from online bank accounts [Deora and Chudasama 2021]. This type of crime is a global problem that cannot be easily combated with national laws and requires joint action by countries to combat it, since, in most cases, cybercrime goes beyond national jurisdictions [Sabillon et al. 2016].

Cyber fraud is a type of cybercrime in which victims are deceived through the use of digital technologies and involves different types of scams and frauds [Sabillon et al. 2016]. In the banking context, fraudsters have shifted their focus from account acquisition to scams in which customers are exploited as the weakest link [KPMG 2019]. This type of fraud is especially concerning because, in addition to financial losses, there can be long-term psychological consequences, and victims of cybercrimes may experience significant psychological distress [Monteith et al. 2021]. Social engineering is a strategy in which cybercriminals exploit the human factor, using psychological tactics. This approach is applied to authorized payments, in which a customer is coerced into transferring their money to an account controlled by the fraudster. This type of scam is still one of the main challenges for the banking sector in fraud prevention, since, most of the time, the customer is accessing their account, and access controls do not detect the fraud [Silva and Vieira 2021; KPMG 2019].

Given that the focus of most digital crimes is on human vulnerabilities, it is essential to understand how individuals routinely interact with technology, as well as the social and cultural factors involved [Monteith et al. 2021]. In Brazil, between September

2024 and March 2025, the number of victims of scams or attempted scams increased from 33% to 38%. The most common scams reported by Brazilians include credit card cloning or card swapping, fraud in which someone impersonates an acquaintance requesting money via WhatsApp, the "fake call center" scam, and the PIX scam [Febraban 2025].

Different computational strategies, such as statistical methods, machine learning, and neural networks, have been explored for the detection and prevention of cybercrimes. These approaches, especially those based on machine learning and artificial intelligence, use datasets as a knowledge base and for training models. However, there is a shortage of reference datasets for detecting cybercrimes, and this is due to the lack of cooperation between law enforcement authorities and researchers in data collection. In addition, the diversity of cybercrimes also contributes to the difficulty of creating comprehensive datasets, as crimes can occur on different platforms and involve different types of data [Al-Khater et al. 2020].

Another challenge in building a comprehensive dataset is the labeling of information, as this process can be complex and time-consuming, often requiring the collaboration of subject matter experts and involving the identification and confirmation of illicit activities [Nicholls, Kuppa and Le-Khac, 2021]. Given this scenario, the present work focuses on cybercrime characterized as banking scams or fraud that victimizes customers of the digital banking system in Brazil. The study describes the construction of the annotated dataset GolpeBR, based on reports of banking scams and fraud. It is hoped that the dataset can contribute to the advancement of academic research aimed at applying computational strategies in the prevention of cybercrime in Brazil.

2. Related Work

The detection and combating of cybercrime through computational learning techniques has been widely explored in different contexts. [Gumma and Peram 2024] conclude that both traditional machine learning (ML) algorithms and advanced deep learning (DL) models are essential to face the dynamic cybersecurity scenario. DL approaches are particularly effective in dealing with unstructured data, providing insights into complex cyber threats. [Dilek et al. 2015] review the application of artificial intelligence (AI) techniques in combating digital crimes, highlighting the increasing vulnerability of cyber infrastructures and the need for more sophisticated defense systems. In this context, AI techniques play an important role in the detection and prevention of cyber attacks, with the potential to improve security in cyberspace.

In [Al-Khater et al. 2020] is presented a comprehensive review of cybercrime detection techniques, exploring different types of cybercrimes, their consequences, and the existing methodologies for their detection and prevention. The authors discuss the weaknesses and strengths of each technique, in addition to providing recommendations to improve the effectiveness of cybercrime detection models. In the banking sector, [Gyamfi and Abdulai 2018] propose the detection of credit card fraud and money laundering using supervised machine learning. [Sarma et al. 2020] address the detection of banking fraud by modeling customer data from graphs. [Balasankula et al. 2024] focus on the detection of fraudulent activities in banking transactions, employing ML and AI to detect abnormal behaviors or suspicious patterns, preventing financial losses and protecting customers.

Large Language Models (LLMs) are used for the detection of text-based cybercrimes - cyberbullying, hate speech, and cyberterrorism. [Ullah et al. 2024] propose the creation of an annotated dataset and the application of pre-trained language models for the classification of criminal texts. [Mallmann, Xavier and Santin 2018] also apply ML to identify text-based cybercrimes. In [Manna, Al-Fayoumi and Al-Fawa'reh 2024], the Bidirectional Encoder Representations from Transformers (BERT) model and natural language processing (NLP) techniques are used for the identification of these crimes. [Plath et al. 2022] address the detection of hate speech against women in texts in Brazilian Portuguese, and the methodology includes the construction of a dataset. Still in the Brazilian context, [Barros, Silva, and Miranda 2020] address the detection of phishing - a cybercrime of data theft by deception - using a rule-based expert system that analyzes the attributes extracted from web pages to identify malicious pages.

The MINA-BR database [Plath et al. 2022] was created from comments automatically extracted from online sources, mainly Twitter and YouTube, using keyword search and profiles of potential victims. The final version of the database contained 6,002 comments, part of these comments were labeled by volunteers, who classified them as offensive or not, and, when offensive, indicated whether they contained hate speech against women, along with a level of confidence. In this study, two methods were used to transform the texts into numerical data: Bag of Words (BoW), which counts the words, and Terem Frequency-Inverse Document Frequency (TF-IDF), which weights the words based on their frequency and relative importance in the dataset. The Removedor de Sufixos da Lingua Portuguesa (RSLP) Stemmer algorithm was applied in some configurations for the extraction of word stems. Finally, classic classification algorithms were evaluated, and the best F1-Score obtained was 0.57 for the Support Vector Machines (SVM) algorithm with TF-IDF representation and without stemming.

Another relevant study, by [Carnaz, Antunes and Nogueira 2021], proposes the creation of an annotated and anonymized corpus of documents in European Portuguese (PT-pt), related to drug trafficking. The dataset served as a resource for the evaluation and comparison of ML and NLP methods and tools in tasks such as sentence detection, named entity recognition, and identification of criminal terms in documents in PT-pt. The corpus was compiled from the extraction of data from police news websites and restricted access reports, using a crawler software to process the various formats available, being later annotated manually. At the end, the corpus contained 163 records, and stratified 10-fold cross-validation was applied, with averages of Precision of 0.784, Recall of 0.768 and F1-Score of 0.771.

The cited studies employ different strategies for detecting cybercrimes in various contexts. While related works focus on other types of cybercrimes, such as hate speech [Plath et al. 2022], trafficking [Carnaz, Antunes and Nogueira 2021], technical phishing [Barros, Silva, and Miranda 2020], or international crimes [Ullah et al. 2024], GolpeBR is a textual dataset built from reports of banking cybercrimes in Brazil. It specifically addresses scams and fraud targeting bank customers, a topic that has not been explored in previous studies.

3. Materials and Methods

The first step for the construction of the GolpeBR dataset was an exploratory study of the sources on the Internet for the extraction of reports of banking cybercrime. In this stage,

the different sources were analyzed regarding the possibility of extracting the information in an automated way and the recognition of the platform as a vehicle for sharing reports of victims of fraud and banking scams.

One of the strategies used to identify textual cybercrimes is the search for keywords and the monitoring of profiles and channels that may be relevant to the research [Plath et al. 2022]. Another strategy is the use of reports in news, bulletins and police reports [Carnaz, Antunes and Nogueira 2021]. However, the isolated use of this last approach may limit the amplitude of the extraction of the dataset. For this reason, the following online sources were selected for data extraction for the composition of the GolpeBR dataset: publications and comments from the online forum Reddit and online news websites from Brave Browser.

On Reddit, the extraction was carried out from the r/golpe subreddit, filtering posts related to the year 2025. For the news indexer search, search terms were defined based on the five most frequent types of scams in Brazil, according to a survey by [Febraban 2025].

For the extraction, normalization, and storage of the contents of news articles and subreddit posts, the libraries Requests, BeautifulSoup, Pandas, and Selenium were used. These tools enabled the automation of data collection processes, navigation through the HTML structure, processing of the extracted data, and organization of the information into dataframes.

The annotation process was performed using the generative artificial intelligence platform Deepseek [Yang et al. 2024]. For this process, a prompt was developed based on the 5W1H methodology (Who, What, Where, When, Why, and How), to extract factual and semantic information from the texts, answering the basic questions about cybercrime [Carnaz, Antunes and Nogueira 2021].

After the annotation process, the data were manually classified by a cybersecurity expert. This step aimed to identify whether the extracted data constituted a report of a banking scam or fraud, categorizing the texts as either a crime that occurred or a crime that did not occur. In cases where a crime occurred, it was indicated whether the crime corresponded to a banking cybercrime or other types of cybercrime.

Finally, for the dataset validation, the Random Forest Classifier, Logistic Regression, and Support Vector Machine (SVM) training algorithms were used, applying hyperparameter tuning and cross-validation with $k = 5$. Additionally, the validation process was executed with two textual formats for the cybercrime description: the first, using the text field, which contains the complete crime reports; and the second, using information annotated through the 5W1H methodology. The evaluation considered the models' accuracy metric, aiming to identify the models' ability to correctly classify cybercrime reports. The best results were obtained with the 5W1H-based approach, especially with the SVM model.

4. Results and Discussion

This section details the construction process of the annotated GolpeBR dataset. The first subsection covers the data extraction process, which was automated using Python libraries. The second subsection addresses the data annotation process using an LLM and the 5W1H methodology, as well as classification by a cybersecurity expert. Finally, the

last subsection presents the dataset validation with the Random Forest Classifier, Logistic Regression, and Support Vector Machine (SVC) algorithms.

4.1. Data Extraction

The first step in building the dataset was conducting an exploratory study of online sources to extract reports of banking fraud and scams. In this stage, the sources were evaluated for the possibility of automated extraction of reports and for their recognition as platforms for sharing information about banking cybercrime. With this in mind, and aiming for diversity of information in the dataset, news articles about banking scams and frauds were selected, as well as the subreddit r/golpes, which is a community within the Reddit forum where users share and discuss content on specific topics.

According to the [Febraban 2025], the most common scams reported by Brazilians include: credit card cloning or card swapping; fraud in which someone impersonates an acquaintance requesting money via WhatsApp; the “central falsa” scam; and the PIX scam. Based on this data, the following search terms were defined for the news articles: “golpe da troca de cartão”, “golpe do cartão clonado”, “golpe do Whatsapp”, “golpe de falsa central”, “golpe do pix” and “golpe do CPF”. For the news search, the Brave browser’s news indexer was used, as its ad-blocking feature was considered an advantage for the search and extraction processes, which were automated using Python libraries. Furthermore, a country search filter was applied, ensuring that results were exclusively from Brazilian news portals.

In the first stage of news extraction, the Requests and BeautifulSoup libraries were used to make GET requests from the links generated with the search terms, navigate the HTML document structure containing the search results, and extract the URL of each listed news article. In total, 890 hyperlinks were extracted, which were normalized and stored in a dataframe using the Pandas library.

In the second stage of the process, the links stored in the dataframe were exported to a CSV file, and 16 domains from the portals that made up the search results were mapped. At this stage, one news article from each portal was manually opened in the browser to inspect the HTML structure and develop a crawler script capable of extracting content from the different portals. Finally, the Selenium framework was used to automate interaction with the news pages and capture the content of each hyperlink. After removing duplicate hyperlinks and private or paid content, 484 news articles were extracted, which were also stored in a Pandas dataframe. Figure 1 illustrates the data extraction process for the news articles.

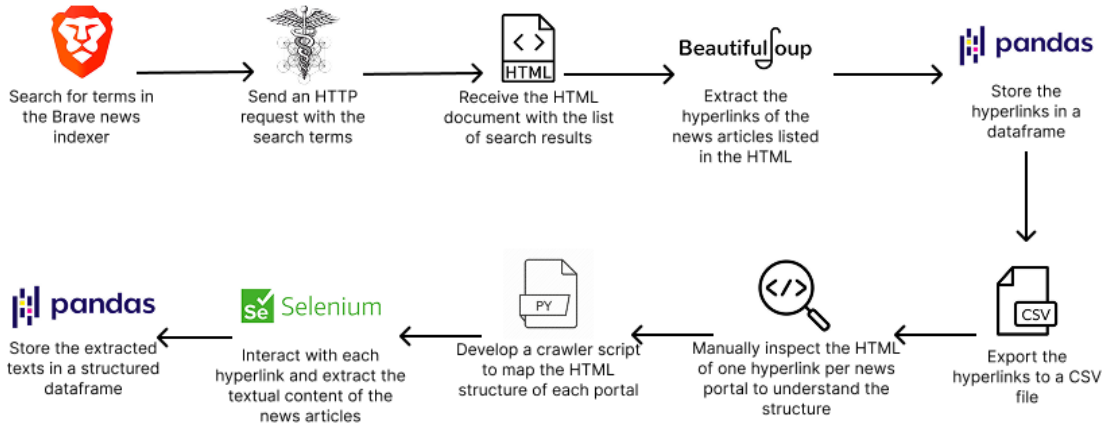

Figure 1. Process of Data Collection and Extraction from Cybercrime News

The process of extracting data from Reddit, in JSON format, was simpler, requiring only the use of the Requests and Pandas libraries. In total, 525 posts were extracted from the *r/golpe* subreddit, and after filtering for textual content only, 373 posts were included in the dataset for the annotation and classification stage.

At the end of the extraction process, 557 textual records were included in the annotation stage.

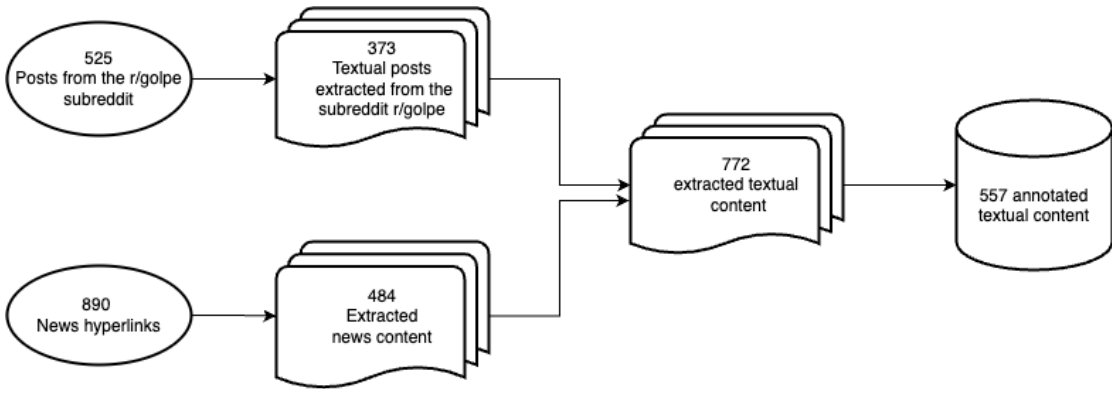

Figure 2. Composition of the dataset for annotation and classification.

4.2. Annotation and Classification of the Data

The data annotation process was carried out using the Deepseek-R1 LLM, which can process extremely long contexts while maintaining robust performance and consistent results across all context window lengths, up to 128,000 tokens. Additionally, the version used is a model specialized in long Chain-of-Thought (CoT) reasoning, which was considered an advantage for processing long texts [Yang et al. 2024], as is the case with the extracted news articles. In this process, a prompt written in Portuguese was used (see Box 1), since the dataset is in the same language, applying the 5W1H method to extract semantic information from the textual content collected from the news articles and the subreddit.

Você receberá entradas no seguinte formato, com colunas separadas por pipe (|):
index|text|flg_crime|cat_crime
Cada linha representa um relato de um possível cibercrime bancário.

Tarefa:
Extraia as informações do campo text utilizando a metodologia 5W1H, conforme as instruções abaixo.

Perguntas (5W1H):
Quem (who): Quem foi a vítima? Quem foi o golpista?
O quê (what): O que aconteceu? Tipo de golpe e prejuízo, se houver.
Quando (when): Quando o golpe ocorreu?
Onde (where): Onde o golpe ocorreu (meio ou canal usado)?
Por quê (why): Por que o golpe funcionou? Qual vulnerabilidade foi explorada?
Como (how): Como o golpe foi executado? Etapas ou método.

Formato de saída esperado:
Para cada linha, gere exatamente uma nova linha com os campos separados por pipe (|) *sem espaços*:
<index>|quem|o_que|quando|onde|por_que|como

Regras obrigatórias:
- Use apenas o campo text como fonte das informações.
- Mantenha o campo index original no início da linha.
- Se alguma informação não estiver claramente presente, deixe o campo vazio, mas mantenha a estrutura com "|"
- Não adicione comentários, títulos ou cabeçalhos.
- Não invente informações.
- Mantenha a ordem original das entradas.

Box 1. *Prompt* for annotation 5W1H.

According [Carnaz, Antunes, and Nogueira 2021], the 5W1H methodology, often used in journalism, is also applied in criminal investigations to analyze facts and identify perpetrators. This approach enables the creation of a structured dataset, which can be used to train supervised learning algorithms. In this way, the data was annotated to answer the following questions: Who was involved? What happened? When did it happen? Where did it happen? Why did it happen? How did it happen?

After this process, the annotated data was manually classified by a cybersecurity expert. At this stage, the texts were classified as either a cybercrime that occurred or did not occur, and if so, it was indicated whether the crime was a banking cybercrime or not. Banking cybercrime was defined as digital scams and fraud targeting customers of financial institutions, with the aim of stealing money. Other cybercrimes refer to any type of digital crime that does not involve scams or fraud against bank customers [Sabillon et al. 2016].

There are various types of biases that can arise in the human data annotation process, with cultural and social biases being among the main types. Annotation bias occurs when the labels assigned by the annotator are systematically skewed [Chen and Joo 2021]. One strategy for dealing with biases suggests that instead of focusing solely on bias removal, the bias should be managed by making it transparent and giving the user a choice of how to adjust it [Demartini, Roitero and Mizzaro 2021]. Considering this, a search was performed in the database for terms related to race, gender, and sexuality, with the objective of detecting any type of bias in the annotation process. Only ten records were identified with mentions of gender, used to characterize the victim or the perpetrator of the crime as a woman or a man. Therefore, no gender-based annotation bias was found.

Table 2 presents examples of data annotated by the LLM and classified by the expert.

	5W1H annotation columns	**Classification columns**
Source	**Index | who | what | when | where | why | how**	**is_crime | type_cybercrime**
Subreddit r/golpe	2|vítima: autor do relato|ligações suspeitas com desconexão automática|recentemente|smartphone (chamadas)|possível exploração de vulnerabilidade em sistema|ligações desconectadas automaticamente	| Y | other cybercrimes
News	397|usuários do Pix no Brasil, Banco Central|novas medidas de segurança para combater golpes e sequestros relâmpago|anunciadas nesta sexta-feira (sem data específica)|transações via Pix e aplicativos bancários|exploração de vulnerabilidades em pagamentos instantâneos|limitação de horários e valores, além de autenticação reforçada para transações	| Y | banking cybercrime
Subreddit r/golpe	13|vítima: autor do relato|produto defeituoso (cooler)|após entrega|site de compras|dificuldade de comprovar defeito|produto não funcionou ao chegar	| N |

Table 2. Annotation and classification.

At the end of the process, 557 records were annotated using the 5W1H methodology. Of these, 90 records were not annotated and were excluded from the classification process because they did not contain sufficient information. Therefore, 467 records were annotated using the LLM and classified by a cybersecurity specialist, with 118 classified as crime not occurring and 349 as crime occurred. Among the records classified as crime, 232 were categorized as other cybercrimes and 117 as banking cybercrime.

4.3. Dataset validation

The authors [Minastireanu and Mesniță 2019] review the most commonly used machine learning algorithms for online fraud detection and highlight that supervised learning algorithms are frequently mentioned and show better accuracy rates in detecting this type of fraud. For this reason, for the validation of the GolpeBR dataset, supervised learning algorithms Random Forest Classifier, Logistic Regression, and SVM were used, with hyperparameter tuning via the GridSearchCV algorithm and cross-validation with K = 5.

For training, records containing the dependent variable categoria_cibercrime with a value equal to "cibercrime bancário" or "outros cibercrimes" were selected, subsequently transformed into numerical values, with 1 for banking cybercrime and 0 for other cybercrimes. To validate the best classification strategy, the models were trained with two different formats of the extracted data, as presented in Table 3. In the first format, the "text" field was used, which corresponds to the full text extracted from the news or subreddit. In the second format, training was performed with the concatenation of fields annotated using the 5W1H method.

The texts used in training were vectorized with the TF-IDF method, which values more relevant words in a document and less frequent ones in the corpus. Thus, 70% of the records were used for training and 30% for testing.

Classifier	Accuracy with the original full text	Accuracy with 5W1H annotation
Random Forest Classifier	0.72	0.83
Logistic Regression	0.69	0.83
SVM	0.76	0.81

Table 3. Accuracy of each algorithm for each text format.

Despite the simple pipeline, the models were able to identify patterns in the texts classified as cybercrimes and perform the classification appropriately. Furthermore, the models trained with data structured by the 5W1H methodology showed better performance in terms of accuracy, especially the Logistic Regression and Random Forest Classifier algorithms, which achieved an accuracy of 0.83.

The dataset and scripts used in the construction of the GolpeBR dataset are available at https://github.com/Jhonata-Tirloni/golpebr_dataset.

5. Final Considerations

Different computational strategies that use datasets as a knowledge base and for model training have been explored for the detection and prevention of cybercrimes. However, [Al-Khater et al. 2020] highlight the lack of reference datasets for cybercrime detection, since data collection is a complex task, and the diversity of cybercrimes also contributes to the difficulty of creating comprehensive datasets [Al-Khater et al. 2020].

Given this challenge, this work presented the process of constructing the GolpeBR database, which was annotated using an LLM and classified by a cybersecurity expert. The dataset contains 309 records, equally distributed among the non-crime, other cybercrime, and banking cybercrime classes.

The annotation process, using the 5W1H methodology, made it possible to extract factual and semantic information from the textual reports, answering the basic questions about cybercrime. This approach enabled the creation of a structured dataset, which can be used to train supervised learning algorithms for the prevention of banking cybercrime in Brazil. Although the use of LLM made the annotation process more economical and efficient, it is recognized that there are risks of introducing bias when compared to human annotation. For this reason, the final classification of cybercrimes was carried out by a cybersecurity expert, increasing the quality and reliability of the dataset.

In the process of data validation with supervised learning algorithms, the annotated data showed better accuracy compared to models trained with the "text" column corpus, especially for the Logistic Regression and Random Forest Classifier algorithms, which achieved an accuracy of 0.83. The results obtained using the "text" field indicate that the unannotated text demands a more robust data treatment pipeline, including steps like stopword removal.

Future work can expand the dataset size by automating the data extraction process, especially the search for terms related to banking cybercrime. These efforts could also involve more specialists in the data annotation process and perform peer validation.

Considering that one of the challenges in preventing banking cybercrimes is the difficulty for banks to identify when a user is making a transaction under duress or has been misled [Silva and Vieira 2021; KPMG 2019], the constructed GolpeBR dataset can be used to train AI agents that, through a conversational interface (WhatsApp, DM), assist potential victims of banking scams in identifying a risky situation. Furthermore, the results of this study can serve as a resource for future research to explore computational strategies aimed at preventing banking fraud and scams in Brazil.

References

Al-Khater, W. A., Al-Maadeed S., Ahmed, A. A., Sadiq, A. S. and Khan, M. K. (2020) "Comprehensive Review of Cybercrime Detection Techniques". In *IEEE Access*, v. 8, p. 137293-137311, doi: 10.1109/ACCESS.2020.3011259.

Barros, M., Silva, C., e Miranda, P. (2020) “Xphide: Um Sistema Especialista para a Detecção de Phishing”. In *Anais do Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)*

Balasankula, U. R., Poojitha, B., Chekurtha, S., Buyya, K., Bala, H. and Rao, P. V. (2024) "Banking Fraud Detection Using Machine Learning Algorithms," In *5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2024, pp. 1228-1233, doi: 10.1109/ICESC60852.2024.10689731.

Carnaz, G., Antunes, M., and Nogueira, V. B. (2021). "An Annotated Corpus of Crime-Related Portuguese Documents for NLP and Machine Learning Processing". In *Data,* doi: https://doi.org/10.3390/data6070071.

Chen, Y. and Joo, J. (2021). “Understanding and mitigating annotation bias in facial expression recognition”. doi: https://doi.org/10.48550/arXiv.2108.08504

Deora, R. S. e Chudasama D. (2021) “Brief Study of Cybercrime on an Internet”. In *Journal of Communication Engineering & Systems*. p. 1-6.

Demartini, G., Roitero, K. and Mizzaro, S. (2021). “Managing bias in human-annotated data: Moving beyond bias removal”. doi: https://doi.org/10.48550/arXiv.2110.13504

Dilek, S., Çakır, H., and Aydın, M. (2015) "Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review”. In *International Journal of Artificial Intelligence & Applications (IJAIA)*, doi: 1502.03552. https://doi.org/10.5121/ijaia.2015.6102.

Febraban – Federação Brasileira de Bancos (2025) “Radar Febraban: Março 2025”. Disponível em: https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/Relatório_Radar%20Febraban_Março_vf.pdf, June.

Gumma, Y. R. and Peram, S. (2024) "Review of Cybercrime Detection Approaches using Machine Learning and Deep Learning Techniques". In *International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, pp. 772-779, doi: 10.1109/ICAAIC60222.2024.10575058.

Gyamfi, N. K.; Abdulai, J.D. (2018) "Bank Fraud Detection Using Support Vector Machine," In *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, pp. 37-41, doi: 10.1109/IEMCON.2018.8614994.

KPMG. (2019) “Pesquisa Global sobre Fraude Bancária. A ameaça multifacetada da fraude: Os bancos estão prontos para enfrentar este desafio?”. Disponível em: https://assets.kpmg.com/content/dam/kpmg/br/pdf/2021/07/pesquisa-global-fraude-bancaria.pdf, June.

Mallmann, J., Xavier, A. dos S., e Santin, A. O. (2018) Detecção de Cibercrime em Redes Sociais: Machine Learning. In *The Tenth International Conference On Forensic Computer Science And Cyber Law*. São Paulo, 2018. p. 44-49, doi: http://dx.doi.org/10.5769/C2018005.

Manna, A., Al-Fayoumi, M. e Al-Fawa'reh, M. (2024) "Detecting Text-Based Cybercrimes Using BERT" In *International Jordanian Cybersecurity Conference (IJCC)*, Amman, Jordan, pp. 111-117, doi: 10.1109/IJCC64742.2024.10847273.

Minastireanu, E.-A., & Mesniță, G. (2019). *An analysis of the most used machine learning algorithms for online fraud detection*. Informatica Economica, 23(1), 5–16. https://doi.org/10.12948/issn14531305/23.1.2019.01.

Monteith, S., Bauer, M., Alda, M., Geddes, J., Whybrow, P. C., and Glen, T. (2021). “Increasing Cybercrime Since the Pandemic: Concerns for Psychiatry”. In *Current Psychiatry Reports 23*, doi: https://doi.org/10.1007/s11920-021-01228-w.

Nicholls, J., Kuppa, A. and Le-Khac, N. A. (2021) "Financial Cybercrime: A Comprehensive Survey of Deep Learning Approaches to Tackle the Evolving Financial Crime Landscape". In *IEEE Access*, v. 9, p. 163965-163986, doi: 10.1109/ACCESS.2021.3134076

Plath, H. O., Paiva, M. E. O., Pinto, D. L. e Costa, P. D. P. (2022). “Detecção de Discurso de Ódio Contra Mulheres em Textos em Português Brasileiro: Construção da Base MINA-BR e Modelo de Classificação”. In *Revista Eletrônica De Iniciação Científica Em Computação*, 20(3).

Sabillon R.; Cavaller V.; Cano J. e Serra-Ruiz J. (2016) "Cybercriminals, cyberattacks and cybercrime". In *IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)*, Vancouver, Canada, p. 1-9, doi: 10.1109/ICCCF.2016.7740434.

Sarma, D., Alam, W., Saha, I., Alam, M. N., Alam, M. J., and Hossain, S. (2020) "Bank Fraud Detection using Community Detection Algorithm," In *Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, p. 642-646, doi: 10.1109/ICIRCA48905.2020.9182954.

Silva, R. L. and Vieira, A. (2021) “Segurança cibernética: o cenário dos crimes virtuais no Brasil”. In *Revista Científica Multidisciplinar Núcleo do Conhecimento* Ano 06, ed. 04, v. 07, p. 134-149, doi: 0.32749/nucleodoconhecimento.com.br/ciencia-da-computacao/crimes-virtuais.

Ullah, F., Faheem, A., Azam, U., Ayub, M. S., Kamiran, F. and Karim, A. (2024). "Detecting Cybercrimes in Accordance with Pakistani Law: Dataset and Evaluation Using PLMs". In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 4717–4728, Torino, Italia. ELRA and ICCL

Yang, Q., Zhang, C., Azenkot, S., Bigham, J. P., Dontcheva, M., Fourney, A., Ju, W., Lee, J., Liao, Q., Lim, B. Y., Nebeling, M., Teevan, J., Wigdor, D., Zhu, J., …. Pan, Z. (2024) "The Future of Human-AI Interaction: A Research Agenda." arXiv preprint, arXiv:2412.19437. https://doi.org/10.48550/arXiv.2412.19437.