STM JOURNALS

E-Commerce for Future & Trends

https://journals.stmjournals.com/ecft

ISSN: 2454-9347
Volume 11, Issue 3, 2024
September–December
DOI (Journal): 10.37591/ECFT

Review ECFT

IntelliGaurd WebScan: Uncovering Dark Patterns on E-Commerce Websites

Yashasvi[1], Vineet Kumar Kankerwal[1], Priyam Gupta[1], Vidhi Khanduja[2,*]

Abstract

Dark patterns are deceptive design elements that influence user behavior online, frequently with unexpected results that go beyond personal experiences. These deceptive methods unintentionally encourage excessive consumption, which can seriously impede sustainability initiatives. This paper presents IntelliGuard WebScan, a system created specially to identify and combat these dishonest strategies. Employing painstaking examinations and assessment of heterogeneous datasets and rigorous experimentation with multiple algorithms, among them a support vector classifier (SVC), we were able to detect dark patterns with an astounding 93.22% accuracy. We propose a framework that uses an AI-powered browser plug-in with a user-friendly interface to find dark patterns in e-commerce websites. It scans the current webpage in real time and warns the user about the presence of dark patterns. Our methodology is designed to identify typical dark patterns that are present on e-commerce websites, such as social proof, obstruction, sneaking, scarcity, false urgency, forced action, and misdirection. Our ultimate objective is to provide users with the information and resources they need to navigate the online environment confidently and wisely, ultimately promoting a responsible online environment that gives priority to mindful consumption and is in line with long-term sustainability objectives.

Keywords: Dark patterns, support vector classifier, UI/UX, IntelliGuard WebScan, e-commerce websites

INTRODUCTION

IntelliGuard WebScan is a powerful tool designed to identify and expose dark patterns on e-commerce websites. Scanning for deceptive design practices that manipulate user behavior helps ensure a more transparent and ethical online shopping experience.

What Are Dark Patterns?

Dark patterns are manipulative and clever designs that target the user's brain and manipulate it to make certain decisions that it was not sure to make earlier. This attacks the user fundamentally by infusing certain fears and insecurities, thereby snatching the freedom to choose between the decisions. This is a very poor trade practice that can potentially harm a brand's reputation for a very long time, leading to great losses of goodwill.

In addition, strict laws are being enacted by governments worldwide to regulate such practices. The rationale behind the legislation is that companies or brands must increase sales/ engagement by offering utility and innovation through their products to attract customers, not by such poor trade practices in which they design certain features in the UI/UX, which are either very ambiguous that the customer becomes trapped or infuses fears and insecurities to boost sales. This is unsustainable and unfair, for which there must be strict compliance with regulating platforms.

***Author for Correspondence**
Vidhi Khanduja
E-mail: vidhikhanduja@hrc.du.ac.in

[1]Student, Department of Computer Science, Hansraj College, University of Delhi, Delhi, India
[2]Assistant Professor, Department of Computer Science, Hansraj College, University of Delhi, Delhi, India

Received Date: May 18, 2024
Accepted Date: July 08, 2024
Published Date: August 27, 2024

Citation: Yashasvi, Vineet Kumar Kankerwal, Priyam Gupta, Vidhi Khanduja. IntelliGaurd WebScan: Uncovering Dark Patterns on E-Commerce Websites. E-Commerce for Future & Trends. 2024; 11(3): 24–32p.

This subject is being closely studied by researchers all over the world, and as much work has not yet been done, there are many areas to explore in this domain. Recent studies have shown limited success in this early stage. The Ministry of Consumer Affairs, Government of India, places special emphasis on dark patterns to improve consumer awareness and encourage a healthy e-commerce experience. A recent study by Singapore University [1] demonstrated the negative effects of dark patterns, as shown in Figure 1.

We have created a novel method for identifying and classifying dark patterns in digital user interfaces called *IntelliGuard WebScan*. Our web browser plug-in analyzes user interactions and detects false design patterns that can reduce user autonomy or result in unintentional behavior. It does this by using sophisticated algorithms and machine learning approaches. The incorporation of sustainability concepts into the development and utilization of our detection tool is a fundamental aspect of our methodology. Our solution aims to protect customer interests while minimizing their environmental impact by prioritizing energy-efficient algorithms and resource-conscious computing techniques. In addition, by offering users educational materials and real-time alerts to assist them in identifying and avoiding dark patterns, our solution supports openness and user empowerment.

Major contributions of the *IntelliGuard WebScan* are:

1. *Data cleaning and feature engineering:* To ensure the effectiveness of our model, we employed data cleaning and feature engineering techniques to optimize and refine the training datasets.
2. *Advanced detection algorithms*: These algorithms employ support vector machines (SVMs) to evaluate user interfaces on websites and detect dark patterns that could otherwise go undetected.
3. *Real-time scanning*: This feature scans websites in real time and notifies users instantly when they encounter misleading design elements.
4. *User-friendly design*: The interface of the created browser extension has an intuitive and user-friendly design that makes it simple for users to comprehend the detected dark patterns and take the necessary action.
5. *Learning*: To better inform the user, it provides static information regarding each dark pattern.

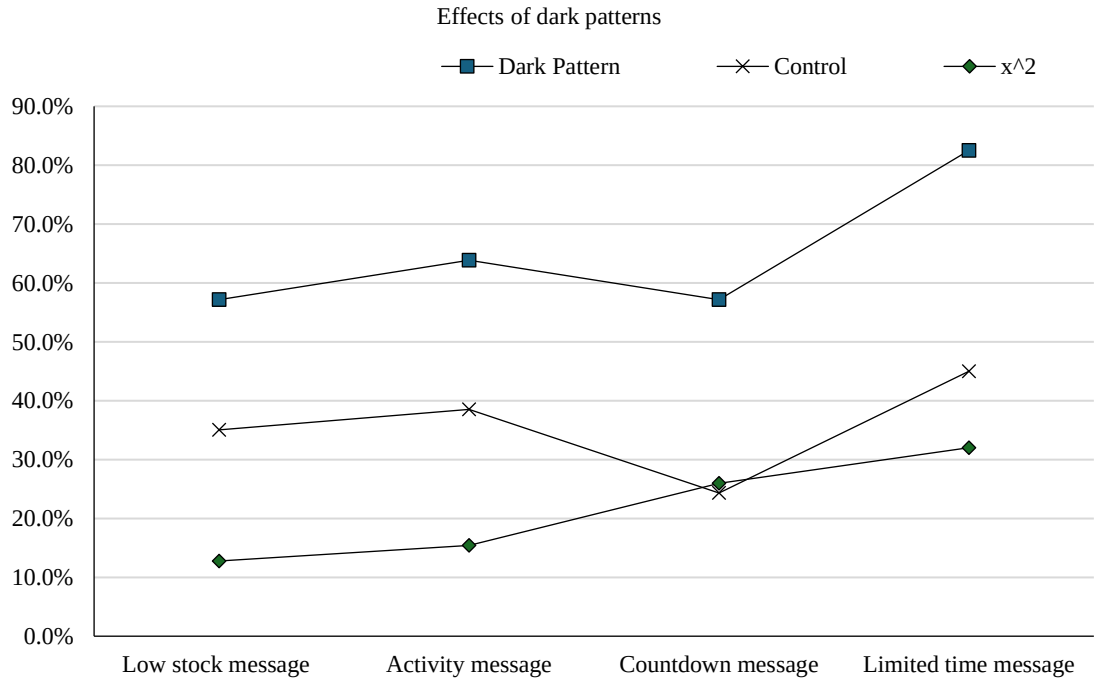

Figure 1. Effects of dark patterns.

The rest of the manuscript is organized as follows: Section 2 introduces the prior state of the art, followed by Section 3, which explains the proposed methodology of our product. Section 4 explains the experimental results and Section 5 concludes the paper.

LITERATURE SURVEY

The widespread use of dark patterns in digital interfaces has led to considerable research on their occurrence in recent years. Geronimo et al. presented two studies that assessed dark patterns in mobile device applications and user perception [2]. The first study analyzed 240 apps on the Google Play Store and found that 95% incorporated one or more ominous dark patterns. The second study involved 584 respondents rating the UI of a subset of apps, revealing that users often cannot perceive malicious design. Mathur et al. developed automated techniques to study dark patterns on a web scale [3]. It simulated user actions on approximately 11 K popular shopping websites, collecting text and screenshots to identify their use of dark patterns. On 11.1% of the websites, the researchers discovered at least one dark pattern, 183 of which included misleading content. On well-known websites, dark patterns were more prevalent and frequently made possible by outside parties, two of which allowed for dishonest activities. Feng et al. created a dark pattern dataset for e-commerce sites to aid researchers in their research on automatic dark pattern detection [4]. Both transformer-based and traditional NLP-based pre-trained language models were used to assess the dataset; the RoBERTa-large model achieved a maximum accuracy of 0.975. Chen et al. analyzed dark pattern taxonomies for mobile platforms and integrated them into a unified taxonomy [5]. It categorizes static and dynamic patterns, identifies six core properties, and develops a UIGuard to detect dark patterns. The UIGuard is a helpful and promising tool that provides a methodical way to reduce the effects of dark patterns for a wide range of applications. Bankel's study investigated the use of dark patterns in donation processes on e-commerce platforms of five NPOs [6]. Three dark patterns, namely, misdirection, confirm shaming, and trick questions, were identified. These patterns contradict NPOs' goals of contributing to social, public, or collective benefits.

Gray et al. [7] provided an overview of dark patterns, a concept originating from UX practitioners, and their current extent. It draws attention to the possibility that UX design and HCI research could help clarify this morally complex and ethical phenomenon. The authors reframe existing notions of dark patterns to reflect designers' strategies for manipulating user and shareholder value and promoting ethical UX practice. Nevala in her eminent work discusses the use of dark patterns in e-commerce and wrote that "dark patterns and their use in e-commerce" can be seen as a shortcut to increase sales, but research has shown that if people become aware of manipulation they encounter online, they start to reassess their attitude toward the retailer resulting in loss of trust and damage for the brand" [8]. Kodandaram et al. investigated the impact of advertisements and promotions on blind screen reader users' web interactions [9]. It was found that blind users are susceptible to misleading advertisements. The algorithm achieved F1 scores of 0.86 and 0.88, making it a potential foundation for future usability improvements in nonvisual ad interaction alternatives.

In this study, we propose an approach that uses an AI-powered browser plug-in to find dark patterns on e-commerce websites. Our methodology is designed to identify typical dark patterns that are present on e-commerce websites, such as sneaking, forced action, social proof, scarcity, false urgency, misdirection, and obstruction.

PROPOSED METHODOLOGY

Figure 2 shows a block diagram of the proposed model. We now explain each module as follows.

Dataset Preparation

Following extensive research, we identified two valuable datasets. We combined these datasets to produce two new datasets. The first dataset is used for dark pattern prediction, whereas the second dataset is used for pattern category prediction.

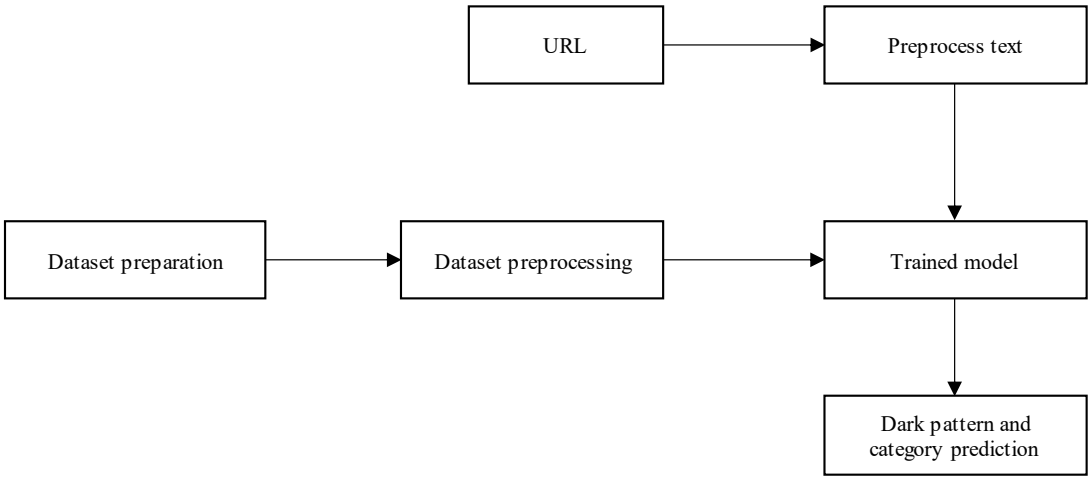

Figure 2. Overall dataflow of the proposed framework.

The first dataset included two labels: 0 and 1. A value of 0 indicates that the text is "not a dark pattern," whereas a value of 1 indicates that it is a "dark pattern." The second dataset contained various types of pattern categories that were labeled from 0 to 6 with the help of *Label Encoder*.

Data Preprocessing

Data preprocessing involves the following steps:

- *Lowering the text*: In this case, we lowered the entire text (using Python. *Lower()* function), which includes all phrases in the dataset. This results in homogeneity and helps train the model more efficiently.
- *Removing stop words*: It is a commonly used word (such as "the," "a," "an," or "in") that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query [10].

The stop words were deleted to increase computational efficiency. This was done to highlight only the most important terms. Python's *NLTK* (Natural Language Toolkit) module allows us to download and delete stop words from textual information.

- *Tokenizing*: This is the process of converting text into smaller components, known as tokens. These tokens may be as small as a single character or word [11]. It is used to extract features, process text, etc. Python uses the *NLTK* package for tokenization.
- *Porter stemming:* This removes suffixes from English words [12]. This was employed to simplify the text. For instance, after stemming, the word "posting" becomes "post." This is done to provide basic textual data that can speed up and improve the effectiveness of the model training.

Model Training

The two models were trained using various classifiers, as illustrated in Figure 3. Whether the text is a dark pattern will be predicted by the first model and the dark pattern category will be predicted by the second.

Extracting Text from the Website

To extract the text from the webpage, the URL is entered through our browser extension *IntelliGuard WebScan*, which produces the preprocessed text as an output, as depicted in Figure 4. We took a scrolling screenshot of the page. We then used *easyOCR* to extract the text and applied the necessary preprocessing steps to the obtained textual data.

Taking the screenshots of the webpage: we use a method that involves capturing a screenshot of the webpage as it is scrolling.

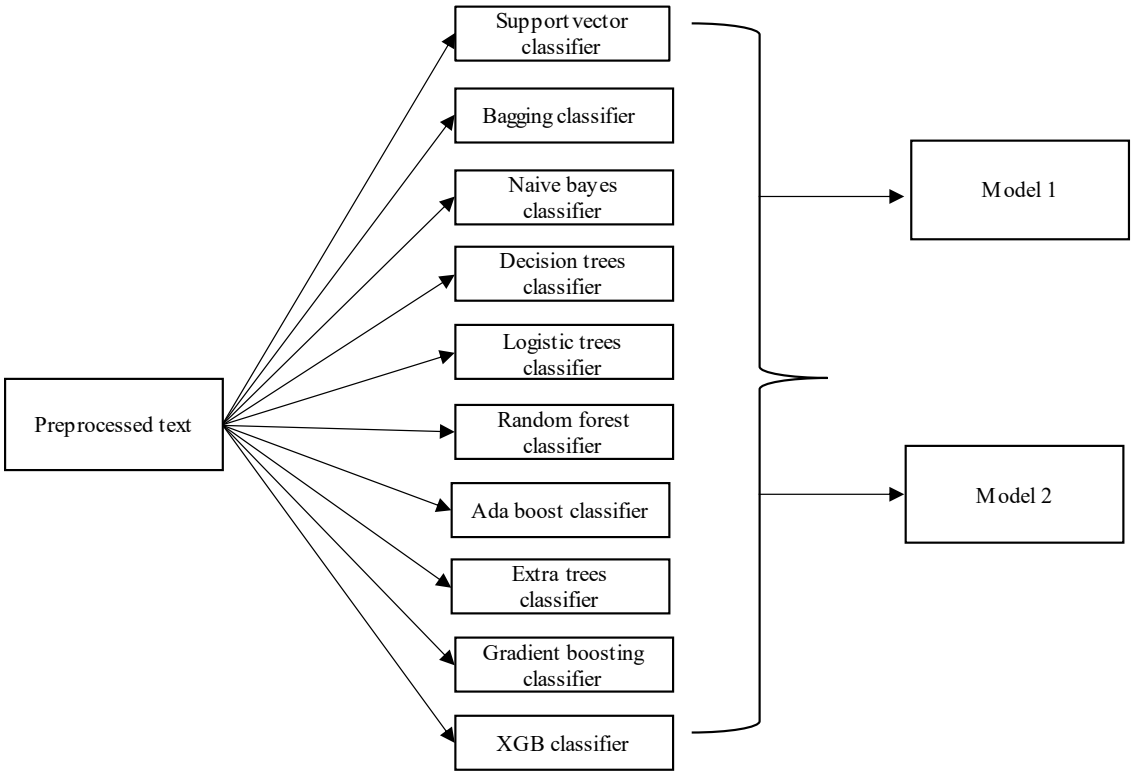

Figure 3. Model training.

Figure 4. Extracting text.

The Python *Selenium* library *Chrome WebDriver* was used to capture the scrolling screenshot. This launches a webpage whose URL is entered first. Subsequently, it browses the entire website, grabs a snapshot of the page that is now visible, and attaches each screenshot to create a single image.

Extracting text: Next, the Python *easyOCR* module was implemented to extract the text. The reader then extracted the screenshot text using Python's *easyOCR* module. Three things came back to us. First, word reading accuracy is the second, and text coordinates are the third. Next, to guarantee that the data is uniform enough to be used for prediction in the machine learning model, we used the same preprocessing steps that we previously followed for the dataset.

Preprocessing: This includes all the steps that we have taken earlier.

Predicting Dark Patterns

Predicting whether a preprocessed text is a dark pattern or not: When the sentences are fed into our model, it determines whether they are a dark pattern. The phrases are in a list that is iterated individually, and the model makes predictions. The outcome was either zero or one. If it is 1, indicating a dark pattern, it is run via our second model, the category predictor, as shown in Figure 5.

Predicting the category of dark pattern: As previously stated, if that sentence is a dark pattern, it is processed through our algorithm to determine its categorization. This returns the category of the dark patterns found. We trained our dataset to categorize the identified dark patterns into seven categories: urgency, forced action, obstruction, scarcity, misdirection, and social proof.

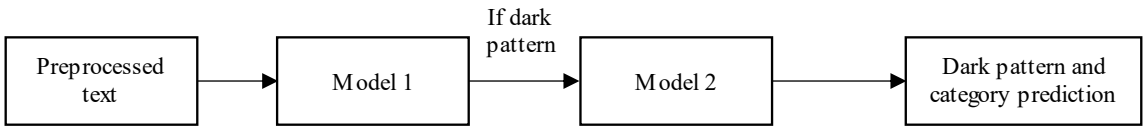

Figure 5. Prediction of dark patterns.

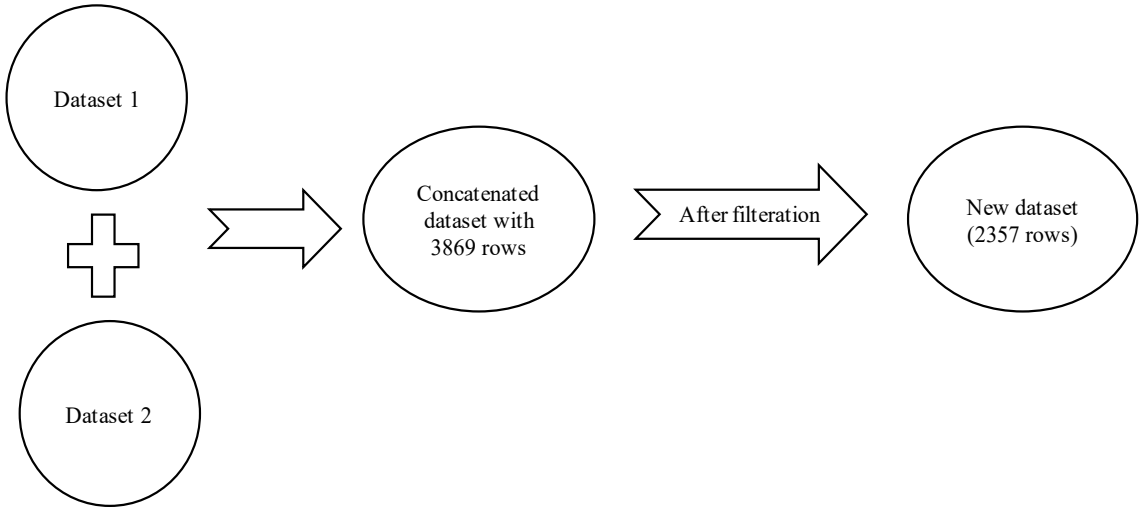

Figure 6. Dataset filtration.

EXPERIMENTAL RESULTS AND ANALYSIS

Dataset Collection

We have two datasets that we concatenated to form a single dataset. After concatenation, we filtered the data by removing all duplicates. The new dataset was created with 2357 rows as shown in Figure 6. The rows include the pattern category, labels for dark patterns, and textual data.

Dark Patterns Detected

IntelliGuard WebScan can detect seven categories of dark patterns which are as follows:

1. *Urgency*: By portraying a deal as having a deadline or creating a sense of urgency through the portrayal of great demand, this dark pattern seeks to accelerate user decision-making and purchases by pressuring consumers to act without giving careful thought.
2. *Social proof*: This dark pattern takes advantage of people's propensity to utilize other people's actions and opinions as a means of validating their own decisions using the power of social media and peer endorsements to expedite user decision-making.
3. *Misdirection*: To draw users' attention away from important information or options and steer them toward a planned choice without full awareness or understanding, misdirection uses deceptive visual clues, language, and emotional pleas.
4. *Scarcity*: By highlighting the limited supply or scarcity of a product or service, scarcity manipulates user behavior by creating a sense of impending doom and forcing users to act quickly to seize the chance.
5. *Obstruction*: Intentionally adding obstacles or complexity to the user's workflow hinders their progress or limits their capacity to carry out intended actions, forcing them to fulfill pre-set tasks or behaviors.
6. *Sneaking*: This is a tactic of secretly hiding or obscuring actions or pertinent information from users, frequently by using misleading design cues or hidden interfaces, which causes users to inadvertently perform actions they may not have wanted or intended.
7. *Forced action*: To access or maintain the desired content, users are required to perform certain activities.

The UI/UX of IntelliGuard WebScan

When the user loads the browser extension on any shopping website, it automatically fetches the URL of the current webpage, scans the entire webpage, and provides the user with a warning that shows the number of dark patterns and distinctive categories, which can be further clicked to show all categories and distinctive dark patterns present on the webpage, as shown in Figure 7.

Results When Tested on Various Websites

In Figure 8. Our browser extension is loaded on the Amazon website, and when the user clicks the Scan button, it displays a warning, and the types of dark patterns included on the website.

Figure 9 depicts the dark patterns on Flipkart. This indicates that two dark patterns have been discovered on this webpage, which, when clicked, display the various types of dark patterns detected.

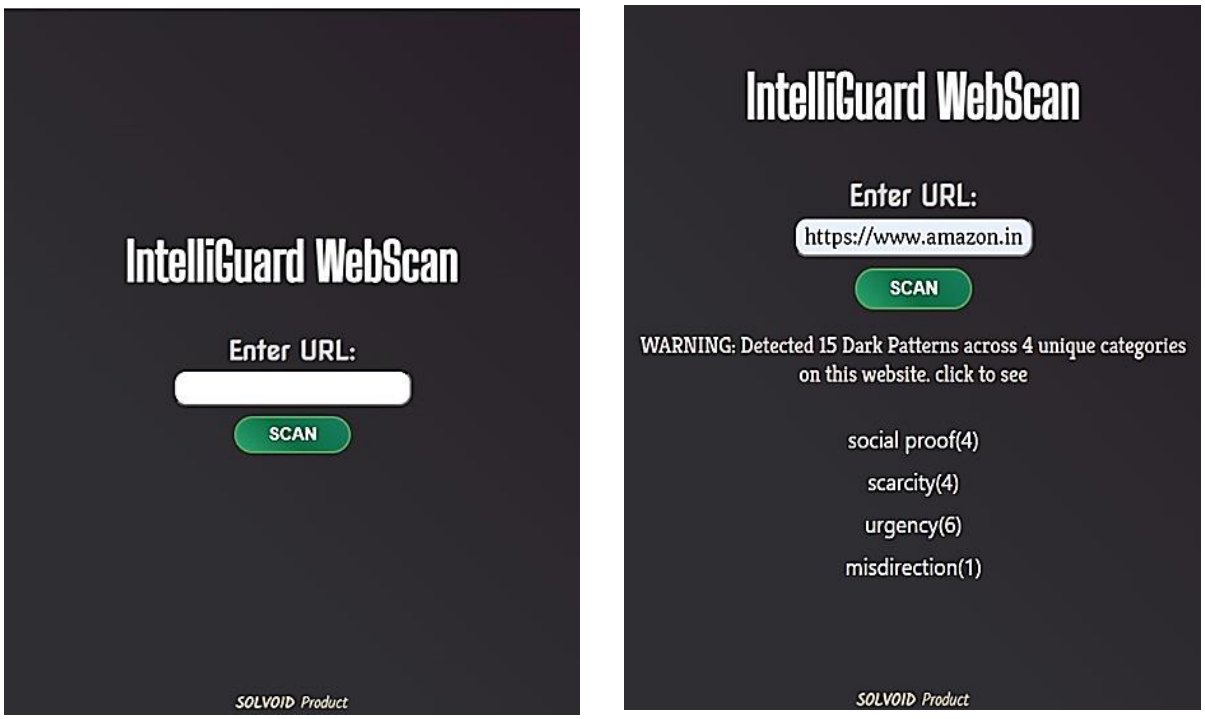

Figure 7. The user interface of IntelliGuard WebScan.

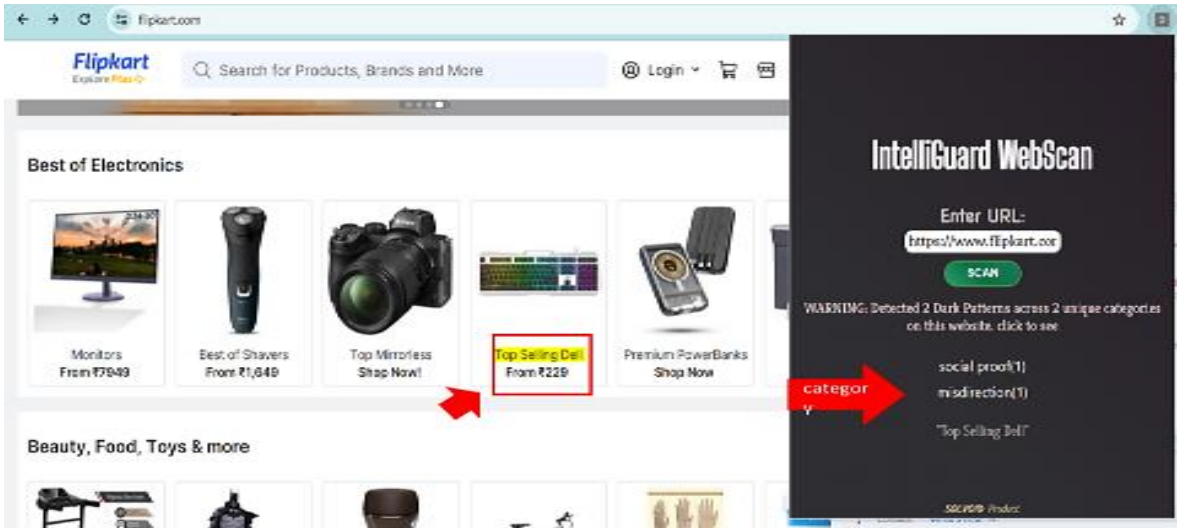

Figure 8. Screenshot depicting dark patterns.

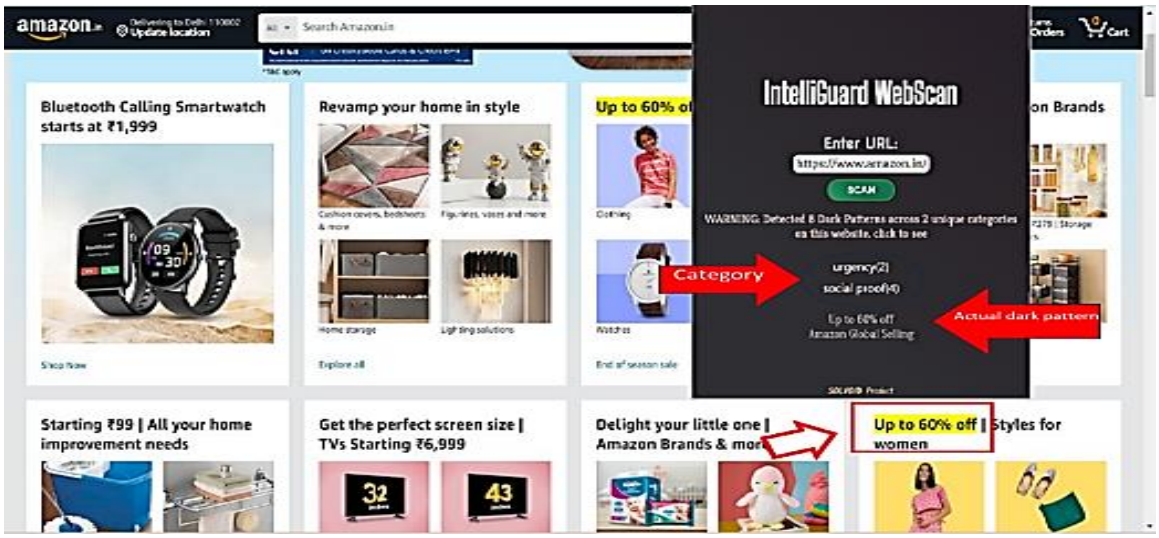

Figure 9. Screenshot depicting dark patterns.

Table 1. Experimental results of various ML algorithms.

Classifier used	Accuracy	Precision	F1 score	Recall
Support Vector Classifier	93.22%	93.75%	93.49%	94.26%
Decision Tree Classifier	82.20%	97.20%	83.76%	98.36%
Logistic Regression Classifier	80.29%	96.03%	91.82%	96.72%
Random Forest Classifier	91.10%	91.02%	92.53%	91.39%
AdaBoost Classifier	92.37%	90.95%	90.50%	91.80%
Bagging Classifier	90.46%	90.66%	90.83%	91.39%
Extra Trees Classifier	92.16%	90.98%	92.33%	91.39%
Gradient Boosting Classifier	87.71%	96.70%	89.13%	97.54%
XGB Classifier	90.08%	93.42%	91.45%	94.26%

Experimental Results

Table 1 revealed that the maximum accuracy obtained was 93.22% with the support vector classifier (SVC). Hence, we chose the SVC to train the final model.

CONCLUSION

In this study, we addressed the important problem of dark patterns—deceptive design cues that take advantage of user psychology to influence online behavior. We suggest *IntelliGuard WebScan*, a tool designed to identify these negative behaviors and protect users from their deceptive influence.

We carefully examined various datasets as part of our research process. To ensure the effectiveness of our model, we optimized and refined the training data using feature engineering techniques and data cleaning. Subsequently, we assessed multiple algorithms and determined that the SVC yielded the best results, detecting dark patterns with an astounding accuracy of 93.22%.

An important step toward creating an online environment that is more moral, open, and sustainable has been taken with the creation of *IntelliGuard WebScan*. This system encourages responsible online interactions and has a positive impact on a sustainable digital ecosystem by enabling users to recognize and steer clear of manipulative design practices. The negative effects of dark patterns include the possibility of increased consumption and resource depletion as a result of pressure tactics, which would ultimately impede long-term environmental and economic sustainability. Examples include impulsive and unnecessary purchases. Nevertheless, the field of dark patterns is dynamic, necessitating ongoing studies and adjustments to keep up with new deceitful strategies.

We will see more developments in *IntelliGuard WebScan* in the future. Our goal is to investigate the possibilities of transferring learning approaches, which enable the model to use information from one domain to improve its performance in another.

REFERENCES

1. Koh WC, Seah YZ. Unintended consumption: The effects of four e-commerce dark patterns. Cleaner Responsible Consum. 2023;11:100145. DOI: 10.1016/j.clrc.2023.100145.
2. Di Geronimo L, Braz L, Fregnan E, Palomba F, Bacchelli A. UI dark patterns and where to find them: A study on mobile applications and user perception. CHI Conf Hum Factors Comput Syst. 2020;1–14. DOI: 10.1145/3313831.3376600.
3. Mathur A, Acar G, Friedman MJ, Lucherini E, Mayer J, Chetty M, et al. Dark patterns at scale: Findings from a crawl of 11K shopping websites. Proc ACM Hum-Comput Interact. 2019;3:1–32. DOI: 10.1145/3359183.
4. Feng JY, Yuki T, Matsumoto N, Fukushima F, Kido H, Yamana H. Dark patterns in e-commerce: A dataset and its baseline evaluations. IEEE Int Conf Big Data. 2022;3015–22.
5. Chen J, Sun J, Feng S, Xing Z, Lu Q, Xu X, Chen C. Unveiling the tricks: Automated detection of dark patterns in mobile applications. 36th Annu ACM Symp User Interface Softw Technol. 2023;1–20. DOI: 10.1145/3586183.3606783.
6. Bankel M. Exploring the use of dark patterns in the donation processes of nonprofit eCommerce. In: Mejtoft T, Söderström U, Norberg O, Freidovich L, editors. Proceedings of the 21st Student Conference in Interaction Technology and Design; 2021 Jun; Umeå, Sweden. Umeå: Umeå University; 2021. p. 65–9.
7. Gray CM, Kou Y, Battles B, Hoggatt J, Toombs AL. The dark (patterns) side of UX design. CHI Conf Hum Factors Comput Syst. 2018;1–14. DOI: 10.1145/3173574.3174108.
8. Nevala E. Dark patterns and their use in e-commerce [Bachelor's thesis]. Jyväskylä: University of Jyväskylä; 2020.
9. Kodandaram SR, Sunkara M, Jayarathna S, Ashok V. Detecting deceptive dark-pattern web advertisements for blind screen-reader users. J Imaging. 2023;9:239. DOI: 10.3390/jimaging 9110239. PubMed: 37998086.
10. Mansur SMH, Salma S, Awofisayo D, Moran K. AidUI: Toward automated recognition of dark patterns in user interfaces. 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE), Melbourne, Australia. 2023. pp. 1958–70. doi: 10.1109/ICSE48619.2023. 00166.
11. GeeksforGeeks. (2017). Removing stop words with NLTK in Python. [online] Available from: https://www.geeksforgeeks.org/removing-stop-words-nltk-python/.
12. Awan AA. (2023). What is tokenization? [online] Datacamp.com. Available from: https://www. datacamp.com/blog/what-is-tokenization.

