30 January, 2024

Pattern Hunters

Nodal Centre: PEC University of Technology, Chandigarh

Team members:

1) Pranav Sharma (22103018)
2) Arnav Singh (22103020)
3) Gaurav Gupta (22103028)
4) Akanksha Narula (22103041)
5) Gurmehar Singh (22103078)

Table of Contents

1) Introduction:

The user experience on e-commerce platforms is crucial in influencing customer behavior in the dynamic world of digital commerce. A growing worry, meanwhile, is the frequency of dishonest design techniques referred to as "dark patterns." These deceptive UI strategies are skillfully designed to trick users into performing activities they might not have chosen voluntarily. So considering the seriousness of the problem, our project aims to provide a novel approach that can detect, classify, and measure the usage of dark patterns on e-commerce platforms.

2) Market analysis and feasibility study:

- **Market Landscape**:

  Over the past few years, the problem of dishonest behaviors on e-commerce platforms has become increasingly apparent. Because customers are becoming more aware of their digital rights and the ethical consequences of manipulative design, there is a growing market for products that address dark patterns. This not only highlights the importance of our initiative but also indicates a need in the market for products that provide people with the ability to defend themselves against dishonest business practices.

- **State of the Art:**

  A survey of the current state of the art reveals many initiatives aimed at mitigating the impact of dark patterns. Existing tools, often relying on traditional machine learning practices, have laid the groundwork for detection mechanisms. However, the limitations of these approaches, coupled with the evolving nature of dark patterns, underscore the need for advanced solutions.

- **Feasibility Study:**

  The feasibility of our dark pattern detection project relies on the convergence of technological advancements and an increasingly discerning user base. The integration of deep learning NLP models, alongside technologies like CNN and OCR, signifies a strategic response to the dynamic nature of deceptive practices. Additionally, the feasibility is substantiated by the willingness of users to adopt browser extensions that enhance their online security and promote transparency.

3) Problem Statement:

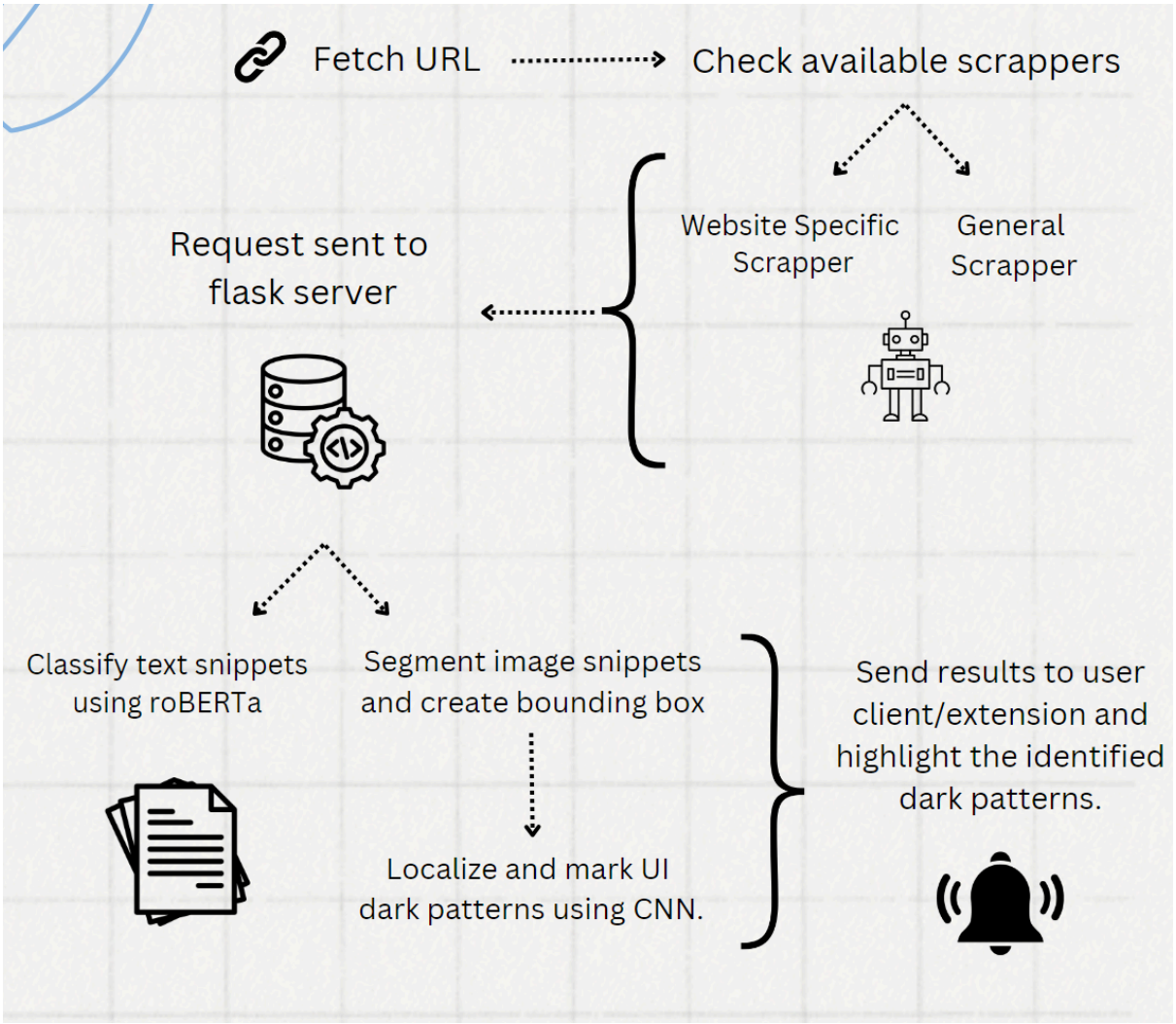
“Design and prototype innovative application or software-based solutions that can detect use, type, and scale of dark patterns on e-commerce platforms.”

A dark pattern is a user interface carefully crafted to trick users into doing a certain action that they wouldn’t have done otherwise. Dark patterns are of many types such as forced action, nagging, confirm shaming, interface interference, false urgency, and basket sneaking to name a few. We aim to

solve this problem by providing a solution that can detect dark patterns on various e-commerce digital platforms to keep our consumers safe through a browser extension by highlighting the dark patterns and providing the users with knowledge about them. It would help protect users from the malicious intentions of developers who aim to increase their profits through these malpractices. We aim to create a transparent digital world.

4) Project Description:

We aim to create a browser extension that uses advanced deep learning algorithms to identify and reveal dark patterns on different e-commerce sites. Giving people the information and understanding they need to safely navigate online environments is the main goal. Our method attempts to safeguard customers from the dishonest tactics used by e-commerce websites to increase their earnings by drawing attention to occurrences of dark patterns.



5) Competitive Context:

To address deceptive practices within e-commerce, various GitHub repositories have emerged, aiming to combat the issue of misleading tactics employed by online platforms to manipulate customer behavior.

- **Existing Solutions:**

  Existing GitHub repositories have focused on the application of traditional machine learning techniques to detect deceptive practices. Although these methods are beneficial, they frequently depend on feature engineering and pre-established guidelines, which restricts their capacity to adjust to changing trends in the ever-changing e-commerce environment.

- **User-Friendly Extension:**

  Our project introduces a user-friendly browser extension which is designed to identify dark patterns seamlessly during the user's online interactions. The extension provides a real-time, accessible, and intuitive solution for users to gain insights into deceptive practices on e-commerce platforms.

- **Current Traditional Machine Learning Model:**

  At present, our extension operates using a traditional machine learning model. However, we acknowledge the potential for higher accuracy achievable through the integration of advanced deep learning and more advanced natural language processing (NLP) models.

- **Advanced Detection Capabilities:**

  Our user-friendly browser extension currently utilizes traditional machine learning. As we're actively exploring advanced deep learning NLP models for better accuracy in identifying subtle dark patterns. Simultaneously, we're integrating Convolutional Neural Networks (CNN) and Optical Character Recognition (OCR) to extend our capabilities, enabling the detection of deceptive practices within textual content, images, pop-ups, and other visual elements. This holistic approach aims to offer users a comprehensive defense against evolving deceptive activities in the e-commerce landscape.

6) Literature Review:

Table 6.1 (Literature review)

Sr.	Title	Advancements	Limitations
1	Dark Patterns at Scale: Findings from a Crawl of 11K	• Developed crawlers for generating the list of shopping websites, the	• Consideration of only text based dark patterns. • Documented dark

	Shopping Websites	product page classifier, and the checkout crawler. • Clustering analysis of collected website data. • Compilation of dataset of over 11k websites.	patterns derived from existing literature only. • Only crawled product pages and checkout pages.
2	Dark patterns in e-commerce: a dataset and its baseline evaluations	• Evaluation of classical NLP models for dark pattern classification.. • Evaluation of transformer based pre-trained language models for dark pattern classification.	• Consideration of only text based dark patterns.
3	AidUI: Toward Automated Recognition of Dark Patterns in User Interfaces	• Proposal for a unified taxonomy of UI Dark Patterns. • Developed a research prototype to detect UI Dark Patterns. • Provided datasets and models for their solution.	• Only 10 specific types of dark patterns were targeted out of 27 classes identified.. • Privacy issues for deployment in applications.

7) Proposed Solution (Methodology):

Our proposed solution involves the use of Natural Language Processing to train the model on existing dataset of various text-based dark patterns. Training the model on the intent of sentences which are labeled as dark patterns helps it to determine a particular piece of text as dark pattern or not in the future. We used both deep-learning and classical machine learning models for training our model. Using deep learning models, we could clearly classify the intent of the statement but deploying them and being able to display the results in a viable amount of time turned out to be a big challenge. Using classification models, the time taken to predict the results turned out to be quite less and model deployment and integration also wasn't an issue but the classical machine learning models often detected dark patterns through the presence of certain "keywords" which often led to erroneous results.

We trained all our proposed models using the same dataset as not much dataset was available on the internet about dark patterns. Initially, we used bag of n grams with Multinomial Naive Bayes, then by Random Forest Classification and then by Bernoulli Naive Bayes. After that, we used Tf-Idf vectorizer with Multinomial Naive Bayes, Random Forest Classification and then Bernoulli Naive Bayes. Also since the dataset we had was limited, we also tried text augmentation using text-attack and NLPaug libraries. The major challenge encountered was that the accuracy was not significantly

increased and also text-attack led to overfitting and NLPaug led to generation of many erroneous inputs, disturbing the original database.

At the end, after careful observation, we decided to go with Bernoulli Naive Bayes, and after applying some specific hyperparameter and threshold values, we were able to achieve somewhat good accuracy. In deep learning, we used the BERT model, implementing it with Tensorflow and Pytorch. The accuracy in Tensorflow turned out to be 0.89. But, we encountered major issues in integrating these models with our server because of the large size of the models and also due to limited GPU availability. Below attached is a table combining all accuracies obtained by trying out a variety of classification and tokenization techniques.

Table 7.1 Without text augmentation:

	MultinomialNB	Random Forest	Bernoulli NB	Gaussian NB
n_grams (1,2)	0.92	0.92	0.94	0.92
n_grams (1,3)	0.92	0.90	0.93	0.95
n_grams (2,3)	0.90	0.85	0.87	0.82
Tf-Idf	0.91	0.96	0.96*	0.82

* after hyperparameter training

Table 7.2 With text augmentation:

	MultinomialNB	Random Forest	Bernoulli NB	Gaussian NB
n_grams (1,2)	0.93	0.91	0.94	0.88
n_grams (1,3)	0.93	0.95	0.91	0.88
n_grams (2,3)	0.91	0.88	0.83	0.85
Tf-Idf	0.91	0.95	0.94	0.79

BERT (before text augmentation) = 0.89
BERT (after text augmentation) = 0.89

After this, the main challenge was to get all the text-based-data from the pages the user was visiting, and for that we used web-scraping. We used the Scrappy framework, which is a python framework that has been used to collect all text data from a website by selecting various CSS selectors. We used the web-link of the current tab which the user is currently viewing, supplied it to the scraper to scrape all the text from that particular link to the scraper using CSS selectors, and then we fed the csv file to our model. After the model predicts the presence of dark patterns, we devise three methodologies to inform the user :

1) Using notification pop-up : We used the chrome extensions default pop-up feature to inform the user about the presence of a dark pattern, also displaying all the dark patterns detected by the model on that page.

2) Using DOM manipulation: The particular text/texts containing dark patterns gets highlighted for the user to be alerted

3) User customized approach: Using this approach, we give users the freedom to select text on his/her own which he/she wants to be examined by dark patterns. In this approach, we only fed the particular text selected by the user to our model and then displayed the respective results.

Our extension was also cross-browser compatible for many browsers, enhancing the user experience. We hosted the model, the scrapper and the extension on the backend flask server. Basically, everytime the user visits a new page, a new link is fed to the scrapper, which then overwrites the previous csv file and that when fed to the model generates a new output, and the user is alerted about the presence of dark patterns in any of the three ways mentioned earlier.

8) Future Scope

We plan to add two significant features in this solution in the future. The first is a feedback mechanism, which allows the user to rate how accurate the prediction was. This shall not only help in reducing erroneous results, but also help in improving the model's accuracy and the entire user experience as a whole. Secondly, we plan to add a chatbot, which alerts the user about the dark pattern it just predicted, its type, how it affects the thinking of an individual, and how to be alert and aware of that particular category of dark pattern. Basically, spreading awareness about dark patterns and enhancing user experience is the main aim which shall be fulfilled using this additional feature.

Lastly, we plan to rely on efficient prediction methods. Predicting using the presence of some specific keywords may turn out to be fast and efficient in the beginning, but in order to avoid erroneous

results, we need to take into consideration the entire context of the statement, which can be achieved with deep learning models involving LSTM or transformer architecture.

Also we plan to integrate a generalized scrapper or crawler which can fetch text from all websites. Also, we plan to integrate UI Dark pattern classification using OCR and Convolutional Neural Networks. A large portion of dark patterns is often observed in pop-ups, advertisements and other design entities. We
plan to detect them as well, beside detecting only text.

9) Directory Structure:

- Augmentation
  - aug_code.ipynb -> Module to augment data using Contextualized Word Embeddings from nlpaug library.
  - aug_data.csv -> Augmented Dataset file in csv format.
- Dataset
  - dataset.tsv -> Dataset on which our NLP model is trained
- Deliverables
  - This module contains all the HTML, CSS and Javascript files of the browser extensions.
- Models
  - This module contains the pickel(.pkl) files of our trained model.
- Scrapper
  - This module contains the files of our scrapper, which is extracting the text from the webpages.
- Training
  - This module contains the code files which we used to train our model.

10) Usage:

For Everyday Users:

- Protect users from deceptive practices while browsing and empower users to control their online experiences.
- Get alerts and manage subscriptions effortlessly.

For Companies:

- Ensure ethical design to build user trust.
- Adhere to industry standards and avoid legal issues.

For Cybersecurity Experts:

- Detecting and preventing online threats.
- Collect insights on emerging dark patterns for proactive cybersecurity.

For Regulatory Bodies and Advocacy Groups:

- Ensure websites follow guidelines and use it as a tool for regulatory enforcement.
- Educate the public on ethical design and user well-being.

11) Conclusion

In our investigation of dark patterns, we've observed their prevalence in both textual and user interface (UI) forms. Employing the Tf-Idf vectorizer for word embeddings and subsequently utilizing the Bernoulli Naive Bayes technique for classification, we achieved an impressive overall accuracy of 0.95. Through further refinement with hyperparameter tuning via RandomizedSearchCV, our accuracy improved to 0.96, with the optimized alpha value set at 0.122527063642198.

Our data collection efforts involved web scraping, allowing us to retrieve not only textual content but also CSS selectors from the target website. These collected data enabled us to effectively identify and highlight instances containing dark patterns.

Despite these achievements, detecting UI-based dark patterns poses an ongoing challenge. To address this, our future strategy involves exploring a combination of image preprocessing techniques and advanced classification methodologies. By integrating these approaches, we aim to enhance our ability to identify and combat deceptive user interface elements.

12) Remarks

The utilization of deceptive interfaces persists as a prevalent technique, especially within the realm of E-commerce platforms. While the development of models based on text and user interface has shown effectiveness to a certain extent, concerns surrounding the precision and accuracy of predictions loom large. This becomes particularly apparent when addressing privacy issues related to the capture of screenshots for optical character recognition (OCR) or image detection. It is imperative to implement effective measures that not only ensure accurate predictions but also prioritize and safeguard user privacy and personal space.

Furthermore, the influence of dark patterns on user buying experiences is a critical concern. Users need to be educated about the extent of manipulation that dark patterns can exert, impacting their decision-making process. Empowering users with knowledge on how to identify and counteract these manipulative tactics becomes paramount for fostering a fair and transparent digital marketplace.

The complexity of these issues underscores the need for governmental intervention. Establishing and enforcing regulations becomes essential to uphold consumer rights within the dynamic landscape of online commerce. Governments play a crucial role in setting the rules and standards that protect users from deceptive practices, ensuring a level playing field and fostering trust in digital transactions. Overall, the multifaceted nature of these challenges emphasizes the importance of comprehensive solutions that address accuracy, privacy, user education, and regulatory frameworks.

13) References

https://github.com/yamanalab/ec-darkpattern
https://github.com/aruneshmathur/dark-patterns
https://github.com/NicholasTung/dark-patterns-recognition
https://github.com/SageSELab/AidUI
https://www.tensorflow.org/guide
https://huggingface.co/docs/transformers/model_doc/bert
https://huggingface.co/docs/transformers/model_doc/roberta
https://nlpaug.readthedocs.io/en/latest/
https://textattack.readthedocs.io/en/latest/apidoc/textattack.augmentation.html
https://pytorch.org/docs/stable/index.html
https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.BernoulliNB.html
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html
https://scikit-learn.org/stable/modules/naive_bayes.html
https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/
https://machinelearning.wtf/terms/bag-of-n-grams/
https://chromewebstore.google.com/
https://docs.scrapy.org/en/latest/
https://flask.palletsprojects.com/en/3.0.x/