DARK PATTERN BUSTER HACKATHON 2023

**A MACHINE LEARNING APPROACH FOR PRODUCT**

**INTEGRITY IN E-COMMERCE**

- - - - - Idea submitted by **TEAM INCEPTION** - - - - -

Vishnu M P

Viswa K

Ranjana G

Divyaa B

Tanush T M

(From Easwari Engineering College, Chennai)

TABLE OF CONTENTS

1. Dark Patterns in E-Commerce
2. Targeted Dark Patterns
3. Scope and Objectives
4. Project Work Flow
5. Data Collection and Pre-processing
6. Feature Engineering
7. Machine Learning Model Development
8. Prototype Development and Implementation
9. Conclusion

DARK PATTERNS IN E-COMMERCE

As India continues to experience an unprecedented surge in the ecommerce market, a deceptive behaviour of online retail firms has slowly started unfolding in the space. These so-called “wrongful practices” known as ‘dark patterns’ are being adopted by some ecommerce companies to manipulate users into taking actions or making decisions that may not be in their best interests.

“Dark patterns shall mean any practices or deceptive design pattern using user interface or user experience interactions on any platform that is designed to mislead or trick users to do something they originally did not intend or want to do, by subverting or impairing the consumer autonomy, decision making or choice, amounting to a misleading advertisement or unfair trade practice or violation of consumer rights,” the Central Consumer Protection Authority (CCPA) said in an order.

The guidelines are applicable to all those platforms offering goods or services in India, including advertisers and sellers.

An Overview of Consumer Protection Authority’s Specified Dark Patterns:

* False Urgency: False Urgency refers to the deceptive practice of falsely conveying or implying a sense of urgency or scarcity to mislead users into making immediate purchases or taking prompt actions, potentially resulting in a purchase.
* Basket Sneaking: Basket Sneaking has been defined as the practice of including additional items, such as products, services, payments to charity, or donations, during the checkout process on a platform without the explicit consent of the user. This results in the total amount payable by the user exceeding the amount intended for the chosen product or service.
* Confirm Shaming: Confirm Shaming is the practice of employing phrases, videos, audio, or any other means to instil a sense of fear, shame, ridicule, or guilt in the user’s mind. The objective is to subtly push the user into a specific action, such as purchasing a product or service from the platform or continuing a subscription, with the primary intent of achieving commercial gains by manipulating consumer choices.
* Forced Action: Forced Action refers to the practice of compelling a user to take an action that necessitates purchasing additional goods, subscribing to unrelated services, signing up for other services, or sharing personal information. This is done in connection with the user’s original intent to buy or subscribe to a particular product or service.
* Subscription Trap: A “Subscription Trap” involves tactics such as making the cancellation of a paid subscription overly complex or impossible, hiding the cancellation option, compelling users to provide payment details for auto-debits even for ostensibly free subscriptions, and employing ambiguous or confusing instructions regarding the cancellation process.
* Interface Interference: Interface Interference refers to a design element that manipulates the user interface by emphasising specific information while obscuring other relevant information relative to that data, with an intention to mislead the user and steer them away from the desired action.
* Bait and Switch: Bait and switch is a deceptive practice that involves advertising a specific outcome based on the user’s action and then providing an alternative outcome that differs from what was initially presented.
* Drip Pricing: Drip pricing is a practice characterised by several deceptive tactics, including not disclosing elements of prices upfront or revealing them subtly within the user experience, or disclosing the price only after the confirmation of purchase, leading to a charge higher than initially indicated at the checkout.
* Disguised Advertisement: Disguised advertisement refers to the practice of presenting advertisements in a manner that conceals their true nature, making them appear as other types of content, such as user-generated content, news articles, or false advertisements. The intention is to seamlessly integrate these ads into the overall interface, tricking customers into clicking on them.

There are many such dark techniques used in multiple places in the web. But the issue that we are focusing on is two-fold:

1. **Misleading Product Information:** Deceptive pricing strategies, specifically the false inflation of prices followed by presenting a seemingly substantial discount, mislead consumers. Our focus lies in recognizing this pattern of misleading crucial information.
2. **Dark Pattern in User Ratings:** The misuse of rating systems leads to misconceptions, where a high number of 5-star ratings might not truly reflect a product's quality compared to a larger set of moderately rated reviews.

[Reference:<https://inc42.com/buzz/govt-issues-list-of-13-dark-patterns-plaguing-ecommerce-websites/:~:text=“Dark%20patterns%20shall%20mean%20any,autonomy%2C%20decision%20making%20or%20choice%2C>]

TARGETED DARK PATTERNS

Our project aims to shed light on two prevalent dark patterns that significantly impact consumers' decision-making processes:

1. Misleading Product Information

* **Deceptive Pricing Strategies**: One of the most common tactics employed by e-commerce platforms involves the manipulation of product prices. This manifests in the form of false inflation of prices, artificially inflating the original price of a product to present an illusion of a substantial discount when the product is offered at its actual market price.
* **Seemingly Substantial Discounts:** Consumers are often lured into making purchases based on the perception of receiving significant discounts. However, these discounts might be misleading as the initial higher price is artificially set, leading buyers to believe they are saving more than they are.
* **Recognition of Misleading Crucial Information:** Our project's primary focus is on recognizing and identifying this pattern of manipulating crucial information highlighting instances where pricing strategies mislead consumers, fostering a more transparent marketplace.

2. Dark Pattern in User Ratings

* **Misuse of Rating Systems:** E-commerce platforms heavily rely on user ratings and reviews as indicators of product quality. However, these systems can be manipulated to create misconceptions. A high number of 5-star ratings might not necessarily represent a product's actual quality, especially when compared to a larger set of moderately rated reviews.
* **Disparity in Rating Perception:** The misuse of ratings can create a skewed perception, where products with fewer, highly positive ratings are perceived as superior, overshadowing products with a larger volume of moderately positive or constructive reviews.
* **Objective of Addressing Rating Misconceptions:** By addressing this dark pattern in user ratings, our project aims to rectify misconceptions, enabling consumers to make more informed decisions based on the actual quality and feedback associated with products.

In summary, these targeted dark patterns, encompassing misleading product information and manipulation of user ratings, profoundly impact consumers' perceptions and decisions in the e-commerce sphere. Our project endeavours to uncover and address these deceptive practices, fostering a more transparent and trustworthy online marketplace for consumers.

SCOPE AND OBJECTIVES

Project Goals:

The primary goal of our project is to develop a robust system capable of identifying and flagging prevalent dark patterns in e-commerce platforms. Specifically, our focus centers on addressing two distinct but interconnected dark patterns:

1. Misleading Product Information:

* 1. Detecting and highlighting instances where deceptive pricing strategies are employed, such as false inflation of prices followed by seemingly substantial discounts.
  2. Empowering consumers by recognizing and flagging products where pricing tactics mislead crucial information, thereby fostering transparency in pricing strategies.

2. Dark Pattern in User Ratings:

* 1. Developing mechanisms to identify manipulation in rating systems, emphasizing instances where a high number of 5-star ratings might not authentically reflect product quality compared to a larger set of moderately rated reviews.
  2. Enabling consumers to make informed decisions by rectifying misconceptions created by the misuse of rating systems.

**PROJECT WORK FLOW**

**Data Collection and Processing:**

* Aggregating data from e-commerce platforms, specifically focusing on product metadata, pricing information, ratings, and reviews.
* Preprocessing and structuring the data to extract essential features necessary for machine learning model ingestion.

**Machine Learning Model Development:**

* Implementing machine learning algorithms to analyze collected data and detect patterns indicative of deceptive practices.
* Training the model to recognize and flag instances of misleading product information and rating manipulation.

**Prototype Development:**

* + Creating an initial prototype in the form of an interactive website to showcase the functionality of the model.
  + Considering scalability options, such as browser extensions or standalone applications, for wider accessibility in the future.

**Feature Engineering:**

* + Extracting critical features from the collected data, focusing on pricing differentials, rating distributions, and other indicative parameters of dark patterns.
  + Utilizing these features as inputs for the machine learning model to effectively identify instances of misleading information and rating manipulations.

**Model Validation and Testing:**

* + Employing rigorous testing methodologies to evaluate the model's accuracy and reliability in detecting targeted dark patterns.
  + Validating the model's performance against known instances of deceptive practices to ensure effectiveness and minimize false positives.

In essence, the project's scope encompasses a comprehensive approach to address prevalent dark patterns in e-commerce by leveraging data-driven methodologies and machine learning techniques. The objectives revolve around developing a robust system capable of identifying, flagging, and ultimately mitigating the impact of deceptive practices on consumer decision-making processes.

DATA COLLECTION AND PRE-PROCESSING

Description of Data Sources:

The dataset utilized for this project is sourced from e-commerce platforms, primarily focusing on Amazon. The data is gathered through scraping techniques, leveraging Amazon's API for metadata collection. This dataset encompasses essential information crucial for identifying targeted dark patterns:

Features in the Dataset:

- name\_of\_product: The name or title of the product listed on the e-commerce platform.

- main\_category : The primary category under which the product is classified.

- sub\_category: A more granular categorization within the main category.

- image: URLs or references to images associated with the product.

- link: Links redirecting to the product page on the platform.

- ratings: The average rating associated with the product.

- no\_of\_ratings: The number of ratings/reviews received by the product.

- discounted\_price: The displayed discounted price of the product.

- actual\_price: The actual price of the product before any discounts.

**NOTE : The dataset contains around 6 million data product entries. This includes 20 main\_categories and around 120 sub\_categories.**

**Data Preprocessing**

In the data preprocessing phase, various techniques are applied to ensure the dataset's cleanliness, uniformity, and readiness for subsequent analysis and machine learning model training. Here is an overview of preprocessing techniques applied to each column in the dataset:

* name\_of\_product:
  + Data Type Handling: Ensuring consistency in the data type (e.g., string) across all entries.
  + Text Cleaning: Removing special characters, leading/trailing whitespaces, or standardizing text formatting if necessary.
  + Normalization: Converting text to lowercase to maintain uniformity.
* main\_category and sub\_category:
  + Categorization Standardization: Verifying and standardizing category names to ensure consistency.
  + One-Hot Encoding: Converting categorical data into numerical form for model compatibility if required.
* image and link:
  + URL Validation: Verifying URL formats and ensuring integrity for image and link columns.
  + Handling Missing Values: Strategies to handle missing or invalid URLs, if present.
* ratings and no\_of\_ratings:
  + Data Type Conversion: Ensuring ratings are stored as numerical values (float or integer).
  + Handling Missing Values: Imputation of missing ratings or no\_of\_ratings if necessary based on domain-specific strategies (e.g., mean, median, mode).
* discounted\_price and actual\_price:
  + Data Type Conversion: Ensuring prices are stored as numerical values (float or integer).
  + Consistency Check: Cross-validation to ensure the discounted\_price is logically lower than or equal to the actual\_price.
  + Handling Missing Values: Imputing missing prices based on appropriate strategies (e.g., median price for similar products).

**Data Cleaning and Null Value Handling:**

- Missing Value Treatment: Identifying and addressing null or missing values across columns.

- Outlier Detection: Identification and handling of outliers that might affect subsequent analysis.

**Feature Engineering:**

- Creating New Features: Deriving additional features like price differentials, discount percentages, or derived rating attributes to enhance model performance.

**Dataset Splitting:**

- Train-Test Split: Partitioning the dataset into training and test sets for model development and evaluation.

By applying these preprocessing techniques to each column in the dataset, we ensure uniformity, cleanliness, and reliability in the data. These steps are crucial in preparing the dataset for subsequent machine learning model training and analysis, ultimately contributing to more accurate and meaningful insights.

FEATURE ENGINEERING

Feature engineering plays a pivotal role in machine learning tasks. Here, we delve into the essential features extracted from the collected data and their significance in identifying the targeted dark patterns.

Features Extracted from the Collected Data:

**1. Price Differentials:**

- Significance: Calculating the difference between discounted\_price and actual\_price for each product.

- Importance: Enables the model to identify instances where the displayed discount might be misleadingly inflated.

**2. Rating Distribution Metrics:**

- Significance: Analyzing the spread and variation in ratings using statistics such as mean, standard deviation, or quartiles.

- Importance: Aids in distinguishing between products with a high number of perfect ratings versus those with more diverse but moderate ratings.

**3. Ratio of Ratings to No. of Ratings:**

- Significance: Computing the ratio of ratings to the number of reviews for each product.

- Importance: Indicates if a product has a suspiciously high number of ratings relative to the total number of reviews.

**4. Category-Based Attributes:**

- Significance: Aggregating statistics (e.g., average price, average ratings) specific to main\_category and sub\_category.

- Importance: Captures patterns within different product categories, helping the model identify discrepancies in pricing or ratings within categories.

**5. Text-Based Features (name\_of\_product):**

- Significance: Utilizing text analysis techniques like word embeddings or TF-IDF to extract meaningful information from product names.

- Importance: May uncover patterns or keywords in product names associated with inflated discounts or misleading descriptions.

Selection and Significance of Features:

* + Discerning Misleading Discounts: Features such as price differentials play a crucial role in identifying instances where discounts might not reflect genuine savings. By analyzing these differences, the model can flag products with potential misleading discount strategies.
  + Rating Manipulation Detection: The distribution metrics and ratio of ratings to the number of reviews aid in identifying patterns where a product might have an unusual number of high ratings compared to the volume of actual reviews. This assists in uncovering potential rating manipulations.
  + Category-Specific Insights: Analyzing category-based attributes helps contextualize patterns within specific product categories, allowing for a more nuanced understanding of deceptive practices that might vary across different product types.
  + Historical Context and Text Analysis: Incorporating historical price data and text-based features provides additional context and insight into irregularities or patterns not solely based on numerical metrics, enhancing the model's ability to detect subtle dark patterns.

By leveraging these carefully selected features extracted from the dataset, the machine learning model gains a comprehensive understanding of various aspects associated with product pricing, ratings, and categorization, enabling it to discern and flag potential instances of dark patterns prevalent in e-commerce platforms.

MACHINE LEARNING MODEL DEVELOPMENT

Overview of Machine Learning Algorithms Used

For the detection of dark patterns in e-commerce data, our project employs a combination of supervised learning techniques and anomaly detection methodologies. The selected algorithms include:

**1. Supervised Learning (Classification):**

- Random Forest Classifier: Utilized for its ability to handle large datasets, handle categorical variables, and discern patterns among numerous features.

- Gradient Boosting Classifier: Employed to improve predictive accuracy by sequentially adding models to correct errors of preceding models.

- Ensemble Learning Techniques: We plan on employing various ML algorithm fused and integrated together to provide the optimum results

**2. Anomaly Detection Techniques:**

- Isolation Forest: Employed for outlier detection, particularly effective in identifying anomalies or irregularities in product pricing and rating distributions.

- One-Class SVM (Support Vector Machine): Utilized to detect deviations from the norm, aiding in recognizing unusual patterns in the dataset.

Model Architecture and Training Process:

- Feature Selection and Preparation: Selected features extracted from the dataset (e.g., price differentials, rating metrics) are prepared and pre-processed for model ingestion. Categorical variables are encoded, and numerical data is scaled for uniformity.

- Model Training: The selected algorithms are trained on the pre-processed dataset, where the data is split into training and validation sets. The models are fed with labelled data, utilizing supervised learning to identify patterns indicative of dark patterns.

- Ensemble Methods: Utilizing ensemble learning techniques, such as Random Forest and Gradient Boosting, where multiple models are combined to enhance accuracy and reduce overfitting.

Fine-tuning and Optimization Techniques:

- Hyperparameter Tuning: Employing techniques like grid search or randomized search to optimize model hyperparameters, ensuring the best model performance.

- Cross-Validation: Employing k-fold cross-validation to validate model robustness, ensuring that the model's performance is consistent across different subsets of the dataset.

- Regularization Methods: Implementing techniques like L1 or L2 regularization to prevent overfitting and improve the model's generalizability.

- Evaluation Metrics: Utilizing metrics such as precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to assess model performance and select the best-performing model.

- Threshold Selection: Setting optimal decision thresholds to balance between model sensitivity and specificity, thereby minimizing false positives and false negatives.

By leveraging a combination of supervised learning techniques and anomaly detection methods, our model aims to effectively identify and flag instances of dark patterns in e-commerce data, fostering a more transparent and trustworthy online marketplace for consumers.

PROTOTYPE DEVELOPMENT AND IMPLEMENTATION

Description of the Initial Prototype (Interactive Website) and Its Functionalities:

The initial prototype developed for this project serves as an interactive website, providing users with a user-friendly interface to interact with the dark pattern detection system. The prototype encompasses the following key functionalities:

**1. Product Input Interface:**

- Users can input the metadata of the product they want to check to our interactive website.

- In the further developed versions of the project, users can input product details or URLs from various e-commerce platforms (e.g., Amazon) into the interface.

- The system then retrieves relevant metadata such as product names, categories, prices, ratings, and other pertinent information for analysis.

**2. Dark Pattern Detection Display:**

- Upon inputting product details, the system processes the information using the trained machine learning models to detect potential dark patterns.

- Detected instances of misleading product information or rating manipulations are flagged and displayed to users.

**3. Visualization and Explanation:**

- The prototype provides visualizations or explanations highlighting the specific features contributing to the identified dark patterns.

- Users are presented with insights or explanations regarding why certain products have been flagged for potential deceptive practices.

**4. User Feedback and Interaction:**

- Users have the option to provide feedback on the flagged products, indicating whether they found the information useful or offering additional insights.

**5. Accessibility and User Guidance:**

- The website includes user guidance or tooltips explaining the system's functionality and providing instructions for optimal use.

Potential Scalability Options (Browser Extensions, Standalone Applications):

**1. Browser Extensions:**

- Expanding the prototype's accessibility by developing browser extensions compatible with popular web browsers (e.g., Chrome, Firefox).

- These extensions could offer real-time detection of dark patterns while users browse e-commerce websites, alerting them to potential deceptive practices.

**2. Standalone Applications:**

- Developing standalone applications for desktop or mobile platforms to widen the system's reach.

- Standalone applications could provide enhanced features, user customization options, and offline functionality.

**3. Cloud Integration and APIs:**

- Exploring options to integrate the system's functionality via APIs or cloud services, allowing integration into third-party applications or platforms.

**4. Enhanced Functionality and Customization:**

- Future iterations could include additional features, such as user preferences, personalized alerts, or the ability to track products for potential dark patterns over time.

The initial prototype serves as a foundational tool demonstrating the capability of the dark pattern detection system. With potential scalability options like browser extensions or standalone applications, the project aims to extend its reach, offering users enhanced accessibility and functionality, ultimately contributing to a more informed and empowered consumer base in the e-commerce domain.

CONCLUSION

**Summary of Key Findings, Achievements, and Contributions of the Project:**

Throughout the development and implementation of the dark pattern detection system in e-commerce, several significant findings, achievements, and contributions have emerged:

1. Effective Dark Pattern Identification: The project successfully developed a machine learning model capable of identifying and flagging instances of misleading product information and rating manipulations across various e-commerce platforms.

2. Feature Engineering Significance: The selection and engineering of features, including price differentials, rating distributions, and category-based attributes, played a pivotal role in enhancing the model's accuracy in detecting deceptive practices.

3. Prototype Development: The creation of an interactive website prototype showcasing the system's functionality provided a user-friendly interface for consumers to identify potential dark patterns in product listings.

4. Scalability and Future Potential: Exploration of potential scalability options, including browser extensions and standalone applications, demonstrated the system's adaptability and potential for wider integration.

5. Empowering Consumers: By highlighting instances of potential deception, the project aimed to empower consumers with the necessary tools and information to make more informed decisions while navigating e-commerce platforms.

**Reinforcement of the Significance of Combating Dark Patterns in E-commerce:**

The prevalence of dark patterns in e-commerce platforms continues to impact consumer trust, decision-making processes, and overall marketplace transparency. The significance of combating these deceptive practices cannot be understated due to the following reasons:

1. Consumer Empowerment: Combatting dark patterns ensures consumers have access to transparent and accurate information, enabling them to make informed purchasing decisions aligned with their preferences and needs.

2. Trust and Integrity: Mitigating deceptive practices fosters a more trustworthy and reliable e-commerce ecosystem, promoting fairness, honesty, and integrity in product presentation and user reviews.

3. Marketplace Transparency: By identifying and flagging potential dark patterns, the project contributes to creating a more transparent and equitable marketplace, encouraging fair competition among sellers.

4. Ethical Responsibility: Upholding ethical standards in e-commerce is crucial. Detecting and addressing dark patterns aligns with ethical responsibilities toward consumers, ensuring they are not misled or manipulated.

In conclusion, the project's findings and contributions underscore the necessity of combatting dark patterns in e-commerce. By leveraging technology, developing robust detection systems, and empowering consumers with tools for transparency, the project aims to foster a more trustworthy and consumer-centric online marketplace.