**Documentation**

A dark pattern refers to a user interface design that is crafted to manipulate or deceive users into taking actions that they might not otherwise choose to do willingly. Dark patterns are often employed on websites and in software interfaces to achieve specific goals, such as tricking users into signing up for services, making purchases, or providing personal information.

These patterns can take various forms, including misleading wording, hidden options, confusing interfaces, and pre-selected checkboxes. The intent behind dark patterns is generally to prioritize the interests of the website or service provider over the user's best interests, often resulting in a frustrating or deceptive user experience.

The term "dark pattern" is used to highlight the unethical or deceptive nature of these design practices. Many advocacy groups and individuals work to raise awareness about dark patterns and promote ethical user interface design.

Dark patterns manifest in various forms, each employing deceptive techniques to manipulate user behavior. Here are some common types:

1. Misdirection: - Hidden Costs: Concealing additional fees until the last step of a transaction. - Sneak into Basket: Adding extra items to a user's cart without clear consent.
2. Obstruction: - Roadblocks: Introducing barriers that make it challenging for users to perform desired actions, such as canceling a subscription.
3. Forced Action: - Confirmshaming: Guilt-tripping users into taking certain actions by making alternatives seem socially unacceptable.
4. Hidden Defaults: - Pre-Selected Options: Automatically opting users into services or subscriptions unless they actively opt out.
5. Interface Interference: - Disguised Ads: Presenting advertisements that resemble native content, leading users to unintentionally click on them.
6. Misleading Language: - Trick Questions: Using ambiguous or confusing language to mislead users during decision-making.
7. Urgency: - Fake Scarcity: Creating a false sense of urgency by suggesting limited availability to prompt immediate action.
8. Privacy Zuckering: - Deceptive Privacy Settings: Manipulating users into sharing more personal information than they intend.
9. Bait and Switch: - Advertising One Thing, Delivering Another: Promising a certain product or service and delivering something less desirable.
10. Friend Spam: - Unauthorized Invitations: Automatically sending invitations or messages to a user's contacts without explicit consent.

In response to the widespread issue of Dark Patterns, which are manipulative design tactics employed on websites to deceive users, our project proposes a comprehensive solution aimed at enhancing transparency and empowering users. Leveraging the extensive "**Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites" dataset by Mathur et al.**, our approach involves identifying and highlighting potential regions of dark patterns on websites.

The dataset, a valuable resource, provides a deep understanding of deceptive design practices across a diverse range of shopping websites.

Our solution consists of a **multi-step process**:

1. Identification of Dark HTML Tags:

- Using the Bernoulli Naïve Bayes algorithm, we classify HTML tags as either dark or non-dark based on their alignment with patterns observed in the dataset. Dark HTML tags are those likely associated with deceptive practices and is identified by the algorithm of Bernoulli Naïve Bayes.

1. Categorization using Multinomial Naïve Bayes:

- If a tag is classified as dark, we employ the Multinomial Naïve Bayes algorithm to further categorize it into specific dark pattern types. This step provides a nuanced understanding of the nature of the detected dark patterns.

1. User Warning System:

- The identified dark patterns and their respective categories are then highlighted on the website interface. This user-friendly approach ensures that users receive immediate visual cues about the potential deceptive elements present.

**Technological Stack**:

1. Flask:

- The Flask web framework is employed to seamlessly integrate our solution into existing website structures, ensuring a smooth user experience.

- It facilitates the creation of an intuitive interface for displaying warnings and detailed information about detected dark patterns.

1. Joblib:

- Joblib is utilized for parallelizing and optimizing computations related to HTML tag classification. This is particularly crucial when dealing with large datasets, ensuring efficient processing.

1. scikit-learn:

- We leverage machine learning algorithms from the scikit-learn library for HTML tag classification. The Bernoulli Naïve Bayes algorithm is utilized for the initial classification of HTML tags, while the Multinomial Naïve Bayes algorithm aids in further categorization.

**Getting Started:**

1. Navigate to the api Directory:

Open your terminal or command prompt.

Change the current directory to the 'api' directory using the 'cd' command.

Command- **cd .\api\**

1. Install Required Dependencies:

Ensure you have Python and pip installed.

Run the following command to install the necessary dependencies listed in the requirements.txt file.

Command- **pip install -r requirements.txt**

1. Run the Main Application:

Execute the main application using the following command:

Command- **python app.py**

This will start the backend server required for the extension.

1. Add the Extension to Your Browser:

Open your browser and go to the Extensions or Add-ons page (this varies by browser).

Enable the developer mode.

Use the "Load Unpacked" feature to add the 'app' folder to your browser extension.

How to Use the Extension:

Once the extension is successfully added to your browser, you can start using it to identify and counteract dark patterns on websites.

1. Navigate to a Website:

Visit a website where you suspect the presence of dark patterns.

1. Activate the Extension:

Click on the extension icon in your browser to activate the tool.

**Detection Process:**

The extension will analyze the website's HTML tags using machine learning algorithms to identify potential dark patterns.

Detected dark patterns will be highlighted or marked on the webpage.

User Warnings:

If dark patterns are found, the extension will provide warnings to the user about the specific type of dark pattern identified.

Users will be informed about deceptive elements present on the webpage.

Counteraction/ Crowdsourced Features:

Depending on the extension's capabilities, users may have options to counteract detected dark patterns, such as reporting, disabling, or providing alternative suggestions.

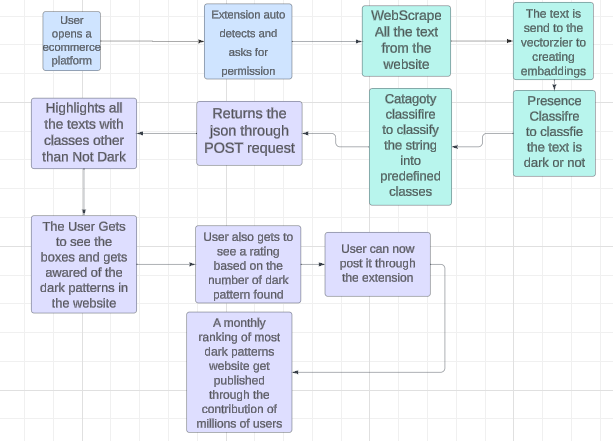
Explore Settings:

Check if the extension has any settings or configurations that can be adjusted based on user preferences.

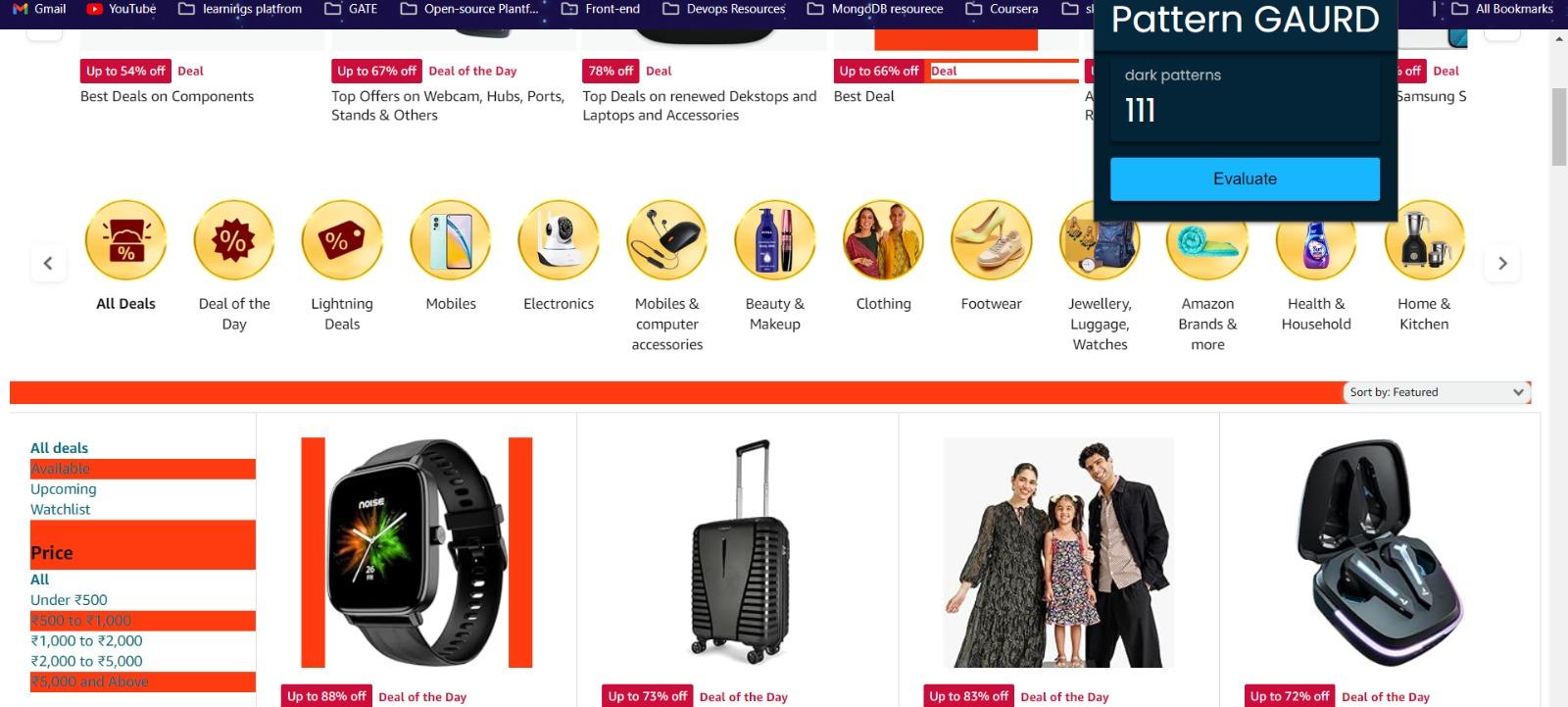
By following these steps, users can effectively utilize the extension to identify, understand, and counteract dark patterns on websites, promoting a more transparent and user-friendly online experience.

**Flowchart:**

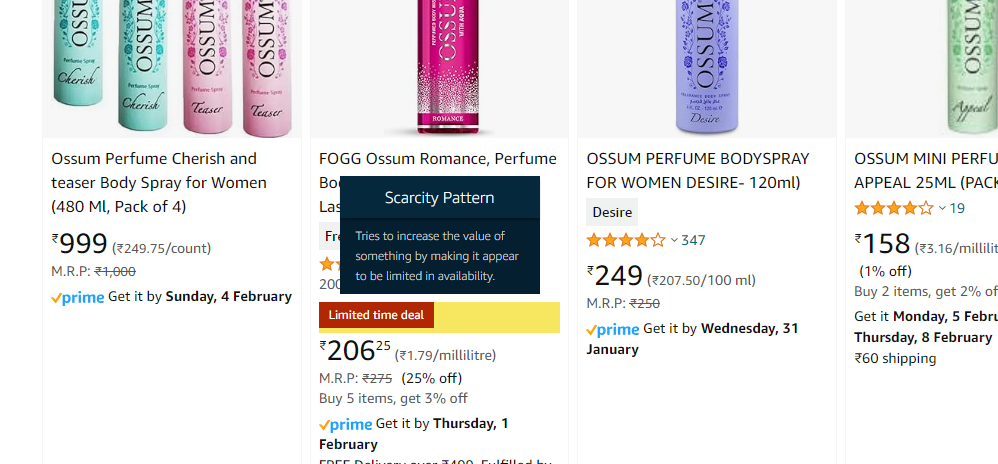
The below flowchart explains the workflow of Pattern Guard:



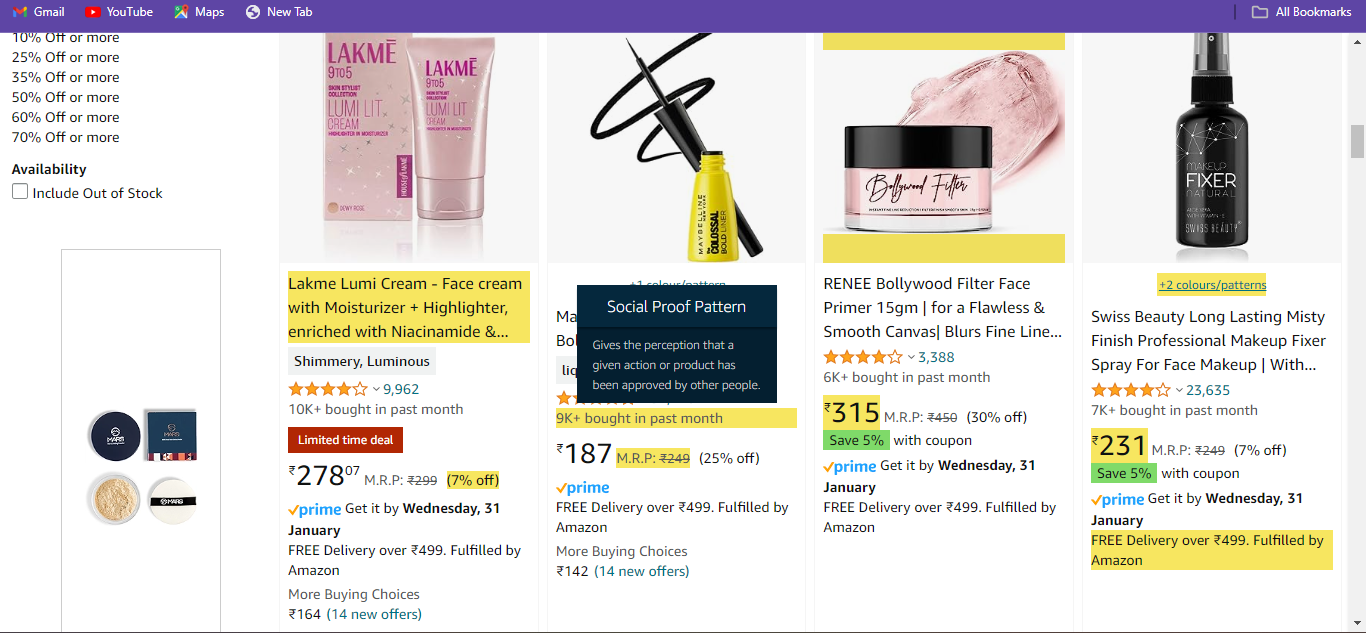
**Working of the extension:**



The extension tells us the number of dark patterns found in the particulate website



Warning regarding scarcity pattern generated by the extension after hovering over the highlighted part



Warning regarding social proof pattern generated by the extension after hovering over the highlighted part

**Business opportunities:**

1. Browser Extension Subscription Model:

- Offer a premium subscription model for the browser extension, providing advanced features and regular updates to users who are particularly concerned about online privacy.

2. API as a Service:

- Provide the dark pattern detection API as a service, allowing businesses to integrate the functionality into their own applications or websites.

3. White Labeling:

- Offer a white-label version of the tool that other businesses or organizations can rebrand and use as part of their services.

4. Collaboration with Browser Companies:

- Explore partnerships with major browser companies to integrate the tool as a built-in feature or extension, increasing its accessibility to a broader user base.

5. Crowdsourced Dark Pattern Database:

- Establish a platform where users can contribute and report new instances of dark patterns, creating a crowdsourced database that continuously improves the tool's accuracy.

**Future Scopes:**

1. Fine-tuning most of the popular open source LLMs like Mistral and Roberta.
2. User Feedback Integration:

Implement a feedback system allowing users to report and provide feedback on the accuracy of detected dark patterns. This data can be used to refine the models and improve detection algorithms.

1. Cross-Browser Compatibility:

Expand the compatibility of the browser extension to support multiple browsers, ensuring a broader user base and greater impact on the identification of dark patterns across different platforms.

1. Mobile Application Integration:

Develop a mobile application version of the tool, allowing users to identify and counteract dark patterns on websites while using their smartphones or tablets.

1. Gamification Elements:

Introduce gamification elements to encourage user engagement. For instance, users could earn points or rewards for actively countering dark patterns, fostering a community-driven approach.

1. API Marketplace:

Create an API marketplace, allowing developers to access and integrate the dark pattern detection API into a variety of applications, expanding its reach beyond browser extensions.

**Outcome:**

- The ultimate goal of our project is to empower users by raising awareness about potential dark patterns on websites.

- By providing clear warnings and categorizing the detected dark patterns, users can make informed decisions, fostering a more transparent and user-centric online experience.

In summary, our solution contributes to ongoing efforts to mitigate the impact of dark patterns on user trust and online interactions. Through the combination of advanced machine learning techniques and user-friendly interfaces, we aim to create a tool that not only detects but also educates users about deceptive design practices, ultimately promoting a safer and more transparent online environment.