**Abstract**

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.

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.

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.