# CMS Feature Extraction and Prediction Task

## Task Description

This task involves reading HTML files from a specified directory, extracting features based on predefined patterns for various CMS platforms, and predicting the CMS for each HTML file using a pre-trained model.

## Steps Involved

1. Read HTML File:   
 - Use a function to read the content of each HTML file from the specified directory.  
  
2. Extract Features:   
 - Implement a feature extraction method that analyzes the content of each HTML file and identifies specific patterns corresponding to various CMS platforms.  
  
3. Prepare Data:   
 - Convert the extracted features into a structured format, such as a DataFrame, ensuring it aligns with the model's expected input format.  
  
4. Predict CMS:   
 - Use the pre-trained model to predict the CMS for each HTML file based on the prepared data. The model processes the feature vector and outputs the predicted CMS.  
  
5. Save Results:   
 - Collect the results, which include the filenames and predicted CMS, and save them to a CSV file for further analysis.

## Results

The predictions have been saved to a CSV file named 'predictions.csv'. The results include the filename and the predicted CMS for each HTML file processed.

## Applicable CMS Platforms

Jekyll, Concrete CMS, Joomla, Ghost, Shopify, Hubspot, BigCommerce, Magento, Blogger, Drupal, WordPress, Typo3, 1-C Bitrix, Adobe Experience Manager, Butter CMS, Contao, Contentful, CraftCMS, Godaddy Website Builder, Grav CMS, Hugo, Jimdo, Liferay, Lithium, NeosCMS, PhpBB, Piwigo, Plone, Prismic, Process Wire, Progress Sitefinity, Sanity, Silverstripe, Sitecore, SMF, Squarespace, Storyblok, Strapi, Tilda, Umbraco, vBulletin, Webflow, Weebly, Wix, XenForo, Plesk, OpenCMS