|  |  |
| --- | --- |
| **Internet Advertisements Data Set** *Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/internet_ads/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/internet_ads/ad.DOCUMENTATION)  **Abstract**: This dataset represents a set of possible advertisements on Internet pages. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 3279 | **Area:** | Computer |
| **Attribute Characteristics:** | Categorical, Integer, Real | **Number of Attributes:** | 1558 | **Date Donated** | 1998-07-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | Yes | **Number of Web Hits:** | 335962 |

**Source:**

Creator & donor:  
  
Nicholas Kushmerick <nick **'@'** ucd.ie>

**Data Set Information:**

This dataset represents a set of possible advertisements on Internet pages. The features encode the geometry of the image (if available) as well as phrases occuring in the URL, the image's URL and alt text, the anchor text, and words occuring near the anchor text. The task is to predict whether an image is an advertisement ("ad") or not ("nonad").

**Attribute Information:**

(3 continous; others binary; this is the "STANDARD encoding" mentioned in the [Kushmerick, 99].)  
  
One or more of the three continous features are missing in 28% of the instances; missing values should be interpreted as "unknown".

**Relevant Papers:**

N. Kushmerick (1999). "Learning to remove Internet advertisements", 3rd Int Conf Autonomous Agents. Available at www.cs.ucd.ie/staff/nick/research/[[Web Link]](https://archive.ics.uci.edu/ml/datasets/download/kushmerick-aa99.ps.gz).  
[[Web Link]](http://rexa.info/paper/2fdc1cee89b7f4f2c9227d6f5d9b05d22c5ab3e9)

**Papers That Cite This Data Set1:**



Dmitriy Fradkin and David Madigan. [Experiments with random projections for machine learning](http://rexa.info/paper/1ddaa5b4eba8093faad88a5426ec586c6ac63f6e). KDD. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/Internet+Advertisements#1ddaa5b4eba8093faad88a5426ec586c6ac63f6e)].  
  
Sergio A. Alvarez and Takeshi Kawato and Carolina Ruiz. [Mining over loosely coupled data sources using neural experts](http://rexa.info/paper/f38cf28d327cab22e1f3c138ae648f6c6776a2d7). Computer Science Dept. Boston College. [[View Context](https://archive.ics.uci.edu/ml/support/Internet+Advertisements#f38cf28d327cab22e1f3c138ae648f6c6776a2d7)].  
  
Shay Cohen and Eytan Ruppin and Gideon Dror. [Feature Selection Based on the Shapley Value](http://rexa.info/paper/811517480cb8dca1073ee39a37c9a343a1179aab). School of Computer Sciences Tel-Aviv University. [[View Context](https://archive.ics.uci.edu/ml/support/Internet+Advertisements#811517480cb8dca1073ee39a37c9a343a1179aab)].

**Citation Request:**

Please refer to the Machine Learning Repository's [citation policy](https://archive.ics.uci.edu/ml/citation_policy.html)