A11.2020.13003_DM_4404

Turkish Headlines Dataset Data Set

Data Set Characteristics:	Text	**Number of Instances:**	4200	**Area:**	Computer
Attribute Characteristics:	N/A	**Number of Attributes:**	7	**Date Donated**	2021-04-14
Associated Tasks:	Classification, Clustering	**Missing Values?**	N/A	**Number of Web Hits:**	22445

Att1 class Att2 class Att3 class

source:

Zekeriya Anil Guven, anilguven1055 '@' gmail.com, Ege University, Department of Computer Engineering

Data Set Information:

Dataset consists of 7 news type labels. These labels are economy, politics, life, technology, magazine, health, sport. This dataset was created by me via Mynet, Milliyet, etc websites. There are 600 headlines for each label in the dataset . Hence, total headlines count is 4200 for dataset.

Attribute Information:

Attributes are economy, politics, life, technology, magazine, health, sport.

Relevant Papers:

GÃ¼ven, Z. A., Diri, B., & Ã‡akaloÇ§lu, T. (2018). Classification of New Titles by Two Stage Latent Dirichlet Allocation. Proceedings - 2018 Innovations in Intelligent Systems and Applications Conference, ASYU 2018. [Web Link]

Guven, Z. A., Diri, B., & Cakaloglu, T. (2019). Comparison of Topic Modeling Methods for Type Detection of Turkish News. UBMK 2019 - Proceedings, 4th International Conference on Computer Science and Engineering. [Web Link]

Citation Request:

Guven, Z. A., Diri, B., & Cakaloglu, T. (2019). Comparison of Topic Modeling Methods for Type Detection of Turkish News. UBMK 2019 - Proceedings, 4th International Conference on Computer Science and Engineering. [Web Link]

Secondary Mushroom Dataset Data Set

Download: Data Folder, Data Set Description

Abstract: Dataset of simulated mushrooms for binary classification into edible and poisonous.

Data Set Characteristics:	Univariate	**Number of Instances:**	61069	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	21	**Date Donated**	2021-04-11
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	36543

Source:

Donor: D. Wagner, dwagner93 '@' gmx.de
Product of bachelor thesis at Philipps-UniversitÃ¤t Marburg, Bioinformatics Division, supervised by Dr. G. Hattab.
Repository containing the related Python scripts and all the data sets: https://mushroom.mathematik.uni-marburg.de/files/
Inspired by the Mushroom Data Set of J. Schlimmer: url:https://archive.ics.uci.edu/ml/datasets/Mushroom.

Data Set Information:

The given information is about the Secondary Mushroom Dataset, the Primary Mushroom Dataset used for the simulation and the respective metadata can be found in the zip.

This dataset includes 61069 hypothetical mushrooms with caps based on 173 species (353 mushrooms
per species). Each mushroom is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended (the latter class was combined with the poisonous class).

The related Python project contains a Python module secondary_data_generation.py used to generate this data based on primary_data_edited.csv also found in the repository.
Both nominal and metrical variables are a result of randomization.
The simulated and ordered by species version is found in secondary_data_generated.csv.
The randomly shuffled version is found in secondary_data_shuffled.csv.

Attribute Information:

One binary class divided in edible=e and poisonous=p (with the latter one also containing mushrooms of unknown edibility).

Twenty remaining variables (n: nominal, m: metrical)
1. cap-diameter (m): float number in cm
2. cap-shape (n): bell=b, conical=c, convex=x, flat=f,
sunken=s, spherical=p, others=o
3. cap-surface (n): fibrous=i, grooves=g, scaly=y, smooth=s,
shiny=h, leathery=l, silky=k, sticky=t,
wrinkled=w, fleshy=e
4. cap-color (n): brown=n, buff=b, gray=g, green=r, pink=p,
purple=u, red=e, white=w, yellow=y, blue=l,
orange=o, black=k
5. does-bruise-bleed (n): bruises-or-bleeding=t,no=f
6. gill-attachment (n): adnate=a, adnexed=x, decurrent=d, free=e,
sinuate=s, pores=p, none=f, unknown=?
7. gill-spacing (n): close=c, distant=d, none=f
8. gill-color (n): see cap-color + none=f
9. stem-height (m): float number in cm
10. stem-width (m): float number in mm
11. stem-root (n): bulbous=b, swollen=s, club=c, cup=u, equal=e,
rhizomorphs=z, rooted=r
12. stem-surface (n): see cap-surface + none=f
13. stem-color (n): see cap-color + none=f
14. veil-type (n): partial=p, universal=u
15. veil-color (n): see cap-color + none=f
16. has-ring (n): ring=t, none=f
17. ring-type (n): cobwebby=c, evanescent=e, flaring=r, grooved=g,
large=l, pendant=p, sheathing=s, zone=z, scaly=y, movable=m, none=f, unknown=?
18. spore-print-color (n): see cap color
19. habitat (n): grasses=g, leaves=l, meadows=m, paths=p, heaths=h,
urban=u, waste=w, woods=d
20. season (n): spring=s, summer=u, autumn=a, winter=w

Relevant Papers:

Dennis Wagner, Dr. G. Hattab, 'Mushroom data creation, curation, and simulation to support classification tasks' in Scientific Reports on 14.04.2021

Citation Request:

If you have no special citation requests, please leave this field blank.