

GreenDB - A Dataset and Benchmark for Extraction of Sustainability Information of Consumer Goods

July 22, 2022

Sebastian Jäger, Alexander Flick, Jessica Adriana Sanchez Garcia, Kaspar von den Driesch, Karl Brendel, Felix Biessmann

GreenDB: A Product-by-Product Sustainability Database

- About 230k unique products (as of 2022-05-16)
- 4 online shops in 2 European countries
- Updated weekly

- Technical details on arxiv [1] or GitHub [2]
- Zenodo: <https://doi.org/10.5281/zenodo.6576662>

[1]: Jäger, S., Greene, J., Jakob, M., Korenke, R., Santarius, T., and Biessmann, F. GreenDB: Toward a Product-by-Product Sustainability Database. Technical Report arXiv:2205.02908, arXiv, May 2022. URL <http://arxiv.org/abs/2205.02908>. arXiv:2205.02908

[2]: <https://github.com/caloo-lab/green-db>

GreenDB: A Product-by-Product Sustainability Database

Products

- automated web-scraping

name	description	brand	category	color	merchant	price	sustainability_labels
MEL - T-shirt à manches longues - grey...	Wood Wood MEL - T-shirt à manc...	Wood Wood	SHIRT	grey melange/gris c...	zalando_fr	79.95	{certificate:GOTS_ORGANIC}
ASOS DESIGN - Jogger oversize à impri...	Jogger par ASOS Design. Un mod...	ASOS DESIGN	PANT	Myrica	asos	11.75	{certificate:BETTER_COTTON_IN...
6 PACK - Chaussettes - schwarz	SNOCKS 6 PACK - Chaussettes - ...	SNOCKS	UNDERW...	schwarz/noir	zalando_fr	22.49	{certificate:GOTS_MADE_WITH_...
Depot Tragetasche »Geschenktüten Niv...	Depot Tragetasche »Geschenktüt...	Depot	BAG	NULL	otto	1.99	{certificate:UNKNOWN}
COLLUSION Unisex - Jogger d'ensembl...	Pantalon de jogging par COLLUSI...	Collusion	PANT	Multicolore	asos	31.95	{certificate:BETTER_COTTON_IN...
FRONT AND BACK POCKETS PLACED P...	Marc O'Polo FRONT AND BACK P...	Marc O'Polo	TRACKS...	pistachio gray/vert	zalando_fr	109.95	{certificate:GOTS_ORGANIC}
BIO SOCKEN MID CUT 3 PACK - Socken...	Natural Vibes BIO SOCKEN MID C...	Natural Vibes	UNDERW...	white/weiß	zalando	24.95	{certificate:GOTS_MADE_WITH_...
Depot Tragetasche »Papiertüten Eva« (...	Depot Tragetasche »Papiertüten E...	Depot	BAG	NULL	otto	2.99	{certificate:UNKNOWN}
Jack & Jones Intelligence - Jogger slim...	Pantalon de jogging par Jack & Jo...	Jack & Jones	PANT	Gris chiné clair	asos	10.0	{certificate:BETTER_COTTON_IN...
ICE CREAM LINER 3 PACK UNISEX - So...	Happy Socks ICE CREAM LINER 3...	Happy Socks	UNDERW...	multi/mehrfarbig	zalando	11.0	{certificate:ORGANIC_CONTENT...

Sustainability Labels

- manually evaluated by experts



id	name	cred...	eco...	social...
certificate:BLUESIGN_APPROVED	bluesign® APPROVED	0	0	0
certificate:BLUESIGN_PRODUCT	Bluesign	79	81	65
certificate:FAIRTRADE_COTTON	Fairtrade - Baumwolle	91	84	79
certificate:FAIRTRADE_COTTON_PROGRAMM	Fairtrade-Baumwollprogramm	0	0	0
certificate:FAIRTRADE_INTERNATIONAL	Fairtrade International	80	64	92
certificate:FAIRTRADE_TEXTILE_PRODUCTION	Fairtrade Textile Production	93	64	92
certificate:GOTS_MADE_WITH_ORGANIC_MATERIA...	GOTS	89	89	87
certificate:GOTS_ORGANIC	GOTS Bio	89	89	87
certificate:ORGANIC_CONTENT_STANDARD_100	OCS 100	0	0	0
certificate:ORGANIC_CONTENT_STANDARD_BLEN...	Organic Content Standard - Blended	0	0	0

Predicting Sustainability Information

- Multi-class classification task: predict the products sustainability label
- *AutoGluon TabularPredictor* trained on the GreenDB
- Features: name, description, brand, category, color, merchant, and price

=> 96% weighted F1

=> 74% macro F1

	f1
BETTER_COTTON_INITIATIVE	0.91
BIORE	0.98
BLUESIGN_APPROVED	0.92
BLUESIGN_PRODUCT	0.96
COTTON_MADE_IN_AFRICA	0.97
CRADLE_TO_CRADLE_BRONZE	0.20
CRADLE_TO_CRADLE_GOLD	0.33
CRADLE_TO_CRADLE_SILVER	0.40
ECOCERT	0.79
EU_ECOLABEL_TEXTILES	0.95
FAIRTRADE_COTTON	0.90
FAIRTRADE_TEXTILE_PRODUCTION	0.89
GLOBAL_RECYCLED_STANDARD	0.89
GOOD_CASHMERE_STANDARD	0.00
GOTS_MADE_WITH_ORGANIC_MATERIALS	0.54
GOTS_ORGANIC	0.82
GREEN_BUTTON	0.92
LEATHER_WORKING_GROUP	0.93
MADE_IN_GREEN_OEKO_TEX	0.96
ORGANIC_CONTENT_STANDARD_100	0.67
ORGANIC_CONTENT_STANDARD_BLENDED	0.49
PRIVATE_LABEL	0.98
RECYCLED_CLAIM_STANDARD_100	0.61
RECYCLED_CLAIM_STANDARD_BLENDED	0.97
RESPONSIBLE_DOWN_STANDARD	0.81
RESPONSIBLE_WOOL_STANDARD	0.62
macro avg	0.74
weighted avg	0.96

Future Work

- Scale the GreenDB (shops, countries, categories)
- Integrate sustainable products without sustainability labels

- End-to-end automated process (without manually parse HTML files)
- Deduplicate/Joining same product from different sources

- Recommender or search engine to find sustainable alternatives for a given product



Thanks!

Mail: sebastian.jaeger@bht-berlin.de
Twitter: [@se-jaeger](https://twitter.com/se-jaeger)
LinkedIn: [se-jaeger](https://www.linkedin.com/in/se-jaeger)
GitHub: [se-jaeger](https://github.com/se-jaeger)
Web: sebastian-jaeger.me