

Table 1: A list of 28 datasets from 2007 until mid-2020 on eCommerce search and recommendation. The wide array of tasks and data types indicates the diversity and richness of the research challenges in the eCommerce domain.

Year	Dataset name and link	Data type	Task
2020	Rakuten multi modal taxonomy dataset <a href="https://sigir-ecom.github.io/data-task.html">https://sigir-ecom.github.io/data-task.html</a>	Catalog	Taxonomy classification Multimodal retrieval
2020	Semantic web embedded product data <a href="https://ir-ischool-uos.github.io/mwpd/index.html">https://ir-ischool-uos.github.io/mwpd/index.html</a>	Catalog	Product matching Product classification
2020	Alibaba eCommerce de-biasing dataset <a href="https://tianchi.aliyun.com/competition/entrance/231785/introduction">https://tianchi.aliyun.com/competition/entrance/231785/introduction</a>	Click logs	Fairness in exposure
2020	Alibaba multimodal recall <a href="https://tianchi.aliyun.com/competition/entrance/231786/introduction">https://tianchi.aliyun.com/competition/entrance/231786/introduction</a>	Catalog	Multimodal retrieval
2019	Alibaba user behavior dataset <a href="http://yongfeng.me/dataset/">http://yongfeng.me/dataset/</a>	Click logs	Recommendations
2019	DeepFashion <a href="https://github.com/switchablenorms/DeepFashion2">https://github.com/switchablenorms/DeepFashion2</a>	Images	Image segmentation Multimodal retrieval
2019	eBay high accuracy recall task <a href="https://sigir-ecom.github.io/ecom2019/data-task.html">https://sigir-ecom.github.io/ecom2019/data-task.html</a>	Search relevance	Product recall
2019	eBay ML data <a href="https://evalai.cloudcv.org/web/challenges/challenge-page/462">https://evalai.cloudcv.org/web/challenges/challenge-page/462</a>	Catalog	Product level equivalence
2018	Alibaba conversational search and recommendation dataset <a href="http://yongfeng.me/dataset/">http://yongfeng.me/dataset/</a>	Conversations	Conversational search and recommendations
2018	Olist Kaggle dataset <a href="https://www.kaggle.com/olistbr/brazilian-ecommerce">https://www.kaggle.com/olistbr/brazilian-ecommerce</a>	Catalog Transactions	Catalog mining Review mining
2018	Rakuten taxonomy dataset <a href="https://sigir-ecom.github.io/ecom2018/data-task.html">https://sigir-ecom.github.io/ecom2018/data-task.html</a>	Catalog	Taxonomy classification
2017	Amazon baby registry <a href="http://yongfeng.me/dataset/">http://yongfeng.me/dataset/</a>	Wishlists	Recommendations
2017	Fashion MNIST – Zalando <a href="https://github.com/zalando-research/fashion-mnist">https://github.com/zalando-research/fashion-mnist</a>	Images	Product classification
2017	Flipkart product catalog data <a href="https://www.kaggle.com/PromptCloudHQ/flipkart-products">https://www.kaggle.com/PromptCloudHQ/flipkart-products</a>	Catalog	Classification
2017	Innerwear dataset <a href="https://www.kaggle.com/PromptCloudHQ/innerwear-data-from-victorias-secret-and-others">https://www.kaggle.com/PromptCloudHQ/innerwear-data-from-victorias-secret-and-others</a>	Catalog	Catalog mining
2017	Lazada product title quality <a href="https://competitions.codalab.org/competitions/16652">https://competitions.codalab.org/competitions/16652</a>	Catalog	Product title grading
2017	Retail rocket dataset <a href="https://www.kaggle.com/retailrocket/ecommerce-dataset">https://www.kaggle.com/retailrocket/ecommerce-dataset</a>	User interactions	Recommendations
2016	Homedepot product search relevance <a href="https://www.kaggle.com/c/home-depot-product-search-relevance">https://www.kaggle.com/c/home-depot-product-search-relevance</a>	Search relevance	Search ranking
2016	Paper doll dataset <a href="https://github.com/kyamagu/paperdoll">https://github.com/kyamagu/paperdoll</a>	Images	Image segmentation Vision
2016	Personalized eCommerce search challenge <a href="https://competitions.codalab.org/competitions/11161">https://competitions.codalab.org/competitions/11161</a>	Search logs Transactions	Search personalization
2015	JD.com reviews and rating dataset <a href="http://yongfeng.me/dataset/">http://yongfeng.me/dataset/</a>	Catalog Ratings and reviews	Recommendations
2015	Online retail dataset of transactions <a href="http://archive.ics.uci.edu/ml/datasets/Online+Retail">http://archive.ics.uci.edu/ml/datasets/Online+Retail</a>	Transactions	Classification, clustering Customer segmentation
2015	Regio - Living Labs dataset <a href="https://bitbucket.org/living-labs/ll-api/src/master/">https://bitbucket.org/living-labs/ll-api/src/master/</a>	Search relevance	Search ranking
2015	Yoochoose dataset <a href="https://2015.recsyschallenge.com">https://2015.recsyschallenge.com</a>	Click logs	Recommendations
2014	Amazon product datasets <a href="https://cseweb.ucsd.edu/~jmcauley/datasets.html">https://cseweb.ucsd.edu/~jmcauley/datasets.html</a> <a href="https://snap.stanford.edu/data/#amazon">https://snap.stanford.edu/data/#amazon</a>	Catalog	Catalog mining Review mining Recommendations
2012	Best Buy Kaggle dataset <a href="https://www.kaggle.com/c/acm-sf-chapter-hackathon-big/data">https://www.kaggle.com/c/acm-sf-chapter-hackathon-big/data</a>	Search logs	Click prediction
2012	Online auctions dataset <a href="http://www.modelingonlineauctions.com/datasets">http://www.modelingonlineauctions.com/datasets</a>	Auctions	Bid price prediction
2007	Multi-domain sentiment dataset <a href="http://www.cs.jhu.edu/~mdredze/datasets/sentiment/">http://www.cs.jhu.edu/~mdredze/datasets/sentiment/</a>	Reviews	Sentiment mining