Table 1
Previously introduced ABSA datasets. The table includes the publication source, the size, number of aspects and domains for each dataset.

Publication source	Dataset name	Size	# aspects	Domains
Ganu et al. (2011) [29]	New York Restaurants	3400 sentences	6	Restaurants
Pontiki et al. (2014) [30]	SemEval-2014	7686 sentences	6	Restaurants, Laptops
Pontiki et al. (2016) [7]	SemEval-2016	5801 sentences	6	Restaurants, Laptops
Saeidi et al. (2016) [17]	SentiHood	5215 sentences	11	Neighbourhoods
Jiang et al. (2019) [27]	MAMS	8879 sentences	8	Restaurants
Lie et al. (2014) [31]	Twitter dataset	6940 tweets	118 (entities)	General domain
Alturaief et al. (2021) [32]	AWARE	11,323 sentences	12	Productivity, Social Networking, Games

Table 2
Size of each data source in the FABSA dataset in terms of number of reviews, sentences, unique orgs and uniqu industries that they contain.

	# reviews	# sentences	# unique orgs.	# unique industries
Trustpilot	1920	5174	3	3
Google Play	6130	9656	14	10
Apple App Store	2524	5716	12	10
Total	10,574	20,546	14	10

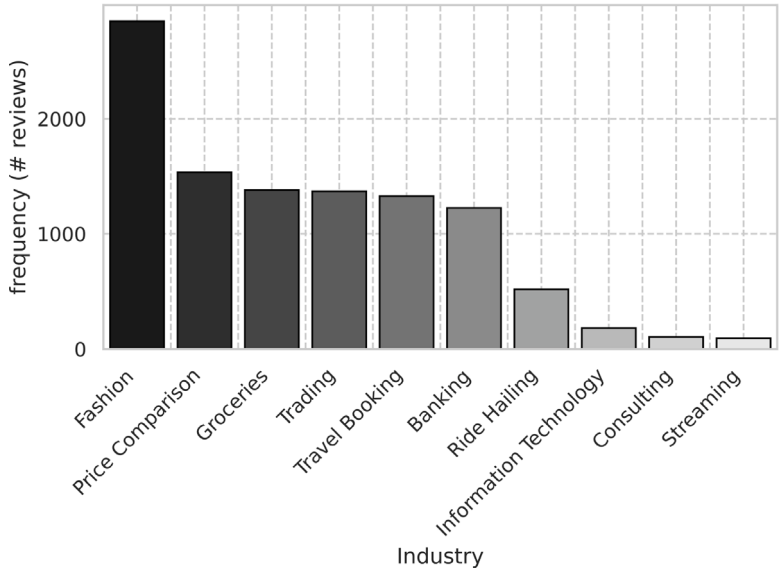

Fig. 1. Size of industries in terms of the number of reviews that they contain.

3.1. Data collection and pre-processing

The FABSA dataset is constructed from three different public data sources: Trustpilot,[2] Google Play[3] and Apple App Store.[4] Table 2 shows (a) the number of reviews, (b) the number of sentences, (c) the unique number of organisations, and (d) the unique number of industries per data source. Overall, we collected 10,574 feedback reviews from 14 organisations that span 10 different industries (e.g. banking, travel booking, fashion). The majority of the reviews are crawled from Google Play ($\sim 6.1K$ reviews) while Trustpilot and the Apple App Store cover $1.9 - 2.5K$ reviews, respectively.

Fig. 1 shows the size of each industry in terms of the number of reviews that they contain. Fashion (retail clothing) is the largest industry in our FABSA dataset, consisting of approximately 2.8K reviews, while Streaming is the smallest industry (94 reviews).

Regarding pre-processing, we mask organisation names with unique identifiers to avoid the identification of specific organisation names in our dataset. As an example, the review "*I love the Nike app!*" is converted into "*I love the ORGXX app!*" wherein the organisation name (i.e. *Nike*) is replaced with its corresponding unique identifier (*ORGXX*).

3.2. Annotation scheme

The FABSA dataset is manually labelled against a hierarchical annotation scheme which consists of 7 parent and 12 child aspect categories (Fig. 2). Each aspect category is associated with a sentiment label (positive, negative and neutral). This creates a total of 36 (12×3) target classification categories.

Following previous work, we adopt a multi-label classification scheme wherein each review is labelled with one or more aspect+sentiment label. Accordingly, a single review may contain multiple different aspects and express different (and in some cases contrasting) polarities. Table 3 shows an example of a review which is associated

[2] uk.trustpilot.com
[3] play.google.com
[4] www.apple.com/uk/app-store/