SemEval (Semantic Evaluation) merupakan sebuah rangkaian evaluasi dari computational semantic analysis system. SemEval 2014 Task 4 merupakan aspect based sentiment analysis yang terdiri dari 4 subtask yaitu aspect term extraction, aspect term polarity, aspect category detection, dan aspect category polarity. Detail in <https://alt.qcri.org/semeval2014/task4/>

Dataset:

Dataset file pada semeval 2014 task 4 terdiri dari dua domain yaitu laptop dan restaurant ada pada <https://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools>

Yang terdiri dari beberapa bagian

1. Laptop trial data: berisikan data mentah yang berisi seluruh kata kata pada ulasan yang diberikan pada domain laptop
2. Restaurant trial data: berisikan data mentah yang berisi seluruh kata kata pada ulasan yang diberikan pada domain restaurant
3. Laptops train data: dapat diakses pada <http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-laptop-reviews-train-data/94748ff4624e11e38d18842b2b6a04d7ca9201ec33f34d74a8551626be122856/>

Data laptops ini terdiri dari 3048 kalimat dalam Bahasa inggris yang diambil dari ulasan pelanggan tentang laptop. Annotator merupakan ornag yang berpengalaman dibidangnya yang menandai untuk aspek kalimat dan polaritasnya.

Format text: xml

1. Restaurant train data: dapat diakses pada

<http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-restaurant-reviews-train-data/479d18c0625011e38685842b2b6a04d72cb57ba6c07743b9879d1a04e72185b8/>

Dataset ini terdiri dari 3044 kalimat dalam Bahasa inggris yang merupakan review dari Ganu et al. (2009). Dataset asli dari Ganu sudah termasuk anotasi untuk aspek kategori dan polaritas kalimat secara keseluruhan. Kemudian dataset tersebut dimodifikasi dengan menandai aspect term pada kalimat, aspect term polarities dan aspect category-specific polarities. Anotasi juga dilakukan oleh orang yang berpengalaman dibidangnya.

Format text: xml

1. Train data (Laptops & Restaurants) V2.0

Dapat diakses pada: <http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-train-data-v20-annotation-guidelines/683b709298b811e3a0e2842b2b6a04d7c7a19307f18a4940beef6a6143f937f0/>

Merupakan dataset yang terdiri dari dua domain yaitu Laptops dan restaurants. Restaurant\_Train\_v0.2 terdiri dari 3041 kalimat dalam Bahasa inggris dari ulasan restaurant dari Ganu et al. dimana dataset aslinya sudah dimodifikasi oleh annotator berpengalaman

Laptop\_Train\_v0.2 terdiri dari 3045 kalimat dalam Bahasa inggris yang diambil dari ulasan pelanggan dari laptop.

Text format: xml

1. Test data: Phase A

Data ini terdiri dari dua domain:

* Restaurants\_Test\_Data\_PhaseA.xml terdiri dari 800 kalimat berbahasa inggris yang diambil dari ulasan restaurant dan akan digunakan untuk subtask 1 (aspect term extraction) dan 3 (aspect category extraction)
* Laptops\_Test\_Data\_PhaseA.xml terdiri dari 800 kalimat berbasa inggris yang diambil dari ulasan pelanggan tentang laptopdan akan digunakan hanya untuk subtask 1 (aspect term extraction) dari phase A ABSA.

1. Test data: Phase A

Data ini terdiri dari dua domain:

* Restaurants\_Test\_Data\_PhaseA.xml terdiri dari 800 kalimat berbahasa inggris yang diambil dari ulasan restaurant dan akan digunakan untuk subtask 2 (aspect term polarity) dan 4 (aspect category polarity)
* Laptops\_Test\_Data\_PhaseA.xml terdiri dari 800 kalimat berbasa inggris yang diambil dari ulasan pelanggan tentang laptopdan akan digunakan hanya untuk subtask 2 (aspect term polarity) dari phase A ABSA.