

# Introduction to Data Science and Analytics

Group Number 1 / Delivery 2

17 April 2019

Table 1: Group Members

Group Members			
<i>Group Member Name</i>	<i>Number</i>	<i>Department</i>	<i>Semester</i>
Şevvalnur Kahraman	150116822	Computer Engineering	4
Nurcihane Köroğlu	150115062	Computer Engineering	4
Emine Feyza Memiş	150114077	Computer Engineering	4
Gizem Aybüke Çeri	150315060	Industrial Engineering	4
Yiğit Hallaç	150714036	Electronics Engineering	4
Amine Boumusou	199518004	Computer Engineering	4

This dataset contains sentences labelled with positive or negative sentiment, extracted from reviews of products, movies, and restaurants with their format like "sentence score".

Score is either 1 (for positive) or 0 (for negative) and each sentences come from three different websites/fields that are imdb.com, amazon.com and yelp.com. For each website, there exist 500 positive and 500 negative sentences. Those were selected randomly for larger datasets of reviews. We attempted to select sentences that have a clearly positive or negative connotation, the goal was for no neutral sentences to be selected.

If you want to see the full datasets you can look for the followings:  
**imdb:** Maas et. al., 2011 'Learning word vectors for sentiment analysis'  
**amazon:** McAuley et. al., 2013 'Hidden factors and hidden topics: Understanding rating dimensions with review text'  
**yelp:** Yelp dataset challenge [http://www.yelp.com/dataset\\_challenge](http://www.yelp.com/dataset_challenge)

Table 2: Features/Attributes

<b>Features/Attributes</b>				
<i>Number</i>	<i>Features</i>	<i>Description</i>	<i>Type</i>	<i>Target</i>
1	Sentence	It contains sentences come from three different websites/fields	String	No
2	Class	It contains score is either 1 (for positive) or 0 (for negative)	Numeric	Yes

Table 3: Class Attribute Values

<b>Class Attribute Values</b>			
<i>Number</i>	<i>Class Value</i>	<i>Number of Instances</i>	<i>Description</i>
1	0	1500	It contain negative sentences
2	1	1500	It contains positive sentences