

Attribute Information:

1...50: Average, standard deviation, min, max and median of the Attributes 51...60 for the source of the current blog post With source we mean the blog on which the post appeared. For example, myblog.blog.org would be the source of the post myblog.blog.org/post_2010_09_10

51: Total number of comments before basetime

52: Number of comments in the last 24 hours before the basetime

53: Let T1 denote the datetime 48 hours before basetime, Let T2 denote the datetime 24 hours before basetime. This attribute is the number of comments in the time period between T1 and T2

54: Number of comments in the first 24 hours after the publication of the blog post, but before basetime

55: The difference of Attribute 52 and Attribute 53

56...60: The same features as the attributes 51...55, but features 56...60 refer to the number of links (trackbacks), while features 51...55 refer to the number of comments.

61: The length of time between the publication of the blog post and basetime

62: The length of the blog post

63...262: The 200 bag of words features for 200 frequent words of the text of the blog post

263...269: binary indicator features (0 or 1) for the weekday (Monday...Sunday) of the basetime

270...276: binary indicator features (0 or 1) for the weekday (Monday...Sunday) of the date of publication of the blog post

277: Number of parent pages: we consider a blog post P as a parent of blog post B, if B is a reply (trackback) to blog post P.

278...280: Minimum, maximum, average number of comments that the parents received

281: The target: the number of comments in the next 24 hours (relative to basetime)

