

# Data Discrepancies

There's more data used for the hedometer evaluation than there is for the ml training and evaluation.

- [The stats](#)

# The stats

The Hedometer data has 3656 entries The ML data has 3343 enttries.

The preprocessing done on the hedometer data makes it difficult to compare the two files.  
Lowercasing, all punctuation removed, spaced injected.