

MINI PROJECT – I
(2018-19)

<Amazon Review>
SYNOPSIS



Institute of Engineering & Technology

Team Members

Tushar Gupta
(171500360)

Apoorvi Goyal
(171500060)

Sarthak Jain
(171500294)

Supervised By

Name

Vinay Agarwal

Department of Computer Engineering & Applications

About the Project: (Minimum 100 words)

The problem being addressed in this project is the poor quality of Amazon reviews at the top of the forum despite the helpfulness rating system. The solution this problem is to pre-rate the reviews as "helpful" or "not helpful" for new reviews and the most helpful reviews are given position at the top. This way, poor quality reviews will be more unlikely to be shown at the top.

This is a binary classification of reviews on the basis of helpfulness as helpful or not helpful. The models will be trained on the basis of features extracted from the reviews. The recent reviews which are not rated as helpful by the users will be of poor quality but are still on top, we tried to make the helpful review to be at the top of the reviews shown to the buyer.

Motivation: (Minimum 30 words)

The reviews are particularly used so that new users can see the rated comments in order to help them make their own purchasing decisions. The quality of the comments made on internet forums has always suffered due to its users. Different websites have tried different methods for extracting more useful comments so that users can see top rated comments.

Future Prospects:

Our future workr will be improving the accuracy of the models implemented. In order to improve the accuracy of our model, the following ac-

tions could be taken:

- There are lots of spelling mistakes in the reviews and abbreva-**
tions are used. This would result in a model that potentially
has less features, as certain spelling errors would have been
corrected/eliminated.
- Reviews with poor grammar, punctuation and improper word**
endings are more difficult to understand and would possibly
lead to people rating them as less helpful. These reviews may
contain useful informations and may be helpful but certain
other factors made them rated as not helpful. Thus preprocess-
ing of data can be improved either by varying the values of
threshold or the reviews that has not been selected because the
product is not much rated.

Requirements:

a) Hardware:

- 8 GB RAM**
- intel i7 7th generation**
- 1 TB ROM**

b) Software:

- **Jupiter notebook**
- **Windows 10**
- **MS Excel**