World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:14, No:07, 2020

An Embarrassingly Simple Semi-supervised Approach to Increase Recall in Online Shopping Domain to Match Structured Data with Unstructured Data

Authors : Sachin Nagargoje

Abstract : Complete labeled data is often difficult to obtain in a practical scenario. Even if one manages to obtain the data, the quality of the data is always in question. In shopping vertical, offers are the input data, which is given by advertiser with or without a good quality of information. In this paper, an author investigated the possibility of using a very simple Semi-supervised learning approach to increase the recall of unhealthy offers (has badly written Offer Title or partial product details) in shopping vertical domain. The author found that the semisupervised learning method had improved the recall in the Smart Phone category by 30% on A=B testing on 10% traffic and increased the YoY (Year over Year) number of impressions per month by 33% at production. This also made a significant increase in Revenue, but that cannot be publicly disclosed.

Keywords : semi-supervised learning, clustering, recall, coverage
Conference Title : ICDM 2020 : International Conference on Data Mining
Conference Location : Dublin, Ireland
Conference Dates : July 30-31, 2020

Open Science Index, Computer and Systems Engineering Vol:14, No:07, 2020 publications.waset.org/abstracts/125267.pdf