

**0.46%** PLAGIARISM
APPROXIMATELY1.03% IN QUOTES q

Report #14194917

INTRODUCTION Background Nowadays, the amount of news has increased every day. They also grow rapidly. This situation makes it difficult for the editor to categorize manually and with this manual categorization makes the categories incorrect for the news. They also wasted the time choosing the category for the news which were published. Even if there is automatic news categorization, the prediction is less accurate. It happens because many categorizations use supervised learning that the dataset is labeled manually with people that spend a lot of time and money. For that problem, people just only classify small partial data and the rest should be done by the computer. So, it can reduce the time and the money for labeling the dataset. This method calls semi-supervised learning with the method pseudo labeling or in the theory of the book called self-training. Self-training does with the only partial dataset labeled by the human and then the classifier learns from that partial dataset and does the prediction which it appends on the

REPORT
#14194917CHECKED
7 DEC 2021, 9:11 AMAUTHOR
ANDRE KURNIAWANPAGE
1 OF 49

training set for the next prediction (real prediction).
Self-training also does with the model or classifier that can be taken from supervised learning. In this research, we proposed news categories classification using machine learning