

ABSTRACT

In a highly developed era like today, the benefits of the internet are very useful in all circles to get information and add insight. All the desired information can be accessed very easily, every day billions of people in the world access it every time. one of the very popular search engines in the world of sata is google. Google has a popular channel for researchers to publish their research results, namely Google Scholar. Google Scholar is made to make it easier for users to search for or browse material and reference education and scientific research in a publication format. This reference can be in the form of scientific journals which are the results of research, research or studies that can be accounted for scientifically.

In this final project, a research that will use the K-Nearest Neighbor method is conducted. This method is a clustering algorithm or grouping of data that is supervised learning. The function of this algorithm is to group data into clusters. Data that will be grouped by the K-Nearest Neighbor method will be crawled first from Google Scholar, then the data in the contents of the file will be processed by text mining so that it can generate cloudwords from keywords in the document.

Data from Google Scholar was successfully crossed with the contents of the document in the form of pdf, then the document will be inputted into the R shiny application so as to produce a word representation of all documents and the most abstract keywords appearing successfully displayed in password. The results of the percentage accuracy of all documents using the K-Nearest Neighbor method in this Final Project is 99%.

Keyword: *Google Scholar, K-Nearest Neighbor, text mining, crawling, clustering*