

## ABSTRAK

Pada zaman yang sangat berkembang seperti sekarang ini, manfaat internet sangat berguna disemua kalangan untuk mendapatkan informasi serta menambah wawasan. Segala informasi yang diinginkan dapat diakses dengan sangat mudah, setiap harinya miliaran orang di dunia mengaksesnya setiap waktu. Salah satu mesin pencarian yang sangat populer di dunia sata ini adalah google. Google memiliki *channel* yang populer bagi kalangan peneliti untuk mempublikasikan hasil penelitiannya, yaitu *Google Scholar*. *Google Scholar* dibuat untuk mempermudah pengguna mencari atau menelusuri materi maupun referensi pendidikan dan penelitian ilmiah dengan format publikasi. Referensi ini dapat berupa jurnal ilmiah yang merupakan hasil-hasil penelitian, riset atau kajian yang dapat dipertanggung jawabkan secara ilmiah.

Pada tugas akhir ini, dilakukan sebuah penelitian yang akan menggunakan metode *K-Nearest Neighbor*. Metode ini adalah salah satu algoritma *clustering* atau pengelompokan data yang bersifat *supervised learning*. Fungsi dari algoritma ini adalah mengelompokkan data kedalam beberapa *cluster*. Data yang akan di kelompokkan dengan metode *K-Nearest Neighbor* sebelumnya akan dicrawling terlebih dahulu dari *Google Scholar*, lalu data yang didapat isi file-nya akan diolah dengan *text mining* sehingga dapat menghasilkan *cloudword* dari *keyword* yang ada di dalam dokumen tersebut.

Data dari *Google Scholar* berhasil dicrawling dengan isi dokumen berbentuk pdf, lalu dokumen tersebut akan diinputkan ke dalam aplikasi R *shiny* sehingga menghasilkan representasi kata dari seluruh dokumen dan kata kunci dari abstrak yang paling banyak keluar berhasil ditampilkan dicloudword. Hasil persentasi keakuratan dari seluruh dokumen dengan menggunakan metode *K-Nearest Neighbor* dalam Tugas Akhir ini adalah 99%.

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