



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB 3

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Penelitian “Implementasi Algoritma *Support Vector Machine* (SVM) Dalam Menentukan Kategori Berita” menggunakan beberapa tahapan dalam implementasi metodologi dan desain sistem. Tahapan penelitian ini dapat dibagi sebagai berikut:

1. Observasi

Observasi dilakukan dengan cara melakukan pengamatan terhadap topik penelitian yang diambil dan mencatat seluruh informasi yang didapat tentang keadaan atau situasi yang terjadi pada masa sekarang.

2. Studi Literatur

Studi literatur dilakukan dengan mencari, mengumpulkan, membaca, Meneliti, memahami dan mengolah dari buku, jurnal ilmiah dan Karya ilmiah lainnya menjadi acuan teoritis untuk penelitian ini. Proses ini bertujuan untuk meningkatkan pemahaman dan mengungkap teori-teori yang berkaitan dengan Text Classification dan Support Vector Machine (SVM), sehingga dapat memperkaya landasan teori dari penelitian.

3. Pengumpulan data

Data akan dikumpulkan dari *dataset* dengan masing-masing kategori berbeda, sehingga dapat menguji tingkat akurasi dan ketepatan dari aplikasi yang akan dirancang. *Dataset* yang digunakan berisi kumpulan dari judul berita yang diperoleh dari portal berita detik.com.

4. Perancangan Aplikasi

Proses perancangan untuk membuat website dimulai dengan membuat *use case diagram*, *data flow diagram* dan rancangan untuk *user interface* dari system atau aplikasi yang akan dibangun.

5. Implementasi

Tahap ini akan dimasukkan hasil penerapan metode pada sistem yang sudah selesai dan hasil perhitungan menggunakan *Text Classification*, *SVM*, metode ekstraksi fitur TF-IDF, dan metode *Confusion Matrix* yang akan menghitung akurasi yang tepat dalam menentukan kategori berita.

6. Pengujian Aplikasi

Proses pengujian aplikasi dilakukan pengujian terhadap sistem yang sudah dibuat, melakukan validasi terhadap hasil pengujian aplikasi serta memperbaiki aplikasi ketika terdapat kesalahan pada saat uji coba aplikasi.

7. Pengukuran performa klasifikasi

Pengukuran akan dilakukan dengan seberapa besar tingkat keberhasilan dari metode ekstraksi fitur TF-IDF yang digunakan dalam proses klasifikasi. Selain itu akan dibandingkan dengan berita yang diambil dari portal berita.

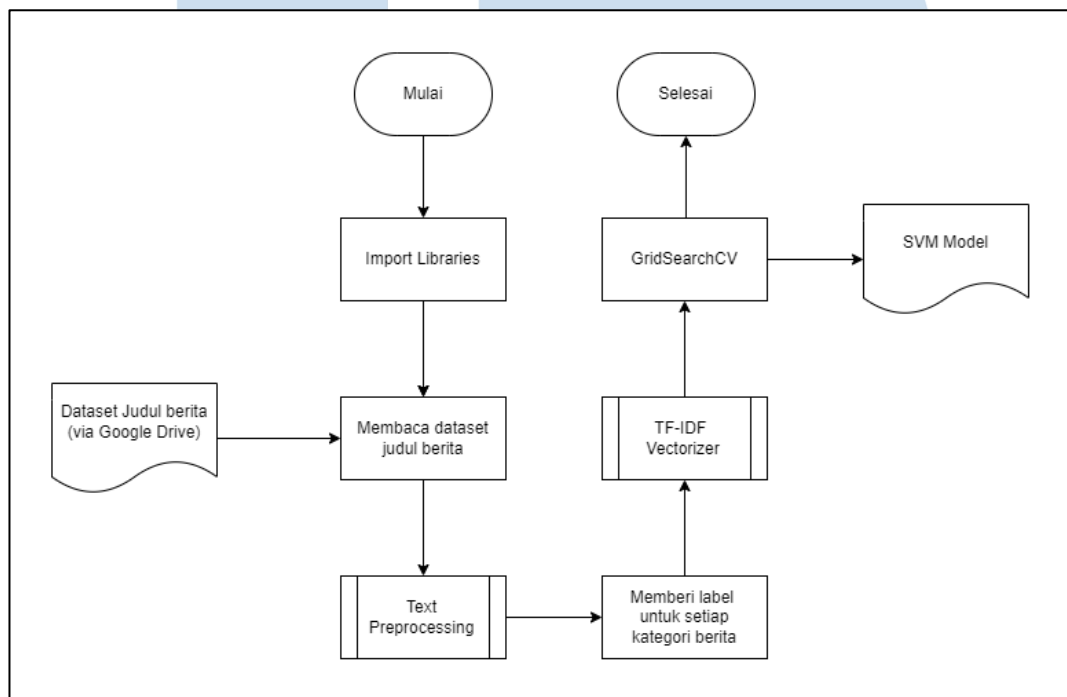
8. Penulisan Laporan

Laporan dibuat sebagai dokumentasi dari proses penelitian, perancangan dan pembuatan aplikasi. Laporan dibuat dari bagian pendahuluan hingga kesimpulan dan saran.

3.2 Perancangan Aplikasi

Perancangan aplikasi dibagi dalam beberapa bagian, yaitu *flowchart* aplikasi pembuatan model SVM, *flowchart* aplikasi *website*, dan rancangan tampilan antarmuka untuk aplikasi *website*.

3.2.1 Flowchart Aplikasi Pembuatan Model SVM

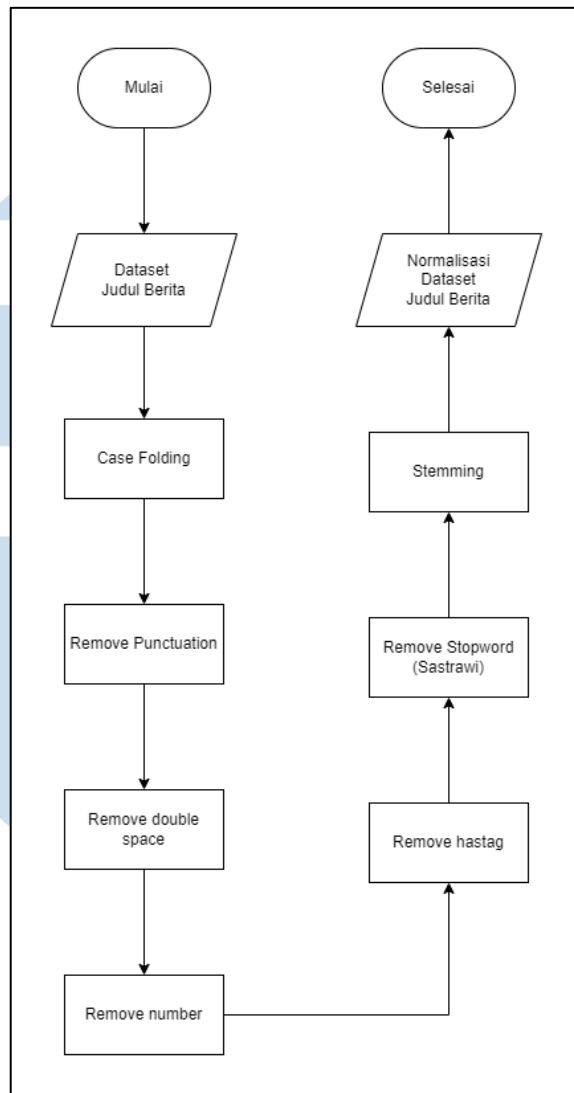


Gambar 3.1 Flowchart Pembuatan Model SVM

Gambar 3.1 merupakan *flowchart* proses pembuatan SVM Model yang akan digunakan untuk klasifikasi judul berita. Hal pertama yang dilakukan dalam proses ini adalah meng-*import* seluruh *library* antara lain TF-IDF, SVM, *pandas*, dan sebagainya yang akan digunakan dalam proses *training dataset*, kemudian *dataset* judul berita akan dimasukkan dan dibaca dalam aplikasi. *Dataset* dalam bentuk file *excel* akan dibaca dan selanjutnya akan melewati proses *text processing*. Hasil dari *text processing* tadi akan diberikan label untuk masing-masing kategori. Terdapat 6 kategori dalam proses pemberian label diantaranya 0 untuk kategori berita

finance, 1 untuk kategori berita *food*, 2 untuk kategori berita *health*, 3 untuk kategori berita otomotif, 4 untuk berita *sport*, dan 5 untuk kategori berita *travel*. Setelah diberikan label, selanjutnya *dataset* akan masuk pada *TF-IDF Vectorizer* yang merupakan proses merubah *dataset* tadi menjadi *vector* sebelum diklasifikasikan menggunakan model SVM. Setelah menjadi nilai *vector* maka *dataset* akan di-*split* seberapa besar data akan dipakai untuk *training* (ukuran data yang dipakai untuk latihan), dan data yang dipakai untuk *test* (ukuran data yang akan digunakan untuk percobaan) menggunakan *GridSearchCV*. Dalam penelitian ini *train size* yang digunakan adalah 80% atau 0.8, sedangkan untuk *test size* yang digunakan adalah 20% atau 0.2.

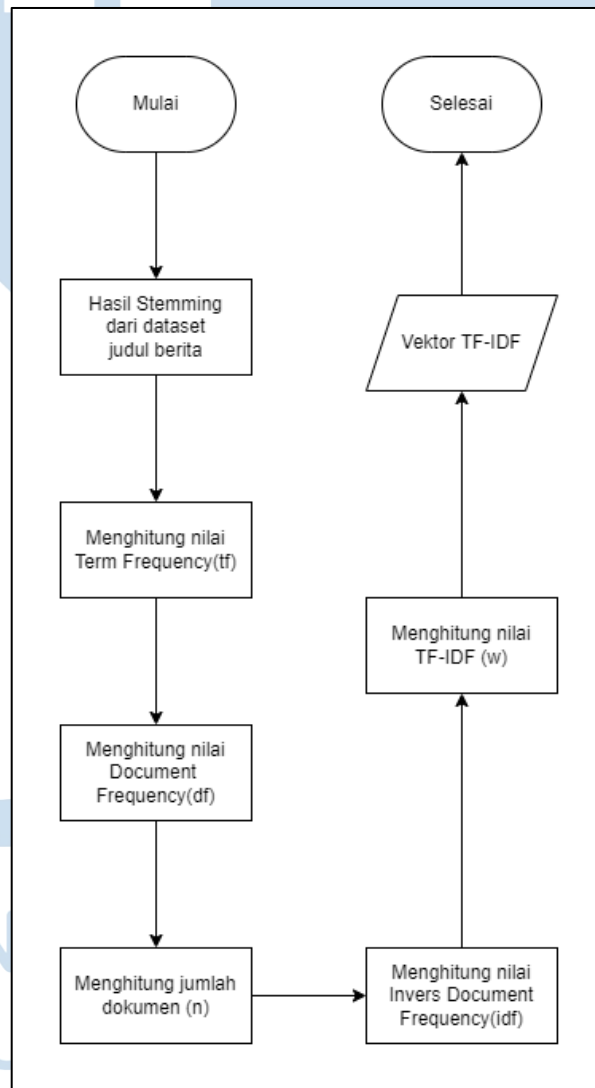
Data yang telah dibagi selanjutnya akan dilatih menggunakan model SVM (Support Vector Machine). Terdapat beberapa nilai dalam proses pelatihan model SVM diantaranya nilai *C* yang adalah nilai optimasi SVM untuk menghindari kesalahan dalam proses pelatihannya, kemudian ada nilai *kernel* yang berfungsi mentransformasikan data ke ruang dimensi yang lebih tinggi, dan nilai *degree* untuk mengontrol fleksibilitas batas keputusan. Setelah proses pelatihan selesai dilakukan, evaluasi proses pelatihan model SVM akan diukur agar memperoleh informasi nilai *F1-Score*. Selanjutnya proses pencarian nilai parameter terbaik akan dilakukan untuk dapat mengetahui nilai terbaik yang bisa dipakai setiap parameter dalam model SVM, sehingga memperoleh prediksi yang akurat. Setelah semuanya selesai maka model SVM akan di-*export* dan disimpan untuk digunakan untuk proses klasifikasi.



Gambar 3.2 Flowchart Text Preprocessing

Gambar 3.2 merupakan gambaran dari beberapa proses *preprocessing* yang dilakukan pada *dataset*. Awalnya *dataset* judul berita akan diambil dan melewati proses *Case Folding* yang bertujuan untuk mengubah semua huruf dalam *dataset* menjadi huruf kecil. Setelah itu, akan melewati proses *Remove Punctuation* yang berfungsi untuk menghilangkan semua tanda baca yang ada dalam *dataset*. Selanjutnya akan melewati proses *Remove double space* memiliki fungsi untuk menghilangkan spasi ganda menjadi spasi tunggal apabila terjadi kesalahan pada pendataan *dataset* yang digunakan. Berikutnya akan dilakukan *Remove number*

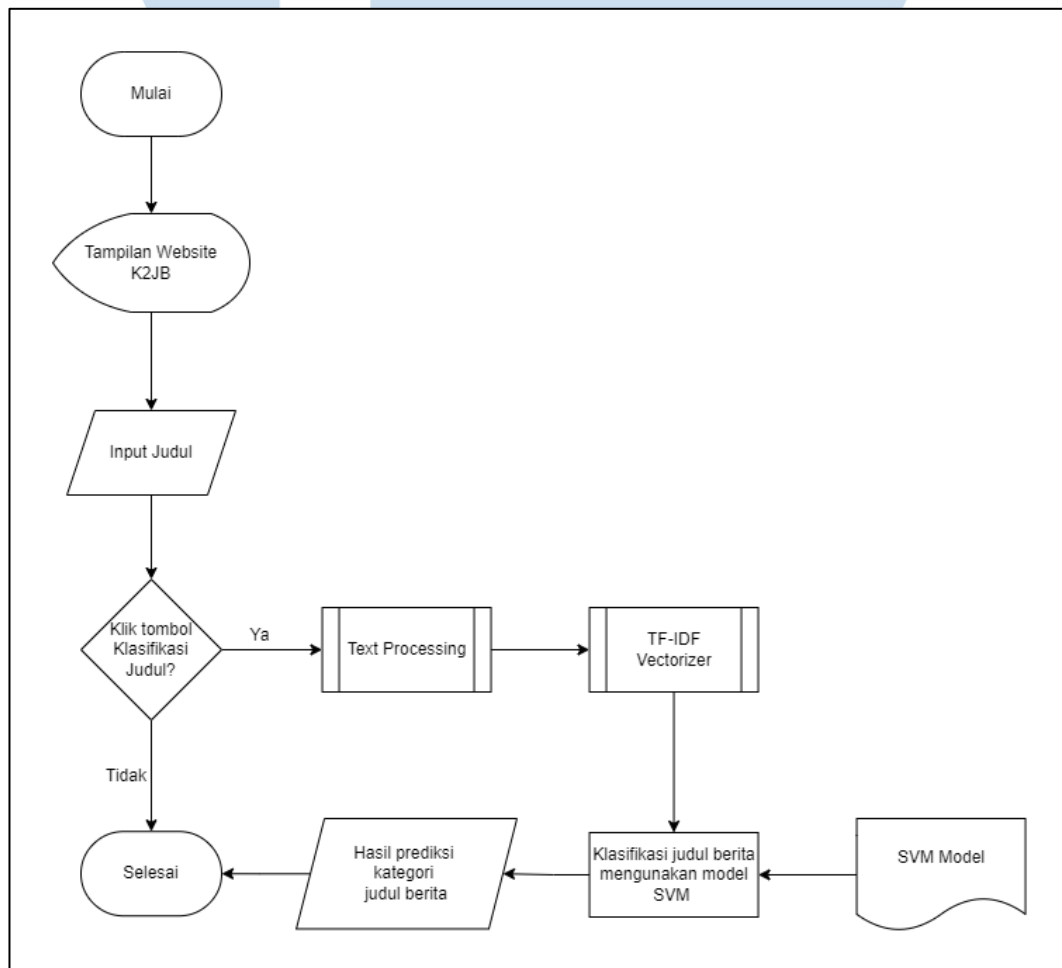
untuk menghilangkan semua angka, dan *Remove hashtag* untuk menghilangkan tagar yang ada pada judul berita. Kemudian dilakukan *Remove Stopword* untuk menghilangkan kata umum yang biasa muncul dan dianggap tidak memiliki makna. Setelah semua selesai, akan dilakukan *Stemming* untuk mencari kata dasar dari setiap kata dalam dokumen dengan membuang imbuhan dan menghasilkan *dataset* yang telah dinormalisasi.



Gambar 3.3 Flowchart TF-IDF

Gambar 3.3 merupakan *flowchart* proses tahapan yang dilewati *dataset* sebelum diubah kedalam bentuk *vector* yang selanjutnya akan diklasifikasikan menggunakan model SVM. Pada awal proses hasil *Stemming* dari *dataset* judul berita akan digunakan dalam perhitungan diantaranya menghitung nilai *Term Frequency*, menghitung nilai *Document Frequency*, menghitung jumlah dokumen, menghitung nilai *Invers Document Frequency*, dan menghitung nilai TF-IDF. Setelah selesai maka akan menghasilkan vektor TF-IDF yang akan digunakan untuk proses klasifikasi selanjutnya.

3.2.2 Flowchart Aplikasi Web

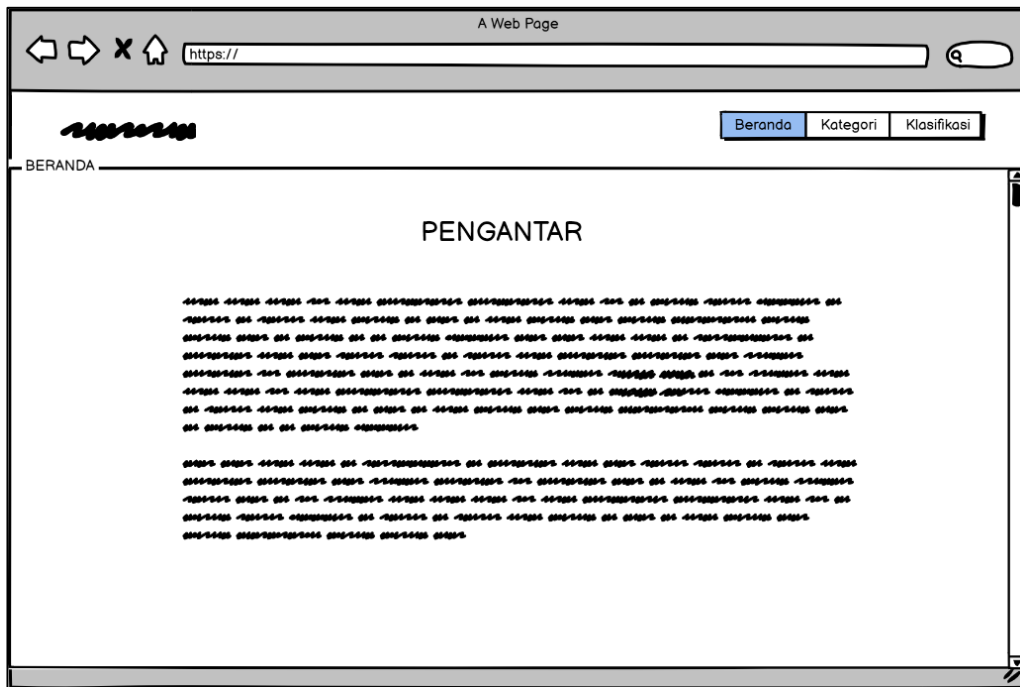


Gambar 3.4 Flowchart Aplikasi Web

Gambar 3.4 adalah gambaran *flowchart* proses dari aplikasi *web* yang digunakan untuk melakukan klasifikasi menggunakan SVM model. Pada saat *user* mengakses *website* Klasifikasi Kategori Judul Berita untuk pertama kali, *user* akan diarahkan ke halaman utama yang berisi beberapa informasi terkait aplikasi *website*, diantaranya fungsi dan kategori apa saja yang dapat diklasifikasikan pada *website* Klasifikasi Kategori Judul Berita. Untuk melakukan klasifikasi kategori judul berita, *user* dapat meng-*input* kata pada kolom yang telah disediakan pada halaman klasifikasi. Ketika *user* menekan tombol klasifikasi judul, maka proses klasifikasi akan dijalankan dengan melewati proses *text processing*, *tf-idf vectorizer* dan klasifikasi menggunakan SVM model untuk menentukan judul berita yang di-*input user* masuk dalam kategori mana. Setelah proses selesai, maka *website* akan memberikan *feedback* berupa label kategori dari hasil klasifikasi judul berita.

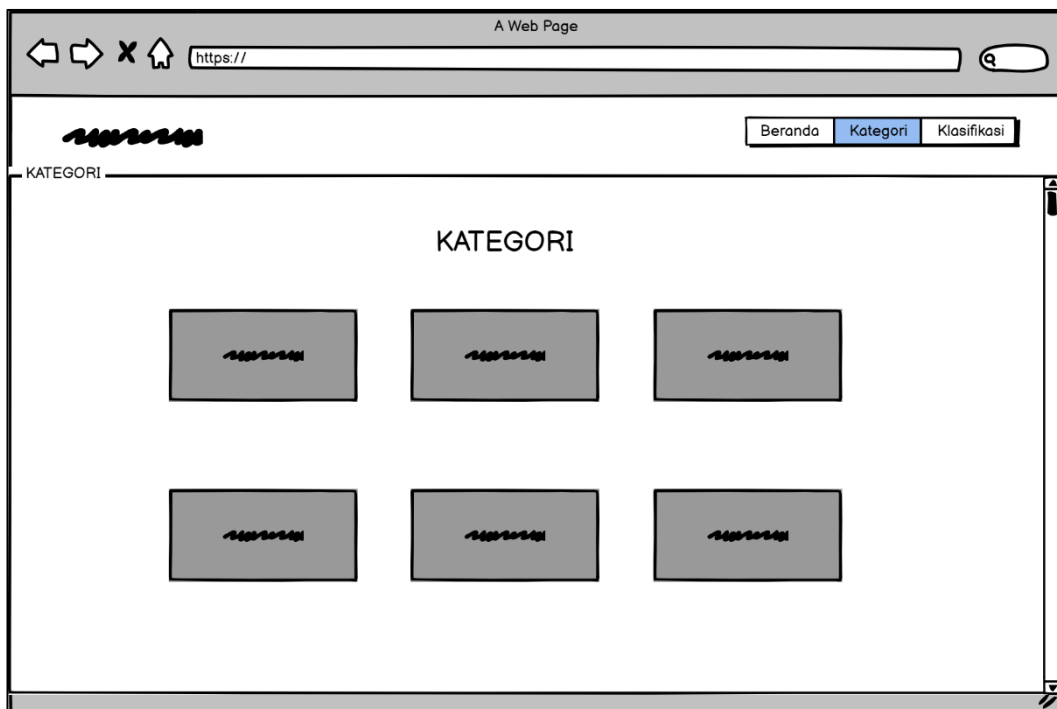
3.2.3 Rancangan Antarmuka Website

Terdapat beberapa rancangan antarmuka dari *website* Klasifikasi Kategori Judul Berita diantaranya adalah rancangan halaman beranda, rancangan halaman kategori, dan halaman klasifikasi.



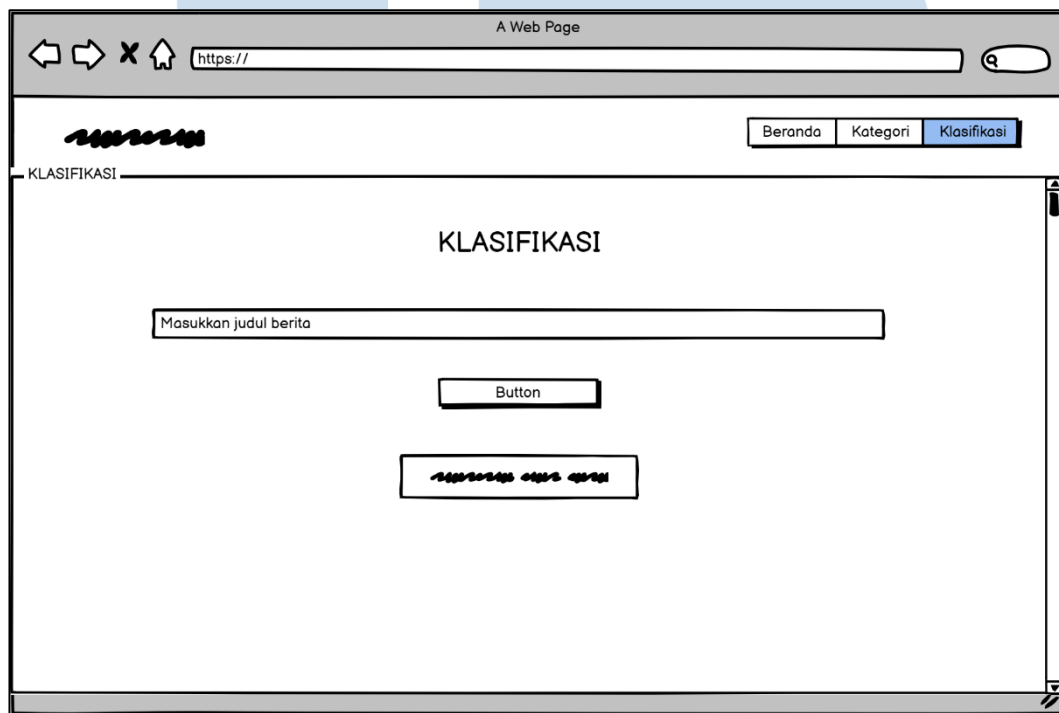
Gambar 3.5 Halaman Beranda

Gambar 3.5 merupakan rancangan antarmuka halaman beranda yang berisi menjelaskan secara singkat fungsi atau kegunaan dari *website* Klasifikasi Kategori Judul Berita.



Gambar 3.6 Halaman Kategori Judul Berita

Gambar 3.6 adalah rancangan halaman kategori yang dapat diklasifikasikan pada aplikasi berbasis web ini. Terdapat beberapa kategori yang ditampilkan pada halaman ini, seperti *finance*, *food*, *health*, otomotif, *sport*, dan *travel*. Ketika *user* menekan tulisan pada kategori yang ada pada halaman diatas, maka akan muncul *modal* yang berisi penjelasan singkat terkait kategori.



Gambar 3.7 Halaman Klasifikasi Judul Berita

Gambar 3.7 merupakan rancangan antarmuka dari halaman klasifikasi yang berisi kolom dan tombol untuk melakukan klasifikasi sesuai dengan *input* yang diberikan oleh *user*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A