

BAB 3

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Penelitian “Implementasi Algoritma Extreme Gradient Boosting dalam Klasifikasi Judul Berita *Clickbait* di Media *Online*” memiliki beberapa tahapan sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan mencari, mengumpulkan, membaca, meneliti, memahami dan mengolah sumber dari buku, jurnal ilmiah, dan karya ilmiah sebagai referensi untuk melakukan penelitian. Proses ini bertujuan untuk meningkatkan pemahaman mengenai teori-teori yang berkaitan dengan algoritma Extreme Gradient Boosting, sehingga dapat memperkaya landasan teori dari penelitian.

2. Pengumpulan Data

Dataset diperoleh dari penelitian yang dilakukan oleh Andika William dan Yunita Sari yang berjudul *CLICK-ID: A novel dataset for Indonesian clickbait headlines* pada tahun 2020. *Dataset* yang didapatkan adalah 980 judul berita (40 judul berita dari setiap sumber seperti *detikNews*, *Kompas*, *Liputan6*, *Okezone*, *SINDOnews*, *Tempo*, dan *Tribunnews*). Dari setiap sumber akan diambil masing-masing 490 judul berita *clickbait* dan 490 judul berita *non-clickbait*. *Dataset* dapat ditambahkan lagi apabila masih diperlukan.

3. Perancangan Aplikasi

Proses perancangan untuk membuat aplikasi dimulai dengan membuat *use case diagram*, *flowchart diagram*, struktur tabel *database*, dan rancangan *mockup* untuk *user interface* dari aplikasi yang akan dibangun.

4. Proses Implementasi

Proses implementasi adalah tahap dimana algoritma Extreme Gradient Boosting diimplementasikan ke dalam aplikasi yang sudah selesai dengan menggunakan bahasa pemrograman Python dan akan menghasilkan jenis klasifikasi judul berita *clickbait* atau *non-clickbait*.

5. Pengujian Aplikasi

Tahap ini bertujuan untuk melakukan perhitungan tingkat akurasi dan efisiensi terhadap aplikasi yang sudah selesai dan melihat apakah terdapat kesalahan pada aplikasi yang dibuat.

6. Penulisan Laporan

Laporan dibuat sebagai dokumen untuk proses penelitian, desain, dan pengembangan aplikasi. Semua hasil perhitungan dan penelitian akan dituliskan ke dalam laporan.

3.2 Perancangan Aplikasi

Perancangan aplikasi dibagi menjadi beberapa bagian, yaitu *flowchart* aplikasi, *flowchart* perhitungan dengan algoritma Extreme Gradient Boosting, perancangan struktur tabel basis data, dan tampilan antarmuka aplikasi web.

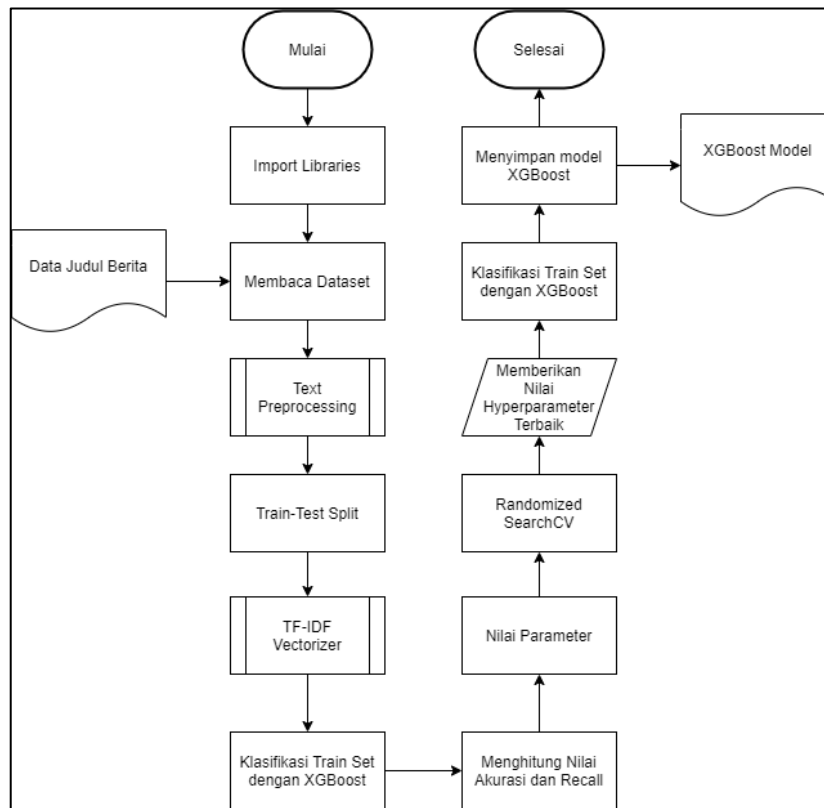
3.2.1 Flowchart Aplikasi

Perancangan *flowchart* aplikasi dibagi menjadi 4, yaitu *flowchart* Klasifikasi dengan XGBoost.

A. Flowchart Klasifikasi dengan XGBoost

Gambar 3.1 merupakan alur dari proses klasifikasi untuk melatih mesin pembelajaran dengan menggunakan algoritma Extreme Gradient Boosting. Langkah pertama dalam melatih mesin pembelajaran ini adalah meng-*import libraries* yang diperlukan, kemudian membaca *dataset* yang berisi kumpulan judul berita. Semua judul berita yang ada di dalam *dataset* akan dilakukan *text preprocessing* agar akurasi dari klasifikasi dapat menjadi lebih maksimal. Setelah tahap *text preprocessing*, judul berita akan dibagi menjadi dua bagian, yaitu dataset *train* dan dataset *test*. Dataset *train* adalah dataset yang dilatih sedangkan *dataset test* adalah *dataset* yang diuji. Kemudian *dataset* yang telah dibagi tadi akan dijadikan vektor dengan menggunakan *library* TF-IDF Vectorizer. Semua vektor tersebut akan dilatih dengan menggunakan algoritma Extreme Gradient Boosting. Setelah proses pelatihan, akan didapatkan nilai akurasi dan *recall*. Kemudian, dengan menggunakan nilai parameter yang telah ditetapkan secara acak, sistem akan masuk ke proses selanjutnya, yaitu proses GridSearchCV. Setelah itu nilai *hyperparameter* terbaik akan didapatkan, lalu *dataset train* akan diklasifikasikan kembali menggunakan nilai *hyperparameter* terbaik yang telah didapatkan tadi. Kemudian, masuk ke proses terakhir yaitu menerapkan algoritma Extreme Gradient Boosting ke dalam sistem.

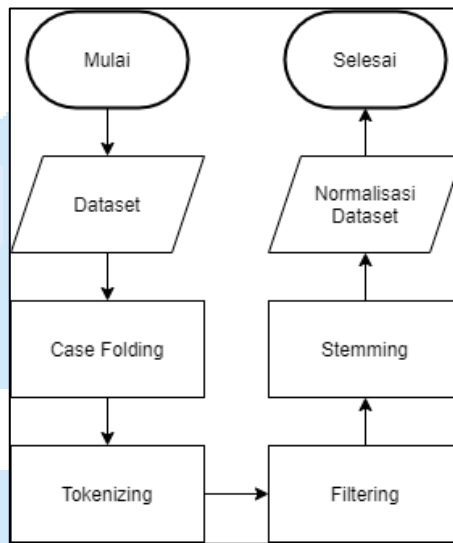
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3. 1 Flowchart Klasifikasi dengan XGBoost

B. Flowchart Text Preprocessing

Gambar 3.2 merupakan alur dari proses *text preprocessing* untuk membersihkan *dataset* agar klasifikasi mendapatkan akurasi yang lebih maksimal. Dimulai dengan mengambil *dataset* yang akan digunakan, kemudian masuk ke tahap *case folding*, yaitu mengubah semua teks yang ada di *dataset* menjadi huruf kecil. Kemudian, tahap *tokenizing* berfungsi untuk membagi teks menjadi beberapa bagian. Lalu, masuk ke tahap *filtering* dimana kata-kata yang penting akan diambil menggunakan algoritma *stoplist*. Setelah tahap *filtering* selesai, teks akan masuk ke tahap terakhir, yaitu tahap *stemming* yang berfungsi untuk mengurangi jumlah indeks yang berbeda. Setelah semua tahap selesai, maka *dataset* akan dinormalisasikan.

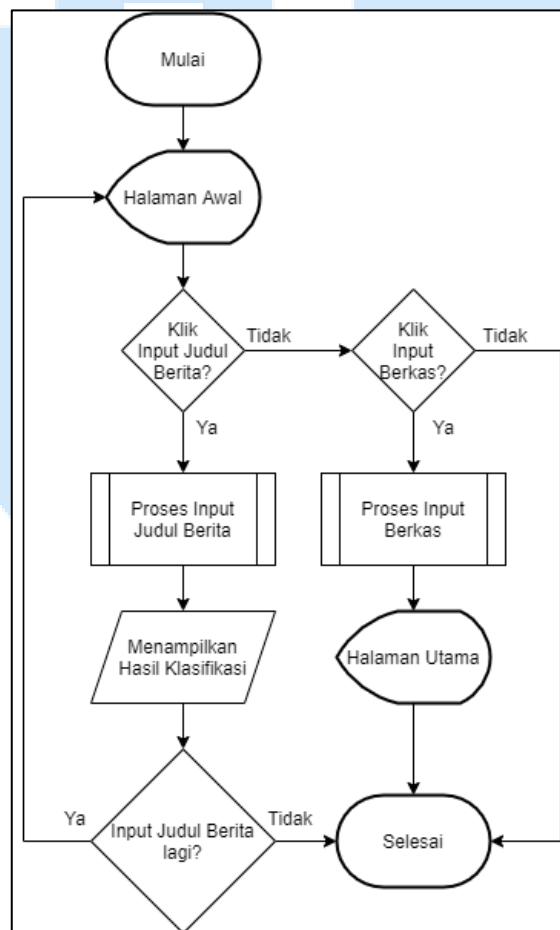


Gambar 3. 2 Flowchart Text Preprocessing

C. Flowchart Aplikasi Web

Gambar 3.3 merupakan alur dari aplikasi web. Pada saat *user* membuka aplikasi *web*, *user* akan masuk ke halaman awal yang berisi tentang penjelasan singkat mengenai berita *clickbait*, cara penggunaan, contoh format berkas, tombol Input Judul Berita, dan tombol Input Berkas. *User* diberikan pilihan untuk memasukkan satu buah judul berita atau lebih dari satu judul berita, jika *user* ingin memasukkan hanya satu buah judul berita, *user* dapat memasukkan judul beritanya secara manual pada kolom yang disediakan. Apabila *user* ingin memasukkan lebih dari satu judul berita, *user* dapat menggabungkan judul-judul berita ke dalam berkas Microsoft Excel dengan format yang sudah ditentukan. Ketika tombol Input Judul Berita ditekan, sebuah modal yang menyediakan kolom untuk judul berita akan muncul. Namun, jika tombol Input Berkas yang ditekan, sebuah modal yang menyediakan kolom untuk memasukkan berkas akan muncul. Setelah proses klasifikasi selesai, hasil

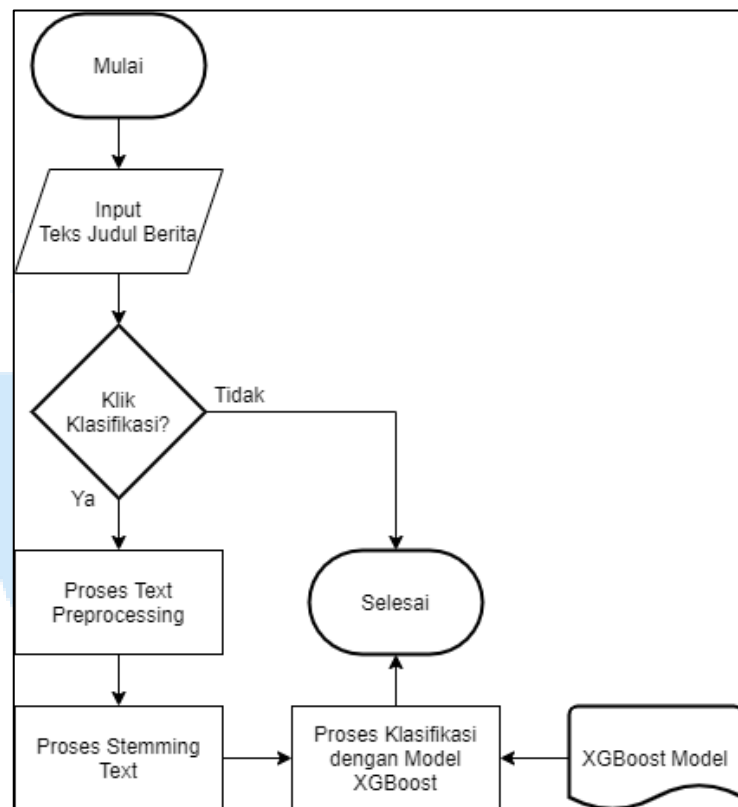
klasifikasi dari data yang dimasukkan oleh *user* akan ditampilkan. Apabila *user* ingin memasukkan judul berita kembali, maka *user* dapat memilih tombol Input Judul Berita Lagi.



Gambar 3. 3 Flowchart Aplikasi Web

D. Flowchart Proses Input Judul Berita

Gambar 3.4 merupakan proses *user* dalam memasukkan teks judul berita yang akan diprediksi. Setelah *user* memasukkan teks judul berita dan menekan tombol Klasifikasi, teks judul berita akan diproses dengan *Text Preprocessing* dan *Stemming Text*. Kemudian, teks judul berita akan diklasifikasi dalam proses Klasifikasi menggunakan Model XGBoost.

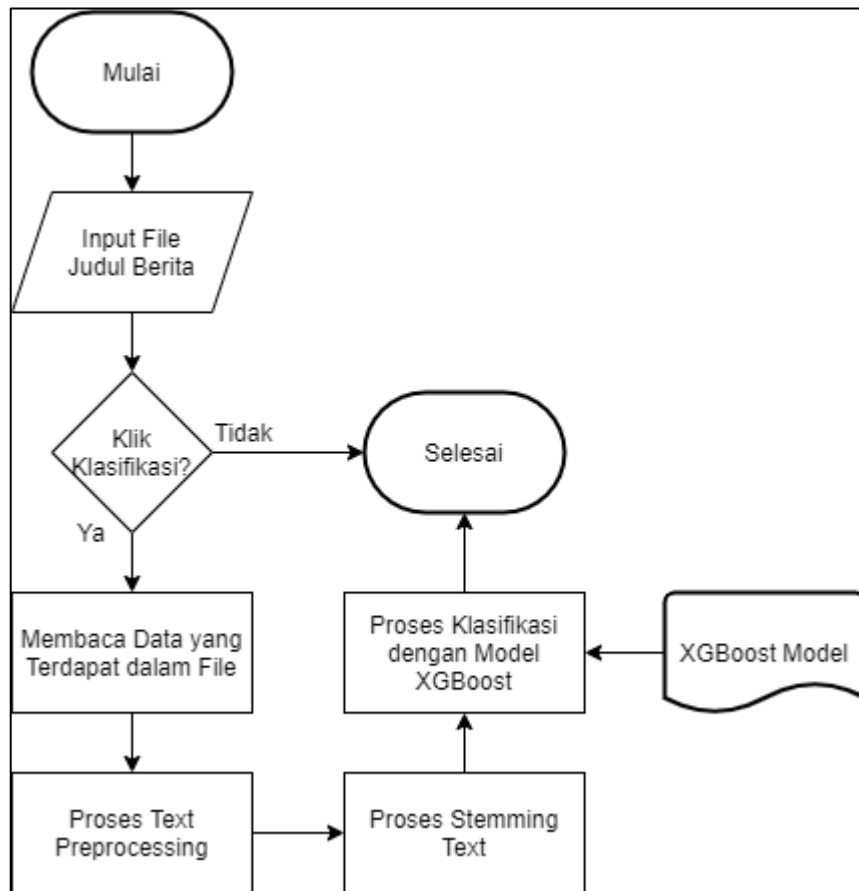


Gambar 3. 4 Flowchart Proses Input Judul Berita

E. Flowchart Proses Input Berkas

Gambar 3.5 merupakan alur dari proses input berkas. Setelah *user* memasukkan berkas yang berisi judul-judul berita dan klik Klasifikasi, semua data yang ada di dalam berkas akan dibaca dan dilakukan proses *Text Preprocessing* dan *Stemming Text*. Kemudian data-data tersebut akan diklasifikasi menggunakan model XGBoost.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



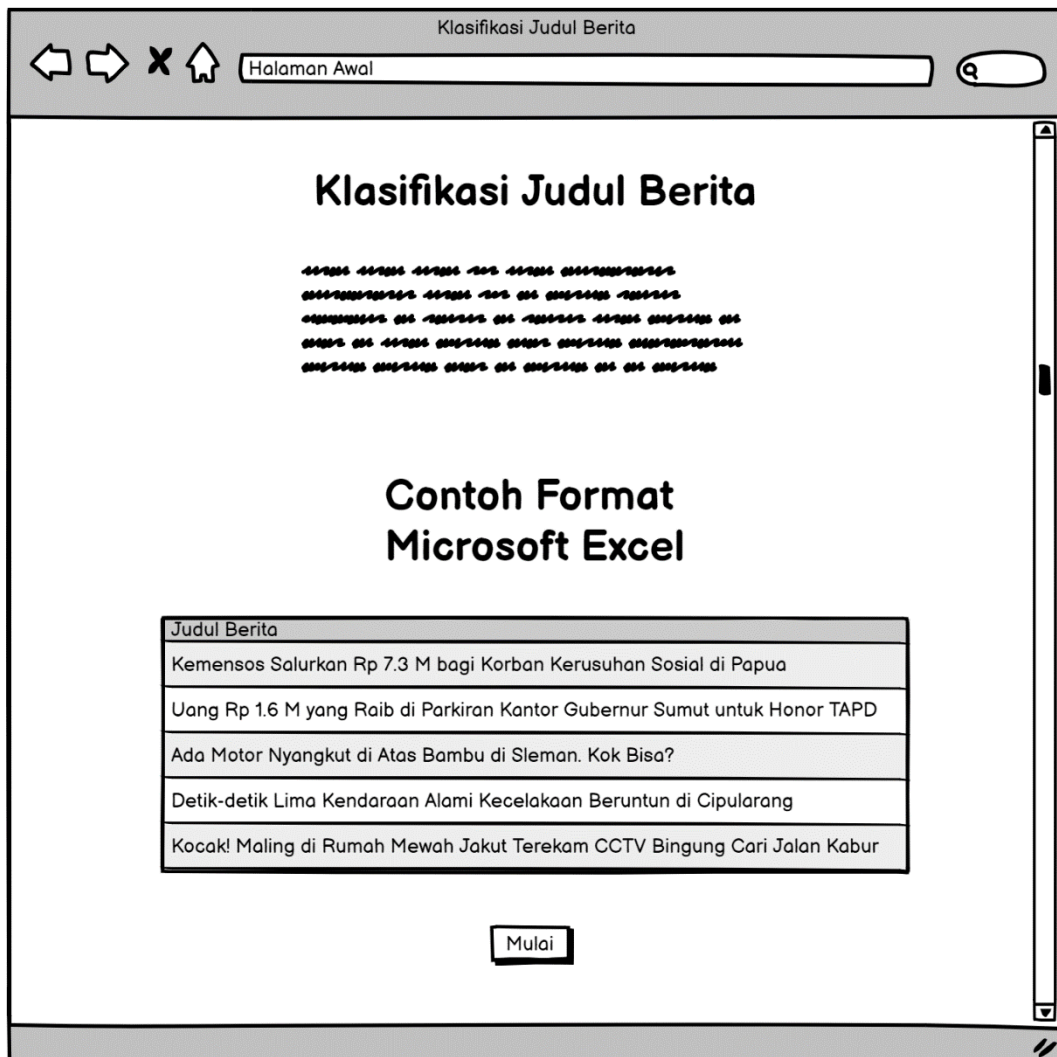
Gambar 3. 5 Flowchart Proses Input Berkas

3.2.2 Tampilan Antarmuka Aplikasi Web

Tampilan antarmuka aplikasi web memiliki 3 halaman, yaitu halaman awal, halaman utama, dan halaman hasil klasifikasi.

A. Perancangan Halaman Awal

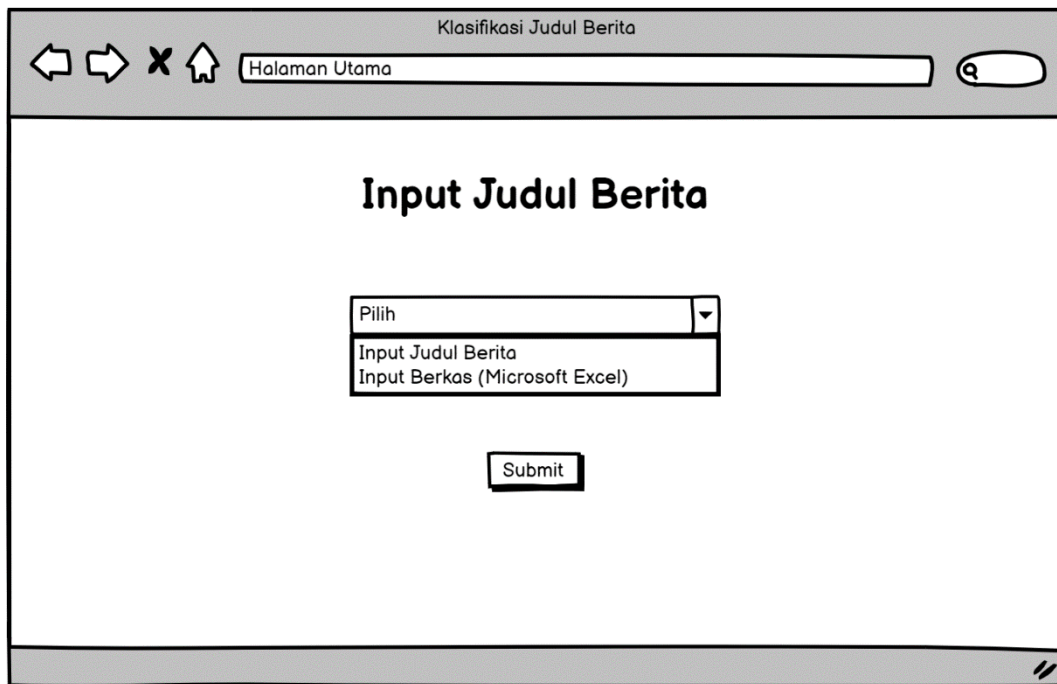
Gambar 3.6 adalah halaman awal untuk aplikasi web yang menjelaskan tentang aplikasi ini dan contoh format berkas (Microsoft Excel) yang dapat dimasukkan ke dalam aplikasi web.



Gambar 3. 6 Perancangan Antarmuka Halaman Awal

B. Perancangan Halaman Utama

Gambar 3.7 adalah halaman utama dimana pengguna aplikasi dapat memilih antara memasukkan judul berita secara manual (hanya satu judul berita) atau memasukkan berkas (Microsoft Excel) untuk banyak judul berita.

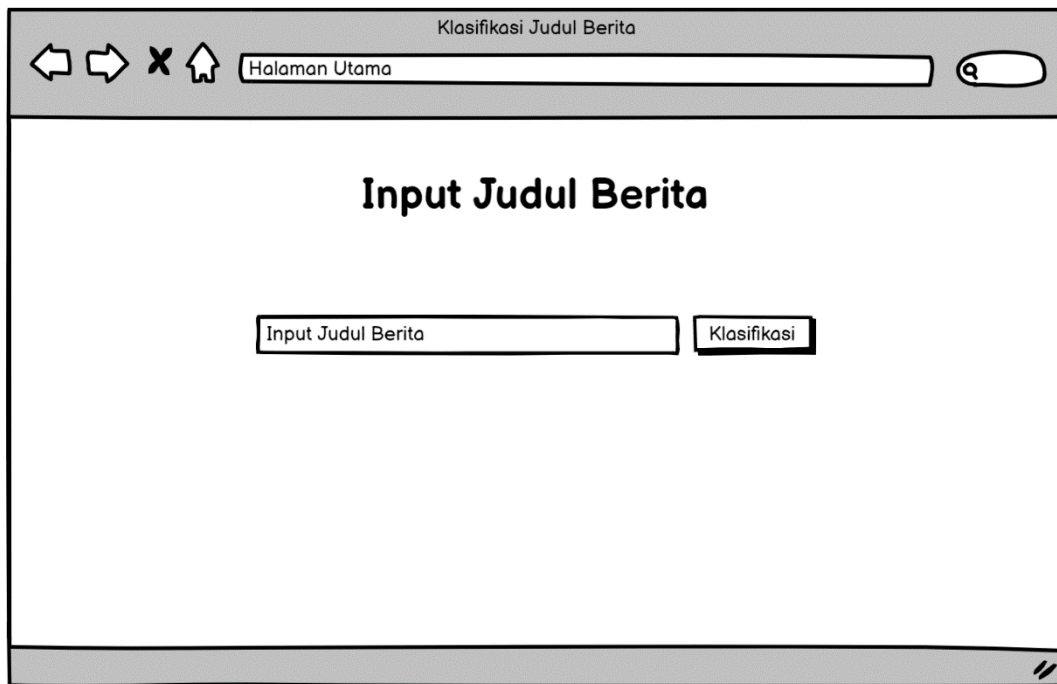


Gambar 3. 7 Perancangan Halaman Utama

C. Perancangan Halaman Utama (Judul Berita)

Gambar 3.8 adalah tampilan antarmuka yang muncul ketika pengguna aplikasi memilih pilihan Masukkan Judul Berita. Halaman ini menyediakan kolom agar pengguna aplikasi dapat memasukkan judul berita untuk diklasifikasi.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

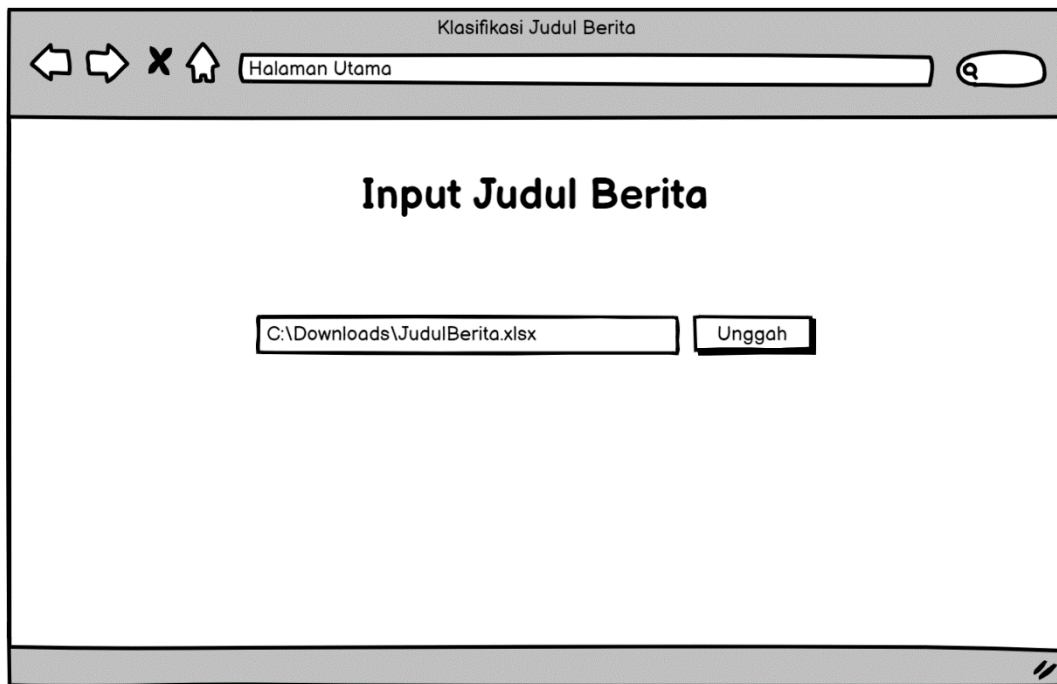


Gambar 3. 8 Perancangan Halaman Utama (Masukkan Judul Berita)

D. Perancangan Halaman Utama (Masukkan Berkas)

Gambar 3.9 adalah tampilan antarmuka yang muncul ketika pengguna aplikasi memilih pilihan Masukkan Berkas. Halaman ini menyediakan kolom agar pengguna aplikasi dapat memasukkan berkas Microsoft Excel yang berisi lebih dari satu judul berita untuk diklasifikasi.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

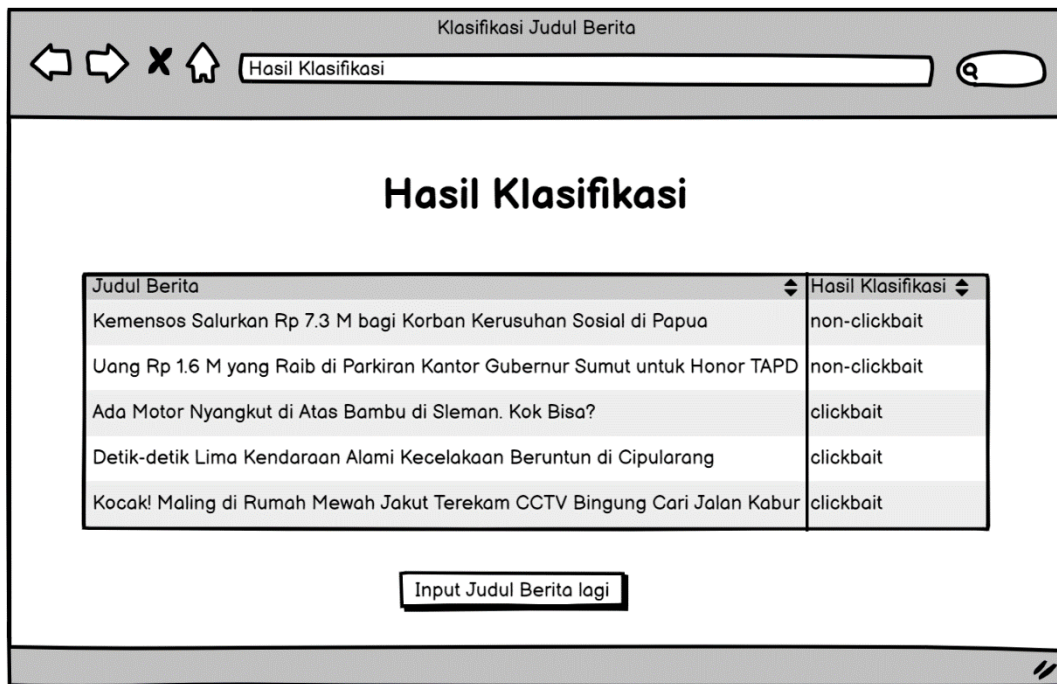


Gambar 3. 9 Perancangan Halaman Utama (Masukkan Berkas)

E. Perancangan Halaman Hasil Klasifikasi

Gambar 3.10 adalah halaman yang menampilkan hasil klasifikasi dari judul-judul berita yang telah dimasukkan oleh pengguna aplikasi. Hasil yang ditunjukkan pada halaman ini adalah clickbait dan non-clickbait.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3. 10 Perancangan Halaman Hasil Klasifikasi

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA