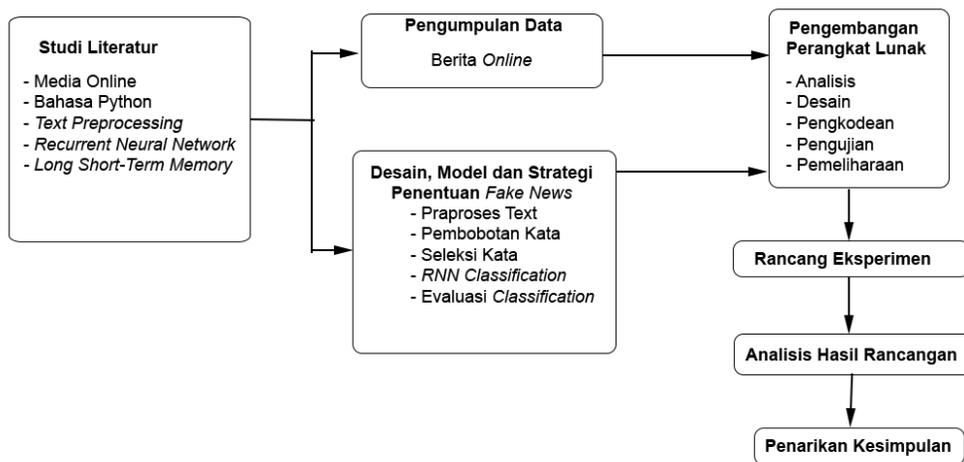


BAB III METODOLOGI PENELITIAN

Pada bab 3 ini penulis akan menjelaskan mengenai metodologi penelitian yang akan digunakan dalam penelitian untuk mendeteksi *fake news* dari data sejenis berita atau artikel bebas yang ada di media *online* dengan pendekatan *long short-term memory*. Pemaparan dan penjelasan mengenai desain penelitian, pengembangan perangkat lunak, serta alat dan bahan yang digunakan akan dijelaskan dalam bab ini.

3.1 Desain Penelitian

Desain penelitian merupakan gambaran garis besar tahapan atau alur kerja dalam penelitian yang akan dilakukan penulis dalam penelitian ini. Tahapan atau alur penelitian diilustrasikan pada gambar 3.1.



Gambar 3.1. Desain Penelitian

Gambar 3.1 menunjukkan desain penelitian yang telah dirancang oleh penulis dan terdiri dari tujuh tahap. Setiap tahap akan dijelaskan pada sub bab berikutnya lebih mendalam guna memberikan gambaran mengenai penelitian ini menjadi lebih jelas.

3.2 Alat dan Bahan Penelitian

Dalam penelitian ini penulis menggunakan dua komponen utama untuk menyelesaikan penelitian yaitu komponen perangkat keras dan perangkat lunak sebagai alat penelitian. Sedangkan bahan yang digunakan untuk penelitian adalah website penyedia berita *online* yang cukup sering dikunjungi masyarakat.

3.2.1. Alat Penelitian

Alat yang digunakan dalam penelitian ini adalah sebuah perangkat keras berupa satu unit Laptop merk *Lenovo* sebagai alat utama guna pengerjaan penelitian ini dengan spesifikasi sebagai berikut :

- AMD A9-9425 Radeon R5
- RAM 4GB
- System Operasi Windows 10 Pro 64-bit
- Mouse

3.2.2. Bahan Penelitian

Bahan yang digunakan dalam penelitian ini bersumber dari sebuah portal berita *online* yang sering digunakan oleh masyarakat agar mendapatkan hasil yang maksimal. Berikut adalah link bahan penelitian ini.

- <https://www.detik.com/>
- <https://tempo.co/>
- <https://cnnindonesia.com/>
- <https://turnbackhoax.id/>
- <https://stophoax.id/>

Bahan yang dibutuhkan dalam penelitian ini selain data berita bahasa Indonesia yang berupa teks, dibutuhkan juga *tools* untuk melakukan perbandingan algoritma, *tools* yang digunakan adalah salah satu library yang sudah ada dalam python yang bernama *scikit-learn*.

Scikit-learn adalah library untuk *machine learning* yang mendukung melakukan beragam pekerjaan dalam *data science* seperti regresi, klasifikasi, *clustering*, data *preprocessing*, *dimensionality reduction*, dan *model selection* (perbandingan, validasi, dan pemilihan

parameter maupun model). *Scikit-learn* ditulis dalam bahasa python, dan menggunakan *numpy* secara ekstensif untuk perhitungan aljabar linier dan operasi array.

Scikit-learn terintegrasi dengan banyak library python lainnya, seperti *matplotlib*, *plotly* untuk plotting, *numpy* untuk vektorisasi array, *panda dataframe*, *scipy* dan banyak lagi.

Matplotlib merupakan library python 2D plotting yang dapat menghasilkan plot dengan kualitas tinggi dalam berbagai format dan dapat digunakan di banyak platform. *Matplotlib* juga dapat digunakan sebagai pembuat grafik dalam berbagai platform, seperti python dan *Jupiter*.

Numpy memiliki kegunaan untuk operasi vektor dan matriks. Fiturnya hampir sama dengan *MATLAB* dalam mengelola *array* dan *array* multidimensi. *Numpy* merupakan salah satu *library* yang digunakan oleh *library* lain seperti *Scikit-Learn* untuk keperluan analisis data.

Panda Dataframe menggunakan sistem *dataframe*, sehingga dapat membuat sebuah file ke dalam tabel virtual seperti spreadsheet dengan menggunakan *pandas*. Dengan menggunakan *pandas*, data juga dapat diolah seperti *join*, *distinct*, *group by*, agregasi, dan teknik lain seperti di *SQL*. *Pandas* juga dapat membaca file dari berbagai format seperti *.txt*, *.csv*, *.tsv*, dan lainnya.

Scipy berguna untuk menangani operasi aljabar dan matriks serta operasi matematika lainnya. *Scipy* juga dapat menangani sejumlah operasi matematika yang lebih kompleks daripada menggunakan library *math* bawaan python.

Dalam *scikit-learn* terdapat beberapa algoritma klasifikasi atau yang biasa disebut *classifier* yang bisa digunakan, diantaranya :

- K-Nearest Neighbors
- Support Vector Machines
- Decision Tree
- Random Forest
- Linear Discriminant Analysis
- Logistic Regression

3.3 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif dimana metode ini memiliki cakupan yang sangat luas. Metode ini didasari oleh eksperimental dimana penelitian ini dilakukan dengan cara eksperimen dengan menguji bahan penelitian untuk mendapatkan persentase akurasi dari aplikasi ini.

3.3.1. Metode Pengumpulan Data

Metode pengumpulan data dilakukan secara manual dengan menyalin berita dari situs-situs portal berita *online* Indonesia yang digunakan. Selain mengumpulkan data untuk di analisa, peneliti juga membuat beberapa *sample* berita palsu yang akan menjadi data *training* dalam proses *classification* antara *valid news* dan *fake news*. Sedangkan untuk data *testing* sendiri penulis mengambil berita dari sumber media *online* yang sudah cukup terpercaya dan beberapa media *online* lain untuk mendapat bahan pengujian. Data berita yang digunakan merupakan artikel berita berbahasa Indonesia dan panjang teks yang diambil dalam tiap berita adalah sepanjang 1 – 3 kalimat.

3.3.2. Metode Pengembangan Perangkat Lunak

Untuk mengembangkan perangkat lunak yang diteliti, penulis terlebih dahulu melakukan analisa kriteria berita palsu yang menjadi acuan validitas, dilanjutkan dengan merancang sistem awal sebagai *prototype* untuk kemudian dikembangkan lagi, lalu melakukan pengujian terhadap *prototype* untuk mendapatkan persentase ketepatan, jika belum mencapai kriteria yang ditentukan maka akan terus dilakukan pengembangan dan perbaikan terhadap aplikasi sampai sesuai dengan kriteria yang ditentukan, setelah sesuai dengan kriteria berikutnya adalah melakukan pemeliharaan terhadap aplikasi secara berkala dan terus dikembangkan lagi untuk menjadi lebih baik.

3.3.3. Metode Pengujian Perangkat Lunak

Untuk menguji perangkat lunak yang telah diteliti, penulis mengukurnya dengan menghitung akurasi, presisi, *recall*, dan *F₁-Measure*. Akurasi adalah ukuran evaluasi yang umum digunakan dalam proses klasifikasi pada *machine learning*. Dalam bidang statistika, akurasi merupakan sistem pengukuran dari tingkat kedekatan pengukuran kuantitas terhadap nilai yang

sebenarnya. (Cohen, 1998). Maka dapat disimpulkan akurasi merupakan tingkat kedekatan nilai yang diprediksi dengan nilai sebenarnya. Untuk mendapatkan nilai akurasi, dapat dihitung menggunakan rumus pada persamaan 8.

$$\text{Akurasi} = \frac{\text{Jumlah Data Benar}}{\text{Seluruh Jumlah Data}} \times 100\% \quad (8)$$

Presisi adalah tingkat ketepatan antara informasi yang diminta dengan jawaban yang sebenarnya (Cohen, 1998). Presisi adalah jumlah sampel berkategori positif diklasifikasi benar dibagi dengan total sampel yang diklasifikasikan sebagai sampel positif. Maka didapatkan rumus untuk menghitung presisi dengan persamaan 9.

$$\text{Presisi} = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Data Satu Kategori}} \times 100\% \quad (9)$$

Recall merupakan ukuran dari tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dengan kata lain, *recall* digunakan pada sebuah sistem temu kembali informasi (*information retrieval*) untuk menilai seberapa besar keberhasilan yang didapat.

$$\text{Recall} = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Data yg Diprediksi Satu Kategori}} \times 100\% \quad (10)$$

F₁-measure (*f₁-score*) adalah salah satu bentuk perhitungan evaluasi dalam sistem *information retrieval* yang mengkombinasikan nilai *recall* dan presisi. Nilai *recall* dan presisi pada suatu keadaan tertentu memiliki bobot yang berbeda. *F₁-measure* ini merupakan ukuran timbal balik antara presisi dan *recall*, dimana didalamnya juga merupakan bobot *harmonic mean* dari presisi dan *recall*.

$$F_1 = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \times 100\% \quad (11)$$