**BAB 1. PENDAHULUAN**

* 1. **Latar Belakang**

Di era *big data* ini informasi tumbuh begitu pesat dan juga terus menyebar di internet, terutama informasi dalam bentuk data tekstual, informasi dalam bentuk data tekstual merupakan salah satu hal yang sangat dibutuhkan oleh setiap individu maupun kelompok, informasi tersebut bisa kita dapatkan hanya cukup dengan membaca. Namun biasanya semakin banyak informasi yang terkandung dalam suatu dokumen tekstual, dokumen tersebut juga memiliki teks yang semakin panjang, tentu hal itu membuat pembaca kesulitan untuk mendapatkan inti dari informasi yang terkandung, karena untuk memahaminya pembaca membutuhkan usaha, ketelitian serta waktu yang cukup lama. Hal tersebut tidak sebanding dengan membaca sebuah ringkasan, dengan membaca ringkasan selain menghemat waktu, pembaca lebih dipermudah untuk mengetahui isi dan makna dari informasi yang terkandung dalam sebuah dokumen. Namun untuk membuat sebuah ringkasan tidaklah semudah yang dibayangkan, bila dilakukan secara manual pasti akan menjadi pekerjaan yang melelahkan. Untuk mengatasi permasalahan tersebut yaitu dengan *text summarization, text summarization* merupakan solusi yang dapat digunakan untuk membuat peringkasan teks dangan tetap mempertahankan kualitas dan makna dari dokumen asli.

*Text summarization* adalah proses mengekstraksi atau mengumpulkan informasi penting dari teks asli dan menyajikan informasi tersebut dalam bentuk ringkasan *(Deepali K dan C. Namrata Mahender, 2016),* berdasarkan studi literature sebelumnya yang dilakukan peneliti, *text summarization* secara luas diklasifikasikan terutama ke dalam dua kategori, yaitu ekstraktif dan abstraktif. ekstraktif bekerja dengan memilih subset dari kata, frasa atau kalimat yang ada dari teks asli untuk membentuk ringkasan, peringkasan ekstraktif mengunakan pendekatan statistik untuk memilih kalimat atau kata kunci untuk menghasilkan sebuah ringkasan. Sedangkan

abstraktif menghasilkan kalimat dari representasi semantik dan kemudian menggunakan teknik generasi bahasa alami untuk membuat ringkasan yang lebih dekat dengan apa yang dihasilkan manusia, teknik abstraktif menggunakan pendekatan linguistik untuk memahami teks asli sehingga dapat menghasilkan ringkasan.

Dari dua pendekatan yang telah dijelaskan, dalam penelitian ini peneliti hanya akan fokus pada pembuatan *text summarization* berbasis ekstraktif dengan mengusulkan metode *TF-IDF* dan proses *preprocessing*, dimana *preprocessing* ditujukan untuk memproses sebuah data sebelum diterapkannya metode *TF-IDF,* serta menganalisa hasil ringkasan menggunakan metode *TF-IDF* sebagai pendekatan ekstraktif untuk menghasilkan ringkasan.

* 1. **Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana memproses dokumen untuk menghasilkan *text summarization*?
2. bagaimana memproses data yang mengandung noise menjadi data yang bersih dari noise sesuai kebutuhan menggunakan proses *text preprocessing* sebelum diterapkannya metode *TF-IDF?*
3. Bagaimana menerapkan metode *TF-IDF* untuk menghasilkan *text summarization* dengan pendekatan ekstraktif?
   1. **Batasan Masalah**

Berdasarkan uraian latar belakang di atas, dapat dirumuskan batasan masalah sebagai berikut :

1. Penelitian ini hanya membahas bagaimana menghasilkan *text summarization ekstraktif*.
2. Penelitian ini hanya membahas bagaimana memproses data yang mengandung noise menjadi data yang bersih dari noise sesuai kebutuhan dengan proses *text preprocessing* sebelum diterapkannya metode *TF-IDF.*
3. Penelitian ini hanya berfokus pada proses pembuatan *text summarization Ekstraktif* dengan metode *TF-IDF*.
   1. **Tujuan**
4. Menganalisa proses pembuatan *text summarization* berbasis ekstraktif.
5. Menerapkan proses *preprocessing* untuk memproses dokumen berbasis data tekstual sebelum di terapkannya metode *TF-IDF*.
6. Menerapkan metode *TF-IDF* untuk menghasilkan *text summarization* berbasis ekstraktif.
   1. **Manfaat**
7. Dengan adanya penelitian diharapkan bisa membantu pengembangkan pengetahuan tentang keilmuan *Natural Language Processing (NLP)* atau pemrosesan bahasa alami kususnya di bidang *text summarization*.
8. Penelitian ini diharapkan dapat membantu seorang pembaca untuk mengetahui isi makna yang terkandung pada sebuah informasi tekstual dengan cukup membaca ringkasan yang sudah mewakili isi dan panjang dari keseluruhan dokumen tekstual.
9. Dengan adanya penelitian ini diharapkan membantu seorang peneliti lain untuk mengetahui proses *text summarization* pada pendekatan ekstraktif dengan penerapan metode *TF-IDF*.