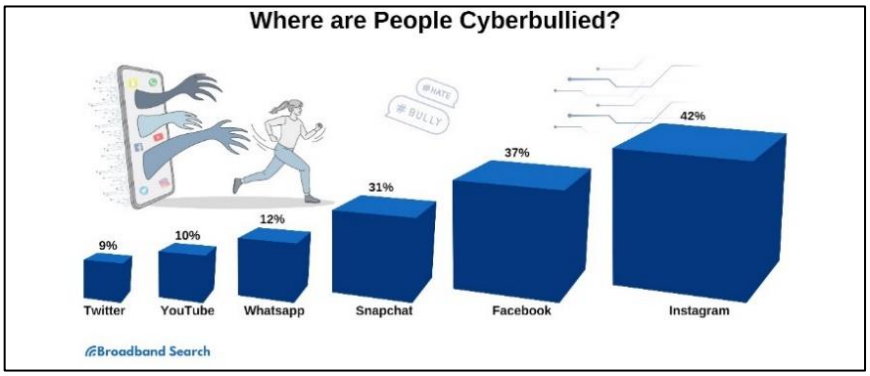
BAB I
PENDAHULUAN

1.1 Latar Belakang

Pada era digital yang berkembang pesat, teknologi dan internet membawa perubahan besar dalam kehidupan sehari-hari, memengaruhi pendidikan, pekerjaan, hingga hiburan. Menurut [1], pengguna internet di Indonesia mencapai 221 juta atau 79,5% dari populasi, naik 1,4% dari periode sebelumnya. Pertumbuhan ini menunjukkan semakin banyaknya orang yang memanfaatkan internet untuk komunikasi, mendapatkan informasi, dan berinteraksi. Salah satu perubahan nyata adalah pertumbuhan media sosial yang pesat, di mana sekitar 139 juta orang di Indonesia aktif menggunakannya dengan rata-rata waktu penggunaan 3 jam per hari [2]. Media sosial tidak hanya sebagai alat komunikasi, tetapi juga sarana untuk menyampaikan pendapat, berbagi pengalaman, dan berekspresi dalam berbagai bentuk seperti tulisan, foto, dan video. *Platform* seperti Facebook, Twitter, dan Instagram menjadi ruang bagi pengguna untuk berpartisipasi dalam diskusi mengenai topik yang penting bagi mereka.

Pada tahun 2024, Instagram menempati peringkat pertama sebagai *platform* media sosial paling banyak digunakan di Indonesia, dengan sekitar 84,8% pengguna internet aktif [3]. *Platform* ini sangat populer di kalangan remaja hingga dewasa, yang menggunakannya untuk berbagi foto, video, dan cerita, serta berinteraksi melalui fitur komentar. Fitur ini memungkinkan pengguna memberikan tanggapan atau pendapat terhadap konten yang diunggah orang lain, tetapi sering disalah gunakan untuk menyebarkan komentar negatif seperti penghinaan dan ujaran kebencian.



Gambar 1. 1 Peringkat Cyberbullying di Setiap Platform
(Sumber: Broadband Search, 2024)

Menurut [4] seperti yang terlihat pada Gambar 1.1, Instagram menempati peringkat pertama sebagai *platform* dengan insiden cyberbullying tertinggi, dengan persentase 42%. Hal ini menunjukkan bahwa meskipun fitur komentar Instagram memungkinkan interaksi yang positif, pengguna juga berisiko terkena *cyberbullying*.

Cyberbullying adalah tindakan perundungan yang dilakukan melalui teknologi digital termasuk media sosial, aplikasi *chating* dan forum *online,* dengan tujuan menakut-nakuti, membuat marah, atau mempermalukan individu lain [5]. Terdapat beberapa bentuk *cyberbullying* yang sering muncul di media sosial, yaitu *Flaming, Denigration, Harassment,* dan *Body Shaming*. Penelitian [6] menemukan bahwa *flaming*, *harassment* dan *denigration* umum muncul dalam komentar, sementara [7] menunjukkan bahwa body shaming juga kerap terjadi di media sosial, dengan dampak buruk pada diri korban. *Flaming* adalah komentar kasar atau hinaan, sedangkan *Denigration* adalah penyebaran rumor dan fitnah [8] [9]. *Harassment* mencakup tekanan ancaman dan pelecehan, sementara *Body Shaming* menargetkan penampilan fisik [10] [7]. Setiap bentuk *cyberbullying* ini memiliki karakteristik unik dalam penggunaan bahasa dan intensinya, yang dapat dianalisis untuk mendeteksi pola-pola kebencian di media sosial.

Cyberbullying memiliki dampak yang sangat merugikan, terutama bagi korban yang sering kali mengalami tekanan psikologis dan sosial. [11] menemukan bahwa remaja yang menjadi korban *cyberbullying* sering merasa tidak berdaya dan kesulitan mengatasi dampak emosional dari bullying online. Dampak seperti kecemasan dan depresi sering muncul pada korban [12]. Di Indonesia, jumlah kasus *cyberbullying* telah meningkat secara signifikan dalam beberapa tahun terakhir. Data dari Komisi Perlindungan Anak Indonesia [13] menunjukkan bahwa dari tahun 2011-2019, terdapat 2.473 laporan *cyberbullying*, dan jumlah ini terus meningkat setiap tahunnya. Meski begitu, banyak korban (31,6%) enggan melaporkan insiden tersebut. Hal ini dikarenakan korban merasa takut, malu, atau kurangnya dukungan dari keluarga dan teman [14].

Dari perspektif pelaku, *cyberbullying* sering dilakukan karena rasa aman yang diciptakan oleh anonimitas di dunia maya. [15] menemukan bahwa banyak pelaku tidak menyadari atau bahkan tidak peduli dengan dampak dari tindakan

mereka, menganggap *cyberbullying* sebagai bentuk hiburan. Di Indonesia, tindakan *cyberbullying* sebenarnya telah diatur dalam Undang-Undang Informasi dan Transaksi Elektronik UU ITE) Pasal 27 Ayat 3, yang dapat menjatuhkan sanksi pidana penjara hingga 6 tahun atau denda maksima Rp 1 miliar kepada pelaku. Meskipun begitu, tantangan utama dalam penegakan hukum adalah kurangnya bukti digital dan banyaknya kasus yang tidak dilaporkan, sering kali karena minimnya dukungan bagi korban [13]. Untuk mengatasi hal ini, sistem deteksi otomatis dapat menyediakan bukti digital untuk meningkatkan perlindungan bagi korban. Berbagai jenis *cyberbullying* dalam satu komentar sering muncul dalam kombinasi, sehingga klasifikasi *multi-label* memungkinkan setiap komentar dikategorikan dengan baik.

Klasifikasi multi-label menjadi salah satu solusi dalam mendeteksi berbagai bentuk *cyberbullying* di media sosial. Klasifikasi *multi-label* adalah pendekatan yang digunakan untuk menangani kasus di mana satu sampel data dapat memiliki dua atau lebih label [17]. Ada dua pendekatan utama dalam klasifikasi *multi-label*, yaitu *Problem Transformation* dan *Algorithm Adaptation*. *Problem Transformation*, seperti *Label Powerset* (LP), mengubah masalah multi-label menjadi beberapa masalah klasifikasi *single-label*, yang dapat memperhitungkan korelasi antar label [18] Sebaliknya, *Algorithm Adaptation* mengubah algoritma single-label agar dapat langsung menangani masalah multi-label, seperti pada ML-KNN, yang memiliki performa lebih baik dalam hal *hamming loss* [18]. Dalam penelitian ini, digunakan dua pendekatan tersebut, yaitu *Label Powerset K-NN dan* ML-KNN, untuk klasifikasi multi-label cyberbullying di Instagram. Fitur diekstraksi menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), yang mengubah teks menjadi format terstruktur dengan menilai setiap kata berdasarkan frekuensi dan relevansinya terhadap klasifikasi [19] Sehingga dengan pendekatan ini, penelitian diharapkan mampu menghasilkan klasifikasi yang akurat.

Klasifikasi *multi-label* telah menjadi metode yang digunakan dalam berbagai bidang penelitian. [NO_PRINTED_FORM] [20] menggunakan pendekatan *multi-label* untuk mendeteksi ujaran kebencian di Twitter Indonesia, menghasilkan akurasi sebesar 66,12% dengan *Label Powerset* (LP) pada *Random Forest Decision Tree*. [21] membandingkan metode *Problem Transformation* dan

Algorithm Adaptation dalam klasifikasi pertanyaan *multi-label* di Kotakode, dengan *Label Powerset*-KNN memberikan hasil terbaik dengan akurasi 86% dan f1-score sebesar 87%. Dalam bidang ekonomi, Penelitian [22] menerapkan ML-KNN untuk klasifikasi *multi-label* fenomena ekonomi, menghasilkan akurasi 63,22%, menunjukkan bahwa ML-KNN efektif dalam menangani data ekonomi yang kompleks. [23] juga menggunakan ML-KNN dalam klasifikasi *multi-label* pengguna media sosial, yang memberikan performa akurasi lebih baik dibandingkan metode lain.

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini berfokus pada penerapan klasifikasi *multi-label* dalam mendeteksi *cyberbullying* di media sosial, khususnya di *platform* Instagram. Penelitian ini membandingkan dua pendekatan utama dalam klasifikasi multi-label, yaitu *Problem Transformation* dan *Algorithm Adaptation*. Dalam *Problem Transformation*, metode *Label Powerset* K-NN digunakan untuk mengubah masalah *multi-label* menjadi beberapa masalah *single-label*. Sedangkan *Algorithm Adaptation* menggunakan ML-KNN, yang merupakan adaptasi dari algoritma K-NN. Penelitian ini menggunakan teknik ekstraksi fitur TF-IDF untuk mengubah teks menjadi data terstruktur, sehingga algoritma dapat memprosesnya dengan akurat. Dengan membandingkan kedua pendekatan ini, penelitian ini bertujuan untuk menemukan metode yang paling akurat dalam mendeteksi berbagai bentuk *cyberbullying* di Instagram, serta memberikan kontribusi untuk meminimalkan dampak negatif *cyberbullying*.

Selain itu, penelitian ini juga bertujuan untuk mengembangkan sebuah sistem yang dapat secara otomatis mengklasifikasikan komentar yang mengandung *cyberbullying*. Sistem ini diharapkan dapat membantu pengguna dalam memoderasi komentar mereka sebelum dipublikasikan di media sosial. Dengan adanya peringatan dini ketika suatu komentar terdeteksi mengandung unsur *cyberbullying*, pengguna dapat lebih berhati-hati dalam berinteraksi secara *online*. Selain sebagai alat pencegah, sistem ini juga diharapkan dapat memberikan alternatif bukti digital yang dapat digunakan oleh korban atau pihak berwenang untuk melaporkan kasus *cyberbullying*. Dengan demikian, sistem ini berpotensi membantu menciptakan lingkungan digital yang lebih aman dan bebas dari ancaman *cyberbullying.*

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya maka dapat dirumuskan beberapa permasalahan seperti berikut:

1. Bagaimana hasil perbandingan antara metode *Label Powerset* K-NN dan ML-KNN dalam mengklasifikasikan *multi-label cyberbullying* pada komentar Instagram?
2. Bagaimana hasil pembuatan sistem yang dapat mengklasifikasikan *cyberbullying* dari teks?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah komentar-komentar yang diambil dari *platform* Instagram. Data diambil menggunakan teknik *scraping* dari akun @doshzn, @raffinagita1717, @prillylatuconsina96, @sithamarino, @anyageraldine, @wanda_haraa, dan @rio_clappy.
2. Data yang digunakan dalam penelitian ini tidak mengandung negasi, sehingga analisis hanya dilakukan pada kalimat tanpa elemen negasi.
3. Data komentar akan dikategorikan dalam lima kelas *cyberbullying*, yaitu *Flaming, Denigration, Harrasment Body Shaming,* dan *Neutral*.
4. Teknik ekstraksi fitur yang digunakan adalah TF-IDF dengan kombinasi 1-3 gram.
5. Metode yang akan digunakan adalah *Label Powerset* K-NN dan ML-KNN untuk klasifikasi *multi-label*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, tujuan dari penelitian ini adalah sebagai sebagai berikut:

1. Mengetahui hasil perbandingan antara metode *Label Powerset* K-NN dan ML-KNN dalam mengklasifikasikan *multi-label cyberbullying* pada komentar Instagram.
2. Membuat sistem yang dapat mengklasifikasikan *cyberbullying* dari teks.

1.5 Manfaat Penelitian

1. Penelitian ini dapat memberikan pemahaman yang lebih baik kepada masyarakat bahwa *machine learning* dapat digunakan untuk mengenali dan mengklasifikasikan berbagai bentuk *cyberbullying* dalam komentar Instagram.
2. Dengan membandingkan algoritma LP-KNN dan ML-KNN serta menggunakan teknik TF-IDF dan n-grams, penelitian ini dapat menentukan model yang memiliki performa terbaik dalam klasifikasi *multi-label cyberbullying*.
3. Hasil penelitian ini dapat menjadi referensi dalam pengembangan model *machine learning* yang lebih akurat untuk klasifikasi *multi-label* pada teks, khususnya dalam konteks *cyberbullying*, serta menjadi dasar bagi penelitian lebih lanjut atau pengembangan sistem deteksi otomatis.