

BAB I PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu negara yang masih terdapat permasalahan serius terkait pernikahan dini. Kini sedang beredar isu terkait penurunan jumlah pernikahan nasional yang mencapai 7,5 persen pada tahun 2023, Indonesia masih dihadapkan pada tantangan serius terkait perkawinan usia anak. Saat ini, ratusan ribu anak di bawah usia 18 tahun telah menjalani pernikahan, yang seringkali dikarenakan beberapa faktor. Salah satu penyebab utama adalah masalah ekonomi yang dihadapi keluarga. Walaupun Deklarasi HAM di tahun 1954 dengan tegas menolak pernikahan dini, praktik nikah dini masih terjadi di seluruh dunia. Hal ini menggambarkan betapa perlunya perlindungan hak asasi anak, khususnya bagi kelompok usia muda, yang masih terabaikan [1].

Pernikahan dini tentunya juga menimbulkan banyak dampak negatif bagi generasi muda, baik dari segi sosial, ekonomi, bahkan psikologis. Pernikahan, sebagai sebuah wadah sosial yang bersifat fundamental, tak hanya memancarkan ikatan antara dua orang, tetapi juga ada dampak yang signifikan terhadap struktur sosial, berkeluarga, dan kesejahteraan masyarakat secara menyeluruh [2]. Secara sosial, pernikahan pada usia muda sering kali menyebabkan terhambatnya pendidikan yang dapat mengurangi peluang untuk mendapatkan pekerjaan yang layak sebagai generasi muda bangsa. Dari segi ekonomi, pasangan yang menikah di usia dini cenderung menghadapi kesulitan finansial karena keterbatasan keterampilan dan pengalaman kerja. Selain itu, secara psikologis, pernikahan dini dapat meningkatkan risiko stres, depresi, dan kekerasan dalam rumah tangga (KDRT) akibat kurangnya kesiapan dalam emosional untuk menjalani kehidupan berumah tangga.

Pada era digital saat ini, masyarakat tentunya sangat melek terhadap media sosial dengan memanfaatkannya sebagai tempat untuk mengekspresikan diri ataupun opini terkait berbagai isu sosial yang terus bermunculan, termasuk pernikahan dini. Platform X (twitter) menjadi salah satu media sosial yang sering digunakan oleh masyarakat untuk melakukan diskusi publik, dimana semua pengguna X bisa memberikan tanggapannya, baik mendukung, menolak, maupun

netral terhadap fenomena pernikahan dini ini. Twitter adalah media sosial yang cukup populer di kalangan umum. Hal tersebut disebabkan oleh aplikasinya yang mudah untuk di gunakan, serta kebebasan bagi pemakai untuk menyampaikan pendapat atau opininya [3]. Akan tetapi, rendahnya pemahaman terhadap pola sentimen masyarakat terkait pernikahan dini ini masih belum banyak diteliti secara mendalam. Salah satu fenomena sosial yang cukup umum di Indonesia adalah pernikahan pada usia muda. Masalah ini muncul di berbagai tempat, di wilayah perkotaan ataupun di pedesaan. Berdasarkan data dari Komnas Perempuan, selama tahun 2021 tercatat ada 59. 709 adalah data kasus yang mendapatkan dispensasi dari pengadilan. Ada berbagai keyakinan di kalangan masyarakat Indonesia bahwa pandangan dan kepercayaan yang keliru dapat mendorong terjadinya pernikahan pada usia muda [4].

Penelitian ini dilakukan untuk mengukur sentimen dengan melihat persepsi masyarakat terhadap pernikahan dini melalui platform X. Metode yang efektif untuk memahami dinamika opini publik yang terus berkembang dalam skala besar seiring berjalannya waktu yaitu analisis sentimen. Melalui teknik pemrosesan teks, metode ini mampu mengklasifikasikan teks ke dalam beberapa kategori, seperti positif, negatif, dan netral. Sentimen positif dapat mencerminkan dukungan atau penerimaan terhadap pernikahan dini, sementara sentimen negatif sering kali menunjukkan kekhawatiran, kritik, atau penolakan terhadap praktik tersebut. Di sisi lain, sentimen netral menyajikan pernyataan yang informatif tanpa kecenderungan emosional tertentu. Dengan melakukan analisis sentimen, kita dapat memperoleh wawasan yang lebih mendalam mengenai pola pikir masyarakat, hal-hal yang mempengaruhi pendapat umum, juga pernikahan dini yang dipandang dalam berbagai diskusi di media sosial.

Berbagai penelitian sebelumnya telah mengungkapkan bahwa *BiLSTM* merupakan metode yang sangat efektif dalam klasifikasi teks, khususnya pada analisis sentimen, penerjemahan mesin, dan berbagai aplikasi pemrosesan bahasa alami *NLP* lainnya. Penelitian sebelumnya menunjukkan bahwa penerapan *BiLSTM* menghasilkan kinerja yang baik, dan juga membandingkannya dengan berbagai metode lainnya [5]. Salah satu alasan utama keunggulan *BiLSTM* dibandingkan dengan metode lainnya adalah kemampuannya dalam memahami

konteks kata dari dua arah yaitu dari urutan kata yang ada sebelum (forward) maupun setelahnya (backward). Sentimen dengan menggunakan metode BiLSTM yang mampu merepresentasikan informasi kontekstual dengan lebih baik dan menganalisis umpan balik pengguna dari kedua arah.

Terdapat salah satu penelitian yang menunjukkan efektifitas dari penggunaan metode *BiLSTM*. Penelitian oleh Qurrotu Aini, M Noer, Fadli Hidayat, dan Abu Tholib yang berjudul "Analisis Sentimen Hasil Pemilu (Quick Count) Calon Presiden dan Wakil Presiden 2024 di Media Sosial Media X Menggunakan Metode Bidirectional Long Short-Term Memory *BiLSTM*". Penelitian tersebut melakukan klasifikasi teks dalam beberapa kategori diantaranya positif, negatif, dan netral yang merupakan fokus utama dalam penelitiannya. Tujuan dari penelitian tersebut yaitu untuk menghitung nilai sentimen serta akurasi dengan menggunakan metode *BiLSTM* dalam analisis sentimen terkait hasil dari pemilu. Data didapat melalui scraping pada media sosial X dengan menggunakan kata kunci hasil quick count pilpres 2024 dan hasil pilpres 2024, menghasilkan total 1.348 tweet. Sebagian besar tweet yang dianalisis menunjukkan model BiLSTM berhasil mencapai akurasi sebesar 86,89%, sementara akurasi keseluruhan yang diperoleh adalah 97% [6].

Penelitian lain yang menunjukkan efektifitas dari penggunaan metode *BiLSTM* oleh Onsu R, Sengkey, dan Kambey Yang berjudul "Implementasi Bi-LSTM dengan Ekstraksi Fitur Word2Vec untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital" menjelaskan terkait layanan publik dalam bentuk digital yaitu Aplikasi Identitas Kependudukan Digital (IKD) yang telah diluncurkan mengalami penurunan rating. Tujuan penelitian ini yaitu melakukan analisis sentimen dari ulasan pengguna IKD dengan menggunakan metode (Bi-LSTM) dan Word2Vec. Data yang dikumpulkan mencakup periode satu tahun, yakni dari bulan Januari 2023 hingga Desember 2023, dengan total sebanyak 14.000 ulasan dan rate yang digunakan untuk aplikasi IKD. Hasil menunjukkan bahwa kombinasi Bi-LSTM dan Word2Vec CBOW dengan dimensi 200 serta proporsi data 80/20 menghasilkan akurasi tertinggi mencapai 96,06%. Selain itu, model ini juga menghasilkan nilai precision sebesar 96,44%, recall sebesar 95,64%, dan f1 score sebesar 96,04%. Semua kombinasi Bi-LSTM dan Word2Vec yang diuji

menunjukkan performa yang lebih baik dibandingkan dengan algoritma machine learning lainnya [7].

Terdapat juga pada sebuah penelitian oleh Rolangon A, Weku A, dan Sandag. Pada tahun 2023. Berjudul "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19" tujuan dari penelitian tersebut yaitu menganalisis sikap masyarakat terhadap layanan rumah sakit secara keseluruhan selama masa pandemi Covid-19. Dan menunjukkan hasil menunjukkan bahwa model BiLSTM mencapai akurasi tertinggi sebesar 86%, diikuti oleh model GRU yang juga mencatatkan akurasi 86%. Sementara itu, model LSTM memperoleh akurasi sebesar 85%, dan model SimpleRNN berada di posisi terakhir dengan akurasi 75%. Selain itu, model BiLSTM juga mencatatkan nilai MCC tertinggi, yakni 71% [8].

Klasifikasi sentimen ini menggunakan pendekatan *BiLSTM*. Yang dipercaya mampu memahami konteks kata dari suatu teks secara lebih efektif. Selain itu, untuk ekstraksi fiturnya menggunakan teknik *Word2Vec* yang dapat menghasilkan representasi kata berbasis vektor yang memiliki makna lebih banyak dibandingkan pendekatan klasik lainnya. Dengan penggunaan kedua metode ini, diharapkan bisa memberi hasil analisis sentimen yang akurat. *BiLSTM* merupakan sebuah pengembangan metode dari model *LSTM* yaitu salah satu metode Deep Learning yang digunakan untuk memproses, memprediksi, dan mengklasifikasikan data. *BiLSTM* bekerja dengan melakukan training sebanyak dua kali atau bisa diartikan dengan memproses informasi dalam dua arah yaitu maju (forward) dan juga mundur (backward), yang memungkinkan menangkap info dari dua arah kata sebelum dan sesudah dalam suatu text.

Meskipun *BiLSTM* memiliki keunggulan dengan memahami konteks kati dari suatu teks secara lebih efektif dan luas karena kemampuannya memproses informasi dalam dua arah, akan tetapi *BiLSTM* juga memiliki keterbatasan diantaranya yaitu Diperlukan jumlah data yang cukup besar untuk menghasilkan model yang optimal. Oleh karena itu, pada kasus dengan dataset yang terbatas, performa model *BiLSTM* mungkin tidak akan mencapai hasil yang maksimal. Selain itu *BiLSTM* juga membutuhkan waktu komputasih yang cukup lama karena

BiLSTM harus memproses informasi dari dua arah secara bersamaan. Meskipun dengan kekurangan tersebut tapi bukan berarti model *BiLSTM* tidak dapat menghasilkan output dengan hasil maksimal. Model *BiLSTM* memang memiliki beberapa kekurangan, antara lain kebutuhan akan data yang besar untuk meminimalkan risiko overfitting dan waktu komputasi yang cukup lama. Namun, kelebihan model *BiLSTM* tidak bisa diremehkan, karena mampu mencapai akurasi yang lebih tinggi, yaitu 91%, dibandingkan dengan model *LSTM* biasa yang hanya mencapai 76% [9].

Pada penelitian ini data didapat melalui proses crawling data dari platform X menggunakan teknik API untuk mengumpulkan opini dari masyarakat tentang pernikahan dini. Analisis sentimen ini menggunakan pendekatan (*CRISP-DM*), yang terdiri dari beberapa tahapan utama: pemrosesan data, feature understanding, processing text, ekstraksi fitur, dan modeling. dengan menggunakan metode ini, penelitian dapat dilakukan secara sistematis untuk menghasilkan analisis sentimen yang lebih tepat dan akurat. Penelitian ini yaitu analisis sentimen multi-kelas terhadap pendapat masyarakat mengenai fenomena pernikahan dini pada platform X, dengan memanfaatkan metode *Bi-LSTM* dan ekstraksi fitur *Word2Vec* dengan menerapkan pendekatan *CRISP-DM*. Pada penelitian ini mengevaluasi performa dari metode *BiLSTM* berdasarkan nilai presisi, recall, dan F1-score.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang diatas maka disusunlah rumusan masalah penelitian ini sebagai berikut:

1.2.1 Permasalahan

Pernikahan dini sampai saat ini tentunya masih menjadi permasalahan yang cukup serius di Indonesia, dipengaruhi oleh faktor ekonomi, sosial, dan budaya. semakin banyak pendapat dan opini masyarakat di media sosial, terutama pada platform X, namun belum banyak yang mendalami sentimen masyarakat secara mendalam dengan metode analisis yang lebih kompleks. Oleh karena itu, diperlukan sebuah pendekatan yang dapat mengklasifikasikan opini masyarakat secara lebih akurat untuk memberikan wawasan yang lebih dalam dan meluas terkait fenomena pernikahan dini ini.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode *BiLSTM* dengan ekstraksi fitur *Word2Vec* untuk dilakukan analisis sentimen multi-class terhadap pendapat masyarakat tentang pernikahan dini pada platform X.

1.2.3 Pertanyaan Penelitian

1. Bagaimana pengaruh text augmentasi dibandingkan dengan metode *SMOTE* dalam mengatasi imbalance data pada dataset pernikahan dini berdasarkan performa *BiLSTM* dengan ekstraksi fitur *Word2Vec*?
2. Bagaimana pengaruh epoch dan batch size pada metode *BiLSTM* dengan ekstraksi fitur *Word2Vec* dalam mengklasifikasi sentimen multi-class terkait pernikahan dini?

1.3 Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah diuraikan diatas, adapun tujuan yang ingin dicapai dalam penelitian ini, dengan berbagai manfaat yang berguna sebagai berikut:

1.3.1 Tujuan Penelitian

Berdasarkan permasalahan yang ada, sebagai berikut:

1. Menganalisis pengaruh penggunaan teknik *SMOTE* dalam penyeimbangan data terhadap performa model klasifikasi sentimen multi-kelas menggunakan *BiLSTM* dan *Word2Vec* terhadap fenomena pernikahan dini pada platform X.
2. Menganalisis seberapa efektif variasi jumlah epoch dan ukuran batch size terhadap akurasi dan performa model *BiLSTM* dengan ekstraksi fitur *Word2Vec* dalam klasifikasi sentimen multi-class terhadap fenomena pernikahan dini.

1.3.2 Manfaat penelitian

Berikut adalah manfaat yang diperoleh melalui penelitian yaitu:

1. Memberi wawasan atau informasi terkait opini publik terhadap pernikahan dini berdasarkan data dari platform x.
2. Memberikan referensi atau pemahaman terkait penggunaan metode *BiLSTM* dan *Word2Vec* dalam analisis sentimen multi-class.

1.4 Batasan Masalah

Berikut ini merupakan batasan masalah yang ada pada tahap pengerjaan penelitian ini:

1. Dataset yang digunakan dalam penelitian ini adalah hasil dari crawling berbasis API Twitter dengan bantuan tool tweet-harvest.
2. Dataset hanya berdasarkan kata kunci tertentu dalam rentang waktu yang ditentukan.
3. Pada tahap crawling dataset yang diambil dibatasi limitnya, *unverified accounts* 600/day, *new unverified accounts* 300/day.

1.5 Sistematika Penulisan

BAB I: Pendahuluan

Berisi Latar Belakang, Perumusan Masalah, Tujuan dan Manfaat, Batasan Masalah, dan Sistematika Penulisan.

BAB II: Kajian Pustaka

Berisi Landasan teori yang membahas terkait Pernikahan Dini, Media Sosial dan Analisis Sentimen , Text Mining, Algoritma *BiLSTM*, Teknik *WordVec*, Penelitian terkait.

BAB III: Metodologi Penelitian

Menjelaskan tentang Tahapan Penelitian, Pengumpulan Data, Preprocessing Data Ekstraksi Fitur dengan *Word2Vec*, Implementasi Model *BiLSTM*, dan Evaluasi Model

BAB IV: Hasil dan Pembahasan

Menampilkan hasil pengumpulan data, pelabelan, hasil *preprocessing*, balancing teks augmentasi, Ekstraksi fitur, *SMOTE*, hasil klasifikasi model *BiLSTM*, hasil evaluasi, serta pembahasan pengaruh parameter epoch dan batch size.

BAB V: Penutup

Berisi kesimpulan yang menjawab tujuan dan pertanyaan penelitian, serta saran untuk penelitian selanjutnya.