

1. Pendahuluan

Latar Belakang

Pemilihan Kepala Daerah (Pilkada) adalah proses di mana masyarakat memilih secara langsung gubernur, bupati, dan/atau walikota, yang telah diatur dalam Undang-Undang Nomor 32 tahun 2004 tentang pemerintahan daerah. Proses pemilihan ini menunjukkan pentingnya demokrasi langsung, di mana rakyat memiliki hak untuk menentukan pemimpinnya. Namun, birokrasi yang buruk dapat menghambat jalannya proses demokrasi, bahkan mengancam hak tersebut, karena birokrasi memegang peranan strategis dalam pengambilan keputusan dan pelayanan publik yang mendukung proses pemilihan tersebut. Dengan demikian, keberhasilan Pilkada sangat bergantung pada kualitas birokrasi yang ada dalam pemerintahan [1].

Pilkada merupakan salah satu momen penting dalam demokrasi Indonesia, hasil Pilkada akan menentukan arah pembangunan dan kesejahteraan masyarakat di daerah selama lima tahun kedepan[19]. Setiap calon berhak memiliki tim kampanye yang bertugas mempromosikan visi misi, program, dan janji kepada masyarakat[19]. Oleh karena itu penting bagi pemangku kepentingan untuk memahami opini publik terhadap Pilkada 2024 agar dapat mengambil keputusan yang tepat. Namun, untuk dapat memilih pemimpin yang tepat masyarakat memerlukan informasi yang akurat dan komprehensif mengenai para kandidat[20].

Opini publik yang beredar di sosial media khususnya X, dapat menjadi informasi yang berharga [2]. Dengan mengetahui sentimen publik yang beredar di media sosial kita dapat mengetahui apa saja yang menjadi perhatian masyarakat, rekam jejak para kandidat, dan harapan masyarakat terhadap calon pemimpinnya. Analisis sentimen dapat membantu untuk mengidentifikasi isu-isu penting yang menjadi perhatian masyarakat, informasi ini dapat digunakan untuk meningkatkan partisipasi pemilih dan meningkatkan kualitas demokrasi[21].

Dalam penelitian sebelumnya yang dilakukan oleh Haque dkk yang membandingkan beberapa metode Neural Network, CNN menunjukkan kinerja yang signifikan dibandingkan dengan LSTM dan LSTM-CNN untuk analisis sentimen[3]. Untuk memprediksi sentimen positif, negatif, dan netral, langkah pertama adalah mengidentifikasi sentimen tersebut, yang dilakukan menggunakan CNN[3]. Menganalisis sentimen dalam data X dibatasi oleh adanya teks tidak terstruktur, namun pendekatan ekstraksi fitur *Term Frequency Inverse Document Frequency* (TF-IDF) digunakan untuk memberi bobot pada setiap kata dan representasi setiap kata[22].

Dalam pemrosesan bahasa alami, vektor kata digunakan untuk merepresentasikan kata dalam format padat. Representasi ini biasanya disebut *word embeddings*, meskipun tidak tergolong baru *word embeddings* populer sejak google merilis Word2vec pada 2013 kemudian muncul *word embeddings* lain seperti glove dan fasttext[16]. CNN tidak dapat menerima masukan bahasa alami manusia secara langsung oleh karena itu fasttext memproses bahasa alami menjadi vektor yang akan diteruskan kedalam program[4]. Penelitian lain yang dilakukan oleh Chalkidis dkk [5] untuk meneliti adaptasi awal analisis hukum menggunakan *deep learning* dan *word embeddings*, menyimpulkan bahwa fasttext dapat menangani out-of-vocabulary (OOV) dan kata-kata langka yang tidak dapat diatasi Word2vec dan glove.

Penelitian yang telah dilakukan oleh RA Rudiyanto dkk [6] menggunakan CNN dan *Particle Swarm Optimization* untuk menganalisis sentimen menyatakan bahwa kinerja akurasi sering mengalami penurunan ketika berhadapan dengan

data yang ambigu, kata-kata yang sulit dikenali, dan kondisi serupa lainnya. Oleh karena itu, diperlukan perpaduan algoritma yang optimal[6]. Penggunaan PSO tidak hanya meningkatkan akurasi dan nilai skor f1 tetapi juga menunjukkan perkembangan yang signifikan dibandingkan dengan metode lain yaitu CNN, TF-IDF, dan Word2vec. Metode *bat algorithm* yang mempunyai kemampuan ekolokasi kelelawar dapat membantu CNN untuk mengoptimalkan hyperparameter CNN, memilih fitur yang relevan, dan mengoptimalkan arsitektur CNN untuk meningkatkan kinerja klasifikasi sentimen[7].

Penelitian sebelumnya yang dilakukan oleh Jahan dkk mengevaluasi lima algoritma machine learning pada dataset yang terdiri dari 100.000 tweet untuk analisis sentimen, menunjukkan bahwa Logistic Regression dan Support Vector Machines (SVM) meraih akurasi tertinggi (86.22%), diikuti oleh Random Forest (82.59%)[23]. Sementara itu, Naive Bayes dan Gradient Boosting memiliki akurasi lebih rendah (masing-masing 70.45% dan 69.96%). Semua model menghadapi kesulitan dalam mengklasifikasikan sentimen negatif, yang menunjukkan tantangan dalam menangkap nuansa emosional pada data media sosial, dan mendorong pengembangan lebih lanjut untuk meningkatkan akurasi analisis sentimen[23].

Dalam penelitian yang dilakukan oleh Reddy dkk *Bat Algorithm* (BAT) digunakan untuk mengoptimalkan *Economic Load Dispatch* (ELD) dalam sistem tenaga listrik[24]. ELD adalah proses yang krusial untuk memastikan operasi pembangkitan tenaga yang efisien dan kost-efektif, dengan tujuan untuk mengurangi biaya operasional dan mendistribusikan beban secara optimal di antara beberapa generator dalam sistem tenaga[24]. BAT, yang terinspirasi oleh perilaku kelelawar dalam mencari mangsa, diterapkan untuk menemukan solusi optimal dalam masalah distribusi beban ini. Dengan demikian, BAT digunakan untuk mengoptimasi sistem tenaga listrik agar lebih efisien, mengurangi biaya operasional, dan meningkatkan kinerja pembangkitan tenaga secara keseluruhan.

Untuk melanjutkan penelitian sebelumnya, penelitian ini bertujuan untuk menyempurnakan dan memperkuat hasil penelitian sebelumnya dengan menyelesaikan permasalahan data yang ambigu, kata-kata yang sulit dikenali, serta optimasi parameter melalui penerapan model CNN yang dipadukan dengan word embedding FastText dan *Bat Algorithm* Optimization dengan harapan mencapai akurasi yang lebih tinggi. Analisis sentimen dalam konteks Pilkada telah banyak dikembangkan, namun sejauh pengetahuan penulis, masih terdapat beberapa aspek yang belum ditangani secara optimal, seperti optimalisasi hyperparameter dan penanganan kata-kata yang jarang muncul atau tidak dikenal. Oleh karena itu, penelitian ini diarahkan untuk mengimplementasikan sistem analisis sentimen Pilkada 2024 yang lebih akurat dengan memanfaatkan deep learning dan algoritma optimisasi. Algoritma Bat, yang masih jarang diterapkan untuk analisis sentimen dalam bahasa Indonesia, diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan kinerja model.

Topik dan Batasannya

Penelitian ini mengimplementasikan sistem analisis sentimen untuk Pilkada 2024 di Indonesia menggunakan Convolutional Neural Network (CNN), ekspansi fitur FastText, dan *Bat Algorithm* sebagai optimasi. Analisis dilakukan pada tweet berbahasa Indonesia dari media sosial X selama periode Pilkada, sementara tweet di luar periode atau berbahasa lain tidak dipertimbangkan.

Tujuan

Tujuan utama dari penelitian ini adalah untuk mengimplementasikan analisis sentimen pada data media sosial X menggunakan metode Convolutional Neural Networks (CNN) yang dioptimalkan dengan *Bat Algorithm* dan Ekspansi Fitur melalui FastText. Penelitian ini juga bertujuan untuk mengukur nilai akurasi

implementasi FastText, serta sebelum dan sesudah optimasi dengan *Bat Algorithm*.