

## Misogyny Text Detection on Tiktok Social Media in Indonesian Using the Pre-trained Language Model IndoBERTweet

Perwira Hanif Zakaria<sup>1</sup>, Dade Nurjanah<sup>2</sup>, Hani Nurrahmi<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>perwiranifz@students.telkomuniversity.ac.id, <sup>2</sup>dadenurjanah@telkomuniversity.ac.id,

<sup>3</sup>haninurrahmi@telkomuniversity.ac.id

---

### Abstrak

Media sosial merupakan platform komunikasi dan informasi yang populer karena kemudahan dan kecepatan aksesnya. Dengan menggunakan media sosial, seseorang dapat mengekspresikan dirinya dengan bebas. Hal ini memicu oknum yang tidak bertanggung jawab untuk melontarkan ujaran kebencian dengan tujuan menjatuhkan seseorang atau sekelompok orang. Misogini adalah salah satu bentuk ujaran kebencian yang ditujukan kepada perempuan. Masalah misogini tidak boleh dianggap remeh karena misogini bisa menjadi salah satu penyebab utama wanita merasa sengsara. Pada penelitian ini akan dibangun sebuah model untuk mendeteksi teks misogini pada media sosial TikTok berbahasa Indonesia dengan menggunakan model IndoBERTweet pre-trained. IndoBERTweet adalah model pra-pelatihan berdasarkan model BERT, yang telah dilatih menggunakan dataset bahasa Indonesia yang diambil dari media sosial Twitter sebelumnya, menghasilkan kinerja yang baik untuk mendeteksi teks misogini di media sosial dengan mengklasifikasikannya. Dataset yang digunakan berupa data teks yang diambil dari komentar misogini dengan memfokuskan pada bentuk misogini berupa stereotip, dominasi, pelecehan seksual, dan mendiskreditkan dalam konten video pendek di akun media sosial TikTok perempuan. Performansi model yang dibangun melakukan pengaturan hyperparameter yang meliputi batch size 16, epochs 10, dan learning rate  $7e-5$  dan dievaluasi menggunakan matriks konfusi dengan hasil akurasi terbaik sebesar 76,89%

**Kata kunci:** Misogini, BERT, *Pre-trained Model*, IndoBERTweet

---

### 1. Pendahuluan

Media sosial merupakan salah satu platform komunikasi dan informasi yang paling populer saat ini. Kepopuleran media sosial tidak lepas dari visual, kemudahan, dan kecepatan akses yang menjadi daya tarik utamanya. Dengan menggunakan media sosial, seseorang dapat mengekspresikan dirinya dengan bebas tanpa ada batasan. Kebebasan berekspresi ini memicu banyak oknum yang tidak bertanggung jawab untuk melakukan kejahatan berupa ujaran kebencian secara online yang bertujuan untuk mem-bully atau menjatuhkan seseorang atau sekelompok orang [1].

Salah satu media sosial yang sedang menjadi trend di Indonesia adalah TikTok. Menurut Statista, situs pengumpul data statistik media sosial di dunia, tercatat hingga Juli 2022, Indonesia menempati urutan kedua jumlah pengguna TikTok dengan total 99 juta pengguna aktif.

Ujaran kebencian dilakukan secara daring melalui media sosial tanpa memandang jenis kelamin. Menurut OXIS, sebuah situs survei Oxford, baik laki-laki maupun perempuan masih bisa menjadi sasaran ujaran kebencian online oleh orang-orang yang tidak bertanggung jawab, namun perempuan masih lebih mungkin menerima ujaran kebencian jika dibandingkan dengan laki-laki.

Misogini adalah salah satu bentuk negativitas yang sering dibagikan atau diberikan di media sosial. Misogini adalah salah satu bentuk ujaran kebencian terhadap perempuan, baik ditujukan kepada individu maupun kelompok. Selain itu, misogini adalah masalah yang tidak bisa dianggap remeh. Ini karena kebencian terhadap wanita adalah alasan utama wanita di seluruh dunia merasa sengsara [1], [2]. Misogyny dapat dikategorikan ke dalam beberapa bentuk perilaku, seperti Stereotype, Dominance, Sexual Harassment, dan Discredit [3]. Perilaku misogini semakin marak seiring dengan berkembangnya media sosial sebagai wadah untuk menyampaikan pendapat secara bebas dan anonim, terutama di masa pandemi COVID-19 [4]. Perilaku misogini yang terus meningkat setiap tahunnya menjadi perhatian yang tidak boleh diremehkan, dan harus dicarikan solusi untuk mengatasinya.

Misogini dapat dikategorikan ke dalam beberapa bentuk perilaku seperti mempermalukan, pelecehan seksual, dominasi, dan mendiskreditkan. Mempermalukan adalah bentuk misogini yang membatasi atau meremehkan perempuan karena beberapa ciri fisik. Pelecehan seksual adalah bentuk kebencian terhadap wanita dalam bentuk permintaan atau pernyataan untuk melakukan tindakan yang diarahkan secara seksual, seperti komentar seksual, lelucon kasar, dan ajakan terus-menerus untuk melakukan perselingkuhan yang dapat membuat Anda tidak nyaman. Diskredit dan dominasi merupakan bentuk misogini berupa ekspresi dan/atau pernyataan yang kasar bahwa laki-laki lebih superior dari perempuan [1], [5], [6].

Deteksi teks misogini adalah tugas yang dilakukan untuk mendeteksi apakah suatu teks atau kalimat merupakan ekspresi misogini atau tidak. Dalam hal deteksi atau pengenalan teks, tugas deteksi teks misogini