

BAB I

PENDAHULUAN

1.1 Latar Belakang

Platform e-commerce merupakan salah satu contoh bentuk belanja *online*, yang memungkinkan pengguna untuk melakukan proses belanja secara cepat dan praktis[1]. Dalam proses belanja *online*, ulasan oleh pengguna menjadi salah satu faktor yang mempengaruhi keputusan pembelian suatu produk. Menurut penelitian yang dilakukan oleh Nurbaiti dkk. [2], menunjukkan bahwa 76% konsumen membaca ulasan sebuah produk sebelum melakukan pembelian. Hal tersebut disebabkan ketika konsumen melakukan pembelian secara *online*, mereka tidak bisa merasakan dan melihat kualitas produk secara langsung seperti halnya belanja secara konvensional[3].

Salah satu *platform* yang menyediakan layanan terkait *ulasan online* adalah Female Daily. Female Daily adalah *platform website* yang menyediakan layanan informasi dan ulasan seputar dunia kecantikan, seperti *makeup*, *skin care*, *hair care* dan *beauty tools*. Melalui *platform* tersebut, pengguna dapat menemukan rekomendasi produk dan membaca artikel seputar dunia kecantikan serta tips dan trik. Pada tahun 2024, Female Daily berhasil mengumpulkan sebanyak 2.506.000 ulasan[4]. Dengan adanya fitur ulasan pelanggan *online* pada *website* Female Daily, hal tersebut memungkinkan pengguna untuk mencari informasi dan mengulas suatu produk.

Namun, seiring dengan meningkatnya popularitas Female Daily sebagai *platform* kecantikan, hal tersebut menimbulkan fenomena terkait dengan adanya ulasan palsu. Hal tersebut seperti yang disampaikan oleh akun @leyxmun di media sosial X pada 17 September 2024, bahwa dia mendapatkan kerja sama dari sebuah *brand* untuk mempromosikan *brand* tersebut melalui konten Video TikTok dan ulasan di Female Daily. Namun pemilik akun tersebut merespon dengan “*Udah berapa kali nolak kerja sama minta sekalian di female daily juga, tentu saja aku tolak karena itu kan platform yg ktanya honest ulasan harus beneran beli pake duit sendiri kann???*”. Hal tersebut menunjukkan bahwa terdapat beberapa *brand* yang mengadakan kampanye berbayar orang untuk memberikan *ulasan* produk di

Female Daily. Keluhan serupa juga disampaikan oleh pemilik akun X @liacwh “aku dah ga percaya review di fd, pada template kebanyakan, gatau mana yg buzzer mana yg asli” pada 18 September 2024.



Gambar 1.1 Keluhan Pengguna Female Daily

Adanya ulasan palsu dapat merusak persepsi konsumen yang objektif dan akurat terhadap suatu produk. Hal tersebut menunjukkan bahwa ulasan palsu dapat mempengaruhi keputusan pembelian dengan mengubah persepsi konsumen terhadap nilai produk[5]. Menurut Song dkk [5], ulasan palsu memiliki beberapa karakteristik, seperti isi konten yang berulang, waktu unggah yang berdekatan, serta gaya bahasa yang terlalu ekspresif atau berlebihan. Ulasan semacam ini biasanya tidak memberikan informasi yang relevan dan cenderung bersifat promosi.

Untuk mengetahui ulasan yang asli atau palsu, maka perlu dibangun sebuah model untuk mengklasifikasikan ulasan yang asli dan palsu. Sebelumnya, telah terdapat berbagai penelitian yang meneliti terkait klasifikasi ulasan asli dan palsu, salah satunya adalah penelitian Analisis Ulasan Palsu pada Produk Female Daily dengan Pendekatan *Natural Language Processing* oleh [6]. Pada penelitian tersebut, peneliti melakukan analisis ulasan palsu pada 10 produk Whitelab di *platform* Female Daily. Jumlah ulasan yang akan diproses adalah sebanyak 3041 ulasan dengan menggunakan algoritma *Decision Tree* dan *Logistic Regression*. Pada penelitian tersebut, peneliti melakukan analisis ulasan palsu yang didasarkan pada analisis sentimen dan *rating* ulasan produk. Hasil dari penelitian tersebut menunjukkan bahwa ulasan dengan rating bintang 5 merupakan ulasan yang paling dominan, namun sebanyak 68,8% ulasan yang ada memiliki sentimen yang tidak terdefinisi sehingga diindikasikan sebagai ulasan palsu. Terdapat penelitian lain terkait dengan klasifikasi ulasan asli dan palsu oleh Kurtcan dkk. [7] yaitu *Classification of Authentic and Fake Online Reviews with Supervised Machine Learning Techniques* oleh. Pada penelitian tersebut, peneliti menggunakan dataset dari *website online ulasan* yaitu Yelp dan TripAdvisor untuk ulasan asli, sedangkan untuk dataset ulasan palsu didapatkan dari Amazon Mechanical Turk. Ulasan tersebut kemudian diproses dengan menggunakan beberapa algoritma *supervised learning* yaitu: *Logistic Regression*, *Linear Discriminant Analysis*, *K-Nearest Neighbors*, *Naive Bayes*, *Decision Tree*, *Bootstrap Aggregating*, *Random Forest*, *Support Vector Machine*, dan *Neural Network*. Dari berbagai algoritma tersebut, model yang memiliki performa paling baik adalah *Random Forest*.

Pelabelan manual terhadap seluruh ulasan membutuhkan waktu dan sumber daya yang besar, sedangkan sebagian besar ulasan yang tersedia belum memiliki label kelas asli atau palsu. Kondisi tersebut sesuai dengan pendekatan *semi-supervised learning* yang menggabungkan sejumlah data berlabel dengan data tidak berlabel. Cascante-Bonilla dkk. [8] melalui penelitian *Curriculum Labeling: Revisiting Pseudo-labeling for Semi-supervised Learning* menunjukkan bahwa *pseudo-labeling* dapat dimanfaatkan pada kondisi label terbatas dan melaporkan eksperimen ImageNet menggunakan 10% data berlabel.

Kualitas *pseudo-label* menjadi hal penting karena prediksi model terhadap data tidak berlabel dapat mengandung kesalahan dan kemudian memperkuat kesalahan tersebut saat digunakan kembali dalam proses *training*. Rizve dkk. [9] mengusulkan seleksi *pseudo-label* berbasis *uncertainty* agar model tidak hanya menggunakan *pseudo-label* berdasarkan prediksi kelas, tetapi juga mempertimbangkan tingkat ketidakpastian prediksi. Berdasarkan hal tersebut, penelitian ini membandingkan *Standard Pseudo-label*, *Confidence Pseudo-label*, *Uncertainty Pseudo-label*, serta kombinasi *Confidence* dan *Uncertainty* dengan *Random Forest* sebagai base classifier.

Berdasarkan uraian tersebut, penelitian ini membangun model klasifikasi ulasan palsu pada *website* Female Daily menggunakan *Random Forest* dengan menggunakan metode *semi-supervised pseudo-labeling*. pada penelitian ini, dataset yang digunakan adalah produk Implora Day to Day Series. Implora Day to Day Series dipilih sebagai objek penelitian dengan mempertimbangkan relevansi merek pada pasar kosmetik Indonesia, keberagaman jenis produk dalam satu seri, serta ketersediaan ulasan konsumen pada Female Daily. Day to Day Series terdiri atas empat produk yang digunakan dalam penelitian, yaitu *Two Way Cake*, *Keychain Lip Bullet*, *Lite Matte Lip Cream*, dan *Brow Pop*. Keempat produk tersebut mewakili beberapa kategori *makeup*, yaitu *complexion*, produk alis, dan produk bibir. Keberagaman jenis produk tersebut memungkinkan penelitian memperoleh ulasan dengan penggunaan yang lebih bervariasi dibandingkan penggunaan satu jenis produk saja.

Pemilihan Implora juga didukung oleh relevansi merek tersebut pada pasar kosmetik Indonesia. Pada tahun 2024, PT Implora Sukses Abadi memperoleh lima penghargaan dalam ajang *Indonesia Best Selling Brand in E-Commerce Market 2024* yang diselenggarakan oleh SWA bersama Kompas.co.id. SWA Media merupakan media bisnis nasional yang telah beroperasi sejak 1985 dan berfokus pada informasi serta analisis bisnis, sedangkan Kompas.co.id merupakan penyedia *e-commerce market intelligence* yang melakukan pemantauan dan analisis performa pasar e-commerce. Penghargaan tersebut meliputi *1st Winner* kategori *Lip Tint*, *3rd Winner* kategori *Eye brow*, *3rd Winner* kategori *Lipstick*, *3rd Winner* kategori *Nail Polish*, dan *Top 10* kategori *Hair Colour* [10]. Selain itu, Implora

memperoleh predikat *Very Good* dalam ajang Indonesia *Original Brand Champion* melalui materi mengenai *Digital Impact* pada kampanye peluncuran Implora Day to Day Series [10]. Pencapaian pada beberapa kategori produk tersebut menunjukkan adanya aktivitas dan relevansi pasar Implora dalam industri kosmetik Indonesia. Aktivitas pasar tersebut memberikan potensi terjadinya interaksi konsumen terhadap produk, salah satunya melalui pemberian ulasan setelah penggunaan produk.

Hasil dari penelitian ini diharapkan dapat meningkatkan akurasi dalam klasifikasi ulasan palsu pada produk kecantikan dan membantu konsumen untuk mengetahui produk yang memang sesuai dengan yang dibutuhkan. Luaran dari penelitian adalah sebuah model klasifikasi untuk mengklasifikasikan ulasan palsu dengan menggunakan algoritma *Random Forest* pada sebuah aplikasi *website* dengan menggunakan framework *Flask* yang memungkinkan pengguna untuk menginput ulasan dan mengetahui hasil klasifikasi secara langsung.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan pada penelitian ini, maka rumusan masalah adalah sebagai berikut:

1. Bagaimana membangun model klasifikasi ulasan palsu pada *website* Female Daily menggunakan algoritma *Random Forest*?
2. Bagaimana performa dari hasil pemodelan klasifikasi ulasan palsu pada *website* Female Daily menggunakan algoritma *Random Forest*?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Membangun model klasifikasi ulasan palsu pada *website* Female Daily menggunakan algoritma *Random Forest*
2. Mengetahui performa dari hasil pemodelan klasifikasi ulasan palsu pada *website* Female Daily menggunakan algoritma *Random Forest*

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini meliputi manfaat teoretis dan manfaat teknis, sebagai berikut:

1. Manfaat Teoretis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan klasifikasi ulasan palsu menggunakan pendekatan *machine learning*, khususnya algoritma Random Forest dengan metode *semi-supervised learning pseudo-labeling*.

2. Manfaat Teknis

Penelitian ini diharapkan menghasilkan model klasifikasi yang mampu memberikan prediksi ulasan asli atau palsu terkait suatu produk yang dapat diimplementasikan melalui aplikasi berbasis *website*, sehingga hal tersebut bisa memberikan pertimbangan kepada pengguna Female Daily terkait keputusan pembelian suatu produk.

1.5 Batasan Masalah

Terdapat beberapa batasan masalah yang diterapkan dalam penelitian ini agar lebih terarah, yaitu:

1. Penelitian ini menggunakan dataset yang didapatkan melalui teknik *web scraping* pada *website* Female Daily dari produk Implora *series Day to Day*, yaitu *Day to Day Series Two Way Cake*, *Day to Day Series Keychain Lip Bullet*, *Day to Day Series Brow Pop*, dan *Day to Day Series Lip Matte Lip Cream*.
2. Data yang dikumpulkan sebanyak 5724 data dengan atribut data yang diambil dari pada rentang 1 Desember 2023 sampai 1 April 2025 melalui *website* Female Daily. Atribut yang diambil adalah nama produk, *shade* produk, tanggal *review*, *username*, *profile age*, *profile description*, *usage period*, *purchase point*, *stars*, *recommend*, *review*.
3. Penelitian ini hanya menggunakan dua kelas, yaitu kelas asli dan kelas palsu. Kelas asli digunakan untuk data yang merupakan ulasan asli, sedangkan kelas palsu digunakan untuk data yang merupakan ulasan palsu.

1.6 Sistematika Penulisan

Dalam penulisan ini, pembahasan disajikan dalam lima bab dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Bab I berisi gambaran umum mengenai latar belakang masalah, rumusan masalah, batasan masalah penelitian, tujuan penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab II berisi teori yang mendukung penelitian dan uraian terkait penelitian terdahulu sebagai referensi dan dasar dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab III berisi metode yang digunakan dalam penelitian ini. Metode tersebut menguraikan rancangan penelitian, teknik pengumpulan data, hingga evaluasi performa model.

BAB IV HASIL DAN PEMBAHASAN

Bab IV berisi mengenai penjelasan dari hasil masing-masing tahapan yang dilakukan dalam penelitian.

BAB V PENUTUP

Bab V berisi kesimpulan dari penelitian yang sudah dilakukan dan rekomendasi dari penulis untuk pengembangan penelitian di masa yang akan datang.

Halaman ini sengaja dikosongkan