

Klasifikasi Jenis Barang Bekas menggunakan Metode Naïve Bayes dengan Seleksi Fitur *Information Gain* (Studi Kasus : Akun Instagram Jual Beli Barang Bekas @infobarkas_Jogja)

Muhammad Fauzan Ziqroh¹, Indriati², Edy Santoso³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email : ¹fziqroh@gmail.com, ²indriati.tif@ub.ac.id, ³edy144@ub.ac.id

Abstrak

Instagram merupakan salah satu media sosial yang populer untuk pemasaran. Berkembangnya fungsi media sosial ini memunculkan istilah baru yaitu *Social Media Influencer*, yaitu agensi periklanan yang memiliki *followers* akun *instagram* dalam jumlah tertentu dimana pengiklan bisa menempatkan iklannya pada akun tersebut dengan membayar sejumlah biaya tertentu. Akun *instagram* @infobarkas_jogja merupakan salah satu *social media influencer* yang menyediakan jasa iklan berbayar dengan konten iklan seputar barang bekas yang ada di Kota Yogyakarta. Pengelolaan akun tersebut terdapat beberapa kendala salah satunya yaitu klasifikasi atau mengelompokkan barang berdasarkan jenis kategorinya. Tujuan penelitian ini yaitu membuat sistem yang mampu mengklasifikasikan jenis barang berdasarkan kategorinya. Penelitian ini akan menggunakan metode *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain*. Data penelitian berupa teks pada *caption* postingan pada akun *instagram* @infobarkas_jogja sebanyak 500 data. Dengan total data latih sebesar 400 data dan data uji sebesar 100 data. Kelas untuk kategori pada penelitian ini dibagi menjadi 5 kelas yaitu properti, kendaraan, pakaian, perabot, dan *gadget*. *Threshold* yang digunakan berjarak 10% mulai dari 10% hingga 90% dan menghasilkan akurasi tertinggi sebesar 98% ketika *threshold* bernilai 10%, 40%, 80%, dan juga 90%. Akurasi tertinggi juga diperoleh ketika klasifikasi tanpa menggunakan seleksi fitur.

Kata kunci: *Caption, Instagram, Klasifikasi, Information Gain, Threshold, Naïve Bayes*

Abstract

Instagram is one of the most popular social media for marketing. The rapid development of this social media function gave rise to a new term, known as Social Media Influencer. Social media influencers are advertising agencies that reach a certain number of Instagram account followers so that advertisers can place their ads on the account by paying a certain amount of fees. Instagram account of @infobarkas_jogja is one of the social media influencers that provides paid advertising services with advertising content about used goods in Yogyakarta City. However, the management of the account has several obstacles, one of which is the classification or grouping of the goods based on the types of category. The purpose of this research is to create a system that is able to classify types of goods based on their categories. This research uses the Naïve Bayes Classifier method using the IG feature selection. The data for this study is in the form of text taken from the caption posts on the Instagram account @infobarkas_jogja with a total of 500 data. With a total of 400 training data and 100 testing data. The classes for the categories in this study were divided into 5 classes, namely property, vehicles, clothing, furniture, and gadgets. The threshold used is 10% ranging from 10% to 90% and produces the highest accuracy of 98% when the threshold is 10%, 40%, 80%, and also 90%. The highest accuracy is also obtained when carrying out classification without resorting to feature selection.

Keywords: *Captions, Instagram, Classification, Information Gain, Threshold, Naïve Bayes*

1. PENDAHULUAN

Akses internet yang semakin mudah menyebabkan pengguna internet meningkat

setiap tahunnya. Penggunaan internet saat ini tidak lepas dari penggunaan social media. Berdasarkan hasil riset Statista Research Department (2022) ketika tahun 2022 jumlah

pengguna internet Indonesia berjumlah 204,7 juta jiwa. Adapun jumlah orang yang menggunakan media sosial di Indonesia berjumlah 61,8% dari jumlah populasi masyarakat Indonesia saat ini. Dari hasil riset tersebut bisa disimpulkan bahwa sosial media di Indonesia sebagai tempat pemasaran memiliki peluang besar karena memiliki jangkauan yang sangat besar.

Media sosial telah menjadi salah satu sarana yang digunakan untuk melakukan pemasaran produk atau yang biasa juga disebut dengan social media marketing (Mileva and Fauzi, 2018). Dengan kemudahan interaksi antara pemilik bisnis dan target pasar membuat media sosial sangat cocok untuk pemanfaatan dalam hal marketing. Dari sekian banyaknya media sosial yang ada saat ini, instagram merupakan salah satu media sosial yang banyak digemari saat ini.

Instagram merupakan media sosial yang dipergunakan untuk membagikan sebuah foto maupun video secara singkat yang nantinya dapat dibagikan dan dilihat oleh sesama pengguna lainnya melalui jaringan internet (Erviahan, 2021). Menurut Erviahan (2021), Instagram berperan sebagai media sosial yang dapat sebagai sarana berbagi foto dan video singkat yang dapat dilihat oleh penggunanya yang lain. Dalam instagram terdapat fitur berbagi foto dan video atau yang bisa disebut sebagai instagram post. Selain itu instagram juga memiliki fitur terbaru yaitu instagram story yang mendukung efektifitas dalam hal pemasaran. Sehingga media sosial yang tepat untuk mengiklankan sebuah produk salah satunya yaitu Instagram. Dengan perkembangan fungsi media sosial instagram sebagai salah satu media pemasaran maka lahirlah social media influencer.

Social media influencer pada instagram sendiri merupakan agensi periklanan yang memiliki followers akun instagram dalam jumlah tertentu dimana pengiklan bisa menempatkan iklannya pada akun tersebut dengan membayar biaya tertentu (Bisma, 2018). Dari sumber data Social Research dan Monitoring Social, Kadin, Kemkominfo, Accenture tahun 2015 dari jumlah pengguna internet di Indonesia produk e-commerce yang populer di cari konsumen adalah pakaian sebesar 67, 10 %, sepatu 20,20 %, Tas 20 %, Jam 7,60 %, Tiket pesawat 5,10 %. Handphone 5,10 %, Aksesoris kendaraan 2,80 %, kosmetik 2,30 % dan Buku 1,80 % (Wellyantama and Krisnadi

2019). Selain dalam bentuk produk baru, barang yang dijual juga dapat berbentuk barang bekas atau yang biasa disebut preloved. Menurut Wellyantama dan Krisnadi (2019), barang bekas merupakan salah satu objek pembelian konsumen di Indonesia yang populer. Konsumen di Indonesia dan negara lain memiliki karakteristik yang tidak berbeda dalam menilai barang bekas berkualitas.

Pengelolaan media sosial khususnya instagram yang menjual barang bekas, masih banyak dijumpai berbagai kendala. Salah satu kendalanya yaitu klasifikasi atau mengelompokkan barang berdasarkan jenis kategorinya. Akun instagram @infobarkas_jogja menyediakan jasa iklan berbayar dengan konten iklan barang bekas atau preloved yang ada di Kota Yogyakarta. Dengan jumlah pengikut 193.000, Instagram @infobarkas_jogja menjadi sebuah akun iklan berbayar yang efektif untuk menyampaikan informasi produk yang akan dijual. Akun instagram @infobarkas_jogja merupakan target pasar pengiklan untuk memasarkan produk barang bekas dengan berbagai macam kategori, barang yang dipromosikan kepada costemer di instagram @infobarkas_jogja sangat beragam mulai dari pakaian, gadget, kendaraan, properti dan lain sebagainya.

Peneliti yang juga saat ini merangkap sebagai owner @infobarkas_jogja merasakan antusiasme followers sangat tinggi terhadap konten @infobarkas_jogja. Terbukti sejak didirikan pada Juli 2018 jumlah konten barang yang dipromosikan sudah mencapai 57.000 kiriman. Instagram @infobarkas_jogja hanya mampu menyalin caption yang diberikan penjual tanpa adanya penambahan kategori barang. Hal ini disebabkan pelayanan masih dilakukan secara manual oleh satu orang yang berposisi sebagai admin. Dengan banyaknya foto dan keterangan barang yang diunggah setiap harinya, sangat tidak memungkinkan untuk admin melakukan proses klasifikasi kategori barang secara manual. Adapun jika penjual diharuskan mengisi kategori barang sendiri, akan berpeluang salah karena adanya beberapa barang yang ambigu dan jumlah kategori yang banyak sehingga diperlukan sebuah machine learning dalam pengambilan keputusannya. Pengolahan data pada caption Instagram sama seperti pengolahan dokumen teks yang merupakan bagian dari ilmu text mining.

Berdasarkan berlimpahnya data iklan pada instagram @infobarkas_jogja, pengolahan

terhadap data barang bekas diharapkan dapat menghasilkan informasi yang berkarakter berdasarkan kategorinya. Pengolahan data barang bekas perlu dilakukan untuk mengetahui informasi secara cepat dan terstruktur yang biasa disebut dengan pengetahuan baru atau knowledge discovery, contohnya informasi mengenai pengklasifikasian data barang bekas dengan kategori-kategori barang yang telah ditentukan. Pengetahuan baru atau yang biasa disebut dengan knowledge discovery tersebut dapat membantu admin instagram @infobarkas_jogja untuk melakukan klasifikasi mengenai tingkat katalogisasi, klasifikasi, penentuan subjek guna mempermudah pihak admin dan costemer untuk mencari info barang bekas, sehingga akan meningkatkan penjualan serta kepuasan kepada pengikut yang menggunakan jasa instagram @infobarkas_jogja.

Untuk menyelesaikan masalah yang telah peneliti jelaskan, maka penelitian kali ini akan menggunakan sebuah metode untuk dapat mengklasifikasikan data berupa teks. Penelitian ini akan menggunakan metode pengklasifikasian sederhana yaitu Naive Bayes Classifier. Menurut Hamzah (2012), metode Naive Bayes Classifier mempunyai dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi.

Selain menggunakan metode Naive Bayes Classifier dalam pengklasifikasian data, untuk meningkatkan akurasi penelitian peneliti menggunakan seleksi fitur Information Gain, dimana Information Gain dapat memberi batasan jumlah fitur yang akan digunakan. Information Gain sendiri merupakan besaran nilai kehadiran dan ketidakhadiran dari suatu kata yang dibatasi dengan threshold, hasil dari Information Gain akan digunakan pada proses klasifikasi selanjutnya.

Penelitian oleh Abdan Syakuro (2017), mengenai analisis sentimen masyarakat terhadap e-commerce pada media sosial menggunakan metode Naive Bayes Classifier (NBC) dengan seleksi fitur Information Gain (IG). Pada penelitiannya menghasilkan tingkat akurasi sebesar 88,8 % dan membuktikan bahwa algoritma tersebut dapat menganalisis sentimen secara otomatis. Hal ini juga membuktikan bahwa kombinasi antara kedua metode yang digunakan terbukti dapat menghasilkan akurasi yang tinggi.

Dari penjabaran kasus diatas peneliti memiliki ketertarikan untuk melakukan penelitian yang diberi judul "Klasifikasi Jenis

Barang Bekas Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Information Gain (Studi Kasus : Akun Instagram Jual Beli Barang Bekas @infobarkas_jogja)".

2. LANDASAN KEPUSTAKAAN

2.1 Media Sosial

Peran media sosial sangat signifikan pada perkembangan bisnis sehingga menjadikannya sebagai alat promosi yang efektif untuk digunakan. Media sosial memiliki kelebihan diantaranya banyak pengguna serta tidak memerlukan biaya untuk membuat akun di sebuah media sosial (Dewi, 2018). Dengan kemudahan berkomunikasi sesama pengguna, saling berbagi informasi, serta dapat membentuk jaringan secara online membuat media sosial banyak digemari dalam hal pemasaran barang atau jasa. Perkembangan fungsi media sosial tersebut menghasilkan istilah baru yang yaitu *social media marketing*.

2.1.1 Social Media Marketing

Sosial media marketing merupakan sebuah teknik marketing yang menggunakan sebuah sosial media sebagai tempat untuk mempromosikan sebuah produk dan sebagai sarana guna membangun target pasar dari bisnis online (Supangat, 2021). Perkembangan dan cakupan pasar yang luas dari media sosial membuat pelaku bisnis tertarik untuk mempromosikan produk dan jasa nya melalui media sosial yang ada. *Social media marketing* sendiri merupakan kegiatan pemasaran dengan memanfaatkan penggunaan media sosial seperti *Instagram*, *Facebook*, *Twitter* dan sebagainya. *Instagram* menjadi salah satu media sosial yang banyak diminati untuk melakukan kegiatan *social media marketing* ini.

2.1.2 Instagram

Instagram memungkinkan penggunanya untuk saling berbagi video atau foto di dalam laman tertentu atau yang bisa disebut sebagai *instagram post*. Selain itu *instagram* juga memiliki fitur terbaru yaitu *Instagram Story* yang mendukung efektifitas dalam hal pemasaran. Sehingga *Instagram* dapat dijadikan pilihan yang tepat untuk mengiklankan sebuah produk. Bahkan Rahman dan Panuju (2017) dalam penelitiannya menyebutkan bahwa kemampuan *instagram* dalam penyajian konten dengan visual yang menarik menjadikan *instagram* sebagai salah satu media sosial yang

sukses digunakan untuk melakukan kegiatan promosi atau berjalan.

2.2 Pre-Processing

Pre-processing data adalah proses pembersihan dan mempersiapkan teks untuk klasifikasi (Haddi dkk, 2013). Data yang diambil dari caption pada *Instagram* merupakan data mentah yang belum layak untuk diolah karena masih memiliki imbuhan, mempunyai karakter tertentu, dan kemunculan beberapa kata yang tidak terlalu mempengaruhi klasifikasi sehingga harus dihapus. *Pre-processing* diawali dengan proses *Case folding* yang merupakan proses untuk mengubah semua huruf menjadi huruf kecil. Selanjutnya proses *Tokenizing* yaitu memecah teks pada dokumen menjadi kata. Selanjutnya ada proses menyeleksi kata-kata penting pada dokumen atau yang biasa disebut dengan *Filtering*. Dan yang terakhir adalah proses *Stemming* dimana imbuhan pada setiap kata dihilangkan sehingga yang tersisa hanyalah kata dasarnya saja.

2.3 Information Gain

Information Gain adalah metode seleksi fitur sederhana dengan cara memeringkat atribut (Chormunge dan Jena, 2016). Menurut Firmansyah dan Gantini (2016) tahapan pertama yang dilakukan yaitu menghitung *entropy* yang dapat dilihat pada Persamaan 2.1, tahapan selanjutnya yaitu menghitung nilai dari *Information Gain* itu sendiri yang bisa dilihat pada Persamaan 2.2.

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i \quad (2.1)$$

Adapun c merupakan total nilai dari kelas yang akan diklasifikasikan dan P_i yaitu total sampel yang digunakan untuk kelas pada indeks ke- i

$$Gain(S, A) = Entropy(S) - \sum_{Values(a)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2.2)$$

Gain (S,A) merupakan nilai *Information Gain* term tertentu. Entropy(S) sendiri merupakan nilai dari entropy total.

$$IG(t) = -(\sum_{i=1}^m P\gamma(ci) \log P\gamma(ci) + P(t) \sum_{i=1}^m P\gamma(ci|t) \log P\gamma(ci|t) + P\gamma(\bar{t}) \sum_{i=1}^m P\gamma(ci|\bar{t}) \log P\gamma(ci|\bar{t}) \quad (2.3)$$

Keterangan :

$P(C_i)$: nilai kemunculan term pada kelas i di dalam seluruh dokumen

$P(t)$: nilai kemunculan term yang ada di dalam suatu dokumen

$P(\bar{t})$: nilai kemunculan term yang tidak ada di dalam suatu dokumen

$P(C_i|t)$: nilai kemunculan term yang muncul pada kelas i di dalam suatu dokumen

$P(C_i|\bar{t})$: nilai ketidakhadiran term pada kelas i di dalam suatu dokumen

2.4 Naïve Bayes Classifier

Menurut Hamzah (2012) Metode Klasifikasi *Naïve Bayes* terbagi menjadi dua tahapan, yaitu tahapan pelatihan dan juga pengujian. Pada tahapan pertama yaitu pelatihan, dilakukan pengenalan pola dengan menghitung nilai kemunculan setiap term terhadap masing-masing dokumen dan juga kelas tertentu. Hasil perhitungan tersebut biasa disebut dengan *Conditional Probability*. Adapaun rumusnya dapat dilihat pada Persamaan 2.4 yaitu sebagai berikut.

$$P(c_j|w_i) = \frac{P(c_j) P(w_i|c_j)}{P(w_i)} \quad (2.4)$$

Keterangan:

$P(c_j|w_i)$: Nilai peluang kelas pada kategori j terhadap kemunculan term i

$P(w_i|c_j)$: Nilai peluang term i masuk ke dalam kelas kategori j

$P(c_j)$: Nilai peluang kemunculan kategori j

$P(w_i)$: Nilai peluang kemunculan suatu term

Setelah dilakukan perhitungan *Conditional Probability*, maka selanjutnya dilakukan perhitungan prior. Adapun prior sendiri merupakan peluang masing-masing kelas kategori pada penelitian kali ini. Rumus untuk menghitung prior tertulis di Persamaan 2.6.

$$P(c_j) = \frac{N_c}{N} \quad (2.5)$$

Keterangan:

N_c : Jumlah dokumen berkategori c_j di dalam dokumen data latih

N : Jumlah keseluruhan dokumen data latih yang digunakan

Tahapan kedua pada *Naïve Bayes* sendiri yaitu tahapan klasifikasi data uji. Adapun data uji harus berbentuk term. Dan untuk mendapatkan hasil dari klasifikasi, hanya term

di data uji yang ada pada data latih saja yang dilakukan perhitungan. Term yang tidak terdapat pada data latih diabaikan saja. Adapun rumusnya terdapat pada persamaan 2.6.

$$P(c_j|w_i) = P(c_j) \times P(w_1|c_j) \times P(w_2|c_j) \dots \\ \times P(w_n|c_j) \quad (2.6)$$

2.5 Akurasi

Akurasi merupakan besaran nilai yang menggambarkan tingkat ketepatan sebuah sistem. Hasil dari akurasi digambarkan dengan persentase. Semakin besar nilai dari akurasi maka semakin besar pula tingkat keakuratan dari suatu sistem. Perhitungan akurasi tertulis di Persamaan 2.7

$$\text{Akurasi} = \frac{(\text{jumlah data benar})}{(\text{jumlah keseluruhan data})} \times 100\% \quad (2.7)$$

3. METODOLOGI PENELITIAN

3.1 Tipe Penelitian

Non-imentatif analitik adalah strategi penelitian yang digunakan. *Non-imentatif* analitik hanya bertujuan untuk menyelesaikan permasalahan menggunakan metode yang ada tanpa harus mengimplimentasikannya menjadi sebuah aplikasi terstruktur.

3.2 Strategi Penelitian

Penelitian Kualitatif merupakan strategi yang digunakan pada penelitian kali ini. Penelitian ini berfokus pada aspek pembahasan secara mendalam yaitu klasifikasi kategori barang pada caption di *instagram* @infobarkas_jogja.

3.3 Lokasi Penelitian

Penelitian bertempat pada Fakultas Ilmu Komputer Universitas Brawijaya Malang. Dimana data yang dikumpulkan berupa data *online* yang bisa diakses melalui jaringan internet.

3.4 Teknik Pengumpulan Data

Penelitian kali ini menggunakan teknik pengumpulan data sekunder. Dimana data tersebut didapatkan dari postingan akun *instagram* @infobarkas_jogja yang saat ini sudah mencapai 50.000 data lebih.

3.5 Data Penelitian

Data pada penelitian kali ini yaitu berfokus pada caption pada foto atau video di Instagram @infobarkas_jogja. Total data yang akan diolah nantinya adalah sebanyak 500 data. Dengan total data latih sebesar 400 data serta data uji sebesar 100 data.

3.6 Implementasi Algoritma

Proses pengimplementasian algoritma dilakukan dengan memproses *caption* pada *instagram* diawali dengan tahap pre-processing. Setelah dilakukan tahap *pre-processing* maka tahapan selanjutnya yaitu seleksi fitur. Kemudian sistem akan melakukan proses klasifikasi menggunakan metode klasifikasi *Naïve Bayes*.

3.7 Teknik Analisis Data

Proses yang dilakukan yaitu menganalisa hasil berupa akurasi dari berbagai macam percobaan yang dilakukan. Percobaan dilakukan menggunakan beberapa variasi *threshold*. Hal ini bertujuan untuk mendapatkan akurasi terbaik pada penelitian kali ini.

4. PERANCANGAN

4.1 Deskripsi Sistem

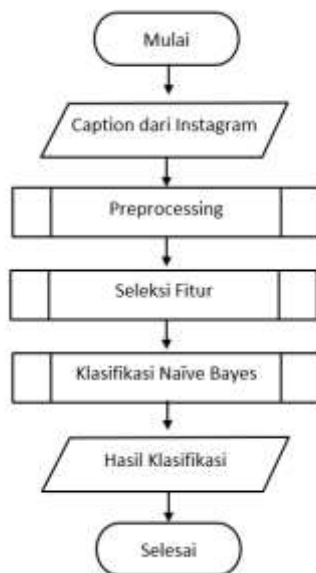
Penelitian ini akan membuat sistem yang dibangun dengan Bahasa Pemrograman *Python*. Sistem ini dibuat untuk mengklasifikasikan *caption* pada *instagram* @infobarkas_jogja dimana akun tersebut merupakan akun penitipan jual-beli barang bekas di Kota Yogyakarta. *Caption* yang diolah nantinya akan diklasifikasikan kedalam beberapa kategori kelas yaitu properti, kendaraan, pakaian, perabot, dan *gadget*.

4.2 Persiapan Data

Data yang diolah terdapat pada akun *instagram* @infobarkas_jogja. Dimana data tersebut diambil secara manual dikarenakan peneliti sendiri lah yang menjadi admin dari akun *instagram* tersebut. Setelah data didapatkan, maka dilakukan *labeling* secara manual oleh peneliti. Jumlah data pada masing-masing kategori peneliti kumpulkan dengan jumlah 100 data pada setiap kategorinya. Untuk data latih peneliti menggunakan 80% dari jumlah keseluruhan data yang berjumlah 500. Sedangkan 20% sisanya akan digunakan untuk data uji.

4.3 Perancangan Proses

Penelitian kali ini digunakan metode *Naïve Bayes* dengan optimalisasi seleksi fitur *Information Gain*. Langkah pertama adalah sistem akan menerima masukan berupa data caption *instagram* yang sudah peneliti kumpulkan. Langkah selanjutnya yaitu *pre-processing*, dalam tahapan ini data akan disederhanakan sehingga siap untuk diolah. Selanjutnya yaitu seleksi fitur yang bertujuan untuk menyeleksi data yang berpengaruh pada penelitian sesuai *threshold* yang ditentukan. Tahapan terakhir yaitu mengolah data yang sudah diseleksi dengan menggunakan metode *Naïve Bayes* menghasilkan klasifikasi berupa kategori kelas dari data tersebut. Adapun alur perancangan prosesnya terdapat di Gambar 4.1.



Gambar 4.1 Diagram Alir Sistem

5. PENGUJIAN DAN ANALISIS

5.1 Pengujian

Bab ini dijelaskan tentang beberapa pengujian yang dicoba dan analisis terhadap hasil pengujian serta pengaruh seleksi fitur pada hasil dari proses klasifikasi. Pengujian menggunakan hanya *Naïve Bayes* saja akan dibandingkan hasilnya ketika ditambahkan dengan seleksi fitur *Information Gain*. Adapun *threshold* yang digunakan mulai dari 10%, 20%, hingga 100%. Pengujian terakhir juga menambahkan metode pengujian *k-fold cross validation* untuk mengetahui akurasi rata-rata ketika *k* bernilai 5 pada *threshold* tertentu.

5.2 Hasil Pengujian

Adapun hasil dari pengujian kali ini dapat dilihat Tabel 5.1. Dimana hasil berupa jumlah data uji yang benar sesuai klasifikasi kelasnya dan juga jumlah data uji yang salah. Dilakukan sebanyak beberapa kali sesuai *threshold* yang digunakan mulai dari 10% hingga 100%.

Tabel 5.1 Hasil akurasi pengujian menggunakan seleksi fitur *Information Gain*

Jumlah term pada seleksi fitur	Jumlah Benar	Jumlah Salah	Akurasi
10%	98	2	98%
20%	96	4	96%
30%	96	4	96%
40%	98	2	98%
50%	97	3	97%
60%	96	4	96%
70%	97	3	97%
80%	98	2	98%
90%	98	2	98%
Tanpa seleksi fitur	98	2	98%
40% + <i>k-fold</i>	97,4	2,6	97,4%

5.3 Analisa Hasil

Dari seluruh pengujian yang telah dilakukan menggunakan beberapa *threshold*. Disimpulkan bahwa Seleksi Fitur *Information Gain* tidak dapat memberikan hasil yang lebih baik ketika dikombinasikan dengan *Naïve Bayes Classifier* namun dapat mengimbangi nilai akurasi ketika tidak menggunakan seleksi fitur. Hal ini terbukti dengan nilai akurasi tertinggi yaitu 98% diperoleh ketika *threshold* 10%, 40%, 80%, dan juga 90% namun nilai akurasi tersebut sama halnya diperoleh ketika tidak menggunakan seleksi fitur yaitu 98%. Hal ini terjadi karena kata yang memiliki nilai *Information Gain* yang tinggi merupakan kata yang dominan pada kelas tertentu sehingga dengan *threshold* 10% saja sudah mampu menghasilkan akurasi yang tinggi.

Percobaan ketika menggunakan Metode *Naïve Bayes* dengan Seleksi Fitur *Information Gain* tetap memberikan hasil yang bagus pada penelitian kali ini. Namun pada penelitian kali ini perpaduan keduanya belum maksimal. Walaupun tingkat akurasi terbaik sama-sama 98% namun pada masing-masing percobaan

menghasilkan dokumen eror yang berbeda. Hal ini juga dikarenakan Naïve Bayes sendiri dipengaruhi oleh faktor data set dan juga data uji. Sehingga hasil pengujian dengan nilai maksimal masih menyebar pada beberapa percobaan *threshold*.

Percobaan ketika menambahkan proses pada pengujian yaitu k-fold cross validation. Pengujian dilakukan menggunakan *threshold* sebesar 40% dan juga k-fold sebesar 5 iterasi menghasilkan akurasi rata-rata sebesar 97,4%. Hal ini menunjukkan bahwa percobaan menggunakan k-fold cross validation sudah mampu menghasilkan akurasi yang tinggi dengan pengujian yang lebih efisien karena pengujian yang dilakukan terhadap semua datasetnya.

6. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan beberapa poin yaitu sebagai berikut :

1. Tingkat akurasi dari klasifikasi jenis barang bekas menggunakan *Naïve Bayes Classifier* saja yaitu sebesar 98%. Akurasi tersebut merupakan nilai tertinggi dalam pengujian kali ini dengan 1 data yang salah di kelas properti dan 1 data yang salah di kelas perabot.

2. Tingkat akurasi dari klasifikasi jenis barang bekas menggunakan *Naïve Bayes Classifier* dan Seleksi Fitur Information Gain tertinggi sebesar 98% juga. Akurasi tersebut diperoleh ketika digunakan *threshold* 10%, 40%, 80%, dan juga 90%. Namun untuk data yang salah berbeda pada masing-masing tingkatan. Dan ketika penambahan dengan pengujian k-fold cross validation menghasilkan nilai akurasi rata-rata sebesar 97,4%.

7. DAFTAR PUSTAKA

- Amin, Daniar Nur, and Ednawati Rainarli. 2019. "Klasifikasi Konten Instagram Berdasarkan Komentar Menggunakan Support Vector Machine".
- Bisma, M.Ardhya. 2018. "Analisis Formulasi Strategi Untuk Social Media Influencer di Instagram dengan Pendekatan TOWS Matrix (Studi Kasus: @Dagelan)".
- Chormunge, S., Jena, S., 2016. "Efficient Feature Subset Selection Algorithm for High Dimensional Data".
- Ervianah, Ermya. 2021. "Strategi Marketing Melalui Media Sosial Instagram (Analisis Digital Social Media Marketing)".
- Haddi, Emma dkk. 2013. "The Role of Text Pre-processing in Sentiment Analysis".
- Hamzah, Amir. 2012. "Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita Dan Abstract Akademis."
- Mileva, Lubiana, and Achmad Fauzi. 2018. "Pengaruh Social Media Marketing Terhadap Keputusan Pembelian (Survei Online Pada Mahasiswa Sarjana Jurusan Ilmu Administrasi Bisnis Angkatan 2014/2015 Fakultas Ilmu Administrasi Universitas Brawijaya Yang Membeli Starbucks Menggunakan Line)." Jurnal Administrasi Bisnis 58.
- Putra, I., Sudarma, M., Kumara, I., 2016, "Klasifikasi Teks Bahasa Bali Dengan Metode Supervised Learning Naive Bayes Classifier". Teknologi Elektro, 14, pp. 81-86
- Rahadi, Dedi Rianto, and Zaniel. 2020. "Social Media Marketing dalam Mewujudkan E-Marketing".
- Rahman, Arif, and Redi Panuju. 2017. "Strategi Komunikasi Pemasaran Produk Fair N Pink Melalui Media Sosial Instagram".
- Statista Research Department. 2022. "Countries with the highest number of internet users 2022". Diakses pada 31 Oktober 2022, dari <https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/>.
- Syakuro, Abdan. 2017. "Analisis Sentimen Masyarakat Terhadap E-Commerce pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) dengan Seleksi Fitur Information Gain (IG)".
- Tsaniputra, Hanif Murtadha. 2021. " Analisis Sentimen Media Sosial Sebagai Rekomendasi Kuliner di Yogyakarta Menggunakan Metode Naive Bayes dengan Seleksi Fitur Information Gain".
- Utama, Heru Sukma dkk. 2019. "Sentimen Analisis Kebijakan Ganjil Genap Di Tol Bekasi Menggunakan Algoritma Naive Bayes dengan Optimalisasi Information

Gain.”

Wellyantama, Prada, and Iwan Krisnadi. 2019.
“Pemanfaatan E-Commerce Untuk Jual-
Beli Barang Bekas Pakai.”