# PENGELOMPOKAN DATA KEPUASAN KONSUMEN PADA APLIKASI TRANSPORTASI ONLINE MENGGUNAKAN ALGORITMA DECISION TREE

Pramudita Rahayu<sup>1\*)</sup>, Ryan Puji Cahyono<sup>2\*)</sup>
<sup>1</sup>Sistem Informasi Akutansi, <sup>2</sup>Teknik Komputer
\*) ryanpujic@gmail.com

#### **Abstrak**

Perkembangan teknologi yang serba digital ini membutuhkan mobilitas yang tinggi dari masyarakat. Banyak perusahaan sekarang menawarkan layanan transportasi online. Transportasi online tidak hanya mengurangi kemacetan lalu lintas, tetapi juga memberikan solusi alternatif dalam kemacetan kendaraan, memungkinkan orang untuk mencapai tempat-tempat yang sulit dijangkau dengan transportasi tradisional. Penelitian ini hanya berfokus pada analisis kepuasan pengguna atau pelanggan Gojek, Grab dan Maxim. Penyedia jasa transportasi online harus tetap fokus pada kenyamanan, keamanan dan kepuasan pelanggan. Tujuan dari penelitian ini adalah untuk mengetahui kepuasan konsumen terhadap transportasi online menggunakan pohon keputusan dengan algoritma C4.5.

Kata Kunci: Teknologi, Online, Klasifikasi, Decision Tree, Algoritma C4.5

# **PENDAHULUAN**

Perkembangan teknologi yang serba digital ini membutuhkan mobilitas yang tinggi dari masyarakat(Saputra & Pasha, 2021). Banyak perusahaan sekarang menawarkan layanan transportasi online. Transportasi online tidak hanya mengurangi kemacetan lalu lintas, tetapi juga memberikan solusi alternatif di antara kemacetan kendaraan, memungkinkan orang untuk mencapai tempat-tempat yang sulit dijangkau oleh transportasi tradisional (Yulianti et al., 2021).

Alat komunikasi yang berkembang pesat saat ini adalah semua dalam satu tangan. Sekarang mudah untuk memesan transportasi mana yang akan digunakan, hanya dengan menggunakan smartphone Anda. Transportasi online siap mengantarkan ke tujuan Anda tanpa menegosiasikan harga. Akibatnya, banyak orang beralih ke transportasi online, yang lebih mudah daripada transportasi tradisional(Mutmainnah, 2020),(Saputra & Pasha, 2021). Perusahaan yang menyediakan jasa transportasi online di Indonesia seperti gojek, grab dan maxim. Penyedia layanan transportasi online tidak hanya perusahaan transportasi online yang berfokus pada taksi, tetapi juga mencakup jenis layanan transportasi lainnya seperti ojek dan mobil pada umumnya(Lazuardi & Sukoco, 2019),(Gerai et al., 2021).

Penelitian ini hanya berfokus menganalisa kepuasan pengguna atau pelanggan gojek, grab, dan maxim. Kepuasan konsumen merupakan situasi yang ditunjukkan oleh konsumen ketika mereka menyadari bahwa kebutuhan dan keinginannya sesuai dengan yang diharapkan serta terpenuhi secara baik(Mutmainnah, 2020). Konsumen yang puas adalah konsumen yang akan berbagi kepuasan dengan produsen. Bahkan, konsumen yang puas akan berbagi rasa dan sipengalaman dengan konsumen lain. Oleh karena itu, baik konsumen maupun produsen akan isama-sama diuntungkan apabila kepuasan terjadi.

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas atau konsep dalam data dengan tujuan menggunakan model klasifikasi untuk membuat prediksi tentang kelas objek yang kelas klasifikasinya tidak diketahui. Salah satu teknik klasifikasi yang paling populer adalah pohon keputusan. Penyedia jasa transportasi online harus tetap fokus pada kenyamanan, keamanan dan kepuasan pelanggan(Balirante, Meylinda et al., 2020). Jika konsumen puas dengan pelayanan dari penyedia jasa transportasi online, maka pelanggan tersebut akan menggunakan layanan tersebut berulang kali (repeat purchase) dan menjadi pelanggan yang loyal (loyal customer) yang lebih maju dan dipercaya oleh pelanggan Indonesia(Febriza & Adrian, 2021),(Primadewi, 2021). Tujuan dari penelitian ini adalah untuk mengetahui kepuasan konsumen terhadap transportasi online menggunakan pohon keputusan dengan algoritma C4.5.

## KAJIAN PUSTAKA

#### **Definisi Decision Tree**

Pohon keputusan adalah metode klasifikasi yang menggunakan struktur pohon di mana setiap simpul mewakili atribut, cabangnya mewakili nilai atribut, dan daun digunakan untuk mewakili kelas. Node teratas dari pohon keputusan ini disebut root(Aldino & Sulistiani, 2020). Kami telah menunjukkan bahwa metode ini sangat populer karena hasil dari model yang terbentuk mudah dipahami. Disebut pohon keputusan karena aturan-aturan yang dibentuknya menyerupai pohon.

Sebuah pohon dibentuk pada grup data dari teknik pengurutan rekursif biner sedemikian rupa sehingga nilai-nilai variabel respons di setiap grup data membuat hasil pengurutan lebih seragam. Konsep pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan keputusan(Fikri et al., 2017). Keuntungan utama menggunakan pohon keputusan adalah mereka menyederhanakan proses pengambilan keputusan yang kompleks

dan membantu pengambil keputusan menginterpretasikan solusi untuk masalah(Nurkholis, n.d.),(Nurkholis & Sitanggang, 2020).

# Pengertian Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk mengklasifikasikan atau mengelompokkan kumpulan data(Nurkholis & Sitanggang, 2020). Dasar dari algoritma C4.5 adalah pembentukan pohon keputusan. Cabang pohon keputusan adalah masalah klasifikasi dan daun adalah kelas atau kelompok. Karena tujuan dari algoritma C4.5 adalah klasifikasi, hasil dari pemrosesan record adalah pengelompokan data ke dalam kelas-kelas tertentu(Edhy Sunanta, 2013),(Wulantina & Maskar, 2019). Algoritma C4.5 merupakan pengembangan lebih lanjut dari algoritma ID3, dan kekurangan dari algoritma ID3 dapat ditutupi oleh algoritma C4.5.

Hal pertama yang harus dilakukan dalam membuat pohon keputusan adalah menentukan atribut/variabel mana yang merupakan akar dari pohon keputusan. Gunakan entropi, gain, informasi split, dan rasio gain untuk menentukan variabel mana yang di-root(Walsh et al., 2019).

Entropi adalah parameter yang mengukur derajat keragaman (heterogenitas) dari suatu dataset(Herlinda et al., 2021). Semakin tinggi nilai entropi, semakin besar keragaman kumpulan data. Rumus untuk menghitung entropi adalah:

Entropy (S) = 
$$-\sum_{i=1}^{m} p_i \log_2(p_i)$$
....(1)

Dimana:

S: himpunan kasus

m: jumlah kelas dari klasifikasi

pi : jumlah proporsi sampel (peluang) untuk kelas i

Gain adalah ukuran keefektifan suatu variabel dalam mengklasifikasikan data. Keuntungan suatu variabel adalah perbedaan antara entropi total dan entropi variabel tersebut. Keuntungan dapat dirumuskan sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|s_i|}{|s|} \times Entropy(S_i)....(2)$$

#### Dimana:

A: Variabel

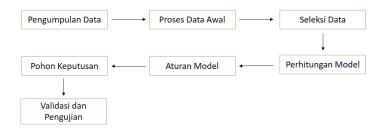
Si: Jumlah sampel untuk i

S: Jumlah sampel untuk seluruh data

Algoritma C4.5 menggunakan nilai gain untuk menentukan variabel yang merupakan node dari pohon keputusan. Variabel dengan gain tertinggi adalah node dari pohon keputusan.

## **METODE**

Metode yang dipakai dalam penerapan Algoritma C4.5, menggunakan desain penelitian yang ditujukan pada Gambar 1 berikut ini :



Gambar 1 Diagram alir penelitian

## Pengumpulan Data

Pengumpulan data adalah pengumpulan data yang digunakan dalam proses algoritma klasifikasi C4.5.

#### **Proses Data Awal**

Proses mentransformasikan, menggabungkan, atau mengonversi data ke dalam bentuk yang sesuai, untuk diproses dengan perhitungan algoritma C4.5(Handayani, 2014).

#### Seleksi Data

Fungsi pilih digunakan untuk memilih data yang digunakan dalam proses algoritma klasifikasi C4.5. Tujuan pemilihan data adalah untuk membuat kumpulan data target, memilih kumpulan data, atau fokus pada subset variabel atau sampel data tempat penemuan dibuat(Mittal & Patidar, 2019),(Saifuddin Dahlan, 2013).

# Perhitungan Model

Hitung semua atribut/variabel, entropi menggunakan persamaan dan kumpulkan informasi menggunakan persamaan untuk menemukan perolehan informasi tertinggi untuk digunakan sebagai simpul akar dalam membangun pohon keputusan(Pratama et al., 2022),(Zakaria et al., 2013).

#### **Aturan Model**

Aturan model atau role model adalah deskripsi penjelasan yang merepresentasikan pohon keputusan.

#### **Pohon Keputusan Decision Tree**

Pohon keputusan merupakan hasil dari proses komputasi entropi dan temu kembali informasi setelah dilakukan pengulangan komputasi hingga semua atribut pohon memiliki kelas dan proses komputasi tidak dapat dilakukan(Nurkholis, n.d.),(Septilia et al., 2020),(Alita et al., 2021).

## Validasi dan Pengujian

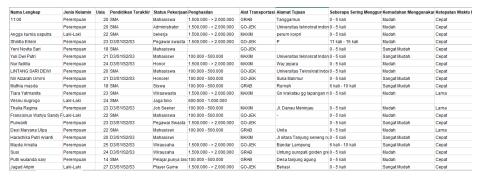
Validasi dan pengujian adalah pengujian yang dilakukan untuk memastikan bahwa semua fungsionalitas berfungsi dengan benar. Validasi dilakukan dengan menggunakan matriks konfusi dan validasi silang 10 kali lipat dilakukan dengan mengacak data dan membagi dataset menjadi 10 segmen yang sama(Amanda, 2017),(Adi et al., 2020). Validasi dan pengujian dilakukan untuk mengetahui akurasi, presisi, dan reproduktifitas hasil prediksi klasifikasi. Akurasi adalah persentase catatan yang diklasifikasikan dengan benar dalam kumpulan data uji(Nasution & Hayaty, 2019). Akurasi adalah persentase data milik model yang baik yang sebenarnya baik. Recall adalah ukuran tingkat pengenalan positif sejati(Oktaviani & Mandasari, 2020).

## HASIL DAN PEMBAHASAN

Hasil dan tahapan proses pembahasan pada algoritma klasifikasi C4.5 adalah sebagai berikut:

#### Pengumpulan Data

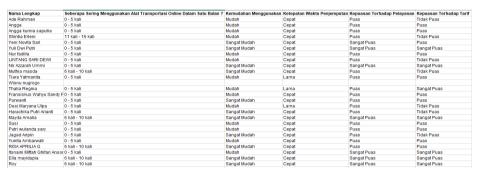
Pengumpulan data didapat dengan melakukan beberapa pertanyaan kepada pengguna puas atau setidaknya dalam menggunakan transportasi online.



Gambar 2 Pengumpulan Data

# Proses Data Awal (Preprocessing data)

Preprocessing data akan dilakukan dengan cara menghilangkan redutant nama, merubah format pada kolom.



Gambar 3 Preprocessing Data

#### Seleksi Data

Variabel yang dipilih untuk menghitung model pohon keputusan adalah kemudahan penggunaan aplikasi, waktu pengambilan, kepuasan terhadap pelayanan, dan kepuasan terhadap harga.

No	Ketepatan Waktu Penjemputan	Kemudahan Menggunakan Aplikasi	Kepuasan Terhadap Pelayanan	Kepuasan Terhadap Tarif	
1	Cepat	Mudah	Puas	Tidak Puas	
2	Cepat	Mudah	Puas	Puas	
3	Cepat	Mudah	Puas	Puas	
4	Cepat	Mudah	Puas	Tidak Puas	
5	Cepat	Sangat Mudah	Sangat Puas	Puas	
6	Cepat	Sangat Mudah	Sangat Puas	Sangat Puas	
7	Cepat	Mudah	Puas	Puas	
8	Cepat	Mudah	Puas	Tidak Puas	
9	Cepat	Sangat Mudah	Sangat Puas	Sangat Puas	
10	Cepat	Sangat Mudah	Puas	Tidak Puas	
11	Lama	Mudah	Puas	Puas	
12	Cepat	Mudah	Puas	Puas	
13	Lama	Mudah	Puas	Sangat Puas	
14	Cepat	Mudah	Puas	Puas	
15	Cepat	Sangat Mudah	Puas	Puas	
16	Lama	Mudah	Puas	Tidak Puas	
17	Cepat	Sangat Mudah	Puas	Tidak Puas	
18	Cepat	Sangat Mudah	Sangat Puas	Sangat Puas	
19	Cepat	Mudah	Puas	Puas	
20	Cepat	Mudah	Puas	Puas	

Gambar 4 Seleksi Data

# **Perhitungan Model**

Anda dapat menggunakan rumus (1) untuk menghitung nilai entropi. Entropi total dihitung dengan menghitung keputusan "sangat puas" dan "puas" dari semua kasus yang ada.

Entropi Total menghitung jumlah keputusan yang sangat memuaskan (4) dan keputusan yang memuaskan (16). dimana 20 adalah jumlah kasus. Kemudian hitung setiap entropi dari semua nilai variabel yang ada. Kemudian Anda dapat menggunakan rumus (2) untuk menghitung nilai temu kembali informasi untuk setiap atribut. Setelah semua nilai entropi dan information gain dihitung, hasil perhitungan tersebut dimasukkan pada Tabel 3. Dari perhitungan nilai entropi dan information retrieval rate di atas, kita dapat melihat bahwa information retrieval rate maksimum adalah recency dengan nilai 0,71081 dan minimum usability dengan nilai 0,3448. Kemudian ambil nilai perolehan informasi maksimum dan jadikan simpul akar. Hapus atribut yang dipilih sebelumnya, ulangi nilai entropi, perhitungan perolehan informasi, pilih perolehan informasi maksimum dan gunakan sebagai simpul bagian dalam pohon. Ulangi perhitungan sampai semua atribut pohon memiliki kelas.

Nod e	Atribut	Nilai	Sum (Nilai )	Sum (Sangat Puas)	Sum (Puas)	Entropi	Gain
1	Ketepatan	Cepat	14	4	10	0,8631	
l	Waktu	Lama	3	0	3	0	
							0,71081
	Kemudaha n Aplikasi	Sangat Mudah	7	4	3	0,98523	
		Mudah	13	0	13	0	
							0,3448
	Tarif	Sangat Puas	3	2	1	0,9182	
		Puas	10	1	9	0,4689	
		Tidak Puas	6	0	6	0	
							0,39183

Gambar 4 Hasil Perhitungan

#### **Decision Tree**

Dari hasil perhitungan entropi dan temu kembali informasi selanjutnya dapat diolah menjadi pohon keputusan. Hasil pohon keputusannya adalah:

# Gambar 5 Pohon Keputusan

# Kesimpulan

Kesimpulan dari penelitian ini adalah:

- 1. Tepat Waktu = Cepat, Pelayanan = Sangat Memuaskan
- 2. Jika ketepatan waktu = lama, kemudahan penggunaan = sangat mudah, pelayanan = puas
- 3. Kekinian = Durasi, Aplikasi Mudah = Sederhana, Harga = Sangat Puas, Pelayanan = Puas

## REFERENSI

- Adi, R. P., Koswara, Y., Tashika, J., Devi, Y., & Saifudin, A. (2020). Pengujian Black Box pada Aplikasi Pertokoan Minimarket Menggunakan Metode Equivalence Partitioning. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, *3*(2), 100. https://doi.org/10.32493/jtsi.v3i2.4695
- Aldino, A. A., & Sulistiani, H. (2020). Decision Tree C4. 5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia). *Edutic-Scientific Journal of Informatics Education*, 7(1).
- Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 17–23.
- Amanda, D. (2017). PENGUJIAN KEPUASAN SEBAGAI VARIABEL INTERVENING ANTARA PENGARUH KEPERCAYAAN DAN ATRIBUT PRODUK TABUNGAN BATARA IB TERHADAP LOYALITAS NASABAH (STUDI PADA PT. BANK TABUNGAN NEGARA (PERSERO) TBK, KANTOR CABANG SYARIAH PALEMBANG). [SKRIPSI]. UIN RADEN FATAH PALEMBANG.
- Balirante, Meylinda, Lefrandt, L., & Kumaat, M. (2020). Analisa Tingkat Kebisingan Lalu Lintas di Jalan Raya Ditinjau Dari Tingkat Baku Mutu Kebisingan Yang Diizinkan. *Jurnal Sipil Statik*, 8(2), 249–256.
- Edhy Sunanta. (2013). Model Integrasi Database Penduduk Indonesia dengan Berbagai Sistem Informasi Berbasis Komputer. Akprind.
- Febriza, M. A., & Adrian, Q. J. (2021). PENERAPAN AR DALAM MEDIA PEMBELAJARAN KLASIFIKASI BAKTERI. *Jurnal BIOEDUIN: Program Studi Pendidikan Biologi*, 11(1), 10–18.
- Fikri, C. M., Agustin, F. E. M., & Mintarsih, F. (2017). Pengelompokan kualitas kerja pegawai menggunakan algoritma K-Means++ dan Cop-Kmeans untuk merencanakan program pemeliharaan kesehatan pegawai di pt. Pln p2b jb depok. *Pseudocode*, *4*(1), 9–17.
- Gerai, S., Donald, M., Indriani, R., & Firmansyah, M. A. (2021). STRATEGI KOMUNIKASI PEMASARAN MELALUI BTS MEAL OLEH RESTORAN MC. DONALDS DAN PERSEPSI KONSUMEN Abstrak. 3(1), 3–12.
- Handayani, M. A. (2014). INOVASI PRODUK SEBAGAI ALTERNATIF KONVERSI AKAD PEMBIAYAAN MURABAHAH DENGAN SISTEM MUSYARAKAH (Studi Kasus Pada Bank Sumsel Babel Syariah Cabang Palembang). *Fordema*, 11(2), 35–47.
- Herlinda, V., Darwis, D., & Dartono, D. (2021). ANALISIS CLUSTERING UNTUK RECREDESIALING FASILITAS KESEHATAN MENGGUNAKAN METODE FUZZY C-MEANS. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 94–99.

- Lazuardi, M. L., & Sukoco, I. (2019). Design Thinking David Kelley & Tim Brown: Otak Dibalik Penciptaan Aplikasi Gojek. *Organum: Jurnal Saintifik Manajemen Dan Akuntansi*, 2(1), 1–11. https://doi.org/10.35138/organum.v2i1.51
- Mittal, A., & Patidar, S. (2019). Sentiment analysis on twitter data: A survey. *ACM International Conference Proceeding Series*, 91–95. https://doi.org/10.1145/3348445.3348466
- Mutmainnah, S. (2020). Pemilihan Moda Transportasi Kereta Api Menuju Pelabuhan Bakauheni. *JICE* (*Journal of Infrastructural in Civil Engineering*), *1*(01), 33. https://doi.org/10.33365/jice.v1i01.854
- Mutmainnah, S. (2020). Pemilihan Moda Transportasi Kereta Api Menuju Pelabuhan Bakauheni. *JICE (Journal of Infrastructural in Civil Engineering)*, *1*(01), 33. https://doi.org/10.33365/jice.v1i01.854
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235. https://doi.org/10.31311/ji.v6i2.5129
- Nurkholis, A. (n.d.). *Model Pohon Keputusan Spasial untuk Evaluasi Kesesuaian Lahan Bawang Putih*. Bogor Agricultural University (IPB).
- Nurkholis, A., & Sitanggang, I. S. (2020). Optimalisasi model prediksi kesesuaian lahan kelapa sawit menggunakan algoritme pohon keputusan spasial. *Jurnal Teknologi Dan Sistem Komputer*, 8(3), 192–200.
- Oktaviani, L., & Mandasari, B. (2020). Powtoon: A digital medium to optimize students' cultural presentation in ELT classroom. *Teknosastik*, 18(1), 33–41.
- Pratama, E. N., Suwarni, E., & Handayani, M. A. (2022). The Effect Of Job Satisfaction And Organizational Commitment On Turnover Intention With Person Organization Fit As Moderator Variable. *Atm*, *6*(1), 74–82.
- Primadewi, A. (2021). *Model Machine Learning untuk Klasifikasi Mutu Telur Ayam Ras Berdasarkan Kebersihan Kerabang*. 8(6), 386–391. https://doi.org/10.30865/jurikom.v8i6.3574
- Saifuddin Dahlan, F. H. (2013). THE INFLUENCES OF PERSONALITY AND COGNITIVE PERCEPTION TOWARDS THE STUDENTS'INTENTION TO USE DATABASE SOFTWARE AT THE COMPUTERIZED ACCOUNTING VOCATIONAL COLLEGES IN LAMPUNG PROVINCE. Universitas Lampung.
- Saputra, V. H., & Pasha, D. (2021). Komik Digital Berbasis Scientific Method Sebagai Media Pembelajaran di Masa Pandemi COVID-19. *Jurnal Riset Teknologi Dan Inovasi Pendidikan (Jartika)*, 4(1), 89–100.
- Septilia, H. A., Parjito, P., & Styawati, S. (2020). Sistem Pendukung Keputusan Pemberian Dana Bantuan menggunakan Metode AHP. *Jurnal Teknologi Dan Sistem Informasi*, 1(2), 34–41.

- Walsh, S., Jong, E. E. C. de, Timmeren, J. E. van, & ... (2019). Decision support systems in oncology. *JCO Clinical Cancer* .... https://doi.org/10.1200/CCI.18.00001
- Wulantina, E., & Maskar, S. (2019). PEMBELAJARAN MATEMATIKA BERBASIS LAMPUNGNESE ETNOMATEMATICS PADA MATERI BANGUN DATAR. *Prosiding Seminar Nasional Pendidikan FKIP Universitas Lampung Tahun*, 793.
- Yulianti, D. T., Damayanti, D., & Prastowo, A. T. (2021). PENGEMBANGAN DIGITALISASI PERAWATAN KESEHATAN PADA KLINK PRATAMA SUMBER MITRA BANDAR LAMPUNG. Jurnal Teknologi Dan Sistem Informasi, 2(2), 32–39.
- Zakaria, R., Howlett, P. G., Piantadosi, J., Boland, J. W., & Moslim, N. H. (2013). Modelling catchment rainfall using sum of correlated gamma variables. *Jurnal Teknologi* (*Sciences and Engineering*), 63(2), 85–88. https://doi.org/10.11113/jt.v63.1918
- Gerai, S., Donald, M., Indriani, R., & Firmansyah, M. A. (2021). STRATEGI KOMUNIKASI PEMASARAN MELALUI BTS MEAL OLEH RESTORAN MC. DONALDS DAN PERSEPSI KONSUMEN Abstrak. 3(1), 3–12.
- Pratama, E. N., Suwarni, E., & Handayani, M. A. (2022). The Effect Of Job Satisfaction And Organizational Commitment On Turnover Intention With Person Organization Fit As Moderator Variable. *Atm*, *6*(1), 74–82.
- Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 17–23.
- Febriza, M. A., & Adrian, Q. J. (2021). PENERAPAN AR DALAM MEDIA PEMBELAJARAN KLASIFIKASI BAKTERI. *Jurnal BIOEDUIN: Program Studi Pendidikan Biologi*, 11(1), 10–18.