

ANALISA DATA MINING ASURANSI KENDARAAN MOBIL MENGGUNAKAN APLIKASI WEKA

¹Nurul Fajriyah, ²Wawan Setiawan

¹Dosen Tetap, Universitas Insan Pembangunan Indonesia

²Dosen Tetap, Universitas Muhammadiyah Tangerang

Email: ¹nurulfajriyah442@gmail.com, ²whawan.s@gmail.com

ABSTRAK

Perkembangan teknologi informasi saat ini sangat pesat, terutama didalam bidang jasa salah satunya yaitu jasa asuransi kendaraan mobil. Salah satu aset utama perusahaan dalam asuransi kendaraan adalah data pelanggan dan riwayat polisnya yang tersedia dalam jumlah besar. Untuk melakukan penguatan strategi bisnis, maka dilakukan analisa data mining, mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan. Selain itu analisa data mining berguna untuk menggali informasi dari basis data pelanggan dan transaksi yang dilakukan pelanggan meliputi pembayaran premi asuransi dan maupun proses klaim yang terjadi sehingga dapat memprediksi seberapa besar kemungkinan pelanggan tersebut akan melakukan *renewal* atas polis asuransi mereka. Pada penelitian ini, akan dilakukan analisis klasifikasi untuk menentukan keputusan pembaharuan (*renewal*) polis pelanggan dengan menggunakan data polis asuransi kendaraan roda empat. Penulis menggunakan komparasi algoritma *Naive Bayes*, C4.5 dan KNN (*K-Nearest Neighbor*). Berdasarkan hasil komparasi use training set, maka diperoleh metode yang terbaik yaitu metode KNN (*K-Nearest Neighbor*), dengan tingkat keberhasilan sebesar 87,1429% dan tingkat error sebesar 12,8571%. Untuk Komparasi *cross validation 10 folds* diperoleh Metode yang terbaik yaitu metode C4.5 dengan tingkat keberhasilan sebesar 82,1429% dan tingkat error sebesar 17,8571%.

Kata kunci : Data Mining, Asuransi, *Naive Bayes*, C4.5, *K-Nearest Neighbor*

PENDAHULUAN

Di era informasi perkembangan teknologi informasi saat ini sangat pesat, terutama didalam bidang jasa salah satunya yaitu jasa asuransi kendaraan (mobil). Salah satu aset utama perusahaan adalah data pelanggan dan riwayat polisnya yang tersedia dalam jumlah besar. Ketersediaan data tersebut menuntut akan adanya teknologi yang dapat memanfaatkan data tersebut untuk melakukan penguatan strategi bisnis.

Prediksi terhadap minat pelanggan sangat penting bagi perusahaan asuransi, karena terhadap minat pelanggan untuk memperbaharui (*renewal*) terhadap polis asuransi mereka akan memberikan kontribusi bagi perusahaan untuk keberlangsungan usaha perusahaan asuransi. Data mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar, melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik,

pembelajaran mesin dan sistem manajemen basis data. Data mining bukanlah suatu bidang analisa yang baru, akan tetapi data mining memiliki akar yang panjang dari bidang ilmu yang berbeda seperti kecerdasan buatan (*artificial intelligent*), machine learning, statistik, database, dan juga information retrieval (Utari et al., 2020). Analisa data mining jasa asuransi kendaraan mobil memiliki peran penting dalam penguatan strategi bisnis, mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan pihak manajemen. Selain itu analisa data mining berguna untuk menggali informasi dari basis data pelanggan dan transaksi yang dilakukan pelanggan meliputi pembayaran premi asuransi dan maupun proses klaim yang terjadi sehingga dapat memprediksi seberapa besar kemungkinan pelanggan tersebut akan melakukan *renewal* atas polis asuransi mereka (Sarwiyah et al., 2020).

Berdasarkan latar belakang tersebut, maka penulis tertarik untuk melakukan penelitian analisa data mining mengenai jasa asuransi kendaraan (mobil). Pada penelitian ini, akan dilakukan analisis klasifikasi untuk menentukan keputusan pembaharuan (*renewal*) polis pelanggan dengan menggunakan data polis asuransi kendaraan roda empat (Novianto, 2019). Penulis menggunakan komparasi algoritma *Naive Bayes*, C4.5 dan KNN (*K-Nearest Neighbor*). Penulis mencoba mengkomparasi data set asuransi kendaraan dengan tiga metode yaitu *Naive Bayes*, C4.5 dan KNN (*K-Nearest Neighbor*) (Karlia et al., 2021). Software yang digunakan dalam penelitian ini adalah Weka (Ariadi, 2020). Tujuan dari penggunaan software ini adalah membandingkan hasil dengan perhitungan secara teoritis dengan hasil yang didapatkan dengan proses di Weka Interface ini.

Adapun tujuan dari penelitian yang dilakukan, yaitu :

1. Melakukan analisa data mining, mengeksplorasi basis data asuransi kendaraan mobil yang sangat berguna untuk mendukung pengambilan keputusan.
2. memperoleh metode yang terbaik dalam mengeksplorasi basis data asuransi kendaraan mobil.

METODE PENELITIAN

DATA UNDERSTANDING

Penulis melakukan proses fase pemahaman data atau data understanding (Bustami, 2014) yaitu untuk proses pengkategorian *renewal* polis, terdiri dari 11 atribut. Dimana 10 atribut prediktor dan 1 atribut hasil atau target. Atribut-atribut yang menjadi parameter terlihat pada tabel 1, yaitu:

Tabel 1. Data Atribut Polis

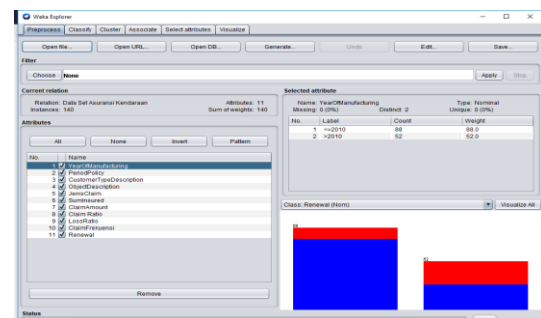
No	Atribut	Nilai
1	Year of Manufacturing	≤ 2010 , >2010
2	Policy Periode	1,2,3,4 dan 5 Tahun
3	Costumer Type Description	Personal dan Company
4	Object Description	Truck,Sedan,Nulti Purpose Vehicle, Pick up dan Minibus.
5	Jenis Claim	<i>Partial Loss Accident</i> , <i>Partial Loss Stolen</i> ,

		dan <i>No Claim</i>
6	Sum Insured	≤ 300 juta dan >300 juta – 1 M
7	Claim Amount	Besar,Kecil, dan <i>No Claim</i> .
8	Loss Ratio	Tinggi dan Rendah
9	Claim Ratio	Tinggi dan Rendah
10	Claim Frekuensi	1x, $>1x$ dan <i>No Claim</i>
11	Renewal	Ya dan Tidak

Sebelum dilakukan pemodelan, diperlukan beberapa teknik pre-processing yang digunakan, yaitu: data validation, data integration and transformation, data size reduction and dicretization (Kadir et al., 2019).

DATA PREPROCESSING

Pada tahap ini penulis melakukan preprocessing data, dengan mengguakan Aplikasi Weka, data yang digunakan untuk analisa adalah data set asuransi kendaraan (Diantika et al., 2021). Berikut ini hasil dari preprocessing data dengan menggunakan Aplikasi Weka



Gambar 1. Preprocessing Data

Dari hasil preprocessing, data yang dihasilkan adalah untuk kendaraan dibawah tahun 2010 sebanyak 88 unit kendaraan yang terasuransi, sementara untuk kendaraan diatas 2010 yaitu sebanyak 52 unit yang sudah terasuransi. Kemudian penulis melakukan Klasifikasi dengan menggunakan Metode *Naive Bayes*, C4.5 dan KNN (Utami et al., 2020).

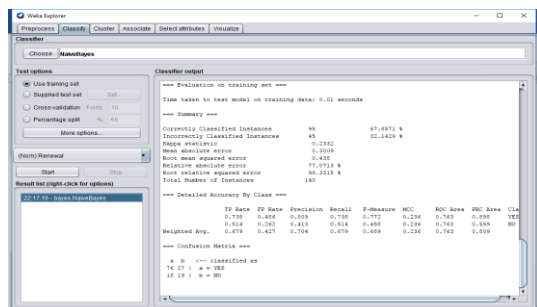
HASIL DAN PEMBAHASAN

Naive Bayes

a. Training Set *Naive Bayes*

Penulis mencoba menggunakan training set pada metode *Naive Bayes*, maka hasilnya

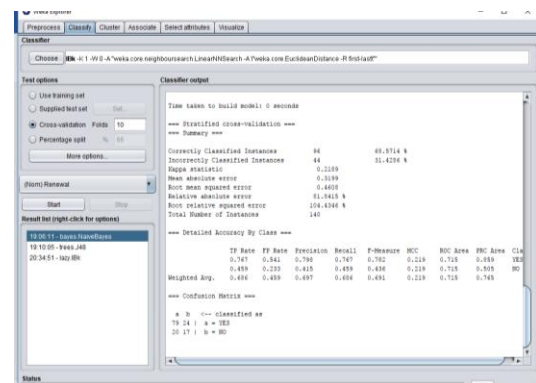
seperti gambar 2. diperoleh hasil keberhasilannya 67,8571 % dengan error 32,1429 %. pengujian tingkat akurasi dengan menggunakan confusion matrix. Baris pertama “74” dan “27” menunjukkan bahwa dari total 140 data, ada 74 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 27 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “18” dan “19” menunjukkan bahwa dari total 140 data, ada 18 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 19 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).



Gambar 2. Set Training Naïve Bayes

b. Cross Validation Naïve Bayes

Penulis mencoba menggunakan cross validation 10 folds pada metode Naive Bayes, maka hasilnya seperti gambar 3. diperoleh hasil keberhasilannya 68,5714 % dengan error 31,4286 %. pengujian tingkat akurasi dengan menggunakan confusion matrix. Baris pertama “79” dan “24” menunjukkan bahwa dari total 140 data, ada 79 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 24 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “20” dan “17” menunjukkan bahwa dari total 140 data, ada 20 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 17 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).

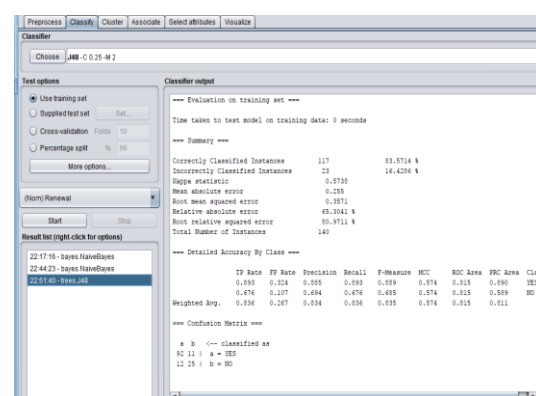


Gambar 3. Cross Validation Naïve Bayes

Metode C4.5

a. Training Set Metode C4.5

Penulis mencoba menggunakan training set pada metode C4.5, maka hasilnya seperti gambar 4. diperoleh hasil keberhasilannya 83,5714 % dengan error 16,4286 %. pengujian tingkat akurasi dengan menggunakan confusion matrix. Baris pertama “92” dan “11” menunjukkan bahwa dari total 140 data, ada 92 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 11 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “12” dan “25” menunjukkan bahwa dari total 140 data, ada 12 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 25 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).

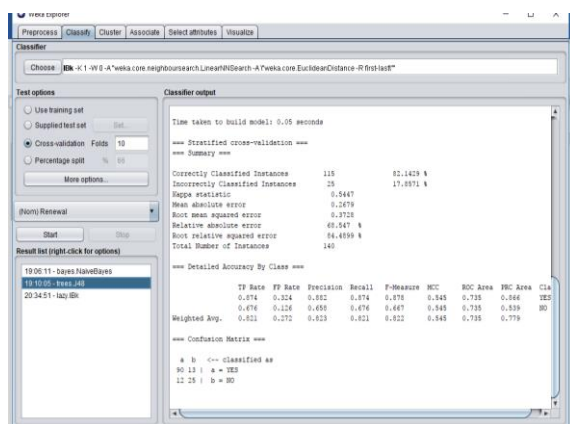


Gambar 4. Training Set Metode C4.5

b. Cross Validation Metode C4.5

Penulis juga mencoba menggunakan decision tree C4.5, maka hasilnya seperti gambar 5. diperoleh hasil keberhasilannya

82,1429 % dengan error 17,8571 %. pengujian tingkat akurasi dengan menggunakan confusion matrix. Baris pertama “90” dan “13” menunjukkan bahwa dari total 140 data, ada 90 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 13 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “12” dan “25” menunjukkan bahwa dari total 140 data, ada 12 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 25 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).

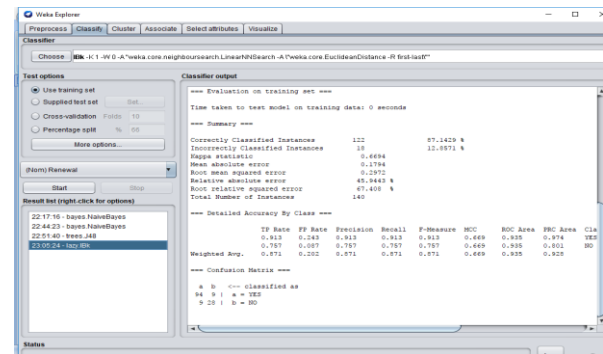


Gambar 5. Cross Validation C4.5

KNN (K-Nearest Neighbor)

a. Training Set Metode KNN

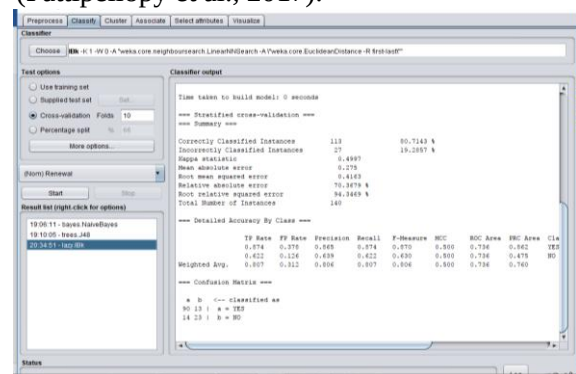
Penulis mencoba menggunakan training set pada metode Metode KNN (Situmorang et al., 2019), maka hasilnya seperti gambar 6. diperoleh hasil keberhasilannya 87,1429 % dengan error 12,8571 %. Pengujian tingkat akurasi dengan menggunakan confusion matrix. Baris pertama “94” dan “9” menunjukkan bahwa dari total 140 data, ada 94 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 9 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “9” dan “28” menunjukkan bahwa dari total 140 data, ada 9 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 28 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).



Gambar 6. Set Training KNN

b. Cross Validation Metode KNN

Penulis mencoba menggunakan KNN (Ulumi et al., 2019), maka hasilnya seperti gambar 7. diperoleh hasil keberhasilannya 80,7143 % dengan error 19,2857 %. Pengujian tingkat akurasi dengan menggunakan confusion matrix (Andarista et al., 2022). Baris pertama “90” dan “13” menunjukkan bahwa dari total 140 data, ada 90 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 13 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA). Baris kedua “14” dan “23” menunjukkan bahwa dari total 140 data, ada 14 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 23 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO) (Pattipeilopy et al., 2017).



Gambar 7. Cross Validation KNN

HASIL KOMPARASI

Dari klasifikasi yang dilakukan mulai dari dengan menggunakan Metode Naïve Bayes, KNN dan Decision Tree C4.5 (Kurniawan et al, 2007), maka hasil dari komparasi yaitu :

a. Komparasi Use Training Set

Tabel 2. Komparasi Use Training Set

	Metode		
	Naïve Bayes	C4.5	KNN
Berhasil	67,86%	83,57%	87,14%
Error	32,14%	16,43%	12,86%

Berdasarkan hasil dari komparasi Tabel 2, maka diperoleh Metode yang terbaik yaitu metode KNN dengan tingkat keberhasilan sebesar 87,1429% dan tingkat error sebesar 12,8571%.

b. Komparasi Cross Validation 10 Folds

Tabel 3. Komparasi Cross Validation 10 Folds

	Metode		
	Naïve Bayes	C4.5	KNN
Berhasil	68,57%	82,14%	80,71%
Error	31,43%	17,86%	19,29%

Berdasarkan hasil dari komparasi Tabel 3, maka diperoleh Metode yang terbaik yaitu metode C4.5 dengan tingkat keberhasilan sebesar 82,1429% dan tingkat error sebesar 17,8571%.

Kesimpulan

Berdasarkan dari hasil analisa yang telah dilakukan maka dapat disimpulkan bahwa dengan menggunakan komparasi use training set maka metode yang terbaik adalah dengan menggunakan metode KNN (K-Nearest Neighbor) yaitu dengan tingkat keberhasilan sebesar 87,1429% dan tingkat error sebesar 12,8571%. Sedangkan Komparasi dengan Cross Validation 10 Folds Metode yang terbaik yaitu metode C4.5 dengan tingkat keberhasilan sebesar 82,1429% dan tingkat error sebesar 17,8571%.

DAFTAR PUSTAKA

Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4. 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29–43. <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1525%0Ahttps://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/download/1525/944>

- Ariadi, F. (2020). Analisa Perbandingan Algoritma DT C.45 dan Naïve Bayes Dalam Prediksi Penerimaan Kredit Motor. *KERNEL: Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika*, 1(1), 1–8. <https://doi.org/10.31284/j.kernel.2020.v1i1.1183>
- Bayes, N., Pt, P., Putera, J., Sarwiyah, Y., Rahaningsih, N., & Basysyar, F. M. (2020). *Data Nasabah Produk Asuransi Kendaraan Menggunakan Algoritma*. 04(03), 124–130.
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898.
- Diantika, S., Subekti, A., Nalatissifa, H., Lase, M., Studi Magister Ilmu Komputer, P., Nusa Mandiri, U., & Margonda Raya No, J. (2021). *Perbandingan Algoritma Klasifikasi K-Nearest Neighbor, Random Forest dan Gradient Boosting untuk Memprediksi Ketertarikan Nasabah pada Polis Asuransi Kendaraan*. 6(3), 463–469. <http://openjournal.unpam.ac.id/index.php/informatika463>
- Kadir, W. O. N., Pramono, B., & Statiswaty. (2019). Penerapan Data Mining Dengan Metode K- Nearest Neighbor (KNN) Untuk Mengelompokan Minat Konsumen Asuransi (PT. Jasaraharja Putera). *Jurnal SemanTIK*, 5(1), 97–104.
- Karlia, J. A., & Nurmansyah, W. (2021). Penerapan Algoritma C4.5 Untuk Penentuan Keterlambatan Pembayaran Premi Asuransi. *Jurnal Ilmiah Teknik Industri Dan Informasi*, Vol.9(2), 100–113.
- Kurniawan, S., & Hidayat, T. (2007). Penerapan data mining dengan metode interpolasi untuk memprediksi minta konsumen asuransi (Studi Kasus Asuransi Metlife). *Media Informatika*, 5(2), 113–128. <https://journal.uui.ac.id/media-informatika/article/view/114>
- NOVIANTO, R. (2019). Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(1), 85–95. <https://doi.org/10.35957/jatisi.v6i1.150>
- Pattipeilopy, W. F., Wibowo, A., & Utari, D. R. (2017). *Pemodelan Dan Prototipe Sistem*

- Informasi Untuk Prediksi Pembaharuan Polis Asuransi Mobil Menggunakan Algoritma C.45. *Prosiding SNATIF*, 791–799.
- Retno Utari, D., & Wibowo, A. (2020). Pemodelan Prediksi Status Keberlanjutan Polis Asuransi Kendaraan dengan Teknik Pemilihan Mayoritas Menggunakan Algoritma-Algoritma Klasifikasi Data Mining. *Prosiding Seminar Nasional Teknoka*, 5(2502), 19–24. <https://doi.org/10.22236/teknoka.v5i.391>
- Situmorang, M. S., Purwadi, P. P., & ... (2019). Implementasi Data Mining Untuk Menganalisa Minat Nasabah Dalam Pemilihan Produk Asuransi Dengan Menggunakan Algoritma K-Means Pada PT. AJ Central *Jurnal Cyber Tech*, x. <https://ojs.trigunadharma.ac.id/index.php/jct/article/view/2942%0Ahttps://ojs.trigunadharma.ac.id/index.php/jct/article/download/2942/713>
- Ulumi, F. M., & Rismayati, R. (2019). Implementasi Metode Perbandingan Eksponensial (MPE) dalam Pendukung Keputusan Penilaian Kepala Madrasah Terbaik di Kota Mataram (Implementation of the Exponential Comparison (MPE) Method in Supporting the Best Madrasah Head Assessment Decisions in th. 1(3), 188–197.
- Utami, W. P., & Handoko, K. (2020). Penerapan K-Means Dalam Pengelompokan Data Polis Asuransi Kendaraan Bermotor Di Pt Jasaraharja Putera. *Jurnal Comasie*, 01.