

INTEGRASI METODE NEIVE BAYES DAN SOFTWARE RAPIDMINER DALAM ANALISIS HASIL USAHA PERUSAHAAN DAGANG

Rika Nofitri, Novica Irawati

Sekolah Tinggi Manajemen Informatika dan Komputer Royal

email: nofitirika15@gmail.com

Abstract: Profits are the main target in a company. Every company wants the benefits achieved in accordance with the desired target. However, the company thinks the target achieved is due to the large sales. The company does not see other factors that affect profit outcomes such as costs, stocks, purchases and other factors in preparing the company's final report. In addition, the data management that has been carried out so far is not comprehensive, because a lot of data must be analyzed so that it requires tools to manage it. This study intends to analyze the factors that influence profit outcomes based on the profit data. By integrating using the Naive Bayes method and rapidminer software tools, this technique will produce information in the form of a pattern of relationships or interrelationships between factors with each other in influencing the achievement of profits. While rapidminer software to simplify the analysis of data on profit. The results of this study are expected to be the basis for predicting the profits of trading companies,

Keywords: Profit Results, Data Mining, Neive Bayes, Rapidminer

Abstrak: Keuntungan menjadi target utama di sebuah perusahaan. Setiap perusahaan menginginkan keuntungan yang dicapai sesuai dengan target yang diinginkan. Akan tetapi perusahaan beranggapan target yang dicapai tersebut didapat karena hasil penjualan yang besar. Perusahaan tidak melihat faktor lainnya yang mempengaruhi hasil keuntungan seperti biaya, stok, pembelian dan faktor-faktor lain dalam menyusun laporan akhir perusahaan. Selain itu, pengelolaan data yang dilakukan selama ini pun tidak menyeluruh, dikarenakan banyak nya data yang harus dianalisis sehingga memerlukan alat bantu dalam pengelolaannya. Penelitian ini bermaksud untuk menganalisa faktor yang mempengaruhi hasil keuntungan berdasarkan data hasil keuntungan. Dengan melakukan integrasi menggunakan metode naive bayes dan alat bantu software rapidminer, teknik ini akan menghasilkan informasi berupa pola hubungan atau keterkaitan antar factor satu dengan lainnya dalam mempengaruhi tercapainya laba. Sedangkan software rapidminer untuk mempermudah dalam menganalisis data hasil keuntungan. Hasil dari penelitian ini diharapkan menjadi landasan untuk memprediksi hasil keuntungan perusahaan dagang,

Kata Kunci : Hasil Keuntungan, Data Mining, Neive Bayes, Rapidminer

PENDAHULUAN

Keuntungan menjadi target utama di sebuah perusahaan. Setiap perusahaan menginginkan keuntungan

yang dicapai sesuai dengan target yang diinginkan. Selama ini perusahaan beranggapan target yang dicapai tersebut didapat karena hasil penjualan yang besar. Perusahaan tidak melihat

faktor lainnya yang mempengaruhi hasil keuntungan seperti biaya, stok, pembelian dan faktor-faktor lain dalam menyusun laporan akhir perusahaan.

Sama halnya dengan perusahaan UD. Arif Jaya yang bergerak pada penjualan makanan ringan, pengusaha hanya terfokus pada factor omset dalam melihat kinerja usahanya. Hal ini jelas mengabaikan factor lain sehingga proses evaluasi dilakukan dengan factor yang tidak utuh. Ketidak utuhan evaluasi ini akan berdampak pada berkurangnya laba yang mestinya dicapai bahkan sampai pada kerugian dan kebangkrutan. Gudang data yang telah disusun bertahun-tahun belum mampu digunakan untuk membantu usaha dalam menemukan informasi yang berharga untuk kemajuan perusahaan.

Berdasarkan permasalahan yang telah dikemukakan, peneliti berkeinginan melakukan kajian terhadap gudang data keuangan perusahaan untuk menemukan barang berharga yang disebut dengan informasi. Proses penggalian informasi menggunakan teknik data mining menggunakan metode naive bayes yang diintegrasikan menggunakan software Rapidminer. Teknik ini akan menghasilkan informasi berupa pola hubungan atau keterkaitan antar factor satu dengan lainnya dalam mempengaruhi tercapainya laba. Sedangkan software rapidminer untuk mempermudah dalam menganalisis data hasil keuntungan.

Teknik data mining dengan metode naive bayes merupakan algoritma yang menerapkan teori Bayes dalam klasifikasi [1]. Teorema keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern

recognition).

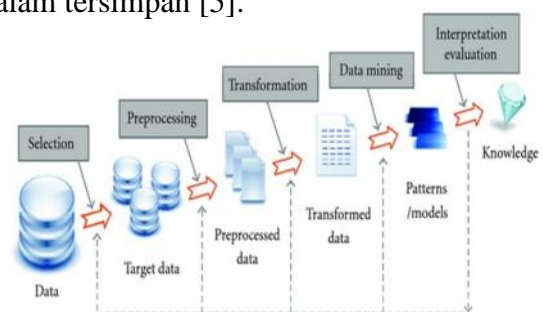
Hasil Keuntungan

Keuntungan adalah diperolehnya kelebihan atas penghasilan selama satu periode akuntansi setelah melakukan perhitungan biaya [2]. Pengertian laba yang diamati oleh struktur akuntansi sekarang ini adalah selisih pengukuran pendapatan dan biaya. Informasi yang didapat dari laporan akuntansi terkait laba ini akan memberikan informasi tentang kinerja perusahaan yang dibutuhkan untuk mengelola perusahaan dimasa depan [3].

Definisi Data Mining

Data mining atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya [4].

Data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan dua metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilihan data. Sebenarnya data mining adalah sebagian dari proses Knowledge Discovery in Database. Melalui beragam algoritma dan teknik, data mining bertujuan untuk mengekstrak pengetahuan dari kumpulan data yang telah tersimpan [5].



Gambar 1. KDD

Gambar 1 menunjukkan bahwa data mining merupakan bagian dari KDD yang memiliki langkah sebagai berikut :

1. 1. Selection : Data dipilih dari kumpulannya sesuai dengan kebutuhan yang akan diteliti.
2. Data Preprocessing : Dimulai dari proses data cleaning (pembersihan data) dari kesalahan, noise, data ganda, tidak konsisten, hilang, rusak. Selanjutnya dilakukan proses enrichment, guna melengkapi data yang ada dengan informasi lain yang dibutuhkan dari eksternal data yang ada.
3. Data Transformation : Melakukan perubahan bentuk menjadi symbol atau coding pada data yang sebelumnya sudah mengalami preprocessing. Hal ini bertujuan untuk mempermudah proses ekstraksi pengetahuan karena data yang akan diproses sudah disesuaikan dengan konsep data mining.
4. Data mining : Mengimplementasikan algoritma dan teknik mining untuk mengekstrak pengetahuan baik berupa penemuan pola, kecenderungan, prediksi dalam lainnya.
5. Evaluation : Menguji kesesuaian dari hasil pengetahuan yang telah diekstrak dengan fakta yang sebenarnya. Proses ini juga dilakukan untuk menyajikan informasi yang dapat dipahami oleh pihak pengambil keputusan

Metode Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan

menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [6].

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu [1]

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

$P(H|X)$: peluang dari hipotesa H berdasar dari kondisi X

X : Sampel data dari kelas (label) yang tidak diketahui

H : merupakan hipotesa bahwa X adalah data dengan kelas (label) C.

$P(H)$: peluang dari hipotesa H

$P(X)$: adalah peluang dari X yang diamati

$P(X|H)$: peluang X, berdasarkan kondisi pada hipotesa H. Jika asumsi atribut tidak saling terkait (conditionally independent), maka:

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$

Bila $P(X|C_i)$ dapat diketahui melalui perhitungan di atas, maka kelas (label) dari data sampel X adalah kelas (label) yang memiliki $P(X|C_i) \cdot P(C_i)$ maksimum.

Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss sebagai berikut.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Di mana :

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas yang dicari

yi : Sub kelas Y yang dicari

μ : mean, menyatakan rata – rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut.

Akurasi = $\frac{\text{Jml Prediksi Yang Benar}}{\text{Jml Prediksi keseluruhan}}$

Error = $\frac{\text{Jml Prediksi Yang Salah}}{\text{Jml Prediksi keseluruhan}}$

Software RapidMiner

RapidMiner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data [7]. RapidMiner pertama kali dinamai Yet Another Learning Environment atau singkat YALE. Pada tahun 2007 akhirnya diganti namanya menjadi RapidMiner.

METODE

Adapun kerangka kerja dari penelitian ini yaitu sebagai berikut :

1. Identifikasi Masalah
Masalah yang diidentifikasi dalam penelitian ini adalah belum digunakannya gudang data keuangan perusahaan untuk mendapatkan informasi yang berguna bagi perusahaan.
2. Mengkaji Bahan Literatur
Ini bertujuan untuk mendapatkan pemahaman yang dalam terhadap masalah yang akan diteliti dan metode yang digunakan sebagai solusi. Bahan literatur yang sangat penting publikasi karya ilmiah baik jurnal maupun proceesing, buku-buku terkait data mining, tutorial

penggunaan software Rapidminer dan referensi lainnya.

3. Pengumpulan Data
Data yang dikumpulkan diperoleh melalui wawancara dan pengamatan secara langsung kelapangan yaitu UD. Arif jaya. Data yang dikumpulkan hanya yang berkaitan dengan data keuangan perusahaan atau data keuangan yang terlebih dahulu ditransformasikan,
4. Menganalisis Data
Data yang dikumpulkan akan dianalisis menggunakan teknik data mining dengan metode neive bayes.
5. Penerapan Sistem
Hasil analisis data yang dilakukan secara manual menggunakan metode neive bayes kemudian diintergrasikan software untuk mengetahui apakah hasil yang diperoleh memiliki kesesuaian dan memiliki tingkat akurasi yang baik.

HASIL DAN PEMBAHASAN

Hasil pengumpulan data akan dilakukan transfromasi data terlebih dahulu yaitu dengan melakukan seleksi data, transformasi data, dan pembersihan data untuk dijadikan sebagai *decission system* yang akan diterapkan kedalam metode *naive bayes.decission system* akan menjelaskan sejumlah n objek, E1, E2, E3,...,E120. Dengan atribut kondisi Omset, Pembelian, Stok dan Biaya. Sedangkan yang menjadi atribut kuputusan adalah Keuntungan.

Adapun *decission system* yang telah diproses tersebut yaitu sebagai berikut :

Tabel 1. Decission System

Objek	Omset	Pembelian	Stok	Biaya	Keuntungan
E1	Target	Besar	Banyak	Efisien	Target Tercapai
E2	Target	Besar	Banyak	Efisien	Target Tercapai
E3	Target	Besar	Banyak	Efisien	Target Tercapai
E4	Target	Besar	Banyak	Efisien	Target Tercapai
E5	Target	Besar	Banyak	Efisien	Target Tercapai
E6	Target	Besar	Banyak	Efisien	Target Tercapai
E7	Target	Besar	Banyak	Efisien	Target Tercapai
E8	Target	Besar	Banyak	Efisien	Target Tercapai
....
E119	Target	Kecil	Sedikit	Tidak Efisien	Target Tercapai
E120	Target	Kecil	Sedikit	Tidak Efisien	Target Tercapai

Selanjutnya menghitung probabilitas dari masing-masing atribut mulai dari omset, pembelian, stok dan biaya serta atribut keputusan yaitu keuntungan yang diperoleh. Adapun tabel probabilitas tersebut sebagai berikut.

Tabel 2. Nilai Probabilitas Omset

Omset	Jumlah Kejadian "Keuntungan"		Probabilitas	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Target	36	10	0,480	0,222
Lebih	31	0	0,413	0,000
Kurang	8	35	0,107	0,778
Jumlah	75	45	0,625	0,375

Tabel 3. Nilai Probabilitas Pembelian

Pembelian	Jumlah Kejadian "Keuntungan"		Probabilitas	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Besar	50	22	0,667	0,489
Kecil	25	23	0,333	0,511
Jumlah	75	45	0,625	0,375

Tabel 4. Nilai Probabilitas Stok

Stok	Jumlah Kejadian "Keuntungan"		Probabilitas	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Banyak	58	22	0,773	0,489
Sedikit	17	23	0,227	0,511
Jumlah	75	45	0,625	0,375

Tabel 5. Nilai Probabilitas Biaya

Biaya	Jumlah Kejadian "Keuntungan"		Probabilitas	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Efisien	58	0	0,773	0
Tidak Efisien	17	45	0,227	1
Jumlah	75	45	0,625	0,375

Tabel 6. Nilai Probabilitas Keuntungan

Dipilih Untuk Keuntungan	Jumlah Kejadian "Keuntungan"		Probabilitas	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Jumlah	75	45	0,625	0,375

Setelah mendapatkan nilai probabilitas diatas, maka akan dilakukan uji test berikut :

1. Apabila diketahui hasil keuntungan dengan omsetnya kurang, pembelian nya kecil, stok nya banyak, biaya nya tidak efisien.

Maka dapat dihitung :

$$\text{Tercapai} = 0,107 * 0,333 * 0,733 * 0,227$$

$$* 0,625 = 0,004$$

$$\text{Tidak Tercapai} =$$

$$0,778 * 0,511 * 0,489 * 1 * 0,375 = 0,073$$

Nilai probabilitas diatas dapat dihitung melakukan normalisasi terhadap likelihood tersebut sehingga jumlah nilai yang diperoleh = 1

$$\text{Probabilitas Tercapai}$$

$$= \frac{0,004}{0,004 + 0,073} = 0,052$$

$$\begin{aligned} & \text{Probabilitas Tidak Tercapai} \\ &= \frac{0,073}{0,073 + 0,004} = 0,948 \end{aligned}$$

2. Apabila diketahui hasil keuntungan dengan omsetnya target, pembelian nya besar, biaya nya tidak efisien. Maka dapat dihitung :

$$\begin{aligned} & \text{Tercapai} \\ &= 0,480 * 0,667 * 0,227 * 0,625 \\ &= 0,246 \end{aligned}$$

Tidak

$$\begin{aligned} & \text{Tercapai} = 0,222 * 0,489 * 1 * 0,375 \\ &= 0,041 \end{aligned}$$

Nilai probabilitas diatas dapat dihitung melakukan normalisasi terhadap likelihood tersebut sehingga jumlah nilai yang diperoleh = 1

$$\begin{aligned} & \text{Probabilitas Tercapai} \\ &= \frac{0,246}{0,246 + 0,041} = 0,857 \end{aligned}$$

$$\begin{aligned} & \text{Probabilitas Tidak Tercapai} \\ &= \frac{0,041}{0,041 + 0,246} = 0,143 \end{aligned}$$

3. Apabila diketahui hasil keuntungan dengan omsetnya kurang, pembelian nya kecil, biaya nya tidak efisien. Maka dapat dihitung :

$$\begin{aligned} & \text{Tercapai} \\ &= 0,107 * 0,333 * 0,227 * 0,625 \\ &= 0,005 \end{aligned}$$

Tidak

$$\begin{aligned} & \text{Tercapai} = 0,778 * 0,511 * 1 * 0,375 \\ &= 0,149 \end{aligned}$$

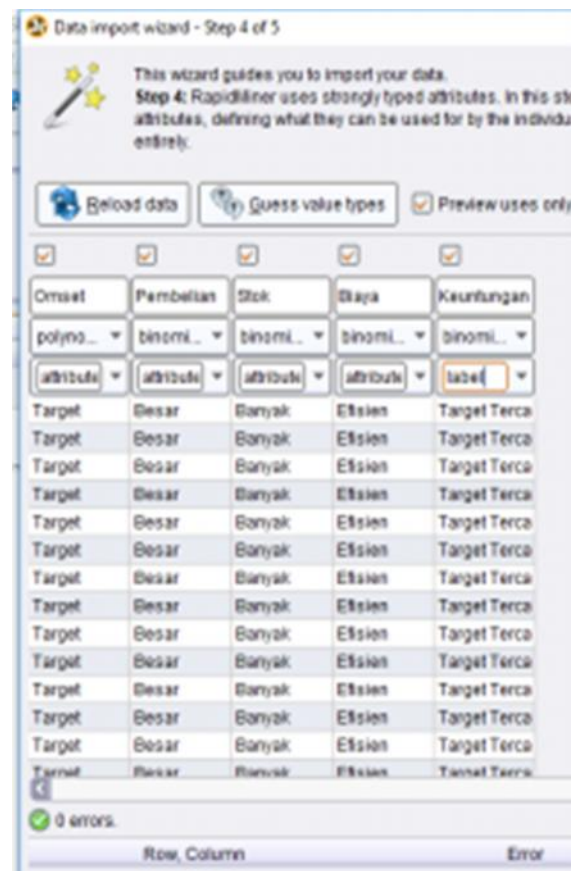
Nilai probabilitas diatas dapat dihitung melakukan normalisasi terhadap likelihood tersebut sehingga jumlah nilai yang diperoleh = 1

$$\begin{aligned} & \text{Probabilitas Tercapai} \\ &= \frac{0,005}{0,005 + 0,149} = 0,032 \end{aligned}$$

Probabilitas Tidak Tercapai

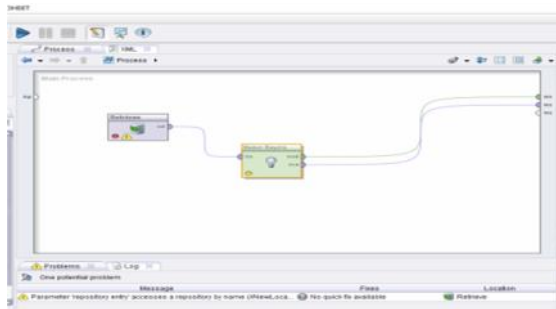
$$= \frac{0,149}{0,149 + 0,005} = 0,968$$

Selanjutnya proses pengujian menggunakan alat bantu *software rapidminer*. Langkah awal melakukan proses import data berupakan *decision system*. Adapun data import yang dimaksud yaitu sebagai berikut :



Gambar 2. Data Import Wizard

Setelah itu dilakukan proses koneksi antara data yang telah diimport dengan operator metode *naive bayes* didalam menu operator. Kemudian men-drag data dan operator tersebut atau melakukan proses pemindahan kedalam jendela *process*. Maka terlihat koneksi tersebut pada gambar sebagai berikut.



Gambar 3. Koneksi *Data Import* dengan Metode *Naive Bayes*

Selanjutnya dilakukan proses menjalankan aplikasi dengan mengklik tombol *run* pada jendela aplikasi. Sehingga dari hasil proses yang dijalankan mendapatkan hasil yaitu sebuah probabilitas kemunculan setiap nilai dari masing-masing atribut kondisi hasil keuntungan. Hasil probabilitas yang telah diproses dapat terlihat pada gambar berikut ini.

Attribute	Parameter	Target Tercapai	Target Tdk Tercapai
Omset	value=Target	0.480	0.222
Omset	value=Lebih	0.413	0.000
Omset	value=Kurang	0.107	0.777
Omset	value=unknown	0.000	0.000
Pembelian	value=Besar	0.667	0.489
Pembelian	value=Kecil	0.333	0.511
Pembelian	value=unknown	0.000	0.000
Stok	value=Banyak	0.773	0.489
Stok	value=Sedikit	0.227	0.511
Stok	value=unknown	0.000	0.000
Biaya	value=Efisien	0.773	0.000
Biaya	value=Tidak Efisien	0.227	1.000
Biaya	value=unknown	0.000	0.000

Gambar 4. Nilai Probabilitas

Gambar tersebut merupakan probabilitas kemunculan dari masing-

masing atribut yang dihasilkan *software rapidminer*. maka selanjutnya akan dilakukan proses test set berikut.

1. Apabila diketahui hasil keuntungan dengan omset target, pembelian nya kecil, stok nya sedikit, biaya nya tidak efisien. Maka dapat dihitung :
Tercapai = $0,480 * 0,333 * 0,227 * 0,227 * 0,625 = 0,005$

Tidak Tercapai = $0,222 * 0,511 * 0,511 * 1 * 0,291 = 0,017$
Maka proses selanjutnya menghitung :

Probabilitas Tercapai

$$= \frac{0,005}{0,005 + 0,017} = 0,227$$

Probabilitas Tidak Tercapai

$$= \frac{0,017}{0,017 + 0,005} = 0,773$$

2. Apabila diketahui hasil keuntungan dengan omset nya kurang, pembelian nya besar, dan biaya nya tidak efisien. Maka dapat dihitung :
Tercapai = $0,107 * 0,667 * 0,227 * 0,625 = 0,010$

Tidak Tercapai = $0,777 * 0,489 * 1 * 0,291 = 0,111$

Maka proses selanjutnya menghitung :

Probabilitas Tercapai

$$= \frac{0,010}{0,010 + 0,111} = 0,083$$

Probabilitas Tidak Tercapai

$$= \frac{0,111}{0,111 + 0,010} = 0,917$$

3. Apabila diketahui hasil keuntungan dengan omsetnya target, pembelian nya kecil. Maka dapat dihitung :
Tercapai = $0,480 * 0,333 * 0,625 = 0,100$

Tidak Tercapai = $0,222 * 0,511 * 0,291 = 0,033$

Maka proses selanjutnya menghitung
:

$$\begin{aligned} &\text{Probabilitas Tercapai} \\ &= \frac{0,100}{0,100 + 0,033} = 0,752 \\ &\text{Probabilitas Tidak Tercapai} \\ &= \frac{0,033}{0,033 + 0,100} = 0,248 \end{aligned}$$

SIMPULAN

Dari hasil integrasi metode *naive bayes* dan *software rapidmineryang* telah dilakukan adalah metode *naive bayes* menghasilkan prediksi yang akan membantu pelaku usaha dalam meningkatkan hasil keuntungan, begitu pun dengan *software rapidminer* yang dapat membantu mempermudah proses prediksi hasil keuntungan.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada DRPM Direktorat Jenderal Penguat Riset dan Pengembangan Kementerian Riset, Teknologi dan Pendidikan Tinggi Sesuai dengan Kontrak Penelitian PDP Tahun Anggaran 2019.

DAFTAR PUSTAKA

- [1] Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *jurnal EECCIS*, 7(1), 59-64.
- [2] Adiwiratama, J. (2012). Pengaruh Informasi Laba, Arus Kas dan Size Perusahaan terhadap return Saham (Studi empiris pada

perusahaan manufaktur yang terdaftar di BEI). *JINAH (Jurnal Ilmiah Akuntansi dan Humanika)*, 2(1).

- [3] Hartono, S. (2012). Pengaruh Informasi Laba Dan Arus Kas Terhadap Harga Saham. *Akurat: Jurnal Ilmiah Akuntansi*, 3(7).
- [4] Pane, D. K. (2013). Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus: Kreditplus). *Pelita Informatika: Informasi dan Informatika*, 4(3).
- [5] Maburur, A. G. (2012). Penerapan data mining untuk memprediksi kriteria nasabah kredit. *KOMPUTA: Jurnal Komputer dan Informatika*, 1(1).
- [6] Saleh, A. (2015). Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga. *Creative Information Technology Journal*, 2(3), 207-217.
- [7] Zulkifli, A. (2016). Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler. *Riau Journal Of Computer Science*, 2(1), 65-76.