

Penerapan Data Mining untuk Memprediksi Minat Nasabah Terhadap Produk Asuransi Meninggal Dunia dengan Metode Naïve Bayes (Studi Kasus : PT. BNI Life Insurance)

Ari Hidayatullah¹, Ena Mudiawati², Muhammad Syafrullah³

^{1,2,3}Magister Ilmu Komputer, Universitas Budi Luhur

e-mail: *¹ arihidayatullah788@gmail.com, ² enamudiaww@gmail.com,
³ mohammad.syafrulloh@budiluhur.ac.id

INTISASI

Pendapatan untuk perusahaan asuransi ditentukan oleh jumlah premi yang dibayar oleh nasabah. faktor penting nasabah berupa premi, premi ditentukan dalam persentase atau tarif tertentu. Pada perusahaan asuransi pasti memiliki jumlah data, dan data tersebut sangat penting bagi perusahaan untuk mengetahui kriteria nasabah yang berminat pada asuransi yang dipasarkan. Dengan adanya informasi dari data nasabah yang ada, perusahaan asuransi dapat mengambil suatu keputusan dalam menerapkan strategi perusahaan diantaranya yaitu menjual produk- produk promo untuk meningkatkan pendapatan perusahaan. Data mining merupakan suatu teknologi yang dapat membantu perusahaan dalam menemukan suatu yang sangat penting dari sekumpulan data. Data mining dapat membentuk suatu pola atau membuat suatu sifat perilaku bisnis yang berguna untuk pengambilan keputusan. Dengan menggunakan metode algoritma Naive Bayes diharapkan bisa membantu perusahaan dalam pengelolaan data nasabah dengan cara mengklasifikasi data nasabah untuk memprediksi minat nasabah dengan tingkat akurasi melebihi 80% dalam memilih produk asuransi meninggal dunia.

Kata Kunci: asuransi, naïve bayes, prediksi, data mining.

ABSTRACT

Income for insurance companies is determined by the amount of premium paid by the customer. Important factors for customers in the form of premiums, premiums are determined in certain percentages or rates. The insurance company certainly has the amount of data, and the data is very important for companies to know the criteria of customers who are interested in the insurance marketed. With the information from existing customer data, the insurance company can make a decision in implementing the company's strategy, which is to sell promo products to increase company revenue. Data mining is a technology that can help companies find a very important set of data. Data mining can form a pattern or create a nature of business behavior that is useful for decision making. By using the Naive Bayes algorithm method, it is expected to be able to assist companies in managing customer data by classifying customer data to predict customer interest with an accuracy rate exceeding 80% in choosing a death insurance product.

Keywords: insurance, naïve bayes, predictions, data mining..

I. PENDAHULUAN

Istilah asuransi sudah menjadi hal yang biasa bagi kalangan masyarakat, Asuransi digunakan untuk bermacam kebutuhan seperti meminimalkan resiko keuangan jika terjadi hal-hal yang tidak terduga seperti meninggal dunia.

Salah satu perusahaan Asuransi yang menyediakan berbagai macam produk asuransi yaitu PT. BNI Life Insurance, produk-produk asuransi yang dikeluarkan BNI Life diantaranya, : Solusi pintar, Solusi Abadi, SmartEdu dan masih banyak lagi seperti

produk-produk promo lainnya yaitu Optima Saving.

Pada awal tahun 2019 PT. BNI Life Insurance mendapatkan peningkatan premi sebesar 10% dari awal tahun 2018, dari peningkatan premi tersebut masih terdapat beberapa produk asuransi yang belum tepat sasaran, seperti produk asuransi Meninggal Dunia, dan produk asuransi mengenai masa tua. jika produk-produk tersebut dikelola dengan baik dan tepat sasaran, penulis mengasumsikan peningkatan premi akan lebih besar dari 10%.

Optima Saving adalah salah satu produk promo PT. BNI Life yang mempunyai benefit meninggal dunia, dimana nasabah akan mendapatkan uang pertanggungan lebih ketika nasabah tersebut meninggal dunia. Dari tahun 2017 –

2018 produk Optima Saving terus di promosikan tetapi pada awal tahun 2019 sampai dengan sekarang promo sudah tidak dipasarkan lagi.

Sehubungan dengan target perusahaan PT. BNI Life untuk meningkatkan premi produk Optim Saving ini sangat membantu dalam meningkatkan premi tersebut. Maka dari itu perlu prediksi yang efektif dalam menargetkan pasar untuk promo produk Optima Saving tersebut.

Berdasarkan uraian di atas diperlukan analisis dari data-data nasabah yang telah membeli produk Optima Saving. Data Mining merupakan teknologi yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Data mining meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatiskan yang dilakukan oleh data mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan. Data Mining dapat menjawab pertanyaan-pertanyaan bisnis yang dengan cara tradisional memerlukan banyak waktu untuk menjawabnya. Data Mining mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi pemrediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka. (Veronika, 2002).

Melihat kondisi diatas maka penulis tertarik untuk melakukan PENERAPAN DATA MINING UNTUK MEMREDIKSI MINAT NASABAH TERHADAP PRODUK ASURANSI MENINGGAL DUNIA

(OPTIMA SAVING) DENGAN METODE NAÏVE BAYES yang di pasarkan oleh BNI Life.

Tinjauan studi dalam penelitian ini mengambil beberapa ringkasan penelitian sebelumnya yang berkaitan dengan metode yang digunakan dalam pendukung keputusan.

1. Penelitian dari (Bustani, 2014) “Penerapan Algoritma Naïve Bayes Classifier untuk Mengklasifikasi Data Nasabah Asuransi” oleh (Bustani, 2014) mengatakan bahwa menggunakan algoritma naïve bayes untuk mengklasifikasi data nasabah asuransi mempunyai tingkat akurasi hasil yang lebih dari 80%.
2. Penelitian dari (Betrisandi, 2017) “Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naïve Bayes berbasis Backward Eliminations” mengatakan bahwa dari sembilan atribut yang digunakan yaitu umur, pekerjaan, penghasilan, jenis asuransi, jangka waktu, cara bayar, mata uang, premi, dan jumlah hari dengan menggunakan algoritma Naive Bayes berbasis Backward Elimination didapatkan delapan atribut weight yaitu umur, jangka waktu, cara bayar, premi, jumlah hari, pekerjaan, penghasilan dan mata uang dalam mengklasifikasi nasabah yang lancar dan tidak lancar dalam pembayaran premi. Secara mandiri tingkat akurasi yang dihasilkan algoritma Naive Bayes adalah 83,02 % dengan menambahkan seleksi fitur Backward Elimination menghasilkan akurasi 85,89 % dalam klasifikasi nasabah lancar dan tidak lancar
3. Penelitian dari (Lailatul M. Chaira, Nurul Hidayat, Inu L. Wibowo, Imam Mukhlash, 2016) “Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Dengan Metode Naive Bayes Classifier” mengatakan bahwa tentang bagaimana menentukan jenis asuransi yang tepat menggunakan task dalam data mining untuk menggali informasi yang berkaitan dengan kebutuhan produk asuransi bagi calon nasabah. Metode yang digunakan untuk klasifikasi adalah Naïve Bayes Classifier. Hasil uji coba menunjukkan bahwa metode NBC mampu mengklasifikasi record dengan tingkat kinerja tertinggi sebesar 94.12% ketika proporsi pembagian data latih 90% dan data uji sebesar 10%.

4. Penelitian dari (Sandy Kurniawan, Taufiq Hidayat, 2007) “Penerapan Data Mining Dengan Metode Interpolasi Untuk Memprediksi Minat Konsumen Asuransi (Studi Kasus Asuransi Metlife)” mengatakan bahwa tentang teknik data mining dapat diterapkan untuk mengetahui besaran premi asuransi yang terbaik bagi setiap nasabah atau calon nasabah sesuai dengan tingkat pendapatan per tahunnya, Semakin besar jumlah data maka proses perhitungan akan semakin lambat karena banyaknya pengolahan data, tetapi hasil data mining akan semakin bermutu tinggi karena berdasarkan data yang besar.
5. Penelitian dari (Rendra Octa Chandra, A. Haidar Mirza, Kurniawa, 2013) “Penerapan Data Mining Untuk Memprediksi Minat Nasabah Pada Ajb Bumiputera 1912 Palembang” mengatakan bahwa Penerapan data mining dengan teknik association rules dan algoritma a priori yang dilakukan menghasilkan sebuah informasi mengenai minat nasabah berdasarkan hubungan antara kriteria nasabah terhadap jenis asuransi yang dipilih, Berdasarkan dua kali penerapan data mining menggunakan XLMiner yang dilakukan dimana pada penerapan yang pertama ditentukan nilai minimum support 50 dan minimum confidence 50% dihasilkan 6 rules dan pada penerapan yang kedua ditentukan nilai minimum support 2 dan minimum confidence 1% dihasilkan 67 rules hal ini menerangkan bahwa semakin tinggi batasan nilai minimum support dan minimum confidence yang ditentukan maka semakin baik pula pola/aturan yang diperoleh. Dan semakin kecil batasan nilai minimum support dan minimum confidence yang ditentukan maka semakin banyak pula pola/aturan yang dapat dihasilkan dan waktu yang diperlukan lebih banyak.
6. Penelitian dari (Andy Victor Pakpahan, Hendra Irawan, 2017) “Penerapan Algoritma Naive Bayes Untuk Menentukan Nasabah Potensial Pada Ajb Bumiputera 1912” mengatakan bahwa Algoritma Naive bayes dapat membantu pihak perusahaan untuk melakukan pemanfaatan data dengan baik seperti untuk melakukan klasifikasi data nasabah potensial, klasifikasi dapat dilakukan untuk data perorangan maupun data berkelompok. Proses klasifikasi untuk data kelompok yang cukup banyak berpengaruh terhadap performa dari aplikasi, apabila di import data lebih dari ratusan data akan memperlama proses klasifikasi.
7. Penelitian dari (Azuralia Abu Bakar dkk 2010) melakukan penelitian tentang metode dalam data mining untuk mengklasifikasikan data nasabah ke dalam 5 kelas produk asuransi dengan metode asosiatif rule. Sistem yang dihasilkan memiliki kinerja sebesar 58%.
8. Penelitian dari (S. Balaji, S. K. Srivasta 2012) mengatakan bahwa dalam penelitian ini teknik klasifikasi untuk preferensi produk asuransi menarik pelanggan. dengan melihat data metode klasifikasi bayes digunakan untuk melakukan kebijakan preferensi pelanggan asuransi jiwa. Navie bayes classifier adalah salah satu metode yang efektif dibandingkan dengan pohon keputusan dan classifier jaringan saraf telah menemukan itu sebanding untuk semua pengklasifikasi lainnya. Hasil analisis menunjukkan bahwa Navie bayes berpotensi dapat menjadi efektif dalam melakukan analisis preferensi pelanggan atas produk asuransi jiwa. Namun, makalah ini mengamati aplikasi KDD / DM di domain asuransi, masalah lain mungkin juga signifikan, seperti mempertimbangkan kebijakan atribut pelanggan lainnya (kelompok usia) preferensi terhadap produk asuransi. Analisis dapat diperpanjang dengan pilihan kombinasi optimal fitur untuk membangun model klasifikasi. Gaya hidup karakteristik pelanggan juga dapat dipertimbangkan preferensi kebijakan terhadap produk asuransi.
9. Penelitian dari (Ari Wibowo, 2015) “Analisis Perbandingan Kinerja Metode Klasifikasi Dalam Data Mining” mengatakan bahwa pengujian untuk semua atribut yang dimasukkan sekaligus pada semua metode klasifikasi. Semua atribut diinput ke dalam model, kemudian dihitung tingkat akurasi, selanjutnya dilakukan analisa terhadap hasil pengujian. Ada 11 atribut yang diuji yaitu atribut income, kelamin, kelas pekerjaan, kode agen, merokok, p_mode, payment, phone, rawat inap, status kawin, dan tahun lahir. Perhitungan akurasi dihitung berdasarkan rata-rata dari masing-masing kelas dan ketepatan prediksi untuk semua kelas. Berdasarkan tingkat akurasi yang dihasilkan metode Bagging CART memberikan tingkat akurasi yang paling baik bila dibandingkan dengan metode CART, dan Random Forest, sedangkan

metode Random Forest memberikan tingkat akurasi yang paling rendah. Hal ini disebabkan karena pada metode Bagging CART ada pembangkitan learning sampel yang akan mereduksi variansi atribut predictor, sehingga ketika dikombinasikan hasilnya lebih baik bila dibandingkan dengan predictor tunggal yang dibangun untuk menyelesaikan masalah yang sama.

10. Penelitian dari (Anjelisa, Lisbeth Mananeke, Mirah Rogi, 2018)“Analisis Pengaruh Strategi Segmentasi, Targeting Dan Positioning (Stp) Terhadap Keputusan Pembelian Produk Bp-Smart Protection Di Ajb Bumiputera 1912 Cabang Sam Ratulangi Manado” mengatakan bahwa Berbagai kondisi dan perubahan yang terjadi dalam lingkungan bisnis akan berdampak pada meningkatnya persaingan. Persaingan di dunia bisnis semakin kompetitif, kondisi ini menuntut perusahaan harus lebih kreatif dan inovatif dengan menawarkan sesuatu yang bernilai lebih, dibandingkan yang dilakukan pesaing. Oleh karena itu perusahaan mulai membuat strategi untuk mendapatkan perhatian dari konsumen. Tujuan dari penelitian ini untuk melihat pengaruh segmentasi, targeting dan positioning terhadap keputusan pembelian. Metode penelitian yang digunakan adalah asosiatif dan teknik pengambilan sampel dilakukan dengan cara sampling jenuh dengan jumlah responden 50. Pengujian hipotesis dilakukan dengan metode analisis linier berganda menggunakan perangkat lunak SPSS versi 20. Hasil penelitian ditemukan bahwa analisis segmentasi, targeting dan positioning secara simultan dan parsial berpengaruh terhadap keputusan pembelian. Pihak perusahaan sebaiknya meningkatkan efektifitas dalam penerapan strategi segmentasi, targeting dan positioning sebagai penunjang dalam peningkatan penjualan yang telah dicapai

II. METODOLOGI PENELITIAN

A. Metode Penelitian

Menurut (Dawson : 2009) ada empat metode penelitian yang umum digunakan yaitu tindakan penelitian, eksperimen, studi kasus dan survey. Dalam konteks penelitian ini menggunakan eksperimen, yaitu suatu metode yang dilakukan dengan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis,

pengujian hipotesis, menafsirkan hasil, dan kesimpulan (Berndtsson: 2008).

Jenis eksperimen dibagi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik (Kothari : 2004). Adapun jenis penelitian ini bersifat eksperimen absolut, merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variable atau attribute.

Penelitian ini jika dilihat dari bentuk data dan informasi yang dikelola, penelitian ini tergolong jenis penelitian kuantitatif. Penelitian kuantitatif adalah penelitian yang hipotesisnya dapat diuji dengan teknik-teknik statistic (Moedjiono :

2016). Dalam penelitian ini penulis menggunakan metode kualitatif dengan studi kasus, metode kualitatif menekankan pada definisi dalam situasi tertentu seperti populasi atau sampel, dalam teknik pengambilan sampel dilakukan secara acak. Sedangkan pendekatan dengan studi kasus merupakan eksplorasi mendalam terhadap program, kejadian, proses, aktifitas, terhadap satu atau lebih.

Pada metode penelitian kuantitatif, peneliti melakukan pengambilan sampel

data nasabah ada pada objek penelitian kemudian data tersebut diolah menggunakan Metode Naïve Bayes kemudian hasil akhirnya adalah mengenasi prediksi nasabah terhadap produk asuransi meninggal dunia.

B. Sampling/Metode Pemilihan Sampel

Teknik sampling adalah teknik yang dilakukan dalam pengambilan sampel yang akan digunakan dalam penelitian. Adapun data yang digunakan pada penelitian ini adalah data sekunder. Data sekunder adalah data yang sudah ada sebelumnya. Adapun data yang dimaksud pada penelitian ini adalah data nasabah (nama, nomor induk, jenis kelamin, status perkawinan, dll).

Populasi dalam penelitian ini adalah pemilihan sampel dengan mengidentifikasi populasi target yaitu populasi yang relevan dengan tujuan masalah dalam penelitian. Populasi dalam penelitian ini adalah data nasabah.

Teknik pengambilan sampel yang digunakan adalah convenience sampling. Convenience sampling adalah pengambilan sampel didasarkan pada ketersediaan elemen dan kemudahan untuk mendapatkannya (Sugiyono : 2001). Sampel yang digunakan pada penelitian ini adalah nasabah PT BNI Life

Pada penelitian ini digunakan data primer yaitu data calon nasabah dari tahun 2017 – 2018, dikarnakan produk yang menjadi bahan penelitian yaitu produk promo, maka data yang di dapatkan tidak begitu banyak.

C. Metode Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan informasi sehingga data yang diperoleh berhubungan dengan penelitian ini. Metode pengumpulan data yang digunakan oleh peneliti, antara lain:

1. Observasi

Observasi atau pengamatan langsung terhadap objek penelitian. Observasi merupakan metode pengumpulan data dengan melakukan pengamatan dan pencatatan secara teliti dan sistematis atas gejala-gejala atau fenomena yang sedang diteliti. Teknik observasi dilakukan dengan observasi berstruktur dengan menyiapkan daftar kebutuhan data dan sumber data.

2. Wawancara

Wawancara merupakan teknik pengumpulan data yang dilakukan melalui tatap muka dan tanya jawab langsung antara peneliti dan narasumber.

3. Studi Pustaka

Mempelajari literatur tentang metode Naïve Bayes serta hal-hal lain yang terkait dengan penelitian.

3.4. Instrumentasi

Berikut rincian instrumentasi yang digunakan atau dibutuhkan dalam penelitian antara lain:

1. Perangkat Keras:

- Processor Inter(R) Core(TM) i7-7500U CPU @ 2.70GHz (4 CPUs),~2.9GHz
- Memory 16 Gb
- Graphics NVIDIA GEFORCE 940Mx
- Storage Internal 500Gb SSD

2. Perangkat Lunak:

- Windows 10 Pro 64 Bit b. SQL Server 2012

- Instrumen yang dibutuhkan dalam pengumpulan data dengan melakukan pengamatan langsung di PT. BNI Life Insurance dan berdiskusi dengan unit terkait dalam studi kasus yang sedang dikerjakan oleh peneliti.
- Instrumen yang dibutuhkan dalam pengumpulan data dengan melakukan Studi Pustaka untuk mempelajari literatur tentang metode naïve bayes serta hal-hal lain yang terkait dengan penelitian.
- Instrumen untuk hasil informasi dari data yang telah dikelola adalah pada tahap UAT yang akan dilakukan oleh unit terkait yaitu unit Bisnis Produk Development dan pemasaran.

D. Teknik Analisis/Rancangan dan Pengujian Data/ Sistem/ Prototipe/ Model/ Rencana Strategi

Rencana Strategi yang dilakukan oleh peneliti yaitu berdiskusi langsung kepada unit terkait mengenai cara penentuan atribut yang digunakan pada saat explorasi data nanti. Atribut yang dipilih berdasarkan pakar dari unit Bisnis

Product Development. Atribut digunakan pada saat pengelolaan data pada Tabel 1.

Tabel 1 Atribut

No	Atribut	Keputusan Pakar
1	Jenis Kelamin	Pria Wanita
2	Status	Lajang Kawin Duda Janda Cerahi
3	Usia	Usia < 30 Usia >= 30 <=50 Usia > 50
4	Pekerjaan	Pegawai swasta Direktur Lainnya PNS Pegawai BUMN Dokter TNI / Polri Wiraswasta Dosen Pensiunan Ibu rumah tangga Notaris Guru Pegawai honorer Pengacara Paramedis Mahasiswa Kontraktor
5	Pendapatan	Uang < 300 Jt Uang >=300 <=600 Jt Uang > 600 Jt
6	Lama Asuransi	1 2 3 4
7	Cara Bayar	Bulanan Semesteran

E. Langkah-langkah Penelitian

Langkah-langkah penelitian yang berisi langkah-langkah penelitian yang direncanakan

oleh peneliti dalam rangka melaksanakan penelitian sesuai dengan kerangka konsep/pola pikir pemecahan masalah dapat dilihat pada diagram alur berikut:



Gambar 1 Diagram Alur Metode Penelitian

F. Pengumpulan Data

Pengambilan data dilakukan di PT. BNI Life Insurance, data ini adalah data historis nasabah yang mengambil produk asuransi meninggal dunia pada tahun

2017-2018. Data tersebut tidak boleh di publis keluar maka penulis mengkombinasikan nama, tanggal lahir maupun alamat nasabah tidak sama dengan data sebenarnya.

G. Preprocessing

Pada umumnya data mentah yang diperoleh masih mengandung beberapa data yang tidak bisa digunakan dalam proses data mining dikarenakan adanya kesalahan atau anomali pada data. Sehingga diperlukan pemrosesan awal data atau Pre-processing sebelum melakukan penggalian informasi terhadap data. Dimana data yang sudah diperoleh akan di normalisasi, dengan kata lain data akan dihilangkan noise-nya. Dalam penelitian ini preprocessing dimulai dengan penghapusan duplikasi data, apabila terdapat 2 atau lebih data dengan nomor polis yang sama dalam satu dataset maka salah satu data tersebut harus dihilangkan.

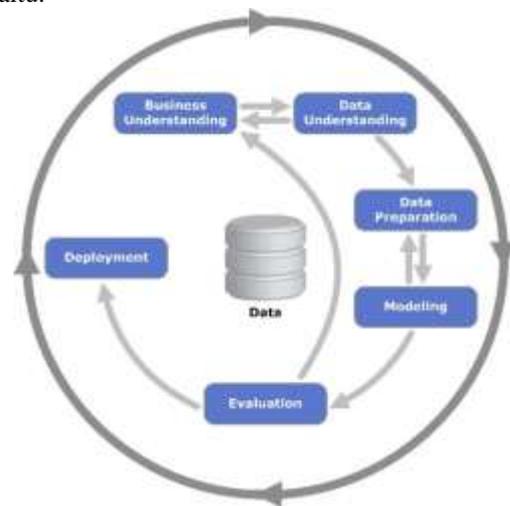
Selanjutnya adalah melakukan penanganan nilai yang hilang, missing value ini bisa terjadi dikarenakan beberapa hal, bisa jadi karena

memang nasabah tidak memilikinya, nasabah lupa mengisikan atau kesalahan inputasi pada sistem database perusahaan.

Untuk beberapa data yang salah satu atributnya tidak terisi akan langsung diganti nilai rata-ratanya bagi atribut yang bersifat numerik (kontinyu) sedangkan untuk atribut yang tergolong kategorikal, nilai yang hilang akan diganti dengan modus dalam atribut tersebut.

H. Langkah-langkah penelitian

Langkah-langkah yang digunakan dalam penelitian ini penulis mengadopsi model CRISP-DM (Cross Standard Industries Process for Data Mining), dimana model ini terdapat 6 tahapan (Chapman et a., 2000) yaitu:



Gambar 2 Model Cross Standard Industries Process for Data Mining

a) Business understanding

Fase awal ini berfokus pada pemahaman tujuan proyek dan persyaratan dari perspektif bisnis, kemudian mengubah pengetahuan ini menjadi definisi masalah penambangan data dan rencana awal yang dirancang untuk mencapai tujuan.

b) Data understanding

Fase pemahaman data dimulai dengan pengumpulan data awal dan berlanjut dengan kegiatan yang memungkinkan Anda untuk menjadi akrab dengan data, mengidentifikasi masalah kualitas data, menemukan wawasan pertama ke dalam data, dan / atau mendeteksi minat berlangganan untuk membentuk hipotesis mengenai informasi tersembunyi.

c) Data Preparation

Tahap persiapan data mencakup semua kegiatan yang diperlukan untuk membangun dataset akhir (data yang akan dimasukkan ke

dalam alat pemodelan) dari data mentah awal. Tugas persiapan data kemungkinan dilakukan beberapa kali dan tidak dalam urutan yang ditentukan. Tugas termasuk tabel, catatan, dan pemilihan atribut, serta transformasi dan pembersihan data untuk alat pemodelan.

d) Modelling

Pada fase ini, berbagai teknik pemodelan dipilih dan diterapkan, dan parameternya dikalibrasi ke nilai optimal. Biasanya, ada beberapa teknik untuk jenis masalah Data Mining yang sama. Oleh karena itu, kembali ke tahap persiapan data seringkali diperlukan.

e) Evaluation

Pada tahap ini dalam proyek, Anda telah membangun sebuah model (atau model) yang tampaknya memiliki kualitas tinggi dari perspektif analisis data. Sebelum melanjutkan ke penyebaran akhir model, penting untuk mengevaluasi secara menyeluruh dan meninjau langkah-langkah yang dijalankan untuk membuatnya, untuk memastikan model dengan benar mencapai tujuan bisnis. Tujuan utamanya adalah untuk menentukan apakah ada beberapa masalah bisnis penting yang belum dipertimbangkan secara memadai. Pada akhir fase ini, keputusan tentang penggunaan hasil penambangan data harus dicapai.

f) Deployment

Penciptaan model pada umumnya bukan akhir dari proyek. Bahkan jika tujuan model adalah untuk meningkatkan pengetahuan tentang data, pengetahuan yang diperoleh akan perlu diatur dan disajikan dengan cara yang dapat digunakan oleh pelanggan. Seringkali melibatkan penerapan model “langsung” dalam proses pengambilan keputusan organisasi – misalnya, personalisasi situs web secara real-time atau penilaian berulang atas basis data pemasaran. Bergantung pada persyaratan, fase penerapan dapat sesederhana menghasilkan laporan atau serumit penerapan proses penambangan data berulang di seluruh perusahaan. Dalam banyak kasus, itu adalah pelanggan, bukan analisis data, yang melakukan langkah-langkah penyebaran. Namun, bahkan jika analisis akan melakukan upaya penyebaran, penting bagi pelanggan untuk memahami terlebih dahulu tindakan apa yang perlu dilakukan untuk benar-benar memanfaatkan model yang dibuat.

I. Pengujian Data

Pada tahap pengujian penulis menggunakan data sample sebanyak 1800

data, data tersebut akan dibagi menjadi dua bagian. Pembagian data terdiri data latih dan data pengujian. Hal ini dilakukan agar lebih efektif untuk belajar dan mengenal pola data di masa lalu, sehingga mampu memberikan hasil prediksi akurasi dengan baik.

Perbandingan metode lain dengan Metode Decision Tree, data yang digunakan adalah data calon nasabah

	Real BUKU	Real BERSIH	Class Precision
pred BUKU	227	60	82.2%
pred BERSIH	41	140	87.1%
Class Recall	87.1%	82.2%	

Gambar 3 Accuracy Naive Bayes 85,97%

Gambar 3 Accuracy Naive Bayes dari hasil akurasi mencapai 85.97%

dengan class recall mencapai 87.11% dan class precision mencapai 82.20%

	Real BUKU	Real BERSIH	Class Precision
pred BUKU	280	100	73.7%
pred BERSIH	21	117	84.6%
Class Recall	88.8%	87.4%	

Gambar 4. Accuracy Tree 73,89%

Gambar 3 Accuracy Naive Bayes dari hasil akurasi mencapai 85.97%

dengan class recall mencapai 87.11% dan class precision mencapai 82.20%.

Dari gambar diatas terlihat hasil perbandingan Naive Bayes dengan Decision Tree lebih unggul 12.08% dengan prediksi memilih, aplikasi yang di gunakan dalam explorasi yaitu RapidMiner.

III.HASIL DAN PEMBAHASAN

A. Data Cleaning

Langkah awal dalam tahapan data mining adalah tahap data cleaning. Tahap ini merupakan suatu proses untuk membersihkan data yang kotor, maksud datayang kotor adalah data yang tidak memiliki nilai, tidak konsisten, tidak relevan atau terdapat kesalahan penulisan saat penginputan. Pembersihan data juga akan mempengaruhi kemampuan dari teknik data mining, karena data yang ditangani akan berkurang jumlah dan kompleksitasnya. Data yang digunakan dalam penelitian ini sudah di-cleaning oleh unit bersangkutan yaitu unit Prodev. Data ini kemudian digunakan pada proses mining dalam penelitian ini.

B. Data Integration

Langkah berikutnya adalah data harus melalui proses integrasi. Data integration merupakan penggabungan data dari berbagai table ke dalam satu table baru. Tidak jarang data yang di perlukan untuk data mining tidak hanya berasal dari satu table, tetapi juga berasal dari beberapa table atau file teks. Pada penelitian ini, tahap integrasi sudah dilakukan pada sumber data yang telah diperoleh.

Data integration dilakukan dengan menggabungkan data dari 6 buah table, untuk table dan query tidak dapat diperlihatkan pada penulisan ini karna berhubungan dengan privasi perusahaan. Data yang diperoleh sebanyak 2520 data yang di ambil untuk produk Optima Saving sendiri.

C. Data Selection

Data Selection merupakan tahap pemilihan data yang relevan yang berada pada table, tidak semua data dipakai, hanya data yang sesuai untuk dianalisis yang akan diambil dari table. Pada penelitian ini setelah tahap data cleaning dan data integration, dilakukan penyeleksian atribut untuk memilih data yang relevan sesuai dengan analisis yang ingin dicapai.

Jumlah keseluruhan atribut yang ada dalam tahap data integration sebanyak 16 atribut, Langkah selanjutnya ialah dilakukan pemilihan terhadap atribut yang relevan dengan target yang ingin dicapai. Atribut yang pilih berdasarkan atribut yang telah diambil oleh unit prodev.

Gambar 5. Data Sebelum Tahap Data Selection

Gambar 6. Data Setelah Tahap Data Selection

D. Pengelompokan dan Analisis Data

Data mining tidak akan pernah dapat dilakukan tanpa tersedianya data yang akan di tambang. Untuk itu dibutuhkan pemahaman akan data yang seperti apa yang akan dipakai dalam proses data mining nanti. Dalam bagian ini akan dibahas mengenai pengolahan data yang diperoleh dari hasil penelitian dilapangan. Data yang berhasil dikumpulkan dari data calon nasabah akan dijadikan sampel dalam penelitian yang akan dijadikan sebagai validasi model sebanyak 2520 record.

Dari 2520 data akan dibagi menjadi sebuah data training, testing sebagai uji coba model melalui beberapa tahapan proses pembentukan model, dan data baru sebagai uji coba model. Sebanyak 720 data diambil untuk prediksi Calon Nasabah yang berminat. Adapun data tersebut terdiri dari 8 atribut. Dimana 7 atribut predictor dan 1 atribut hasil. Penelitian ini bertujuan untuk menentukan akurasi prediksi kelulusan tepat waktu atau yang tidak tepat waktu dengan menggunakan klasifikasi naïve bayes. Selanjutnya algoritma tersebut.

Data yang akan dipakai dalam proses data mining, perlu dipersiapkan dan diubah dalam bentuk model yang lebih sederhana agar bisa dianalisis dengan memakai teknik data mining. Data yang dipakai berupa data nasabah :

1. Jenis Kelamin
2. Status
3. Usia
4. Pekerjaan
5. Pendapatan
6. Lama asuransi
7. Cara bayar

E. Data Set Awal

Dataset awal adalah dataset yang telah dan akan dijadikan obyek penelitian ini.

Data set awal berisikan 8 atribut, yang terdiri dari 7 atribut predictor dan 1 atribut hasil. Masing-masing atribut dan nilai dapat dilihat dalam table dibawah ini.

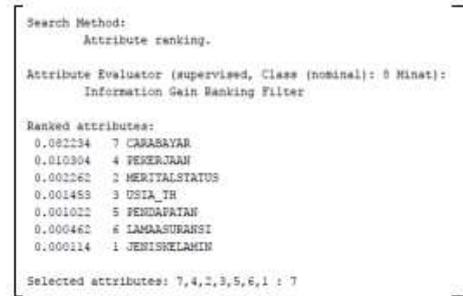
Tabel 2. Data Variabel

No	Atribut	Nilai
1	Jenis Kelamin	Pria
		Wanita
2	Status	Lajang
		Kawin
		Duda
		Janda
		Cerari
3	Usia	Usia < 30
		Usia >= 30 <=50
		Usia > 50
4	Pekerjaan	Pegawai swasta
		Direktur

No	Atribut	Nilai
		Lainnya
		PNS
		Pegawai BUMN
		Dokter
		TNI / Polri
		Wiraswasta
		Dosen
		Pensiunan
		Ibu rumah tangga
		Notaris
		Guru
		Pegawai honorer
		Pengacara
		Paramedis
		Mahasiswa
		Kontraktor
Pendeta		
5	Pendapatan	Uang < 300 Jt
		Uang >=300 <=600 Jt
		Uang > 600 Jt
6	Lama Asuransi	1 Tahun
		2 Tahun
		3 Tahun
		4 Tahun
7	Cara Bayar	Bulanan
		Semesteran
		Tahunan
		Semesteran

F. Uji Coba Data Sample

Pada pengujian data sample yang berisikan 1800 data dengan 7 atribut predictor dan 1 atribut hasil, dengan menggunakan algoritma Selector Gain Ratio Attribute melalui tools WEKA ranking teratas, yaitu:



Gambar 7. Output Atribut Selector dari Tools WEKA

Atribut yang dipilih disajikan dalam bentuk tabel berikut ini :

Tabel 3. Rangking

No	Ranking	Atribut
1	0.082234	Cara Bayar
2	0.010304	Pekerjaan
3	0.002262	Merital Status
4	0.001453	Usia
5	0.001022	Pendapatan
6	0.000462	Lama Asuransi
7	0000114	Jenis Kelamin

G. Pemodelan dengan Algoritma Naive Bayes

Penggunaan algoritma Naive Bayes menggunakan data training dimulai dengan melakukan perhitungan probabilitas prior untuk mengetahui nilai yang Memilih dan Tidak untuk semua record.

Proses untuk melakukan perhitungan minat nasabah dalam memilih asuransi menggunakan data yang telah dipilih, Langkah pertama mencari nilai prior probabilitas untuk semua kasus Memilih dan Tidak. Jumlah data memilih sebanyak 963 dan jumlah data yang tidak memilih sebanyak 837 sehingga jumlah data yang diipakai adalah sebanyak 1800 data.

$$P(\text{Memilih}) = \frac{963}{1800} = 0.5350$$

Nilai prior probabilitas untuk memilih didapatkan jumlah data memilih yaitu sebanyak 963 dibagi jumlah data keseluruhan yaitu 1800 maka akan didapatnka angka yaitu 0.5350.

$$P(\text{Tidak}) = \frac{837}{1800} = 0.4650$$

Nilai prior probabilitas untuk tidak memilih didapatkan jumlah data tidak memilih yaitu sebanyak 837 dibagi jumlah data keseluruhan yaitu 1800 maka akan didapatkan angka yaitu 0.4650.

Tabel 4. Probabilitas Prior Data

Atribut	Nilai	Jumlah Data	P(x c1)		P(x c2)		
			Memilih	Tidak	Memilih	Tidak	
Total		1800	963	837	0.5350	0.4650	
Jenis Kelamin	Pria	789	417	352	0.4330	0.4205	
	Wanita	1031	546	485	0.5670	0.5795	
Status	Janda	69	39	30	0.0405	0.0358	
	Duda	9	3	6	0.0031	0.0072	
	Kawin	1536	832	704	0.8640	0.8411	
	Cerai	4	3	1	0.0031	0.0012	
	Lajang	182	86	96	0.0893	0.1147	
Usia	Usia <30	138	64	74	0.0665	0.0884	
	Usia >=30 <=50	999	548	451	0.5691	0.5388	
	Usia >50	663	351	312	0.3645	0.3728	
Pekerjaan	Pegawai swasta	1066	575	491	0.5971	0.5866	
	Direktur	10	7	3	0.0073	0.0036	
	Lainnya	94	39	55	0.0405	0.0657	
	Pns	215	122	93	0.1267	0.1111	
	Pegawai BUMN	118	67	51	0.0696	0.0609	
	Dokter	28	11	17	0.0114	0.0203	
	Tni / polri	21	12	9	0.0125	0.0108	
	Wiraswasta	96	49	47	0.0509	0.0562	
	Dosen	24	12	12	0.0125	0.0143	
	Pensiunan	29	17	12	0.0177	0.0143	
Ibu rumah tangga		70	33	37	0.0343	0.0442	
	Notaris	4	2	2	0.0021	0.0024	
Atribut	Nilai	Jumlah	Memilih	Tidak	P(x c1)		
	Guru	8	5	3	0.0052	0.0036	
	Pegawai honorer	2	1	1	0.0012	0.0012	
	Pengacara	2	2	0	0.0021	0.0000	
	Paramedis	1	1	0	0.0012	0.0000	
	Mahasiswa	5	2	3	0.0021	0.0036	
	Kontraktor	1	1	0	0.0012	0.0000	
	Pendeta	2	2	0	0.0021	0.0000	
	Pelaut	1	0	1	0.0000	0.0012	
	Bankir	1	1	0	0.0012	0.0000	
	Petani	2	2	0	0.0021	0.0000	
	Pendapatan	Uang <300	875	485	390	0.5036	0.4659
		Uang >=300 <=600	796	411	385	0.4268	0.4600
		Uang >600	129	67	62	0.0696	0.0741
LamaAsuransi	1 Tahun	138	77	61	0.0800	0.0729	
	2 Tahun	345	191	154	0.1983	0.1840	
	3 Tahun	439	229	210	0.2378	0.2509	
	4 Tahun	878	466	412	0.4839	0.4922	
Cara Bayar	Tahunan	524	302	222	0.3136	0.2652	
	Sekaligus	752	336	386	0.3801	0.4612	
	Bulanan	136	136	0	0.1412	0	
	Semesteran	388	139	229	0.1651	0.2736	

dan empat. Dari table terdapat beberapa langkah untuk menghitung yaitu :

Tabel 5. Probabilitas Posterior Data X

Data X	Nilai	P(x c1)	
		Memilih	Tidak
Jenis Kelamin	Pria	0.4330	0.4205
Status	Janda	0.0405	0.0358
Usia	Usia <30	0.0665	0.0884
Pekerjaan	Pegawai swasta	0.5971	0.5866
Pendapatan	Uang <300	0.5036	0.4659
LamaAsuransi	1 Tahun	0.0800	0.0729
Cara Bayar	Tahunan	0.3136	0.2652

Dari table diatas terdapat beberapa langkah untuk menghitung, yaitu :

Tabel 6. Tabel Perhitungan P(X | Ci)P(Ci)

Urutan	Rumus	Hasil
a.	$P(X C_1)$	P(remark = Memilih) $= 0.4330 * 0.0405 * 0.0665 * 0.5971 * 0.5036 * 0.0800 * 0.3136$ $= 0.00000879758$
b.	$P(X C_2)$	P(remark = Tidak) $= 0.4205 * 0.0358 * 0.0884 * 0.5866 * 0.4659 * 0.0729 * 0.2652$ $= 0.00000703132$
c.	$P(X C_1)P(C_1)$	P(X, remark=Memilih)P(remark=Memilih) $= 0.5350 * 0.00000879758$ $= 0.0000047057053$
d.	$P(X C_2)P(C_2)$	P(X, remark=Tidak)P(remark=Tidak) $= 0.4650 * 0.00000703132$ $= 0.0000032695638$

Dari hasil perhitungan diatas, didapat nilai $P(X|Ci) P(Ci)$ lebih besar untuk remark = Memilih sehingga dapat disimpulkan bahwa data training tersebut “Memilih”.



Gambar 8. Halaman Add Data Rapid Miner

Untuk menentukan kasus baru termasuk Memilih atau Tidak Memilih, dilakukan perhitungan probabilitas pada Tabel 4.3 Perhitungan probabilitas posterior untuk menentukan data testing termasuk klasifikasi mana terdapat disajikan pada Tabel 4.4 Misalkan diambil sebuah data testing X dengan nilai seperti pada Tabel 4.3 kolom dua untuk menentukan kelas mana, dilakukan perhitungan probabilitas posterior yang hasilnya terdapat pada Tabel 4.4 kolom tiga

Gambar 5 merupakan gambar dari tools Rapid Miner untuk add data atau untuk mencari dataset yang akan diupload untuk di proses, caranya adalah dengan Klik tombol import data lalu mencari dataset yang akan digunakan.



Gambar 9. Halaman untuk mencari dataset

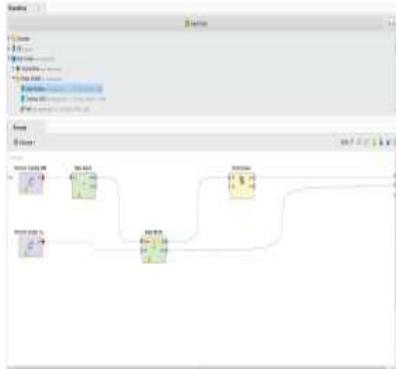
Gambar 9. merupakan tampilan untuk mencari dataset yang akan kita gunakan, jika memilih My Computer maka akan mengarah ke local penyimpanan,

jika memilih database maka akan langsung mengarah ke database yang ada di server.



Gambar 10. Pemilihan Type data

Pada Gambar 10. merupakan tampilan pada saat import data dan memilih type data, ada beberapa type data yang dapat di pilih diantaranya : Polynominal, Binominal, Real, Integer, Date_Time, Date, Time.



Gambar 11. Proses Data Rapid Miner

Pada gambar 11. merupakan tampilan proses data yang ada pada rapid miner, dengan drag and drop data training dan data testing, lalu memilih metode yang naïve bayes dengan dihubungkan dari data training ke metode naïve bayes, memilih operator Apply Modal untuk menghubungkan data training dan data testing, ditambahkan dengan operator Performance Proses data rapid miner sudah bisa dijalankan.

H. Evaluasi dan Validasi

Hasil dari pengujian metode yang telah dilakukan yaitu dengan algoritma Naïve Bayes, dilakukan pengujian tingkat akurasi dengan menggunakan bantuan software Rapid Miner untuk mencari confusion matrix dan kurva ROC/AUC (Area Under Cove). Setelah itu dilakukan pengujian hasil dari pemilihan atribut dengan menggunakan algoritam Gain Ration Attribute Evaluator dari tolls Weka. Pemodelan dengan Algoritma Naïve Bayes

a) Confusion Matrix Data Set Original

Pengujian confusion matrix untuk dataset yang diolah menggunakan algoritma Naïve Bayes dengan 1800 data Training dan 720 data data Testing dilihat pada tabel dibawah ini:

Tabel 7 Confusion Matrix Data Trainigng Naïve Bayes

	True Memilih	True Tidak	Class Precision
Pred. Memilih	277	60	82.20%
Pred. Tidak	41	342	89.30%
Class recall	87.11%	85.07%	

Perhitungan dari tabel diatas adalah sebagai berikut :

1. Nilai presisi (menunjukkan kualitas) dari kelas Memilih

$$precision = \frac{TP}{TP + FP} = \frac{277}{277 + 60} = 82.20\%$$

2. Nilai recall (menunjukkan kuantitas) dari kelas Memilih

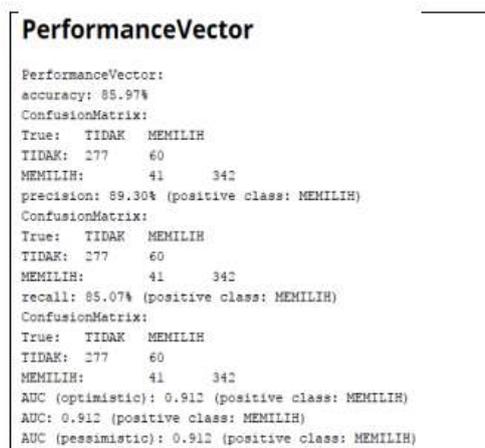
$$recall = \frac{TP}{TP + FN} = \frac{277}{277 + 41} = 87.11\%$$

3. Nilai akurasi yang didapat secara keseluruhan

$$recall = \frac{TP + TN}{TP + TN + FP + FN} = \frac{277 + 342}{277 + 342 + 60 + 41} = 85.97\%$$

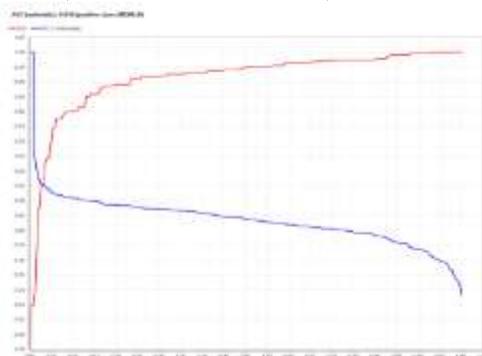
b) Performance Vector Data Set Original

Dari pengujian model yang telah dilakukan dengan menggunakan software rapid miner, didapatkan performance vector dan kuva ROC sebagai berikut :



Gambar 12 Performance Vector

Dari Gambar 12 Performance Vector hasil akurasi mencapai 85.97% dengan Precision 89.30% (Positive class : Memilih) dan Recall 85.07% (Positive class : Memilih)



Gambar 13. Kurva ROC

IV. KESIMPULAN

Pada penelitian ini bahwa data yang di gunakan untuk prediksi minat nasabah terhadap produk asuransi meninggal dunia (Optima Saving) dapat diproses menggunakan metode Naïve Bayes, hasil akurasi mencapai 85.97%. Penggunaan yang tepat ketika menerapkan algoritma naïve bayes dapat meningkatkan akurasi yang lebih baik dalam memprediksi minat nasabah tersebut, terutama dalam menyimpulkan informasi yang akan diambil oleh unit terkait dalam melakukan pemasaran. Pengembangan akan dilakukan dengan menggunakan aplikasi pendukung seperti matlab, RapidMiner sebagai prototype dari perhitungan Naïve Bayes yang nantinya diterapkan pada sistem operational objek penelitian.

UCAPAN TERIMA KASIH

Terima kasih untuk semua pihak yang telah membantu untuk pembuatan jurnal ini.

REFERENSI

- [1] Danarti, Dessy. (2011). *Jurus Pintar Asuransi Ed 1*, Yogyakarta: G-Media
- [2] Lailatul M. Chaira , Nurul Hidayat , Inu L. Wibowo , Imam Mukhlash. (2016) “Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Dengan Metode Naïve Bayes Classifier,” ISSN: 1829-605X Vol. 13, No. 2, Nopember 2016, 11-20
- [3] Han, J., & Kamber, M, *Data Mining: Concepts and Techniques*, 3rd Edition. Morgan Kauffman Publisher, San Fransisco (2012).
- [4] Bustami, B. Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI. Jurnal Penelitian Teknik Informatika*, Vol: 3 No. 2 (2013) 127-146.
- [5] Betrisandi (2017) “Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination,” ISSN print 2087-1716, ISSN online 2548-7779 . *Jurnal Ilmiah Volume 9 Nomor 1 April 2017*.
- [6] Taufiq Hidayat, Sandy Kurniawan (2007) “Penerapan Data Mining Dengan Metode Interpolasi Untuk Memprediksi Minat Konsumen Asuransi (Studi Kasus Asuransi Metlife),” 113-128 ISSN: 0854-4743, Vol. 5, No. 2, Desember 2007
- [7] Hendra Irawan , Andy Victor Pakpahan (2017) “Penerapan Algoritma Naive Bayes Untuk Menentukan Nasabah Potensial Pada Ajb Bumiputera 1912,” Vol.10 No.2, Desember 2017
- [8] Kurniawa, A. Haidar Mirza , Rendra Octa Chandra(2013) “Penerapan Data Mining Untuk Memprediksi Minat Nasabah Pada Ajb Bumiputera 1912 Palembang” Vol.x No.x, April 2013:1-10
- [9] Bakar, A. A., Othman, Z., Yusoff M. S. N. M., & Ismail, R, “Development of Knowledge Model for Insurance Product Decision using the Associative Classification Approach,” in *10th International on Intelgent System Design and Application, Malaysia (2010)*.
- [10] S. K. Srivasta , S. Balaji (2012) “Naive Bayes Classification of Public Health Data with Greedy Feature Selection,” 2013 Volume 13 Issue 2, *Communications of the IIMA ©2013*
- [11] Turban, Efraim, Jay E Aronson, and Ting-Peng Liang. 2007. *Decision Support Systems and Intelligent Systems*.

- Decision Support Systems and Intelligent Systems. Vol. 7. <http://www.amazon.co.uk/dp/0131230131>.
- [12] Anjelisa, Lisbeth Mananeke, Mirah Rogi. 2018. Analisis Pengaruh Strategi Segmentasi, Targeting Dan Positioning (Stp) Terhadap Keputusan Pembelian Produk Bp-Smart Protection Di Ajb Bumiputera 1912 Cabang Sam Ratulangi Manado.
- [13] Informasi Keuangan PT. Bni Life <https://www.bni-life.co.id/id/laporan-keuangan> (diakses pada 01-Mei-2019 10:33)
- [14] Informasi Produk Asuransi <https://www.bni-life.co.id/id/product/bisnis/jiwa/blifeoptima-executive-saving> (diakses pada 10-Mei-2019 14:33).