

Klasifikasi Minat Nasabah Terhadap Penjualan Asuransi Kesehatan dengan Metode Naïve Bayes Clasifier

Yonan Mulyana^{1,*}, Puspita Nurul Sabrina², Herdi Ashaury³

^{1,2,3}Informatika, Sains dan Informatika, Universitas Jenderal Achmad Yani,
Cimahi Indonesia, Kode Pos 40513

[1yonanmulyana18@if.unjani.ac.id](mailto:yonanmulyana18@if.unjani.ac.id), [2puspita.sabrina@lecture.unjani.ac.id](mailto:puspita.sabrina@lecture.unjani.ac.id)

[3herdi.ashaury@lecture.unjani.ac.id](mailto:herdi.ashaury@lecture.unjani.ac.id)

Abstrak

Dalam dunia perbankan terdapat produk yang di tawarkan kepada nasabahnya salah satunya yaitu asuransi, asuransi yang ditawarkan kepada nasabah yaitu asuransi kesehatan. Asuransi kesehatan merupakan asuransi yang memberikan santunan kesehatan kepada seseorang (tertanggung) yang berupa biaya perawatan dan pengobatan jikalau, seseorang (tertanggung) tersebut diserang penyakit kritis. Asuransi kesehatan akan memberi perlindungan ketika seseorang (tertanggung) tersebut mengalami risiko kesehatan di kemudian hari. Perusahaan asuransi menyimpan data nasabah, data tersebut sangat penting bagi perusahaan perbankan untuk mengetahui kriteria nasabah yang berminat terhadap asuransi yang tawarkan. Dengan terdapatnya informasi dari data nasabah yang ada, data tersebut apabila digali dengan tepat dapat di ketahui pola untuk dilakukan data mining dengan metode klasifikasi, Data mining adalah teknik yang dapat membantu bisnis untuk menemukan sesuatu yang sangat penting dari sekumpulan data. Dengan melakukan klasifikasi menggunakan algoritma Naive Bayes classifier diharapkan minat nasabah dapat diprediksi dengan mengklasifikasikan data nasabah serta dilakukan eksperimen dan analisa untuk peningkatan akurasi naïve bayes dengan menerapkan metode feature selection. Hasil dari penelitian ini dapat disimpulkan bahwa memiliki ketepatan kinerja yang kurang baik dalam pengklasifikasikan data minat nasabah dikarenakan tingkat akurasi yang baik di dapatkan dari tingkat akurasi 80% ke atas.

Kata Kunci: *Data Mining, Asuransi, Klasifikasi, Naïve Bayes, feature selection.*

A. Pendahuluan

Kesehatan merupakan hak bagi setiap umat manusia. Manusia umumnya ingin menjaga tubuhnya tetap sehat. Namun, risiko sakit tidak ada yang tahu kapan dan kepada siapa hal itu akan terjadi. Oleh karena itu, untuk mengatasi risiko penyakit Ini maka hal tersebut perlu dipersiapkan dengan mendaftarkan diri ke asuransi kesehatan.

Asuransi kesehatan merupakan asuransi yang memberikan santunan kesehatan kepada seseorang (tertanggung) yang berupa biaya perawatan dan pengobatan jikalau, seseorang (tertanggung) tersebut diserang penyakit kritis. Asuransi kesehatan akan memberi perlindungan ketika seseorang (tertanggung) tersebut mengalami risiko kesehatan di kemudian hari.

Salah satu faktor yang menentukan keberhasilan target penjualan produk asuransi adalah dalam pengelolaan segala jenis sumber daya yang berhubungan dengan tenaga penjualan yang akan mempengaruhi dan mendukung keberhasilan perusahaan dalam pencapaian tujuannya. Perusahaan asuransi perlu mendeteksi kriteria nasabah yang memiliki

keterminatan terhadap minat atau tidak nya pada produk asuransi yang di tawarkan tiap tahunnya yang relatif rendah karna adanya pesaing dari instansi lain.

Database perusahaan asuransi menyimpan data nasabah dan data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan untuk dilakukan klasifikasi. Serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining. Klasifikasi adalah sebuah metode untuk mengelompokan data. Klasifikasi merupakan teknik *data mining* yang melihat atribut dari kelompok data yang sudah didefinisikan sebelumnya. Atribut-atribut ini digunakan sebagai variabel dalam penentuan kelas suatu objek baru. Tujuan klasifikasi yaitu untuk menentukan kelas dari suatu objek yang belum diketahui kelasnya dengan akurat.

“Berdasar penelitian sebelumnya algoritma klasifikasi Data Mining tentang klasifikasi minat nasabah terhadap asuransi salah satunya menggunakan Naïve bayes. Metode Naïve Bayes dipilih karena metode ini memperhatikan seluruh fitur pada data latih sehingga membuat metode ini optimal dalam melakukan proses klasifikasi”.

“Metode Naïve Bayes merupakan metode penggolongan berdasarkan propabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung” (Zulkifli.2017)

Pada penelitian kali ini selain bereksperimen terhadap peningkatan akurasi juga bertujuan mendapatkan model atribut yang berpengaruh dengan cara menerapkan *Feature Selection* dikarnakan akurasi yang di dapatkan oleh algoritma naïve bayes masih dapat ditingkatkan dengan menerapkan metode *feature selection*. *Feature Selection* adalah salah satu cara untuk menentukan atribut yang paling berpengaruh di dalam dataset. *Feature Selection* berperan memilih subset yang tepat dari set fitur asli, karena tidak semua fitur/atribut relevan dengan masalah. Berdasarkan latar belakang yang telah dijabarkan di atas, penulis akan mengkaji Klasifikasi Minat Nasabah Terhadap Penjualan Asuransi Kesehatan dengan Metode Naïve Bayes Clasifier dan bereksperimen terhadap peningkatan akurasi yang bertujuan untuk mendapatkan model atribut yang berpengaruh terhadap peningkatan akurasi dengan cara menerapkan metode *Feature Selection*.

B. Metode

Metode Penelitian dalam penelitian ini adalah pendekatan deskriptif analitis. Metode teknik ini berfungsi untuk menggambarkan tentang suatu objek yang di teliti melalui kesimpulan yang bersifat umum. Bisa di bilang penelitian deskriptif analitis ini mengambil masalah atau memusatkan perhatian terhadap masalah - masalah yang sebagaimana adanya saat dilakukanya penelitian, hasil penelitian tersebut kemudian diolah dan dianalisa untuk diambil kesimpulannya.

“Penelitian ini menggunakan metode waterfall yang pertama kali dikembangkan oleh Winston Royce pada tahun 1970. Metode tersebut memiliki pendekatan yang sistematis dengan menerapkan daur hidup dalam pengembangan sistem perangkat lunaknya”.(Maliki and Dangkoa 2019).

Model pengembangan ini bersifat linier dari tahap awal pengembangan sistem yaitu tahap perencanaan hingga tahap terakhir pengembangan sistem yaitu tahap pemeliharaan.

Tahap selanjutnya tidak dilanjutkan sampai tahap sebelumnya selesai, dan tahap sebelumnya tidak dapat dikembalikan atau diulang.

1.1 Metode Pengumpulan Data

Metode yang digunakan dalam pengumpulan data yaitu:

- a. Studi Pustaka untuk mengumpulkan informasi dari penelitian sebelumnya dalam bentuk buku atau jurnal dengan sumber tertulis yang berhubungan dengan asuransi, data mining, dan Naive Bayes. Kemudian mengevaluasi dan mengumpulkan data yang diperlukan berdasarkan topik yang diambil, yang dapat digunakan sebagai referensi atau untuk memecahkan masalah dalam penelitian ini yang melibatkan penggunaan metode Naive Bayesian untuk klasifikasi minat nasabah.
- b. Observasi, dilakukan dengan mengamati metode dan kriteria yang digunakan untuk peminatan nasabah dilakukan dan mengetahui tujuan penelitian sehingga hasil akhir dapat dibandingkan dengan hasil yang diperoleh setelah kemudian mendokumentasikan hasil kriteria yang digunakan untuk peminatan nasabah secara rinci, yang telah diidentifikasi sebagai data yang akan diproses dalam metode Naive Bayes.

1.2 Tahap Presprocessing Data

“Pada tahap ini merupakan langkah-langkah yang dilakukan sebelum langkah inti dilakukan, yang dimana sebagian besar data sebelumnya ialah data mentah, atau data yang tidak lengkap, dengan adanya noise atau tidak konsisten. Oleh karena itu, tahap preprocessing data sangat penting, dikarenakan suatu data yang tidak berkualitas akan menghasilkan kualitas data mining yang tidak baik, berikut 3 teknik dalam preprocessing data yang dilakukan” :

- a. Data Cleanning
Pada teknik ini dilakukan pembersihan data yang kosong atau hilang, serta data yang mengandung kesalahan.
- b. Data Integration
Teknik ini merupakan proses menggabungkan atau menyatukan dua atau lebih sebuah data dari berbagai sumber database yang berbeda ke dalam sebuah penyimpanan seperti gudang data
- c. Data Transformation
Pada teknik ini dilakukan proses perubahan data menjadi data dalam bentuk yang sesuai.

1.3 Klasifikasi menggunakan algoritma Naïve Bayes Classifier

Pada tahap ini dilakukan pengklasifikasian dengan algoritma Naïve Bayes Classifier pada data set Jenis Kelamin, Status, Usia, Pekerjaan, Pendapatan/Tahun, Produk, Lama Asuransi, Premi dan Hasil. Dengan label Minat dan Tidak Minat.

1.4 Eksperimen Untuk Mendapatkan Akurasi Terbaik

Untuk mendapatkan akurasi terbaik dengan mencoba bereksperimen terhadap jumlah record data set dengan memisahkan data mana saja yang termasuk kedalam data uji dan data latih, lalu selanjutnya bereksperimen terhadap atribut mana saja yang di gunakan dan mana saja yang dihilangkan.

1.5 Analisa faktor penentu akurasi terbaik

Pada tahap ini dilakukan Analisa terhadap atribut – atribut mana yang paling mempengaruhi terhadap hasil akurasi akhir sehingga akurasi terbaik bisa di dapatkan.

C. Hasil dan Pembahasan

Implementasi

Tahap Implementasi merupakan penerapan algoritma terhadap mesin dari semua analisis dan perancangan sistem yang telah disusun secara fungsional, maka dilakukan implementasi sistem kedalam bentuk program python dengan cara menerjemahkan hasil perancangan secara rinci kedalam bahasa pemrograman.

Persiapan Implementasi

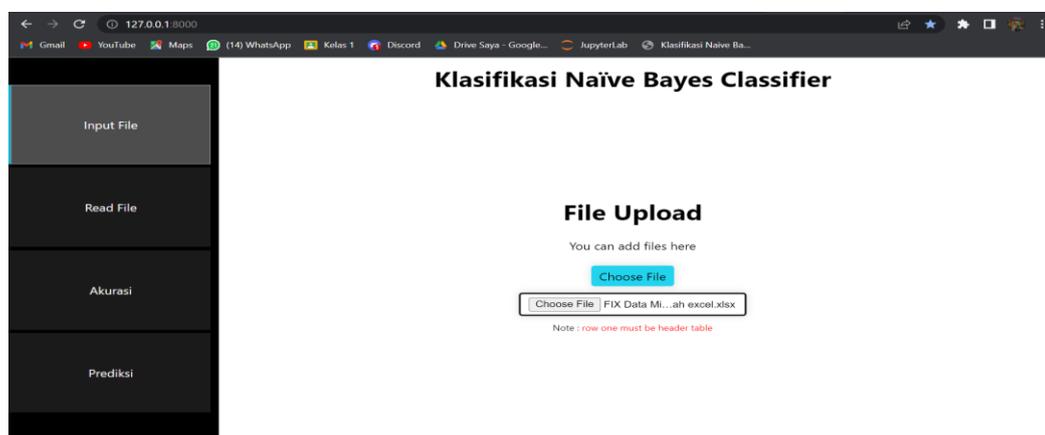
Sebelum menjalankan sistem ada hal yang harus diperhatikan yaitu kebutuhan sistem. Dalam penelitian ini system yang di gunakan adalah mesin pemrograman dengan bahasa python dengan menggunakan *jupyter lab*, berdasarkan kebutuhan dengan menggunakan bahasa pemrograman tersebut *Google Chrome* digunakan sebagai media *web browser*. adapunn langkah yang perlu di siapkan untuk memulai menjalankan Bahasa pemrograman python, yaitu seperti instalasi library dan lain sebagainya seperti :

- pip jupyterlab
- pip install pandas
- pip install -U scikit-learn
- pip install numpy
- pip install matplotlib
- pip install seaborn
- pip install pyplot
- pip install scipy

Proses Implementasi Antarmuka

a. Input File

Hal pertama yang dilakukan adalah dengan menginput data file yang berformat .xlsx ke dalam aplikasi contohnya seperti pada gambar 1 di bawah ini.



Gambar 1. Implementasi Halaman Input Data

b. Read File

Selanjutnya data di di baca oleh aplikasi dan di ditampilkan. Untuk selengkapnya dapat di lihat pada gambar 2 di bawah ini.

Jenis kelamin	Status	Usia	Pekerjaan	Pendapatan pertahun	Produk	Label
2	1	1	10	1	1	0
1	1	2	1	2	1	0
1	1	1	1	2	1	0
2	2	2	1	2	1	1
1	1	1	1	2	1	0
1	1	2	1	2	1	0
1	2	2	1	2	2	1
1	1	2	10	1	2	1
2	2	2	8	1	1	0
2	1	2	10	1	1	0
1	1	2	6	3	1	1
2	1	1	6	3	1	0

Gambar 2. Implementasi Halaman Read File

c. Akurasi

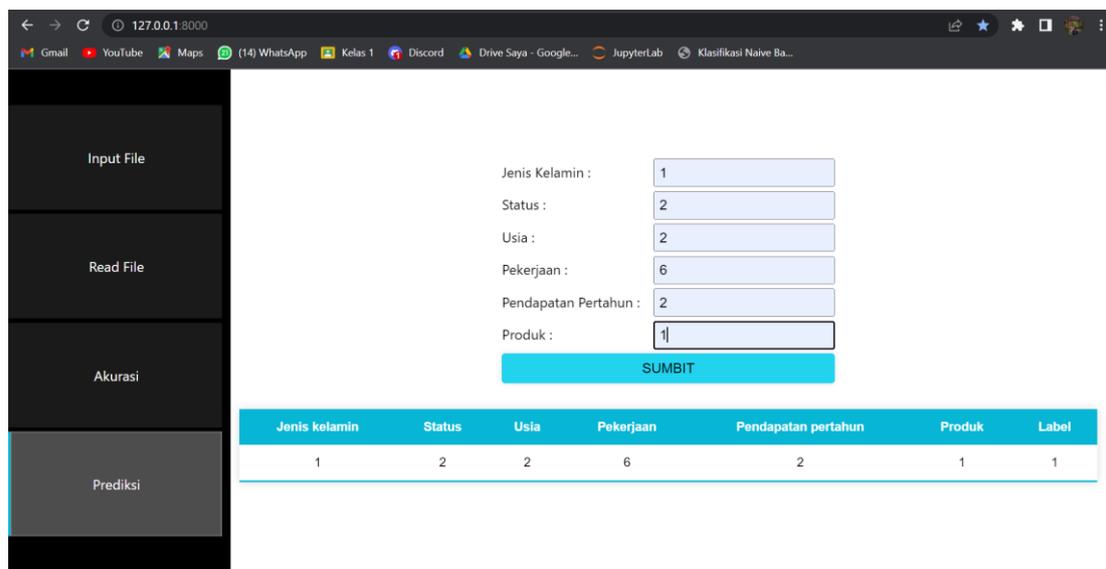
Selanjutnya implementasi halaman akurasi untuk menampilkan hasil eksperimen terhadap klasifikasi data dengan membandingkan akurasi dari beberapa atribut yang di pilih dengan menggunakan *feature selection* dan yang tidak menggunakan *feature selection*. Selengkapnya dapat dilihat pada gambar 3 di bawah ini.

<p>Model Feature Selection Support : [True True False False True False] Ranking : [1 1 4 3 1 2]</p> <p>Selection Feature Dimensi dari X_train : 764 Dimensi dari Y_train : 764 Dimensi dari X_test : 85 Dimensi dari Y_test : 85 Accuracy Score : 0.7411764705882353</p>	<p>Not Selection Feature Dimensi dari X_train : 764 Dimensi dari Y_train : 764 Dimensi dari X_test : 85 Dimensi dari Y_test : 85 Accuracy Score : 0.7176470588235294</p>
--	--

Gambar 3. Implementasi Halaman Akurasi

d. Prediksi

Selanjutnya implementasi halaman prediksi untuk menampilkan hasil prediksi dari inputan yang telah di masukan oleh user, selengkapnya seperti pada gambar 4 di bawah ini.



Jenis kelamin	Status	Usia	Pekerjaan	Pendapatan tahunan	Produk	Label
1	2	2	6	2	1	1

Gambar 4 Implementasi Halaman Prediksi

Berdasarkan dari kedua hasil eksperimen yang telah dilakukan dimana untuk eksperimen pertama hasil akurasi yang di dapatkan adalah 0.7176 dengan menggunakan metode *non feature selection* dengan memakai ke enam atribut yaitu atribut “jenis_kelamin”, “status”, “usia”, “pekerjaan”, “pendapatan_pertahun”, dan “produk”. sedangkan

Dengan menggunakan metode *feature selection* yang dapat berguna untuk membuang data yang tidak ada korelasinya, sehingga akan meningkatkan skor akurasi data dalam melakukan prediksi. Dan hasil nya terdapat support dan ranking seperti pada gambar 5.5 di bawah ini.

```
Support = [ True True False False True False]
Ranking = [1 1 4 3 1 2]

x1= df.loc[:, [ 'jenis_kelamin', 'status', 'pendapatan_pertahun' ]]
```

Gambar 5. 5 Atribut Feature Selection

Dimana rangking yang bernilai 1 bersifat true dan rangking 2,3,4 bersifat false sehingga atribut yang dapat mempengaruhi tingkat akurasi yaitu atribut “jenis_kelamin”, “status”, dan “pendapatan_pertahun”. Dan hasil akurasi yang di dapatkan dari metode ini adalah 0.7412.

Dapat diambil informasi bahwa secara keseluruhan aplikasi klasifikasi *naïve bayes* pada pemrograman mesin python dengan menggunakan *feature selection* memiliki kecenderungan tingkat akurasi yang lebih tinggi yang dimana hasil tingkat akurasi lebih

besar didapatkan pada eksperimen ke dua dengan atribut yang di pilih adalah atribut "jenis_kelamin", "status", dan "pendapatan_pertahun".

D. Kesimpulan

Penelitian ini akan melakukan analisa terhadap klasifikasi data minat nasabah dengan menggunakan metode naïve bayes dan untuk mengetahui atribut mana yang sangat berpengaruh terhadap peningkatan akurasi skor yang akan di dapatkan. Berdasarkan hasil klasifikasi data minat nasabah dengan menggunakan dataset sebanyak 849 data dengan di bagi menjadi 2 yaitu data latih sebanyak 764 data dan data uji sebanyak 85 data, dengan perbandingan menggunakan metode feature selection dan non feature selection. Untuk yang menggunakan feature selection dengan atribut yang di pilih yaitu atribut "jenis_kelamin", "status", dan "pendapatan_pertahun", menghasilkan tingkat akurasi yang lebih tinggi terhadap metode Naive Bayes sebesar 74,12%, di bandingkan dengan metode non feature selection dengan atribut yang di pakai adalah atribut "jenis_kelamin", "status", "usia", "pekerjaan", "pendapatan_pertahun", dan "produk" yang menghasilkan tingkat akurasi 0.7176.

Hasil dari penelitian ini dapat disimpulkan bahwa pengklasifikasian menggunakan naïve bayes pada penelitian ini memiliki ketepatan kinerja yang kurang baik dalam pengklasifikasikan data minat nasabah dikarenakan tingkat akurasi yang baik di dapatkan dari tingkat akurasi 80% ke atas.

Daftar Pustaka

- Anita, S. (2015). Data mining memprediksi minat masyarakat terhadap asuransi jiwa dengan metode interpolasi linier. JURIKOM (Jurnal Riset Komputer), 2(6), 62-65.
- 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.
- Han, J., & Kamber, M, Data Mining: Concepts and Techniques, 3rd Edition. Morgan Kauffman Publisher, San Fransisco (2012).
- Ian H. Witten, frank Eibe, and Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Asma Stephan and Burlington, Eds. United States of America: Morgan Kaufmann, 2011.
- Maimoon, Oded, and Lior Rokach. 2010. Data Mining and Knowledge Discovery Handbook. Vol. 40. doi:10.1002/1521-3773(20010316)40:63.3.CO;2-C.

- Maliki, Ona, and Fandi Dangkoa. 2019. "Sistem Pakar Tipe Perumahan Menggunakan Metode Forward Chaining." *Jurnal Informatika Upgris4(2)*: 150–57.
- Srimuddawamah, Ika. 2015. Sistem Pakar Diagnosa Penyakit Kulit Pada Anak Menggunakan Metode Naïve Bayes. Program Teknologi dan Ilmu Komputer. Universitas Brawijaya, Malang.
- Sugiyono. (2011). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Teas, R. Kenneth et al. "A Path Analysis of Causes and Consequences of Salespeople's Perceptions of Role Clarity." *Journal of Marketing Research* 16 (1979): 355 - 369.
- Zulkifli, S. (2017). Decision Support System Pemberian Bonus Tahunan Pada Karyawan Berdasarkan Kinerja Karyawan Menggunakan Metode Simple Additive Weighting (Study Kasus: Stmik Pringsewu). *Jurnal TAM (Technology Acceptance Model)*, 7(0), 67–73. <http://ojs.stmikpringsewu.ac.id/index.php/JurnalTam/article/view/74/74>