

## Perbandingan Algoritma Klasifikasi K-Nearest Neighbor, Random Forest dan Gradient Boosting untuk Memprediksi Ketertarikan Nasabah pada Polis Asuransi Kendaraan

Sri Diantika<sup>1</sup>, Agus Subekti<sup>2</sup>, Hiya Nalatisifa<sup>3</sup>, Mareanus Lase<sup>4</sup>

<sup>1,2,3</sup> Program Studi Magister Ilmu Komputer, Universitas Nusa Mandiri, Jalan Margonda Raya No. 545, Pondok Cina, Depok, Jawa Barat 16424, (021) 31908575  
e-mail: <sup>1</sup>14002378@nusamandiri.ac.id, <sup>2</sup>agus@nusamandiri.ac.id, <sup>3</sup>14002363@nusamandiri.ac.id

<sup>4</sup> Program Studi Teknik Informatika, STMIK Nusa Mandiri, Jalan Margonda Raya No. 545, Pondok Cina, Depok, Jawa Barat 16424, (021) 31908575  
e-mail: <sup>4</sup>mareanus.mle@nusamandiri.ac.id

Submitted Date: February 04<sup>th</sup>, 2021  
Revised Date: September 23<sup>rd</sup>, 2021

Reviewed Date: June 01<sup>st</sup>, 2021  
Accepted Date: October 12<sup>th</sup>, 2021

### Abstract

An insurance policy provides coverage for compensation for specified loss, damage, illness, or death in exchange for premium payments. Likewise for vehicle insurance, every year the customer needs to pay a premium to the insurance company so that if an accident occurs that is not profitable for the vehicle, the insurance company provides compensation to the customer. The purpose of this research is to classify the health insurance cross-sell prediction dataset so that certain patterns or relationships can be found between the data to become valuable information and build a model to predict whether policyholders (customers) from the previous year will also be interested in insurance. Vehicles provided by the company. The researcher uses the K-nearest neighbor classification algorithm, Random Forest, and gradient boosting classifier as well as Python data mining tools. After doing the research, it was found that the K-nearest neighbor classification algorithm produces a higher accuracy of 91%, when compared to the Random Forest algorithm which is 87% and the boosting classifier algorithm is 88% in classifying customer interest in taking a vehicle insurance policy.

Keywords: Klasifikasi; K-nearest neighbor; Random Forest; Gradient Boosting Classifier; Data Mining

### Abstrak

Polis asuransi memberikan jaminan kompensasi atas kerugian, kerusakan, penyakit, atau kematian yang ditentukan sebagai imbalan atas pembayaran premi. Begitupun pada asuransi kendaraan, setiap tahun nasabah perlu membayar premi kepada perusahaan penyedia asuransi sehingga jika terjadi kecelakaan yang tidak menguntungkan bagi kendaraan, perusahaan penyedia asuransi memberikan kompensasi kepada nasabah. Tujuan dari dilakukannya penelitian ini adalah untuk mengklasifikasikan *dataset health insurance cross sell prediction* sehingga dapat ditemukan pola-pola ataupun hubungan keterkaitan tertentu antara data untuk menjadi informasi yang berharga dan membangun model untuk memprediksi apakah pemegang polis (pelanggan) dari tahun sebelumnya juga akan tertarik dengan Asuransi Kendaraan yang disediakan oleh perusahaan. Peneliti menggunakan algoritma klasifikasi *K-nearest neighbor*, *Random Forest* dan *gradient boosting classifier* dan juga *tools data mining* Python. Setelah dilakukannya penelitian, didapatkan hasil bahwa algoritma klasifikasi *K-nearest neighbor* menghasilkan akurasi yang lebih tinggi yaitu sebesar 91%, jika dibandingkan dengan algoritma *Random Forest* yaitu sebesar 87% dan algoritma *boosting classifier* sebesar 88% dalam mengklasifikasikan ketertarikan nasabah untuk mengambil polis asuransi kendaraan

Kata Kunci: Klasifikasi; *K-nearest neighbor*; *Random Forest*; *Gradient Boosting Classifier*; *Data Mining*

## 1. Pendahuluan

Polis asuransi merupakan surat yang berisikan kesepakatan antara pihak yang akan berperan sebagai pengguna asuransi dengan sebuah perusahaan penyedia asuransi (Soemitra, 2017). Polis asuransi dapat memberikan jaminan kompensasi atas kerugian, kerusakan, penyakit, atau kematian yang ditentukan sebagai bentuk imbalan atas pembayaran premi yang telah dilakukan sebelumnya. Premi sendiri merupakan banyaknya uang atau dana yang wajib dibayarkan oleh setiap pemegang polis kepada penanggung karena sebelumnya sudah ada perjanjian pertanggungan yang telah disepakati bersama (Soemitra, 2017).

Sama seperti asuransi kesehatan, ada asuransi kendaraan dimana setiap tahun nasabah perlu membayar premi sebesar-besarnya kepada perusahaan penyedia asuransi sehingga jika terjadi kecelakaan yang tidak menguntungkan oleh kendaraan, perusahaan penyedia asuransi memberikan kompensasi kepada nasabah tersebut.

Perusahaan WXZ salah satu badan usaha yang berkecimpung dalam bidang asuransi. Perusahaan WXZ sudah memiliki jumlah nasabah atau pemegang polis yang terhitung banyak. Data dari nasabah Perusahaan WXZ kemudian diproses menggunakan teknik *data mining*.

Data *mining* adalah suatu rangkaian proses yang dijalankan menggunakan beberapa teknik *machine learning* atau pembelajaran komputer guna mencari tahu dan merinkas sebuah pengetahuan atau *knowledge* (Mustafa, Ramadha, & Thenata, 2017) atau dapat juga dikatakan bahwa Data *mining* menemukan pola atau informasi yang dirasa menarik di dalam sebuah data dengan memanfaatkan suatu model. Model atau algoritma dalam data mining sangatlah beragam dan bervariasi (Muzakir & Wulandari, 2016).

Dengan menjalankan rangkaian data *mining*, diharapkan dapat mengetahui acuan ataupun hubungan yang ada antara data agar dapat dijadikan suatu informasi berharga dan membangun model untuk memprediksi apakah pemegang polis (pelanggan) dari tahun sebelumnya juga akan tertarik dengan Asuransi Kendaraan yang disediakan oleh perusahaan. Dari pola-pola ataupun hubungan keterkaitan data ini kemudian juga dapat dimanfaatkan perusahaan untuk merencanakan strategi komunikasinya sehingga dapat menjangkau pelanggan tersebut dan mengoptimalkan model bisnis dan pendapatannya.

Dalam beberapa tahun ini, data *mining* banyak menyita perhatian masyarakat, karena bisa mengubah data yang luas dan dalam jumlah yang besar menjadi suatu informasi yang berguna. Salah satu tugas yang bisa dilakukan dengan teknik data *mining* adalah melakukan proses pengklasifikasian. klasifikasi sendiri dapat diartikan sebagai suatu hal yang dikerjakan untuk memberi nilai pada data pokok untuk digolongkan kedalam kelompok *group* tertentu dari jumlah *group* yang tersedia atau dapat juga dikatakan rangkaian prosedur yang melakukan pelatihan ataupun pembelajaran terhadap fungsi target yang memetakan setiap set atribut ke satu jumlah label *group* yang sudah dibuat (Utomo, 2020)

Aplikasi data *mining* yang dimanfaatkan di penelitian ini yaitu aplikasi *python*. *Python* dapat diartikan sebagai bahasa pemrograman *interpretatif* multiguna. *Python* sendiri berbeda dengan yang lainnya. Bahasa pemrograman yang lain cenderung lebih susah untuk diaplikasikan dan dipahami, *python* berfokus pada penulisan kode agar memudahkan untuk memahami perintah yang ditulis dalam bentuk coding. Berdasarkan hal tersebut maka menjadikan aplikasi *Python* sangat mudah dipelajari oleh pemula maupun bagi yang sudah menguasai bahasa pemrograman selain *python* sebelumnya (Harismawan, Kharisma, & Afirianto, 2018)

Beberapa tinjauan studi yang telah penulis lakukan yaitu pada penelitian yang memprediksi Status penggunaan Asuransi Kendaraan dengan menerapkan teknik pemilihan mayoritas (Utari & Wibowo, 2020) dan juga pada penelitian mengenai suatu Prediksi Keputusan yang akan diambil calon nasabah yang dinilai berpotensi menggunakan teknik Particle Swarm Optimization hasil yang didapat yaitu *decision tree* C4.5 berbasis PSO (*Particle Swarm Optimization*) memberikan nilai akurasi sebanyak 82.92% (Dwiasnati & Devianto, 2019)

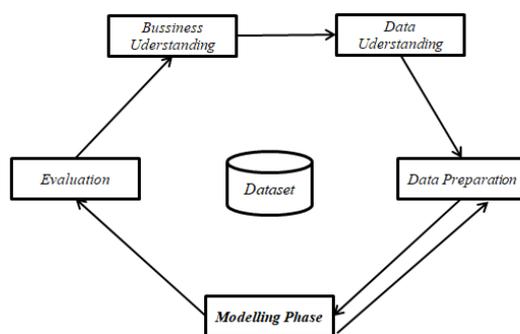
Selain itu ada juga penelitian dengan tujuan untuk memperkirakan suatu tingkat kemudahan pembayaran kredit menggunakan model *Naive Bayes* berlandaskan alur *Forward Selection* memperoleh nilai akurasi sebesar 71,97% (Hasan, 2017)

Dari *literature review* yang dilakukan, maka peneliti tertarik meneliti tentang ketertarikan Nasabah pada Asuransi kendaraan menggunakan beberapa algoritma data *mining* seperti algoritma *K-nearest neighbor*, algoritma *random forest*, dan *Gradient Boosting*. Dari beberapa metode ini

kemudian akan diusulkan satu metode atau algoritma yang dianggap mampu dan memiliki akurasi tertinggi dalam mengklasifikasikan prediksi ketertarikan nasabah pada asuransi kendaraan.

## 2. Metode Penelitian

Metode penelitian yang diterapkan pada penelitian ini berdasar pada *Cross-Standard Industry for Data Mining* (CRISP-DM) sebagai suatu bentuk penyeragaman pendekatan *data mining*. Adapun Metode penelitian adalah sebagai berikut:



Gambar 1. Metode penelitian CRISP-DM

Berikut Penjelasan dari fase penelitian CRISP-DM (Fadillah, 2015):

- Tahap Pengenalan Bisnis (*Business Understanding*)**  
Pemahaman bisnis ini adalah untuk menentukan tujuan dari bisnis, memberikan penilaian terhadap situasi, dan juga untuk menentukan tujuan *data mining* diterapkan.
- Tahap Pengenalan Data (*Data Understanding*)**  
Beberapa *point* yang perlu ditekankan pada saat pemahaman data seperti mengumpulkan data yang pertama digunakan, menjabarkan data, mengeksplor data, dan memvalidasikan keungulan dari data.
- Tahap Perencanaan Data (*Data Preparation*)**  
Beberapa perihal yang perlu disiapkan antara lain, mendeskripsikan *dataset*, memilih data yang mau digunakan, kemudian membangun data serta mengintegrasikan data dan melakukan pembersihan data.
- Tahap Pemodelan (*Modelling Phase*)**  
Tahapan pemodelan merupakan tingkatan yang secara langsung menyertakan *data mining* dalam pemilihan Teknik, algoritma maupun parameter dengan nilai yang optimal. Pada tahapan penentuan model, ada beberapa perihal yang diterapkan antara lain,

memutuskan teknik pemodelan, merumuskan model, dan menilai model. Fase yang dilakukan menunjuk pada penelitian tersebut

### e. Tahap Evaluasi

Tahap terakhir pada fase penelitian ini adalah tahap evaluasi. Proses dimana kita menginterpretasikan hasil penelitian *data mining* yang telah selesai dilakukan. Proses mengevaluasi dilakukan secara menyeluruh dengan sasaran agar hasil pada fase pembangunan model dapat sesuai dengan sasaran pada tahap *business understanding*. Ada beberapa *point* yang perlu dikerjakan yaitu: mengkaji hasil yang diperoleh (*Evaluate Results*), melakukan tinjauan terhadap prosedur (*Review Process*), dan juga menentukan tahapan yang perlu dilakukan selanjutnya.

### 2.1. K-Nearest Neighbor (K-NN)

Algoritma *K-Nearest Neighbor* (K-NN) diartikan sebagai model klasifikasi yang menentukan kategori kelas dari sebuah objek baru didasarkan pada kelas terbanyak berdasarkan dari jarak terdekat  $k$  dalam *group* data yang telah dilatih (Wijaya & Ridwan, 2019)

Algoritma *K-Nearest Neighbor* (K-NN) menjalankan perintah klasifikasi berdasarkan pada jarak terdekat dari sampel uji ke sampel terlatih untuk menetapkan KNNnya. Setelah *K-Nearest Neighbor* (K-NN) terkumpul kemudian barulah diambil mayoritas dari *K-Nearest Neighbor* (K-NN) yang selanjutnya dijadikan sebagai prediksi dari sampel ujinya. Jarak objek terdekat biasanya dihitung berdasarkan jarak Eucledian. Dibawah ini akan diperlihatkan tata cara untuk menghitung metode *K-Nearest Neighbor* (Wijaya & Ridwan, 2019):

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Keterangan:

- $x$  = Sampel dari data
- $x_2$  = Data yang akan diujikan
- $I$  = Parameter dari data
- $d$  = Selisih atau Jarak
- $p$  = Luas dari data

Langkah-langkah dari teknik *K-Nearest Neighbor* yaitu mulai dari input data terlatih, kemudian label data terlatih, nilai  $k$ , dan juga data pengujian (Wijaya & Ridwan, 2019)

- Seluruh data yang pengujian, hitung terlebih dahulu jaraknya ke setiap data terlatih
- Kemudian menentukan nilai  $k$  data terlatih yang memiliki jarak paling dekat dengan data pengujian
- Setelah itu cek label dari  $k$  data ini
- Tentukan juga label yang frekuensinya paling banyak
- Selain itu masukan data pengujian ke kelas yang mempunyai frekuensi terbanyak
- Kemudian *stop* dan label untuk semua data pengujian bisa diperoleh.

## 2.2. Random Forest

*Random Forest Algorithm* merupakan suatu algoritma *machine learning* yang dimanfaatkan untuk proses pengklasifikasian. *Decision Tree* merupakan bagian dari *Random forest*, yaitu merupakan suatu pohon keputusan untuk membedakan suatu data (Sanjaya, Renata, Budiman, Anderson, & Ayub, 2020)

*Random Forest Algorithm* dapat juga dianggap sebagai bentuk perkembangan dari metode klasifikasi *Decision Tree* yang didasarkan dari pemilihan atribut *random* disetiap *node* untuk membentuk klasifikasi. Proses pengklasifikasinya ini didasarkan pada jumlah suara terbanyak dari pohon keputusan yang didapat (Ratnawati & Sulistyaningrum, 2019)

Berdasarkan teknis yang diterapkan oleh *Random Forest*, membuat hasil prediksi yang lebih efisien. Secara umum, *Random Forest* memiliki kelebihan antara lain (Wiraguna et al., 2019):

- Dapat mengatasi masalah *over-fitting*
- Tidak terlalu sensitif terhadap data *outlier*
- Parameter dapat diubah.

## 2.3. Gradient Boosting

*Gradient Boosted Trees* menjadi salah satu dari teknik mesin pembelajaran yang dimanfaatkan untuk mengatasi klasifikasi maupun regresi. *Gradient Boosted Trees* menghasilkan model prediksi *ensemble* yaitu dari model prediksi yang menghasilkan performance lemah, kemudian metode *Gradient Boosted Trees* memberikan asumsi sebuah nilai real bernilai  $y$  (Pertiwi, Adiwisatra, & [3], 2019). Metode ini menggabungkan pembelajar lemah menjadi pembelajar tunggal yang kuat secara berulang-ulang (Salam, Musaruddin, & Moku, 2020).

## 3. Hasil

### 3.1. Tahap Pemahaman Bisnis (*Bussiness Understanding*)

Penelitian ini sumber informasi yang digunakan yaitu dari kaggle. Pemahaman terhadap objek penelitian dilakukan dengan mencari dan menggali informasi dari objek penelitian itu sendiri yaitu dari data *health insurance cross sell prediction* agar dapat dilakukan klasifikasi ketertarikan nasabah pada asuransi kendaraan

### 3.2. Tahap Pemahaman Data (*Data Understanding*)

Sebelum pemahaman data maka perlu dilakukannya pengumpulan, pengidentifikasi, dan pemahaman data yang akan diproses. Data yang diterapkan dalam penelitian diperoleh dari data pengguna asuransi yang didapatkan dari <https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction>. Dataset ini memiliki 127038 *number of instances*, 11 *number of attributes* dan 2 *attribute class*.

### 3.3. Tahap Mempersiapkan Data (*Data Preparation*)

Menyiapkan data awal untuk perangkat permodelan, data penelitian ini sebanyak 127038 data yang kemudian diproses menggunakan algoritma klasifikasi untuk mengklasifikasikan data ketertarikan nasabah untuk membeli polis asuransi kendaraan. Kelas atau hasil dari data set ini yaitu *variable response* yang memiliki nilai 1 dan 0, dimana 1 berarti tertarik untuk melakukan *cross selling*, dan 0 berarti tidak mau melakukan *cross selling*. Didalam proses ini, data diklasifikasikan sesuai variabel yang telah dibuat. Beberapa komponen variabel yang digunakan, yaitu:

Tabel 1. Variabel Data *Health Insurance Cross Sell Prediction*

Nama Variabel	Type	Keterangan
<i>Gender</i>	<i>Numeric</i>	Jenis Kelamin
<i>Age</i>	<i>Numeric</i>	Usia
<i>Driving License</i>	<i>Numeric</i>	SIM
<i>Region Code</i>	<i>Numeric</i>	Kode POS
<i>Previously insured</i>	<i>Numeric</i>	Pernah mengasuransikan kendaraan sebelumnya
<i>Vehicle age</i>	<i>Numeric</i>	Usia Kendaraan
<i>Vehicle damage</i>	<i>Numeric</i>	Kerusakan Kendaraan
<i>Annual Premium</i>	<i>Numeric</i>	Premi tahunan
<i>Policy Sales</i>	<i>Numeric</i>	Kebijakan penjualan

Vintage	Numeric	Keantikan
Response	Numeric	Ketertarikan

### 3.4. Tahap Pemodelan (Modelling Phase)

Menentukan atau membangun model *data mining* yang digunakan untuk mengolah data yang sudah disiapkan. Tujuan dari penerapan teknik data mining ini adalah untuk melakukan klasifikasi keterterikan konsumen untuk mengambil polis kendaraan menggunakan algoritma *K-nearest neighbor*, *random forest* dan *gradient boosting classifier*. Tahap pemodelan dapat dilihat pada gambar 2

```
[ ] from sklearn.linear_model import LinearRegression
# Split Feature Vector and Label
X = df.drop(['Response'], axis = 1) # menggunakan semua feature kecuali target
y = df['Response'] # target / label

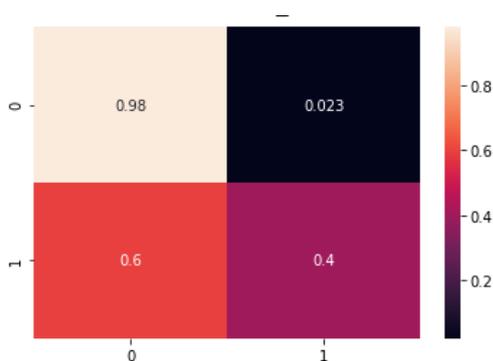
#Splitting the data into Train and Test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,
                                                    y,
                                                    test_size = 0.3,
                                                    random_state = 789)
```

Gambar 2. Tahap pemodelan klasifikasi

Data yang sudah di *preprocessing* kemudian dilakukan perhitungan dengan menggunakan *tools* Python agar dihasilkan performa dari masing-masing algoritma yang diuji, penelitian ini melakukan pembagian persentasi data testing dan data training. Berikut penjelasan dari masing-masing algoritma tersebut:

#### a. K-nearest neighbor

Berikut *confusion matrix* yang diperoleh algoritma *K-nearest neighbor* berdasarkan pengujian menggunakan *tools* Python dapat dilihat pada gambar 3



Gambar 3. Confusion Matrix K-Nearest Neighbor

Terlihat pada tabel 2 hasil perhitungan *confusion matrix* yang dilakukan dengan metode *K-nearest neighbor*. Dengan menerapkan metode *K-nearest neighbor* hasil akurasi yang didapatkan sebesar 91%, nilai *presisi* 90% dan *recall* 91%

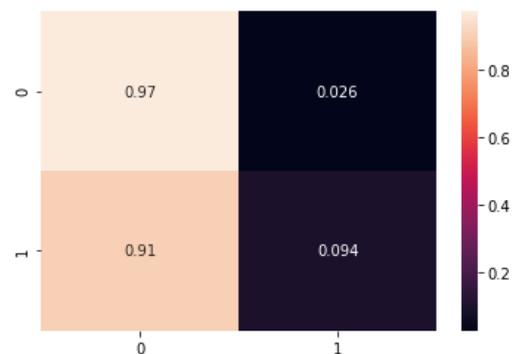
Tabel 2 Clasification Report K-Nearest Neighbor

	Precision (%)	Recall (%)	F1-Score (%)
0	92	98	95
1	71	40	51
Accuracy			91
Macro Avg	82	69	73
Weighted Avg	90	91	90

Berdasarkan tabel di atas, hasil perhitungan nilai rata-rata algoritma *K-nearest neighbor* diperoleh akurasi sebesar 91%, presisi 90% dan *recall* sebesar 91% dan *F1-Score* sebesar 90%

#### b. Random Forest

Berikut *confusion matrix* yang diperoleh algoritma *K-nearest neighbor* berdasarkan pengujian menggunakan *tools* Python dapat dilihat pada gambar 4



Gambar 4. Confusion Matrix Random Forest

Dari hasil perhitungan *confusion matrix* yang dilakukan dengan metode *Random Forest*, maka dihasilkan ringkasan *classification report* yang dapat dilihat pada tabel 3. Terlihat bahwa hasil perhitungan nilai rata-rata algoritma *Random Forest* diperoleh akurasi 87%, presisi 82% dan *recall* sebesar 87% dan *F1-Score* sebesar 83%

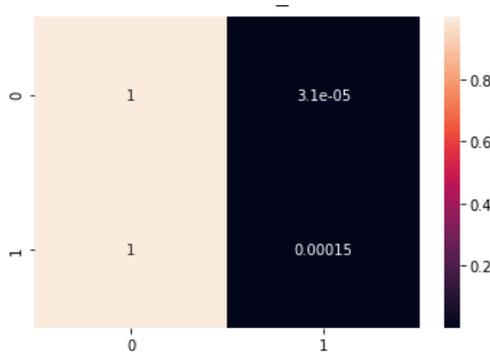
Tabel 3 Clasification Report Random Forest

	Precision (%)	Recall (%)	F1-Score (%)
0	89	97	93
1	33	9	15
Accuracy			87
Macro Avg	61	53	54
Weighted Avg	82	87	83

#### c. Gradient Boosting Classifier

Berikut *confusion matrix* yang diperoleh algoritma *gradient boosting classifier* berdasarkan

pengujian menggunakan tools Python dapat dilihat pada gambar 5



Gambar 5. Confusion Matrix gradient boosting

Dari hasil perhitungan *confusion matrix* yang dilakukan dengan metode *gradient boosting classifier*, maka dihasilkan ringkasan *classification report* yang dapat dilihat pada tabel 4. Terlihat bahwa hasil perhitungan nilai rata-rata algoritma *gradient boosting classifier* diperoleh akurasi 88%, presisi 82% dan recall sebesar 88% dan F1-Score sebesar 82%

Tabel 4 Classification Report Gradient Boosting

	Precision (%)	Recall (%)	F1-Score (%)
0	88	100	94
1	40	0	0
Accuracy			88
Macro Avg	64	50	47
Weighted Avg	82	88	82

### 3.5. Evaluasi (Evaluation)

Dapat diketahui dari proses data mining tersebut hasil pengukuran keakuratan (akurasi) yang dicapai oleh model yang dibangun sebagai bentuk representasi hasil proses klasifikasi. Tabel 5 menunjukkan perbandingan Hasil pengujian dari algoritma *K-nearest neighbor*, *Random Forest* dan *gradient boosting classifier*.

Tabel 5 Perbandingan Hasil Uji

Algoritma	Akuracy (%)	Precision (%)	Recall (%)
<i>K-Nearest Neighbor</i>	91	90	91
<i>Random Forest</i>	87	82	87
<i>Gradient Booster</i>	88	82	88

## 4. Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan oleh peneliti mengenai klasifikasi ketertarikan nasabah pada asuransi kendaraan menggunakan algoritman klasifikasi *K-nearest neighbor*, *Random Forest* dan *gradient boosting classifier*, maka dapat diperoleh sebuah kesimpulan, sebagai berikut:

- Hasil *accuracy* menunjukkan hasil yang berbeda tergantung dengan jenis data, nilai *k-folds* dan jumlah *instance*, label *class*, *Percentage split* yang digunakan. Pada penelitian ini peneliti menggunakan *Percentage split*, 70% *data training* dan 30% *data test*
- Algoritma klasifikasi *K-nearest neighbor* menghasilkan akurasi yang lebih tinggi yaitu sebesar 91%, jika dibandingkan dengan algoritman *Random Forest* yaitu sebesar 87% dan algoritma *boosting classifier* sebesar 88% dalam megklasifikasikan ketertarikan nasabah untuk mengambil polis asuransi kendaraan

## Referensi

- Dwiasnati, S., & Devianto, Y. (2019). *Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swarm Optimization*. 6(2), 286–292.
- Fadillah, A. P. (2015). *Penerapan Metode CRISP-DM Untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah ( Studi Kasus Universitas XYZ )*. 1, 260–270.
- Harismawan, A. F., Kharisma, A. P., & Afrianto, T. (2018). *Analisis Perbandingan Performa Web Service Menggunakan Bahasa Pemrograman Python , PHP , Dan Perl Pada Client Berbasis Android*. 2(1), 237–245.
- Hasan, M. (2017). Menggunakan algoritma naive bayes berbasis. 9, 317–324.
- Mustafa, M. S., Ramadha, M. R., & Thenata, A. P. (2017). *Implementasi Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. 4(2).
- Muzakir, A., & Wulandari, R. A. (2016). *Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan Dengan Teknik Decision Tree*. 3(1), 19–26.
- Pertiwi, M. W., Adiwisatra, M. F., & [3], D. S. (2019). *Analisa Komparasi Menggunakan 5 Metode Data Mining Dalam Klasifikasi Persentase Wanita Sudah Menikah Di Usia*

- 15-49 Yang Memakai Alat KB (Keluarga Berencana). (November).  
<https://doi.org/10.31294/Jki.V7i1.5741>
- Ratnawati, L., & Sulistyningrum, D. R. (2019). Penerapan Random Forest Untuk Mengukur Tingkat Keparahan Penyakit Pada Daun Apel. 8(2).
- Salam, D., Musaruddin, M., & Mokui, H. T. (2020). Kombinasi Analisis Teknis Dan Algoritma Gradient Boosting Untuk Memprediksi Indeks Harga Saham Gabungan. 1–6.
- Sanjaya, J., Renata, E., Budiman, V. E., Anderson, F., & Ayub, M. (2020). Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest Dan Adaptive Boosting. 6(April), 50–60.
- Soemitra, A. (2017). *Bank & Lembaga Keuangan Syariah*. Jakarta: Kencana.
- Utari, D. R., & Wibowo, A. (2020). *Pemodelan Prediksi Status Keberlanjutan Polis Asuransi Kendaraan Dengan Teknik Pemilihan Mayoritas Menggunakan Algoritma-Algoritma Klasifikasi Data Mining*. 5(2502), 19–24.  
<https://doi.org/10.22236/Teknoka.V5i.391>
- Utomo, D. P. (2020). Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung. 4(April), 437–444.  
<https://doi.org/10.30865/Mib.V4i2.2080>
- Wijaya, N., & Ridwan, A. (2019). Klasifikasi Jenis Buah Apel Dengan Metode K-Nearest Neighbors. 8, 74–78.
- Wiraguna, A., Faraby, S. Al, Sc, M., Adiwijaya, P., Si, S., & Si, M. (2019). Klasifikasi Topik Multi Label Pada Hadis Bukhari Dalam Terjemahan Bahasa Indonesia Menggunakan Random Forest. 6(1), 2144–2153.