

**IMPLEMENTASI K-OPTIMAL PADA ALGORITMA *K-NEAREST
NEIGHBOR* UNTUK KLASIFIKASI KETERTARIKAN MASYARAKAT
DALAM PEMBELIAN ASURANSI PERJALANAN**

(Skripsi)

Oleh

**QORRY INDIRA DEVINA
NPM 1917031098**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

K-OPTIMAL IMPLEMENTATION OF THE K-NEAREST NEIGHBOR ALGORITHM FOR CLASSIFICATION OF PUBLIC INTEREST IN PURCHASING TRAVEL INSURANCE

By

Qorry Indira Devina

The rapid development of science and technology in the era of globalization provides benefits and progress, especially in the transportation sector. The development of the transportation sector, in addition to having advantages, also has risks that often cause substantial losses. One of the risk transfer agencies that can accommodate and take over all losses arising from these risks, including a travel insurance company. This research examines an analysis to classify who might be interested in purchasing travel insurance so that it can assist companies in optimizing product marketing based on their historical database using the K-Nearest Neighbor algorithm. The data used is travel insurance prediction data in India. From the modelling results using the division of 90% train data and 10% test data, the most optimal number of nearest neighbours is four ($k=4$) with a model accuracy rate of 83%. Therefore, travel insurance prediction data with $k=4$ can be used to classify and predict customer groups.

Keywords : Classification, K-Nearest Neighbor, Travel Insurance.

ABSTRAK

IMPLEMENTASI K-OPTIMAL PADA ALGORITMA *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI KETERTARIKAN MASYARAKAT DALAM PEMBELIAN ASURANSI PERJALANAN

Oleh

Qorry Indira Devina

Perkembangan ilmu pengetahuan dan teknologi yang sangat pesat pada era globalisasi memberikan manfaat serta kemajuan, khususnya pada sektor transportasi. Perkembangan sektor transportasi tersebut, selain memiliki keuntungan juga memiliki risiko yang tidak jarang menimbulkan kerugian yang cukup besar. Salah satu lembaga pengalihan risiko yang dapat menampung dan mengambil alih segala kerugian yang timbul dari risiko-risiko tersebut, diantaranya adalah perusahaan asuransi perjalanan. Pada penelitian ini dikaji analisis untuk mengklasifikasikan siapa saja yang sekiranya akan tertarik dalam pembelian asuransi perjalanan agar dapat membantu perusahaan dalam mengoptimalkan pemasaran produk berdasarkan sejarah databasenya, menggunakan algoritma *K-Nearest Neighbor*. Data yang digunakan adalah data *travel insurance prediction* di India. Dari hasil pemodelan menggunakan pembagian 90% data *train* dan 10% data *test* didapatkan jumlah tetangga terdekat yang paling optimal sebesar 4 tetangga ($k = 4$) dengan tingkat akurasi model yang didapatkan sebesar 83%. Oleh karena itu, dapat disimpulkan bahwa data *travel insurance prediction* dengan $k = 4$ dapat digunakan untuk klasifikasi dan memprediksi golongan *customer*.

Kata Kunci : Klasifikasi, *K-Nearest Neighbor*, Asuransi Perjalanan.

**IMPLEMENTASI K-OPTIMAL PADA ALGORITMA *K-NEAREST
NEIGHBOR* UNTUK KLASIFIKASI KETERTARIKAN MASYARAKAT
DALAM PEMBELIAN ASURANSI PERJALANAN**

Oleh

QORRY INDIRA DEVINA

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi

**: IMPLEMENTASI K-OPTIMAL PADA
ALGORITMA K-NEAREST NEIGHBOR
UNTUK KLASIFIKASI
KETERTARIKAN MASYARAKAT
DALAM PEMBELIAN ASURANSI
PERJALANAN**

Nama Mahasiswa

: Qorry Indira Devina

Nomor Pokok Mahasiswa

: 1917031098

Jurusan

: Matematika

Fakultas

: Matematika dan Ilmu Pengetahuan Alam



1. Komisi Pembimbing

Dr. Ahmad Faisol, S.Si., M.Sc.
NIP 19800206 200312 1 003

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

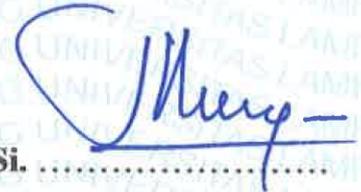
MENGESAHKAN

1. Tim Penguji

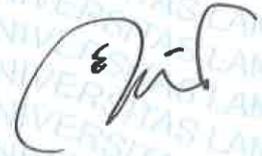
Ketua : Dr. Ahmad Faisol, S.Si., M.Sc.



Sekretaris : Dr. Aang Nuryaman, S.Si., M.Si.



Penguji : Drs. Eri Setiawan, M.Si.



Bukan Pembimbing



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Supto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 8 Februari 2023

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Qorry Indira Devina**
Nomor Pokok Mahasiswa : **1917031098**
Jurusan : **Matematika**
Judul Skripsi : **IMPLEMENTASI K-OPTIMAL PADA ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI KETERTARIKAN MASYARAKAT DALAM PEMBELIAN ASURANSI PERJALANAN**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 8 Februari 2023

Yang menyatakan,



Qorry Indira Devina

NPM 1917031098

RIWAYAT HIDUP

Qorry Indira Devina lahir di Serang, Provinsi Banten pada tanggal 21 April 2001, dari pasangan Bapak Robinson dan Ibu Retnaningsih. Penulis merupakan anak sulung dari tiga bersaudara, yakni Farell Abdiel Fayedh dan Fathir Sebastian.

Pendidikan Taman Kanak-kanak (TK) Aisyah Serang diselesaikan tahun 2007, Sekolah Dasar (SD) diselesaikan di SD Negeri 2 Kota Serang pada tahun 2013, Sekolah Menengah Pertama (SMP) di SMP Negeri 1 Kota Serang pada tahun 2016, dan Sekolah Menengah Atas (SMA) di SMA Negeri 2 Kota Serang pada tahun 2019.

Tahun 2019, penulis terdaftar sebagai mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis aktif mengikuti organisasi Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila. Pada bulan Juni sampai bulan Juli tahun 2022, penulis melakukan kerja praktik di Badan Pusat Statistik (BPS) Kota Yogyakarta. Pada bulan Agustus sampai bulan Desember 2022, penulis mengikuti Studi Independen yang diselenggarakan oleh Kampus Merdeka dengan program Magang dan Studi Independen Bersertifikat di PT. Mitra Talenta Grup (Celerates) dengan tema *Big Data & Business Intelligence*.

KATA INSPIRASI

*“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”
(Q.S Al-Baqarah: 286)*

*“Barang siapa bertaqwa kepada Allah, niscaya Dia akan mengadakan baginya jalan keluar. Dan memberinya rezeki dari arah yang tiada disangka-sangkanya”
(Q.S Ath-Thalaq: 2)*

*Allah akan mendatangkan hari-hari yang membuat kita Bahagia, sebab hidup takkan selamanya
sulit
Jangan pernah berhenti berdoa*

*Happiness is not something readymade
It comes from your own actions*

*Life is inherently risky.
There is only one big risk you should avoid at all costs, and that is the risk of doing nothing
So, take the risk or lose the chance*

*Impossible is just an opinion
When there is a will, there is way*

PERSEMBAHAN

Alhamdulillah, puji syukur kepada Allah SWT. yang telah memberikan petunjuk dan kekuatan juga memberikan penerangan dalam ilmu pengetahuan. Hanya karena-Nya lah skripsi ini bisa penulis selesaikan dengan rasa syukur dan bahagia. Dengan segala kerendahan hati, penulis persembahkan karya sederhana ini kepada:

Orang Tua Tercinta

Sebagai tanda terima kasih karena selalu mencurahkan doa, tenaga, pikiran, dan dukungannya untuk keberhasilan penulis dalam menuntut ilmu serta menjadi penyemangat terbaik sehingga penulis bisa menyelesaikan skripsi ini.

Dosen Pembimbing dan Pembahas

Yang senantiasa memberikan bimbingan, arahan, dan ilmu yang bermanfaat bagi penulis.

Almamaterku Tercinta, Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT. Atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul “Implementasi K-Optimal pada Algoritma *K-Nearest Neighbor* untuk Klasifikasi Ketertarikan Masyarakat dalam Pembelian Asuransi Perjalanan” dengan sebaik – baiknya dan selesai tepat pada waktunya. Skripsi ini disusun sebagai salah satu syarat memperoleh gelar Sarjana Matematika (S.Mat) pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan dukungan, bimbingan dan motivasi. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada :

1. Bapak Dr. Ahmad Faisol, S.Si., M.Sc., selaku dosen pembimbing 1 yang telah memberikan arahan, bantuan dan saran kepada penulis dalam menyusun Skripsi ini.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku dosen pembimbing 2 dan Ketua Jurusan Matematika yang telah memberikan bantuan dan saran yang bermanfaat dalam penyusunan Skripsi ini.
3. Bapak Drs. Bapak Drs. Eri Setiawan, M.Si. selaku dosen penguji yang telah memberikan arahan dan saran yang bermanfaat dalam penyusunan Skripsi ini.
4. Ibu Dr. Fitriani, S.Si., M.Sc. selaku dosen pembimbing akademik yang telah memberikan arahan dalam penyusunan Skripsi ini.
5. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Drajat yang selalu mendukung dalam pelaksanaan seminar dan sidang skripsi saya.

7. Seluruh dosen, staf dan Civitas Akademik Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Ayah, Mama, Abang, Ating, Iyo, Iyi, Eyo, Puyen 1, Puyen 2, Undel, Ntem dan keluarga besar yang selalu memberikan semangat, dukungan, motivasi, amunisi dan do'a kepada penulis.
9. Partner terbaikku Shidqi Rozaq yang senantiasa memberikan semangat dan dukungan dalam menyelesaikan skripsi ini.
10. Para Sahabat tersayangku Bela, Ija, Disa, Kinan, Meyin, Hana, Ali dan Balkum.
11. Teman seperjuanganku dari studi independen celerates, seminar proposal, seminar hasil, kompre hingga wisuda, yaitu Aul.
12. Teman – teman tersayangku Pebi, Nada, Shella, Deswita, Yeye, Poet, Fii, Eca, Listra, Ale, Aulia, Meli, Hijer, Debi, Encik, Manda dan Sugar yang telah memberi semangat, dukungan dan do'a sehingga penulis dapat menyelesaikan skripsi.
13. Seluruh teman Jurusan Matematika 2019, terimakasih atas kebersamaannya.
14. Semua pihak yang tidak bisa saya sebutkan satu – persatu yang telah membantu penulis dalam menyelesaikan skripsi.

Semoga kebaikan yang telah diberikan kepada penulis dapat terbalaskan dengan kebaikan. Penulis menyadari bahwa skripsi ini masih banyak kekurangan baik dalam penyajian maupun teknik penulisan. Oleh karena itu, kritik dan saran sangat penulis harapkan agar dapat menjadi pelajaran dan perbaikan kedepannya. Semoga skripsi ini dapat bermanfaat untuk siapapun.

Bandar Lampung, 8 Februari 2023
Penulis

Qorry Indira Devina
NPM. 1917031098

DAFTAR ISI

Halaman

ABSTRAK	ii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	5
2.1 Asuransi	5
2.2 Klasifikasi	8
2.3 <i>Artificial Intelligence (AI)</i>	9
2.4 <i>Machine Learning</i>	10
2.5 Algoritma K-Nearest Neighbor	11
2.5.1 Perhitungan Jarak	11
2.5.2 <i>Confusion Matrix</i>	15
III. METODOLOGI PENELITIAN.....	18
3.1 Waktu dan Tempat Penelitian.....	18
3.2 Data Penelitian.....	18
3.3 Metode Penelitian	18
IV. HASIL DAN PEMBAHASAN.....	21
4.1 Deskripsi Data	21
4.2 <i>Data Preparation</i>	23
4.3 <i>Data Processing</i>	26
4.3.1 Pembagian Data <i>Train</i> dan Data <i>Test</i>	27
4.3.2 Penentuan Parameter k	28
4.3.3 <i>Training Data</i>	32
4.4 Evaluasi Data	34

V. PENUTUP	37
5.1 Kesimpulan.....	37
5.2 Saran	37
DAFTAR PUSTAKA	38
LAMPIRAN.....	41

DAFTAR TABEL

Tabel	Halaman
1. Contoh Data <i>Train</i> dan Data <i>Test</i>	12
2. Interpretasi Nilai <i>Error</i>	17
3. Data <i>Travel Insurance Prediction</i>	22
4. Sepuluh Data Pertama pada Data <i>Travel Insurance Prediction</i>	24
5. Statistika Deskriptif Data <i>Travel Insurance Prediction</i>	25
6. Contoh Perhitungan Manual K-Nearest Neighbor	32
7. <i>Classification Report</i>	35

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Confusion Matrix</i>	16
2. Plot <i>Error Rate</i> Data <i>Train</i> 70% dan Data <i>Test</i> 30%	29
3. Plot <i>Error Rate</i> Data <i>Train</i> 80% dan Data <i>Test</i> 20%	30
4. Plot <i>Error Rate</i> Data <i>Train</i> 90% dan Data <i>Test</i> 10%	31
5. <i>Heatmap Confusion Matrix</i>	34

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Perkembangan ilmu pengetahuan dan teknologi yang sangat pesat pada era globalisasi pada saat ini memberikan manfaat serta kemajuan khususnya pada sektor transportasi. Pada awalnya perkembangan transportasi bergerak secara perlahan, yang awalnya manusia hanya dapat berjalan kaki untuk berpergian, penemuan roda untuk transportasi lalu memanfaatkan hewan untuk ditunggangi seperti kuda dan sapi. Berkembangnya sektor transportasi dalam hal ini transportasi darat baik dalam kendaraan pribadi seperti sepeda, motor dan mobil serta kendaraan umum seperti angkutan kota, bus dan travel membuat aktivitas manusia menjadi jauh lebih mudah. Transportasi darat dapat membuat segala aktivitas manusia menjadi lebih praktis dan efisien waktu (Asri, dkk., 2017).

Semakin meningkatnya inovasi pada perkembangan transportasi darat terdapat juga bahaya yang selalu mengintai pengendara maupun penumpang. Perkembangan sektor transportasi tersebut tentu disertai dengan risiko yang tidak jarang menimbulkan kerugian dari kerugian kecil bahkan sampai kerugian yang sangat besar. Salah satu bahaya tersebut adalah kecelakaan lalu lintas yang sering ditemui. Secara umum terdapat empat faktor utama yang menyebabkan kecelakaan lalu lintas, yaitu faktor pengguna jalan seperti kondisi fisik pengemudi mabuk, lelah atau sakit, faktor kendaraan seperti kondisi mesin, ban, lampu, rem ataupun muatan kendaraan, faktor lingkungan jalan seperti medan jalan yang rusak, bergelombang berlubang ataupun rambu lalu lintas yang rusak dan faktor cuaca seperti hujan, asap, kabut, bersalju dan lain sebagainya. Faktor-faktor yang menyebabkan terjadinya kecelakaan tersebut saling berkaitan dalam terjadinya kecelakaan lalu lintas yang

merupakan risiko yang dapat timbul ketika manusia menggunakan transportasi, baik transportasi pribadi ataupun transportasi umum. Per 3 Oktober 2022 Korlantas Polri mencatat angka kecelakaan lalu lintas yang terjadi di Indonesia pada 2022 masih cukup tinggi. Terjadi 6.707 kasus dengan korban meninggal dunia sebanyak 452 orang, yang mengalami luka berat sebanyak 972 dan yang mengalami luka ringan sebanyak 6.704 orang dengan total kerugian material sekitar Rp 13 miliar lebih. Salah satu hal yang menyebabkan kecelakaan adalah rendahnya kesadaran pengemudi dalam berlalu lintas. Untuk mencegah segala kemungkinan yang terjadi di jalanan sewaktu-waktu maka sangat diperlukan suatu lembaga pengalihan risiko yang dapat menampung dan mengambil alih segala kerugian yang timbul dari risiko-risiko tersebut, misalnya perusahaan asuransi perjalanan (Megasari, 2022).

Asuransi merupakan sebuah perjanjian antara penanggung dan tertanggung yang mewajibkan tertanggung membayar premi dengan kesepakatan dua belah pihak untuk memberikan penggantian atas kerugian, kerusakan, kehilangan keuntungan maupun kematian yang tidak diharapkan yang mungkin sewaktu-waktu akan terjadi dikarenakan suatu kejadian yang tidak terduga yang dapat terjadi kapanpun dan dimanapun. Istilah asuransi berasal dari Bahasa Inggris, yaitu '*insurance*' yang artinya pertanggungan dan bahasa Belanda, yaitu '*assurantie*' atau '*verzekering*' yang berarti pertanggungan, yang kemudian muncul istilah '*assuradeur*' yang artinya penanggung sedangkan '*greassureede*' yang artinya tertanggung. Asuransi memegang peranan penting, karena asuransi memberikan perlindungan terhadap kemungkinan-kemungkinan buruk yang dapat terjadi. Sebagai aset paling berharga dalam hidup, sudah seharusnya setiap orang mendapatkan perhatian paling utama. Salah satunya yaitu dengan membekali diri dengan asuransi. Banyak kejadian atau musibah yang tidak diinginkan yang dapat terjadi dalam hidup dan setiap orang harus selalu siap ketika kejadian atau musibah buruk datang. Oleh karena itu diperlukan suatu analisis mendalam untuk mengklasifikasi dan memprediksi siapa saja yang sekiranya akan tertarik untuk membeli polis asuransi, khususnya asuransi perjalanan, salah satunya menggunakan metode *K-Nearest Neighbor* (KNN) (Betris, 2017).

Algoritma *K-Nearest Neighbor* merupakan sebuah algoritma yang menghitung jarak antara kasus baru dan kasus lama berdasarkan pencocokan bobot. *K-Nearest Neighbor* merupakan algoritma berbasis pembelajaran dimana *dataset* terbagi menjadi dua bagian yaitu dataset *train* dan data *test*. *Dataset train* disimpan sebagai acuan data *test*, sehingga klasifikasi data *test* baru yang tidak terklasifikasi diperoleh dengan membandingkan jarak data *train* terdekat yang paling mirip dengan kumpulan pembelajaran data *train* (Mustafa & Simpen, 2014).

Penelitian sebelumnya telah banyak dilakukan dengan klasifikasi menggunakan *K-Nearest Neighbor*. Seperti pada penelitian Mustakim dan Oktaviani (2016) yang melakukan penelitian *Algoritma K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. Penelitian ini bertujuan untuk mengklasifikasikan dan memprediksi kasus penentuan predikat prestasi Mahasiswa dengan empat kelas predikat, yaitu pujian, sangat memuaskan, memuaskan dan cukup penelitian ini dilakukan dikarenakan jumlah mahasiswa baru yang semakin tahun semakin banyak dan semakin tidak menutup kemungkinan bahwa mahasiswa lama dapat menyelesaikan perkuliahan dengan tepat waktu sehingga, mengakibatkan jumlah mahasiswa program studi sistem informasi di UIN Sultan Syarif (Suska) Riau akan semakin banyak. Penelitian ini menghasilkan akurasi sebesar 82%.

Kasus lainnya yang serupa dibahas oleh Betrisandi (2017) dengan penelitian *Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination*. Penelitian ini bertujuan untuk mengklasifikasikan nasabah asuransi yang lancar dan tidak lancar membayar premi asuransi menggunakan *profile* data nasabah untuk pengukuran agar dapat membantu perusahaan asuransi dalam mengambil keputusan menerima atau menolak calon nasabah asuransi nantinya. Penelitian ini menghasilkan nilai akurasi menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* sebesar 85,89% , nilai *precision* sebesar 86,03% dan *recall* sebesar 99,77%. Sehingga dapat disimpulkan bahwa data nasabah asuransi jiwa tersebut dinyatakan akurat menggunakan metode *Naive Bayes* berbasis *Backward elimination*.

Penelitian-penelitian yang sudah dilakukan yaitu klasifikasi nasabah asuransi jiwa. Oleh karena itu, penulis tertarik untuk melakukan penelitian klasifikasi ketertarikan masyarakat dalam pembelian asuransi perjalanan. Penelitian ini bertujuan untuk mengklasifikasikan masyarakat mana saja yang sekiranya akan membeli asuransi perjalanan berdasarkan sejarah *databasenya* agar dapat membantu perusahaan asuransi perjalanan dalam mengambil keputusan untuk menawarkan asuransi perjalanan dengan penawaran yang maksimal kepada kepada calon *customer*.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan k-optimal untuk mendapatkan hasil klasifikasi terbaik.
2. Mengklasifikasikan masyarakat yang akan tertarik dan tidak tertarik untuk membeli polis asuransi perjalanan berdasarkan sejarah *databasenya* dengan algoritma *K-Nearest Neighbor* dengan menggunakan nilai k terbaik.

1.3 Manfaat Penelitian

Adapun manfaat dari kerja praktik ini adalah:

1. Menambah pengetahuan dan wawasan tentang algoritma *K-Nearest Neighbor*.
2. Mengetahui nilai *error* terkecil yang didapatkan dengan nilai k-optimal.
3. Mengetahui nilai akurasi maksimal dengan nilai k-optimal untuk mendapatkan hasil klasifikasi terbaik.

II. TINJAUAN PUSTAKA

2.1 Asuransi

Perusahaan jasa memiliki karakteristik yang berbeda dengan perusahaan non-jasa. Asuransi merupakan salah satu dari banyak perusahaan yang bergerak dibidang jasa dan merupakan Lembaga Keuangan Non-Bank. Asuransi adalah suatu bisnis perlindungan finansial atau dengan kata lain ganti rugi secara finansial untuk jiwa, kesehatan, properti dan lainnya untuk mendapatkan penggantian kerugian dari kejadian-kejadian yang tak terduga yang dapat terjadi pada setiap orang seperti kematian, kehilangan, kerusakan atau sakit, yang mana melibatkan pembayaran premi secara teratur, seperti bulanan ataupun tahunan dalam jangka waktu tertentu sebagai ganti polis yang menjamin perlindungan tersebut (Rahmadyarti, dkk., 2017).

Terdapat dua pihak yang terlibat didalam asuransi, ada tertanggung dan penanggung. Istilah tertanggung pada asuransi menunjuk pada orang atau pihak yang memperoleh jaminan penggantian kerugian dari penyedia asuransi ketika terjadi risiko-risiko yang tidak diinginkan yang tertulis dan disetujui oleh tertanggung dan penanggung yang ada di dalam perjanjian polis asuransi serta tertanggung membayarkan premi ke penanggung untuk membalas jasanya yang berupa pembayaran uang yang jumlahnya relatif kecil dan dibayarkan secara periodik kepada penanggung. Sedangkan penanggung merupakan orang yang menjamin dirinya akan menanggung kerugian-kerugian dari kejadian-kejadian yang tidak diinginkan yang dihadapi oleh tertanggung.

Terdapat tiga elemen dalam asuransi, yaitu:

1. Premi

Premi adalah pembayaran yang wajib dibayarkan oleh pihak tertanggung atau nasabah kepada pihak penanggung atau penyedia layanan asuransi selama jangka waktu yang telah disepakati oleh kedua belah pihak sebagai jasa pengalihan risiko. Pada umumnya premi dibayarkan secara bulanan, semesteran hingga tahunan. Pembayaran premi wajib dibayarkan oleh pihak tertanggung kepada pihak penanggung untuk mendapatkan manfaat asuransi pada saat diperlukan.

2. Polis Asuransi

Polis asuransi merupakan sebuah dokumen yang legal dan sah yang menjadi dasar hukum yang menghubungkan pihak tertanggung dan pihak penanggung yang berisi dan mengatur tentang perjanjian asuransi yang telah disepakati oleh kedua belah pihak dan harus dibuat secara tertulis. Polis asuransi bertindak sebagai dasar untuk pemembayaran biaya ganti rugi atas kehilangan atau kerusakan yang dialami oleh pihak tertanggung.

3. Klaim

Klaim merupakan proses permohonan atau pengajuan resmi yang diajukan oleh pihak tertanggung atau nasabah kepada pihak perusahaan asuransi untuk mendapatkan pembayaran sebagai bentuk ganti rugi atas kehilangan atau kerusakan berdasarkan ketentuan polis asuransi yang sudah disepakati oleh kedua belah pihak secara resmi dan tertulis. Untuk mendapatkan pembayaran atau ganti rugi, penanggung harus memeriksa validitas klaim terlebih dahulu untuk menghindari kecurangan-kecurangan yang mungkin terjadi.

Terdapat beberapa jenis asuransi yang ditawarkan oleh perusahaan asuransi yang dibutuhkan oleh masyarakat. Berikut merupakan jenis-jenis asuransi:

1. Asuransi Jiwa

Asuransi jiwa merupakan asuransi yang digunakan untuk perlindungan dari kerugian finansial yang tak terduga, kehilangan pendapatan seseorang yang dikarenakan terdapat kematian anggota keluarga tertanggung yang menjadi tulang punggung bagi keluarga.

2. Asuransi Kesehatan

Asuransi kesehatan adalah asuransi yang memberikan jaminan dari pihak perusahaan asuransi kepada pihak tertanggung untuk mengganti biaya-biaya yang dibutuhkan untuk kesehatan tertanggung. Asuransi kesehatan dapat *cover* biaya-biaya pengobatan.

3. Asuransi Pendidikan

Asuransi pendidikan adalah jenis asuransi yang memberikan proteksi atas pendidikan anak di masa yang akan mendatang dan yang sekaligus investasi bagi pihak terjamin yang sekaligus menjadi ahli waris pemegang polis. Biasanya asuransi pendidikan sudah mencakup asuransi jiwa juga.

4. Asuransi Kendaraan

Kendaraan merupakan salah satu barang yang berharga yang perlu dilindungi. Peristiwa yang tak terduga dapat terjadi kapan saja dan dimana saja, seperti kecelakaan atau kehilangan. Pemilik kendaraan yang memberikan kendaraannya asuransi setidaknya akan lebih tenang apabila terjadi hal-hal yang tidak terduga pada kendaraannya. Karena dengan memberikan kendaraan asuransi, pihak perusahaan asuransi ikut serta membantu untuk meringankan kerugian.

5. Asuransi Kecelakaan

Kecelakaan merupakan musibah yang tidak dapat diprediksi. Seseorang dapat mengalami kecelakaan dimanapun dan kapanpun, misalnya kecelakaan ditempat kerja atau dijalan saat berkendara. Asuransi kecelakaan merupakan asuransi yang menjamin kerugian atas risiko cacat permanen ataupun meninggal dunia akibat kecelakaan dengan menanggung biaya perawatannya.

6. Asuransi Korporasi

Karyawan merupakan salah satu aset yang berharga bari perusahaan. Oleh karena itu, karyawan biasanya mendapatkan perlindungan dengan asuransi korporasi. Asuransi korporasi data berupa asuransi jiwa berkelompok maupun asuransi kesehatan berkelompok yang anggotanya merupakan karyawan tetap di perusahaan.

7. Asuransi Hari Tua

Ada beberapa jenis pekerjaan yang tidak mendapatkan dana pensiun. Oleh karena itu, diciptakannya asuransi hari tua agar orang yang tidak mendapatkan dana pensiun dapat memperoleh uang hasil investasinya untuk masa tua mereka.

8. Asuransi Perjalanan

Asuransi perjalanan adalah sebuah asuransi yang akan memberikan perlindungan untuk risiko yang mungkin akan terjadi ketika seseorang sedang dalam perjalanan. Asuransi perjalanan berbeda dengan asuransi kecelakaan, karena asuransi perjalanan mencakup asuransi kecelakaan dan otomatis mencakup asuransi jiwa, sedangkan asuransi kecelakaan belum tentu memberikan perlindungan terhadap risiko atau peristiwa buruk dalam perjalanan, contohnya keterlambatan keberangkatan. Asuransi perjalanan sangat penting untuk setiap orang yang sering berpergian.

2.2 Klasifikasi

Klasifikasi merupakan sebuah kata yang berasal dari bahasa Belanda, yaitu '*classificatie*'. Kemudian kata '*classificatie*' tersebut berasal dari Bahasa Prancis, yaitu '*classification*' yang berarti 'klasifikasi' atau 'pengelompokan'. Klasifikasi merupakan sebuah teknik pengelompokan atau penggolongan untuk mengelompokkan sebuah data dengan ciri-ciri atau karakteristik yang serupa sesuai dengan aturan yang ditetapkan. Secara harfiah, klasifikasi merupakan pengelompokan segala sesuatu berdasarkan kelas-kelasnya. Terdapat banyak algoritma yang digunakan untuk mengklasifikasi data, yaitu *K-Nearest Neighbor* (KNN), *Fuzzy*, *Support Vector Machine* (SVM), dan *Artificial Neural Network* (Wafiyah, dkk., 2017).

Klasifikasi memiliki tujuan, yaitu:

1. Dapat menentukan karakteristik data sehingga memudahkan pengelompokan.
2. Mengumpulkan data yang memiliki karakteristik yang sama dalam satu kelas.

Klasifikasi memiliki manfaat, yaitu:

1. Dapat mempermudah mengenali, membandingkan dan mempelajari karakteristik data.
2. Mengelompokkan data berdasarkan persamaan dan perbedaan karakteristik.
3. Informasi yang terdapat pada data dapat dipecah-pecah menjadi kategori yang relatif sedikit.

2.3 Artificial Intelligence (AI)

Artificial Intelligence atau kecerdasan buatan adalah kecerdasan yang ditambahkan, ditanamkan dan diatur oleh manusia ke dalam suatu teknologi untuk berpikir seperti manusia yang akan meniru tindakan mereka. Teknologi ini yang akan dikembangkan untuk mengolah dan menafsirkan data dengan benar. *Artificial Intelligence* dapat menyelesaikan suatu masalah secara optimal sehingga dapat menentukan sendiri Tindakan terbaik yang harus diambil. Penelitian terdahulu menyatakan *Artificial Intelligence* merupakan sebuah sistem yang dirancang agar dapat berinteraksi kepada manusia menggunakan suatu kemampuan khusus serta kebiasaan seperti manusia (Goralski & Tan, 2020).

Munculnya *Artificial Intelligence* merupakan salah satu kemajuan penting dalam kehidupan manusia karena efektivitas pekerjaan manusia semakin tercapai. Sistem ini mampu bekerja dengan akurat dan cepat sehingga dapat meminimalisir kesalahan-kesalahan yang dapat disebabkan oleh manusia, seperti kelelahan dan *human error*. *Artificial Intelligence* banyak digunakan untuk memecahkan berbagai masalah seperti bisnis, robotika, matematika, diagnosis medis, analisis keuangan, sains, game, persepsi dan penalaran. Bukan hanya tentang robot, tetapi *Artificial Intelligence* juga tentang memahami sifat tindakan serta pemikiran yang cerdas menggunakan komputer sebagai perangkat eksperimental. *Artificial Intelligence* sudah banyak digunakan pada kehidupan sehari-hari di berbagai bidang industri diantaranya industri kesehatan, pendidikan, manufaktur, bisnis, jasa dan produk (Sunarya, dkk., 2015).

2.4 Machine Learning

Machine Learning didefinisikan sebagai algoritma matematika dan aplikasi komputer yang dijalankan dengan proses pembelajaran dari beberapa data yang akan menghasilkan prediksi yang bermanfaat untuk masa mendatang. Pembelajaran dapat dilakukan dengan dua proses, yaitu *training* dan *testing*. *Machine learning* mampu belajar dengan sendirinya agar dapat memutuskan suatu keputusan tanpa harus berulang kali diprogram oleh manusia sehingga komputer dapat menjadi lebih cerdas karena belajar dari pengalaman data-data yang dimiliki. Aplikasi ilmu kecerdasan buatan ini dapat digunakan pada berbagai macam industri di dunia dan terus digunakan oleh industri-industri besar dan para peneliti agar terus berkembang (Huang, dkk., 2006).

Terdapat beberapa cara belajar *machine learning* agar dapat menganalisa data dengan baik berdasarkan data yang diberikan saat awal pengembangan. Berikut merupakan teknik *machine learning*.

1. *Supervised Machine Learning*

Supervised machine learning merupakan teknik yang dapat digunakan untuk menerima informasi dan menganalisis data yang telah memiliki beberapa kelas. Artinya klasifikasi datanya sudah jelas dan pada algoritmanya hanya tinggal memprediksi pola datanya saja.

2. *Unsupervised Machine Learning*

Unsupervised machine learning merupakan suatu teknik yang dapat diaplikasikan pada data yang tidak mempunyai informasi yang dapat diterapkan yang tidak mempunyai kelas. Teknik ini akan dihadapkan dengan sekumpulan data mentah kemudian datanya akan dianalisis untuk menemukan pola data atau struktur tersembunyi pada data yang tidak memiliki kelas atau label.

3. *Semi-Supervised Machine Learning*

Semi-supervised machine learning merupakan teknik yang dapat digunakan apabila ingin menggunakan teknik *supervised learning* tetapi pelabelan data yang dimiliki tidak cukup. Teknik ini dapat menggabungkan jumlah data yang berlabel yang jumlahnya sedikit dan data yang tidak berlabel yang jumlahnya banyak.

2.5 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan salah satu metode yang digunakan dalam pengklasifikasian dan merupakan salah satu metode yang menggunakan teknik *supervised machine learning* yang mencari jarak terdekat terhadap data yang akan diuji dengan data k tetangga terdekatnya dalam data latih. Metode *K-Nearest Neighbor* merupakan algoritma yang sangat sederhana dan tidak memperhatikan distribusi dari data yang ingin dikelompokkan. Algoritma *K-Nearest Neighbor* menerapkan *lazy learning* atau *instant based learning*, yang artinya tidak menggunakan titik data *training* untuk membuat sebuah model dan merupakan algoritma non parametrik, yaitu tidak membuat asumsi apapun terhadap data. Algoritma ini dapat digunakan untuk klasifikasi dan regresi (Sadli, dkk., 2018).

2.5.1 Perhitungan Jarak

Terdapat beberapa cara penghitungan jarak pada algoritma *K-Nearest Neighbor* untuk menghitung jarak dari data baru ke data latih pada sekumpulan data sebagai berikut.

1. Jarak *Euclidean*

Pada umumnya, untuk menghitung jarak antar dua data pada algoritma *K-Nearest Neighbor* menggunakan penghitungan jarak *euclidean*. Jarak *euclidean* merupakan salah satu metode untuk penghitungan untuk mengukur jarak antar dua titik yang ada dalam *euclidean space* yang mempelajari hubungan antara sudut dan jarak. Jarak *euclidean* dapat digunakan untuk mengukur dua titik dalam satu dimensi seperti penghitungan *Pythagoras*. Jarak *euclidean* dapat digunakan pada *1-dimensional space*, *2-dimensional space* dan *multi-dimensional space* atau lebih dari dua variabel (Mustofa & Suasana, 2018).

Secara umum, formula *euclidean distance* pada *1-dimensional space* sebagai berikut.

$$d_{12} = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2}$$

Keterangan :

d_{12} = Jarak antara objek 1 dengan 2

x_{1i} = Nilai objek 1 pada variabel ke- i

x_{2i} = Nilai objek 2 pada variabel ke- i

n = Banyaknya variabel yang diamati

Formula tersebut dapat digunakan apabila jumlah variabel bebas pada data hanya terdapat satu variabel. Ketika terdapat dua atau lebih variabel bebas pada data akan digunakan formula sebagai berikut.

$$d_{12} = \sqrt{(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots}$$

Keterangan :

d_{12} = Jarak antara objek 1 dengan 2

x_{1i} = Nilai objek 1 pada variabel ke- i

x_{2i} = Nilai objek 2 pada variabel ke- i

y_{1i} = Nilai objek 1 pada variabel ke- i

y_{2i} = Nilai objek 2 pada variabel ke- i

Berikut merupakan contoh perhitungan jarak *Euclidean* menggunakan 8 variabel. Diberikan 3 data *train* dan 1 data *test* sebagai berikut.

Tabel 1. Contoh Data *Train* dan Data *Test*.

Data	Age	Employment Type	Graduate OrNot	Annual Income	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Travel Insurance
Train	31	1	1	400000	6	1	0	0	No
Train	31	2	1	1250000	7	0	0	0	No
Train	34	2	1	500000	4	1	0	0	Yes
Test	33	2	1	1400000	4	0	1	1	No

Didapatkan formula sebagai berikut.

$$d_{12} = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + (x_{13} - x_{23})^2 + (x_{14} - x_{24})^2 + \sqrt{(x_{15} - x_{25})^2 + (x_{16} - x_{26})^2 + (x_{17} - x_{27})^2 + (x_{18} - x_{28})^2}}$$

Jarak data *train* 1 dengan data *test* 1

$$\begin{aligned} d_{11} &= \sqrt{(1 - 2)^2 + (31 - 33)^2 + (1 - 1)^2 + (400.000 - 1.400.000)^2 + \sqrt{(6 - 4)^2 + (1 - 0)^2 + (0 - 1)^2 + (0 - 1)^2}} \\ &= 1.000.000,000006 \end{aligned}$$

Jarak data *train* 2 dengan data *test* 1

$$\begin{aligned} d_{12} &= \sqrt{(2 - 2)^2 + (31 - 33)^2 + (1 - 1)^2 + (1.250.000 - 1.400.000)^2 + \sqrt{(7 - 4)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 1)^2}} \\ &= 150.000,000005 \end{aligned}$$

Jarak data *train* 3 dengan data *test* 1

$$\begin{aligned} d_{13} &= \sqrt{(2 - 2)^2 + (34 - 33)^2 + (1 - 1)^2 + (500.000 - 1.400.000)^2 + \sqrt{(4 - 4)^2 + (1 - 0)^2 + (0 - 1)^2 + (0 - 1)^2}} \\ &= 900.000,000002222 \end{aligned}$$

2. Jarak *Manhattan* atau *Cityblock*

Jarak *Manhattan* digunakan untuk menghitung perbedaan absolut (mutlak) antara koordinat sepasang objek. Jarak *Manhattan* merupakan metode perhitungan jarak pada ruang dengan menerapkan konsep selisih mutlak (Buaton, dkk., 2016). Penghitungan jarak *manhattan* akan menghasilkan jarak yang mirip dengan jarak *euclidean* pada beberapa kasus. Kelebihan jarak ini yaitu dapat mendeteksi keberadaan *outlier* dengan baik (Kurniawan & Mufti, 2018).

Secara umum, formula jarak *manhattan* sebagai berikut.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Keterangan :

$d(x, y)$ = Jarak antara x dan y

x_i = Nilai objek x pada variabel ke- i

y_i = Nilai objek y pada variabel ke- i

n = Banyaknya variabel yang diamati

Jarak *Manhattan* memiliki tingkat akurasi yang lebih lemah dibandingkan dengan jarak *euclidean* dengan persentase tertinggi yaitu dapat mencapai 100%, sedangkan jarak *manhattan* hanya dapat mencapai 80%.

Berikut merupakan contoh perhitungan jarak *Manhattan* menggunakan 8 variabel berdasarkan data pada Tabel 1.

Jarak data *train* 1 dengan data *test* 1

$$\begin{aligned} d_{11} &= |1 - 2| + |31 - 33| + |1 - 1| + |400.000 - 1.400.000| + |6 - 4| + \\ &\quad |1 - 0| + |0 - 1| + |0 - 1| \\ &= 1.000.008 \end{aligned}$$

Jarak data *train* 2 dengan data *test* 1

$$\begin{aligned} d_{12} &= |2 - 2| + |31 - 33| + |1 - 1| + |1.250.000 - 1.400.000| + |7 - 4| + \\ &\quad |0 - 0| + |0 - 1| + |0 - 1| \\ &= 150.007 \end{aligned}$$

Jarak data *train* 3 dengan data *test* 1

$$\begin{aligned} d_{13} &= |2 - 2| + |34 - 33| + |1 - 1| + |500.000 - 1.400.000| + |4 - 4| + \\ &\quad |1 - 0| + |0 - 1| + |0 - 1| \\ &= 900.004 \end{aligned}$$

3. Jarak *Minkowski*

Jarak *Minkowski* merupakan sebuah metrik dalam ruang vektor dimana suatu norma didefinisikan sekaligus dianggap sebagai *generalisasi* dari jarak *Euclidean* dan jarak *Manhattan*. Pada pengukuran jarak objek jarak ini biasanya menggunakan nilai p yaitu 1 atau 2 (Nishom, 2019).

Secara umum, formula jarak *minkowski* sebagai berikut.

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

Keterangan :

$d(x, y)$ = Jarak antara x dan y

x_i = Nilai objek x pada variabel ke- i

y_i = Nilai objek y pada variabel ke- i

n = Banyaknya variabel yang diamati

p = Power

Berikut merupakan contoh perhitungan jarak *Minkowski* dengan $p = 3$ menggunakan 8 variabel berdasarkan data pada Tabel 1.

Jarak data *train* 1 dengan data *test* 1

$$\begin{aligned} d_{11} &= [(1 - 2)^3 + (31 - 33)^3 + (1 - 1)^3 + (400.000 - 1.400.000)^3 + \\ &\quad (6 - 4)^3 + (1 - 0)^3 + (0 - 1)^3 + (0 - 1)^3]^{1/3} \\ &= -3,33333E + 17 \end{aligned}$$

Jarak data *train* 2 dengan data *test* 1

$$\begin{aligned} d_{12} &= [(2 - 2)^3 + (31 - 33)^3 + (1 - 1)^3 + (1.250.000 - 1.400.000)^3 + \\ &\quad (7 - 4)^3 + (0 - 0)^3 + (0 - 1)^3 + (0 - 1)^3]^{1/3} \\ &= -1,125E + 15 \end{aligned}$$

Jarak data *train* 3 dengan data *test* 1

$$\begin{aligned} d_{12} &= [(2 - 2)^3 + (34 - 33)^3 + (1 - 1)^3 + (500.000 - 1.400.000)^3 + \\ &\quad (4 - 4)^3 + (1 - 0)^3 + (0 - 1)^3 + (0 - 1)^3]^{1/3} \\ &= -2,43E + 17 \end{aligned}$$

2.5.2 Confusion Matrix

Confusion matrix merupakan salah satu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep *Machine Learning*. *Confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem

atau model dengan hasil klasifikasi yang sebenarnya. Dalam *confusion matrix* terdapat istilah *precision* yang artinya tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, *recall* yang artinya tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi, *f1-score* yang merupakan salah satu perhitungan evaluasi dalam informasi temu kembali yang mengkombinasikan *recall* dan *precision* dan *accuracy* yang artinya tingkat kedekatan antara nilai prediksi dengan nilai aktual (Putra, dkk., 2021).

Pengujian akurasi digunakan untuk menemukan persentase ketepatan dalam proses pengklasifikasian terhadap data testing yang diuji. *Confusion matrix* merupakan *tool* yang dapat digunakan untuk mengevaluasi model klasifikasi agar dapat memperkirakan objek yang benar atau salah. Sebuah matrik dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi (Mustakim & Oktaviani, 2016).

Terdapat 4 istilah sebagai representasi hasil klasifikasi pada *confusion matrix*, yaitu:

1. *True Positive* (TP), yaitu data positif yang diprediksi benar.
2. *True Negative* (TN), yaitu data negatif yang diprediksi benar.
3. *False Positive* (FP) – *Type I Error*, yaitu data negatif namun diprediksi sebagai data positif.
4. *False Negative* (FN) – *Type II Error*, yaitu data positif namun diprediksi sebagai data negatif.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <small>Type I Error</small>
	0 (Negative)	FN (False Negative) <small>Type II Error</small>	TN (True Negative)

Gambar 1. *Confusion Matrix*

Pada beberapa kasus, *Type II Error* lebih berbahaya karena data yang sebenarnya yaitu data yang positif tetapi prediksinya menunjukkan data tersebut negatif. Misalnya pada kasus *covid-19*, terdapat pasien yang terbukti positif *covid-19* akan tetapi prediksi datanya salah, pada prediksi menunjukkan bahwa pasien negatif terkena *covid-19*. Pasien tersebut tidak akan menyadari bahwa sebenarnya dia positif dan pasien itu tetap beraktivitas seperti biasanya. Pasien tersebut dapat menyebarkan virus *covid-19* kepada keluarga, kerabat dan orang-orang disekitarnya, seperti tidak menjaga jarak dengan keluarga di rumah, tidak mencuci tangan sehabis bersin bahkan ketika makan dengan menggunakan peralatan makan yang sama dengan keluarga di rumah. Oleh karena itu, dapat dikatakan bahwa *Type II Error* lebih berbahaya.

Tabel 2. Interpretasi nilai *Error*.

Nilai <i>Error</i>	Interpretasi
< 10%	Klasifikasi dengan akurasi tinggi
10% – 20%	Klasifikasi yang baik
20% – 50%	Klasifikasi yang layak
> 50%	Klasifikasi yang tidak akurat

Berikut merupakan formula yang terdapat pada *confusion matrix*.

$$Precision = \frac{TP}{(TP+FP)} \times 100\%$$

$$Recall = \frac{TP}{(TP+FN)} \times 100\%$$

$$F1\ Score = \frac{2 \times precision \times recall}{(precision+recall)}$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100\%$$

$$Error = \frac{(FP+FN)}{(TP+TN+FN+FP)} \times 100\%$$

Keterangan :

TP = True Positive

FP = False Positive

TN = True Negative

FN = False Negative

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun akademik 2022/2023 yang bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data *Travel Insurance Prediction* di India. Jumlah data yang digunakan pada penelitian ini sebanyak 1.987 data dan terdapat 9 variabel yaitu, *employment type*, *age*, *graduate or not*, *annual income*, *family members*, *chronic diseases*, *frequent flyer*, *ever travelled abroad*, dan *travel insurance*. Data yang digunakan pada penelitian ini diperoleh dari *website* Kaggle.

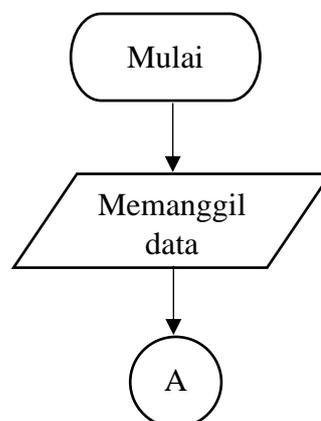
3.3 Metode Penelitian

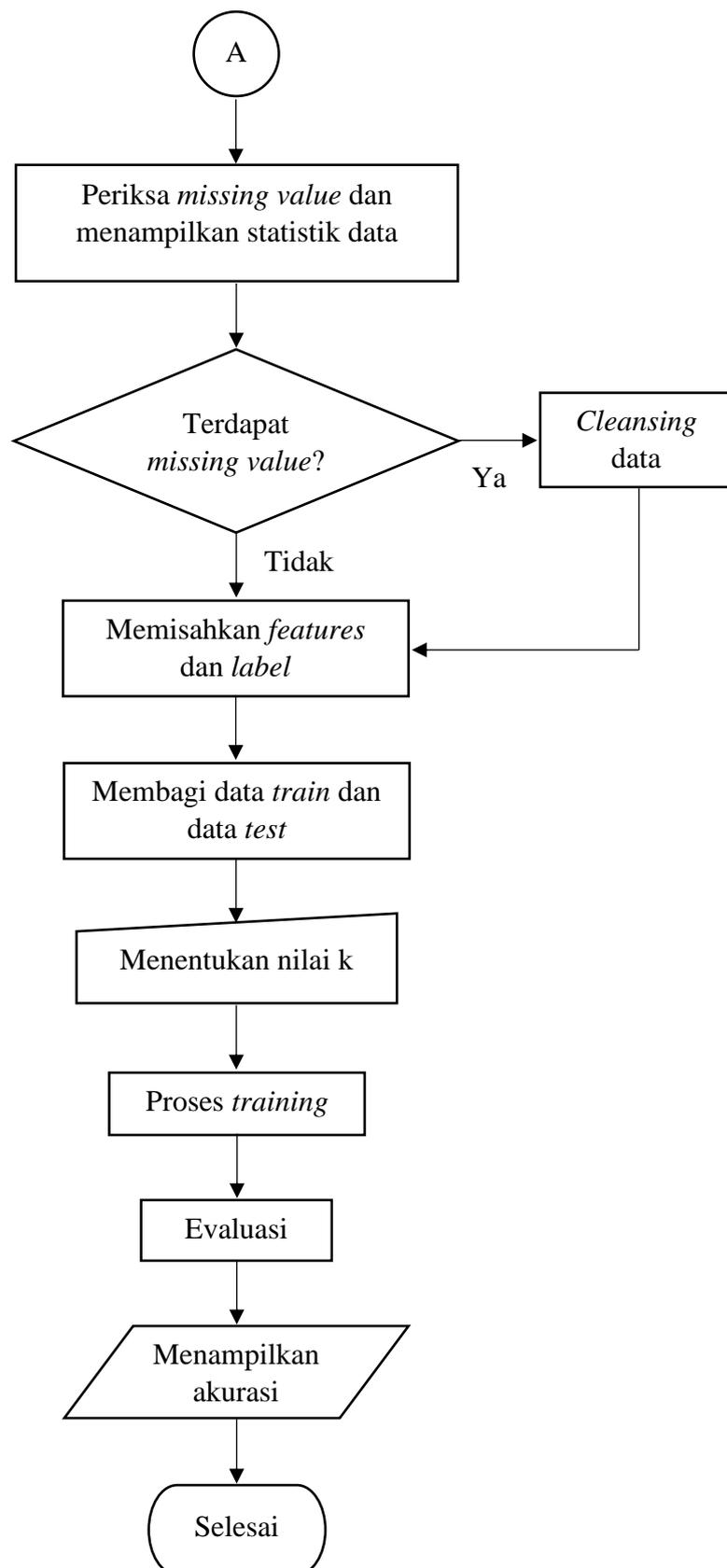
Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut:

1. Memanggil data *Travel Insurance Prediction* dengan format *Comma Separated Value* (CSV) ke dalam Python dengan bantuan Google Colaboratory.

2. Memeriksa *missing values* dan menampilkan statistik data.
 - a. Apabila terdapat *missing values*, dapat menghapus baris yang terdapat *missing values* tersebut atau mengganti *missing values* dengan rata-rata variabel yang terdapat *missing values* tersebut.
 - b. Jika tidak terdapat *missing values*, maka lanjutkan ke proses selanjutnya.
3. Memisahkan data *features* atau variabel bebas, yaitu *employment type, age, graduate or not, annual income, family members, chronic diseases, frequent flyer, ever travelled abroad* dan data *label* atau variabel terikat, yaitu *travel insurance*.
4. Membagi data *train* dan data *test* dengan pembagian :
 - a. 70% data *train* dan 30% data *test*
 - b. 80% data *train* dan 20% data *test*
 - c. 90% data *train* dan 10% data *test*
5. Mencari dan menentukan parameter *k*, yaitu jumlah tetangga terdekat yang optimal pada data latih dengan melihat *error rate* terkecil dari masing-masing pembagian data *train* dan data *test* sebelumnya.
6. Melakukan *training* pada data *train* dengan menggunakan parameter *k* yang telah ditentukan dengan menggunakan algoritma K-Nearest Neighbor.
7. Melakukan evaluasi dengan menampilkan laporan klasifikasi dan *confusion matrix*.
8. Menampilkan akurasi model data uji.

Berikut merupakan flowchart klasifikasi menggunakan algoritma K-Nearest Neighbor.





V. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dari analisis data *Travel Insurance Prediction* diperoleh nilai *f1-score* dari model dalam memprediksi data yang positif sebesar 0.72 dan *f1-score* dari model dalam memprediksi data yang negatif sebesar 0.88. Sedangkan untuk akurasi model didapatkan 83% . Artinya penelitian ini dapat digunakan sebagai acuan agar dapat membantu perusahaan asuransi perjalanan dalam melakukan penawaran ke target ke target yang tepat dan maksimal kepada masyarakat yang masuk kedalam klasifikasi membeli asuransi perjalanan.

5.2 Saran

Berikut merupakan saran yang dibuat untuk pengembangan penelitian selanjutnya:

1. Dapat menggunakan algoritma pengklasifikasian yang lain agar dapat membandingkan tingkat akurasinya.
2. Dapat mengganti dengan variabel lain atau menambah variabel.

DAFTAR PUSTAKA

- Asri, K. N., Saptono, H. & Njatrijani, R. 2017. Pelaksanaan Asuransi Sosial pada PT. Jasa Raharja (Persero) terhadap Korban Kecelakaan Lalu Lintas di Kota Semarang. *Diponegoro Law Journal*. **6**(2): 1-17.
- Betris. 2017. Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naïve Bayes Berbasis Backward Elimination. *ILKOM Jurnal Ilmiah*. **9**(1): 96-101.
- Buaton, R., Sundari, Y., & Maulita, Y. 2016. Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean. *Media Informasi Analisa dan Sistem*. **1**(2): 47-53
- Goralski, M. A. & Tan, T. K. 2020. Artificial Intelligence and Sustainable Development. *International Journal of Management Education*. **18**(1): 1-14.
- Huang, G. B., Zhu, Q. Y., & Siew, C. K. 2006. Extreme Learning Machine: Theory and Applications. *Neurocomputing*. **70**(1-3): 489-501.
- Kurniawan, E. D., & Mufti, M. 2018. Implementasi Algoritma K-Nearest Neighbor Dengan Metode Klasifikasi dan Pengukuran Jarak Manhattan Distance untuk Prediksi Kelulusan UN Berdasarkan Hasil Nilai TryOut Berbasis Java Desktop Pada SMA Harapan Jaya 2. *SKANIKA*. **1**(1): 76-81.

- Megasari, P. N. 2022. Polri Catat 6.707 Kasus Kecelakaan Sepanjang 2022, 452 Orang Tewas. <https://news.detik.com/berita/d-6325521/polri-catat-6707-kasus-kecelakaan-sepanjang-2022-452-orang-tewas#:~:text=Firman%20memaparkan%20jumlah%20kasus%20kecelakaan,material%20sekitar%20Rp13%20miliar%20lebih>. Diakses pada 4 November 2022.
- Mustafa, M. S. & Simpen, I. W. 2014. Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIK Dipanegara Makassar). *Creative Information Technology Journal*. **1**(1): 270-281.
- Mustakim, M., & Oktaviani, G. F. 2016. Algoritma K-Nearest Neighbor Classification sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. *Jurnal Sains, Teknologi dan Industri*. **13**(2): 195-202.
- Mustofa, Z., & Suasana, I. S. 2018. Algoritma Clustering K-Medoids pada E-Government Bidang Information and Communication Technology dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi dan Komunikasi*. **9**(1): 1-10.
- Nishom, M. 2019. Perbandingan Akurasi Eulidean Distance, Minkowski Distance dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*. **4**(1): 20-24.
- Putra, T. W., Triayudi, A., & Andrianingsih. 2021. Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, KNN dan Decision Tree. *Jurnal Teknologi Informasi dan Komunikasi*. **6**(1): 20-26.
- Rahmadyarti, A., Njatrijani, R. & Saptono, H. 2017. Pelaksanaan Asuransi Jiwa Terhadap Penumpang Online (Studi PT. GO-JEK Indonesia dan PT. Asuransi Allianz Utama Indonesia). *Diponegoro Law Journal*. **6**(2): 1-13.

- Retnaningtias, N., Efendi, R., & Susilo, B. 2015. Aplikasi Biometrika Pencocokan Citra Garis Telapak Tangan Dengan Metode Transformasi Wavelet dan Mahalanobis Distance. *Jurnal Rekursif*. **3**(2): 95-106.
- Sadli, M., Fajriana, F., Fuadi, W., Ermatita, E., & Pahendra, I. 2018. Penerapan Model K-Nearest Neighbor dalam Klasifikasi Kebutuhan Daya Listrik Untuk Masing-Masing Daerah di Kota Lhokseumawe. *Jurnal ECOTIPE*. **5**(2): 11-18.
- Sunarya, A., Santoso, S., & Sentanu, W. 2015. Sistem Pakar untuk Mendiagnosa Gangguan Jaringan Lan. *CCIT Journal*. **8**(2): 1-11.
- Wafiyah, F., Hidayat, N., & Perdana, R. S. 2017. Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **1**(10): 1210-1219.