



Klasifikasi Kategori Notifikasi Pengiriman Pesan *One-Time Password*

Novendri Isra Asriny

21917037

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer

Konsentrasi Sains Data

Program Studi Informatika Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia

2024

Lembar Pengesahan Pembimbing

Klasifikasi Kategori Notifikasi Pengiriman Pesan *One-Time Password*

Novendri Isra Asriny

21917037



الجامعة الإسلامية
الاندونيسية

Pembimbing 1

Pembimbing 2

Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D.

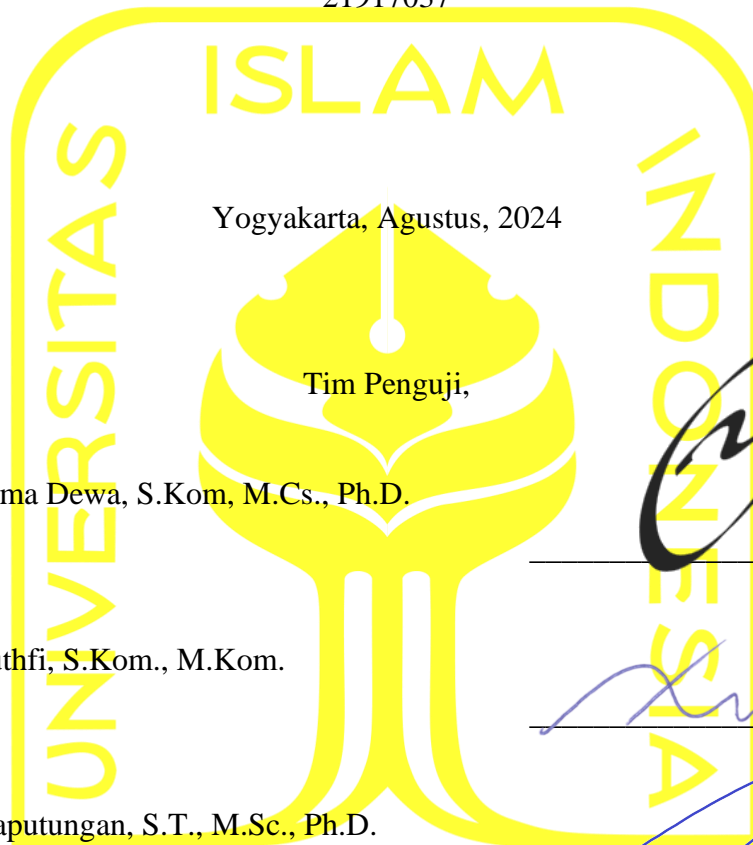
Dr. Ahmad Luthfi, S.Kom., M.Kom.

Lembar Pengesahan Penguji

Klasifikasi Kategori Notifikasi Pengiriman Pesan *One-Time Password*

Novendri Isra Asriny

21917037



Yogyakarta, Agustus, 2024

Tim Penguji,

Chandra Kusuma Dewa, S.Kom, M.Cs., Ph.D.

Ketua

Dr. Ahmad Luthfi, S.Kom., M.Kom.

Anggota I

Irving Vitra Paputungan, S.T., M.Sc., Ph.D.

Anggota II

09/09/2024

الجامعة الإسلامية
Indonesia
Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Dr. Irving Vitra Paputungan, S.T., M.Sc.

Abstrak

Klasifikasi Kategori Notifikasi Pengiriman Pesan *One-Time Password*

Pada ekonomi modern saat ini, pendorong utamanya yakni berasal dari uang digital yang mana secara global mengadopsi *smartphone* yang memudahkan pengguna untuk mengakses layanan keuangan digital. Dengan perkembangan internet yang sangat pesat pada saat ini, banyak muncul penyedia jasa seperti penyedia jasa transportasi, jasa toko *online* dan banyak lagi jasa-jasa yang memanjakan pengguna untuk semakin tidak dapat terlepas dari penggunaan *smartphone*. Seiring dengan hal tersebut, transaksi *online* pun meningkat pesat saat ini, berbagai macam penyedia jasa pembayaran elektronik seperti *credit/debit card*, *e-wallet* dan masih banyak macam pembayaran kontak tanpa langsung. Hal ini, mengakibatkan terdapat sejumlah masalah keamanan terkait kata sandi dan meluasnya pencurian kata sandi dalam transaksi *online*. Akibatnya, banyak bisnis mencari alternatif untuk menanggulangnya. Untuk meminimalisir hal tersebut, muncul pendekatan alternatif *one-time password*. *One-time password* (OTP) merupakan skema otentikasi terhadap kata sandi atau penambahan validasi yang digunakan untuk setiap sesi otentikasi pengguna. Apabila sesi otentikasi telah habis atau digunakan, maka kata sandi tersebut tidak valid lagi. Kemudian, apabila terdapat pengiriman ulang pada *one-time password* maka yang dapat digunakan untuk sesi otentikasi selanjutnya adalah kode otentikasi terbaru untuk digunakan. Dalam pengiriman SMS *one-time password* sangat mungkin terjadi kegagalan pada pengirimannya, baik dikarenakan jaringan yang mana permasalahan muncul dari operator atau permasalahan teknis seperti *request time out* pada server oleh pihak ketiga yang mengirimkan pesan atau bahkan perusahaan yang menyediakan jasa pengiriman SMS OTP. Hasil penelitian yang telah dilakukan, diketahui bahwa pada metode *bayesian network* (BN) variabel-variabel yang signifikan yakni *delivered*, *undelivered*, *average notification per minutes*, *rate undelivered*. Sedangkan untuk status yang setidaknya memiliki 3 *child nodes* yakni *not delay*, *unknown summary*, *long delay*, *normal*, *maybe issue from operator* dan *not normal*. Kemudian dilanjutkan pengujian pada metode *machine learning* (*naïve bayes* [NB]) dan *deep learning* (*multi layer perceptron* [MLP]) bahwa secara evaluasi matriks (akurasi) didapatkan MLP dengan variabel terpilih dari BN lebih tinggi sedikit dibandingkan NB. Hasil dari metode NB dengan variabel terpilih BN sebesar 87,74% dengan waktu komputasi selama 0,006 detik. Pada MLP variabel terpilih BN sebesar 88,12% dengan waktu komputasi selama 15,82 detik. Walaupun waktu latih MLP lebih lama, metode *deep learning* menjadi pilihan terbaik karena tingkat akurasi yang lebih tinggi.

Kata kunci

transaksi online, *one-time password*, transaksi SMS, *machine learning*, *deep learning*, *bayesian network*, *naïve bayes*, *multilayer perceptron*.

Abstract

Classification of One-Time Password Message Delivery Notification Categories

In today's modern economy, the main driver comes from digital money, where the global adoption of smartphones makes it easier for users to access digital financial services. With the very rapid development of the internet at this time, many service providers have emerged, such as transportation service providers, online shop services, and many other services that pamper users who increasingly cannot be separated from using smartphones. In line with this, online transactions are increasing rapidly nowadays, with various types of electronic payment service providers such as credit/debit cards, e-wallets, and many other types of non-direct contact payments. This has resulted in a number of security problems related to passwords and widespread password theft in online transactions. As a result, many businesses are looking for alternatives to cope. To minimize this, an alternative one-time password approach has emerged. A one-time password (OTP) is a password authentication scheme or additional validation used for each user authentication session. If the authentication session has expired or been used, the password is no longer valid. Then, if there is a re-transmission of the one-time password, what can be used for the next authentication session is the latest authentication code to use. When sending a one-time password SMS, it is possible for failure to occur in the delivery, either due to network problems arising from the operator or technical problems such as requesting a time out on the server by the third party sending the message or even the company that provides OTP SMS sending services. The results of the research that has been done show that in the Bayesian Network (BN) method, the significant variables are delivered, undelivered, average notification per minute, and rate undelivered, for the status that has at least three child nodes, namely not delay, unknown summary, long delay, normal, maybe issue from the operator, and not normal. Then, continued testing on the machine learning method (naïve Bayes [NB]) and deep learning (multi-layer perceptron [MLP]) that, in terms of matrix evaluation (accuracy), obtained MLP with selection variables from BN is slightly higher than NB. The results of the NB method with BN selection variables are 87.74% with a computing time of 0.006 seconds. In MLP, the BN selection variable is 88.12%, with a computing time of 15.82 seconds. Although the MLP training time is longer, the deep learning method is the best choice because of its higher level of accuracy.

Keywords

online transactions, one-time passwords, SMS transactions, machine learning, deep learning, Bayesian network, naïve Bayes, multilayer perceptron.

Pernyataan Keaslian Tulisan

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, Agustus, 2024



Novendri Isra Asriny, S.Stat.

Daftar Publikasi

Publikasi yang menjadi bagian dari tesis

Asriny, N. I., Dewa, C. K., & Luthfi, A. (2024). Modeling OTP Delivery Notification Status through a Causality and Apriori. *Indonesian Journal of Computing and Cybernetics Systems*, 18(1), 61–72.

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Novendri Isra Asriny	Mendesain eksperimen (60%) Menulis <i>paper</i> (70%)
Chandra Kusuma Dewa	Mendesain eksperimen (25%) Menulis dan mengedit <i>paper</i> (15%)
Ahmad Luthfi	Mendesain eksperimen (25%) Menulis dan mengedit <i>paper</i> (25%)

Halaman Kontribusi

Pada penelitian ini terdapat beberapa kontribusi dari pihak lain, yakni :

1. Dhiya Mahdi Asriny, yang telah mendukung penelitian dan menjelaskan terkait bagaimana data *flow* proses bisnis pada layanan jasa pengiriman pesan SMS OTP (*One-Time Password*). Serta menjadi pihak yang memberikan pendeskripsian dan arahan dari sisi praktisi yang berpengalaman terhadap pendekatan metode yang digunakan dalam penelitian ini.
2. Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D. sebagai pembimbing pertama yang memberikan masukan sejak diskusi pertama penelitian ini akan dilakukan, masukan pada alternatif metode yang digunakan dalam publikasi yang mengharuskan metode tambahan sebagai basis metode utama yang digunakan. Hingga, masukan-masukan dalam penyusunan draf tesis yang berdatap meningkatkan kualitas interpretasi hasil penelitian.
3. Dr. Ahmad Luthfi, S.Kom., M.Kom. sebagai pembimbing kedua yang memberikan masukan secara mendalam pada penulisan metode Bayesian Network hingga interpretasi. Masukan yang diberikan sangat membantu dalam interpretasi hasil yang lebih kompleks dan tentunya sangat meningkatkan gaya dan bahasa tulisan pada makalah yang di publikasi.

Halaman Persembahan

Laporan tesis ini penulis persembahkan kepada pihak-pihak yang telah memberikan dukungan, bantuan secara materi dan moral selama masa pendidikan di Program Studi Informatika, Program Magister, Universitas Islam Indonesia

Pertama, kepada diri saya sendiri, Novendri Isra Asriny. Terima kasih untuk segala usaha, pemikiran, ikhtiar yang telah dilakukan dalam perjalanan ini dengan penuh keteguhan dan keyakinan. Setiap langkah perjalanan ini adalah cerminan ketangguhan, cinta, dan rasa syukur yang telah kau semai dalam hati. Teruslah melangkah dengan penuh keberanian dan kasih sayang, karena mereka disekitarmu membutuhkan cahaya dan semangatmu. Tetaplah menjadi bintang yang selalu menerangi langit, memberikan inspirasi dan harapan untuk dirimu sendiri dan orang lain. Teruslah beribadah, bermimpi, berjuang, dan mencintai sepenuh hati. Jaya di laut, darat, dan udara!.

Selanjutnya, orang tua saya. Ayah Masri dan Ibu Neny Desriana (Asriny), terima kasih atas keyakinan kalian yang tulus, kesabaran yang tanpa batas, serta cinta kasih yang melimpah adalah dorongan yang tak ternilai. Doa-doa yang tiada henti menjadi kekuatan bagi saya dalam menyelesaikan pendidikan magister ini dengan baik. Segala upaya dan dukungan kalian adalah kunci utama bagi keberhasilan saya. Terima kasih atas segalanya, Ayah dan Ibu.

Selanjutnya, abang dan adik saya, Dhiya Mahdi Asriny dan Fidela Elvina Asriny yang selalu memberikan dukungan, cinta kasih, motivasi dan selalu mengawal setiap langkah perjalanan pendidikan magister ini. Terima kasih yang tulus! Insha Allah, kita akan terus saling menyayangi, menjaga, dan mendukung satu sama yang lain, agar kita dapat tumbuh dan membanggakan keluarga Asriny lebih lagi.

Selanjutnya, terima kasih sebesar-besarnya kepada Bapak Muhammad Muhajir, S.Si., M.Sc. dan Ibu Dr. techn., S.Si., M.Si. Rohmatul Fajriyah. Dukungan dan nasihat-nasihat berharga kalian telah menjadi pendorong bagi saya dalam menyelesaikan pendidikan magister ini.

Selanjutnya, terima kasih banyak teruntuk Keluarga Juragan Muda (KJM) yang selalu memberikan dukungan penuh dan senyuman yang tak pernah pudar, membantu penulis untuk dapat menyelesaikan pendidikan magister ini dengan penuh semangat.

Dukungan tulus dari semua pihak, sangat berarti bagi penulis dalam menyelesaikan laporan tesis ini. Terima kasih, semoga kita semua selalu dilimpahkan keberkahan, rahmat dan kebahagiaan dari Allah SWT. Aamiin yaa rabbal alamin.

Kata Pengantar

Bismillahirrahmanirrahim,

Assalammualaikum Warrahmatullahi Wabarakatuh.

Terimakasih atas rahmat, ridha dan karunia-Mu yang telah memberikan kesempatan untuk membuat hingga menyelesaikan laporan penelitian (tesis) yang berjudul “**Klasifikasi Kategori Notifikasi Pengiriman Pesan One-Time Password**”.

Laporan tesis ini disusun sebagai syarat untuk memperoleh gelar Magister Komputer (M.Kom) pada Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia, Sleman, Yogyakarta.

Tesis ini dapat terwujud dan terselesaikan atas bantuan, dukungan, bimbingan, harapan serta doa yang telah diberikan oleh banyak pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang setulus-tulusnya kepada :

1. Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D., selaku Ketua Program Studi Magister Informatika Universitas Islam Indonesia, yang telah memberikan bimbingan, fasilitas dan masukan pada sidang progres laporan tesis.
2. Bapak Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D. selaku dosen pembimbing utama yang bersedia memberikan ilmu, arahan, kritik, serta nasihat kepada penulis selama penyusunan laporan tesis serta proses-proses dibelakangnya.
3. Bapak Dr. Ahmad Luthfi, S.Kom., M.Kom. selaku dosen pembimbing kedua yang bersedia memberikan ilmu, arahan, kritik serta nasihat kepada penulis selama penyusunan laporan tesis pada metode bayesian network dan penulisan makalah yang sangat kritis dan terstruktur.
4. Dosen Program Studi Magister Informatika Universitas Islam Indonesia yang telah memberikan ilmu, bimbingan dan bantuan selama masa studi pascasarjana.
5. Seluruh pihak dan rekan-rekan yang telah berkontribusi kepada penulis dalam menyelesaikan laporan tesis ini.

Penulis memohon maaf apabila terdapat kesalahan. Oleh sebab itu, penulis sangat terbuka pada kritik dan saran yang sifatnya membangun untuk selalu dapat bermanfaat bagi pembaca.

Sleman, Juli 2024


Novendri Isra Asriny

Daftar Isi

Lembar Pengesahan Pembimbing	i
Lembar Pengesahan Penguji.....	ii
Abstrak.....	iii
Abstract.....	iv
Pernyataan Keaslian Tulisan	v
Daftar Publikasi	vi
Halaman Kontribusi.....	vii
Halaman Persembahan	viii
Kata Pengantar.....	ix
Daftar Isi.....	x
Daftar Tabel.....	xii
Daftar Gambar	xiv
Glosarium	xv
1.1 Pendahuluan	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	7
1.6 Metodologi Penelitian	7
1.7 Sistematika Penulisan	8
2.1 Konsep Pengetahuan	10
2.1.1 Klasifikasi Pengiriman Pesan SMS OTP.....	10
2.1.2 <i>Machine Learning</i>	11
2.1.3 <i>Deep Learning</i>	11
2.1.4 <i>Bayesian Network</i>	12

2.1.5	<i>Naïve Bayes</i>	13
2.1.6	<i>Multilayer Perceptron</i>	14
2.1.7	Matriks Evaluasi	16
2.2	Tinjauan Pustaka	17
3.1	Pengumpulan Data	32
3.1.1	Populasi dan Sampel Data	32
3.1.2	Sumber Data	33
3.1.3	Aliran Data	33
3.1.4	Variabel Penelitian	34
3.2	Metodologi yang Diusulkan.....	35
4.1	Data Eksplorasi	38
4.2	Kategorisasi Data	41
4.3	<i>Bayesian Network</i>	42
4.3.1	<i>Conditional Probability Distributions (CPDs)</i>	46
4.3.2	Pemilihan Variabel dan Status Signifikan	62
4.3.3	<i>Evaluation Model Output</i>	63
4.4	<i>Naïve Bayes</i>	64
4.4.1	Seluruh Variabel	64
4.4.2	Variabel Signifikan.....	66
4.5	<i>Multilayer Perceptron</i>	68
4.5.1	Pengujian Pada Seluruh Variabel	68
4.5.2	Pengujian Pada Variabel Signifikan	69
4.6	Performa Metode <i>Naïve Bayes</i> dan <i>Multilayer Perceptron</i>	70
5.1	Kesimpulan	73
5.2	Saran	75

Daftar Tabel

Tabel 2.1 <i>Confusion Matriks</i>	17
Tabel 2.2 Ulasan Kritis Tema 1.....	23
Tabel 2.3 Ulasan Kritis Tema 2.....	23
Tabel 2.4 Ulasan Kritis Tema 3.....	27
Tabel 3.1 Variabel Penelitian.....	34
Tabel 3.2 Variabel Status (<i>Summary</i>) Penelitian	34
Tabel 4.1 Menentukan Kategori Baru dengan Menggunakan Persentil Data	42
Tabel 4.2 <i>Parent</i> dan <i>Child nodes</i>	44
Tabel 4.3 Nilai CPDs pada variabel kategori <i>delivered</i>	47
Tabel 4.4 Nilai CPDs pada variabel kategori <i>undelivered</i>	47
Tabel 4.5 Nilai CPDs pada variabel kategori <i>Average Notification per Minutes</i>	48
Tabel 4.6 Nilai CPDs pada variabel kategori <i>Rate Undelivered</i>	48
Tabel 4.7 Nilai CPDs pada variabel kategori <i>Not Normal</i> dan variabel kategori <i>Day</i>	48
Tabel 4.8 Nilai CPDs pada variabel kategori <i>Unknown Summary</i> dan variabel kategori <i>Long Delay</i>	49
Tabel 4.9 Nilai CPDs pada variabel kategori <i>Normal Delay</i> dan variabel kategori <i>Normal</i>	49
Tabel 4.10 Nilai CPDs pada variabel kategori <i>Average Delay</i> dan <i>Maybe Issue From</i> <i>Operator</i>	49
Tabel 4.11 Nilai CPDs pada variabel status <i>Maybe Delay From 3rd Party</i>	49
Tabel 4.12 Hasil Klasifikasi Kategori <i>Delivered</i>	50
Tabel 4.13 Hasil Klasifikasi Kategori <i>Undelivered 1</i>	50
Tabel 4.14 Hasil Klasifikasi Kategori <i>Undelivered 2</i>	51
Tabel 4.15 Hasil Klasifikasi Kategori <i>Undelivered 3</i>	51
Tabel 4.16 Hasil Klasifikasi Kategori <i>Undelivered 4</i>	52
Tabel 4.17 Hasil Klasifikasi Kategori <i>Undelivered 5</i>	52
Tabel 4.18 Hasil Klasifikasi Kategori <i>Undelivered 6</i>	53
Tabel 4.19 Hasil Klasifikasi Kategori <i>Undelivered 7</i>	53
Tabel 4.20 Hasil Klasifikasi Kategori <i>Undelivered 8</i>	53
Tabel 4.21 Hasil Klasifikasi Kategori <i>Average Notification Per Minutes 1</i>	54
Tabel 4.22 Hasil Klasifikasi Kategori <i>Average Notification Per Minutes 2</i>	54
Tabel 4.23 Hasil Klasifikasi Kategori <i>Average Notification Per Minutes 3</i>	55

Tabel 4.24 Hasil Klasifikasi Kategori <i>Average Notification Per Minutes</i> 4	56
Tabel 4.25 Hasil Klasifikasi Kategori <i>Average Notification Per Minutes</i> 5	56
Tabel 4.26 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 1	56
Tabel 4.27 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 2	57
Tabel 4.28 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 3	57
Tabel 4.29 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 4	58
Tabel 4.30 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 5	58
Tabel 4.31 Hasil Klasifikasi Kategori <i>Rate Undelivered</i> 6	58
Tabel 4.32 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 1	59
Tabel 4.33 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 2	60
Tabel 4.34 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 3	60
Tabel 4.35 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 4	61
Tabel 4.36 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 5	61
Tabel 4.37 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 6	62
Tabel 4.38 Hasil Klasifikasi Status <i>Maybe Issue From 3rd Party</i> 7	62
Tabel 4.39 Pemilihan Variabel dan Status Signfikan	63
Tabel 4.40 Nilai Akurasi Berdasarkan Kategori	63
Tabel 4.41 Distribusi data uji pada <i>summary status</i> pada seluruh variabel.....	64
Tabel 4.42 Distribusi data uji pada <i>summary status</i> pada parsial variabel.....	66
Tabel 4.43 Performa partisi ukuran data latih	70
Tabel 5.1 Kategorisasi Berdasarkan Hasil <i>Machine Learning (Naïve Bayes)</i>	74

Daftar Gambar

Gambar 1.1 Transaksi berdasarkan metode pembayaran (Xendit, 2022).....	1
Gambar 2.1 Struktur <i>Multilayer Perceptron</i>	15
Gambar 2.2 Arsitektur <i>Multilayer Perceptron</i>	16
Gambar 2.3 Diagram <i>Prisma Protocol</i> (da Silva et al., 2019)	18
Gambar 3.1 Diagram Alir Penelitian (Tarigan et al., 2023).	31
Gambar 3.2 Diagram Alir Data Transaksi SMS OTP	33
Gambar 3.3 Diagram Alir Bayesian Network.	36
Gambar 3.4 Diagram Alir Naïve Bayes.....	36
Gambar 3.5 Diagram Alir Multi-Layer Perceptron.	37
Gambar 4.1 Jumlah Transaksi Berdasarkan Kategori Waktu.....	38
Gambar 4.2 Jumlah Transaksi Berdasarkan Kategori Ringkasan	39
Gambar 4.3 Korelasi Antar Variabel Transaksi Pengiriman Pesan OTP	40
Gambar 4.4 Box plot untuk identifikasi <i>outlier</i>	41
Gambar 4.5 Visualisasi <i>network</i> dari model Bayesian Network	43
Gambar 4.6 <i>Confusion matrix</i> metode <i>naïve bayes</i> pada seluruh variabel.....	65
Gambar 4.7 <i>Confusion matrix</i> <i>naïve bayes</i> pada parsial variabel.....	67

Glosarium

BN	- Bayesian Network
CPDs	- Conditional Probability Distributions
E-Wallet	- Electronic Wallet (Dompet Digital)
MLP	- Multi-Layer Perceptron
OTP	- One-Time Password
QRIS	- Quick Response Code Indonesian Standard
SMS	- Short Message Service
VA	- Virtual Account (Rekening bank virtual yang dapat menjadi rekening tujuan pembayaran transaksi)

BAB 1

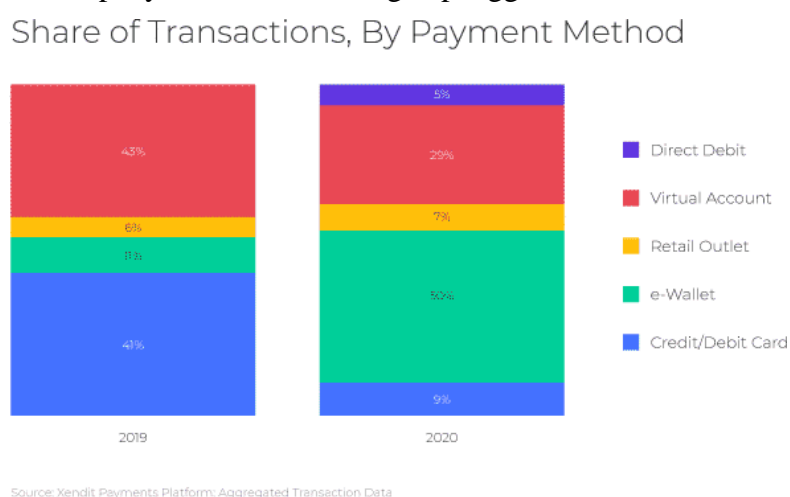
Pendahuluan

1.1 Pendahuluan

Salah satu pendorong terbesar ekonomi modern yakni dari uang digital yang mana secara global adopsi smartphone yang memudahkan pengguna untuk mengakses layanan keuangan digital di negara berkembang. Dimana ada uang, maka perlu ada keamanan, tetapi sebelumnya pada tingkat kerentanan dari transaksi uang digital mengecilkan hati pengguna untuk menggunakannya pada transaksi sehari-hari (Hartsook, 2000).

Perkembangan internet hingga saat ini sangat pesat dimana banyak muncul inovasi baik penyedia jasa transportasi, jasa toko online (e-commerce) dan masih banyak lagi jasa-jasa yang memanjakan pengguna untuk semakin tidak bisa terlepas dari penggunaan smartphone. Selain itu, muncul berbagai macam pilihan pembayaran elektronik seperti credit/debit card, e-Wallet, dan metode pembayaran tanpa kontak langsung lainnya (Bakar et al., 2017).

Berdasarkan data dari (Xendit, 2022) penggunaan dari e-Wallet meningkat signifikan pada tahun 2020 yang mana memunculkan produk QRIS yang memungkinkan interoperabilitas antar penyedia e-Wallet dengan pengguna dalam transaksi sehari-hari.



Gambar 1.1 Transaksi berdasarkan metode pembayaran (Xendit, 2022).

Berdasarkan Gambar 1.1 dapat dilihat pada tahun 2019 tren penggunaan pembayaran melalui virtual account (VA) dan credit/debit card mendominasi dalam pembayaran

transaksi dibandingkan e-Wallet, akan tetapi pada tahun 2020 transaksi menggunakan e-Wallet melampaui metode transaksi VA dan credit/debit card di tahun 2019. Hal ini mengartikan bahwa transaksi dari e-Wallet mulai menjadi primadona di tahun 2020.

Semakin meningkatnya penggunaan teknologi pembayaran melalui *smartphone*, banyak perusahaan kini berlomba-lomba untuk mengembangkan fitur-fitur yang tidak hanya mempermudah pengguna dalam bertransaksi, tetapi juga meningkatkan keamanan dalam proses tersebut. Salah satu pendekatan alternatif untuk meningkatkan keamanan yang dapat digunakan adalah kata sandi satu kali (*One-time Password* [OTP]), yang hanya bagus untuk proses otentikasi pengguna dalam satu sesi (Chow et al., 2015). Dengan adanya skema ini tentunya dapat meningkatkan kepuasan pelanggan dan merasa aman ketika bertransaksi. Dari sisi perusahaan, merupakan salah satu poin penting yang memberikan dampak signifikan terhadap keberlanjutan perusahaan dan kepercayaan perusahaan (Barlas et al., 2023).

One-time Password (OTP) merupakan skema otentikasi terhadap kata sandi atau penambahan validasi yang digunakan untuk setiap sesi otentikasi pengguna. Apabila sesi otentikasi telah habis atau *one-time password* telah digunakan, maka tidak valid lagi. Selain *one-time password* digunakan, apabila ada pengiriman ulang pada *one-time password* maka yang dapat digunakan untuk sesi otentikasi selanjutnya adalah kode otentikasi terbaru untuk digunakan. Skema ini direalisasikan dengan menjamin secara komputasi tidak mungkin untuk menentukan kata sandi berikutnya dari kata sandi saat ini, pun apabila peretas membutuhkan jumlah yang cukup besar secara waktu dan sumber daya komputasi agar dapat membedakan data yang dapat digunakan pada pengiriman *One-Time Password* yang masih berstatus antri (*queue*) (Raya, 2011).

Keuntungan dari penggunaan OTP adalah (Abousteit et al., 2020) :

1. Kata sandi yang berbeda dari sebelumnya yang aman dari serangan yang ingin menggunakan kata sandi yang sebelumnya, hal ini diakibatkan karena tidak lagi valid untuk login untuk sesi mendatang.
2. Umumnya penggunaan OTP diterima melalui SMS, sehingga tidak perlu memiliki akses ke *email*
3. OTP memungkinkan perusahaan untuk tidak hanya meningkatkan pengalaman pengguna tetapi juga mengurangi biaya operasional

karena pengguna sudah terbiasa dengan ponsel dan tidak memerlukan perangkat tambahan untuk menerima kode.

Apabila OTP tidak tersedia untuk mengimbangi teknologi saat ini, maka risiko terjadinya *fraud* dalam transaksi uang digital akan meningkat secara signifikan, yang pada akhirnya dapat menyebabkan penurunan tingkat kepercayaan dan kepuasan pelanggan. Sehingga diperlukan penelitian yang berkaitan terhadap pengiriman pesan satu kali kata sandi (*One-Time Password*) untuk meningkatkan efektifitas dan efisiensi dalam kesuksesan atau kegagalan pada pengiriman pesannya.

Dalam upaya membangun model analitik, diperlukan pemrograman yang mengintegrasikan hubungan, prosedur, dan logika keputusan secara eksplisit ke dalam sistem cerdas, baik melalui aturan yang disusun secara manual maupun masukan dari pakar. Pendekatan tradisional ini mendorong terciptanya kerangka pemrograman baru yang lebih praktis. Hal ini semakin relevan mengingat saat ini terdapat ketersediaan data yang semakin beragam, volume data yang besar, serta akses yang semakin luas terhadap daya komputasi. Sehingga, model analitik saat ini semakin banyak dibangun secara umum yang disebut sebagai *machine learning* (Russell & Novig, 2020).

Machine learning mengurangi beban kerja manusia dalam menjelaskan dan memformalkan pengetahuan ke dalam bentuk yang dapat diakses oleh mesin, serta memungkinkan pengembangannya melalui sistem cerdas secara efisien. *Machine learning* juga menggambarkan kapasitas sistem untuk belajar dari data historis guna mengotomatiskan proses pembuatan model analitik dan menyelesaikan tugas-tugas terkait. Sedangkan *deep learning* merupakan konsep dari *machine learning* dengan basis konsep dari jaringan saraf tiruan. Banyak pengaplikasian bahwa, model yang dihasilkan dari *deep learning* mengungguli dari pada model pembelajaran mesin dan pendekatan analisis data secara tradisional (Díaz-Ramírez, 2021).

Penelitian ini akan melakukan klasifikasi data SMS berdasarkan status pengiriman SMS OTP dengan menggunakan data transaksi pengiriman SMS *One-Time Password* di Indonesia pada bulan juli hingga bulan november tahun 2021 dengan menggunakan tiga metode *bayesian network*, *naïve bayes*, dan *multilayer perceptron*. Kemudian, penggunaan metode *machine learning* dan *deep learning* sebagai perbandingan untuk mengetahui apakah pada penelitian ini cukup dengan metode *machine learning* saja atau perlu hingga ke

penggunaan *deep learning* dalam pengklasifikasian data transaksi SMS *One-Time Password*. Secara tahapan, penelitian ini dimulai dengan mengetahui karakteristiknya terlebih dahulu untuk mendapatkan gambaran secara umum kondisi dari data yang digunakan, hal ini guna membantu peneliti dalam penyeleksian tahap pertama pada variabel yang digunakan. Setelah mengetahui karakteristik data, maka dilanjutkan untuk melihat kausalitas antar variabel guna mengetahui karakteristik data lebih spesifik, penyeleksian variabel tahap dua, serta mengetahui seberapa besar pengaruh (probabilitas) sebuah variabel sebab dalam mempengaruhi variabel akibat dengan menggunakan metode *bayesian network*. Kemudian dilanjutkan dengan mencoba melatih data dengan metode *machine learning* (*naïve bayes*) dan metode *deep learning multilayer perceptron* untuk mengetahui klasifikasi prediksi yang mana hasil evaluasi model terbaik akan yang terpilih. Dari sisi biaya pada implementasi komputasi, *deep learning* cenderung lebih mahal dibandingkan metode *machine learning*.

Selanjutnya, penelitian terdahulu menjadi landasan dalam penelitian ini agar tidak terjadi penelitian yang serupa. Penelitian dengan menggunakan metode *bayesian network*, *naïve bayes*, dan *multilayer perceptron* sudah banyak dilakukan dan diterapkan terhadap berbagai dataset untuk pengklasifikasian data (Alharbi & Alghazzawi, 2019; El Hlouli et al., 2020; Farooq, 2020; Kasi, 2019; Kumari et al., 2014; Naik et al., 2016; Ncibi et al., 2017; Ning et al., 2019; Oyegoke et al., 2021; Pryor et al., 2022; Sable & Kalavadekar, 2016; Umair et al., 2020), namun peneliti tidak menemukan penelitian terdahulu yang melakukan pengklasifikasian data sms dengan menggunakan *bayesian network*, *naïve bayes*, dan *multilayer perceptron*.

Tujuan dari penelitian ini adalah untuk membantu bagi perusahaan penyedia layanan jasa pengiriman pesan OTP dalam memprediksi kapan akan terjadi kegagalan saat pengiriman SMS yang diminta oleh *merchant* ke pengguna, memberikan informasi variabel apa saja yang secara signifikan mempengaruhi dalam kesuksesan/kegagalan dari transaksi pengiriman SMS, serta mengusulkan kategori baru dalam pengiriman pesan dari data yang ada. Harapannya akan membantu dalam pengambilan keputusan untuk *switch provider* ketika banyak terjadi kegagalan dalam pengiriman SMS.

Secara spesifik tujuan pada penelitian ini menggunakan metode *bayesian network* yakni dengan menggunakan metode tersebut diyakini bahwa dapat merepresentasikan

kompak dari distribusi probabilitas bersama serta hubungan probabilitas kausal antara sejumlah variabel acak dan kondisi dependensi (Cai et al., 2019). Kemudian pada *naïve bayes* bertujuan untuk mengklasifikasikan objek dengan asumsi memiliki independensi yang kuat antar objek, kecenderungan *naive bayes* juga banyak digunakan untuk filter *spam* (Lv et al., 2020). Selain itu, *multilayer perceptron* salah satu algoritma sebagai pengklasifikasi untuk pemfilteran yang mana algoritma ini memungkinkan pengintegrasian fitur data secara spontan dari tingkat atas ke tingkat yang lebih rendah (dalam), yang kemudian juga dapat menghasilkan kinerja yang lebih baik dalam presisi pada data latih yang lebih besar (Dada et al., 2019).

Kontribusi penelitian ini pada sisi ilmiah dapat memberikan pengetahuan kepada peneliti lainnya terkait proses dibalik transaksi SMS *one-time password* serta klasifikasi apa saja yang ada diperusahaan penyedia jasa pengiriman pesan SMS *one-time password*, Selain itu, klasifikasi dapat membantu identifikasi pola dan struktur dalam data pengiriman pesan SMS OTP. Sedangkan pada sisi sosial dapat membantu perusahaan penyedia jasa untuk menentukan kategori-kategori yang berpengaruh signifikan atau kategori yang menjadi utama berdasarkan data yang ada.

1.2 Rumusan Masalah

1. Bagaimana karakteristik waktu transaksi pengiriman pesan berdasarkan data transaksi pengiriman pesan SMS OTP?
2. Apa saja variabel yang menjadi komponen utama dalam menentukan sukses/gagalnya pengiriman pesan SMS OTP?
3. Faktor-faktor apa saja yang menjadi kausalitas kegagalan pengiriman pesan SMS OTP?
4. Bagaimana pengembangan model *machine learning* serta bagaimana hasil klasifikasi pada data transaksional pengiriman pesan SMS OTP?

1.3 Batasan Masalah

Batasan masalah merupakan ruang lingkup penelitian yang bertujuan untuk memperjelas fokus penelitian dan menjaga agar penelitian tetap terarah dan relevan dengan tujuan penelitian, maka pada penelitian ini diberikan batasan-batasan sebagai berikut :

1. Data penelitian yakni Data Transaksi Pengiriman Pesan SMS OTP pada tahun 2021 yang diperoleh dari PT XYZ (perusahaan yang bergerak dibidang jasa pengiriman pesan SMS OTP).
2. Objek pengamatan penelitian adalah transaksi pengiriman pesan SMS OTP di Indonesia pada bulan Juli hingga November 2021.
3. Variabel penelitian yang digunakan yakni *transaction date*, *transaction time*, jumlah pesan terkirim, jumlah pesan dikirim, jumlah pesan tidak dikirim, jumlah pesan dengan status tidak diketahui, jumlah pesan dengan pemberitahuan tidak diterima oleh perusahaan PT XYZ dari pihak ketiga, jumlah pesan dengan pemberitahuan pengiriman status batas waktu dari penyedia pihak ketiga ke PT XYZ, rata-rata notifikasi dari penyedia pihak ketiga perdetik, *rate delivered*, *rate undelivered*, rata-rata notifikasi dari penyedia pihak ketiga per menit, *summary status*, dan *day category*. Dengan variabel status (*summary*) yakni *not delay notification*, sangat normal *notification*, *unknown summary notification*, *normal delay notification*, *long delay notification*, *maybe delay from 3rd party*, *normal notification*, *timeout to 3rd party*, *average delay notification*, *maybe delay from operator*, dan *not normal notification*.
4. Metode penelitian yang digunakan yakni *Bayesian Network*, *Naive Bayes*, dan *Multilayer Perceptron*.
5. Pengolahan data penelitian menggunakan *Python 3.9.7*.

1.4 Tujuan Penelitian

1. Menganalisis karakteristik waktu transaksi pengiriman pesan berdasarkan data transaksi pengiriman pesan SMS OTP.
2. Mengidentifikasi dan menentukan variabel-variabel penyusun utama dalam studi kasus data transaksi pengiriman pesan SMS OTP.
3. Mengidentifikasi faktor-faktor kausal yang menyebabkan kegagalan pengiriman pesan pesan SMS OTP.
4. Mengembangkan model *machine learning* untuk klasifikasi data transaksional pengiriman pesan dan mengevaluasi hasil klasifikasinya.

1.5 Manfaat Penelitian

Penelitian ini harapannya bermanfaat untuk menjadi referensi perusahaan yang menyediakan jasa layanan pengiriman pesan OTP dalam mengetahui karakteristik dari pengguna kecenderungan menggunakan layanan pesan OTP ini saat kapan, variabel-variabel apa saja yang menentukan dalam kesuksesan/gagalnya pengiriman pesan OTP, serta metode-metode yang dapat digunakan dalam memprediksi klasifikasi. Dengan hal tersebut memudahkan bagi perusahaan untuk menelusuri akar permasalahan isu kegagalan pengiriman pesan SMS one-time password dengan kategori yang telah spesifik serta prediksi dari model yang telah dibangun.

Kemudian, manfaat untuk para peneliti, harapannya dengan menggunakan klasifikasi data dapat menambah *insight* pada pola dan struktur data transaksi pengiriman pesan SMS OTP. Kemudian dapat mengetahui hubungan antar variabel dan status pengiriman pesan SMS OTP serta fenomena transaksi pengiriman pesan SMS OTP. Selain itu, kedepannya dengan adanya penelitian ini diharapkan dapat memungkinkan pengorganisasian informasi secara sistematis, sehingga memudahkan peneliti untuk mengakses, menganalisis, dan membangun pengetahuan yang ada.

1.6 Metodologi Penelitian

Tahapan metodologi yang dilakukan secara menyeluruh yakni:

1. Studi Literatur

Mengumpulkan dan melakukan studi terkait penelitian terdahulu dengan sumber buku, jurnal, *web article* (laporan / defenisi), karya tulis ilmiah.

2. Persiapan Data

Pada penelitian ini tahapan dari pre-proses yakni menghapus NA/nilai kosong, membagi data latih 80% dan data uji 20% dengan pembagian secara acak, teknik dan model menggunakan klasifikasi dan prediksi dengan metode *naïve bayes* dan *multilayer perceptron*

3. Pengumpulan Data

Data yang akan digunakan pada penelitian ini adalah data transaksi pengiriman pesan SMS OTP dari perusahaan penyedia layanan jasa pengiriman pesan OTP. Perusahaan itu sendiri yang mengelola data menggunakan Oracle sebagai alat kueri. Data tersebut diperoleh dengan menggunakan database kueri (SQL).

4. Pengolahan Data

Pada penelitian ini menggunakan mengkategorisasi data dengan menggunakan pendekatan persentil data dengan asumsi distribusi *uniform*, memilah memilah variabel dan status yang signifikan dengan *bayesian network* dan algoritma *naïve bayes (machine learning)* dan *multilayer perceptron (deep learning)* dengan berbagai parameter hingga ditemukan akurasi yang optimal yang kemudian akurasi antar algoritma tersebut diperiksa dan dipilih.

5. Evaluasi

Pada penelitian ini untuk mengevaluasi model dengan melihat nilai akurasi yang dihasilkan serta *confusion matrix*

1.7 Sistematika Penulisan

Untuk memberikan gambaran mengenai isi penulisan, berikut sistematika penulisan:

BAB 1 PENDAHULUAN

Bab ini mencakup penjelasan pendahuluan yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan luaran penelitian.

BAB 2 TINJAUAN PUSTAKA

Bab ini memaparkan teori-teori dan penelitian-penelitian yang berkaitan dengan OTP / pengiriman pesan OTP, termasuk metode *machine learning* dan *deep learning* yang digunakan

BAB 3 METODOLOGI

Bab ini menjelaskan tahapan-tahapan penelitian terkait pengiriman pesan SMS OTP, mulai dari pengumpulan data, hingga metodologi yang diusulkan

BAB 4 HASIL DAN PEMBAHASAN

Bab ini membahas karakteristik transaksi pengiriman pesan SMS OTP di Indonesia serta hasil analisis pengujian yang dilakukan dengan menggunakan metode *machine learning* dan *deep learning*.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian serta saran yang mempertimbangkan keterbatasan temuan dalam penelitian dan rekomendasi untuk penelitian selanjutnya.

BAB 2

Tinjauan Pustaka

2.1 Konsep Pengetahuan

2.1.1 Klasifikasi Pengiriman Pesan SMS OTP

One-time Password (OTP) merupakan skema otentikasi terhadap kata sandi atau penambahan validasi yang digunakan untuk setiap sesi otentikasi pengguna. Apabila sesi otentikasi telah habis atau *one-time password* telah digunakan, maka tidak valid lagi. Selain *one-time password* digunakan, apabila ada pengiriman ulang pada *one-time password* maka yang dapat digunakan untuk sesi otentikasi selanjutnya adalah kode otentikasi terbaru untuk digunakan. Skema ini direalisasikan dengan menjamin secara komputasi tidak mungkin untuk menentukan kata sandi berikutnya dari kata sandi saat ini, pun apabila peretas membutuhkan jumlah yang cukup besar secara waktu dan sumber daya komputasi agar dapat membedakan data yang dapat digunakan pada pengiriman *One-Time Password* yang masih berstatus antri (*queue*) (Raya, 2011).

Dalam perjalanan pengiriman pesan SMS OTP ke pengguna (*user*) menghasilkan kemungkinan-kemungkinan yang terjadi. Yang mana direpresentasikan dalam kategori-kategori status yang disusun seperti *pending* (tertunda), *sent* (terkirim), *delivered* (tersampaikan), *failed* (gagal), *undelivered* (tidak tersampaikan), *expired* (kadaluwarsa), *rejected* (ditolak), *unknown* (tidak diketahui) (Platform, 2024).

Klasifikasi merupakan salah satu metode pembelajaran terarah, yang memetakan pada suatu masukan terhadap keluaran yang diinginkan (Drummond, 2011). Klasifikasi dapat digunakan untuk memastikan detail pengiriman pesan, seperti apakah pesan berhasil dikirim, tidak berhasil, atau tertunda, saat mengirim pesan SMS OTP (*One-Time Password*). Upaya untuk meningkatkan efektivitas dan ketepatan pengiriman SMS OTP, yang sering digunakan dalam transaksi pengiriman pesan SMS OTP menjadikan metode klasifikasi ini sangat relevan.

2.1.2 *Machine Learning*

Pembelajaran mesin adalah salah satu penggunaan AI yang menggunakan data untuk membuat prediksi. Arthur Samuel pertama kali mempresentasikan pembelajaran mesin pada tahun 1952. Komputer menggunakan proses yang disebut pembelajaran mesin untuk belajar dari data; Tanpa data, komputer tidak mampu belajar. Teknik belajar lebih banyak, menggunakan data masa lalu yang terdiri dari data yang dikumpulkan secara elektronik yang dapat diakses untuk pembelajaran untuk meningkatkan kinerja atau membuat prediksi yang tepat berdasarkan pengalaman yang diinginkan terhadap data sebelumnya yang dapat diambil dan dimanfaatkan sebagai data pembelajaran (Mohri et al., 2018).

Dalam pembelajaran mesin, empat mata pelajaran umum dieksplorasi, seperti:

- a. Pendidikan terbimbing salah satu cara untuk belajar secara terarah adalah melalui pembelajaran terarah yang melibatkan pembuatan hubungan antara masukan dan hasil yang diinginkan. Dapatkan sekumpulan data yang telah diberi label sebagai data pelatihan, dan perkirakan semua titik tak terlihat (seperti klasifikasi).
- b. Pembelajaran tidak terarah: Metode ini berusaha menggunakan input dalam bentuk model kumpulan data. Sebagai data pelatihan, ia diberikan sekumpulan data berlabel dan tidak berlabel, dan menggunakan data ini untuk memperkirakan semua titik tak terlihat, seperti pengelompokan.
- c. Pembelajaran Semi-diawasi: Teknik ini menciptakan fungsi dengan memadukan pembelajaran yang diawasi dan tidak diawasi.
- d. Belajar melalui Penguatan Bagian pembelajaran dan pengujian dari suatu pendekatan diselesaikan secara bersamaan, yang dikenal sebagai pembelajaran penguatan. Data tentang pembelajaran dikumpulkan melalui Ada empat belas metode untuk terlibat secara aktif dengan lingkungan sekitar untuk mengukur bagaimana setiap langkah pembelajaran diterima.

2.1.3 *Deep Learning*

Deep learning adalah jenis teknik pembelajaran mesin yang melatih komputer untuk menyelesaikan tugas yang diharapkan dilakukan orang dengan menggunakan konsep dari jaringan saraf buatan. Deep learning akan menjadi alat yang sangat berguna dalam pembelajaran yang diawasi untuk bangunan arsitektur dan dapat diterapkan pada tantangan

yang melibatkan penggunaan beberapa kumpulan data. Ada beberapa aplikasi untuk pembelajaran imersif, termasuk klasifikasi langsung teks, audio, dan visual. Dalam pembelajaran mendalam, ada metode untuk mengekstraksi fitur dari algoritme dan data pelatihan. Teks, suara, dan visual dikategorikan dalam pendidikan khusus. Melalui penggunaan pembelajaran mendalam, komputer dapat dilatih dengan kumpulan data yang telah menjalani pelabelan ekstensif. Kumpulan data ini selanjutnya dapat diubah menjadi nilai piksel gambar. Kategorisasinya berguna untuk mengenali atau mengkategorikan pola input (LeCun et al., 2015)

Deep learning semakin banyak digunakan dalam industri dan penelitian untuk menangani tantangan data besar termasuk pengenalan suara, visi komputer, dan pemrosesan bahasa saraf. Ini karena deep learning selalu berkembang. Ada rekayasa sifat dalam pembelajaran mendalam. Untuk membantu model mengidentifikasi antara kelas dan prosedur yang efektif untuk mencapai hasil prediksi yang baik, teknik fitur diterapkan. Tetapi menguasai rekayasa fitur itu menantang karena Berbagai prosedur rekayasa harus digunakan ketika beberapa kumpulan data dari berbagai jenis sedang digunakan. Teknik pemanfaatan konvolusi Jaringan saraf akan menjadi yang paling efektif dan membantu untuk mengidentifikasi fitur dalam gambar yang akan diteruskan ke lapisan berikutnya, membentuk hipotesis nonlinier yang dapat meningkatkan kompleksitas model. Proses di mana menggunakan model kompleksitas jaringan saraf konvolusional akan membutuhkan waktu pengujian yang signifikan untuk mendapatkan pemahaman menyeluruh tentang penggunaan unit pemrosesan grafis (GPU) (Thohari & Hertantyo, 2018).

2.1.4 Bayesian Network

Bayesian Network (BN), atau disebut juga sebagai jaringan kepercayaan, adalah model probabilistik yang digambarkan secara grafis yang menggambarkan hubungan antara variabel acak. Untuk menawarkan metode yang efektif dalam simulasi ketergantungan dan ketidakpastian statistik dalam berbagai bidang, seperti penilaian risiko, diagnosis medis, dan kecerdasan buatan, Bayesian Network (BN) memadukan teori grafik dengan teori probabilitas (Polotskaya et al., 2024).

Berdasarkan (Ojha et al., 2018) rumus dasar dari bayesian network didasarkan pada teoroma bayes dan aturan rantai untuk probabilitas gabungan. Jika suatu jaringan dengan

variabel acak X_1, X_2, \dots, X_n maka probabilitas gabungan dari semua variabel dalam jaringan dinotasikan sebagai berikut.

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \quad (1)$$

Dimana, X_i adalah variabel pada jaringan, $\text{parents } x_i$ adalah himpunan variabel-variabel utama langsung dari X_i dalam graf.

Atau dalam penggunaannya, jika terdapat variabel A,B, dan C dalam jaringan dengan struktur graf A mempengaruhi B ($A \rightarrow B$), A mempengaruhi C ($A \rightarrow C$), maka probabilitas gabungan dihitung sebagai berikut

$$P(A, B, C) = P(A) \cdot P(B|A) \cdot P(C|A) \quad (1)$$

2.1.5 Naïve Bayes

Naïve Bayes (NB) adalah salah satu algoritma klasifikasi yang paling dikenal dalam penambangan data. Algoritma ini menentukan probabilitas bahwa sebuah contoh baru termasuk dalam kelas tertentu dengan asumsi bahwa semua atribut bersifat independen satu sama lain dalam konteks kelas tersebut. Asumsi ini diperlukan untuk memperkirakan probabilitas multivariat dari data pelatihan. Dalam praktiknya, banyak kombinasi nilai atribut mungkin tidak ditemukan dalam data pelatihan atau tidak muncul dalam jumlah yang memadai. Oleh karena itu, estimasi langsung dari probabilitas multivariat yang relevan menjadi tidak dapat diandalkan. Naïve Bayes mengatasi masalah ini dengan mengadopsi asumsi independensi bersyarat. Meski asumsi independensi yang diterapkan cukup ketat, Naïve Bayes tetap merupakan pengklasifikasi yang sangat efektif dalam berbagai aplikasi dunia nyata.

Misalkan X_1, X_2, \dots, X_n adalah vektor n-dimensi variabel acak (fitur) dari domain D_x dan X_1, X_2, \dots, X_n adalah instans yang sesuai. Misalkan Y adalah variabel acak tak teramati dari domain $D_y = 0,1$. Meskipun tidak diketahui, diasumsikan bahwa ada fungsi D_x ke D_y . Tujuannya adalah untuk memperkirakan nilai target y_i dari Y untuk instans tertentu x_i dari X . Dengan kata lain, tujuannya adalah untuk memilih kelas Y yang memaksimalkan probabilitas posterior, $P(Y = y | X = x)$. Disini, $P(Y = y)$ dan $P(X = x | Y = y)$ adalah probabilitas kondisional kelas dan prior. Karena *naive bayes* mengasumsikan independensi

konsional, $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | Y = y) = \prod_{i=1}^n P(X_i = x_i | Y = y)$.

Misalkan C adalah jumlah kelas Y. Berdasarkan teorema Bayes sebagai berikut.

$$P(Y = y | X = x) \tag{1}$$

$$= \frac{P(Y=y, X=x)}{P(X=x)} \tag{2}$$

$$= \frac{P(Y=y)P(X=x | Y=y)}{P(X=x)} \tag{3}$$

$$= \frac{P(Y=y)P(X_1=x_1, X_2=x_2, \dots, X_n=x_n | Y=y)}{\sum_i^C P(y_i, X=x)} \tag{4}$$

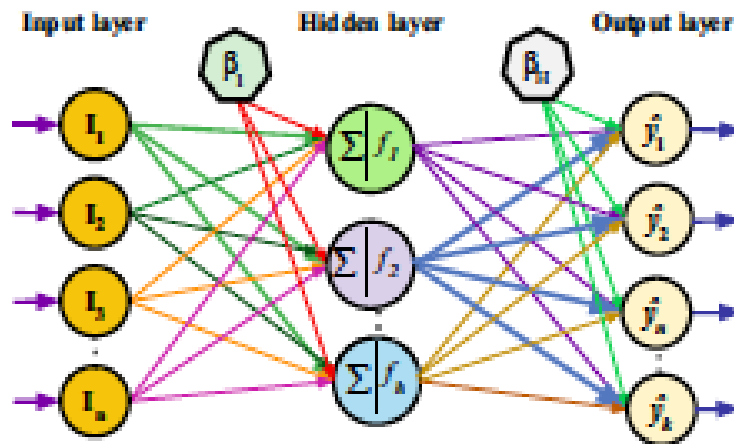
$$= \frac{P(Y=y) \prod_{i=1}^n P(X_i=x_i | Y=y)}{\sum_i^C P(y_i, X=x)} \tag{5}$$

Dalam banyak kasus praktis $P(X = x | Y = y)$ diasumsikan mengikuti distribusi Gaussian, meskipun beberapa literatur mencatat pengecualian terhadap asumsi ini. George dan koleganya (1995) menggantikan asumsi Gaussian yang fleksibel dengan estimasi kepadatan kernel. Selain asumsi Gaussian, salah satu asumsi penting adalah bahwa nilai dari satu fitur tidak mempengaruhi nilai fitur lainnya. Meski ada perdebatan mengenai seberapa realistis asumsi ini, asumsi tersebut telah membuat algoritme lebih praktis. Selain itu, asumsi Naïve dapat dimanfaatkan untuk mempercepat eksekusi algoritme. Sebagai contoh, probabilitas atribut dapat dihitung secara paralel dengan menggunakan CPU, mesin, atau kluster yang berbeda dalam aplikasi berbasis Naïve Bayes di dunia nyata. Keberhasilan algoritme Naïve Bayes ini telah mendorong komunitas peneliti untuk mengeksplorasi variasi dari metode tersebut (Wickramasinghe & Kalutarage, 2021).

2.1.6 *Multilayer Perceptron*

MLP adalah contoh dari FFNN yang banyak digunakan. Dalam MLP, elemen pemrosesan diatur dalam arah satu arah. Jaringan ini mengolah informasi melalui komunikasi antara tiga jenis lapisan: lapisan input, lapisan tersembunyi, dan lapisan output. Lapisan-lapisan ini saling terhubung melalui bobot yang nilainya bervariasi dalam rentang [-1, 1]. Setiap simpul dalam MLP menjalankan dua fungsi, yaitu fungsi penjumlahan dan fungsi aktivasi. Nilai

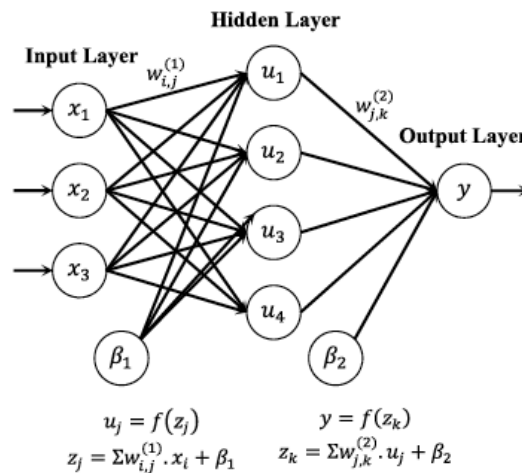
hasil kali antara nilai masukan, bobot, dan bias dihitung menggunakan fungsi penjumlahan seperti yang dijelaskan dalam persamaan (Du et al., 2023).



Gambar 2.1 Struktur *Multilayer Perceptron*

Multi-Layer Perceptron Neural Networks (MLP-NNs) adalah jenis jaringan saraf umpan maju yang menggunakan metode pembelajaran terawasi untuk menyelesaikan masalah regresi dan klasifikasi (Alickovic & Subasi, 2020). MLP-NN terdiri dari setidaknya tiga lapisan: lapisan input, lapisan tersembunyi, dan lapisan output. Dalam konteks masalah klasifikasi, jumlah simpul pada lapisan input dan lapisan output disesuaikan dengan jumlah fitur dan kelas dalam data. Selain itu, konfigurasi MLP-NN melibatkan pengaturan parameter tambahan seperti jumlah lapisan tersembunyi, jumlah simpul di setiap lapisan tersembunyi, jumlah epoch pembelajaran, dan metode pembelajaran bobot. Simpul pada lapisan tersembunyi menerima sinyal dari lapisan input melalui koneksi berbobot, kemudian menghitung fungsi aktivasi dan mengirimkan hasilnya ke lapisan output. Di antara berbagai algoritma pelatihan untuk MLP-NN, algoritma backpropagation adalah yang paling umum digunakan (Salehi et al., 2020). Dalam metode ini, setiap sampel pelatihan diproses oleh MLP-NN untuk menghasilkan output yang dihitung melalui langkah maju. Output dari setiap simpul diperoleh dengan menjumlahkan nilai input yang tertimbang dan menerapkan fungsi aktivasi seperti sigmoid. Kesalahan kuadrat dari output yang diprediksi dibandingkan dengan output aktual dihitung, dengan ukuran kesalahan yang paling sering digunakan adalah mean squared error. Dalam konteks MLP-NN, satu epoch berarti satu siklus penuh melalui seluruh dataset pelatihan, dan pelatihan sering memerlukan beberapa epoch. Satu

epoch menunjukkan jumlah iterasi yang diperlukan untuk menggunakan semua vektor pelatihan guna memperbarui bobot. Gambar 1 menunjukkan arsitektur MLP-NN dengan konfigurasi 3–4–1, yaitu tiga simpul di lapisan input, empat simpul di satu lapisan tersembunyi, dan satu simpul di lapisan output.



Gambar 2.2 Arsitektur *Multilayer Perceptron*

Koneksi antar lapisan dalam arsitektur MLP dihubungkan menggunakan bobot U_{ij} untuk koneksi dari lapisan input (x_i) ke lapisan tersembunyi (Z_j), bobot V_{jk} untuk koneksi antara lapisan tersembunyi (Z_j) yang satu ke lapisan tersembunyi (Z_j) berikutnya, dan bobot W_{kl} untuk koneksi dari lapisan tersembunyi (Z_j) ke lapisan output (Y_k). Tujuan dari proses pembelajaran dalam MLP adalah untuk menentukan bobot sinaptik yang paling optimal guna mengklasifikasikan data latih dan data validasi. Algoritma yang umum digunakan untuk pelatihan MLP adalah *backpropagation*, tetapi pada penelitian ini menggunakan *feed forwarding*.

Dalam penelitian ini menggunakan fungsi aktivasi Sigmoid. Berdasarkan (Goodfellow et al., 2016) diketahui bahwa sigmoid lebih cocok untuk hasil luaran probabilistik yang juga hasil luaran lebih stabil dan cocok pada klasifikasi dibandingkan dengan fungsi aktivasi lainnya. Sehingga, pada penelitian ini menggunakan fungsi aktivasinya adalah sigmoid yang lebih sesuai.

2.1.7 Matriks Evaluasi

Matriks evaluasi model studi ini menggunakan model *confusion matrix*, yang merupakan model evaluasi visual yang digunakan dalam *deep learning* dan salah satu model pengukuran

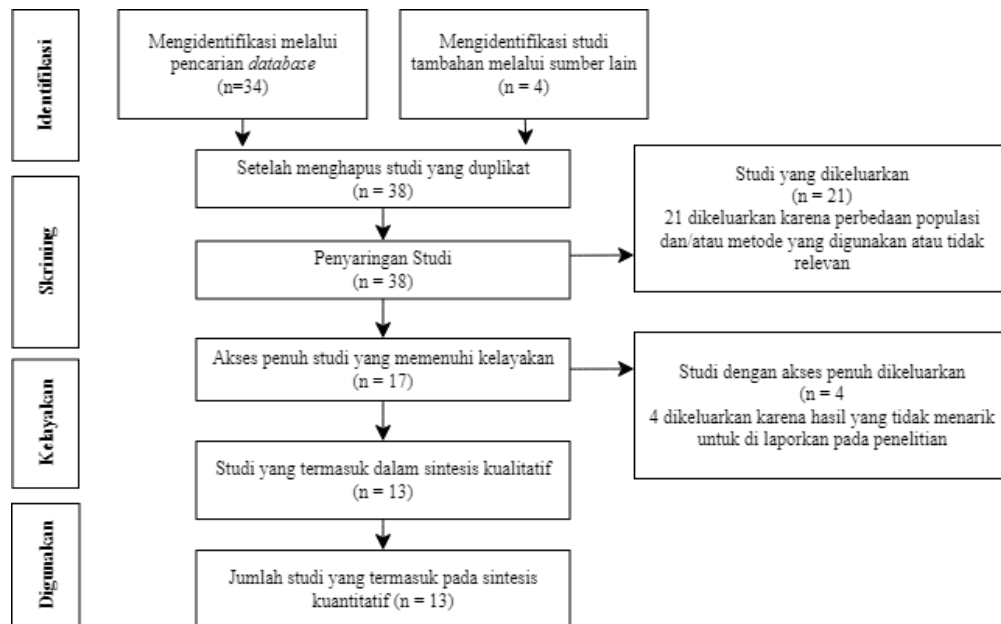
keputusan dalam kasus *supervised learning*. Tingkat *confusion* model *deep learning* di banyak kelas direpresentasikan secara visual oleh model *confusion matrix*. Menurut (Suryadewiansyah et al., 2020), model *confusion matrix* menghitung semua kelas untuk setiap label dengan menggabungkan hasil kelas aktual dan kelas yang berbeda. Model *confusion matrix* multikelas adalah model yang digunakan dalam penelitian ini. Karena ada lebih dari dua kelas dalam penelitian ini, *confusion matrix* multikelas dapat digunakan untuk menyelesaikan masalahnya. *Confusion matrix* multikelas memiliki 4 kombinasi nilai prediksi dan nilai aktual yang berbeda, yaitu *true positive*, *true negative*, *false positive*, dan *false negative* (Rahmad et al., 2020). *True positive*, *true negative*, *false positive*, dan *false negative* dapat menghasilkan akurasi, presisi, *recall*, dan *f1-score* (Markoulidakis et al., 2021).

Tabel 2.1 *Confusion Matriks*

Nama Matriks	Keterangan
<i>True Positive</i> (TP)	Hasil prediksi positif dan itu benar
<i>True Negative</i> (TN)	Hasil prediksi negatif dan itu benar
<i>False Positive</i> (FP)	Hasil prediksi positif dan itu salah
<i>False Negative</i> (FN)	Hasil prediksi negatif dan itu salah

2.2 Tinjauan Pustaka

Dalam meninjau penelitian-penelitian terdahulu, terdapat proses dalam melakukannya yang mana proses tersebut dapat dilihat pada Gambar 2.3.



Gambar 2.3 Diagram *Prisma Protocol* (da Silva et al., 2019)

Dalam suatu penelitian tentu memerlukan referensi untuk memperkuat teori-teori dalam penelitian. Selain itu, akan membantu pemahaman peneliti dalam melakukan penelitian dan dapat diketahui hubungan antara penelitian terdahulu dengan penelitian yang akan dilakukan. Penelitian terkait klasifikasi data SMS telah dilakukan sebelumnya oleh peneliti terdahulu, seperti klasifikasi data SMS spam dan klasifikasi data SMS spam atau bukan pada bahasa Inggris dan bahasa Romawi Urdu. Klasifikasi merupakan salah satu metode pembelajaran terarah, yang memetakan pada suatu masukan terhadap keluaran yang diinginkan (Drummond, 2011). Memperoleh serangkaian data yang telah dilabeli sebagai data latih dan membuat prediksi untuk semua poin yang tidak terlihat. Berikut merupakan beberapa referensi dan penelitian yang berkaitan dengan *one-time password*, *machine learning*, serta *deep learning* yang telah penulis rangkum guna membantu dalam melakukan penelitian.

Pertama, penelitian terkait dengan kerangka otentikasi dua faktor dengan menggunakan OTP-SMS yang berbasis *blockchain* (Alharbi & Alghazzawi, 2019). Pada penelitian ini menyatakan bahwa OTP-SMS merupakan salah satu skema yang sering digunakan dengan tujuan untuk menambah lapisan keamanan pada saat proses *login*. Hal ini tentunya memecah kerentanan yang hanya menggunakan hanya satu faktor untuk otentikasi. Tujuan dari penelitian ini adalah untuk menunjukkan keefektifan metode 2FA baru, OTP-

SMS, untuk mendeteksi berbagai jenis komunikasi, khususnya *Man In The Middle* (MITM) dan komunikasi antar organisasi antar pemerintah. Diketahui bahwa kerangka kerja yang diajukan lebih aman dan efektif dari serangan MITM dan pihak ketiga, serta kompleksitas dan waktu komputasinya lebih rendah daripada kerangka kerja pesaing.

Kedua, penelitian terkait dengan analisis kekuatan kata sandi *real time* pada aplikasi *web* menggunakan beberapa pendekatan *machine learning* (Farooq, 2020) Pada penelitian ini menyebutkan bahwa mudahnya penyerang untuk masuk kedalam akun dengan menebak kata sandi, hal ini dikarenakan mudahnya prediksi pola yang biasanya manusia gunakan dengan acuan seperti kata kamus, pola papan ketik, nama orang, tempat, frase yang dikenal, dan lainnya. Tujuan dari penelitian ini adalah implementasi beberapa algoritma dari *machine learning* seperti *Naïve Bayes*, *Decision Tree*, *Linear Regression*, *Neural Network* dan *Random Forest* pada aplikasi berbasis *web* secara *real time* yang memberikan saran kepada pengguna untuk memilih kata sandi yang kuat. Didapatkan bahwa *Naïve Bayes* memiliki nilai akurasi terendah yakni sebesar 87%, kemudian *Decision Tree* sebesar 99% dengan data akun sebanyak 250 akun yang mana 150 akun mengandung kekuatan kata sandi yang kuat dan 100 akun mengandung kata sandi yang lemah.

Ketiga, penelitian yang membahas tentang evaluasi skema otentikasi pengguna menggunakan biometrik perilaku dan *machine learning* (Pryor et al., 2022). Pada penelitian ini menyebutkan bahwa penyimpanan data pada perangkat ponsel meningkat pesat dalam beberapa tahun terakhir sehingga penelitian ini memiliki tujuan mengevaluasi skema biometrik otentikasi pengguna dengan menggunakan algoritma *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbor*, *Naïve Bayes*, *Logistic Regression*, *Long Short-Term Memory-Recurrent Neural Network*, dan *Multilayer Perceptron*. Didapatkan bahwa dari berbagai kombinasi dinamika sentuhan dan data sensor gerakan ponsel, model yang terbaik yakni dari algoritma *Random Forest Classifier* dengan rata-rata akurasi sebesar 97,7% tetapi *recall score* tertinggi dimiliki oleh MLP sebesar 99,9% dengan menggunakan data yang ditangkap oleh akselerometer dan magnetometer.

Keempat, penelitian yang membahas tentang “*Real Time Credit Card Transaction Analysis*” (Naik et al., 2016) Pada penelitian ini menyebutkan bahwa penipuan pada kartu kredit mengalami peningkatan secara mendunia, diketahui bahwa metode pembayaran *online* yang sudah menjadi biasa digunakan. Tujuan dari penelitian ini deteksi penipuan

kartu kredit otomatis melalui pembelajaran mesin yang mana pada penelitian ini menggunakan teorema *Naïve Bayes*. Diketahui dari penelitian ini bahwa *Bayesian Network* membantu untuk mendapatkan cakupan penipuan yang tinggi dengan tanda sinyal palsu yang rendah. Didapatkan hasil deteksi penipuan yang jauh lebih baik dalam arti terjadi deteksi yang lebih sensitif pada transaksi yang tinggi/besar setelah adanya evaluasi implementasi.

Kelima, penelitian berjudul “*A Multi-Layer Perceptron Model for Classification of E-mail Fraud*” (Oyegoke et al., 2021). Hal yang melatarbelakangi penelitian ini adanya peningkatan yang sangat signifikan pada *spam e-mail rate* oleh *fraudster* ditahun 2014 yang mencapai 70%. Deteksi *fraud* diperlukan untuk melindungi informasi, aset, akun, hingga transaksi *real time* pengguna dan perusahaan. Pada penelitian ini menggunakan metode ML untuk mendeteksi *fraud* dengan data CLAIR. Didapatkan hasil bahwa deteksi *fraud* menggunakan MLP dengan mengadopsi *back-propagation true positif* (TP) yang teridentifikasi yakni 91,4%, 83,3%, dan 99,7% pada *Spam, Ham, dan fraudulent e-mails*. Sedangkan nilai presisinya secara berturut-turut yakni 89,4%, 86%, dan 99,9%.

Keenam, penelitian berjudul “*Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures*” (El Hlouli et al., 2020). Latar belakang dari penelitian yakni diketahui bahwa semakin meningkatnya transaksi *credit card* maka meningkat pula *fraud* yang terjadi. Berdasarkan laporan BBC pada tahun 2015, *credit card fraud* sebesar \$21 juta diseluruh dunia dan meningkat pada tahun 2020 menjadi \$ 31 juta. Penelitian ini bertujuan untuk mendeteksi transaksi *fraud* menggunakan MLP dan ELM. Didapatkan hasil bahwa ELM sangat cepat untuk memprediksi transaksi *fraud* terbaru. Tetapi nilai akurasi hanya mencapai 95,46% sedangkan MLP mencapai 97,84% tingkat keakurasiannya.

Ketujuh, penelitian berjudul “*A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia*” (Chai et al., 2021). Diketahui bahwa dari data yang dikumpulkan 30.1% merupakan hipertensi dan 69.9% merupakan normal. Tujuan dari penelitian ini adalah pengembangan model jaringan saraf *multilayer perceptron* untuk klasifikasi hipertensi remaja yang berfokus pada penggunaan data antropometri dan sosiodemografi yang dikumpulkan dari studi penelitian *cross-sectional*. Didapatkan hasil

dengan menggunakan *hidden-layer* 50 mendapatkan nilai sensitifitas sebesar 41%, spesifitas 91%, presisi 65%, *F-score* 50%, akurasi 76%, dan ROC-AUC sebesar 75%.

Kedelapan, penelitian berjudul “*Spam SMS Detection Using Naïve Bayes Classifier*” (Vijay & Kumar, 2021). Hal yang melatarbelakangi penelitian ini adalah Spam SMS telah menjadi permasalahan pada pengguna *mobile phone* saat ini, seketika bahwa *mobile phone* mendapat banjir dengan SMS spam, menyebabkan pesan yang penting dan asli menjadi terlewatkan dari penglihatan pengguna. Tujuan dari penelitian ini adalah mengimplementasikan sejumlah data untuk mendeteksi SMS Spam dengan menggunakan metode *supervised machine learning naïve bayes*. Didapatkan hasil bahwa nilai akurasi yang dihasilkan sebesar 97%, kemudian sensitifitas sebesar 98% dan spesifitas 93%.

Kesembilan, penelitian berjudul “*Spam Message Classification Based on the Naïve Bayes Classification Algorithm*” (Ning et al., 2019). Hal yang melatarbelakangi penelitian ini adalah adanya peningkatan populasi dari servis pesan singkat dimana pesan *spam* juga meningkat dan serius. Tujuan dari penelitian ini adalah untuk mengklasifikasikan pesan *spam* dengan menggunakan algoritma *Naive Bayes* dengan preproses nya dengan TF-IDF. Didapatkan hasil bahwa akurasi pada *commercial* sebesar 83,68%, *other mesaage* 86,32%, *other spam* 97,86%, *mafia* 80,36%, *prostitution* 97,73%, *fraud* 93,85%, dan *political* 99,93%.

Kesepuluh, penelitian berjudul “*Extraction of Bank Transaction Data and Classification using Naïve Bayes*” (Kasi, 2019). Hal yang melatarbelakangi penelitian ini adalah klasifikasi dapat digunakan untuk menganalisis data tersebut berdasarkan PKS dan akibatnya menggunakan informasi ini untuk berbagai aplikasi. Pemrosesan data ini dapat dibuat efisien dengan mengubah data menjadi bentuk yang sesuai untuk analisis menggunakan langkah-langkah pra-pemrosesan. Tujuan dari penelitian ini adalah untuk melakukan pendekatan dengan mengekstraksi data transaksional, pra-proses menggunakan menemukan pola dan menerapkan pengklasifikasian menggunakan *Naïve Bayes*. Didapatkan hasil bahwa nilai akurasi dengan menggunakan klasifikasi *Naïve Bayes* sebesar 90,8% dan tingkat kesalahan sebesar 0,92% dengan tanpa asumsi kelas mayoritas. Didapatkan juga nilai presisi sebesar 92,7%.

Kesebelas, penelitian berjudul “*Multi-class of Bi-lingual SMS using Naïve Bayes Algorithm*” (Umair et al., 2020). Latar belakang penelitian ini adalah klasifikasi terhadap

data SMS dengan menggunakan bahasa inggris dan bahasa romawi urdu untuk mengetahui isi SMS tersebut adalah *spam* atau bukan. Tujuan dari penelitian ini adalah mengklasifikasikan data SMS kedalam *multi-class* dengan kategori *friendly*, *services*, dan *spam* dengan menggunakan algoritma *naïve bayes*. Hasil akurasi pengklasifikasian data dalam penelitian tersebut adalah sebesar 98,87%.

Keduabelas, penelitian berjudul “*Integrating Bayesian network and generalized raking for population synthesis in Greater Jakarta*” (Ilahi & Axhausen, 2019). Latar belakang penelitian ini adalah adanya permasalahan pengumpulan data tentang seluruh populasi yang kurang efisien, karena membutuhkan *cost* yang mahal dan membutuhkan waktu terutama dalam jumlah populasi yang besar. Tujuan dari penelitian ini mengefisiensikan dalam pengumpulan data populasi dengan menggunakan pendekatan yang mengintegrasikan *bayesian network* dan *generalized raking multilevel iterative proportional fitting* (IPF). Dihilangkan bahwa *bayesian network* dapat menghasilkan data yang merepresentasikan distribusi probabilitas data sampel yang berguna bagi setiap peneliti yang memiliki sampel terbatas untuk memulai dan *IPF* dapat mencocokkan menggunakan data sensus yang teragregat.

Ketigabelas, penelitian berjudul “*An interaction-Based Bayesian Network Framework for Surgical Workflow Segmentation*” (Luo et al., 2021). Hal yang melatarbelakangi penelitian ini adalah perlunya pengenalan dan pengelompokan alur kerja bedah untuk menilai keterampilan bedah serta keefektifan rumah sakit, tentunya hal ini memegang peran penting dalam memelihara dan meningkatkan sistem bedah dan perawatan kesehatan. Tujuan dari penelitian ini adalah untuk mengusulkan kerangka kerja berbasis interaksi untuk mengenali alur kerja bedah dan mengintegrasikan *bayesian network* untuk menyelesaikan masalah ketidakpastian pada pergerakan staf bedah secara *real-time* selama bedah saraf. Didapatkan hasil kinerja metode *bayesian network* yang baik dengan tingkat akurasi yang tinggi sebesar 70% yang selanjutnya secara semantik menjelaskan interaksi dan kerja sama di antara staf bedah.

Berdasarkan dari ketigabelas penelitian yang ada menunjukkan bahwa metode *machine learning* lebih dianjurkan dalam pembuatan pemodelan prediksi klasifikasi dengan struktur data yang kolomnya sedikit tetapi jumlah barisnya yang banyak. Peneliti tidak menemukan adanya penelitian terdahulu mengenai klasifikasi untuk kategori notifikasi

pengiriman pesan satu kali kata sandi (*One-Time Password*). Berdasarkan penelitian yang telah disebutkan, maka penelitian yang akan dilakukan adalah tentang klasifikasi kategori notifikasi pengiriman pesan *One-Time Password* dengan menggunakan metode *bayesian network* untuk memilih variabel dan status yang berpengaruh terhadap data transaksi dan *machine learning* untuk memperdiksi klasifikasi data, seperti: *naïve bayes*, dan *multilayer perceptron*. Tujuan penggunaan *naïve bayes*, dan *multilayer perceptron* untuk mengetahui perbandingan akurasi yang dihasilkan terhadap klasifikasi kategori notifikasi pengiriman pesan *One-Time Password*. Selain itu, metode *machine learning* maupun *deep learning* secara keberlanjutan akan lebih fleksibel ketika terdapat rencana *mendeploy* model kedalam server.

Tabel 2.2 Ulasan Kritis Tema 1.

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
1	<i>One-Time Password</i>	<i>Two Factor Authentication, One Time Password, Blockchain, Smart Contract, Ethereum, Man in the Middle Attack, Third Party</i>	Pada penelitian ini mengangkat topik tentang penerapan <i>blockchain</i> terhadap OTP-SMS, sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS.	(Alharbi & Alghazzawi, 2019)	Kerangka kerja yang diajukan lebih aman dan efektif dari serangan MITM dan pihak ketiga

Tabel 2.3 Ulasan Kritis Tema 2.

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
2	<i>Machine Learning</i>	<i>Passwords, password strength, password analysis, machine learning.</i>	Pada penelitian ini mengangkat topik tentang analisis kekuatan kata sadi <i>real time</i> dengan implementasi beberapa algoritma <i>machine learning</i> (<i>Naïve Bayes, Decision Tree, Linear Regression, Neural Network, dan Random Forest</i>), sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan menggunakan metode <i>Bayesian Network, Naïve Bayes, dan Multilayer Perceptron</i>	(Farooq, 2020)	<i>Naïve Bayes</i> memiliki nilai akurasi terendah yakni sebesar 87%, kemudian <i>Decision Tree</i> sebesar 99%
3	<i>Machine Learning</i>	<i>Behavioral biometrics, machine learning, phone movement, touch dynamics, user authentication</i> 1.	Pada penelitian ini melakukan evaluasi skema autentika pengguna menggunakan biometrik perilaku dan <i>machine learning</i> (<i>Random Forest, SVM, KNN, Naïve Bayes, Logistic Regression, LSTM-RNN, Multilayer Perceptron</i>), sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan	(Pryor et al., 2022).	<i>Random Forest Classifier</i> dengan rata-rata akurasi sebesar 97,7% tetapi <i>recall score</i> tertinggi dimiliki oleh MLP sebesar 99,9%

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
			menggunakan metode <i>Bayesian Network, Naïve Bayes, dan Multilayer Perceptron</i>		
No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
4	<i>Machine Learning</i>	<i>Online banking, Credit card fraud detection, Data mining, OTP, Naïve Bayes, Hadoop file system, Map reduce, Machine learning.</i>	Pada penelitian ini melakukan deteksi penipuan kartu kredit dengan menggunakan teorema <i>Naïve Bayes</i> dan <i>Bayesian Network</i> , sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan menggunakan metode <i>Bayesian Network, Naïve Bayes, dan Multilayer Perceptron</i>	(Naik et al., 2016)	Didapatkan dengan <i>bayesian network</i> dapat membantu cakupan penipuan yang tinggi dengan tanda sinyal palsu yang rendah yang artinya deteksi yang lebih sensitif pada transaksi yang besar
5	<i>Machine Learning</i>	<i>An Interaction-Based Bayesian Network Framework for Surgical Workflow Segmentation</i>	Pada penelitian ini melakukan pengelompokkan pada alur kerja bedah untuk menilai keterampilan bedah serta keefektifan rumah sakit dengan menggunakan metode <i>Bayesian Network</i>	(Luo & Nara, 2021)	Nilai akurasi didapatkan sebesar 70%

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
6	<i>Machine Learning</i>	<i>Integrating Bayesian Network and Generalized Raking for Population Synthesis in Greater Jakarta</i>	Pada penelitian ini membahas tentang mengefisinsikan pengumpulan data populasi dengan menggunakan metode <i>Bayesian Network, Generalized Raking Multilevel Iterative Proportional Fitting (IPF)</i>	(Ilahi & Axhau sen, 2019)	Data yang di representasikan distribusi probabilitas data sampel yang berguna bagi setiap peneliti yang memiliki sampel terbatas
7	<i>Machine Learning</i>	<i>Text message classification, Roman Urdu text, Feature extraction, Multinomial Naive Bayes, Multiple classes</i>	Pada penelitian ini membahas tentang klasifikasi data SMS kedalam <i>multi-class</i> dengan kategori <i>friendly, services</i> , dan <i>spam</i> dengan menggunakan algoritma <i>Naive Bayes</i> , sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan menggunakan metode <i>Bayesian Network, Naive Bayes</i> , dan <i>Multilayer Perceptron</i>	(Umair et al., 2020)	Hasil akurasi pengklasifikasian data dalam penelitian tersebut adalah sebesar 98,87%.
8	<i>Machine Learning</i>	<i>Extraction of Bank Transaction Data and Classification using Naive Bayes</i>	Pada penelitian ini membahas tentang mengekstraksi data transksional, pra-proses menggunakan pola dan menerapkan pengklasifikasian menggunakan algoritma <i>naive bayes</i>	(Kasi, 2019)	Nilai akurasi sebesar 90,8%, tingkat kesalahan sebesar 0,92% dan nilai presisi sebesar 92,7%

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
9	<i>Machine Learning</i>	<i>Naïve Bayesian, spam message, classification</i>	Pada penelitian ini membahas tentang klasifikasi pesan spam dengan menggunakan algoritme <i>Naïve Bayes</i> yang pada pre-prosesnya menggunakan TF-IDF, sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan menggunakan metode <i>Bayesian Network, Naïve Bayes, dan Multilayer Perceptron</i>	(Ning et al., 2019)	Akurasi pada <i>commercial</i> sebesar 83,68%, <i>other mesaage</i> 86,32%, <i>other spam</i> 97,86%, <i>mafia</i> 80,36%, <i>prostitution</i> 97,73%, <i>fraud</i> 93,85%, dan <i>political</i> 99,93%.
10	<i>Machine Learning</i>	<i>Spam SMS Detection Using Naïve Bayes Classifier</i>	Pada penelitian ini membahas tentang implementasi sejumlah data untuk mendeteksi SMS Spam dengan menggunakan metode <i>supervised machine learning naïve bayes</i> .	(Vijay & Kumar, 2021)	Nilai akurasi yang dihasilkan sebesar 97%, kemudian sensitifitas sebesar 98% dan spesifitas 93%.

Tabel 2.4 Ulasan Kritis Tema 3.

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
11	<i>Deep Learning</i>	<i>A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric</i>	Pada penelitian ini mengimplementasi penggunaan <i>online e-banking system</i> dengan meningkatkan keamanan yang terenkripsi dan OTP dengan menggunakan metode <i>Naïve Bayes</i>	(Chai et al., 2021)	Nilai sensitifitas sebesar 41%, spesifitas 91%, presisi 65%, F-score 50%, akurasi 76%, dan ROC-AUC sebesar 75%.

No.	Sub Tema	Keywords	Ulasan Kritis	Pustaka	Hasil
		<i>Measurements: A Cross- Sectional Study in Sarawak, Malaysia</i>			
12	<i>Deep Learning</i>	<i>A Multi-Layer Perceptron Model for Classification of E-mail Fraud</i>	Pada penelitian ini membahas tentang deteksi <i>fraud</i> pada transaksi <i>real time</i> dengan menggunakan metode MLP dengan adopsi <i>back-propagation</i>	(Oyegoke et al., 2021)	<i>True positif (TP)</i> yang teridentifikasi yakni 91,4%, 83,3%, dan 99,7% pada <i>Spam, Ham,</i> dan <i>fraudulent e-mails</i> . Sedangkan nilai presisinya secara berturut-turut yakni 89,4%, 86%, dan 99,9%.
13	<i>Deep Learning</i>	<i>Credit card fraud, MLP, ELM</i>	Pada penelitian ini mendeteksi <i>fraud</i> menggunakan <i>Multilayer Perceptron</i> dan <i>Extreme Learning Machine</i> , sementara dalam penelitian ini akan melakukan pengklasifikasian data SMS dengan menggunakan metode <i>Bayesian Network, Naïve Bayes,</i> dan <i>Multilayer Perceptron</i>	(El Hlouli et al., 2020)	Akurasi hanya mencapai 95,46% menggunakan ELM sedangkan MLP mencapai 97,84% tingkat keakurasiannya.

Berdasarkan dari penelitian yang telah dilakukan oleh (Alharbi & Alghazzawi, 2019) menggunakan studi kasus *two factor authentication OTP, blockchain* sedangkan pada penelitian ini menggunakan data pengiriman pesan singkat OTP dengan menggunakan metode *bayesian network, naïve bayes* dan MLP, maka kebaruan dari penelitian ini adalah usual variabel utama dalam penyusun kesuksesan/gagal pengiriman pesan singkat OTP dan prediksi dari klasifikasi/kategori mendatang pada transaksi pengiriman pesan.

Berdasarkan dari penelitian yang telah dilakukan oleh (Farooq, 2020; Ilahi & Axhausen, 2019; Kasi, 2019; Luo et al., 2021; Naik et al., 2016; Pryor et al., 2022; Umair et al., 2020; Vijay & Kumar, 2021) menggunakan metode *bayesian network* dan *naïve bayes*, yang membedakan dari penelitian sebelumnya dengan penelitian adalah menggunakan data pengiriman pesan singkat OTP dengan menggunakan metode *bayesian network, naïve bayes* dan MLP, maka kebaruan dari penelitian ini adalah usul variabel utama dalam penyusun kesuksesan/gagal pengiriman pesan singkat OTP dan prediksi dari klasifikasi/kategori mendatang pada transaksi pengiriman pesan.

Berdasarkan dari penelitian yang telah dilakukan oleh (Chai et al., 2021; El Hlouli et al., 2020; Oyegoke et al., 2021) menggunakan metode *multilayer perceptron*, yang membedakan dari penelitian sebelumnya dengan penelitian adalah menggunakan data pengiriman pesan singkat OTP dengan menggunakan metode *bayesian network, naïve bayes* dan MLP, maka kebaruan dari penelitian ini adalah usual variabel utama dalam penyusun kesuksesan/gagal pengiriman pesan singkat OTP dan prediksi dari klasifikasi/kategori mendatang pada transaksi pengiriman pesan.

Selain dari ketigabelas tinjauan pustaka yang telah dilakukan, terdapat tambahan studi literatur sebagai penguat dalam hasil dan pembahasan.

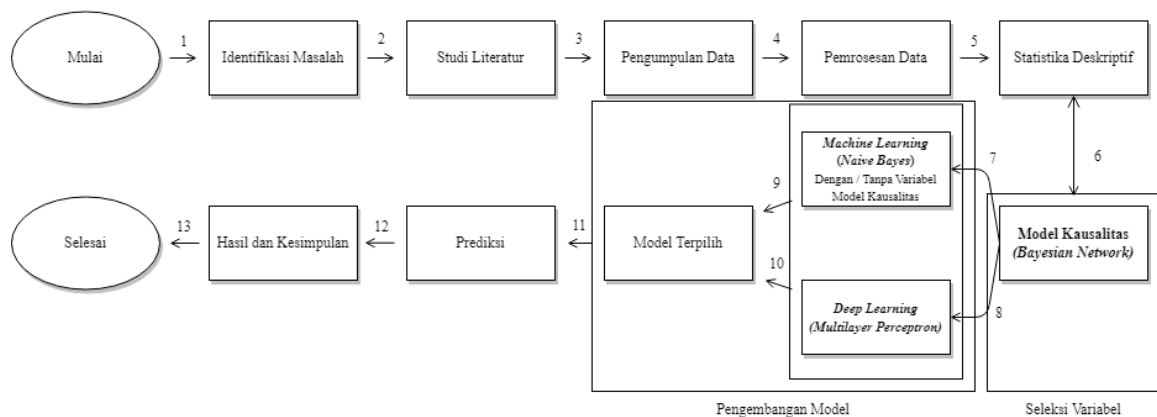
Berdasarkan (Adugna et al., 2006; Domingos & Pazzani, 1997; Haykin, 2009) pada implementasi yang mudah untuk mengetahui hubungan antar fitur, efisiensi secara komputasi, menangani data kategorik, asumsi dari fitur-fiturnya independen, dan hasil prediksi yang lebih stabil maka metode *naive bayes* menjadi pilihan yang tepat untuk digunakan. Sedangkan pada metode *multilayer perceptron* digunakan untuk mempelajari pola hubungan yang kompleks pada input fiturnya, kemudian dapat menangani tipe data yang bercampur dan tidak ada asumsi yang ketat yang mana MLP dapat mempelajari pola data secara langsung dari data yang dilatih. Jika dibandingkan metode *deep learning*

lainnya, metode MLP ini memiliki potensi akurasi yang tinggi karena menggunakan multi-layer dalam iterasi yang dilakukan kemudian dapat di tingkatkan jumlah lapisan dan *neuron* nya untuk menangkap pola yang lebih kompleks, sehingga memungkinkan untuk mendapatkan akurasi yang tinggi.

BAB 3

Metodologi

Pada penelitian ini mencoba untuk mereduksi variabel dengan melihat hubungan kausalitas antar variabel yang digunakan menggunakan metode *Bayesian Network* yang kemudian dilanjutkan dengan menggunakan metode *machine learning* yakni *Naïve Bayes* untuk mendapatkan model klasifikasi ringkasan akhir dari penggabungan data yang dibandingkan dengan *deep learning* yakni *Multi-Layer Perceptron*. Perbandingan metode dilakukan untuk mengetahui berdasarkan dari data yang ada apakah cukup dengan metode *machine learning* dengan hasil reduksi variabel menggunakan *Bayesian Network* atau memerlukan *deep learning* dalam mereduksi serta mengklasifikasi transaksi pengiriman SMS *one-time password*. Secara umum alur penelitian yang akan dilakukan sebagai berikut.



Gambar 3.1 Diagram Alir Penelitian (Tarigan et al., 2023).

Keterangan:

- (1). *Input* : Identifikasi masalah topik penelitian.
- (2). *Output* : Rumusan masalah, tujuan penelitian, hipotesis penelitian
Input : Metode penelitian.
- (3). *Output* : Penelitian-penelitian terdahulu yang berhubungan dengan *output* pada poin 2.
Input : Variabel-variabel yang digunakan pada penelitian.
- (4). *Output* : Data terkumpulkan.
Input : *Raw data*.
- (5). *Output* : Data telah terorganisir serta sudah memenuhi kebutuhan metode yang digunakan.

- Input* : Data yang telah terorganisir.
- (6). *Output* : Ringkasan terkait data yang telah di *input* pada poin 5.
Input : Data yang telah terorganisir.
- (7). *Output* : Probabilitas kausalitas variabel-variabel dan features
Input : Data yang telah terorganisir dengan pembagian 80% data latih, 20% data uji.
- (8). *Output* : Probabilitas kausalitas variabel-variabel dan features
Input : Data yang telah terorganisir dengan pembagian 80% data latih, 20% data uji.
- (9). *Output* : Model prediktif klasifikasi serta nilai akurasi model dengan menggunakan metode *machine learning (naïve bayes classifier)*.
Input : Nilai akurasi model.
- (10). *Output* : Model prediktif klasifikasi serta nilai akurasi model dengan menggunakan metode *deep learning (multilayer perceptron)*
Input : Nilai akurasi model.
- (11). *Output* : Nilai akurasi model tertinggi yang dipilih.
Input : Data uji 20%.
- (12). *Output* : Hasil prediksi klasifikasi yang dihasilkan dari pelatihan model yang terpilih.
Input : Hasil penelitian yang dilakukan (model, nilai akurasi model, evaluasi model).
- (13). *Output* : Hasil dan kesimpulan topik penelitian.

3.1 Pengumpulan Data

3.1.1 Populasi dan Sampel Data

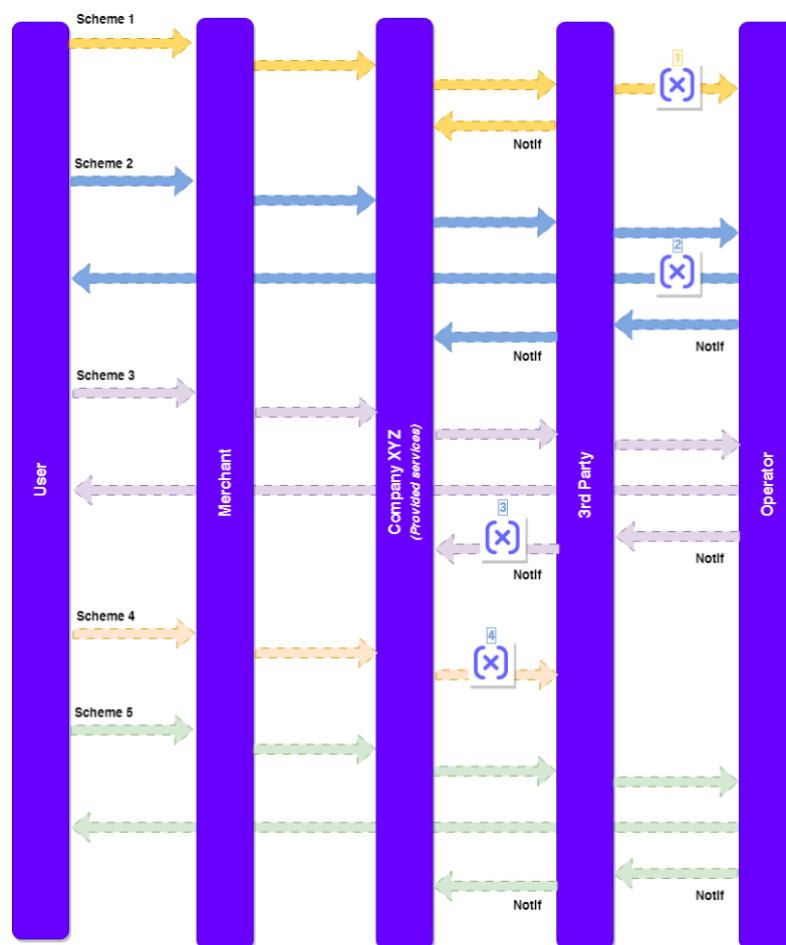
Penelitian menggunakan populasi data transaksi pengiriman SMS *one-time password* di Indonesia pada perusahaan XYZ. Sampel yang digunakan yakni data transaksi pengiriman SMS *one-time password* di Indonesia pada bulan Juli hingga November 2021, dengan jumlah variabel sebanyak 14, jumlah baris data sebanyak 259.839 dengan kombinasi tipe data numerik dan kategorik.

3.1.2 Sumber Data

Data yang digunakan pada penelitian ini berasal dari perusahaan XYZ yang bergerak dibidang teknologi IT dan salah satu produknya menyediakan pengiriman SMS *one-time password*.

3.1.3 Aliran Data

Berikut aliran data yang dimiliki oleh perusahaan XYZ dalam mendapatkan data.



Gambar 3.2 Diagram Alir Data Transaksi SMS OTP

Berdasarkan Gambar 3.2 dapat diketahui sebagai berikut :

- Skema 1: Ketika pihak ketiga request ke operator untuk mengirimkan pesan (tidak dapat respon/gagal) maka pihak ketiga akan mengirimkan notifikasi gagal ke perusahaan XYZ.
- Skema 2: Ketika pihak ketiga sudah berhasil mengirimkan ke operator namun,

operator gagal mengirimkan ke *user*.

- Skema 3: Pesan telah berhasil dikirimkan oleh operator. Tetapi, perusahaan XYZ telat mendapatkan *delivery status* dari pihak ketiga.
- Skema 4: Perusahaan XYZ gagal mengirimkan *request* (kirim SMS) ke pihak ketiga
- Skema 5: SMS OTP yang telah berhasil dikirimkan oleh pihak ketiga

3.1.4 Variabel Penelitian

Variabel yang digunakan pada penelitian ini terdapat 14 variabel, semua variabel yang digunakan didapatkan dari PT XYZ dengan rekomendasi dan olahan dari beberapa variabel yang mempunyai nilai ambang batas tersendiri dari ahli *engineer* nya.

Tabel 3.1 Variabel Penelitian

No	Variabel	Defenisi	Tipe Data
1	Trans Date	Tanggal transaksi SMS OTP dilakukan	Time
2	Trans Time	Waktu transaksi SMS OTP dilakukan	Time
3	Sent	Jumlah status SMS OTP yang dikirim atas permintaan merchant (status pengiriman)	Varchar2
4	Delivered	Jumlah Status Akhir SMS OTP saat sudah terkirim dan diterima oleh pengguna (<i>delivery status</i>)	Varchar2
5	Undelivered	Jumlah status SMS terakhir saat tidak terkirim SMS OTP (status pengiriman)	Varchar2
6	Unknown	Jumlah status SMS terakhir ketika nomor ponsel tujuan SMS OTP (status pengiriman) tidak diketahui	Varchar2
7	Fail_DN	Jumlah pemberitahuan status pengiriman tidak diterima oleh PT XYZ dari penyedia pihak ketiga	Varchar2
8	Fail_TO	Jumlah pemberitahuan pengiriman status batas waktu dari penyedia pihak ketiga ke PT XYZ	Varchar2
9	AVG_NOTI_PER_SECOND	Rata-rata notifikasi dari penyedia pihak ketiga per detik	Varchar2
10	RATE_DELIV	Nilai ambang batas transaksi SMS OTP yang berhasil terkirim	Varchar2
11	RATE_UNDELIVERED	Nilai ambang batas terhadap transaksi OTP SMS yang gagal terkirim	Varchar2
12	AVG_NOTI_PER_MINUTE	Rata-rata notifikasi dari penyedia pihak ketiga per menit	Varchar2
13	SUMMARY	Ringkasan penggabungan data dari <i>summary_status</i> dan <i>summary_noti_from_3rdParty</i>	Varchar2
14	DAY_CAT	Kategori transaksi dilakukan pagi atau sore hari	Varchar2

Tabel 3.2 Variabel Status (*Summary*) Penelitian

No	Variabel	Defenisi	Tipe Data
1	<i>Not Delay Notification</i>	Tidak ada Delay yang terjadi pada notifikasi (pengiriman)	Nominal

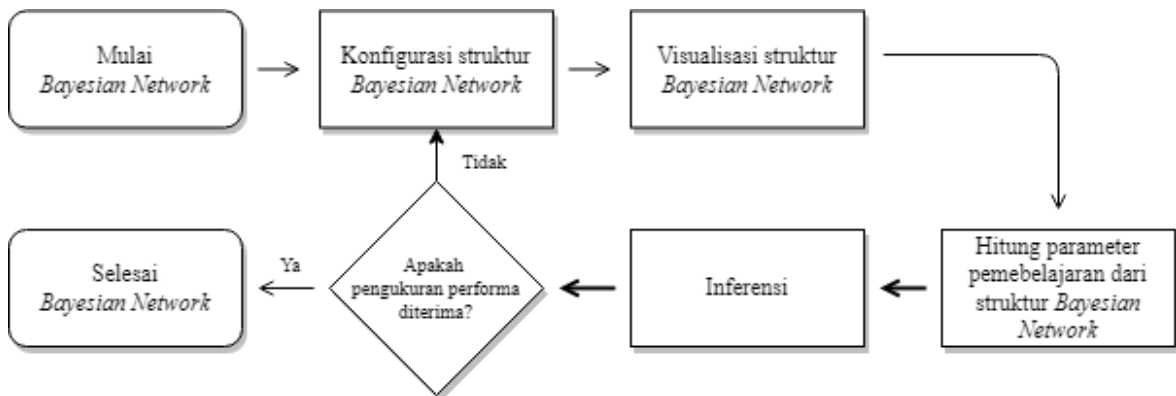
No	Variabel	Defenisi	Type Data
		cepat) <4 detik, (tetapi kecenderungan banyak dalam millisecond)	
2	<i>Sangat Normal Notification</i>	Tidak ada delay yang terjadi pada notifikasi (sangat normal) 5 detik	Nominal
3	<i>Unknown Summary Notification</i>	Tidak dapat didefinisikan karena tidak tahu status notifikasi baik dari 3rd party maupun operator, tetapi dari sisi company XYZ sudah terkirim	Nominal
4	<i>Normal Delay Notification</i>	Pengiriman notifikasi selama 8 detik	Nominal
5	<i>Long Delay Notification</i>	Pengiriman notifikasi selama 25 - 45 menit	Nominal
6	<i>Maybe Delay From 3rd Party</i>	Pengiriman notifikasi dalam kurun waktu 1 - 10 menit	Nominal
7	<i>Normal Notification</i>	Pengiriman notifikasi di antara 5-7 detik	Nominal
8	<i>Timeout to 3rd Party</i>	Pesan gagal di request oleh company XYZ ke 3rd Party	Nominal
9	<i>Average Delay Notification</i>	Pengiriman notifikasi 8-15 detik	Nominal
10	<i>Maybe Issue From Operator</i>	Pesan gagal di request oleh 3rd Party atau Operator gagal mengirimkan pesan ke user	Nominal
11	<i>Not Normal Notification</i>	Terjadi delay yang tidak menentu (diatas rata-rata delay notifikasi) pada notifikasi (pengiriman cepat dalam milisecond) >45 menit	Nominal

3.2 Metodologi yang Diusulkan

Pada Gambar 3.2 menunjukkan proses transaksi pengiriman pesan OTP yang dibagi menjadi beberapa skema. Setiap skema yang terjadi terdapat proses *input* hingga *output*, hal ini mengartikan bahwa setiap proses mengandung sifat sebab dan akibat yang di representasikan pada ke 5 skema yang dihasilkan dari proses-proses tersebut. Untuk menguji dan identifikasi proses bisnis ini secara ilmiah, dengan menggunakan data yang dihasilkan dari setiap proses (bisnis OTP), maka metode *bayesian network* merupakan metode yang dapat digunakan untuk menjawab tujuan tersebut. Selain dari hasil pengujian, metode *bayesian network* juga akan digunakan untuk mereduksi variabel status transaksi dengan mengambil variabel yang signifikan berdasarkan *parent nodes* dengan jumlah *child nodes* yang mempengaruhinya.

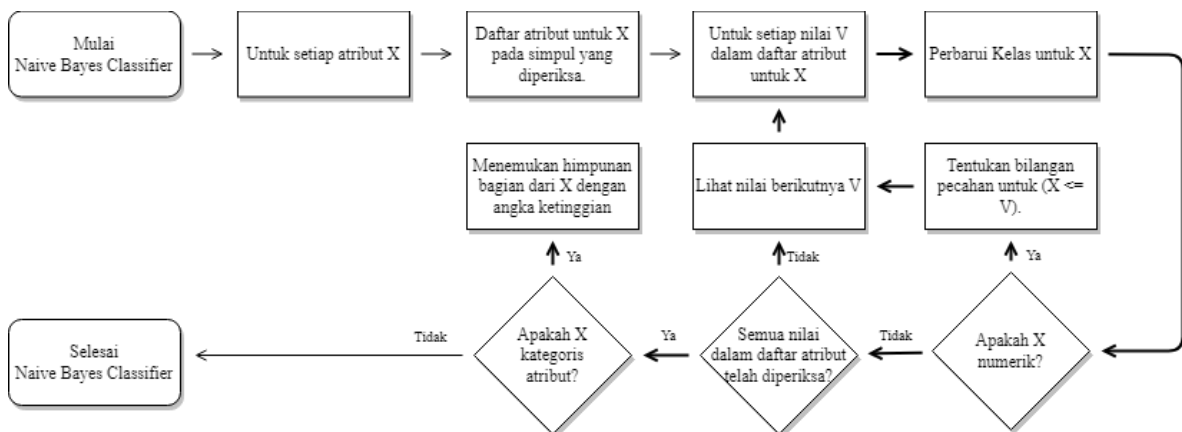
Bayesian Network adalah representasi grafis dari hubungan probabilitas antara variabel yang relevan. Model grafis memberikan sejumlah manfaat bila digabungkan dengan teknik statistik untuk pemodelan data. Pertama, model dengan mudah mengakomodasi contoh di mana entri data tertentu tidak ada karena mengkodekan ketergantungan antara semua variabel. Selanjutnya dengan menggunakan Bayesian Network untuk mempelajari

hubungan sebab akibat, seseorang dapat memahami area masalah dan memperkirakan efek dari suatu tindakan. Kemudian, Model adalah representasi sempurna untuk menggabungkan pengetahuan sebelumnya (yang sering berbentuk hubungan sebab akibat) dan data karena memiliki semantik kausal dan probabilistik. Terakhir, teknik statistik Bayesian dikombinasikan dengan Bayesian Network memberikan strategi praktis dan efektif untuk mencegah overfitting data (Heckerman, 1997).



Gambar 3.3 Diagram Alir Bayesian Network.

Naïve Bayes adalah algoritma pembelajaran langsung yang memanfaatkan aturan Bayes dan anggapan mendasar, mengingat kelasnya, atributnya independen secara kondisional. Terlepas dari kenyataan bahwa dalam praktiknya asumsi independensi ini sering dipatahkan, Naïve Bayes sering kali menghasilkan akurasi klasifikasi yang kompetitif. Selain itu, keefektifan komputasinya dan banyak karakteristik menarik lainnya, berkontribusi pada penggunaan luas Naïve Bayes dalam praktik (Webb, 2017).

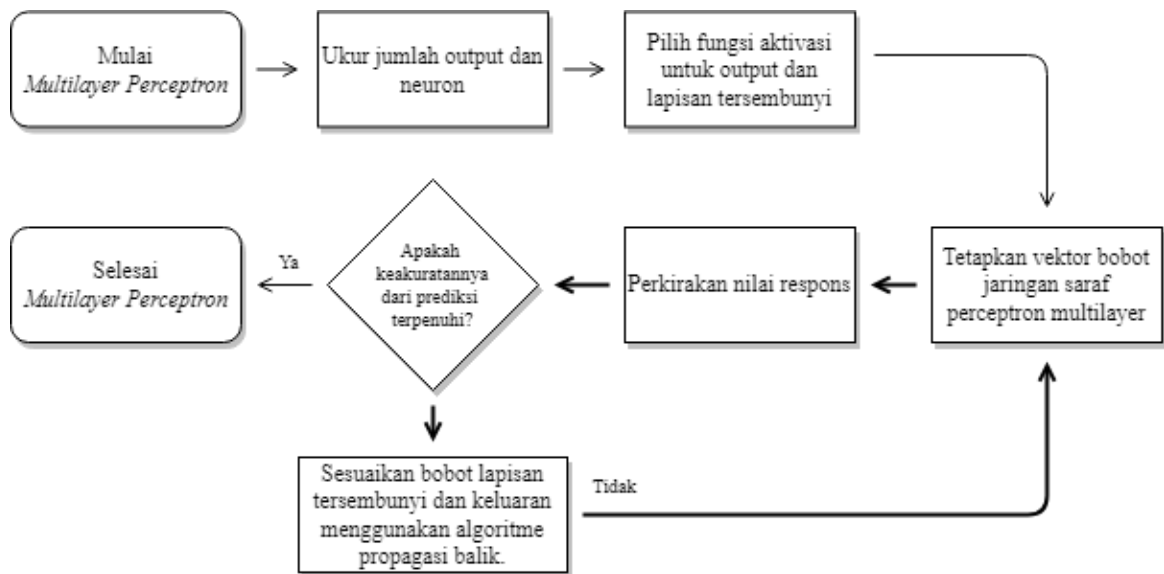


Gambar 3.4 Diagram Alir Naïve Bayes.

Multi-Layer Perceptron (MLP) merupakan salah satu popular algoritma yang digunakan dalam pembuatan model untuk supervised learning dengan menggunakan arsitektur dari algoritma Artificial Neural Networks (ANNs).

Algoritma pembelajaran MLP yang terkenal adalah metode backpropagation, yang merupakan generalisasi dari aturan Least Mean Squared, di antara banyak algoritma lainnya. Backpropagation adalah proses penyebaran kesalahan dari lapisan ke lapisan, dimulai dengan lapisan keluaran dan bergerak mundur, untuk memperbaiki bobot.

Selain itu, MLP juga memiliki metode *feedforward* multilayer yang mana Informasi mengalir secara tidak langsung dari lapisan input ke lapisan output melalui lapisan tersembunyi. Setiap jaringan membawa bobotnya sendiri dengan menggunakan fungsi aktivasi perceptron pada lapisan yang sama identik, ini adalah sigmoid untuk lapisan tersembunyi secara umum. Lapisan keluaran juga dapat berupa sigmoid atau fungsi liner, tergantung pada pengaplikasiannya.



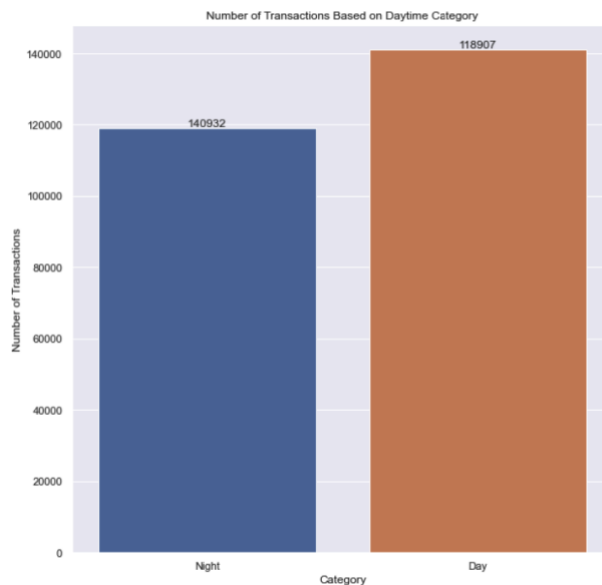
Gambar 3.5 Diagram Alir Multi-Layer Perceptron.

BAB 4

Hasil dan Pembahasan

4.1 Data Eksplorasi

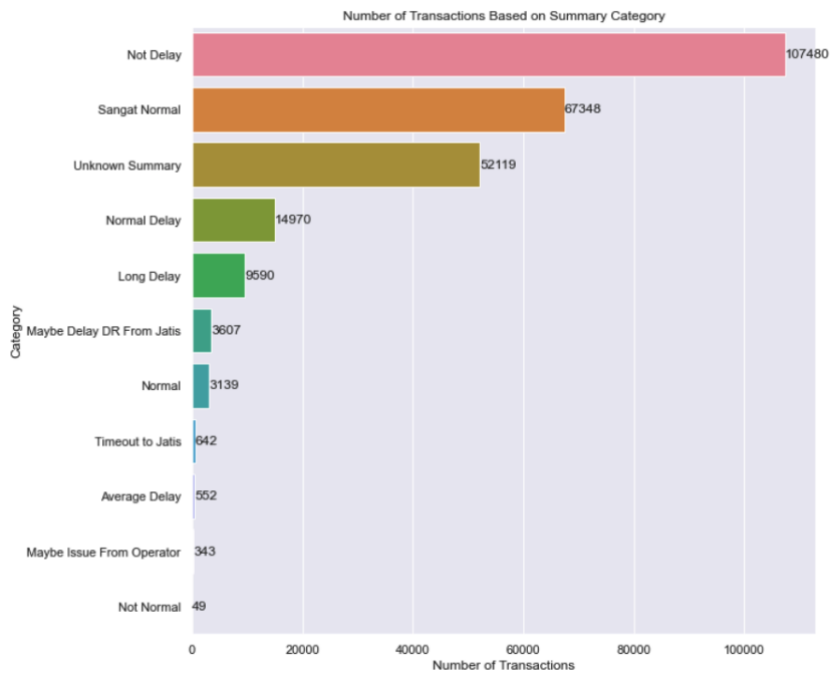
Pada analisis data eksplorasi ini memiliki tujuan untuk mendapatkan pemahaman umum terkait dengan karakteristik pada *dataset* yang digunakan. Berikut merupakan hasil yang didapatkan.



Gambar 4.1 Jumlah Transaksi Berdasarkan Kategori Waktu

Berdasarkan Gambar 4.1 dapat diketahui bahwa transaksi banyak dilakukan dipagi hari (54.2%), sedangkan transaksi dimalam hari (45.76%).

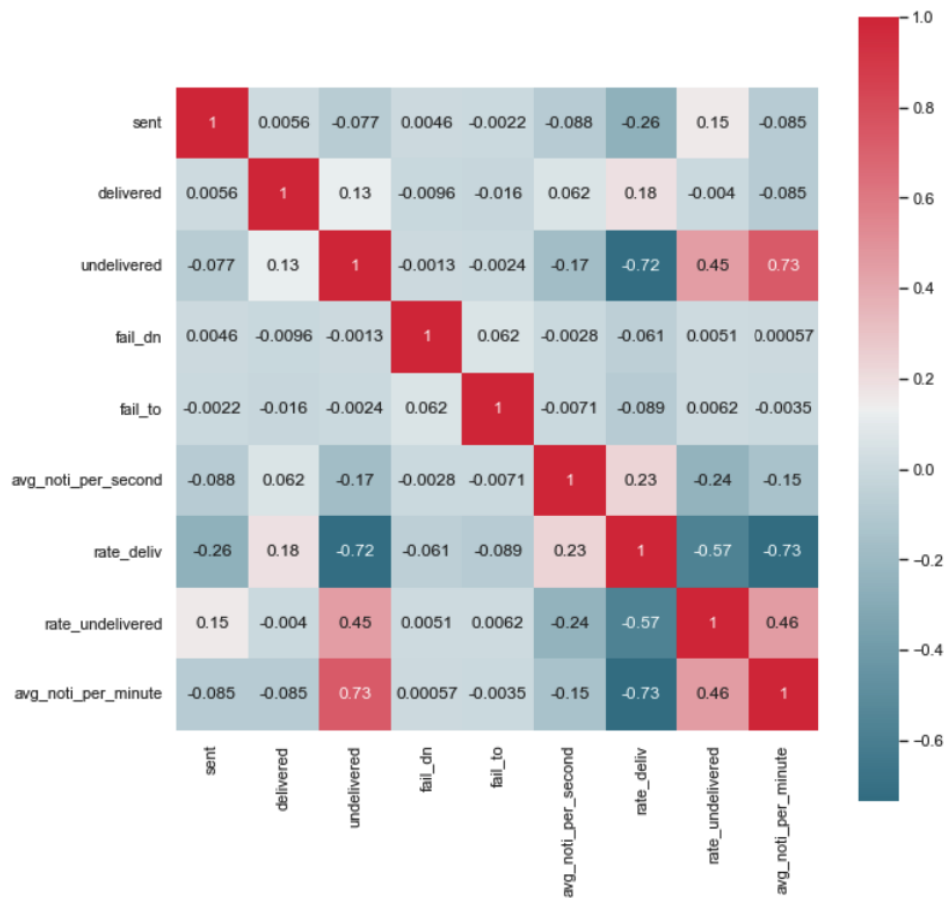
Secara detail diketahui bahwa waktu transaksi 3 teratas terjadi pada waktu jam 19:00, 20:00, 21:00. Kemudian dilanjutkan pada waktu 13:00, 15:00, dan 18:00. Secara selisih persentase, tidak begitu signifikan gap antar waktu transaksi yang terjadi, akan tetapi memang distribusi pengguna transaksi pesan OTP kecenderungan hampir merata pada sejak jam 9:00 – 18:00 dikisaran ~20rb transaksi perharinya. Hal ini pula didukung pada (Lidwina, 2020) menyatakan bahwa lonjakan transaksi (aktivitas belanja konsumen di *e-commerce*) terjadi sejak jam 9:00 hingga 15:00 kemudian melandai beberapa saat, yang kemudian terjadi lonjakan transaksi lagi pada jam 18:00 hingga 21:00.



Gambar 4.2 Jumlah Transaksi Berdasarkan Kategori Ringkasan

Berdasarkan Gambar 4.2 dapat diketahui bahwa transaksi pengiriman pesan OTP pada 2 teratas pengiriman secara lancar (Tidak delay [41.36%], Sangat Normal [25.92%]), tetapi pada peringkat ke 3 tidak ketahui statusnya (antara kegagalan pengiriman dari pihak ke 3 ke operator ataupun dari operator mengirimkan status ke pihak ke 3) sebesar 20.06%.

Diketahui juga dari Gambar 4.2, kecenderungan dari pengiriman pesan OTP masih sesuai dengan kebutuhan bisnis, akan tetapi pada status yang tidak diketahui perlu di telusuri lebih lanjut karena tidak ada keterangan yang secara eksplisit apakah gagal atau berhasil yang mana status ini terbentuk adanya pengiriman pesan dari perusahaan penyedia jasa OTP tetapi tidak mendapatkan umpan balik baik dari pihak operator atau dari pihak ke-3.



Gambar 4.3 Korelasi Antar Variabel Transaksi Pengiriman Pesan OTP

Berdasarkan Gambar 4.3 dapat diketahui bahwa variabel yang memiliki hubungan kuat yakni *rate delivery* dengan *undelivered* (-72%), kemudian rata-rata notifikasi per menit dengan *undelivered* (73%), dan rata-rata notifikasi per menit dengan *rate delivery* (-73%).

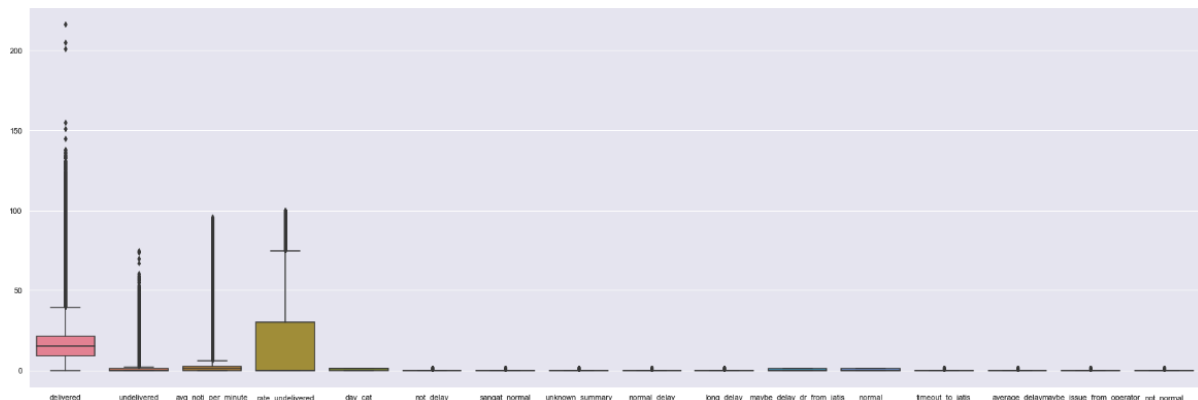
Sedangkan yang memiliki korelasi sedang yakni *rate undelivered* dengan *rate delivery* (-57%), rata-rata notifikasi per menit dengan *rate undelivered* (46%), *rate undelivered* dengan *undelivered* (45%), kemudian untuk korelasi lemah yakni *rate delivery* dengan rata-rata notifikasi per detik (23%) *rate undelivered* dengan rata-rata notifikasi per detik (-24%), dan rata-rata notifikasi per menit dengan rata-rata notifikasi per detik (-15%).

Berdasarkan (Schober & Schwarte, 2018) disebutkan bahwa korelasi yang berada di 10-39% merupakan korelasi lemah, 40-69% merupakan korelasi sedang, kemudian 70-80% merupakan korelasi kuat. Hal ini mengartikan bahwa setiap perubahan yang terjadi pada variabel *rate delivery* akan mempengaruhi *undelivered* secara negatif begitu juga berlaku sebaliknya (apabila *rate delivery* meningkat maka jumlah *undelivered* akan menurun).

Selanjutnya, setiap perubahan yang terjadi pada rata-rata notifikasi per menit akan mempengaruhi signifikan terhadap perubahan pada variabel *undelivered*. Kemudian, setiap perubahan pada variabel rata-rata notifikasi per menit akan mempengaruhi *rate delivery* secara negatif.

Pada korelasi sedang merepresentasikan bahwa pengaruh antar variabel yang saling terikat bisa terjadi untuk meramalkan satu variabel secara akurat berdasarkan yang lain dan/atau tidak mempengaruhi antar variabel, karena banyak faktor lain yang juga dapat mempengaruhi hubungan antar variabel. Dalam hal ini variabel *rate undelivered* dengan *rate delivery* yang masuk kedalam kategori korelasi sedang dimana ada kemungkinan bahwa setiap perubahan yang terjadi pada variabel *rate undelivered* akan mempengaruhi *rate delivery* dan berlaku sebaliknya. Selanjutnya ada rata-rata notifikasi permenit dengan *rate undelivery* yang masuk kedalam kategori korelasi sedang yang menagartikan ada kemungkinan setiap perubahan yang terjadi pada variabel rata-rata notifikasi permenit dengan *rate undelivery* dan berlaku sebaliknya.

Pada korelasi lemah merepresentasikan bahwa pengaruh antar variabel yang saling terikat sangat kecil kemungkinannya untuk mempengaruhi antar variabel (sangat banyak faktor yang lain juga dapat mempengaruhi hubungan antar variabel).



Gambar 4.4 Box plot untuk identifikasi *outlier*

Diketahui pada Gambar 4.4 terdapat banyak data *outlier* pada variabel *delivered*, *undelivered*, rata-rata notifikasi permenit, dan *rate undelivered*.

4.2 Kategorisasi Data

Diketahui dari Gambar 4.4 data yang digunakan mengandung banyak *outlier* (data ekstrim), untuk memitigasi hal ini, maka dapat menggunakan kategorisasi data supaya persebaran data lebih baik dan model yang dihasilkan optimal.

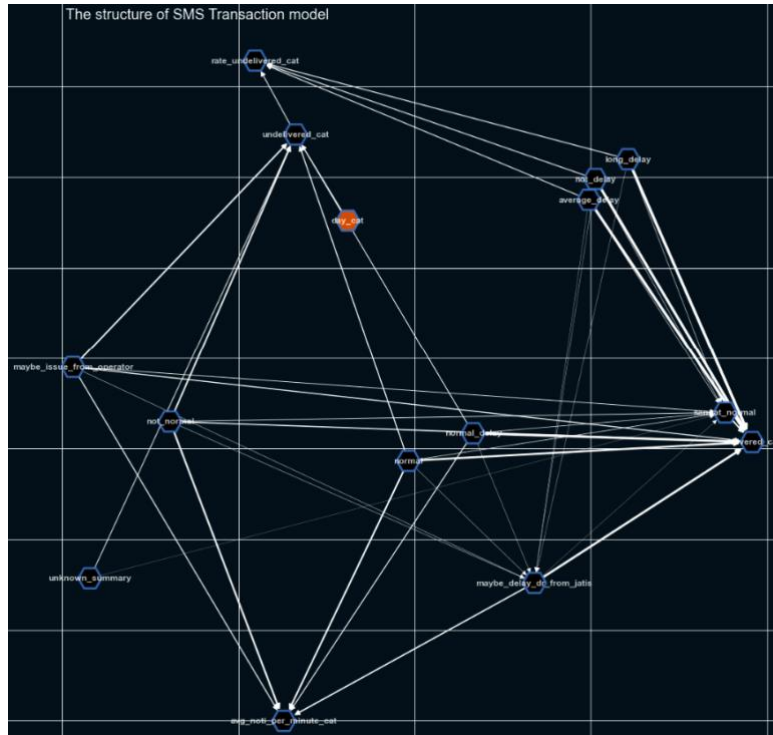
Tabel 4.1 Menentukan Kategori Baru dengan Menggunakan Persentil Data

Variabel	Kategori Angka	Rentang Nilai	Anotasi	Persentil	Label
Delivered Category	1	0 - 5	<i>Delivered Message</i>	10%	<i>Very Low</i>
	2	6-11	<i>Delivered Message</i>	35%	<i>Low</i>
	3	12-18	<i>Delivered Message</i>	65%	<i>Average</i>
	4	19-26	<i>Delivered Message</i>	89%	<i>High</i>
	5	27-16	<i>Delivered Message</i>	100%	<i>Very High</i>
Undelivered Category	1	0 - 1	<i>Undelivered Message</i>	75%	<i>Low</i>
	2	2-7	<i>Undelivered Message</i>	100%	<i>High</i>
Average Notification per minutes	1	0 - 0.01	<i>Minutes</i>	25%	<i>Direct</i>
	2	0.02 - 2.42	<i>Minutes</i>	75%	<i>Normal</i>
	3	2.43 - 96	<i>Minutes</i>	100%	<i>Long Delay</i>
Rate Undelivered	1	0-5	<i>Percent</i>	65%	<i>Very Low</i>
	2	6-100	<i>Percent</i>	100%	<i>Very High</i>

Dalam mencacah data menjadi beberapa kategori (pada kolom kategori numerik dan lable) pada **Table 4.1** menggunakan metode persentil. Pendekatan dalam penentuan persentase angka irisan merujuk kepada kurva distribusi normal, nilai distribusi yang dilihat adalah pendistribusian banyaknya data (*modus*).

4.3 Bayesian Network

Bayesian network digunakan dalam penelitian ini untuk mendapatkan mereduksi variabel-variabel yang digunakan dalam penelitian ini, selain itu juga dengan menggunakan metode ini dapat mengeksplorasi variabel-variabel apa saja yang memiliki hubungan kausalitasnya.



Gambar 4.5 Visualisasi *network* dari model Bayesian Network

Berdasarkan Gambar 4.5 dapat diketahui bahwa terdapat 6 variabel yang menjadi *parent* atau variabel utama, variabel tersebut yakni kategori *undelivered*, *rate undelivered category*, *average notification per minutes category*, *status maybe delay from 3rd party*, status sangat normal dan *delivered category*.

Variabel kategori *undelivered* mengartikan bahwa sejumlah pesan yang tidak terkirim, yang mana variabel ini menjadi variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut yakni status *maybe issue from operator*, status *not normal*, status *unknown summary*, kategori *day* dan status normal.

Kemudian pada kategori *rate undelivered* mengartikan bahwa persentase pesan yang tidak terkirim, yang mana kategori ini menjadi variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut yakni kategori *undelivered*, status *average delay*, status *not delay* dan status *long delay*.

Kemudian kategori *average notification per minutes* mengartikan bahwa rata-rata pesan notifikasi terkirim dalam menit, yang mana kategori ini menjadi variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut

yakni status *not normal*, status *maybe issue from operator*, status *normal*, status *normal delay*, dan status *maybe delay from 3rd party*.

Kemudian status *maybe delay from 3rd party* mengartikan bahwa status pengiriman pesan yang telah dikirim perusahaan penyedia layanan OTP mendapatkan umpan balik dari pihak ke-3 yang mengirimkan durasinya diatas rata-rata waktu sukses pengiriman pesan otp, yang mana status ini menjadi salah satu variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut yakni status *not normal*, status *maybe issue from operator*, status *normal*, status *normal delay*, status *average delay*, status *not delay*, dan status *long delay*.

Kemudian status sangat normal mengartikan bahwa status dari pengiriman pesan OTP berjalan sangat lancar tanpa ada jeda waktu yang signifikan atau jeda waktu yang membuat pengguna menunggu. Status ini menjadi salah satu variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut yakni status *average delay*, status *not delay*, status *long delay*, status *normal delay*, status *maybe delay from 3rd party*, status *maybe issue form operator*, status *normal*, status *unknown summary* dan status *not normal*.

Terakhir, kategori *delivered* mengartikan bahwa pengiriman sejumlah pesan OTP yang berhasil diterima oleh pengguna. Kategori ini menjadi salah satu variabel utama yang disusun atau dipengaruhi oleh beberapa variabel turunannya (*child nodes*), variabel tersebut yakni status *normal delay*, status *normal*, status *maybe delay from 3rd party*, status *not normal*, status *maybe issue from operator*, status sangat normal, status *long delay*, status *not delay*, dan status *average delay*.

Secara lebih detail pembagian antara *parent* dan *child nodes* dapat dilihat pada Table 4.3.

Tabel 4.2 *Parent* dan *Child nodes*

Titik Induk	Titik Turunan
<i>undelivered_cat</i>	<i>rate_undelivered_cat</i>
<i>day_cat</i>	<i>undelivered_cat</i>
	<i>delivered_cat</i>
<i>not_delay</i>	<i>rate_undelivered_cat</i>
	<i>sangat_normal</i>

Titik Induk	Titik Turunan
	<i>maybe_delay_dr_from_3rd party</i>
<i>sangat_normal</i>	<i>delivered_cat</i>
<i>unknown_summary</i>	<i>undelivered_cat</i>
	<i>sangat_normal</i>
	<i>delivered_cat</i>
<i>normal_delay</i>	<i>undelivered_cat</i>
	<i>avg_noti_per_minute_cat</i>
	<i>sangat_normal</i>
	<i>maybe_delay_dr_from_3rd party</i>
	<i>delivered_cat</i>
<i>long_delay</i>	<i>rate_undelivered_cat</i>
	<i>sangat_normal</i>
	<i>maybe_delay_dr_from_3rd party</i>
<i>maybe_delay_dr_from_3rd party</i>	<i>delivered_cat</i>
	<i>avg_noti_per_minute_cat</i>
	<i>sangat_normal</i>
	<i>delivered_cat</i>
<i>normal</i>	<i>undelivered_cat</i>
	<i>avg_noti_per_minute_cat</i>
	<i>sangat_normal</i>
	<i>maybe_delay_dr_from_3rd party</i>
	<i>delivered_cat</i>
<i>average_delay</i>	<i>rate_undelivered_cat</i>
	<i>sangat_normal</i>
	<i>maybe_delay_dr_from_3rd party</i>
	<i>delivered_cat</i>
<i>maybe_issue_from_operator</i>	<i>undelivered_cat</i>
	<i>avg_noti_per_minute_cat</i>
	<i>sangat_normal</i>

Titik Induk	Titik Turunan
	<i>maybe_delay_dr_from_3rd party</i>
	<i>delivered_cat</i>
	<i>undelivered_cat</i>
<i>not_normal</i>	<i>avg_noti_per_minute_cat</i>
	<i>sangat_normal</i>
	<i>maybe_delay_dr_from_3rd party</i>

Berdasarkan Table 4.3 dapat diketahui bahwa:

1. Hampir seluruh *parent nodes* dipengaruhi oleh variabel *delivered category* kecuali *day category* dan *undelivered category*.
2. Banyak *parent nodes* dipengaruhi oleh *undelivered category*, *parent nodes* nya yakni : *day category*, status *unknown summary*, status *normal delay*, status *normal*, status *maybe issue from operator*, dan status *not normal*
3. Variabel *maybe delay from 3rd party* mempengaruhi beberapa *parent nodes* seperti status *not delay*, status *unknown summary*, status *normal delay*, status *long delay*, status *normal*, status *average delay*, status *maybe issue from operator*, dan status *not normal*.
4. Transaksi pada saat pagi / malam hari ditentukan oleh *undelivered category*.
5. Status pengiriman sangat normal hanya dipengaruhi oleh variabel *delivered category*.
6. Status pengiriman normal hanya dipengaruhi oleh variabel *delivered category*, *undelivered category*, kategori rata-rata notifikasi per menit, status sangat normal, dan status *maybe delay from 3rd party*.
7. Status *unknown summary* dipengaruhi oleh variabel *undelivered category* dan status sangat normal.

4.3.1 Conditional Probability Distributions (CPDs)

Conditional Probability Distributions merupakan analisis untuk membantu mengetahui nilai probabilitas dari masing-masing variabel yang memiliki kausalitas antar satu sama yang lain atau probabilitas kejadian dari 1 variabel.

Tabel 4.3 Nilai CPDs pada variabel kategori *delivered*

Variabel	Status
average_delay	Tidak
long_delay	Tidak
maybe_delay_dr_3rd_party	Tidak
maybe_issue_from_operator	Tidak
normal	Tidak
normal_delay	Tidak
not_delay	Tidak
not_normal	Tidak
sangat_normal	Tidak

Variabel	Probabilitas
delivered_cat	
Sangat Rendah	93.48%
Rendah	3.62%
Sedang	1.45%
Tinggi	0.72%
Sangat Tinggi	0.72%

Tabel 4.4 Nilai CPDs pada variabel kategori *undelivered*

Variabel	Status							
day_cat	Malam	Malam	Malam	Malam	Pagi	Pagi	Pagi	Pagi
maybe_issue_from_operator	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Ya
normal	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
normal_delay	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
not_normal	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak
unknown_summary	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	Tidak

Variabel	Probabilitas							
undelivered_cat								
Rendah	98.93%	95.65%	0.21%	12.05%	98.62%	97.30%	0.78%	3.59%
Tinggi	1.07%	4.35%	99.79%	87.95%	1.38%	2.70%	99.22%	96.41%

Tabel 4.5 Nilai CPDs pada variabel kategori *Average Notification per Minutes*

Variabel	Status				
maybe_delay_dr_3rd_party	Tidak	Tidak	Tidak	Tidak	Ya
maybe_issue_from_operator	Tidak	Tidak	Tidak	Ya	Tidak
normal	Tidak	Tidak	Ya	Tidak	Tidak
normal_delay	Tidak	Tidak	Tidak	Tidak	Tidak
not_normal	Tidak	Ya	Tidak	Tidak	Tidak

Variabel	Probabilitas				
avg_noti_per_minute_cat					
Langsung	93.48%	0.05%	0.01%	0.14%	0.00%
Normal	6.44%	0.15%	24.07%	99.17%	96.56%
Tertunda lama	0.08%	99.80%	75.92%	0.69%	3.43%

Tabel 4.6 Nilai CPDs pada variabel kategori *Rate Undelivered*

Variabel	Status					
average_delay	Tidak	Tidak	Tidak	Tidak	Ya	Ya
long_delay	Tidak	Tidak	Tidak	Ya	Tidak	Tidak
not_delay	Tidak	Tidak	Ya	Tidak	Tidak	Tidak
undelivered_cat	Rendah	Tinggi	Rendah	Rendah	Rendah	Tinggi

Variabel	Probabilitas					
rate_undelivered_cat						
Rendah	91.50%	5.18%	7.32%	9.05%	5.48%	28.57%
Tinggi	8.50%	94.82%	92.68%	90.95%	94.52%	71.43%

Tabel 4.7 Nilai CPDs pada variabel kategori *Not Normal* dan variabel kategori *Day*

Variabel	Probabilitas
not_normal	
Tidak	96.23%
Ya	3.77%

Variabel	Probabilitas
day_cat	
Malam	45.73%
Pagi	54.27%

Tabel 4.8 Nilai CPDs pada variabel kategori *Unknown Summary* dan variabel kategori *Long Delay*

Variabel	Probabilitas	Variabel	Probabilitas
unknown_summary		long_delay	
Tidak	99.76%	Tidak	98.77%
Ya	0.24%	Ya	1.23%

Tabel 4.9 Nilai CPDs pada variabel kategori *Normal Delay* dan variabel kategori *Normal*

Variabel	Probabilitas	Variabel	Probabilitas
normal_delay		normal	
Tidak	99.75%	Tidak	74.15%
Ya	0.25%	Ya	25.85%

Tabel 4.10 Nilai CPDs pada variabel kategori *Average Delay* dan *Maybe Issue From Operator*

Variabel	Probabilitas	Variabel	Probabilitas
average_delay		maybe_issue_from_operator	
Tidak	99.85%	Tidak	98.61%
Ya	0.15%	Ya	1.39%

Tabel 4.11 Nilai CPDs pada variabel status *Maybe Delay From 3rd Party*

Variabel	Status						
<i>average_delay</i>	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya
<i>long_delay</i>	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak
<i>maybe_issue_from_operator</i>	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak
<i>Normal</i>	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak
<i>normal_delay</i>	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
<i>not_delay</i>	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
<i>not_normal</i>	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
Variabel	Probabilitas						
<i>maybe_delay_dr_3rd_party</i>							
Tidak	99.95%	99.97%	99.25%	99.99%	99.86%	99.84%	98.72%
Ya	0.05%	0.03%	0.75%	0.01%	0.14%	0.16%	1.28%

Diketahui dari Table 4.3 hingga Table 4.11 bahwa:

1. Ketika variabel *average delay*, *long delay*, *maybe delay from 3rd party*, *maybe issue from operator*, *normal*, *normal delay*, *not delay*, *not normal*, sangat normal didapatkan probabilitas tertinggi kategori *delivered* yakni sangat rendah dengan nilai sebesar 93.48%. Hasil klasifikasi probabilitas terhadap Table 4.3 dengan kategori *delivered* ditunjukkan pada Tabel 4.12.

Tabel 4.12 Hasil Klasifikasi Kategori *Delivered*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Pesan yang terkirim akan terklasifikasi kedalam kategori Sangat Rendah dan jumlah pesan terkirim berada di rentang nilai 0-5.
<i>long_delay</i>	Tidak	
<i>maybe_delay_dr_3rd_party</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>sangat_normal</i>	Tidak	

2. Kausalitas yang terjadi pada variabel kategori *undelivered* sebagai berikut:
 - a. Ketika variabel kategori *day* adalah malam, kemudian *maybe issue from operator*, *normal*, *normal delay*, *not normal*, dan *unknown summary* dengan status tidak, maka didapatkan kategori *undelivered* rendah sebesar 98.93% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered 1* ditunjukkan pada Tabel 4.13.

Tabel 4.13 Hasil Klasifikasi Kategori *Undelivered 1*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Malam	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Rendah dan jumlah pesan yang tidak terkirim berada di rentang nilai 0-1.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Tidak	

- b. Ketika variabel kategori *day* adalah malam, kemudian *maybe issue from operator*, normal, *normal delay*, dan *not normal* dengan status tidak, tetapi variabel *unknown summary* dengan status ya, maka didapatkan kategori *undelivered* rendah sebesar 95.65% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered 2* ditunjukkan pada Tabel 4.14.

Tabel 4.14 Hasil Klasifikasi Kategori *Undelivered 2*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Malam	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Rendah dan jumlah pesan yang tidak terkirim berada di rentang nilai 0-1.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Ya	

- c. Ketika variabel kategori *day* adalah malam, kemudian *maybe issue from operator*, normal, *normal delay*, dan *unknown summary* dengan status tidak, tetapi variabel *not normal* dengan status ya, maka didapatkan kategori *undelivered* tinggi sebesar 99.79% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered 3* ditunjukkan pada Tabel 4.15.

Tabel 4.15 Hasil Klasifikasi Kategori *Undelivered 3*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Malam	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Tinggi dan jumlah pesan yang tidak terkirim berada di rentang nilai 2-7.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Ya	
<i>unknown_summary</i>	Tidak	

- d. Ketika variabel kategori *day* adalah malam, kemudian normal, *normal delay*, *not normal*, dan *unknown summary* dengan status tidak, tetapi variabel *maybe issue from operator* dengan status ya, maka didapatkan kategori *undelivered*

tinggi sebesar 87.95% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered* 4 ditunjukkan pada Tabel 4.16.

Tabel 4.16 Hasil Klasifikasi Kategori *Undelivered* 4

Variabel Status	Status	Simpulan
<i>day_cat</i>	Malam	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Tinggi dan jumlah pesan yang tidak terkirim berada di rentang nilai 2-7.
<i>maybe_issue_from_operator</i>	Ya	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Tidak	

- e. Ketika variabel kategori *maybe issue from operator*, *normal*, *normal delay*, *not normal*, dan *unknown summary* dengan status tidak, tetapi variabel *day* dengan status pagi, maka didapatkan kategori *undelivered* rendah sebesar 98.62% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered* 5 ditunjukkan pada Tabel 4.17.

Tabel 4.17 Hasil Klasifikasi Kategori *Undelivered* 5

Variabel Status	Status	Simpulan
<i>day_cat</i>	Pagi	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Rendah dan jumlah pesan yang tidak terkirim berada di rentang nilai 0-1.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Tidak	

- f. Ketika variabel *day* adalah pagi, kemudian *maybe issue from operator*, *normal*, *normal delay*, *not normal* dengan status tidak, kemudian *unknown summary* dengan status ya, maka didapatkan kategori *undelivered* rendah sebesar 97.30% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered* 6 ditunjukkan pada Tabel 4.18.

Tabel 4.18 Hasil Klasifikasi Kategori *Undelivered 6*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Pagi	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Rendah dan jumlah pesan yang tidak terkirim berada di rentang nilai 0-1.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Ya	

- g. Ketika variabel *day* adalah pagi, kemudian *maybe issue from operator*, *normal*, *normal delay*, *unknown summary* dengan status tidak, kemudian *not normal* dengan status ya, maka didapatkan kategori *undelivered* tinggi sebesar 99.22% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered 7* ditunjukkan pada Tabel 4.19.

Tabel 4.19 Hasil Klasifikasi Kategori *Undelivered 7*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Pagi	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Tinggi dan jumlah pesan yang tidak terkirim berada di rentang nilai 2-7.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Ya	
<i>unknown_summary</i>	Tidak	

- h. Ketika variabel *day* adalah pagi, kemudian *not normal*, *normal*, *normal delay*, *unknown summary* dengan status tidak, kemudian *maybe issue form operator* dengan status ya, maka didapatkan kategori *undelivered* tinggi sebesar 96.41% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.4 dengan kategori *undelivered 8* ditunjukkan pada Tabel 4.20.

Tabel 4.20 Hasil Klasifikasi Kategori *Undelivered 8*

Variabel Status	Status	Simpulan
<i>day_cat</i>	Pagi	

Variabel Status	Status	Simpulan
<i>maybe_issue_from_operator</i>	Ya	Pesan yang tidak terkirim akan terklasifikasi kedalam kategori Tinggi dan jumlah pesan yang tidak terkirim berada di rentang nilai 2-7.
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	
<i>unknown_summary</i>	Tidak	

3. Kausalitas yang terjadi pada variabel kategori *average notification per minute* sebagai berikut:

- a. Ketika variabel kategori *maybe delay 3rd party*, *maybe issue from operator*, *normal*, *normal delay*, *not normal* dengan status tidak, maka didapatkan kategori langsung terkirim pada variabel *average notification per minute* sebesar 93.48% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.5 dengan kategori *average notification per minutes 1* ditunjukkan pada Tabel 4.21.

Tabel 4.21 Hasil Klasifikasi Kategori *Average Notification Per Minutes 1*

Variabel Status	Status	Simpulan
<i>maybe_delay_dr_3rd_party</i>	Tidak	Rata-rata notifikasi pesan per menit akan terklasifikasi kedalam kategori Langsung dan durasi pengiriman (dalam menit) berada di rentang nilai 0-0.01.
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- b. Ketika variabel kategori *maybe delay 3rd party*, *maybe issue from operator*, *normal*, *normal delay* dengan status tidak, kemudian *not normal* dengan status ya, maka didapatkan kategori tertunda lama (*long delay*) pengiriman pada variabel *average notification per minute* sebesar 99.80% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.5 dengan kategori *average notification per minutes 2* ditunjukkan pada Tabel 4.22.

Tabel 4.22 Hasil Klasifikasi Kategori *Average Notification Per Minutes 2*

Variabel Status	Status	Simpulan
<i>maybe_delay_dr_3rd_party</i>	Tidak	Rata-rata notifikasi pesan per menit akan terklasifikasi kedalam kategori
<i>maybe_issue_from_operator</i>	Tidak	

Variabel Status	Status	Simpulan
<i>normal</i>	Tidak	Tertunda lama dan durasi pengiriman
<i>normal_delay</i>	Tidak	(dalam menit) berada di rentang nilai
<i>not_normal</i>	Ya	2.43-96.

- c. Ketika variabel kategori *maybe delay 3rd party*, *maybe issue from operator*, *normal delay*, *not normal* dengan status tidak, kemudian *normal* dengan status ya, maka didapatkan kategori tertunda lama (*long delay*) pengiriman pada variabel *average notification per minute* sebesar 75.92% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.5 dengan kategori *average notification per minutes 3* ditunjukkan pada Tabel 4.23.

Tabel 4.23 Hasil Klasifikasi Kategori *Average Notification Per Minutes 3*

Variabel Status	Status	Simpulan
<i>maybe_delay_dr_3rd_party</i>	Tidak	Rata-rata notifikasi pesan per menit
<i>maybe_issue_from_operator</i>	Tidak	akan terklasifikasi kedalam kategori
<i>normal</i>	Ya	Tertunda lama dan durasi pengiriman
<i>normal_delay</i>	Tidak	(dalam menit) berada di rentang nilai
<i>not_normal</i>	Tidak	2.43-96.

- d. Ketika variabel kategori *maybe delay 3rd party*, *normal delay*, *not normal*, *normal* dengan status tidak, kemudian *maybe issue from operator* dengan status ya, maka didapatkan kategori normal waktu pengiriman pada variabel *average notification per minute* sebesar 99.17% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.5 dengan kategori *average notification per minutes 4* ditunjukkan pada Tabel 4.24.

Tabel 4.24 Hasil Klasifikasi Kategori *Average Notification Per Minutes 4*

Variabel Status	Status	Simpulan
<i>maybe_delay_dr_3rd_party</i>	Tidak	Rata-rata notifikasi pesan per menit akan terklasifikasi kedalam kategori
<i>maybe_issue_from_operator</i>	Ya	
<i>normal</i>	Tidak	Tertunda lama dan durasi pengiriman (dalam menit) berada di rentang nilai 0.02-2.42.
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- e. Ketika variabel kategori *maybe issue from operator*, *normal delay*, *not normal*, normal dengan status tidak, kemudian *maybe delay 3rd party* dengan status ya, maka didapatkan kategori normal waktu pengiriman pada variabel *average notification per minute* sebesar 96.56% kemungkinan yang akan terjadi. Hasil klasifikasi probabilitas terhadap Table 4.5 dengan kategori *average notification per minutes 5* ditunjukkan pada Tabel 4.25.

Tabel 4.25 Hasil Klasifikasi Kategori *Average Notification Per Minutes 5*

Variabel Status	Status	Simpulan
<i>maybe_delay_dr_3rd_party</i>	Ya	Rata-rata notifikasi pesan per menit akan terklasifikasi kedalam kategori
<i>maybe_issue_from_operator</i>	Tidak	
<i>normal</i>	Tidak	Tertunda lama dan durasi pengiriman (dalam menit) berada di rentang nilai 0.02-2.42.
<i>normal_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

4. Kausalitas yang terjadi pada variabel kategori *rate undelivered* sebagai berikut:
- a. Ketika variabel kategori *average delay*, *long delay*, *not delay* dengan status tidak, kemudian *undelivered* rendah, maka didapatkan kategori sangat rendah pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 91.50%. Hasil klasifikasi probabilitas terhadap Table 4.6 dengan kategori *rate undelivered 1* ditunjukkan pada Tabel 4.26.

Tabel 4.26 Hasil Klasifikasi Kategori *Rate Undelivered 1*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Persentase pesan tidak terkirim akan terklasifikasi kedalam kategori Sangat Rendah dan persentase pesan yang
<i>long_delay</i>	Tidak	
<i>not_delay</i>	Tidak	

Variabel Status	Status	Simpulan
<i>undelivered_cat</i>	Rendah	tidak terkirim berada di rentang nilai 0-5.

- b. Ketika variabel kategori *average delay*, *long delay*, *not delay* dengan status tidak, kemudian *undelivered* tinggi, maka didapatkan kategori sangat tinggi pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 94.82%. Hasil klasifikasi probabilitas terhadap Table 4.6 dengan kategori *rate undelivered 2* ditunjukkan pada Tabel 4.27.

Tabel 4.27 Hasil Klasifikasi Kategori *Rate Undelivered 2*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Persentase pesan tidak terkirim akan
<i>long_delay</i>	Tidak	terklasifikasi kedalam kategori Sangat
<i>not_delay</i>	Tidak	Tinggi dan persentase pesan yang tidak
<i>undelivered_cat</i>	Tinggi	terkirim berada di rentang nilai 6-100.

- c. Ketika variabel kategori *average delay*, *long delay* dengan status tidak, kemudian *undelivered* rendah, dan *not delay* dengan status ya, maka didapatkan kategori sangat tinggi pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 92.68%. Hasil klasifikasi probabilitas terhadap Table 4.6 dengan kategori *rate undelivered 3* ditunjukkan pada Tabel 4.28.

Tabel 4.28 Hasil Klasifikasi Kategori *Rate Undelivered 3*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Persentase pesan tidak terkirim akan
<i>long_delay</i>	Tidak	terklasifikasi kedalam kategori Sangat
<i>not_delay</i>	Ya	Tinggi dan persentase pesan yang tidak
<i>undelivered_cat</i>	Rendah	terkirim berada di rentang nilai 6-100.

- d. Ketika variabel kategori *average delay*, *not delay* dengan status tidak, kemudian *undelivered* rendah, dan *long delay* dengan status ya, maka didapatkan kategori sangat tinggi pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 90.95%. Hasil klasifikasi probabilitas

terhadap Table 4.6 dengan kategori *rate undelivered* 4 ditunjukkan pada Tabel 4.29.

Tabel 4.29 Hasil Klasifikasi Kategori *Rate Undelivered* 4

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Persentase pesan tidak terkirim akan
<i>long_delay</i>	Ya	terklasifikasi kedalam kategori Sangat
<i>not_delay</i>	Tidak	Tinggi dan persentase pesan yang tidak
<i>undelivered_cat</i>	Rendah	terkirim berada di rentang nilai 6-100.

- e. Ketika variabel kategori *long delay*, *not delay* dengan status tidak, kemudian *undelivered* rendah, dan *average* dengan status ya, maka didapatkan kategori sangat tinggi pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 94.52%. Hasil klasifikasi probabilitas terhadap Table 4.6 dengan kategori *rate undelivered* 5 ditunjukkan pada Tabel 4.30.

Tabel 4.30 Hasil Klasifikasi Kategori *Rate Undelivered* 5

Variabel Status	Status	Simpulan
<i>average_delay</i>	Ya	Persentase pesan tidak terkirim akan
<i>long_delay</i>	Tidak	terklasifikasi kedalam kategori Sangat
<i>not_delay</i>	Tidak	Tinggi dan persentase pesan yang tidak
<i>undelivered_cat</i>	Rendah	terkirim berada di rentang nilai 6-100.

- f. Ketika variabel kategori *long delay*, *not delay* dengan status tidak, kemudian *undelivered* rendah, dan *average* dengan status ya, maka didapatkan kategori sangat tinggi pada variabel *rate undelivered* dengan kemungkinan yang akan terjadi sebesar 71.43%. Hasil klasifikasi probabilitas terhadap Table 4.6 dengan kategori *rate undelivered* 6 ditunjukkan pada Tabel 4.31.

Tabel 4.31 Hasil Klasifikasi Kategori *Rate Undelivered* 6

Variabel Status	Status	Simpulan
<i>average_delay</i>	Ya	Persentase pesan tidak terkirim akan
<i>long_delay</i>	Tidak	terklasifikasi kedalam kategori Sangat
<i>not_delay</i>	Tidak	Tinggi dan persentase pesan yang tidak
<i>undelivered_cat</i>	Tinggi	terkirim berada di rentang nilai 6-100.

5. Pengiriman pesan OTP dengan kemungkinan dikategori tidak di variabel *not normal* kemungkinan yang akan terjadi sebesar 96.23%.
6. Pengiriman pesan OTP dengan kemungkinan dikategori transaksi di pagi divariabel *day* kemungkinan yang akan terjadi 54.27% lebih besar 8.54% dari transaksi di malam hari.
7. Pengiriman pesan OTP dengan kemugkinan dikategori tidak di variabel *unknown summary* kemugkinan yang akan terjadi sebesar 99.76%.
8. Pengiriman pesan OTP dengan kemungkinan dikategori tidak di variabel normal kemungkinan yang akan terjadi sebesar 74.15%.
9. Pengiriman pesan OTP dengan kemungkinan dikategori tidak di variabel *average delay* yang akan terjadi sebesar 99.85%.
10. Pengiriman pesan OTP dengan kemungkinan dikategori tidak di variabel *maybe issue from operator* yang akan terjadi sebesar 98.61%.
11. Kausalitas yang terjadi pada variabel kategori *maybe delay from 3rd party* sebagai berikut:
 - a. Ketika variabel *average delay, long delay, maybe issue from operator, normal, normal delay, not delay* dengan status tidak, kemudian *not normal* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.95%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 1 ditunjukkan pada Tabel 4.32.

Tabel 4.32 Hasil Klasifikasi Status *Maybe Issue From 3rd Party* 1

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Ya	

- b. Ketika variabel *average delay, long delay, maybe issue from operator, normal, normal delay, not normal* dengan status tidak, kemudian *not delay*

dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.97%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 2 ditunjukkan pada Tabel 4.33.

Tabel 4.33 Hasil Klasifikasi Status *Maybe Issue From 3rd Party* 2

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Ya	
<i>not_normal</i>	Tidak	

- c. Ketika variabel *average delay*, *long delay*, *maybe issue from operator*, *normal*, *not delay*, *not normal* dengan status tidak, kemudian *normal delay* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.25%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 3 ditunjukkan pada Tabel 4.34.

Tabel 4.34 Hasil Klasifikasi Status *Maybe Issue From 3rd Party* 3

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Ya	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- d. Ketika variabel *average delay*, *long delay*, *maybe issue from operator*, *normal delay*, *not delay*, *not normal* dengan status tidak, kemudian *normal* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.99%. Hasil klasifikasi

probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 4 ditunjukkan pada Tabel 4.35.

Tabel 4.35 Hasil Klasifikasi Status *Maybe Issue From 3rd Party* 4

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Ya	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- e. Ketika variabel *average delay*, *normal*, *maybe issue from operator*, *normal delay*, *not delay*, *not normal* dengan status tidak, kemudian *long delay* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.86%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 5 ditunjukkan pada Tabel 4.36.

Tabel 4.36 Hasil Klasifikasi Status *Maybe Issue From 3rd Party* 5

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Ya	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- f. Ketika variabel *average delay*, *normal*, *maybe issue from operator*, *normal delay*, *not delay*, *not normal* dengan status tidak, kemudian *long delay* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 99.84%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party* 6 ditunjukkan pada Tabel 4.37.

Tabel 4.37 Hasil Klasifikasi Status *Maybe Issue From 3rd Party 6*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Tidak	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Ya	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

- g. Ketika variabel *long delay*, *normal*, *maybe issue from operator*, *normal delay*, *not delay*, *not normal* dengan status tidak, kemudian *average delay* dengan status ya, maka didapatkan nilai kemungkinan dikategori tidak pada variabel *maybe delay from 3rd party* sebesar 98.72%. Hasil klasifikasi probabilitas terhadap Table 4.11 dengan kategori *maybe delay from 3rd party 7* ditunjukkan pada Tabel 4.38.

Tabel 4.38 Hasil Klasifikasi Status *Maybe Issue From 3rd Party 7*

Variabel Status	Status	Simpulan
<i>average_delay</i>	Ya	Kemungkinan <i>issue</i> dari pihak ke-3 pada pengiriman pesan SMS OTP akan terklasifikasi kedalam status Tidak.
<i>long_delay</i>	Tidak	
<i>maybe_issue_from_operator</i>	Tidak	
<i>Normal</i>	Tidak	
<i>normal_delay</i>	Tidak	
<i>not_delay</i>	Tidak	
<i>not_normal</i>	Tidak	

4.3.2 Pemilihan Variabel dan Status Signifikan

Berdasarkan (Chickering, 1996; Heckerman et al., 1995), *parent nodes* dengan setidaknya 3 *child nodes* dianggap berpengaruh dengan antar hubungan variabel sehingga dapat menghindari variabel dan fitur yang berpotensi redundan, sehingga dengan memilih setidaknya 3 *child nodes* dalam 1 *parent nodes* kecenderungan tidak memberikan informasi yang redundan dan dapat mengelola kompleksitas dari variabel dan fitur yang dimiliki pada data.

Sehingga berdasarkan dari Tabel 4.1 dan Tabel 4.2 hasil pemilihan variabel dan status sebagai berikut :

Tabel 4.39 Pemilihan Variabel dan Status Signfikan

Seluruh Variabel dan Status	Hasil Seleksi Variabel dan Status
Kategori <i>Day</i>	Kategori <i>Day</i>
Kategori <i>Delivered</i>	Kategori <i>Delivered</i>
Kategori <i>Rate Undelivered</i>	Kategori <i>Rate Undelivered</i>
Kategori <i>Average Notification per Minutes</i>	Kategori <i>Average Notification per Minutes</i>
Kategori <i>Not Delay</i>	Kategori <i>Not Delay</i>
Kategori Sangat Normal	Kategori Sangat Normal
Kategori <i>Unknown Summary</i>	Kategori <i>Unknown Summary</i>
Kategori <i>Normal Delay</i>	Kategori <i>Normal Delay</i>
Kategori <i>Long Delay</i>	Kategori <i>Long Delay</i>
Kategori <i>Maybe Delay From 3rd Party</i>	Kategori <i>Maybe Delay From 3rd Party</i>
Kategori Normal	Kategori <i>Average Delay</i>
Kategori <i>Timeout to 3rd Party</i>	Kategori <i>Not Normal</i>
Kategori <i>Average Delay</i>	
Kategori <i>Maybe Issue From Operator</i>	
Kategori <i>Not Normal</i>	

4.3.3 Evaluation Model Output

Untuk mengetahui hasil dari pemodelan (perhitungan nilai probabilitas) dapat diukur dari tingkat akurasi.

Tabel 4.40 Nilai Akurasi Berdasarkan Kategori

Keterangan	Nilai Akurasi
Kategori <i>Day</i>	54.23%
Kategori <i>Delivered</i>	35.37%
Kategori <i>Undelivered</i>	89.77%
Kategori <i>Rate Undelivered</i>	92.20%
Kategori <i>Not Delay</i>	94.21%
Kategori Sangat Normal	99.98%
Kategori <i>Unknown Summary</i>	99.75%
Kategori <i>Normal Delay</i>	99.80%

Keterangan	Nilai Akurasi
Kategori <i>Long Delay</i>	98.80%
Kategori <i>Maybe Delay From 3rd party</i>	79.77%
Kategori Normal	74.06%
Kategori <i>Average Delay</i>	99.87%
Kategori <i>Maybe Issue From Operator</i>	98.61%
Kategori <i>Not Normal</i>	96.33%
Kategori <i>Timeout to 3rd Party</i>	99.82%

Berdasarkan Table 4.40 dapat diketahui bahwa kategori yang memiliki akurasi yang rendah (dibawah 80%) yakni kategori *day*, kategori *delivered*, kategori *maybe delay from 3rd party*, kategori normal. Dari tingkat akurasi yang rendah ini dapat mengartikan bahwa nilai probabilitas yang dihasilkan belum menangkap pola data yang kemungkinan besar akibat dari data yang kurang bervariasi nilainya.

4.4 *Naïve Bayes*

Pengujian pada metode ini memisahkan dataset dengan 80% data latih dan 20% data uji dengan pengujian sebanyak 2 kali menggunakan seluruh variabel penelitian dan variabel yang signifikan berdasarkan dari hasil pengujian sebelumnya menggunakan metode Bayesian Network (memiliki tingkat probabilitas kejadian diatas 80%). Selain itu pula, semakin besar porsi data latih yang digunakan maka akan memberikan wawasan yang memungkinkan relevan dalam pengambilan keputusan pada data uji (Joseph et al., 2022).

4.4.1 **Seluruh Variabel**

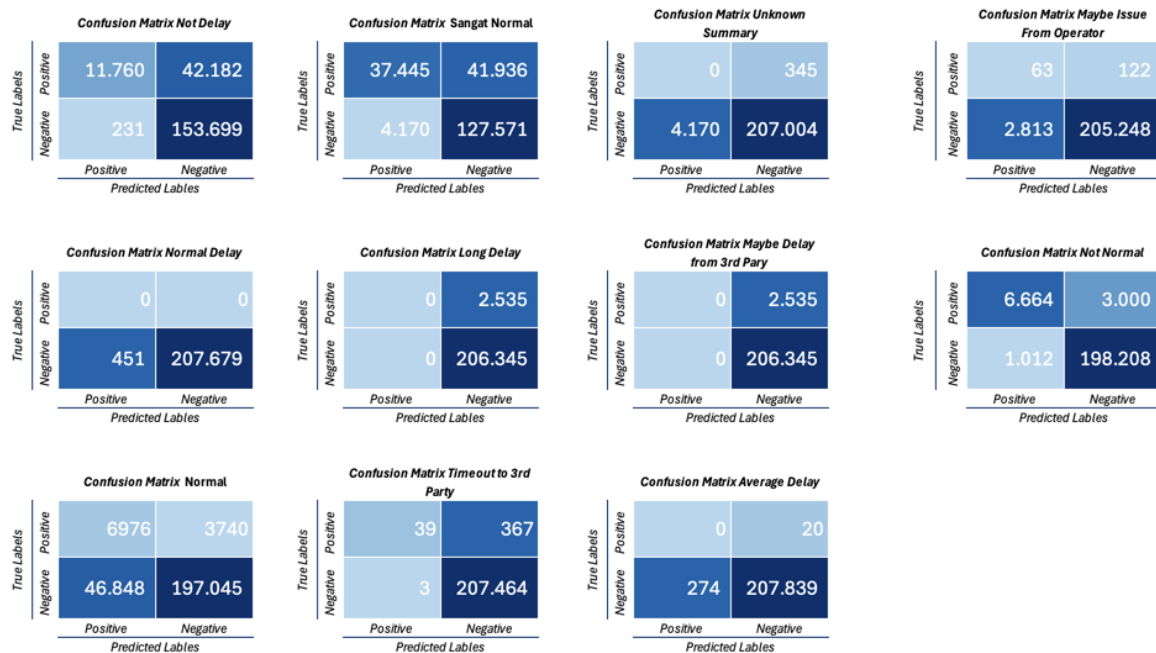
Pada pengujian dengan menggunakan seluruh variabel pada metode naive bayes didapatkan nilai akurasi sebesar 53.48% pada data latih sedangkan pada data uji 53.45%. Secara distribusi persebaran data sebagai berikut:

Tabel 4.41 Distribusi data uji pada *summary status* pada seluruh variabel

Status	Jumlah Data
<i>Maybe Delay From 3rd Party</i>	86.065
<i>Normal</i>	53.824

Status	Jumlah Data
<i>Very Normal</i>	41.615
<i>Not Delay</i>	11.991
<i>Not Normal</i>	7.676
<i>Maybe Issue From Operator</i>	2.876
<i>Long Delay</i>	2.535
<i>Unknown Summary</i>	523
<i>Normal Delay</i>	451
<i>Average Delay</i>	274
<i>Timeout to 3rd Party</i>	42

Berdasarkan table 4.41 dapat diketahui bahwa persebaran pada data uji paling banyak terjadi pada *maybe delay from 3rd party* (41.40%), normal (25,89%), sangat normal (20.02%), *not delay* (5.77%), *not normal* (3.69%), *maybe issue from operator* (1.38%), *long delay* (1.22%), selanjutnya dibawah 1% persebarannya.



Gambar 4.6 *Confusion matrix* metode *naïve bayes* pada seluruh variabel

Pada Gambar 4.6 dapat diketahui bahwa:

1. Pada status *Not Delay* sebanyak 79.60% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 20.40%;

2. Pada status Sangat Normal sebanyak 78.16% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 21.84%;
3. Pada status *Unknwon Summary* sebanyak 97.87% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 2.13%;
4. Pada status *Normal Delay* sebanyak 99.78% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 0.22%;
5. Pada status *Long Delay* sebanyak 98.89% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.11%;
6. Pada status *Maybe Delay from 3rd party* sebanyak 82.39% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 17.61%;
7. Pada status Normal sebanyak 80.13% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 19.87%;
8. Pada status *Timeout to 3rd Party* sebanyak 99.82% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 0.18%;
9. Pada status *Average Delay* sebanyak 99.86% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 0.14%;
10. Pada status *Maybe Issue from Operator* sebanyak 98.59% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.41%;
11. Pada status *Not Normal* sebanyak 98.08% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.92%.

4.4.2 Variabel Signifikan

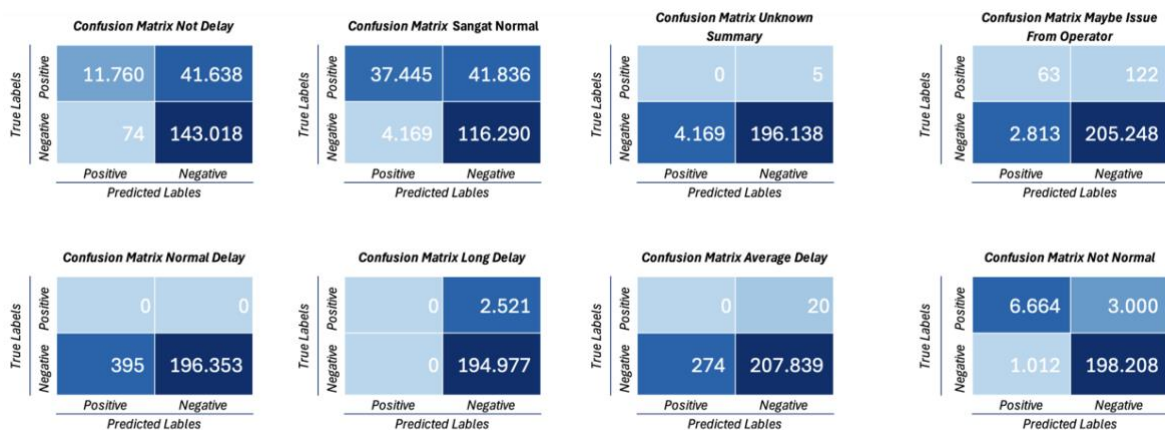
Pada pengujian dengan menggunakan sebagian variabel pada metode naïve bayes didapatkan nilai akurasi sebesar 87.73% pada data latih sedangkan pada data uji 87.74%. Secara distribusi persebaran data sebagai berikut:

Tabel 4.42 Distribusi data uji pada *summary status* pada parsial variabel

Status	Jumlah Data
<i>Very Normal</i>	31.196
<i>Not Delay</i>	9.043
<i>Not Normal</i>	5.784
<i>Maybe Issue From Operator</i>	2.154

Status	Jumlah Data
<i>Long Delay</i>	1.847
<i>Unknown Summary</i>	391
<i>Normal Delay</i>	343
<i>Average Delay</i>	220

Berdasarkan table 4.42 dapat diketahui bahwa persebaran pada data uji paling banyak terjadi pada sangat normal (61.20%), *not delay* (17,74%), *not normal* (11.35%), *maybe issue from operator* (4.23%), *long delay* (3.62%), *long delay* (3.62%), selanjutnya dibawah 1% persebarannya.



Gambar 4.7 *Confusion matrix* naïve bayes pada parsial variabel

Pada Gambar 4.7 dapat diketahui bahwa:

1. Pada status *Not Delay* sebanyak 78.77% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 21.23%;
2. Pada status Sangat Normal sebanyak 76.97% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 23.03%;
3. Pada status *Unknown Summary* sebanyak 97.92% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 2.08%;
4. Pada status *Normal Delay* sebanyak 98.80% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 0.20%;
5. Pada status *Long Delay* sebanyak 98.72% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.28%;
6. Pada status *Average Delay* sebanyak 99.87% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 0.13%;

7. Pada status *Maybe Issue from Operator* sebanyak 98.59% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.41%;
8. Pada status *Not Normal* sebanyak 98.08% yang diprediksikan sama dengan aktual statusnya, sedangkan yang salah prediksi sebesar 1.92%.

4.5 Multilayer Perceptron

Parameter pada metode ini digunakan *hidden layer* sebanyak 12 lapisan dengan jumlah iterasi sebanyak 500 dan *learning rate* sebesar 0.006%, parameter yang ditentukan setelah melakukan perulangan berkali-kali untuk mendapatkan hasil yang optimal dari sisi konsumsi waktu untuk proses serta hasil akurasi yang didapatkan. Selain itu pula, pengaturan untuk parameter didukung dengan alasan ilmiah seperti *hidden layer* yang banyak akan membantu meningkatkan kapasitas representasi model (Mitchell, 1997), kemudian untuk iterasi sebanyak 500 sebagai penyeimbang dari jumlah *hidden layer* yang ditentukan agar tidak terjadi *overfitting* (Xu et al., 2023), kemudian untuk *learning rate* yang kecil membantu pembelajaran model secara bertahap yang dapat mengurangi *overfitting* dan dapat mempertahankan generalisasi hasil (Senior et al., 2013).

4.5.1 Pengujian Pada Seluruh Variabel

Pengujian pada metode ini memisahkan dataset dengan 80% data latih dan 20% data uji dengan menggunakan variabel x nya yakni *delivered category*, *undelivered category*, *average notification per minute*, *rate undelivered category*.

Parameter yang digunakan yaitu:

- *Hidden activation* : Sigmoid
- *Output activation* : Softmax
- Jumlah *hidden layer*: 12;
- Jumlah iterasi: 500;
- *Learning rate*: 0.00006

Didapatkan waktu pengujian selama 1 menit 4.7 detik dan akurasi sebesar 87.91% ~ 88%. Parameter yang digunakan merupakan parameter yang pada saat ini sudah mencapai optimalnya setelah melakukan *trial* parameter sebanyak 58 kali.

Dengan nilai akurasi 87.91%, status yang dapat ditangkap dengan baik hanya status sangat normal, normal, *maybe delay from 3rd party*, *not delay*, mungkin *issue* dari operator, dan tidak normal. Sedangkan status *long delay*, *normal delay*, *average delay*, *unknown summary*, *timeout to 3rd party* belum dapat ditangkap pada data yang digunakan.

Pada status-status yang sudah dapat menangkap pada kategori data, diketahui dari 207.871 data terdapat kesalahan prediksi kategori sebesar 12.09% (25.131)

4.5.2 Pengujian Pada Variabel Signifikan

Pengujian pada metode ini memisahkan dataset dengan 80% data latih dan 20% data uji dengan menggunakan variabel x nya yakni *undelivered category*, *rate undelivered category*.

Parameter yang digunakan yaitu:

- *Hidden activation* : Sigmoid
- *Output activation* : Softmax
- Jumlah *hidden layer*: 12
- Jumlah iterasi: 500
- *Learning rate*: 0.00006

Didapatkan waktu pengujian selama 24.9 detik dan akurasi sebesar 88.12%. Parameter yang digunakan merupakan parameter yang pada saat ini sudah mencapai optimalnya berdasarkan dari pelatihan sebelumnya.

Dengan nilai akurasi 88.12%, status yang dapat ditangkap dengan baik hanya status sangat normal, normal, *maybe delay from 3rd party*, *not delay*, mungkin *issue* dari operator, dan tidak normal. Sedangkan status *long delay*, *normal delay*, *average delay*, *unknown summary*, *timeout to 3rd party* belum dapat ditangkap pada data yang digunakan.

Pada status-status yang sudah dapat menangkap pada kategori data, diketahui dari 67.969 data terdapat kesalahan prediksi kategori sebesar 11.87% (8070).

4.6 Performa Metode *Naïve Bayes* dan *Multilayer Perceptron*

Setelah melakukan pengujian pada metode *Naïve Bayes* dan *Multilayer Perceptron*, maka diperlukan melihat perbandingan performa secara metode.

Tabel 4.43 Performa partisi ukuran data latih

Metode	Penggunaan Data	Ukuran Data Uji	Akurasi	Waktu (detik)
Naïve Bayes	Overall	20%	52,82%	0,053
		40%	53,02%	0,037
		60%	52,78%	0,029
		80%	53,48%	0,013
		90%	54,18%	0,007
	Partial	20%	28,44%	0,004
		40%	28,47%	0,003
		60%	87,69%	0,014
		80%	87,74%	0,006
		90%	87,71%	0,009
MLP	Overall	20%	5,86%	154,96
		40%	25,99%	121,9
		60%	25,93%	84,02
		80%	87,91%	49,88
		90%	87,95%	27,22
	Partial	20%	88,52%	57,900
		40%	88,06%	44,740
		60%	88,10%	29,730
		80%	88,12%	15,820
		90%	87,73%	8,110

Berdasarkan Table 4.43 dapat diketahui bahwa apabila ukuran data latih berada dibawah 80% akurasi yang didapatkan sangat rendah dibandingkan ketika data latih yang digunakan sebesar 80%. Hal ini sejalan dengan data yang digunakan pada penelitian yang digunakan merupakan data aktual transaksi yang memungkinkan terdapat anomali pada waktu tertentu.

Semakin banyak *child nodes* atau variabel turunannya maka semakin kompleks hubungan kausalnya, dimana semakin kompleks hubungan kausalnya maka akan

mempermudah identifikasi akar masalah (Spirtes & Zhang, 2019). Hal ini tentunya dapat membantu pemilik bisnis dalam mengevaluasi pengiriman pesan OTP.

Perbedaan waktu komputasi secara signifikan antara *naive bayes* dan *multilayer perceptron* karena mekanisme dasar dari masing-masing model yang mana, kesederhanaan mekanisme *naive bayes* melakukan perhitungan/perkiraan probabilitas dengan cepat tanpa memerlukan peoptimalan yang iteratif (berulang). Sedangkan kompleksitas dari metode MLP memerlukan beberapa kali iterasi dan perhitungan gradien, yang meningkatkan beban waktu komputasi (Raschka, 2018; Zhang, 2004). Sehingga pada performa metode *machine learning* dan *deep learning* ini merepresentasikan selisih akurasi dan waktu komputasi sebagai referensi peneliti selanjutnya dan bagi *stakeholder* untuk mempertimbangkan dalam penggunaan metode yang akan dipakai berdasarkan *training set* dan *test set* dengan studi kasus klasifikasi yang tipe data numerik dan kategorik. Pada penelitian ini memilih metode *naive bayes* secara hasil akhir, karena akurasi yang dihasilkan tidak begitu jauh dengan metode *multilayer perceptron*.

Berdasarkan penelitian yang telah disebutkan, maka penelitian ini memberikan wawasan yang lebih komprehensif mengenai berbagai klasifikasi notifikasi pengiriman pesan SMS OTP, membantu peneliti dan praktisi memahami pola (fenomena) dan karakteristik transaksi pengiriman pesan SMS OTP dalam berbagai konteks, serta mengetahui hubungan antar variabel dan status transaksi pengiriman pesan SMS OTP yang lebih statis. Dengan demikian, memberikan kontribusi penting dalam pengembangan metode dan alat dalam sains data, khususnya dalam analisis dan pengelolaan data transaksi pengiriman pesan SMS OTP.

Keterbatasan dari penelitian ini belum adanya studi literatur yang membahas tentang klasifikasi status pengiriman transaksi pesan SMS OTP sehingga tidak bisa membandingkan relevansi variabel dan status yang terpilih dari hasil metode bayesian network yang mengartikan hasil variabel dan status yang terpilih sudah baik/optimal. kemudian, data yang digunakan pada penelitian ini hanya berasal dari 1 perusahaan penyedia layanan jasa pengiriman pesan OTP, serta dengan jumlah data sebanyak 259.839 dengan pembagian data latih sebanyak 80% dan data uji 20% didapatkan akurasi sebesar 87,74% pada metode naive bayes dengan variabel dan status terpilih dari bayesian network. Sedangkan pada multilayer perceptron didapatkan akurasi sebesar 88,12% dengan variabel dan status sama seperti

metode naive bayes. Menurut peneliti, jika diimplementasikan kedalam bisnis maka akan lebih baik didapatkan akurasi prediksi berada di rentang 95-99% dengan toleransi kesalahan setidaknya 5%. Sehingga untuk meningkatkan akurasi, perlunya penambahan data yang digunakan.

BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

1. Berdasarkan dari data yang ada, diketahui bahwa transaksi yang paling banyak dilakukan pada saat pagi. Dengan status yakni status *Not Delay*, *Sangat Normal* dan *Unknown Summary* yang mengartikan bahwa banyak transaksi pada perusahaan penyedia jasa pesan OTP berjalan dengan lancar, kecuali terdapat status yang masih belum diketahui apakah pesan terkirim atau gagal.
2. Pada metode Bayesian Network diketahui bahwa variabel yang signifikan pada model yakni *delivered category*, *undelivered category*, *average notification per minutes*, dan *rate undelivered category*. Kemudian pada status yang signifikan (setidaknya 1 *parent nodes* memiliki pengaruh ke 3 *child nodes*) yakni *not delay*, *unknown summary*, *long delay*, *normal*, *maybe issue from operator*, dan *not normal*.
3. Faktor-faktor yang menjadi kausalitas kegagalan pengiriman pesan SMS (*Undelivered*) yakni:
 - a. Ketika variabel kategori day adalah malam, kemudian *maybe issue from operator*, *normal*, *normal delay*, dan *unknown summary* dengan status tidak, tetapi variabel *not normal* dengan status ya
 - b. Ketika variabel kategori day adalah malam, kemudian *normal*, *normal delay*, *not normal*, dan *unknown summary* dengan status tidak, tetapi variabel *maybe issue from operator* dengan status ya
 - c. Ketika variabel day adalah pagi, kemudian *maybe issue from operator*, *normal*, *normal delay*, *unknown summary* dengan status tidak, kemudian *not normal* dengan status ya
 - d. Ketika variabel day adalah pagi, kemudian *not normal*, *normal*, *normal delay*, *unknown summary* dengan status tidak, kemudian *maybe issue form operator* dengan status ya

4. Berikut hasil performa prediksi klasifikasi data:

- a. Dengan menggunakan metode *naïve bayes* didapatkan dengan menggunakan data latih pada seluruh variabel dengan porsi 80% memiliki akurasi sebesar 53.48%.
- b. Dengan menggunakan metode *naïve bayes* didapatkan menggunakan data latih pada variabel yang signifikan (dari metode Bayesian Network) dengan porsi 80% memiliki akurasi sebesar 87.73% (meningkat 34.25% dibandingkan dengan menggunakan seluruh variabel)
- c. Dengan menggunakan metode *Multi Layer Perceptron* didapatkan dengan menggunakan data latih pada seluruh variabel dengan porsi 80% memiliki akurasi sebesar 87.91%
- d. Dengan menggunakan metode *Multi Layer Perceptron* didapatkan dengan menggunakan data latih pada variabel yang signifikan (dari metode Bayesian Network) dengan porsi 80% memiliki akurasi sebesar 88.12% (meningkat 0.21% dibandingkan dengan menggunakan seluruh variabel).

Selain dari performa prediksi klasifikasi data, diketahui pula kategori yang tersusun sebagai berikut:

Tabel 5.1 Kategorisasi Berdasarkan Hasil *Machine Learning* (*Naïve Bayes*)

Kategori Variabel	Kategori Delivered	Kategori Undelivered	Kategori Average Notification Per Minutes	Kategori Rate Undelivered
Notifikasi Normal	<i>Average</i>	<i>High</i>	<i>Long Delay</i>	<i>Very High</i>
Notifikasi Maybe Delay From 3rd Party	<i>Above Average</i>	<i>Low</i>	Normal	<i>Very Low</i>
Notifikasi Sangat Normal	<i>Below Average</i>	<i>Low</i>	<i>Direct</i>	<i>Very Low</i>
Notifikasi <i>Not Delay</i>	<i>Average</i>	<i>Low</i>	<i>Direct</i>	<i>Very High</i>
Notifikasi <i>Not Normal</i>	<i>Below Average</i>	<i>High</i>	<i>Long Delay</i>	<i>Very High</i>
Notifikasi <i>Maybe Issue From Operator</i>	<i>Low</i>	<i>High</i>	Normal	<i>Very High</i>
Notifikasi <i>Long Delay</i>	<i>Average</i>	<i>Low</i>	<i>Direct</i>	<i>Very High</i>
Notifikasi <i>Normal Delay</i>	<i>Average</i>	<i>Low</i>	<i>Direct</i>	<i>Very High</i>
Notifikasi <i>Average Delay</i>	<i>Above Average</i>	<i>Low</i>	<i>Direct</i>	<i>Very High</i>
Notifikasi <i>Unknown Summary</i>	<i>Low</i>	<i>Low</i>	<i>Direct</i>	<i>Very Low</i>
Notifikasi <i>Timeout to 3rd Party</i>	<i>Low</i>	<i>Low</i>	<i>Direct</i>	<i>Very Low</i>

Untuk mengetahui nilai rentang berdasarkan kategorisasi data merujuk ke Tabel 4.1

5.2 Saran

Model yang didapatkan pada penelitian ini masih perlu diuji lebih lanjut bersama para praktisi dibidang pengiriman jasa pesan SMS OTP agar model yang didapat melekat pada proses bisnis pengiriman pesan SMS OTP dengan baik. Sehingga hasil dari penelitian ini dapat bermanfaat untuk membantu dalam memonitoring performa jasa yang sediakan yang memiliki dampak keberlanjutan perusahaan penyedia jasa layanan pengiriman pesan SMS OTP.

Untuk peneliti selanjutnya, klasifikasi status pengiriman pesan SMS OTP pada penelitian ini dapat digunakan atau disesuaikan dengan pada penelitian selanjutnya agar klasifikasi status ini dapat statis atau menjadi standar untuk penelitian selanjutnya. Kemudian, untuk meningkatkan performa akurasi dan prediksi dapat dilakukan untuk penambahan dataset yang lebih banyak dan beragam datanya, tetapi lebih diprioritaskan untuk penambahan data dari sisi ragam datanya.

Selain itu, dapat membandingkan metode bayesian network dengan menggunakan metode lainnya agar dapat mengetahui dan mengukur kebaikan metode *Bayesian Network* dan *Naïve Bayes*, metode yang dapat digunakan seperti *Association Rule*, *Decision Tree*, *Random Forest*, *Markov Random Fields* (MRF), *Hidden Markov Models* (HMM), *Conditional Random Fields* (CRF), *Dynamic Bayesian Network* (DBN), *Factor Graph*, *Probabilistic Graphical Models* (PGM), *Tree Augmented Naïve Bayes* (TAN) dan algoritma lainnya. Sedangkan pada metode *Multi Layer Perceptron* (MLP) dapat menggunakan metode *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), *Deep Belief Network* (DBN) atau dengan kendala dataset yang lebih rendah dapat menggunakan *Generative Adversarial Networks* (GAN).

Daftar Pustaka

- Abousteit, M. H. S., Tammam, A. F., & Wahdan, A. M. (2020). A novel approach for generating one-time password with secure distribution. *Proceedings of the World Conference on Smart Trends in Systems, Security and Sustainability, WS4 2020*, 461–466. <https://doi.org/10.1109/WorldS450073.2020.9210322>
- Adugna, T. D., Ramu, A., & Haldorai, A. (2006). A Review of Pattern Recognition and Machine Learning. In *Journal of Machine and Computing* (Vol. 4, Issue 1). <https://doi.org/10.53759/7669/jmc202404020>
- Alharbi, E. T., & Alghazzawi, D. (2019). Two Factor Authentication Framework Using OTP-SMS Based on Blockchain. *Transactions on Machine Learning and Artificial Intelligence*, 7(3), 17–27. <https://doi.org/10.14738/tmlai.73.6524>
- Alickovic, E., & Subasi, A. (2020). Normalized neural networks for breast cancer classification. In *IFMBE Proceedings* (Vol. 73). Springer International Publishing. https://doi.org/10.1007/978-3-030-17971-7_77
- Bakar, R. A., Aziz, N. A., Muhammad, A., & Muda, M. (2017). Perceived Ease of Use, Security and Privacy of Mobile Banking. *International Journal of Business, Economics and Law*, 13(2), 56–62.
- Barlas, A., Valakosta, A., Katsionis, C., Oikonomou, A., & Brinia, V. (2023). The Effect of Corporate Social Responsibility on Customer Trust and Loyalty. *Sustainability (Switzerland)*, 15(2), 1–14. <https://doi.org/10.3390/su15021036>
- Cai, B., Kong, X., Liu, Y., Lin, J., Yuan, X., Xu, H., & Ji, R. (2019). Application of Bayesian Networks in Reliability Evaluation. *IEEE Transactions on Industrial Informatics*, 15(4), 2146–2157. <https://doi.org/10.1109/TII.2018.2858281>
- Chai, S. S., Cheah, W. L., Goh, K. L., Chang, Y. H. R., Sim, K. Y., & Chin, K. O. (2021). A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia. *Computational and Mathematical Methods in Medicine*, 2021. <https://doi.org/10.1155/2021/2794888>
- Chickering, D. M. (1996). *Learning Bayesian Networks is NP-Complete* (D. Fisher & H.-J. Lenz (eds.); pp. 121–130). Springer New York. <https://doi.org/10.1007/978-1-4612->

- Chow, Y. W., Susilo, W., Au, M. H., & Barmawi, A. M. (2015). A visual one-time password authentication scheme using mobile devices. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8958, 243–257. https://doi.org/10.1007/978-3-319-21966-0_18
- da Silva, M. M. O., Teixeira, J. M. X. N., Cavalcante, P. S., & Teichrieb, V. (2019). Perspectives on how to evaluate augmented reality technology tools for education: a systematic review. *Journal of the Brazilian Computer Society*, 25(1). <https://doi.org/10.1186/s13173-019-0084-8>
- Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, 5(6). <https://doi.org/10.1016/j.heliyon.2019.e01802>
- Díaz-Ramírez, J. (2021). Machine Learning and Deep Learning. *Ingeniare*, 29(2), 182–183. <https://doi.org/10.4067/S0718-33052021000200180>
- Domingos, P., & Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*, 29, 103–130. <https://link.springer.com/article/10.1023/A:1007413511361>
- Drummond, C. (2011). Classification BT - Encyclopedia of Machine Learning. In C. Sammut & G. I. Webb (Eds.), *Springer* (p. 171). Springer US. https://doi.org/10.1007/978-0-387-30164-8_111
- Du, P., Zhao, S., Xing, C., Chen, X., Hu, H., Ren, F., Zhang, M., Xie, L., Huang, X., & Wen, J. (2023). Hydrodynamic detection based on multilayer perceptron and optimization using dynamic mode decomposition. *Ocean Engineering*, 278, 114258. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.114258>
- El Hlouli, F. Z., Riffi, J., Mahraz, M. A., El Yahyaouy, A., & Tairi, H. (2020). Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures. *2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020*, 3–7. <https://doi.org/10.1109/ISCV49265.2020.9204185>
- Farooq, U. (2020). Real Time Password Strength Analysis on a Web Application Using Multiple Machine Learning Approaches. *International Journal of Engineering*

- Research & Technology*, 9(12), 359–364. www.ijert.org
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning Book. *MIT Press*, 326. <https://doi.org/10.1038/nmeth.3707>
- Hartsook, R. F. (2000). Let's talk money. *Fund Raising Management*, 31(7), 31–32. <https://doi.org/10.4324/9781003346463-14>
- Haykin, S. (2009). Neural Networks and Learning Machines. In *Pearson Education, Inc.* (3rd ed., Vols. 1–3). Pearson. <https://doi.org/10.1016/B978-0-12-809633-8.20339-7>
- Heckerman, D. (1997). Bayesian Networks for Data Mining. *Kluwer Academic Publishers*, 119, 79–119.
- Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data DAVID. *Machine Learning*, 20, 197–243.
- Ilahi, A., & Axhausen, K. W. (2019). Integrating Bayesian network and generalized raking for population synthesis in Greater Jakarta. *Regional Studies, Regional Science*, 6(1), 623–636. <https://doi.org/10.1080/21681376.2019.1687011>
- Joseph, V. R., Joseph, V. R., & Stewart, H. M. (2022). Optimal ratio for data splitting. *March*, 531–538. <https://doi.org/10.1002/sam.11583>
- Kasi, U. (2019). Extraction of Bank Transaction Data and Classification using Naive Bayes. *International Journal of Science and Research (IJSR)*, 9(5), 3–8. <https://doi.org/10.21275/SR20510143811>
- Kumari, M., Vohra, R., & Arora, A. (2014). Prediction of Diabetes Using Bayesian Network. *International Journal of Computer Science and Information Technologies*, 5(4), 5174–5178. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.640.3573>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lidwina, A. (2020). *Jadwal Belanja Konsumen di E-Commerce Kebanyakan di Siang Hari*. Data Boks. <https://databoks.katadata.co.id/datapublish/2021/06/09/jadwal-belanja-konsumen-di-e-commerce-kebanyakan-di-siang-hari>
- Luo, N., Nara, A., & Izumi, K. (2021). An interaction-based bayesian network framework for surgical workflow segmentation. *International Journal of Environmental Research and Public Health*, 18(12). <https://doi.org/10.3390/ijerph18126401>

- Lv, T., Yan, P., Yuan, H., & He, W. (2020). Spam Filter Based on Naive Bayesian Classifier. *Journal of Physics: Conference Series*, 1575(1).
<https://doi.org/10.1088/1742-6596/1575/1/012054>
- Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4).
<https://doi.org/10.3390/technologies9040081>
- Mitchell, T. M. (1997). Machine Learning. In *Natural Computing Series*. McGraw-Hill Science/Engineering/Math. https://doi.org/10.1007/978-3-031-17922-8_9
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning* (2nd ed.). MIT Press.
- Naik, A., Phulmamdikar, K., Pradhan, S., & Thorat, S. (2016). Real Time Credit Card Transaction Analysis. *I(11)*, 1663–1666.
- Ncibi, K., Sadraoui, T., Faycel, M., & Djenina, A. (2017). A Multilayer Perceptron Artificial Neural Networks Based a Preprocessing and Hybrid Optimization Task for Data Mining and Classification. *International Journal of Econometrics and Financial Management*, 5(1), 12–21. <https://doi.org/10.12691/ijefm-5-1-3>
- Ning, B., Junwei, W., & Feng, H. (2019). Spam message classification based on the naïve Bayes classification algorithm. *IAENG International Journal of Computer Science*, 46(1).
- Ojha, R., Ghadge, A., Tiwari, M. K., & Bititci, U. S. (2018). Bayesian network modelling for supply chain risk propagation. *International Journal of Production Research*, 56(17), 5795–5819. <https://doi.org/10.1080/00207543.2018.1467059>
- Oyegoke, T. O., Akomolede, K. K., Aderounmu, A. G., & Adagunodo, E. R. (2021). A Multi-Layer Perceptron Model for Classification of E-mail Fraud. *European Journal of Information Technologies and Computer Science*, 1(5), 16–22.
<https://doi.org/10.24018/compute.2021.1.5.24>
- Platform, C. (2024). *SMS Delivery Statuses*. Bird. <https://docs.bird.com/connectivity-platform/sending-sms/sms-delivery-statuses>
- Polotskaya, K., Muñoz-Valencia, C. S., Rabasa, A., Quesada-Rico, J. A., Orozco-Beltrán, D., & Barber, X. (2024). Bayesian Networks for the Diagnosis and Prognosis of

- Diseases: A Scoping Review. *Machine Learning and Knowledge Extraction*, 6(2), 1243–1262. <https://doi.org/10.3390/make6020058>
- Pryor, L., Mallet, J., Dave, R., Seliya, N., Vanamala, M., & Sowell-Boone, E. (2022). Evaluation of a User Authentication Schema Using Behavioral Biometrics and Machine Learning. *Computer and Information Science*, 15(3), 1. <https://doi.org/10.5539/cis.v15n3p1>
- Rahmad, F., Suryanto, Y., & Ramli, K. (2020). Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification. *IOP Conference Series: Materials Science and Engineering*, 879(1). <https://doi.org/10.1088/1757-899X/879/1/012076>
- Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. <http://arxiv.org/abs/1811.12808>
- Rayes, M. O. (2011). *One-Time Password BT - Encyclopedia of Cryptography and Security* (H. C. A. van Tilborg & S. Jajodia (eds.); pp. 885–887). Springer US. https://doi.org/10.1007/978-1-4419-5906-5_785
- Russell, S., & Novig, P. (2020). *Artificial Intelligence: A Modern Approach 4th Edition* (4th ed.). Pearson.
- Sable, A. S., & Kalavadekar, P. N. (2016). SMS Classification Based on Naive Bayes Classifier and Semi-Supervised Learning. *International Journal of Modern Trends in Engineering and Research*, 3(7), 16–25.
- Salehi, M., Farhadi, S., Moieni, A., Safaie, N., & Ahmadi, H. (2020). Mathematical Modeling of Growth and Paclitaxel Biosynthesis in *Corylus avellana* Cell Culture Responding to Fungal Elicitors Using Multilayer Perceptron-Genetic Algorithm. *Frontiers in Plant Science*, 11(August), 1–12. <https://doi.org/10.3389/fpls.2020.01148>
- Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Senior, A., Heigold, G., Ranzato, M., & Yang, K. (2013). An Empirical Study of Learning Rates in Deep Neural Networks for Speech Recognition. *New York*, 1, 6724–6728. <https://doi.org/10.1109/ICASSP.2013.6638963>
- Spirtes, P., & Zhang, K. (2019). Causal discovery and inference: concepts and recent

- methodological advances. *Applied Informatics*, 3(1). <https://doi.org/10.1186/s40535-016-0018-x>
- Suryadewiansyah, M. K., Endra, T., & Tju, E. (2020). Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 8(2), 81–88.
- Tarigan, U. P. P., Novemingsen, R., & Lu, D. (2023). Implementation of Lean Six Sigma to Reduce Work Time Waste in the Goods Transportation Department. *JKIE (Journal Knowledge Industrial Engineering)*, 10(1), 1–11. <https://doi.org/10.35891/jkie.v10i1.3730>
- Thohari, A., & Hertantyo, G. B. (2018). Implementasi Convolutional Neural Network untuk Klasifikasi Pembalap MotoGP Berbasis GPU. *Proceedings on Conference on Electrical Engineering, Telematics, Industrial Technology, and Creative Media*, 50–55.
- Umair, M., Saeed, Z., Ahmad, M., Amir, H., Akmal, B., & Ahmad, N. (2020). Multi-class Classification of Bi-lingual SMS using Naive Bayes Algorithm. *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*. <https://doi.org/10.1109/INMIC50486.2020.9318153>
- Vijay, & Kumar, S. (2021). Spam SMS Detection Using Naive Bayes Classifier. *International Journal of Scientific Research and Engineering Development*, 4(1), 561–563.
- Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes : applications , variations and vulnerabilities : a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277–2293. <https://doi.org/10.1007/s00500-020-05297-6>
- Xendit. (2022). *e-Wallet Payments usage and trends in Indonesia*. Xendit.
- Xu, C., Coen-Pirani, P., & Jiang, X. (2023). Empirical Study of Overfitting in Deep Learning for Predicting Breast Cancer Metastasis. *Cancers*, 15(7), 6. <https://doi.org/10.3390/cancers15071969>
- Zhang, H. (2004). The optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004*, 2, 562–567.

LAMPIRAN A

{Judul Lampiran}

{Isi lampiran, pastikan semua lampiran telah diacu di isi/body tesis. Perhatikan bahwa penomoran lampiran menggunakan huruf dan kemudian tercantum di dalam daftar isi}