

Integrasi *Business Analytics* dalam Manajemen Performa Jaringan Seluler 4G

Eka Kosasih*¹, Tanty Oktavia²

^{1,2} Information System Management Department, BINUS Graduate Program – Bina Nusantara University, Indonesia 11480

e-mail: *1eka.kosasih@binus.ac.id, 2toktavia@binus.edu

Abstrak

Analitik bisnis berperan penting dalam mengoptimalkan kinerja jaringan telekomunikasi. Penelitian ini menggunakan metodologi data mining, khususnya pendekatan Knowledge Discovery Database (KDD), bersama algoritma machine learning untuk memprediksi dan meningkatkan kualitas sinyal 4G di Indonesia. Data real-time dari proyek lokasi baru salah satu operator telekomunikasi digunakan dalam studi ini. Dengan SQL ClickHouse untuk pemrosesan data, Python untuk machine learning, dan Tableau untuk visualisasi, penelitian ini mengembangkan model analisis data berbasis hidden Markov models (HMM). Topik ini dipilih karena meningkatnya kebutuhan akan jaringan seluler yang andal dan berkualitas tinggi, yang memerlukan teknik analitik canggih untuk memantau dan meningkatkan kinerja jaringan. HMM diimplementasikan untuk memberikan prediksi akurat dan wawasan mengenai perbaikan sinyal. Hasil awal menunjukkan model ini efektif dalam mengidentifikasi indikator kinerja utama seperti akurasi, presisi, dan recall, serta tren kualitas sinyal jaringan. Akurasi model mencapai 93,06%, dengan presisi 96,97%, recall 95,52%, dan F1-score 96,24%. Studi ini menunjukkan potensi besar integrasi analitik bisnis dan machine learning dalam sektor telekomunikasi, meningkatkan kualitas layanan dan efisiensi operasional.

Kata kunci—3-5 Business Analytics, Optimasi Jaringan, Hidden Markov Model, Data Mining, Prediksi Kualitas Sinyal

Abstract

Business analytics plays a crucial role in optimizing telecommunication network performance. This research focuses on using data mining methodologies, specifically the Knowledge Discovery Database (KDD) approach, combined with machine learning algorithms, to predict and enhance 4G signal quality for a mobile operator in Indonesia. The study leverages real-time data from a new site project of an Indonesian telecom operator. Employing SQL ClickHouse for data processing, Python for machine learning, and Tableau for visualization, the research develops a comprehensive data analysis model using hidden Markov models (HMM). The selection of this topic stems from the increasing demand for reliable and high-quality mobile networks, which necessitates advanced analytical techniques to monitor and improve network performance. By implementing HMM, the research aims to provide accurate predictions and actionable insights into signal quality improvements. Preliminary results demonstrate the model's effectiveness in identifying key performance indicators such as accuracy, precision, and recall, as well as trends in network signal quality. The model achieved an accuracy of 93,06%, precision of 96,97%, recall of 95,52%, and F1-score of 96,24%. This study highlights the significant potential of integrating business analytics and machine learning in the telecommunications sector, offering valuable contributions to the industry by improving service quality and operational efficiency.

Keywords—Business Analytics, Network Optimization, Hidden Markov Model, Data Mining, Signal Quality Prediction.

1. PENDAHULUAN

Dalam era digital saat ini, tengah pesatnya perkembangan teknologi komunikasi, kualitas sinyal merupakan aspek krusial dalam menentukan keberhasilan layanan jaringan telekomunikasi. Penggunaan teknologi 4G yang semakin meluas menuntut operator seluler untuk terus meningkatkan performa jaringan agar dapat memenuhi ekspektasi pengguna yang tinggi akan kecepatan dan stabilitas koneksi [1]. Dalam konteks ini, analitik bisnis dan teknik data mining telah muncul sebagai alat yang sangat berharga untuk memahami dan mengoptimalkan kinerja jaringan.

Pendekatan *Knowledge Discovery Database* (KDD)[2], yang merupakan salah satu metode dalam *data mining*, memungkinkan analisis mendalam dari kumpulan data besar dan kompleks. Ketika dikombinasikan dengan algoritma *machine learning* seperti *hidden Markov models* (HMM) [3], metode ini dapat menghasilkan prediksi yang akurat dan *insight* yang bermanfaat mengenai kualitas sinyal. Hasil analisis ini sangat penting untuk mengidentifikasi dan mengatasi potensi masalah dalam jaringan, serta untuk merancang strategi optimasi yang efektif.

Penelitian terdahulu telah menunjukkan berbagai aplikasi *machine learning* dan analitik data dalam optimasi jaringan telekomunikasi. Misalnya, Haider mendeteksi anomali KPI menggunakan *machine learning*, dengan teknik deteksi anomali menggunakan outlier dan perhitungan IQR (Interquartile Range) serta ELBO (Evidence Lower Bound) [4]. Penelitian ini menunjukkan bagaimana *machine learning* dapat digunakan untuk mendeteksi ketidaknormalan dalam data KPI, yang relevan dalam konteks analisis performa jaringan. Penelitian berikutnya menerapkan teknik *unsupervised machine learning* untuk menilai kinerja jaringan LTE melalui analisis *Key Performance Indicators* (KPI) dan menemukan kelompok sel yang menunjukkan kinerja serupa [5]. Pendekatan ini menekankan pentingnya evaluasi kinerja jaringan dengan menggunakan *clustering* untuk mengidentifikasi sel yang memiliki performa yang tidak diinginkan, yang sangat berguna dalam optimasi jaringan. Penelitian Omer memantau dan mengotomatisasi kinerja Radio Access Networks menggunakan Markov chain, yang menunjukkan akurasi 94,61% dalam memprediksi status KPI [6]. Penelitian ini memperlihatkan bagaimana penggunaan model *markov chain* dapat meningkatkan akurasi prediksi kinerja jaringan, khususnya dalam hal aksesibilitas dan retensibilitas.

Hendrawan menggunakan *Discrete Time Markov Chain* (DTMC) untuk memprediksi degradasi KPI *accessibility* [7]. Ini memberikan landasan bagi penggunaan model Markov untuk prediksi degradasi kinerja jaringan, meskipun dengan pendekatan yang sedikit berbeda dibandingkan dengan penelitian ini yang menggunakan HMM. Sementara penelitian Ming Yan meningkatkan akurasi prediksi dengan model Markov berbobot berdasarkan klasifikasi pengguna seluler [8]. Kedua studi ini menunjukkan berbagai aplikasi model Markov dalam prediksi mobilitas pengguna, yang penting untuk optimasi handover dan kualitas layanan.

Pada penelitian kali ini bertujuan untuk menerapkan metodologi KDD dan HMM dalam konteks peningkatan kualitas sinyal 4G untuk operator telekomunikasi di Indonesia. Dengan menggunakan data *real-time* dari proyek lokasi baru, penelitian ini memanfaatkan SQL *ClickHouse* untuk pemrosesan data [9], *Python* untuk pengembangan model *machine learning* [10], dan Tableau untuk visualisasi hasil. Fokus utama dari penelitian ini adalah untuk mengembangkan model analisis data yang dapat memprediksi dan meningkatkan kualitas sinyal, serta memberikan wawasan yang bermanfaat bagi pengambilan keputusan dalam optimasi jaringan.

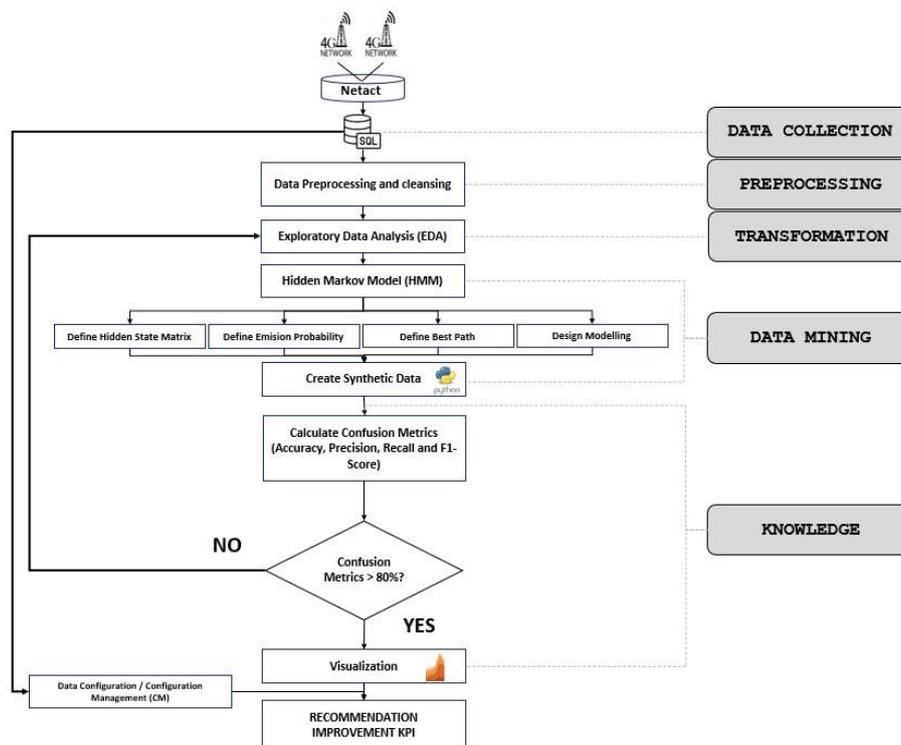
Latar belakang penelitian ini didorong oleh kebutuhan mendesak untuk jaringan seluler yang lebih andal dan berkualitas tinggi. Dengan integrasi teknik analitik canggih, penelitian ini berupaya memberikan kontribusi yang signifikan terhadap peningkatan layanan dan efisiensi operasional di sektor telekomunikasi.

2. METODE PENELITIAN

Metodologi penelitian didefinisikan sebagai pencarian informasi yang relevan secara ilmiah dan sistematis tentang suatu topik tertentu. Bagian sub bab ini berhubungan dengan bagaimana data dikumpulkan dan dianalisis. Penelitian ini fokus pada membuat model prediksi KPI LTE menggunakan algoritma *machine learning*. Terdapat tiga metode dalam mengolah data menjadi pengetahuan yang berguna. Ketiga metode itu adalah *knowledge Discovery in Database (KDD)*, *Cross-Industry Standard Process for Data Mining (CRISP-DM)* dan SEMMA (*Stage, Explore, Modify, Model, Access*).

Pada penelitian ini metodologi yang dilakukan adalah menggunakan pendekatan metodologi *Knowledge Discovery in Database (KDD)* [11]. KDD memiliki pendekatan yang lebih tradisional dan akademis. Meskipun metodologi ini digunakan oleh para spesialis dari berbagai bidang, namun pada dasarnya mengikuti prinsip yang sama dari analisis data yang bersifat sekuensial dan iteratif. Gambar 1 merupakan alur framework konseptual penelitian ini.

Setiap tahapan dalam KDD, mulai dari pemilihan data hingga evaluasi hasil, dirancang untuk memastikan bahwa setiap aspek analisis data dilakukan dengan hati-hati dan cermat. Tahapan awal melibatkan pemahaman dan pemilihan data yang relevan untuk diolah, diikuti dengan tahap pra-pemrosesan data untuk membersihkan dan mempersiapkan data mentah. Transformasi data dilakukan untuk mengubah data menjadi format yang lebih cocok untuk analisis, kemudian dilanjutkan dengan penerapan algoritma *machine learning* untuk menemukan pola dan menghasilkan model prediktif yang akurat. Proses ini diakhiri dengan evaluasi, yaitu dengan menghitung *confusion metrics* yang terdiri dari empat komponen (*accuracy, precision, recall dan F1-score*) dimana model dianggap baik jika hasilnya lebih dari 80%, dan interpretasi hasil untuk memastikan bahwa model yang dihasilkan dapat memberikan wawasan yang berguna dan aplikatif dalam manajemen performa jaringan seluler 4G. Berikut penjelasan tahapan-tahapannya dengan menggunakan pendekatan metodologi KDD.



Gambar 1 Framework Penelitian dengan KDD

2.1 Data collection

Tahap awal melibatkan pengumpulan data *real-time* dari eNB jaringan seluler LTE, yang disimpan dalam *database Network Management System (NMS)* Nokia bernama Netact. Data Netact, yang terbagi berdasarkan operator, ditransfer ke *server SQL* lokal dan berisi KPI untuk *monitoring* performa jaringan LTE [12], parameter konfigurasi sel, dan *counter* untuk analisa performa jaringan. Dataset ini digunakan untuk pemodelan prediksi KPI jika ada perubahan konfigurasi. Pada penelitian ini, data diambil dari salah satu perusahaan *vendor* telekomunikasi dengan *granularity* selama 6 bulan dan merupakan data terbaru dari salah satu proyek yang sedang berjalan. Data diagregasi pada tingkat *site* dengan mengambil sampel sebanyak 1600 *new site*. Atribut yang diambil terdiri dari *timestamp daily* sebagai *timeseries*, informasi *site* 4G, atribut parameter, atribut KPI, dan atribut *counter-counter* sebagai data pendukung.

Data yang diolah dalam penelitian ini mencakup beberapa jenis informasi penting yang dikategorikan dalam tabel 1 berikut.

Tabel 1 Kategori dataset penelitian

Tipe	Nama Atribut	Tipe Data
KPI	<i>Success rate of RRC connection setup</i>	<i>float64</i>
	<i>Success rate of eRAB setup</i>	<i>float64</i>
	<i>Success rate of SI signaling</i>	<i>float64</i>
	<i>Drop rate of eRAB</i>	<i>float64</i>
	<i>Success rate of intra-frequency handover</i>	<i>float64</i>
	<i>Success rate of inter-frequency handover</i>	<i>float64</i>
Parameter Cell	<i>Cell_ID</i>	<i>Integer</i>
	<i>eNB_ID</i>	<i>Integer</i>
	<i>PCI: Physical Cell ID</i>	<i>Integer</i>
	<i>Sudut antena (angle)</i>	<i>Integer</i>
	<i>Tacking Area Code (TAC)</i>	<i>Integer</i>
	<i>Bandwidth</i>	<i>Integer</i>
Counter	<i>avg_cell_range_n2</i>	<i>float64</i>
	<i>avg_ho_cellpair_distance</i>	<i>float64</i>
	<i>high_ta_share</i>	<i>float64</i>
	<i>irat_total_share</i>	<i>float64</i>
	<i>e_rab_stp_att</i>	<i>float64</i>
	<i>RACH_STP_COMPLETION_SR_LTE_5569a</i>	<i>float64</i>
	<i>RACH_STP_ATT</i>	<i>float64</i>
	<i>COMP_CONT_BASED_RACH_STP_SR_LTE_5670a</i>	<i>float64</i>
	<i>ul_interference_dbm</i>	<i>float64</i>
Timeseries	<i>xDate</i>	<i>Date</i>

2.2 Preprocessing

Tahap kedua melibatkan pemrosesan dan pembersihan dataset menggunakan SQL. Nilai setiap KPI diperiksa dan dibersihkan, di mana *missing data* atau nilai *NULL* diisi dengan nol menggunakan fungsi SQL *IFNULL (expression, value)*. Selain itu, dilakukan agregasi nilai KPI dari *level cell* ke *level site* menggunakan SQL. Tahap ini juga membentuk parameter observasi yang diperlukan untuk prediksi jaringan, seperti indikasi masalah *coverage* dan *interference*.

2. 3 Transformasi

Tahap ketiga dilakukan *exploratory data analysis* (EDA) [13] untuk memahami isi dan parameter dalam dataset. Pada tahap ini, dilakukan *encoding* data KPI, di mana nilai-nilai yang memenuhi *threshold* KPI akan diberi status "*improved*", nilai di antara *threshold* dan *threshold*-5% akan diberi status "*maintain*", dan nilai di bawah *threshold*-5% akan diberi status "*degraded*". Ketiga status ini merupakan *hidden state* yang tidak dapat diamati langsung. Selain itu, dibuat *observation state* berdasarkan *parameter* observasi yang dipersiapkan sebelumnya, yaitu "*coverage*" dan "*interference*", yang bisa diamati langsung dari data namun bergantung pada *hidden state*.

2. 4 Data mining

Tahap keempat merupakan proses pembuatan *modelling* dengan membuat *synthetic feature* menggunakan algoritma *Hidden Markov Model* (HMM). HMM sering digunakan dalam pemrosesan sinyal, pengenalan pola, dan berbagai aplikasi lain di mana sistem yang diobservasi dapat dijelaskan melalui probabilitas transisi antara *state* yang tidak dapat diamati langsung.

Pada HMM terdapat beberapa tahapan. Tahapan pertama adalah menghitung probabilitas perpindahan *hidden state* dari *state i* ke *state j*. Secara matematis, probabilitas tersebut dapat dituliskan sebagai berikut:

$$a_{ij} = P(S_{t+1} = j | S_t = i) \quad (1)$$

Tahapan yang kedua adalah menghitung probabilitas emisi yaitu dengan mengamati simbol k , dalam hal ini adalah masalah yang diobservasi, pada *state j*. Secara matematis dapat dituliskan seperti rumus berikut:

$$a_j(k) = P(O_t = k | S_t = j) \quad (2)$$

Selanjutnya adalah menghitung vektor probabilitas awal (π) dan algoritma viterbi untuk mendapatkan urutan *hidden state* yang paling optimal. Proses tersebut dapat di hitung dengan menggunakan persamaan 3 dan 4 berikut.

$$\pi_i = P(S_1 = i) \quad (3)$$

$$\delta_t(i) = \max_{\pi} P(O_1, O_2, \dots, O_t, S_t = i | \lambda) \quad (4)$$

Berikut adalah proses yang dilakukan pada tahapan *data mining*.

2. 4.1 Hidden state

Hidden state pada penelitian ini mencerminkan kondisi fluktuasi dari nilai KPI yang dikategorikan sebagai "*improved*", "*maintain*", dan "*degraded*". Ini penting untuk memahami dinamika performa jaringan dan bagaimana perubahan kecil dapat berdampak besar. Setelah menentukan *hidden state*, selanjutnya adalah membuat matrix probabilitas *hidden state*. Matriks ini menunjukkan kemungkinan perpindahan dari satu *hidden state* ke *hidden state* lainnya dalam periode waktu tertentu. Dengan menghitung probabilitas transisi ini, model dapat memprediksi bagaimana performa jaringan kemungkinan besar akan berubah dari waktu ke waktu. Ini memberikan wawasan penting tentang stabilitas jaringan dan potensi perbaikan atau degradasi. Matriks ini juga membantu dalam memahami pola dan tren jangka panjang dalam data performa jaringan.

2. 4.2 Probabilitas emisi

Setelah menghitung matrix probabilitas *hidden state*, maka selanjutnya adalah menentukan *observation state* dan menghitung matrix probabilitas emisi. *Observation state* adalah kondisi umum masalah di jaringan LTE, seperti masalah "*coverage*" dan "*interference*". Setelah mendapatkan *observation state*, Langkah selanjutnya adalah menghitung matrix probabilitas emisi dari *observation state* ke *hidden state*. Matriks ini menunjukkan kemungkinan

terjadinya *observation state* tertentu berdasarkan *hidden state* yang ada. Ini berarti model dapat memperkirakan masalah apa yang mungkin muncul jika jaringan berada dalam kondisi tertentu. Probabilitas emisi membantu menghubungkan kondisi tersembunyi dengan gejala yang teramati, yang merupakan langkah penting dalam diagnosa masalah jaringan.

2. 4.3 Vektor probabilitas awal (π)

Vektor probabilitas awal (π) digunakan sebagai titik awal dalam model untuk menentukan distribusi awal dari *hidden state*. Probabilitas ini menunjukkan kemungkinan awal setiap *hidden state* sebelum data observasi digunakan.

2. 4.4 Viterbi

Setelah mendapatkan matrix probabilitas *hidden state*, matrix probabilitas emisi dan nilai π , maka selanjutnya dilakukan pemodelan prediksi dengan mencari urutan *hidden states* yang paling mungkin berdasarkan *observation state*. Hasil ini bisa dilakukan dengan menggunakan algoritma viterbi [14]. Berikut tahapan-tahapannya.

- **Inisialisasi** yaitu menentukan nilai awal untuk matriks probabilitas *hidden state*, matriks probabilitas emisi, dan π . Langkah ini menetapkan dasar bagi proses prediksi dengan memberikan nilai awal yang diperlukan untuk perhitungan selanjutnya.
- **Rekursi** adalah mengulangi proses inisialisasi mulai dari waktu t hingga $t+n$ untuk memperbaharui nilai parameter. Proses ini melibatkan perhitungan berulang untuk memperbaharui nilai parameter berdasarkan data baru yang masuk, memastikan model tetap relevan dan akurat.
- **Terminasi**, menghitung dan menyimpulkan hasil akhir dari perhitungan probabilitas atau jalur optimal. Ini adalah tahap akhir dari proses perhitungan, di mana model menghasilkan prediksi yang siap untuk dievaluasi.
- **Path backtracking** yaitu menelusuri kembali jalur untuk menemukan urutan *hidden states* yang paling mungkin berdasarkan *observation state*. Langkah ini penting untuk memahami urutan peristiwa yang mengarah pada kondisi jaringan saat ini, memungkinkan analisis yang lebih mendalam dan prediksi yang lebih akurat.

2. 5 Knowledge

Hasil dari *path backtracking* digabungkan dengan data nyata dan dibandingkan untuk menghasilkan data sintesis yang dievaluasi dan divisualisasikan. Proses ini melibatkan validasi model dengan membandingkan prediksi dengan data aktual untuk memastikan akurasi dan keandalan model. Evaluasi ini juga mengidentifikasi area yang memerlukan penyesuaian atau peningkatan pada model. Salah satu metode evaluasi utama adalah *confusion matrix*, yang membantu dalam mengukur kinerja model. *Confusion matrix* menyajikan informasi tentang jumlah *true positives* (TP), *false positives* (FP), *true negatives* (TN), dan *false negatives* (FN) [15]. Metrik evaluasi utama yang dihitung dari *confusion matrix* meliputi:

- Akurasi (*Accuracy*), yaitu untuk mengukur seberapa baik model dalam mengklasifikasikan data secara keseluruhan. Akurasi dihitung sebagai rasio jumlah prediksi yang benar terhadap total prediksi yang dibuat. Akurasi dapat dihitung menggunakan formula berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

Di mana TP adalah jumlah prediksi benar sebagai positif, TN adalah jumlah prediksi benar sebagai negatif, FP adalah jumlah prediksi salah sebagai positif, dan FN adalah jumlah prediksi salah sebagai negatif.

- Presisi (*Precision*) digunakan untuk mengukur seberapa akurat model dalam mengidentifikasi positif. Presisi adalah rasio prediksi benar positif terhadap semua prediksi positif yang dibuat. Presisi tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

- *Recall* atau disebut juga sebagai *sensitivity*, yaitu untuk mengukur seberapa baik model dalam mendeteksi semua kasus positif yang sebenarnya. *Recall* adalah rasio prediksi benar positif terhadap total kasus positif yang sebenarnya. *Recall* tinggi menunjukkan bahwa model berhasil mendeteksi sebagian besar dari kasus positif yang ada.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

- *F1 Score* dihitung sebagai rata-rata harmonis dari presisi dan *recall*. *F1 Score* memberikan gambaran tentang keseimbangan antara presisi dan *recall*, dan berguna ketika ada ketidakseimbangan antara kelas positif dan negatif.

$$F1\ Score = 2 \times \frac{PRECISION \times RECALL}{PRECISION+RECALL} \quad (8)$$

Visualisasi hasil evaluasi dilakukan untuk memudahkan pemahaman data dan pengambilan keputusan. Grafik dan plot dari metrik evaluasi memberikan wawasan yang mendalam tentang performa model dan area yang memerlukan penyesuaian lebih lanjut.

3. HASIL DAN PEMBAHASAN

Setelah melalui tahapan pemodelan dengan pendekatan metodologi *Knowledge Discovery in Database* (KDD), berikut adalah hasil yang diperoleh dari penelitian ini. Namun, bagian pembahasan hanya mencakup proses *data mining* yang dilakukan.

3.1 Menentukan matriks probabilitas transisi *hidden state*

Matriks probabilitas transisi *hidden state* di hitung berdasarkan perpindahan dari satu *hidden state* ke *hidden state* lainnya dalam periode waktu t hingga $t+n$. Analisis ini secara khusus berfokus pada data yang diekstraksi dari *site* dengan id 01SNB0015, dengan observasi KPI adalah *Intra Frekuensi Handover Succes Rate* (*Intra Freq HOSR*). Dari hasil percobaan diperoleh probabilitas transisi dari *improved* ke *maintain* adalah 0,1489, dari *maintain* ke *degraded* adalah 0,526, dan dari *degraded* ke *improved* adalah 0,80. Hal ini mengindikasikan bahwa jaringan yang berada dalam kondisi baik (*improved*) cenderung tetap stabil, namun ada kemungkinan penurunan ke kondisi *maintain* atau bahkan *degraded*. Dengan menggunakan *python*, data transisi *hidden state* yang dihasilkan seperti pada tabel 2 berikut.

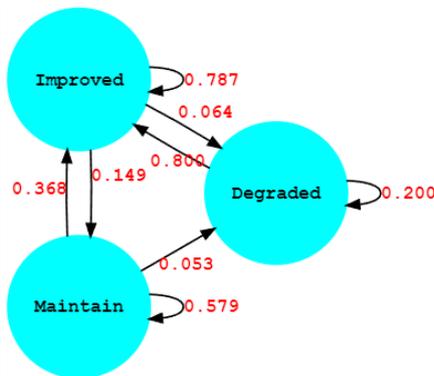
Tabel 2 Matriks probabilitas transisi *hidden state*

	<i>Degraded</i>	<i>Maintain</i>	<i>Improved</i>
<i>Degraded</i>	20,00%	0,00%	80,00%
<i>Maintain</i>	5,26%	57,89%	36,84%
<i>Improved</i>	6,38%	14,89%	78,72%

Secara matematis, matriks transisi untuk *hidden state* kasus diatas dirumuskan sebagai berikut:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0,200 & 0,000 & 0,800 \\ 0,053 & 0,579 & 0,368 \\ 0,064 & 0,149 & 0,787 \end{bmatrix} \quad (9)$$

Dengan menggunakan *library network package* dari *graphviz* [16] diagram Markov dapat divisualisasikan seperti yang ditunjukkan dalam gambar 2.



Gambar 2 Probabilitas *hidden state*

3. 2 Menentukan matriks probabilitas emisi

Matriks probabilitas emisi menunjukkan bahwa masalah *coverage* lebih sering terjadi dalam kondisi *maintain* dengan probabilitas 0,9474. Namun perlu diingat bahwa kondisi *maintain* masih dibawah nilai *threshold* KPI. Tabel 3 adalah hasil perhitungan matriks probabilitas emisi.

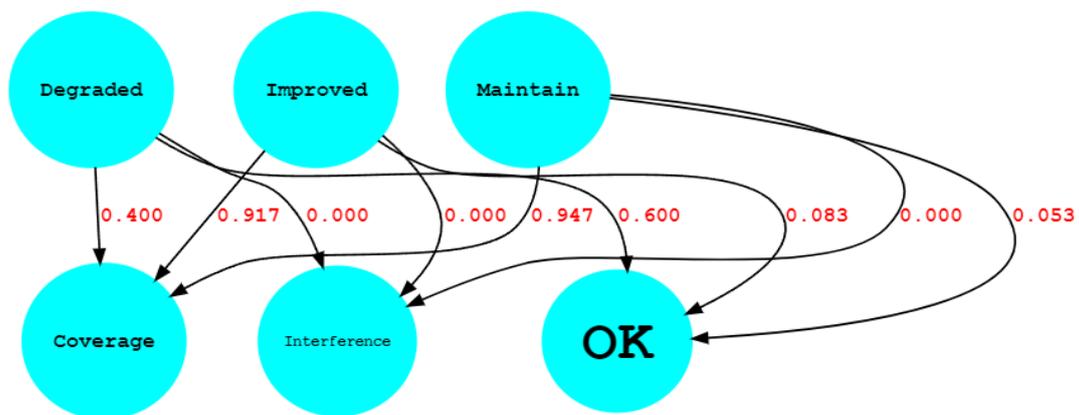
Tabel 3 Matriks probabilitas emisi

	<i>Coverage</i>	<i>Interference</i>	<i>OK</i>
<i>Degraded</i>	40,00%	0,00%	60,00%
<i>Maintain</i>	94,74%	0,00%	5,26%
<i>Improved</i>	91,67%	0,00%	8,33%

Transposisi ke format matriks sebagai berikut:

$$B = \begin{bmatrix} b_{1v_1} & b_{1v_2} & b_{1v_3} \\ b_{2v_1} & b_{2v_2} & b_{2v_3} \\ b_{3v_1} & b_{3v_2} & b_{3v_3} \end{bmatrix} = \begin{bmatrix} 0,40 & 0,00 & 0,60 \\ 0,9474 & 0,00 & 0,0526 \\ 0,9167 & 0,00 & 0,0833 \end{bmatrix} \tag{10}$$

Visualisasi diagram Markov untuk probabilitas emisi dapat dilihat pada gambar 3.



Gambar 3 Probabilitas emisi

3. 3 Menentukan probabilitas awal (π)

Berdasarkan perhitungan probabilitas emisi, distribusi probabilitas awal dapat diperoleh sebagai berikut.

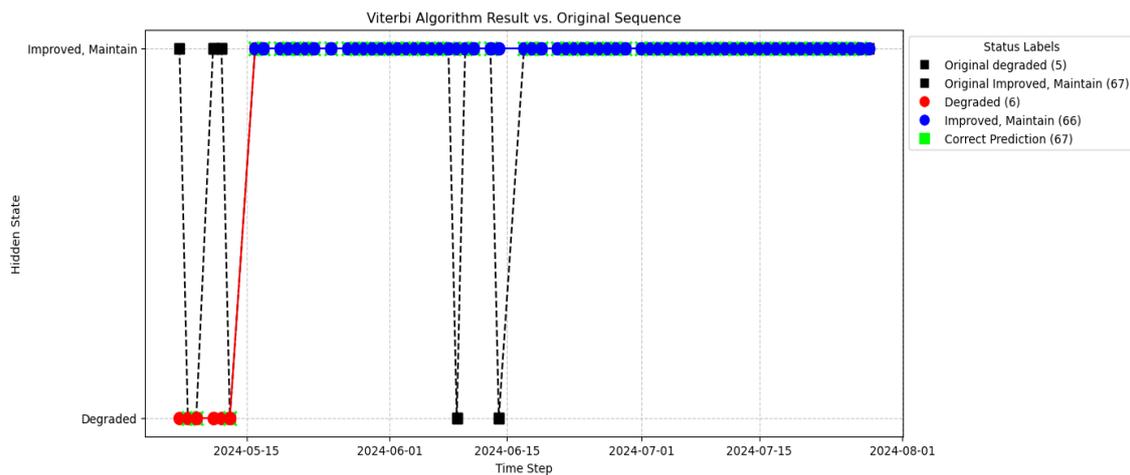
$$\pi = \begin{bmatrix} 0,069444 \\ 0,263889 \\ 0,666667 \end{bmatrix} \quad (11)$$

3. 4 Model prediksi dengan viterbi

Tahap selanjutnya adalah mencari urutan *hidden state* yang optimal menggunakan algoritma Viterbi [14]. Algoritma ini bekerja dengan mengidentifikasi sekuen *hidden state* yang paling mungkin berdasarkan sekuen *observation state*. Pada setiap langkah waktu, algoritma Viterbi meningkatkan nilai secara iteratif dengan menentukan probabilitas maksimum dari setiap jalur yang mencapai keadaan i pada waktu t . Ini dilakukan dengan memastikan urutan observasi yang benar hingga waktu t , sehingga algoritma ini dapat secara dinamis melacak keadaan dengan probabilitas tertinggi pada setiap tahap.

Setelah menyelesaikan urutan, algoritma Viterbi melakukan *backtrack* untuk memilih *state* yang paling dominan pada setiap waktu, membentuk jalur atau sekuen *hidden state* yang paling mungkin menuju sekuen observasi. Data observasi dalam penelitian ini diperoleh dengan menyaring secara acak dataset berdasarkan *siteid*. Dengan menggunakan *observation state* seperti "coverage" dan "interference", algoritma ini mampu memprediksi kondisi jaringan yang tersembunyi dan tidak dapat diamati langsung.

Proses penentuan sekuen *hidden state* yang optimal melibatkan beberapa langkah penting, seperti inialisasi, rekursi, identifikasi jalur terbaik (*best path*) sekuen *hidden state* pada waktu T , dan menetapkan hasil dari jalur terbaik tersebut. Langkah-langkah ini memastikan bahwa analisis yang dilakukan akurat dan dapat diandalkan, memberikan gambaran yang jelas tentang performa jaringan dan membantu dalam pengambilan keputusan untuk peningkatan kualitas jaringan LTE. Gambar 4 adalah barisan prediksi *hidden state* yang terbentuk dari algoritma viterbi. Hasil ini merupakan data sintesis yang akan dijadikan sebagai dasar prediksi analisa peningkatan performa jaringan seluler 4G.

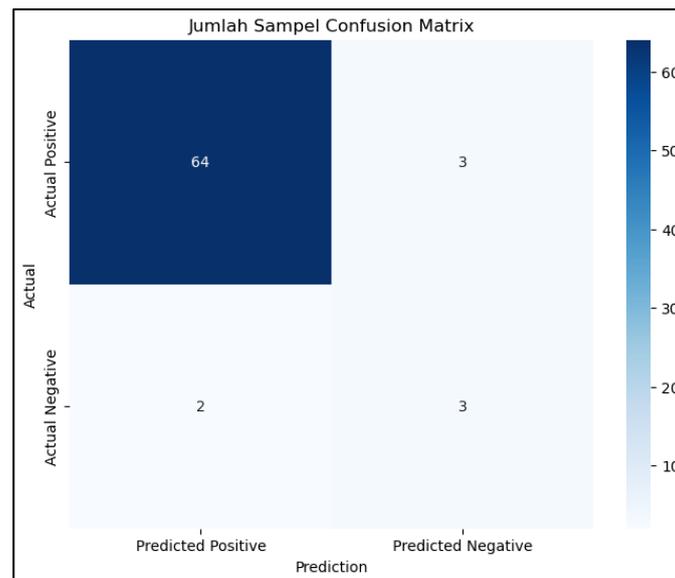


Gambar 4 Barisan prediksi best-path hasil dari algoritma viterbi

3. 4 Evaluasi dan visualisasi

Tahap selanjutnya adalah evaluasi hasil prediksi menggunakan *confusion matrix* [17]. *Confusion matrix* membantu mengukur kinerja model dengan membandingkan prediksi *hidden state* dengan kondisi sebenarnya. Matriks ini memberikan informasi tentang jumlah *true positives*, *false positives*, *true negatives*, dan *false negatives*, yang kemudian digunakan untuk menghitung metrik kinerja seperti akurasi, presisi, *recall*, dan *F1-score*.

Pada penelitian ini, hasil nilai masing-masing dari *true positif (TP)*, *False Positives (FP)*, *True Negatives (TN)* dan *False Negatives (FN)* adalah 64, 2, 3 dan 3, seperti yang ditunjukkan pada gambar 5. Dengan menggunakan persamaan 5,6,7 dan 8 maka dapat diperoleh hasil perhitungan seperti pada tabel 4.



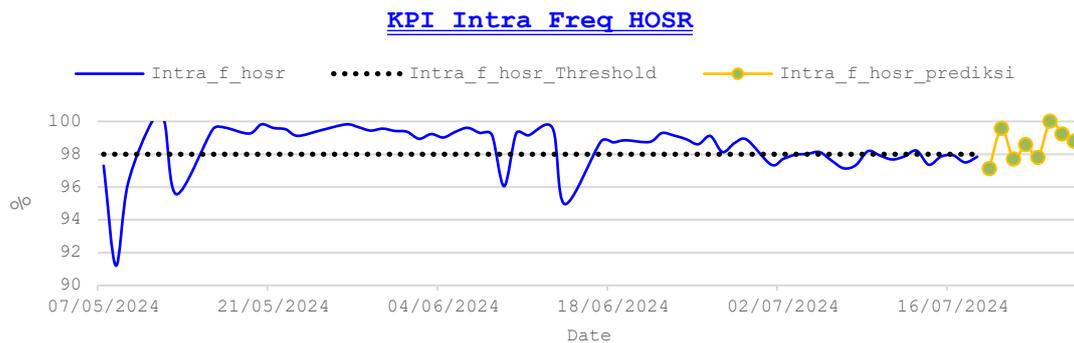
Gambar 5 Jumlah sampel hasil penelitian

Hasil evaluasi ini memberikan gambaran jelas tentang seberapa baik model memprediksi kondisi jaringan, serta area yang perlu ditingkatkan. Visualisasi hasil prediksi membantu dalam memahami pola dan tren data secara lebih intuitif, memfasilitasi pengambilan keputusan yang lebih baik.

Tabel 4 Confusion matrix model prediksi

Metrics	Percentage %
<i>Accuracy</i>	93,06%
<i>Precision</i>	96,97%
<i>Recall</i>	95,52%
<i>F1 Score</i>	96,24%

Visualisasi hasil prediksi membantu memahami pola dan tren data secara lebih intuitif, memfasilitasi pengambilan keputusan yang lebih baik. Gambar 5 menyajikan wawasan dari hasil prediksi yang digunakan untuk memberikan rekomendasi atau membuat keputusan yang tepat dalam meningkatkan kinerja jaringan seluler 4G. Visualisasi ini memungkinkan identifikasi area yang memerlukan perhatian khusus, sehingga tindakan perbaikan dapat diambil dengan lebih efektif.



Gambar 6 Pola dan tren dari data sintetis (gabungan hasil prediksi dan *real data*)

4. KESIMPULAN

Dari hasil yang diperoleh, dapat disarikan beberapa poin penting yang mencerminkan kontribusi dan dampak penelitian ini. Berikut adalah poin-poin yang dapat disimpulkan.

1. Penelitian ini berhasil menerapkan metodologi *Knowledge Discovery in Database* (KDD) yang dikombinasikan dengan algoritma *Hidden Markov Model* (HMM) untuk memprediksi dan meningkatkan kualitas sinyal 4G.
2. Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model yang dikembangkan memiliki akurasi tinggi dengan nilai 93,06%, presisi 96,97%, *recall* 95,52%, dan *F1 score* 96,24%.
3. Model prediksi yang dihasilkan efektif dalam mengidentifikasi indikator kinerja kunci dan tren, sehingga dapat memfasilitasi pengambilan keputusan yang lebih baik untuk optimasi jaringan.
4. Visualisasi hasil prediksi memberikan wawasan yang berguna untuk perencanaan dan peningkatan performa jaringan seluler 4G.

5. SARAN

Penelitian ini memberikan peluang untuk studi lanjutan di masa depan. Berikut adalah beberapa rekomendasi yang perlu dipertimbangkan untuk penelitian berikutnya.

- Untuk meningkatkan performa model prediksi, disarankan untuk terus memperbarui dan menambahkan data *real time* dari berbagai lokasi proyek. Oleh karena itu, penelitian selanjutnya perlu menambahkan objek-objek observasi lainnya yang berpengaruh terhadap peningkatan KPI jaringan seluler.
- Penggunaan teknik *data mining* dan *machine learning* lainnya, seperti *Random Forest*, *Neural Networks*, atau *Support Vector Machines* (SVM), dapat dieksplorasi untuk dibandingkan dengan HMM guna menemukan metode yang paling efektif untuk kasus ini.
- Implementasi model prediksi ini dapat diperluas ke jaringan 5G di masa depan untuk memastikan kualitas layanan yang lebih baik dengan teknologi yang lebih baru.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Nokia Indonesia yang telah menyediakan data dan fasilitas penelitian. Semoga hasil penelitian ini bermanfaat bagi pengembangan ilmu manajemen performa jaringan seluler.

DAFTAR PUSTAKA

- [1] S. Pramono, M. D. Ariyanto, L. Alvionita, and M. E. Sulisty, "Analysis and optimization of 4G long term evolution (LTE) network in urban area with carrier aggregation technique on 1800 MHz and 2100 MHz frequencies," *AIP Conf. Proc.*, vol. 2217, no. April, 2020, doi: 10.1063/5.0000731.
- [2] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc. Sci. Res.*, vol. 110, no. October 2022, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [3] P. Dymarski, *Hidden Markov Model*, Edited by Przemyslaw Dymarski, no. November. 2011. [Online]. Available: <http://www.intechopen.com/books/show/title/hidden-markov-models-theory-and-applications>
- [4] M. Thesis and M. E. Haider, "Machine Learning and KPI Analysis applied to Time-Series Data in Physical Systems: Comparison and Combination," no. May, 2021.
- [5] R. M. M. Santos, "Machine Learning Techniques using Key Performance Indicators for the Configuration Optimization of 4G Networks," *Pdfs.Semanticscholar.Org*, vol. 800, pp. 1–10, 1800, [Online]. Available: <https://pdfs.semanticscholar.org/eda9/e7b8de05d3f5b11fff25bfce8f31b2175325.pdf>
- [6] A. S. Omer, T. A. Yemer, and D. H. Woldegebreal, "Hybrid K-Mean Clustering and Markov Chain for Mobile Network Accessibility and Retainability Prediction," *Eng. Proc.*, vol. 18, no. 1, pp. 1–11, 2022, doi: 10.3390/engproc2022018009.
- [7] Hendrawan, "Accessibility degradation prediction on LTE/SAE network using discrete time markov chain (DTMC) model," *J. ICT Res. Appl.*, vol. 13, no. 1, pp. 1–18, 2019, doi: 10.5614/itbj.ict.res.appl.2019.13.1.1.
- [8] M. Yan, S. Li, C. A. Chan, Y. Shen, and Y. Yu, "Mobility prediction using a weighted markov model based on mobile user classification," *Sensors*, vol. 21, no. 5, pp. 1–20, 2021, doi: 10.3390/s21051740.
- [9] S. Chen *et al.*, "Large Vector Spatial Data Storage and Query Processing Using Clickhouse," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 48, no. M-1–2023, pp. 65–72, 2023, doi: 10.5194/isprs-archives-XLVIII-M-1-2023-65-2023.
- [10] Y. Hasija and R. Chakraborty, *Python for Data Analysis*. 2021. doi: 10.1201/9781003090113-4-4.
- [11] C. Zhang and J. Han, *Data Mining and Knowledge Discovery*. 2021. doi: 10.1007/978-981-15-8983-6_42.
- [12] C. Johnson, L. Serna, T. Novosad, N. Hathiramani, L. Serna, and C. Johnson, "LTE FDD Optimization Guidelines," pp. 1–255, 2022.
- [13] A. T. Jebb, S. Parrigon, and S. E. Woo, "Exploratory data analysis as a foundation of inductive research," *Hum. Resour. Manag. Rev.*, vol. 27, no. 2, pp. 265–276, 2017, doi: 10.1016/j.hrmr.2016.08.003.
- [14] N. G. dan A. Firdaniza, "Hidden Markov Model," *Hidden Markov Model*, vol. 106, no. 99, pp. 43–48, 2006, doi: 978–979–16353–9–1.
- [15] G. Donald Allen and D. Goldsby, "Confusion Theory and Assessment," *IJISSET-International J. Innov. Sci. Eng. Technol.*, vol. 1, no. 10, pp. 436–443, 2014, [Online]. Available: www.ijiset.com
- [16] A. Hagberg, D. Schult, and P. Swart, "NetworkX Reference (Python)," *Python Packag.*, p. 464, 2011.
- [17] A. Ardiansyah, J. Triloka, K. kunci-Pengolahan Citra, D. Kesegaran Buah, and K. Akurasi, "Evaluasi Kinerja Model YOLOv8... 357," pp. 357–368, 2024.