



**Pembuatan Model Klasifikasi Teks Dengan BERT dan GRU
untuk Pengkodean ICD-10 Diagnosis Utama Penyakit pada
Proses Klaim BPJS di RSUP Persahabatan**

Aqge Priwibowo
22917025

*Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer
Konsentrasi Sains Data
Program Studi Informatika Program Magister
Fakultas Teknologi Industri
Universitas Islam Indonesia*

2025

Lembar Pengesahan Pembimbing

Pembuatan Model Klasifikasi Teks Dengan BERT dan GRU untuk Pengkodean ICD-10 Diagnosis Utama Penyakit pada Proses Klaim BPJS di RSUP Persahabatan

Aqge Priwibowo

22917025



Pembimbing Pertama

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

Pembimbing Kedua

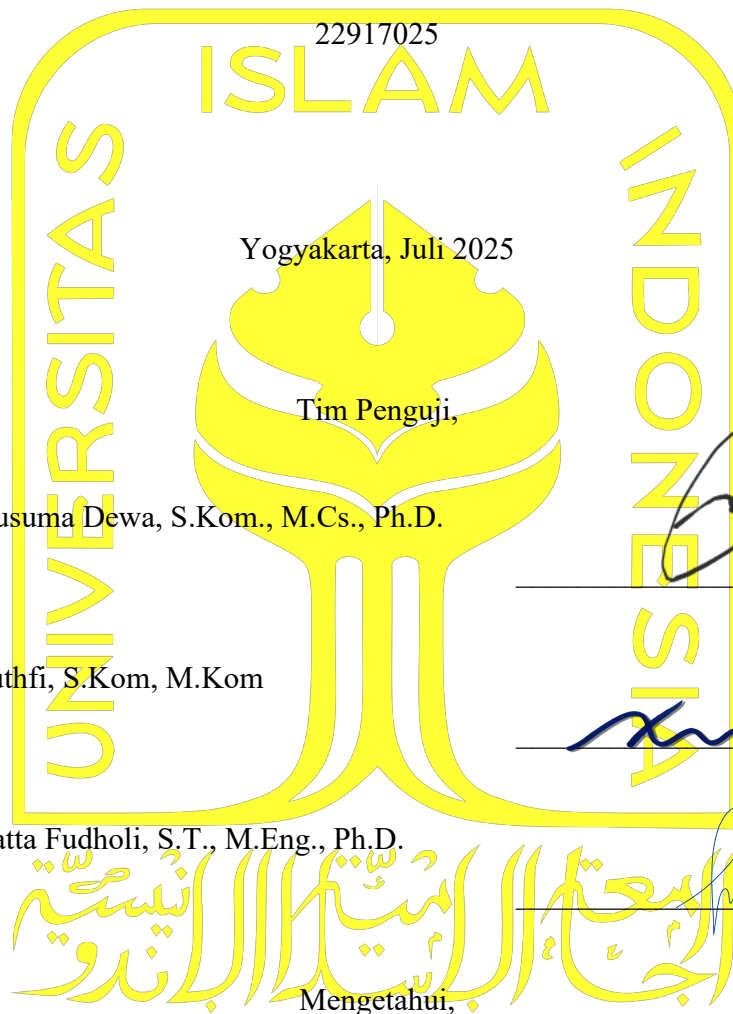
Dr. Ahmad Luthfi, S.Kom., M.Kom

Lembar Pengesahan Penguji

Pembuatan Model Klasifikasi Teks Dengan BERT dan GRU untuk Pengkodean ICD-10 Diagnosis Utama Penyakit pada Proses Klaim BPJS di RSUP Persahabatan

Aqge Priwibowo

22917025



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

Ketua

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

Anggota I

Ir. Dthomas Hatta Fudholi, S.T., M.Eng., Ph.D.

Anggota II

Mengetahui,

Ketua Program Studi Informatika Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia



Ir. Irving Vitra Papatungan, S.T., M.Sc., Ph.D.

Abstrak

Pembuatan Model Klasifikasi Teks Dengan BERT dan GRU untuk Pengkodean ICD-10 Diagnosis Utama Penyakit pada Proses Klaim BPJS di RSUP Persahabatan

Peran rumah sakit tidak hanya pada pelayanan kesehatan, tetapi juga sebagai tulang punggung sistem kesehatan dunia. Kemampuan rumah sakit untuk tetap berkelanjutan secara finansial sangat terkait dengan kemampuannya mengelola klaim asuransi. Pembayaran klaim asuransi BPJS didasarkan pada pengelompokan diagnosis penyakit dan prosedur, dan ini sangat dipengaruhi oleh hasil dari kodefikasi (koding). Ketepatan koding ini, akan menentukan pengelompokan penagihan.

Aspek teknis yang menjadi fokus penelitian ini adalah pada penanganan kegiatan kodefikasi oleh tim rekam medis. Dari sudut pandang sains data, hal ini merupakan implementasi dari teknik klasifikasi teks. Sehingga dapat diangkat rumusan masalah yaitu, bagaimana membangun model untuk melakukan kodefikasi teks diagnosis kedalam kode ICD-10. Model memfokuskan pada 8 kelas teks (label) urutan teratas, yaitu kelompok yang diidentifikasi memiliki performa klasifikasi diatas 50% pada eksperimen awalan yang melibatkan 24 kelas. Sedangkan kelas dengan performa dibawah 50% secara sengaja diberi kode baru sebagai NN. Metode yang dipilih untuk klasifikasi teks ini, adalah dengan menggunakan arsitektur BERT dan BiGRU yang dikenal sangat efektif untuk pengolahan data sequensial. Lapisan awal dari arsitektur ini berupa BERT untuk memperoleh vektor word embedding yang kontekstual. Diatasnya ditambahkan lapisan BiGRU mengekstraksi fitur sepanjang teks dan dapat dikelompokkan sesuai labelnya.

Model yang dikembangkan menunjukkan performa yang baik, *precision* 82,18%, dan *recall* 81,59%, *F1-score* 82,66%, dan akurasi 81,92%, yang menandakan efektivitasnya dalam mengklasifikasikan teks diagnosis. Namun, beberapa hasil masih dipengaruhi oleh ketidakseimbangan kelas, label yang tidak konsisten, serta tumpang tindih fitur pada kelas gabungan. Temuan ini menekankan pentingnya peningkatan kualitas data, serta membuka potensi eksplorasi lanjutan untuk lebih meningkatkan performa model.

Kata kunci

arsitektur berlapis, bert, bigru, diagnosis, icd10, klasifikasi teks, teks medis

Abstract

Development of a Text Classification Model Using BERT and GRU for ICD-10 Main Disease Diagnosis Coding in BPJS Claim Processing at RSUP Persahabatan.

Hospitals play a vital role not only in providing healthcare services but also as a backbone of the global healthcare system. Their financial sustainability is closely tied to their ability to manage insurance claims effectively. In Indonesia, BPJS insurance claim payments are based on the grouping of disease diagnoses and procedures, which is heavily influenced by the accuracy of codification (coding). The precision of this coding directly affects how claims are classified and processed.

This study focuses on the technical aspect of codification handled by the medical records team. From a data science perspective, this process is seen as an implementation of text classification techniques. Therefore, the research question addressed is how to build a model capable of codifying diagnostic texts into ICD-10 codes. The model focuses on the top 8 diagnosis text classes identified to have classification performance above 50% in an initial experiment involving 24 categories. Meanwhile, classes with performance below 50% were deliberately assigned a new label, NN. The method selected for this classification task employs a BERT and BiGRU architecture, which is known to be highly effective for sequential data processing. The initial layer uses BERT to generate contextual word embeddings, followed by a BiGRU layer that extracts features across the text to classify them into the appropriate labels.

The developed model shows strong performance, achieving precision of 82.18%, recall of 81.59%, F1-score of 82.66%, and accuracy 81.92% indicating its effectiveness in classifying diagnostic texts. However, performance was still affected by class imbalance, inconsistent labeling, and feature overlap in merged classes. These findings highlight the importance of improving data quality and suggest further exploration to enhance model performance.

Keywords

bert, bigru, diagnosis, medical text, text classification, icd-10, layered architecture

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, Juli 2025



Aqge Priwibowo, ST

Daftar Publikasi

Publikasi yang menjadi bagian dari tesis

Publikasi berikut menjadi bagian dari Bab 3

Priwibowo, A., Dewa, C.K., & Luthfi, A. (2025), Enhancing Disease Diagnosis Coding: A Deep Learning Approach with Bidirectional GRU For ICD-10 Classification, *JURNAL INFOTEL*.

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Aqge Priwibowo	Mendesain eksperimen (80%) Menulis <i>paper</i> (75%)
Chandra Kusuma Dewa	Mendesain eksperimen (10%) Menulis dan mengedit <i>paper</i> (10%)
Ahmad Luthfi	Mendesain eksperimen (10%) Menulis dan mengedit <i>paper</i> (15%)

Halaman Kontribusi

Penelitian ini merupakan hasil dari kolaborasi berbagai pihak yang berperan dalam berbagai aspek penelitian. Pembimbing, Candra Kusuma Dewa, S.Kom, M.Cs, Ph.D dan Dr. Ahmad Luthfi, S.Kom, M.Kom, memberikan bimbingan serta masukan yang berperan dalam membentuk arah penelitian ini.

Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. berkontribusi dalam penyempurnaan metodologi dengan memberikan kritik konstruktif serta saran yang mendukung validitas model klasifikasi teks diagnosis. Pendekatan dalam menentukan label kelas berdasarkan eksperimen awal menjadi landasan utama dalam pengembangan model ini.

Selain itu, dr. Andi Yussianto, S.H., M.Epid. memberikan wawasan mendalam terkait tantangan dalam pengkodean penyakit di rumah sakit, yang menjadi salah satu aspek penting dalam penelitian ini. Agus Wiyanto, S.Kom. berperan dalam penyediaan dataset serta dukungan teknis yang diperlukan dalam pengolahan data dan pengujian model.

Dengan kontribusi dari berbagai pihak, penelitian ini dapat tersusun secara sistematis dan menghasilkan temuan yang relevan dalam bidang klasifikasi teks diagnosis medis.

Halaman Persembahan

Terima kasih yang sebesar-besarnya kepada orang tua saya, yang selalu menjadi sumber inspirasi dan motivasi dalam setiap langkah perjalanan akademik ini. Dukungan dan doa mereka telah memberikan kekuatan bagi saya untuk terus maju. Juga berterima kasih kepada keluarga saya, yang dengan penuh kesabaran selalu memberikan dorongan serta energi positif selama proses penelitian ini.

Kepada pembimbing saya, yang telah memberikan arahan dalam penelitian ini serta bimbingan yang sangat berharga, terima kasih atas dedikasi mereka dalam membantu menyelesaikan penelitian ini. Terima kasih kepada para dosen, atas masukan berharga dalam memperkaya wawasan dan memperdalam pemahaman terhadap topik penelitian.

Tak lupa, untuk seluruh rekan sejawat, yang telah menjadi teman diskusi dalam menyelesaikan berbagai permasalahan akademik. Pertukaran ide telah membantu saya menghindari kebingungan serta menemukan solusi terbaik dalam penelitian ini. Semoga semua dukungan dan kebaikan yang diberikan mendapatkan balasan yang setimpal.

Kata Pengantar

Alhamdulillah, puji syukur kehadiran Allah SWT, yang senantiasa telah memberikan perlindungan serta melimpahkan rahmat, taufik, dan hidayahnya kepada penulis sehingga bisa menyelesaikan Tesis ini dengan judul “Pembuatan Model Klasifikasi Teks Dengan BERT dan GRU untuk Pengkodean ICD-10 Diagnosis Utama Penyakit pada Proses Klaim BPJS di RSUP Persahabatan”. Tesis ini merupakan sebagai syarat untuk mendapatkan gelar Magister Konsentrasi Sains Data Jurusan Teknik Informatika Program Mesgister Fakultas Teknologi Industri di Universitas Islam Indonesia Yogyakarta.

Dalam proses penyusunan Tesis atau tugas akhir ini tidak lepas dari bantuan banyak pihak, baik itu berupa bimbingan, kritik, saran, dukungan dan motivasi maupun doa dari orang-orang sekitar. Oleh sebab penulis mengucapkan banyak terima kasih kepada pihak-pihak yang membantu dalam penyelesaian tesis ini.

Penelitian ini berfokus pada mengusulkan kode ICD-10 pada diagnosis primer pasien rawat inap, yang merupakan langkah awal dalam pengembangan model klasifikasi teks diagnosis di lingkungan rumah sakit. Sebagai penelitian rintisan di bidang ini, masih terdapat banyak potensi yang dapat dieksplorasi lebih lanjut. Dengan semakin banyaknya data yang tersedia, penelitian di masa depan dapat menghasilkan model yang lebih akurat, lebih luas cakupannya, serta lebih bermanfaat bagi sistem informasi kesehatan.

Saya berharap penelitian ini dapat menjadi dasar bagi penelitian-penelitian selanjutnya yang ingin mengembangkan lebih jauh model klasifikasi diagnosis berbasis teks. Masih banyak ruang untuk inovasi dan penyempurnaan, baik dari segi metode, data, maupun penerapannya dalam dunia medis. Rumah sakit juga terbuka bagi para peneliti yang memiliki komitmen untuk mengeksplorasi lebih jauh bidang ini, guna menghasilkan solusi yang lebih optimal dalam sistem pengkodean diagnosis dan pengelolaan data kesehatan.

Saya menyadari bahwa tesis ini masih jauh dari kesempurnaan. Oleh karena itu, saya terbuka terhadap kritik dan saran yang membangun demi perbaikan lebih lanjut. Semoga penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan serta menjadi referensi bagi penelitian-penelitian selanjutnya.

Akhir kata, saya berharap tesis ini dapat memberikan kontribusi positif, baik bagi dunia akademik maupun praktik di bidang kesehatan, serta menginspirasi para peneliti lain untuk terus mengembangkan penelitian ini ke arah yang lebih luas dan mendalam.

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.....	xiii
Daftar Gambar	xiv
Glosarium	xvi
BAB 1 Pendahuluan	1
1.1 Latar Belakang.....	1
1.1.1 Jaminan Kesehatan Nasional	1
1.1.2 Dampak Kodefikasi Diagnosa terhadap Sistem Rumah Sakit	2
1.1.3 Proses Bisnis.....	5
1.1.4 Kontribusi Penelitian	6
1.2 Rumusan masalah	7
1.3 Tujuan penelitian	8
1.4 Batasan masalah.....	8
1.5 Manfaat Penelitian	9
1.6 Metodologi Penelitian.....	9
1.7 Sistematika Penulisan	11

BAB 2 Tinjauan Pustaka	12
2.1 Pendekatan Tata Kelola	12
2.2 Klasifikasi Teks Medis	13
2.3 BERT dan BiGRU pada Klasifikasi Teks.....	16
2.4 Adaptasi Arsitektur BERT-BiGRU untuk Teks Diagnosis Medis	19
2.5 Landasan Konsep.....	20
2.5.1 Kode ICD - 10	20
2.5.2 Text Preprocessing	21
2.5.3 Bidirectional Encoder Representations from Transformers (BERT)	21
2.5.4 Recurrent Neural Network (RNN)	22
2.5.5 Bidirectional Gated Recurrent Unit (Bi-GRU).....	22
BAB 3 Metodologi	24
3.1 Pendekatan Penelitian dalam Penyelesaian Klaim BPJS.....	24
3.2 Tahapan Penelitian.....	25
3.2.1 Pemahaman bisnis	25
3.2.2 Pengambilan data.....	25
3.2.3 Persiapan dataset.....	26
3.2.4 Preprocessing	27
3.2.5 Pemodelan	27
3.2.6 Evaluasi	29
3.3 Dataset	31
BAB 4 Hasil dan Pembahasan.....	33
4.1 Pemahaman Bisnis	33
4.2 Persiapan Data	34
4.3 Pelabelan Dataset.....	35
4.4 Distribusi Token	39
4.5 Pelatihan Model	40

4.6	Optimisasi Hyperparameter	42
4.7	Vector Embedding	43
4.8	Evaluasi Model	45
4.9	Implikasi dan Kontribusi Penelitian	49
BAB 5 Kesimpulan dan Saran.....		51
5.1	Kesimpulan	51
5.2	Saran	51
Daftar Pustaka		52
LAMPIRAN A		58

Daftar Tabel

Tabel 1.1 Rangkuman literatur tentang hubungan kausal klaim BPJS.....	3
Tabel 2.1 Ringkasan penelitian terkait kalsifikasi teks medis.....	14
Tabel 2.2 Ringkasan studi terkait klasifikasi teks dengan BERT dan BiGRU.....	18
Tabel 4.1 Distribusi frekuensi kode ICD-10 pada dataset awal	36
Tabel 4.2 Performa setiap kelas dari hasil evaluasi pada eksperimen awalan	37
Tabel 4.3 Distribusi frekuensi kelas ICD-10 paska pemetaan label.....	38
Tabel 4.4 Hyperparameter yang digunakan untuk pelatihan model.....	41
Tabel 4.5 Evaluasi performa model menggunakan BERT- <i>base</i> , <i>learning rate</i> 1E-4 dan beberapa variasi <i>drop rate</i> dan <i>n-layer</i>	43
Tabel 4.6 Evaluasi performa model dengan variasi penggunaan <i>vector embedding</i>	45
Tabel 4.7 Evaluasi performa model untuk tiap kelas	48

Daftar Gambar

Gambar 1.1 Ilustrasi model hubungan kausal pada RSUP Persahabatan. Sumber interpretasi literatur dan interview pakar	5
Gambar 1.2 Ilustrasi model dan notasi proses bisnis pada RSUP Persahabatan untuk klaim BPJS. Sumber dari interview pakar pada RSUP Persahabatan.	6
Gambar 2.1 Diagram petal (<i>flower diagram</i>) menunjukkan penelitian terkait sebagaimana diuraikan pada bagian sebelumnya.....	20
Gambar 2.2 Struktur kode ICD-10. Sumber: Sheikh Shams Azam (2020).....	21
Gambar 2.3 Struktur gerbang GRU. Sumber: Ying Tian (2023)	23
Gambar 2.4 Struktur Bi-GRU. Sumber: Ying Tian (2023).	23
Gambar 3.1 Mengatasi koding oleh manusia (3) akan berdampak hingga klaim BPJS (11).	24
Gambar 3.2 Diagram alur penelitian secara umum	25
Gambar 3.3 Aktifitas rinci pada penelitian yang perlu diperhatikan untuk pemahaman lebih detail	26
Gambar 3.4 Diagram alur training model menggunakan training dataset, mengevaluasi model dengan validation dataset, dan mengukur performa model dengan testing dataset.	29
Gambar 3.5 <i>Screen capture</i> ilustrasi dataset dalam bentuk file worksheet.....	32
Gambar 4.1 Ilustrasi data dengan temuan karakter operasi.....	34
Gambar 4.2 Ilustrasi data dengan temuan format json	35
Gambar 4.3 Ilustrasi data dengan temuan data fitur dianggap kosong.....	35
Gambar 4.4 Visualisasi F1-score tiap kelas pada eksperimen awal menunjukkan kelas dengan performa di atas 50%, untuk dijadikan target untuk pelatihan model utama..	37
Gambar 4.5 Distribusi frekuensi kelas ICD-10 paska pemetaan.....	38
Gambar 4.6 Histogram distribusi frekuensi panjang token untuk seluruh dataset.....	39
Gambar 4.7 Distribusi frekuensi panjang token untuk setiap label kelas ICD-10 pada dataset sebelum proses split.....	40
Gambar 4.8 Penurunan nilai loss selama proses pelatihan 300 epoch	42
Gambar 4.9 Komparasi peningkatan performa pada <i>training dataset</i> dan <i>validation dataset</i>	42
Gambar 4.10 Evaluasi performa model menggunakan BERT-base, Indo-BERT, dan BioClinical-BERT	44

Gambar 4.11 Peningkatan performa pada training dan validation dataset. Meliputi precision, recall, dan F1-score	46
Gambar 4.12 <i>Confusion Matrix</i> menunjukkan performa model memprediksi setiap kelas	47

Glosarium

BPJS	- Badan Penyelenggara Jaminan Sosial
BERT	- <i>Bidirectional Encoder Representations from Transformers</i>
BiGRU	- <i>Bidirectional Gated Recurrent Unit</i>
Critical-Path	- Jalur sebab-akibat yang diidentifikasi berdampak besar
Down-Claim	- Klaim dari pihak rumah sakit dengan nilai kurang dari yang seharusnya
INA-CBG's	- <i>Indonesian-Case Based Groups</i>
Klaim	- Tuntutan yang harus dipenuhi oleh BPJS sesuai dengan perjanjian yang telah disepakati
Koleksi	- Kumpulan data
Semantik	- Semantik mengacu pada pemahaman dan representasi makna dari suatu teks atau ujaran, bukan hanya strukturnya
Up-Claim	- Klaim dari pihak rumah sakit dengan nilai melebihi dari yang seharusnya

BAB 1

Pendahuluan

1.1 Latar Belakang

Dalam konteks global maupun lokal, peran rumah sakit tidak hanya terbatas pada pelayanan kesehatan semata, rumah sakit bahkan menjadi tulang punggung sistem kesehatan di seluruh dunia dalam menjaga kesejahteraan masyarakat. Kemampuan rumah sakit untuk tetap berkelanjutan secara finansial sangat terkait dengan kemampuannya dalam mengelola klaim asuransi dengan sukses (Tama & Kusmiyatno, 2019). Klaim pembiayaan seperti BPJS Kesehatan bukan hanya menjadikan rumah sakit tetap mampu menyediakan layanan kesehatan, tetapi juga menguatkannya dalam menjaga kesehatan masyarakat.

Sebagai institusi profesional, rumah sakit dan fasilitas pelayanan kesehatan lainnya, membutuhkan tenaga dokter, perawat dan tenaga ahli kesehatan lainnya, dalam operasionalnya menyediakan pelayanan kesehatan. Pendapatan rumah sakit dalam operasional tersebut bersumber dari masyarakat yang memanfaatkan pelayanan kesehatan yang disediakan rumah sakit itu sendiri (Rifandi & Shofiani, 2019).

Klaim asuransi dianggap sebagai salah satu pendapatan rumah sakit dari pihak asuransi kesehatan untuk menanggung sejumlah biaya pelayanan kesehatan yang diterima (Rifandi & Shofiani, 2019; Sari et al., 2023). Secara umum pendapatan operasional terbesar di rumah sakit diperoleh dari klaim asuransi, baik swasta maupun Pemerintah (Sari et al., 2023).

Penundaan klaim dapat menyebabkan penundaan pembayaran pelayanan kesehatan, mengakibatkan penurunan pendapatan rumah sakit dan dampak negatif pada operasional rumah sakit. Keterlambatan pembayaran klaim BPJS Kesehatan juga mengganggu aliran kas rumah sakit, terutama karena sebagian besar pasien rumah sakit menggunakan BPJS Kesehatan (Nabila et al., 2020).

1.1.1 Jaminan Kesehatan Nasional

Pemerintah Indonesia menggunakan program Jaminan Kesehatan Nasional melalui BPJS Kesehatan untuk meningkatkan kesehatan masyarakat secara nasional (Udin, 2022). Program pemerintah melalui Permenkes nomor 71 tahun 2013 memungkinkan fasilitas kesehatan, rumah sakit, puskesmas hingga klinik di Indonesia dapat bekerja sama dengan

BPJS Kesehatan. Dua tahun setelah dicanangkannya program tersebut di tahun 2014, jumlah peserta bpjs meningkat menjadi 168 512 237 jiwa atau 67% (Pradani et al., 2017).

BPJS Kesehatan melakukan pembayaran kepada fasilitas kesehatan rujukan tingkat lanjutan, seperti rumah sakit, dalam paket pembayaran yang disebut Tarif *Indonesian-Case Based Groups* atau Tarif INA-CBG's. Pembayaran ini didasarkan pada pengelompokan diagnosis penyakit dan prosedur, yang mencakup semua sumber daya rumah sakit yang digunakan dalam pelayanan medis dan nonmedis (Kemkes, 2016).

Rumah sakit yang menjadi mitra BPJS dapat menagihkan biaya pelayanan dengan metode pembayaran retrospektif dan metode pembayaran prospektif (Ariyanti & Gifari, 2019; Kemkes, 2016). Telah menjadi kesepakatan oleh seluruh rumah sakit, BPJS dan pemangku kepentingan lainnya untuk melakukan penerapan model pembayaran prospektif. Penagihan prospektif ini dilakukan dengan memilih dari basis kasus pembiayaan (*case base payment*) yang sebagaimana ditetapkan oleh kementerian kesehatan (Ariyanti & Gifari, 2019; Kemkes, 2016; Rifandi & Shofiani, 2019).

Metode pembiayaan prospektif sangat dipengaruhi oleh hasil dari koding. Kegiatan koding yang dimaksud adalah kodefikasi data resume medis pasien meliputi diagnosis utama, sekunder dan tindakan. Ketepatan kodefikasi ini akan menentukan hasil pengelompokan dalam aplikasi INA-CBG's (Anyaprita et al., 2020; Kemkes, 2016). Hal terpenting dalam kegiatan koding adalah penentuan kodefikasi ICD-10 untuk diagnosis pasien. Ketepatan kode diagnosis ini, dari sudut pandang pembiayaan, yang akan menjadi acuan utama pada besarnya biaya pelayanan kesehatan yang akan ditagihkan (Trisetyawan & Eryani, 2020).

Pengkodean diagnosis dilakukan oleh petugas rekam medis dengan menganalisa tulisan diagnosis dari dokter, dan berkas lainnya (Maimun, n.d.; Trisetyawan & Eryani, 2020). Dengan banyaknya jumlah pasien yang harus di laporkan dalam durasi terbatas, diperlukan sejumlah petugas rekam medis untuk melakukan kegiatan koding untuk seluruh data rekam medis agar kegiatan berjalan tepat waktu.

1.1.2 Dampak Kodefikasi Diagnosa terhadap Sistem Rumah Sakit

Beberapa penelitian telah menyoroti berbagai dampak dari kesalahan pengkodean diagnosis, mulai dari klaim asuransi BPJS kesehatan yang tertunda, kesalahan alokasi sumber daya, hingga perencanaan finansial rumah sakit. Beberapa diantaranya juga mengidentifikasi penyebab dari kesalahan dapat kodefikasi ICD-10 ini.

Tabel 1.1 Rangkuman literatur tentang hubungan kausal klaim BPJS

No.	Sub Tema	Pemikiran Kausal	Pustaka
1	Keterlambatan Berkas Verifikasi BPJS: Apa Akar Masalah dan Solusinya?	Potensi penurunan kualitas pelayanan. Klaim perlu tahap pengumpulan dokumen. Menyebabkan penundaan pembayaran	Estri Aditya Pradani, 2017
2	Pengaruh Pembayaran Asuransi BPJS Terhadap Akuntansi Pendapatan Rumah Sakit	Ada Regulasi yang mengatur mutu pelayanan. Klaim asuransi merupakan pendapatan, sumber arus kas dan biaya operasional dan sebagainya	Muhamad Rifandi, 2019
3	Analisis Persetujuan Klaim BPJS Kesehatan pada Pasien Rawat Inap	Penagihan prospektif melibatkan pemilihan pembayaran berbasis kasus, INA-CBG's, sesuai ketentuan Kementerian Kesehatan.	Muhammad Tijar Gifari, 2019
4	Hubungan Ketepatan Kode Diagnose Dyspepsia Dengan Klaim Bpjs	Klaim Bpjs membutuhkan pengkodean ICD-10 dengan sistem paket INA CBG's. Kesalahan menyebabkan kerugian bagi Rumah sakit	Rendi Trisetyawan, 2020
5	Penyebab Keterlambatan Pelayanan Klaim Asuransi (Bpjs) Di Rumah Sakit	Kekurangan SDM khusus koding, sehinga pengkodean terbengkalai. Ada Banyak data rekam medis yang perlu dikodifikasi. Peran petugas koding adalah mengkode diagnosis sesuai ICD 10.	Nur Maimun, 2020
6	Dampak Keterlambatan Klaim Bpjs Kesehatan Terhadap Mutu Pelayanan Rumah Sakit	Pengajuan klaim dipengaruhi oleh, pengumpulan berkas, pengkodean, pengiriman berkas klaim. Keterlambatan mempengaruhi arus kas dan mutu pelayanan Rumah sakit.	Devi Anyaprita, 2020
7	Analisis Penyebab Pending Klaim Akibat Koding Berkas Rekam Medis	Pending klaim mengakibatkan pemasukan rumah sakit akan menurun dan berdampak pada operasional. Pending klaim BPJS menyebabkan <i>cash flow</i> rumah sakit terganggu.	Salma Firyal Nabila, 2020

8	Analysis Pending Klaim Pembayaran BPSJ	Koding sangat menentukan besarnya biaya yang dibayarkan ke rumah sakit. Pending klaim penyebabnya yaitu ketidaktepatan koding dan ketidaklengkapan berkas	Hendra Rohman, 2021
9	Kasus Sengketa dan Klaim yang Belum Diproses di Rumah Sakit pada BPJS	Kasus klaim tertunda disebabkan oleh ketidaksesuaian kode dengan berkas. Kasus klaim mempengaruhi alokasi anggaran, kebijakan perencanaan, dan kualitas layanan rumah sakit.	Yastori, S.Si, M.Si, 2021
10	Faktor-faktor yang Menyebabkan Tertundanya Klaim di Rumah Sakit	Penyebab pengembalian berkas Klaim antara lain, kesalahan pengkodean, kesalahan input, salah diagnosis, dan rekam medis yang tidak lengkap.	Irwanto Ali Udin, 2022
11	Analisis Pengakuan Pendapatan Klaim Berdasarkan PSAK 23	Salah satu pendapatan terbesar Rumah sakit adalah Klaim Asuransi	Immu Puteri Sari, 2023

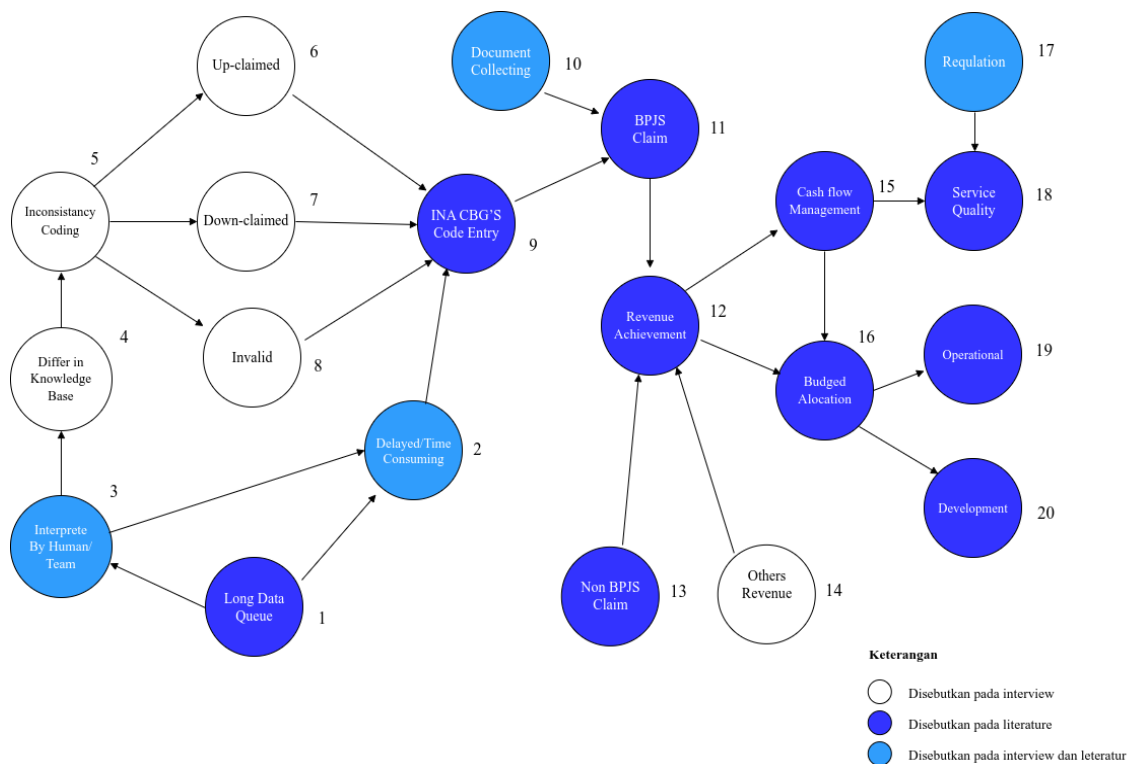
Selain studi literatur Tabel 1.1, dilakukan juga penelusuran informasi melalui *interview* dengan pakar yang kompeten. Hasil *interview* menyebutkan bahwa sepertiga dari masalah penagihan klaim tertunda akibat banyaknya kasus yang harus dikodekan dan dua pertiga masalah lainnya berupa kesalahan dalam kodefikasi oleh tim koder (Yussianto, 2024). Yussianto mengelompokkan kesalahan dalam kodefikasi dibagi menjadi tiga kelompok, *up-claim* sekitar 17% yaitu mengkodekan penagihan lebih tinggi dari yang diterima pasien, *down-claim* sekitar 80% dimana penagihan justru lebih rendah, dan sisanya adalah *miss-claim (invalid claim)* sekitar 3% atau salah dalam mengkodekan.

Dari ketiga kasus tersebut penundaan dan *down-claim* menjadi sorotan utama karena dinilai sangat merugikan pihak rumah sakit. Kerugian yang dimaksud adalah secara finansial, yang kemudian akan berpengaruh pada operasional dan pelayanan rumah sakit. Tetapi tidak mengesampingkan masalah *up-claim* juga bisa berpotensi akan mempengaruhi kepercayaan dari pihak BPJS (Yussianto, 2024).

Tim petugas koding, dalam menginterpretasikan teks diagnosis dan tindakan, memiliki basis pengetahuan yang tidak seragam. meskipun sudah dengan dukungan usaha dari rumah sakit, akan tetapi inilah yang menjadi sumber ketidak konsistenan yang muncul pada koding. Salah satu usaha untuk menseragamkannya adalah dengan melibatkan instalasi

verifikasi yang memverifikasi diagnosis sebelum dilimpahkan pada tim koding (Yussianto, 2024).

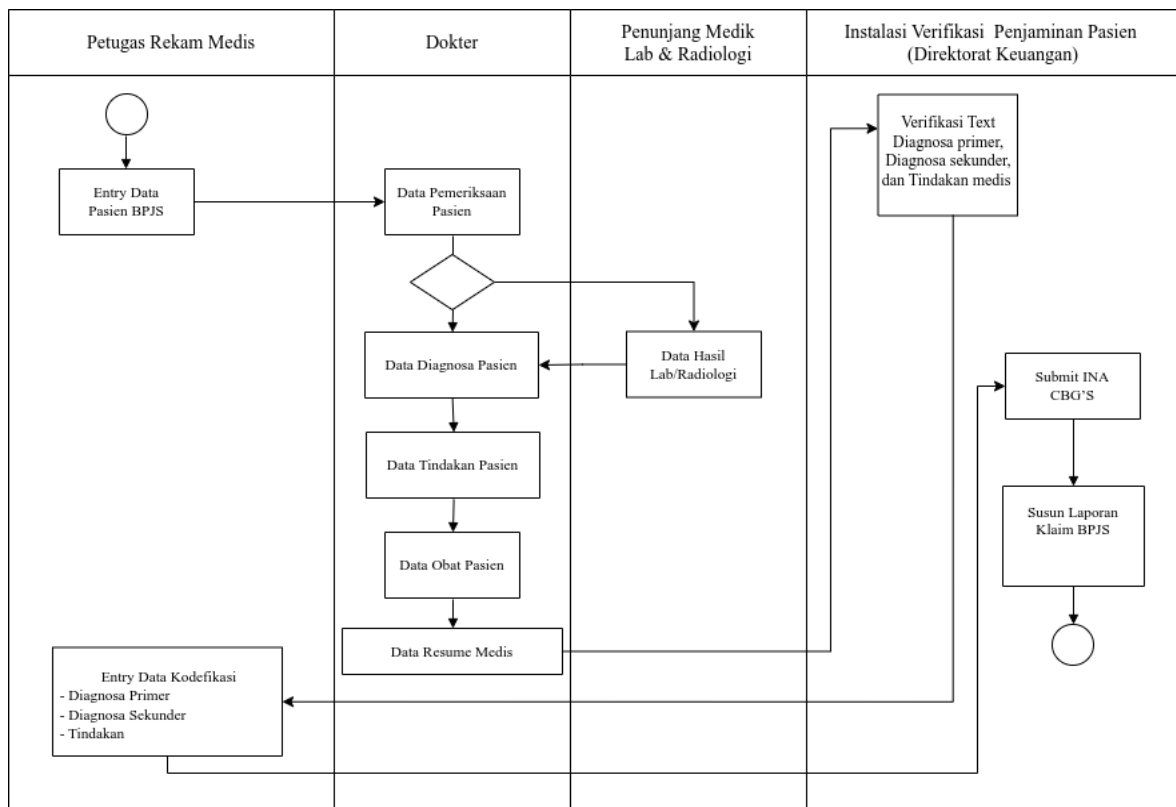
Uraian diatas dapat disusun menjadi sebuah perkiraan hubungan kausal untuk menelusuri akar masalah dalam hal klaim biaya pelayanan kesehatan pada rumah sakit. Ilustrasi perkiraan hubungan kausal yang sejalan dengan permasalahan pada RSUP Persahabatan dapat ditunjukkan sebagaimana Gambar 1.1. Kegiatan kodefikasi oleh tim petugas rekam medis, pada node nomor 3, akan menjadi fokus dari penelitian.



Gambar 1.1 Ilustrasi model hubungan kausal pada RSUP Persahabatan. Sumber interpretasi literatur dan interview pakar

1.1.3 Proses Bisnis

Seperti tampak pada Gambar 1.2 proses bisnis pada rumah sakit, khususnya RSUP Persahabatan, menagihkan klaim pembiayaan dilakukan paska kodifikasi selesai dilakukan. Salah satu upaya untuk menstandarkan redaksional teks diagnosis, baik primer maupun sekunder, adalah dengan melakukan verifikasi diagnosis dan tindakan (*procedure*) sebelum kodefikasi oleh tim koder (Wiyanto, 2023). Dokumen pemeriksaan dikumpulkan sebagai data dukung penagihan Klaim ke BPJS (Pradani et al., 2017; Yussianto, 2023).



Gambar 1.2 Ilustrasi model dan notasi proses bisnis pada RSUP Persahabatan untuk klaim BPJS. Sumber dari interview pakar pada RSUP Persahabatan.

Penagihan ini dapat saja dikembalikan lagi oleh pihak BPJS disebabkan oleh kesalahan atau kurang lengkapnya pengisian berkas rekam medis oleh petugas, dan juga adanya pemahaman yang berbeda pada internal rumah sakit (Udin, 2022; Yastori, 2022). Penyebab penundaan pembayaran yang dapat terjadi adalah kesalahan dalam proses pengkodean, kesalahan dalam proses input, penempatan diagnosis yang salah, dan rekam medis yang tidak lengkap (Udin, 2022). Rohman dan Yastori di tahun 2021 juga menyebutkan hal yang sejenis yaitu penyebab pending klaim disebabkan oleh ketidaktepatan koding dan ketidaklengkapan dokumen.

1.1.4 Kontribusi Penelitian

Tujuan dari penelitian ini adalah membangun model klasifikasi teks kedalam kode ICD-10 dari data rekam medis menggunakan arsitektur *Bidirectional Gated Recurrent Unit (BiGRU)* dan *Bidirectional Encoder Representations from Transformers (BERT)*. Tujuan tersebut dicapai dengan mengintegrasikan konteks proses klaim BPJS ke dalam perumusan

arah penelitian. Hal ini secara langsung mempengaruhi strategi pemilihan label ICD-10 dan fokus pemilihan data diagnosis primer pada metodologi penelitian.

Dengan arahan penekanan pada performa, model yang dikembangkan tetap mempertahankan kelas-kelas ICD-10 dengan performa tinggi sebagai label kemudian untuk menangani kelas dengan performa rendah diterapkan strategi pelabelan kelas “NN” (Not Named) untuk menjaga keterwakilan data tanpa perlu memaksakan pemetaan yang tidak akurat.

Penggunaan teks diagnosis primer yang otentik dari RSUP Persahabatan sebagai dataset merepresentasikan variasi dan gaya bahasa pencatatan klinis yang lebih bebas, informal, dan mencerminkan realitas operasional dalam praktik rumah sakit sehari-hari. Penelitian ini memberikan perspektif terhadap studi klasifikasi medis dengan nuansa lokal dibandingkan dengan penelitian sejenis yang menggunakan dataset berbahasa Inggris, Tiongkok, atau Bengali yang memiliki karakteristik bahasa dan struktur yang berbeda.

Kontribusi penelitian ini dapat dilihat dalam tiga sudut pandang utama. Pertama, secara teoretis, penelitian ini memperluas cakupan studi klasifikasi teks medis dengan menunjukkan efektivitas arsitektur BERT-BiGRU pada teks diagnosis berbahasa Indonesia yang otentik dan tidak terstruktur. Secara metodologis, pendekatan penggabungan label dengan performa rendah (di bawah 50% F1-score) menjadi kelas gabungan menawarkan sebuah strategi dalam menghadapi tantangan ketidakseimbangan label secara khusus dalam implementasi klasifikasi teks diagnosa kedalam kode ICD-10. Sementara itu, secara praktis, model klasifikasi yang dikembangkan memberikan solusi aplikatif untuk mendukung proses koding ICD-10 di rumah sakit, khususnya dalam konteks optimalisasi klaim BPJS melalui pendekatan sains data. Keberadaan model bisa menjadi fondasi awal bagi pengembangan sistem pendukung keputusan yang lebih cerdas dan sesuai dengan konteks operasional rumah sakit di Indonesia.

1.2 Rumusan masalah

Berdasarkan uraian pada bagian latar belakang diketahui bahwa point penting yang dapat menjadi fokus penelitian ini adalah pada penyelesaian kasus kegiatan kodefikasi oleh tim rekam medis. Dari sudut pandang sains data, kegiatan kodefikasi tersebut merupakan implementasi dari teknik klasifikasi teks. Rumusan masalah yang dapat diangkat pada penelitian ini adalah,

- 1 Bagaimana membangun model yang dapat melakukan kodefikasi teks diagnosis, yaitu mengklasifikasikan teks diagnosis kedalam kelas kode ICD-10

2 Apakah model yang dibangun dapat melakukan kodefikasi dengan baik

1.3 Tujuan penelitian

Tujuan dari penelitian ini adalah membangun model klasifikasi teks diagnosis utama menjadi kode ICD-10 pada data rekam medis menggunakan *Bideractional Gated Recurrent Unit* (BiGRU) dan *Bidirectional Encoder Representations from Transformers* (BERT), yang diharapkan dapat menjadi rekomendasi solusi dalam kegiatan coding pada proses penagihan klaim BPJS.

1.4 Batasan masalah

Untuk menjaga ketajaman fokus dan meningkatkan potensi pencapaian performa model yang optimal, penelitian ini dirancang dengan ruang lingkup yang terbatas pada pengembangan model klasifikasi teks terhadap kelompok kelas diagnosis yang menunjukkan potensi hasil klasifikasi yang menjanjikan. Pemilihan ruang lingkup ini didasarkan pada pertimbangan praktis dan masukan dari pihak pakar dari pihak RSUP Persahabatan, yang merekomendasikan agar pengembangan awal difokuskan pada kelas-kelas dengan tingkat keakuratan yang memadai, dibandingkan langsung mencakup keseluruhan kode ICD-10 yang sangat bervariasi. Pendekatan ini juga mengilustrasikan adanya strategi bertahap, di mana proses penelitian dapat dilanjutkan pada tahap berikutnya dengan cakupan kelas yang lebih luas, seiring dengan peningkatan volume dan keragaman data yang tersedia. Untuk pertimbangan tersebut, maka dipilih batasan ruang lingkup sebagai berikut,

- 1 Model klasifikasi teks memfokuskan pada coding/pengkodean diagnosis utama dari data resume medis paska verifikasi ke dalam kode ICD-10
- 2 Kelas teks yang akan menjadi fokus penelitian adalah 8 kelas urutan teratas, yaitu kelompok yang memiliki performa melebihi 50% pada eksperimen awal dengan rekaman data RSUP Persahabatan.
- 3 Meskipun penelitian ini mendapatkan data dan wawasan dari RSUP Persahabatan, pihak rumah sakit tidak terlibat dalam proses pengembangan atau evaluasi teknis model.

Arah pengembangan model memang dipengaruhi oleh masukan dari pihak rumah sakit terkait kebutuhan praktis di lapangan. Namun demikian, penelitian ini dilaksanakan secara independen dan bukan merupakan bagian dari agenda atau program penelitian rumah

sakit. Pengujian atau evaluasi lanjutan terhadap model ini oleh pihak RSUP merupakan kegiatan terpisah di luar ruang lingkup penelitian ini.

1.5 Manfaat Penelitian

Penelitian ini dilakukan dengan harapan dapat memberikan kontribusi nyata bagi pengembangan pengetahuan dan penerapan teknologi di bidang klasifikasi teks medis, khususnya dalam konteks kodefikasi diagnosis di lingkungan rumah sakit di Indonesia. Dengan memanfaatkan pendekatan pembelajaran mesin modern dan data klinis otentik, penelitian ini dirancang untuk menghasilkan pemahaman, metode, dan alat yang relevan dan aplikatif dalam praktik layanan kesehatan nasional. Manfaat yang diharapkan dari penelitian ini antara lain:

- 1 Memberikan landasan awal dalam pemanfaatan arsitektur BERT-BiGRU untuk klasifikasi teks diagnosis medis berbahasa Indonesia, yang masih jarang dieksplorasi dalam penelitian terdahulu.
- 2 Menyediakan pendekatan berbasis data lokal yang otentik, yang mencerminkan karakteristik bahasa pencatatan klinis di rumah sakit Indonesia, sebagai alternatif dari studi sebelumnya yang umumnya menggunakan data asing atau sumber daring.
- 3 Mengusulkan strategi penanganan label dengan performa rendah, melalui penggabungan kelas untuk meningkatkan distribusi data dan mempertahankan stabilitas performa model, yang dapat direplikasi dalam studi klasifikasi serupa.
- 4 Menghasilkan model klasifikasi yang dapat digunakan sebagai alat bantu dalam proses koding ICD-10, sehingga mendukung akurasi dan efisiensi proses klaim BPJS di rumah sakit.
- 5 Memberikan pijakan bagi pengembangan sistem pendukung keputusan berbasis kecerdasan buatan, yang dapat diintegrasikan ke dalam sistem informasi rumah sakit di masa depan untuk membantu proses klasifikasi diagnosis secara otomatis.

1.6 Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model klasifikasi teks diagnosis medis ke dalam kodefikasi ICD-10. Proses metodologis yang diterapkan bertujuan untuk menghasilkan model yang dapat digunakan sebagai alat bantu dalam proses klasifikasi otomatis, khususnya untuk mendukung keperluan klaim BPJS di lingkungan rumah sakit.

Langkah-langkah utama dalam metodologi penelitian ini meliputi:

1 Pemahaman Masalah dan Konteks Bisnis

Penelitian ini dilatarbelakangi oleh kebutuhan rumah sakit untuk meningkatkan efisiensi proses kodifikasi diagnosis dalam rangka pengajuan klaim ke BPJS. Proses manual yang selama ini dilakukan memiliki risiko keterlambatan, ketidakkonsistenan, dan potensi kesalahan klasifikasi yang berdampak pada penolakan klaim.

2 Pengumpulan Data Diagnosis

Data yang digunakan berasal dari catatan diagnosis pasien di sebuah rumah sakit rujukan. Dataset ini mencerminkan kondisi nyata, termasuk penggunaan bahasa Indonesia sehari-hari dengan istilah medis yang tidak selalu baku atau terstruktur.

3 Pra-pemrosesan Teks

Tahapan ini mencakup pembersihan teks diagnosis, normalisasi karakter, tokenisasi, serta transformasi teks menjadi representasi vektor menggunakan pretrained embedding dari model BERT.

4 Perancangan Arsitektur Model

Penelitian ini mengusulkan arsitektur gabungan antara BERT dan BiGRU (Bidirectional Gated Recurrent Unit) untuk menangkap konteks semantik dari teks diagnosis secara mendalam. Model ini dirancang agar mampu memahami arah konteks dari dua sisi (forward dan backward) serta mengoptimalkan pemrosesan urutan kata dalam data medis.

5 Strategi Penggabungan Kelas Minoritas

Untuk mengatasi masalah ketidakseimbangan distribusi kelas, kelas-kelas dengan performa rendah atau jumlah sampel terbatas digabungkan ke dalam kelas baru. Strategi ini bertujuan untuk memperkaya fitur dan mempertahankan performa model secara keseluruhan.

6 Evaluasi Kinerja Model

Model yang dikembangkan dievaluasi menggunakan metrik klasifikasi seperti precision, recall, dan F1-score. Penilaian dilakukan pada beberapa konfigurasi model dan distribusi kelas untuk memperoleh pendekatan terbaik dalam klasifikasi ICD-10 berbasis teks diagnosis.

Metodologi ini secara keseluruhan dirancang untuk merepresentasikan pendekatan yang praktis, adaptif, dan dapat direplikasi untuk kasus serupa di institusi layanan kesehatan lainnya.

1.7 Sistematika Penulisan

Penelitian ini menggunakan sistematika penulisan yang digambarkan sebagai berikut:

BAB I PENDAHULUAN

Pada pendahuluan menguraikan latar belakang masalah terkait pentingnya klasifikasi teks diagnosis medis dalam mendukung proses kodefikasi ICD-10 untuk klaim BPJS. Bab ini juga mencakup rumusan masalah, tujuan, batasan, serta kontribusi dan manfaat penelitian dalam konteks pengembangan model klasifikasi berbasis arsitektur BERT-BiGRU dengan data diagnosis otentik dari rumah sakit.

BAB II TINJAUAN PUSTAKA

Pada tinjauan pustaka dibahas literatur yang relevan mengenai klasifikasi teks, penggunaan model berbasis deep learning (seperti BERT dan BiGRU), serta studi-studi sebelumnya yang menggunakan data medis dalam tugas klasifikasi ICD-10. Bagian ini juga menyajikan identifikasi celah riset dan posisi kebaruan penelitian ini dalam lanskap penelitian terkait.

BAB III METODOLOGI PENELITIAN

Pada metodologi penelitian dijelaskan tahapan-tahapan teknis dalam membangun model klasifikasi, mulai dari pengumpulan dan pembersihan data diagnosis, tokenisasi dan embedding teks, perancangan arsitektur model BERT-BiGRU, hingga strategi penggabungan kelas untuk meningkatkan performa klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan disajikan evaluasi kinerja model terhadap data uji, termasuk metrik seperti precision, recall, dan F1-score untuk masing-masing label. Analisis juga dilakukan terhadap kelas-kelas dengan performa tinggi serta dampak dari strategi penggabungan kelas terhadap kualitas klasifikasi secara keseluruhan.

BAB V KESIMPULAN DAN SARAN

Pada kesimpulan dan saran dirangkum temuan utama dari penelitian, seperti efektivitas arsitektur BERT-BiGRU dalam klasifikasi teks diagnosis medis berbahasa Indonesia dan potensi pemanfaatannya dalam proses kodefikasi ICD-10. Bab ini juga memberikan saran terkait pengembangan sistem lanjutan dan potensi integrasi ke dalam sistem informasi rumah sakit.

BAB 2

Tinjauan Pustaka

2.1 Pendekatan Tata Kelola

Beberapa penelitian telah mengusulkan pendekatan manajerial untuk mengatasi keterlambatan klaim BPJS, khususnya dengan meningkatkan sumber daya manusia dan efisiensi administrasi (Maimun, n.d.; Nabila et al., 2020). Nabila dan Maimun menyarankan penambahan jumlah petugas coding guna menangani volume klaim yang semakin meningkat secara lebih efektif. Solusi yang ditawarkan Nabila dan Maimun sejalan dengan penelitian lain (Rohman et al., 2021), menekankan pentingnya pengembangan dan pelatihan berkelanjutan bagi petugas coding agar proses klaim dapat berjalan lebih optimal. Selain itu, pendekatan manajerial lainnya mencakup pembuatan deskripsi kerja yang jelas bagi petugas coding untuk memastikan pembagian tugas yang lebih terstruktur dan menghindari tumpang tindih pekerjaan.

Selain peningkatan sumber daya manusia, beberapa studi juga menyoroti pentingnya perbaikan administrasi dan pemanfaatan teknologi (Rohman et al., 2021; Sakinah & Yunengsih, 2021). Sakinah dan Rohman menggarisbawahi perlunya peningkatan kelengkapan dan keterbacaan resume medis serta penulisan diagnosis yang lebih jelas, karena kesalahan dalam dokumentasi sering kali menjadi penyebab utama penolakan klaim. Sementara itu, beberapa penelitian juga mengusulkan pengembangan teknologi untuk meningkatkan efektivitas dan efisiensi kerja petugas coding, sehingga dapat mempercepat proses verifikasi dan pengajuan klaim BPJS.

Penelitian ini berupaya mengeksplorasi pendekatan berbasis *machine learning* sebagai alternatif solusi teknologi untuk membantu proses coding ICD-10 secara lebih akurat dan cepat. Dengan mengotomatisasi klasifikasi diagnosis, teknologi ini diharapkan dapat mengurangi kesalahan coding yang sering menjadi penyebab keterlambatan atau penolakan klaim. Pendekatan ini tidak dimaksudkan untuk menggantikan strategi manajerial yang telah ada, tetapi lebih kepada melengkapinya agar dapat meningkatkan tingkat keberhasilan klaim serta mendukung keberlanjutan finansial rumah sakit.

2.2 Klasifikasi Teks Medis

Sejumlah penelitian mengenai klasifikasi teks telah banyak dilakukan dalam kurun waktu 10 tahun terakhir ini. Tantangan utama dalam klasifikasi teks medis adalah mengekstraksi informasi dari deskripsi medis yang tidak terstruktur, terutama dalam bahasa medis yang khusus (Parjono & Kusumadewi, 2023). Parjono menerapkan *Clinical Text Mining* dengan *Natural Language Processing* untuk mengatasi tantangan tersebut. Menggunakan data dari Rumah Sakit bagian poliklinik penyakit dalam sebanyak 3787 data dengan 5 kategori kode ICD-10 yang berbeda, proses ekstraksi data melibatkan beberapa tahapan, seperti *case folding*, kontraksi, tokenisasi, dan penggunaan *word2vec* untuk konversi kata ke dalam vektor. Hasil pemodelan dengan neural network menunjukkan akurasi model sebesar 86.8%.

Guo et al. (2019) Memanfaatkan beberapa bobot yang berbeda sekaligus untuk kata pada tahap *word embedding* sesuai dengan kelas dokumennya, dan menggunakan TextCNN untuk proses klasifikasi teks. Hasilnya menunjukkan bahwa akurasi klasifikasi metode yang diusulkan melebihi metode lainnya secara signifikan di atas CNN *static* dan CNN dengan menggunakan pembobotan tunggal. Lebih lanjut yang dilakukan Md. Aslam Parwez (2023) adalah menyusun data teks yang representatif pada domain *biomedical* berdasarkan distribusi, relasi dan sintak. Kemudian mengimplementasikan klasifikasi teks menggunakan metode *deep learning* dan mengevaluasi perbandingannya. Model BiLSTM dan LSTM menunjukkan performa yang lebih baik dari pada CNN dan CNN-LSTM.

Penelitian lain menggunakan *Dense Connected CNN* untuk mendapatkan kemampuan fitur *multi-scale attention*. Model mengambil output dari seluruh *layer* di atasnya sebagai input dan menghasilkan *feature map* melalui mekanisme filter untuk layernya. Masalah ukuran teks, perbedaan jumlah kata diatasi dengan menggunakan *multi-scale attention*. Arsitektur ini memberikan hasil dengan tingkat akurasi paling tinggi dibandingkan dengan LSTM, *Self attention* dan model CNN lainnya (Wang et al., 2018).

Klasifikasi biner sendiri sudah lama dimulai untuk mengklasifikasikan dokumen klinis dan artikel riset, dibangun dengan menggunakan CNN dan menerapkan beberapa jenis filter. Dengan ide dasar menemukan *convolution filters* (CFs) yang tepat untuk mengekstrak fitur yang bermanfaat. Model CNN yang memanfaatkan *pre-trained vector* memberi hasil yang lebih baik daripada penyatuan CNN untuk mengklasifikasikan kelas yang sering dijumpai. Akan tetapi untuk kelas yang lebih jarang, penggabungan CNN lebih baik (Rios & Kavuluru, 2015).

Disisi lain, penelitian yang lebih spesifik pada domain medis untuk mengklasifikasikan teks diagnosis pada kode ICD telah dikerjakan oleh Francisco Duarte di

tahun 2018 untuk mengklasifikasikan penyebab kematian pada kode ICD-10, dikombinasi *word embedding* dari teks sertifikat pada susunan hierarki sejumlah *Bidirectional-GRU* dan melanjutkan dengan *neural attention*. Klasifikasi ini juga mempertimbangkan pengaruh dari multi output untuk meningkatkan keakuratan hasilnya. Dari berbagai variasi didapatkan hasil menggabung ketiga tahap tersebut memberikan hasil paling baik untuk klasifikasi ICD dari level *chapter*, level *block*, hingga level seluruh kode.

Kode ICD ditandai dengan hanya menggunakan deskripsi dari diagnosis nya saja, tanpa melibatkan informasi tambahan seperti resep dan sejenisnya. Dengan menggunakan distribusi *word embedding (word2vec)* untuk memaksimalkan makna kata sesuai urutannya. Klasifikasi Tidak hanya mempertimbangkan aspek semantik kata dalam teks, tetapi juga memisahkan proses klasifikasi secara bertingkat untuk meningkatkan akurasi. Merumuskan proses secara berjenjang (hierarki), dengan membatasi jumlah kelas pada tiap level dengan mentransformasi proses klasifikasi hingga ke tingkat karakter (Azam et al., 2020).

Hasil yang tidak begitu baik pada pengkategorian diagnosis medis ke dalam kelompok ICD-10 dengan melibatkan *text-preprocessing* dan melakukan multi-class classification dengan *Naive Bayes*, *Support Vector Machine* dan *Logistic regression* (Amin et al., 2021). *Support vector machine* dan *logistic regression* memiliki tingkat akurasi 52% sedangkan *naive bayes* hanya 49% .

Tabel 2.1 Ringkasan penelitian terkait kalsifikasi teks medis

No	Pustaka	Tujuan	Metode	Data	Hasil
1	Anthony Rios, 2015	Klasifikasi teks biner	CNN + <i>pretrained vector</i>	Artikel biomedical (Inggris)	CNN memberikan F1-score 3% lebih baik dari logistic log dan naïve bias
2	Francisco Duarte, 2018	Kodefikasi ICD-10 penyebab kematian pada teks sertifikat kematian	Mengkombinasikan <i>word embedding</i> , GRU dan <i>neural attention</i> .	Koleksi sertifikat kematian (Spanyol)	Akurasi ICD dari level <i>chapter</i> 89%, level <i>block</i> 81%, level seluruh kode 76%

3	Shiyao Wang, 2018	Klasifikasi Teks	Menggunakan <i>Dense Connected</i> CNN dan multi-scale attention.	Koleksi rekam medis (Inggris)	Arsitektur <i>Dense CNN + multi-scale attention</i> ini memberikan tingkat akurasi paling tinggi hingga 83%.
4	Bao Guo, 2019	Mengukur pemanfaatan beberapa bobot untuk kata pada klasifikasi teks	Menggunakan TextCNN dengan <i>word embedding</i> .	Koleksi rekam medis (Inggris)	Menunjukkan akurasi 86.6% melebihi metode lainnya
5	Sheikh Shams Azam, 2020	Pemetaan otomatis kode ICD-10 dari dokumen klinis	LSTM dengan arsitektur bertingkat	Koleksi dokumen klinis (Inggris)	LSTM dengan hirarki berjenjang mencapai F1score 88.05%
6	Zulius Akbar Amin, 2021	Klasifikasi teks multi-kelas	<i>Naive Bayes</i> , <i>Support Vector Machine</i> dan <i>Logistic regression</i>	Rekam Medis (Indonesia)	Akurasi 52% sedangkan naive bayes hanya 49%
7	Md. Aslam Parwez, 2023	Klasifikasi teks	berbagai metode <i>deep learning</i>	Koleksi artikel medis (Inggris)	BiLSTM dan LSTM menunjukkan akurasi 88.94% yang lebih baik dari pada CNN dan CNN-LSTM
8	Parjono dan Kusumadewi, 2023	Klasifikasi teks medis kedalam kategori ICD-10	Menerapkan <i>Clinical Text Mining</i> dengan ANN	Rekam Medis (Indonesia)	Pemodelan dengan <i>Neural Network</i> menunjukkan akurasi model sebesar 86.8%.

Dari berbagai penelitian tersebut, pada Tabel 1.1, terlihat bahwa pendekatan *deep learning* memberikan hasil yang lebih unggul dibandingkan metode konvensional dalam

klasifikasi teks medis, terutama dalam pengkodean ICD-10. Model seperti CNN, LSTM, BiLSTM, dan GRU telah terbukti mampu menangkap pola bahasa yang kompleks serta meningkatkan akurasi prediksi dibandingkan metode berbasis aturan atau model pembelajaran mesin klasik seperti *Naïve Bayes* dan SVM. Sejalan dengan temuan tersebut, penelitian ini akan melanjutkan eksplorasi dengan penerapan *deep learning* pada klasifikasi teks medis, dengan harapan dapat menghasilkan model yang lebih akurat dan dapat diimplementasikan secara praktis.

2.3 BERT dan BiGRU pada Klasifikasi Teks

Klasifikasi teks pada bidang medis merupakan topik dengan pemanfaatan yang paling banyak baik jumlah dan variasinya. Li et al. pada tahun 2023 menggunakan kanal ganda untuk mengekstrak informasi dari teks. Secara khusus menggunakan *word2vec* CBOV dengan melakukan proses training pada dataset webMedQA dan CMDDD. Vektor diproses melalui 2 jalur layer yang berbeda, *convolution* untuk mendapatkan informasi butiran lokal seperti gejala, nomenklatur, dan divisinya. Dan diproses dengan jalur lain layer *BiGRU-Attention* untuk mendapatkan informasi lebih global di sepanjang kalimat. Menggabungkan kedua output jalur layer untuk diproses dengan *classification softmax* untuk menentukan klasifikasi teks-nya. Li mengklaim telah mengatasi ekstraksi fitur dengan akurasi yang rendah, dan mampu memberi rekomendasi divisi (poliklinik) dengan efektif.

Modifikasi GRU secara selektif memilih data *sequence* kemudian melewatkannya (*skip*) oleh Huang et al. (2023) pada penerapan SA-SGRU untuk mengklasifikasikan teks. Eksperimen ini melibatkan tiga dataset publik. Model ini menggunakan *Skip-GRU* untuk melewati konten yang tidak relevan dan menangkap informasi global yang efektif dari teks. Untuk mengatasi limitasi dalam distribusi bobot pada data teks yang dalam (panjang), digunakan mekanisme *self-attention*. Pada jalur lainnya, terdapat layer CNN yang dioptimalkan untuk mengekstrak fitur lokal dari teks. Kedua vektor dari *Skip-GRU* dan CNN tersebut disatukan sebagai input vektor untuk layer klasifikasi. Akhirnya, *softmax* digunakan untuk mengklasifikasikan teks.

Pola klasifikasi pada teks medis dengan memanfaatkan dua jalur kanal secara paralel telah dilakukan sebelumnya pada tahun 2021 oleh S. K. Prabhakar dan D. Won. Prabhakar dan Won menggunakan beberapa model *convolution* dan *BiGRU* sebagai dua kelompok jalur untuk memproses dataset Hallmarks dan AIM. Arsitektur tersebut memberikan hasil yang baik dengan akurasi sebesar 95.76%.

Layer pertama merupakan *word embedding layer* yang mengkonversi teks menjadi vektor yang representatif terhadap konteks semantiknya. Layer CNN mengekstraksi fitur lokal dari teks. Beberapa segmen *convolution* digunakan pada layer ini, akan tetapi CNN mampu enable parallel computing sehingga tidak menambah durasi training. Keluaran digabungkan dengan maksimal pooling dan average pooling untuk mengambil fitur informasinya. Jalur BiGRU digunakan untuk mengekstraksi informasi *long-term* di sepanjang urutan teks. Multihead attention melanjutkan output dari BiGRU dan dapat memfokuskan pada bagian yang berbeda-beda. Bobot yang dihasilkan merupakan gabungan yang representatif dari semua head dengan menggunakan maksimal *pooling* dan *average pooling*. *Softmax classifier* sebagai lapisan terakhir yang melakukan klasifikasi. Arsitektur jalur dua kanal ini membutuhkan sejumlah segmen *convolution* dan arsitektur yang berlapis-lapis. Perlu diperhatikan semakin kompleks arsitekturnya, maka membutuhkan sumber daya komputasional yang semakin tinggi (C. Chen et al., 2020).

Ide klasifikasi teks yang lebih sederhana tidak sebagai segmented jalur kanal-kanal, tetapi hanya berbentuk *stacked layer* (tumpukan berlapis). Dengan penggunaan *word Embedding* yang lebih baik seperti BERT, BiGRU mampu bekerja lebih baik. Dengan Mengkombinasikan BERT model dan BiGRU menghasilkan kombinasi yang powerfull untuk menyelesaikan masalah *text classification* dengan performa yang sangat baik pada klasifikasi teks dalam bahasa china (Yu et al., 2021).

Qing Yu (2021) menunjukkan arsitektur sederhana dengan menambahkan *embedding vector* BERT, yang dapat merepresentasi kata yang menyesuaikan konteks informasinya dengan lebih baik dengan capaian F1-skore diatas 90%. Bidirectional GRU terhubung pada keluaran layer BERT memastikan sebelum proses ekstraksi informasi, vektor merupakan representasi semantik dari kata. BiGRU Sebagai layer yang berfungsi untuk mengekstraksi features dari dua arah secara simultan dari teks menggunakan word vector BERT. Dan untuk menyelesaikan klasifikasi teks maka ditambahkan layer *softmax*. Arsitektur ini menunjukkan akurasi hingga 95% pada berbagai domain teks.

Pola sederhana ini juga diterapkan pada klasifikasi teks medis lainnya. Pada Teks medis *word embedding vector* BioBERT memiliki *out-of-vocabulary* lebih kecil (sedikit) dibandingkan BERT biasa. Rancangan model BGA dan B2GA (BERT-GRU-Attention), untuk klasifikasi teks medis, keduanya menggunakan BioBERT untuk memahami makna kata dalam teks medis. Mencapai peningkatan akurasi hingga 38% dibandingkan dengan BERT biasa (W. Chen et al., 2022). Model lain yang dikenal dengan BiGRU-MA, terdiri

dari Layer BiGRU dan *self-Attention*, juga mampu mengekstraksi fitur semantik untuk tujuan klasifikasi teks dengan baik (Jiang & Wang, 2022).

Menggunakan BERT dan GRU telah terbukti efektif baik dalam pendekatan *multi-channel* maupun *stacked layer*. Meskipun *multi-channel* mampu menangkap fitur lokal dan global dengan lebih baik, arsitekturnya yang kompleks memerlukan biaya komputasi yang lebih tinggi. Sebaliknya, pendekatan *stacked layer* menawarkan solusi yang lebih sederhana dengan efisiensi komputasi yang lebih baik tanpa mengorbankan akurasi secara signifikan. Oleh karena itu, penelitian ini akan berfokus pada pendekatan *stacked layer*, karena arsitektur yang lebih ringkas dan biaya komputasi yang lebih rendah menjadikannya pilihan yang lebih praktis untuk klasifikasi teks medis.

Tabel 2.2 Ringkasan studi terkait klasifikasi teks dengan BERT dan BiGRU

No	Pustaka	Tujuan	Metode	Data	Hasil
1	Qing Yu, 2021	Klasifikasi teks	Menggunakan BERT-BiGRU	Artikel berita (Mandarin)	Akurasi hingga 95%
2	S. K Prabhakar, 2021.	Klasifikasi pada teks medis	Memfaatkan dua jalur kanal. CNN dan BiGRU. Softmax classifier	Abstraksi artikel medis (Inggris)	akurasi sebesar 95.76%.
3	TianTian Jiang, 2022	Klasifikasi teks medis	Lapisan BiGRU-MA, <i>bidirectional</i> GRU and <i>self-attention mechanism</i>	Artikel teknologi (Mandarin)	Menunjukkan hasil yang kompetitif
4	Weidong Chen, 2022	Klasifikasi teks medis	Menggunakan BioBERT-GRU- <i>Attention</i>	Artikel medis (Inggris)	Peningkatan akurasi hingga 38%
5	Xiaoli Li, 2023	Rekomendasi klinik tujuan	Menggunakan kanal ganda CNN dan BiGRU-attention.	Koleksi webMedQA (Mandarin)	Memperbaiki akurasi, recall and F1-score sebesar 10.65%, 8.94% and 11.62%

6	Yuan Huang, 2023	Klasifikasi Teks dan Sentiment	Menggunakan Kanal ganda CNN dan Skip-GRU.	Koleksi Teks review	Memiliki performa yang lebih unggul pada 3 dataset
---	------------------	--------------------------------	---	---------------------	--

2.4 Adaptasi Arsitektur BERT-BiGRU untuk Teks Diagnosis Medis

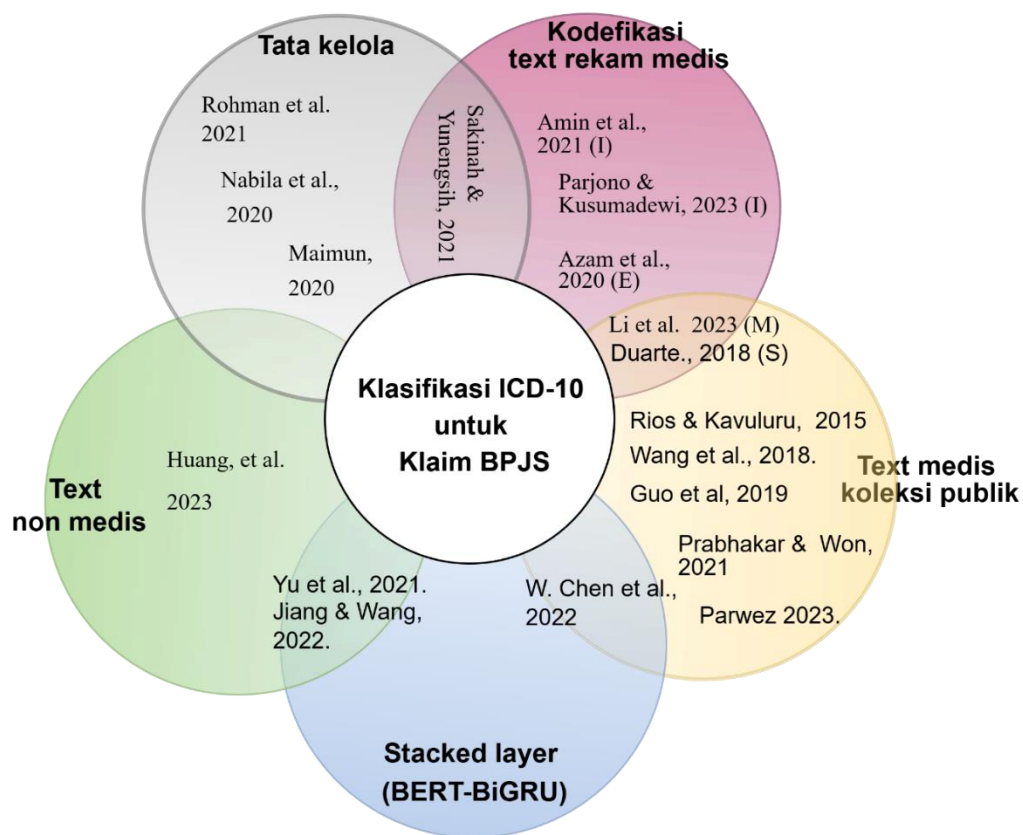
Diagram petal pada Gambar 2.1 menunjukkan ilustrasi sejumlah penelitian sebelumnya yang berhubungan dengan penelitian yang dilakukan saat ini. Diagram menunjukkan penelitian terkait penanganan masalah klaim BPJS dan juga sejumlah penelitian dengan beberapa metode yang fokus pada klasifikasi teks, termasuk didalamnya teks medis dan non medis. Penelitian tentang klasifikasi teks rekam medis sebelumnya sebagaimana terangkum dalam Tabel 2.1 diketahui menggunakan data berbahasa asing seperti Inggris, Spanyol, atau Mandarin, dan sering kali berasal dari koleksi teks daring. Penelitian tersebut memanfaatkan arsitektur LSTM, CNN dan GRU. Banyak dari studi tersebut melaporkan akurasi di atas 90%.

Sementara itu, studi yang menggunakan data asli dari rumah sakit ditemukan bersumber dari departemen klinis tertentu dengan variasi label yang sempit dan terbatas. Penelitian ini menggunakan teknik *ANN fully connected* dan mampu menunjukkan akurasi di atas 80%, untuk 5 kelas terpilih. Penelitian lain yang memanfaatkan data sejenis dilakukan menggunakan teknik *Naive bayes* dan SVM melakukan klasifikasi 22 kelas ICD-10 kode X00 dengan akurasi 52%.

Beberapa penelitian sebelumnya yang terangkum dalam Tabel 2.2 mengeksplorasi penggunaan arsitektur BiGRU, baik secara tunggal, berlapis (*stacked*), maupun dalam bentuk kanal ganda (*dual channel*) bersama dengan CNN, untuk klasifikasi teks non medis seperti review produk dan artikel berita. Pendekatan-pendekatan ini umumnya menunjukkan hasil dengan performa tinggi untuk model klasifikasi teks. Penggabungan BiGRU dengan dengan *word embedding* atau *pretrained model* juga dilakukan, meskipun sebagian besar penelitian tersebut pada domain non-medis. Dengan demikian, terdapat peluang untuk mengadaptasi pendekatan serupa pada teks diagnosis medis yang memiliki karakteristik bahasa dan struktur yang berbeda.

Penelitian ini mengisi celah penelitian sebelumnya, yang belum mengeksplorasi penerapan arsitektur BERT-BiGRU pada teks diagnosis medis otentik berbahasa Indonesia untuk klasifikasi ICD-10, dengan menggunakan data asli dari rumah sakit di Indonesia. Kebaruan pada penelitian ini terletak pada penggunaan arsitektur BERT-BiGRU untuk

melatih model klasifikasi teks diagnosis medis ke dalam kode ICD-10, menggunakan data diagnosis asli dari rumah sakit. Dataset yang digunakan merupakan data lokal yang merepresentasikan kondisi nyata, termasuk variasi bahasa dan struktur yang tidak terstandarisasi. Selain itu, penelitian ini juga mengusulkan strategi penggabungan kelas-kelas yang memiliki jumlah sampel kecil dan cenderung menghasilkan performa rendah (dibawah 50%). Penggabungan ini dimaksudkan agar kelas-kelas tersebut menjadi kelas yang memiliki populasi data lebih besar dan fitur lebih kaya, dan diharapkan akan mempertahankan performa model secara keseluruhan.



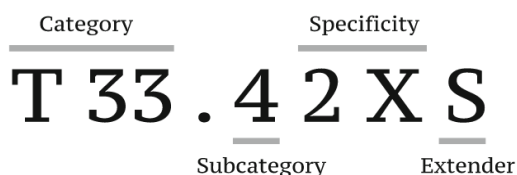
Gambar 2.1 Diagram petal (*flower diagram*) menunjukkan penelitian terkait sebagaimana diuraikan pada bagian sebelumnya.

2.5 Landasan Konsep

2.5.1 Kode ICD - 10

Kode ICD-10 terdiri dari dua sistem yaitu ICD-10 *Clinical Modification* untuk pengkodean diagnostik, dan ICD-9 PCS untuk pengkodean prosedur atau tindakan. ICD-10 menggunakan konvensi pengkodean tujuh karakter mencakup abjad dan numerik

sebagaimana ditunjukkan pada Gambar 2.2. Pada karakter pertama yang menunjukkan jenis diagnosis, dua karakter berikutnya dalam bentuk angka menunjukkan diagnosis yang spesifik, dan 3 tambahan karakter untuk tingkat keparahan, dan sejenisnya (Azam et al., 2020).



Gambar 2.2 Struktur kode ICD-10. Sumber: Sheikh Shams Azam (2020)

Kode ICD-10 memerlukan setidaknya 3 karakter pertama untuk menentukan diagnosis penyakit. Revisi yang berkelanjutan terus dilakukan untuk tujuan meningkatkan akurasi dan relevansi kode. Untuk menentukan tarif BPJS kode ICD-10 diperlukan 3 karakter untuk kategori diagnosis dan ditambah 1 karakter untuk memperinci (Kemkes, 2016).

2.5.2 Text Preprocessing

Text preprocessing adalah proses untuk merapikan, membersihkan, dan menstandarisasi data teks menjadi bentuk yang dapat dikonsumsi oleh sistem yang didukung oleh *machine learning* dan *deep learning*. Teknik umum preprocessing melibatkan *text cleaning*, *tokenizing*, *removing*, konversi, koreksi, penghapusan *stopwords*, *stemming*, dan *lemmatisasi*. Ide utamanya adalah menghapus konten yang tidak perlu agar mendapatkan dokumen teks yang bersih (Sarkar, 2019).

2.5.3 Bidirectional Encoder Representations from Transformers (BERT)

Klasifikasi teks terbagi dalam dua bagian, pertama merupakan proses *feature engineering* yang didalamnya terdapat *word embedding* dan kedua adalah mengklasifikasikannya. BERT merupakan metode *embedding* yang memiliki kelebihan karena kontekstual dan dapat generate embedding untuk kata-kata yang *Out-Of-Vocabulary* (Shreyashree et al., 2022).

Untuk mengklasifikasikan teks digunakan BERT sebagai tahap ekstraksi *embedding*-nya. BERT sendiri telah melalui tahap *pre-trained* terhadap 30 000 kosakata. BERT

menggunakan dua token khusus yaitu [CLS] dan [SEP], untuk menunjukkan pemisahan kalimat dan urutannya.

Model BERT-*base* terdiri dari *encoder* dengan 12 blok *transformer*, 12 *self-attention heads*, dan ukuran *hidden* sebesar 768. Model BERT mengambil masukan berupa urutan dengan panjang maksimum 512 token dan menghasilkan representasi dari urutan tersebut. Pada pengklasifikasi teks, BERT menggunakan *hidden state* terakhir (h) dari token pertama [CLS] sebagai representasi dari seluruh urutan (Sun et al., 2019).

Multilingual BERT (m-BERT) juga merupakan model BERT-Base yang telah dilatih untuk 104 bahasa dunia (Sourav et al., 2023).

2.5.4 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN), ditemukan secara khusus memiliki tujuan untuk mengatasi masalah pemrosesan data *sequential*. RNN mengambil output dari layer sebelumnya sebagai input untuk layer berikutnya. Klasifikasi teks yang merupakan data *sequential* sangat tepat dengan menggunakan RNN (Shreyashree et al., 2022).

Ada dua varian RNN yang tidak mengalami masalah *exploding/vanishing gradient*, yaitu (*Long-Short Term Memory*) dan GRU (*Gated Recurrent Units*). LSTM dilengkapi dengan tiga *gate* yaitu, *forget gate* untuk menghapus semua informasi tidak penting dari memory, *input gate* untuk mengupdate informasi baru kedalam memory, dan *output gate* yang menentukan informasi penting untuk dilewatkan sebagai output. GRU secara sederhana adalah varian dari LSTM.

2.5.5 Bidirectional Gated Recurrent Unit (Bi-GRU)

Gated Recurrent Unit (GRU), merupakan arsitektur RNN yang memiliki mekanisme *gate*, yaitu gerbang pengaturan yang memungkinkan model untuk belajar melupakan (*forget*) terhadap hal yang kurang penting, dan belajar untuk mengingat (*remember*) hal yang lebih penting (Moons et al., 2020).

Dalam bentuk formal dapat ditunjukkan pada persamaan berikut

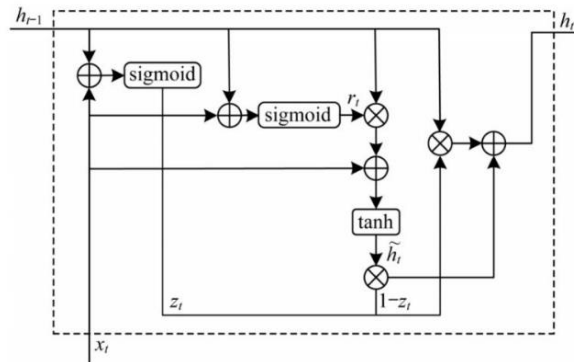
$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (2.1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2.2)$$

$$h_t^* = \tanh(W_h \cdot [r_t \times h_{t-1}, x_t] + b_h) \quad (2.3)$$

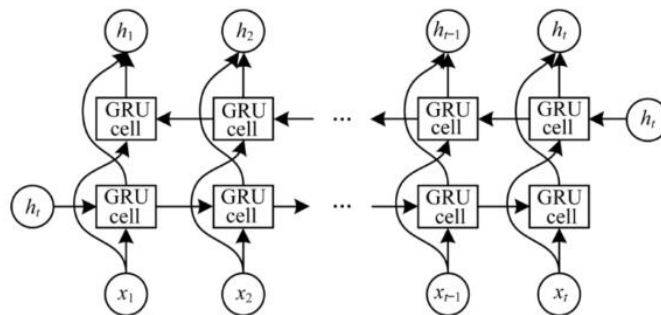
$$h_t = (1 - z_t) \times h_{t-1} + z_t \times h_t^* \quad (2.4)$$

Dengan x_t sebagai vektor input, z_t sebagai update vektor, r_t untuk *reset vector*, dan h_t sebagai *output vector* pada waktu t . Pada persamaan tersebut juga terlihat matriks-matriks W dan nilai bias b yang dapat diperoleh dengan proses training.



Gambar 2.3 Struktur gerbang GRU. Sumber: Ying Tian (2023)

BiGRU merupakan variasi dari GRU yang bekerja dua arah, model struktur GRU terlihat pada Gambar 2.3. BiGRU bekerja dengan memproses data dari depan hingga belakang, dan sebaliknya dari belakang ke depan. Gambar 2.4 menunjukkan keduanya menggunakan vektor kata yang sama, tetapi parameter keduanya independen satu sama lain. Vektor label input dapat dianggap sebagai urutan input. Urutan input melewati *forward* GRU dan *backward* GRU secara berturut-turut (Tian, 2023).



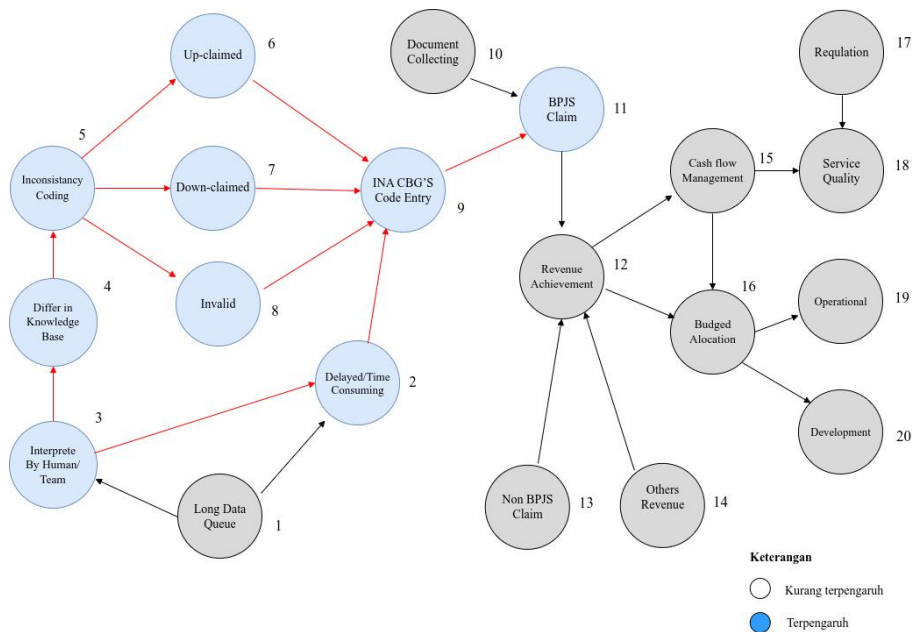
Gambar 2.4 Struktur Bi-GRU. Sumber: Ying Tian (2023).

BAB 3

Metodologi

3.1 Pendekatan Penelitian dalam Penyelesaian Klaim BPJS

Sebagaimana dijelaskan pada bab pendahuluan, *root cause* dari masalah klaim/penagihan BPJS adalah kodifikasi diagnosis utama kedalam kode ICD-10 oleh tim rekam medis atau koding. Gambar 3.1 dapat diamati bahwa permasalahan klaim BPJS, khususnya pada RSUP Persahabatan, dapat ditelusuri kembali ke salah satu faktor utama, yaitu node *interpretate by human*.

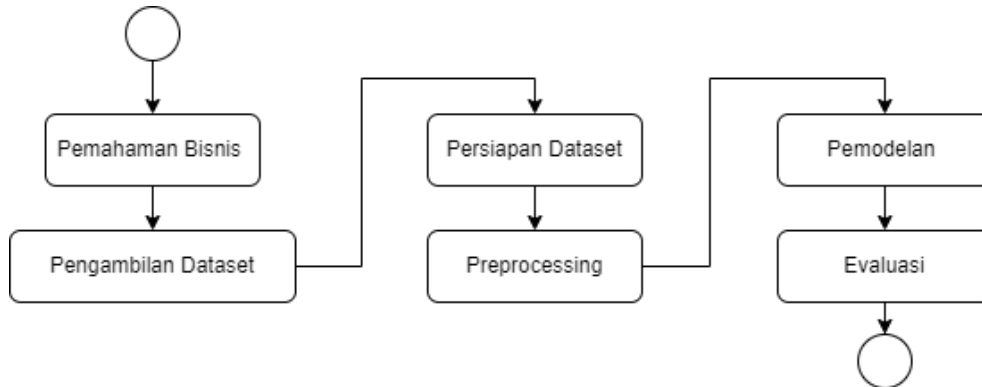


Gambar 3.1 Mengatasi koding oleh manusia (3) akan berdampak hingga klaim BPJS (11).

Metodologi yang digunakan dalam penelitian ini difokuskan pada penyelesaian masalah pada *node* koding tersebut. Dengan menerapkan pendekatan berbasis *deep learning*, penelitian ini bertujuan untuk mengembangkan model klasifikasi yang mampu mengotomatisasi proses pengkodean diagnosis utama. Model yang dihasilkan diharapkan dapat mengurangi subjektivitas dan inkonsistensi dalam interpretasi, sehingga meningkatkan akurasi pengkodean dan secara langsung berdampak pada efisiensi proses klaim BPJS.

3.2 Tahapan Penelitian

Secara garis besar langkah-langkah pelaksanaan penelitian mengikuti pola penelitian sains data yang umum dipraktikkan. Diilustrasikan seperti yang terlihat pada Gambar 3.2.



Gambar 3.2 Diagram alur penelitian secara umum

3.2.1 Pemahaman bisnis

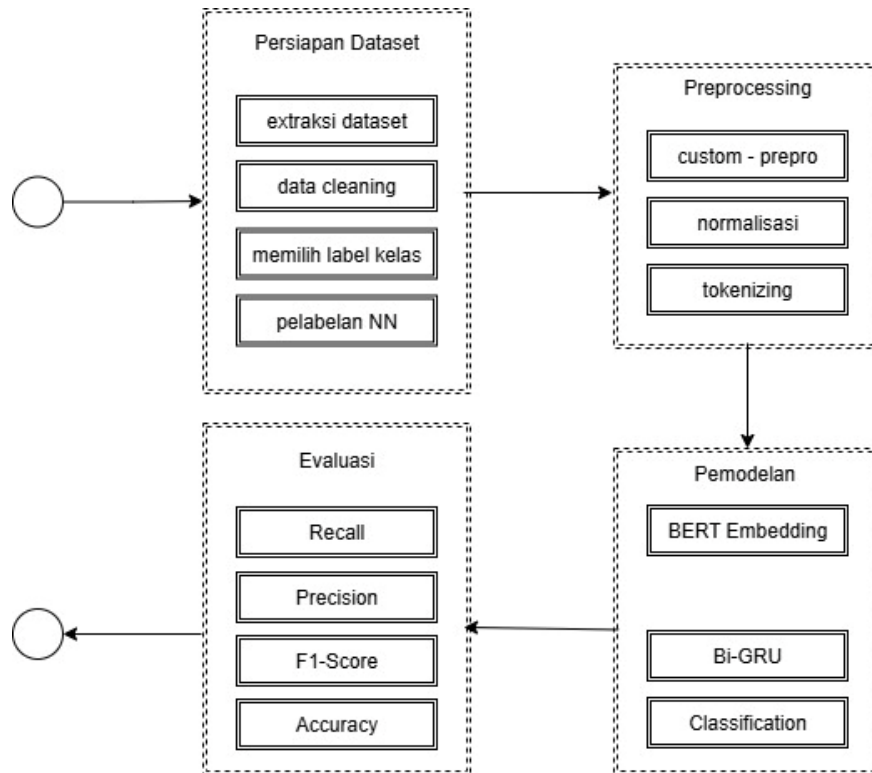
Tahap ini bertujuan untuk memahami konteks operasional dan permasalahan yang dihadapi oleh RSUP Persahabatan, khususnya terkait dengan proses klaim BPJS yang menjadi fokus utama penelitian. Untuk mendukung pemahaman ini, dilakukan kegiatan berikut:

1. Wawancara informal dengan pakar, yaitu Manajer IT dan Direktur Pelayanan Medik.
2. Pertemuan dengan tim perwakilan dari unit pelayanan kesehatan.

Melalui proses ini, diperoleh pemahaman tentang alur proses bisnis terkait pengelolaan klaim BPJS, serta dilakukan pelacakan terhadap kemungkinan hubungan sebab-akibat yang diidentifikasi berkontribusi pada masalah klaim BPJS. Selain itu, dari interaksi dengan para pemangku kepentingan, muncul insight penting yang turut membentuk arah penelitian ini. Kebutuhan terhadap alat bantu, dalam riset ini berupa model klasifikasi, yang memiliki performa tinggi pada label-label tertentu, lebih diutamakan dibandingkan model yang berusaha mencakup seluruh label ICD-10 namun dengan akurasi yang rendah. Temuan ini menjadi dasar dalam perumusan strategi pemilihan label dan rancangan model klasifikasi yang dikembangkan dalam penelitian ini.

3.2.2 Pengambilan data

Data yang akan digunakan pada penelitian, diambil dari database SIMRS RSUP Persahabatan. Proses pengambilan/ekstraksi data dilakukan oleh personel yang memiliki akses database.



Gambar 3.3 Aktifitas rinci pada penelitian yang perlu diperhatikan untuk pemahaman lebih detail

3.2.3 Persiapan dataset

Tahap persiapan dataset (pada Gambar 3.3) diawali dengan mengekstraksi data keluaran dari dalam database menjadi format file *external* yang lebih mudah untuk diproses. Dataset dipersiapkan dalam format file sheet yang mudah dibedakan antara kolom fitur dan kolom label.

Tahap selanjutnya adalah membersihkan dataset, yaitu membuang *record (raw data)* yang rusak, tidak kompatible untuk diproses, dan data kosong. Perlu diingat bahwa dataset bersumber dari sistem yang aktif digunakan sebagai bagiana dari dev-ops dalam kurun waktu yang lama. Dalam prosesnya sangat mungkin terjadi perubahan fitur dan dinamika lainnya. Assumsi ini membuka potensi adanya variasi format data yang terekam dari tiap-tiap perubahan dan penyesuaian sistem.

Untuk mengoptimalkan performa model yang akan dibangun, perlu memperhatikan tahap pemilihan label kelas ICD-10 sebelum memulai tahap pelatihan model. Menentukan label kelas dari daftar kelas-kelas dengan performa tertinggi dapat menjadi pertimbangan yang relevan untuk memperoleh model dengan performa terbaik. Tahap ini disandarkan pada

hasil evaluasi dari percobaan awal yang dilakukan terhadap seluruh label kelas yang tersedia (Buda et al., 2018; Chawla et al., 2002). Eksperimen awal ini dilakukan dengan menggunakan arsitektur model yang sama seperti model utama, yaitu kombinasi antara BERT dan BiGRU, dan dilatih selama 60 epoch. Seluruh label ICD-10 dalam dataset dimasukkan tanpa penyaringan. Nilai-nilai *hyperparameter* yang digunakan mengacu pada nilai standar yang umum digunakan dalam berbagai penelitian serupa, tanpa proses tuning lebih lanjut, guna memastikan *baseline* yang konsisten antar kelas.

Dari hasil evaluasi inilah kemudian ditentukan label-label dari kelas dengan performa di atas 50%, yang diasumsikan masih memiliki potensi performa yang baik untuk dijadikan sebagai label kelas. Sementara itu, kelas-kelas dengan performa di bawah 50% disatukan membentuk kelas baru sebagai kelas yang belum dipetakan (Buda et al., 2018). Dalam penelitian ini digunakan label “NN” sebagai pengganti dari gabungan kelas ICD-10 tersebut. Pendekatan ini bertujuan untuk menjaga keterwakilan data dan memberikan ruang bagi kategori dengan performa rendah tanpa menurunkan performa model secara keseluruhan. Dampak terhadap peningkatan performa model hanyalah manfaat tambahan yang tidak menjadi motivasi utama.

3.2.4 Preprocessing

Meskipun dataset bersumber dari database SIMRS yang sudah memiliki aturan dan validasi penginputan data, akan tetapi masih memungkinkan adanya *inconsistency* saat entry data oleh user. Dari Gambar 3.3 terlihat hanya ada tahap standarisasi teks dan *tokenizing* untuk memisahkan teks menjadi elemen kata-kata terkecil.

Standarisasi dilakukan untuk mengubah teks menjadi bentuk yang standar/seragam dari temuan data yang beragam akan tetapi merupakan data yang identik. Process ini untuk mengatasi penulisan data yang tidak konsisten (Sarkar, 2019), seperti penggunaan singkatan dan variasi penulisan angka. Tahap *tokenizing* akan memberi output yang memudahkan untuk mengenali kata apa saja yang bisa diidentifikasi sebagai nilai. Process tokenisasi pada penelitian ini memanfaatkan fitur yang terdapat pada BERT yang telah dilengkapi dengan *case folding*.

3.2.5 Pemodelan

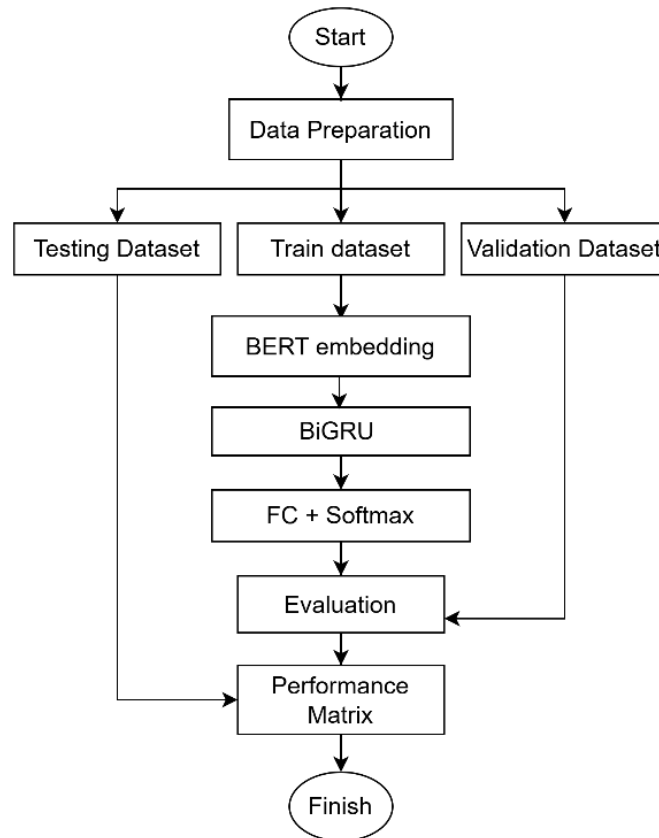
Teks diagnosis yang melewati *preprocessing* selanjutnya ditambahkan data vektor yang dikenal dengan *word embedding*. *Vector embedding* yang digunakan adalah BERT. BERT

merupakan representasi kata dalam bentuk vektor yang didalamnya terkandung makna secara semantik, yaitu hubungan kata dalam konteks makna teks. BERT untuk bahasa Indonesia juga sudah tersedia berbentuk IndoBERT. Setelah tahap ini, langkah berikutnya merupakan training model sehingga perlu memisahkan dataset menjadi training dataset, validation dataset, dan testing dataset. Ketiga dataset digunakan untuk membangun model dengan arsitektur Bi-GRU dan layer *classifier* (Ni et al., 2021; Ouyang et al., 2022; Zhang et al., 2023).

Pendekatan *fixed splitting* dengan membagi data menjadi *training*, *validation*, dan *test set* merupakan praktik umum dalam pengembangan model deep learning, terutama saat data tidak terlalu kecil. Metode ini efisien secara komputasi dan mencerminkan skenario dunia nyata di mana model hanya dilatih sekali dan diuji pada data baru. Seperti dijelaskan oleh Chollet (2021), penyetelan *hyperparameter* dilakukan berdasarkan performa pada data validasi, dan jika dilakukan berulang, berisiko menyebabkan *overfitting* terhadap data validasi. Oleh karena itu, data testing harus tetap terpisah dan hanya digunakan sekali di akhir untuk mengukur kemampuan generalisasi model. Setelah proses tuning selesai, model akhir dilatih ulang menggunakan seluruh data non-test untuk memperoleh hasil pelatihan yang optimal.

Gambar 3.4 menunjukkan setelah lapisan BERT, digunakan lapisan BiGRU (*Bidirectional Gated Recurrent Unit*) sebagai lapisan ekstraktor fitur untuk menangkap pola dan dependensi sekuensial pada teks diagnosis. BiGRU dipilih karena kemampuannya dalam memproses data sekuensial dan menangkap informasi dari dua arah (maju dan mundur) dalam teks (Tian, 2023). Dibandingkan dengan model rekuren lain seperti LSTM, GRU lebih sederhana dan efisien secara komputasi, tetapi tetap memiliki performa yang kompetitif. Kombinasi antara pemahaman kontekstual dari BERT dan kemampuan pemrosesan sekuensial dari BiGRU menjadikan arsitektur ini sesuai untuk tugas klasifikasi teks diagnosis medis.

Proses training model akan melibatkan pengaturan sejumlah hyperparameter dan epoch hingga proses mencapai/mendekati konvergen. Validation dataset berperan dalam memilih pengaturan hyperparameter yang paling optimal. Dan sebagai langkah akhir testing dataset digunakan sebagai pengujian untuk menentukan seberapa baik performa model. Gambar 3.4 menunjukkan bagaimana dataset ditambahkan *vector embedding* dan ekstraksi fitur oleh layer BiGRU untuk selanjutnya dikelompokkan dengan layer *fully connected* dan *softmax* sebagai kalsifikatornya.



Gambar 3.4 Diagram alur training model menggunakan training dataset, mengevaluasi model dengan validation dataset, dan mengukur performa model dengan testing dataset.

3.2.6 Evaluasi

Penelitian ini tidak melibatkan RSUP Persahabatan secara langsung dalam pelaksanaan evaluasi model, sebagaimana telah menjadi batasan lingkup penelitian. Hal ini dilakukan dengan mempertimbangkan bahwa pada tahap awal pengembangan model, fokus utama adalah untuk mengevaluasi performa model menggunakan pendekatan standar. Keterlibatan institusi RSUP akan lebih relevan pada tahap lanjutan ketika model telah menunjukkan kinerja yang stabil dan layak untuk diuji dalam konteks spesifik seperti klaim BPSJ rumah sakit. Selain itu, melibatkan institusi dalam proses evaluasi akan memerlukan persetujuan administratif dan etis yang menjadi lebih kompleks, dan belum menjadi fokus utama dalam ruang lingkup penelitian ini. Oleh karena itu, evaluasi dilakukan secara terpisah dari konteks operasional RSUP Persahabatan, tetapi tetap menjaga relevansi dunia nyata melalui struktur dan karakteristik data yang digunakan.

Dalam evaluasi model sangat penting untuk mampu menentukan label data teks diagnosis sesuai kelasnya, yaitu menentukan label ICD-10 yang sesuai. Untuk fokus kebutuhan ini maka perlu untuk mengevaluasi nilai *recall*. Akan tetapi juga perlu untuk

memperhatikan nilai *precision* agar tetap seimbang. Sebagaimana sudah menjadi praktek umum, untuk model klasifikasi teks untuk mengukur kedua nilai tersebut. Pengukuran kedua nilai performa tersebut dilakukan sepanjang proses training untuk setiap epoch, pada dataset *testing*, dan *validation* agar dapat memantau peningkatan performa model (Batu et al., 2024).

Penelitian ini menggunakan metrik evaluasi yang umum digunakan dalam klasifikasi teks, yaitu precision (P), recall (R), dan F1-score (F1) (Qin et al., 2021). *Precision* menunjukkan tingkat ketepatan dari prediksi model, *recall* menggambarkan jumlah kelas yang berhasil diidentifikasi oleh model, dan *F1-score* digunakan ketika diperlukan keseimbangan antara *precision* dan *recall*. Ketiga metrik evaluasi ini, sebagaimana dijelaskan oleh (Prabhakar & Won, 2021), dihitung menggunakan rumus berikut:

$$Precision = \frac{T_p}{T_p + F_p} \quad (3.1)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (3.2)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3.3)$$

Dimana TP, TN, FP, dan FN secara berurutan mewakili jumlah *true positive*, *true negative*, *false positive*, dan *false negative*. *True Positive* (TP) menunjukkan jumlah prediksi benar untuk kelas positif, *False Negative* (FN) menunjukkan jumlah data yang salah diklasifikasikan sebagai negatif padahal sebenarnya positif, *False Positive* (FP) adalah jumlah data yang salah diklasifikasikan sebagai positif padahal sebenarnya negatif, dan *True Negative* (TN) adalah jumlah prediksi benar untuk kelas negatif (Sarkar, 2019).

Meskipun *recall* dan *precision* sudah cukup mewakili performa model dalam klasifikasi teks, *F1-score* juga digunakan untuk memberikan gambaran yang lebih seimbang antara keduanya. Sementara itu, meskipun *accuracy* bukanlah metrik utama dalam klasifikasi multi-kelas, terlebih yang tidak seimbang, metrik ini tetap dihitung agar hasil penelitian dapat dibandingkan dengan studi lain yang menggunakan *accuracy* sebagai tolok ukur. Namun, *accuracy* tidak menjadi fokus selama proses pelatihan dan penyetalan model. Nilainya hanya dihitung pada data *testing*, dengan menggunakan *confusion matrix*, semata-mata untuk melengkapi evaluasi secara umum. Formula untuk menghitung akurasi terlihat pada persamaan berikut:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (3.4)$$

Dimana TP, TN, FP, dan FN secara berurutan mewakili jumlah *true positive*, *true negative*, *false positive*, dan *false negative*.

3.3 Dataset

Dataset yang digunakan pada penelitian ini bersumber dari data yang telah terkumpul pada database rumah sakit. Pemanfaatan SIMRS oleh seluruh personel rumah sakit selama kurang lebih 5 tahun telah merekam puluhan ribu dataset dengan bermacam variasi jumlah dan jenis. Dataset yang dikeluarkan sebagai sumber data penelitian merupakan data pasien rawat inap dari tahun 2021, 2022, dan 2023. Proses mengekstraksi data dari record database menjadi dataset dilakukan oleh personel rumah sakit yang memiliki akses terhadap database.

Dataset kemudian disimpan dalam format file sheet dan disusun dalam representasi kolom yang terpisah antara fitur dan label agar memudahkan pengolahan selanjutnya. Dua pasangan kolom yang menjadi kebutuhan dasar penelitian ini adalah kolom berisi teks diagnosis utama (primer) dan kolom label kode ICD-10. Terlihat pada Gambar 3.5 kode ICD-10 yang digunakan sebagai label adalah kode yang telah dilakukan pemetaan (kolom Mapp), yaitu label yang telah melewati proses pemilihan.

Model akan dilatih untuk membaca teks diagnosis primer dan memisahkannya sesuai dengan label Mapp pada dataset. Pemisahan dataset sebagai training sebanyak 70%, sebagai validation 20% dan sebagai testing 10%.

ICD10	Mapp	Primer
I25.1	I25.1	Angina Pectoris Stabil ec Coronary Artery Disease 1 Vessel Disease
I12.0	NN	GEA dehidrasi sedang-berat
A09.9	NN	- gea dehidrasi ringan sedang perbaikan
K30	NN	- Observasi febris h5 e.c viral infection
I63.8	NN	- stroke infark tromboembolik frontotemporo-parietalis kanan curiga transformasi per
A15.0	NN	- TB paru bakterologis kasus baru HIV (-) dalam pengobatan OAT bulan ke-1 (TMT 2
E11.9	NN	- tumor mandibula anterior suspek ganas
J47	J47	1. Gagal napas tipe 1
A16.2	A16.2	1. Gagal napas tipe 1 ec HAP high risk 2. Syok sepsis ec HAP high risk
C50.9	NN	1. Gagal napas tipe 2 2. Syok sepsis dd/hipovolemik 3. CAP dengan PSI CI V 4. Ca
E11.5	NN	1. Gangren pedis dextra post amputasi transtibial dextra POD 4 (29/9/23)
A09.9	NN	1. GEA dehidrasi sedang perbaikan
I63.8	NN	1. Hemiparesis dextra et afasia motorik dengan riwayat penurunan kesadaran ec CV
J16.8	J16.8	1. Hemorrhoid grade 4
J16.8	J16.8	1. Pansitopenia dengan riwayat neutropenia berat curiga infiltrasi HIV pada bone ma
K30	NN	1. Pansitopenia dengan riwayat neutropenia curiga infiltrasi HIV pada bone marrow
J16.8	J16.8	CAP
J16.8	J16.8	cap
J90	NN	CAP (104) Usia, jantung, hematokrit, Efusi pleura
J16.8	J16.8	CAP (Community Acquired Pneumonia)
A16.2	A16.2	CAP (Severe) PSI 99 (usia jenis kelamin, RR, natrium)
E11.5	NN	CAP + Hematemesis
J16.8	J16.8	CAP CURB 65-0, PSI Class III
J16.8	J16.8	CAP CURB 65-1

Gambar 3.5 *Screen capture* ilustrasi dataset dalam bentuk file worksheet

BAB 4

Hasil dan Pembahasan

Pada Bab 4 ini menunjukkan hasil yang diperoleh dari pelaksanaan tahapan penelitian serta analisis terhadap temuan yang muncul. Pembahasan akan mencakup anomali, pola/tren, serta faktor yang berpengaruh yang ditemukan.

4.1 Pemahaman Bisnis

Tahap awal penelitian, dilakukan pertemuan dengan pihak RSUP Persahabatan, termasuk dengan tim rekam medis, tim SIMRS dan manajemen rumah sakit. Juga dilakukan diskusi/interview lanjutan dengan para pakar RSUP Persahabatan. Tujuan dari diskusi ini adalah untuk memahami proses pencatatan diagnosis, pengkodean ICD-10, dan pengajuan klaim BPJS secara umum seperti ditunjukkan pada Gambar 1.2. Tujuan lain sebagaimana disebutkan pada bab 3 juga bertujuan untuk mengidentifikasi masalah seputar penagihan klaim BPJS. Gambar 1.1 merupakan ilustrasi masalah yang menjadi penyebab dan akibat, beberapa diantaranya diidentifikasi melalui proses wawancara.

Pada tahap awal, terdapat ekspektasi bahwa pihak RSUP akan menunjukkan antusiasme yang tinggi terhadap inisiatif penelitian berbasis sains data ini, mengingat relevansi terhadap efisiensi pengkodean dan proses klaim. Namun, kenyataan di lapangan menunjukkan adanya keraguan dan bahkan resistensi dari sebagian pihak, terutama terkait dengan pemahaman konsep dasar pemodelan berbasis NLP. Salah satu pernyataan yang muncul dalam diskusi menyiratkan keraguan terhadap nilai riset apabila model yang dikembangkan belum mampu mencapai tingkat akurasi di atas 95%. Persepsi ini menjadi tantangan tersendiri dalam memahami antara pendekatan ilmiah yang bersifat eksperimental dan ekspektasi praktis di lingkungan rumah sakit. Meskipun demikian, izin untuk mengakses dan menggunakan data tetap diberikan, meskipun dengan ekspektasi bahwa hasil akhir harus menunjukkan performa yang signifikan agar dinilai layak untuk dilanjutkan lebih jauh.

Dari hasil diskusi tersebut, juga diperoleh beberapa masukan penting. Salah satunya adalah, jika pendekatan teknologi sains data dapat menjadi solusi maka pihak RSUP lebih mengutamakan kinerja model dapat diandalkan meskipun hanya untuk beberapa kode ICD-10 tertentu, dibandingkan cakupan kode ICD-10 yang luas. Salah satu narasumber menyatakan bahwa rumah sakit tetap dapat memperoleh manfaat dari penelitian ini, “...

meskipun ini penelitian ya, tetapi Rumah sakit bisa dapat manfaatnya, jika memang tidak semua ICD bisa ditebak, maka bisa kita ambil yang bagus-bagus saja dulu. Nanti jika itu bisa terbukti baik bisa kita lanjutkan...” (Yussianto, 4 September 2023).

Berdasarkan masukan ini, arah penelitian kemudian disesuaikan untuk memprioritaskan performa model daripada mengklasifikasikan seluruh kode ICD-10 yang tersedia. Alih-alih melatih model pada 24 kelas ICD-10, penelitian ini difokuskan pada 7 kelas teratas yang menunjukkan performa klasifikasi >50% pada eksperimen awal. Pendekatan ini dianggap lebih bermanfaat secara praktis karena model yang mampu membantu pengkodean diagnosis umum secara andal dinilai lebih bernilai bagi rumah sakit dibandingkan model dengan cakupan luas tetapi performa rendah.

4.2 Persiapan Data

Penelitian ini mengumpulkan data rekam medis dari Sistem Informasi Manajemen Rumah Sakit (SIMRS) di sebuah rumah sakit rujukan nasional, sebanyak 10.000 raw data, dimana masing-masing terdiri dari kode ICD-10 yang dipasangkan dengan teks diagnosis utama. Selama proses ekstraksi, ditemukan beberapa baris data yang rusak atau tidak lengkap, sehingga tidak dapat digunakan. Setelah dilakukan proses pembersihan data secara menyeluruh, sebanyak 9.982 baris data tetap layak untuk dianalisis. Dataset ini kemudian diekspor dalam format CSV atau XLS.

Temuan data yang perlu dibersihkan diantaranya adalah data yang dianggap rusak karena memiliki format berbeda, memiliki karakter operasional misalnya, '=' dan '-' seperti ditunjukkan oleh **Error! Reference source not found.**, data dengan format json pada Gambar 4.2, dan data yang dianggap kolom kosong seperti pada Gambar 4.3. Ilustrasi pada gambar-gambar tersebut **Error! Reference source not found.** menunjukkan potongan informasi contoh data rusak yang perlu dibuang/dibersihkan.

C50.9	pro opreksisi luas mamae kiri	C	leura, nodul kontralateral, tulang) mutasi EGFR
J47	=- hemoptisis perbaikan - pro embolisasi	F	=- pemeriksaan swab antigen tidak dilakukan
J16.8	sepsis	I	M1: CT Scan toraks kontras

Gambar 4.1 Ilustrasi data dengan temuan karakter operasi

133)\nHFpEF (58%	USG Abdomen"]	USG Abdomen	{ "kode": "", "nama": "USG Abdomen" }	{ "transportasi_pulang": "mobil", "tujuan_pulang": "1", "diagnosa_awal_dokter": "\u2022 Dipepsia without Alarm Symptoms\n\u2022 TB Paru on OAT bulan ke 8\n\u2022 Hiponatremia Hiposmolar Euvolemik (Na:133)\nHFpEF (58%", "diagnosa_utama_icd": "", "diagnosa_sekunder_icd": "", "tindakan_icd": "", "umm_kesadaran": "baik", "umm_vital": "baik", "tgl_jam_resume": "2023-07-28 15:00:00", "dl": "" }
------------------	---------------	-------------	---------------------------------------	---

kolosis	{ "transportasi_pulang": "kendaraan pribadi", "tujuan_pulang": "1", "diagnosa_awal_dokter": "Bronkiectasis terinfeksi pada bekas TB\nTerduga TB kambuh", "diagnosa_utama_icd": null, "diagnosa_sekunder_icd": "", "tindakan_icd": "", "umm_kesadaran": "CM", "umm_vital": "TSS", "tgl_jam_resume": "", "dl": "" }
---------	---

Gambar 4.2 Ilustrasi data dengan temuan format json

I63.8	\N
I63.8	\N
C34.9	\N
C34.9	\N
A16.2	\N
C34.9	\N
A15.0	\N
C34.9	\N
J90	\N
J90	\N
J16.8	\N
J16.8	\N
K30	\N

Gambar 4.3 Ilustrasi data dengan temuan data fitur dianggap kosong

4.3 Pelabelan Dataset

Untuk mendapatkan wawasan awal, dataset dikelompokkan sesuai label ICD-10-nya untuk menganalisis distribusi frekuensi masing-masing kode dan teks diagnosis. Total kelas pada dataset awal adalah 24 kelas, mengacu pada masing-masing kode ICD-10 pada table Tabel 4.1 Distribusi frekuensi kode ICD-10 pada dataset awal. Dalam kasus dataset yang tidak seimbang (*imbalanced dataset*), sangat penting untuk melakukan evaluasi pada tahap awal eksperimen guna menilai kinerja masing-masing model (Buda et al., 2018). Evaluasi ini memberikan gambaran menyeluruh mengenai performa model dalam mengklasifikasikan setiap kelas dengan baik. Dengan menyertakan seluruh dataset dalam eksperimen awal, dapat diperoleh tolok ukur kinerja dasar yang menjadi acuan dalam menentukan efektivitas model serta kelas-kelas yang memiliki akurasi prediksi yang lebih baik dibandingkan yang lain.

Tabel 4.1 Distribusi frekuensi kode ICD-10 pada dataset awal

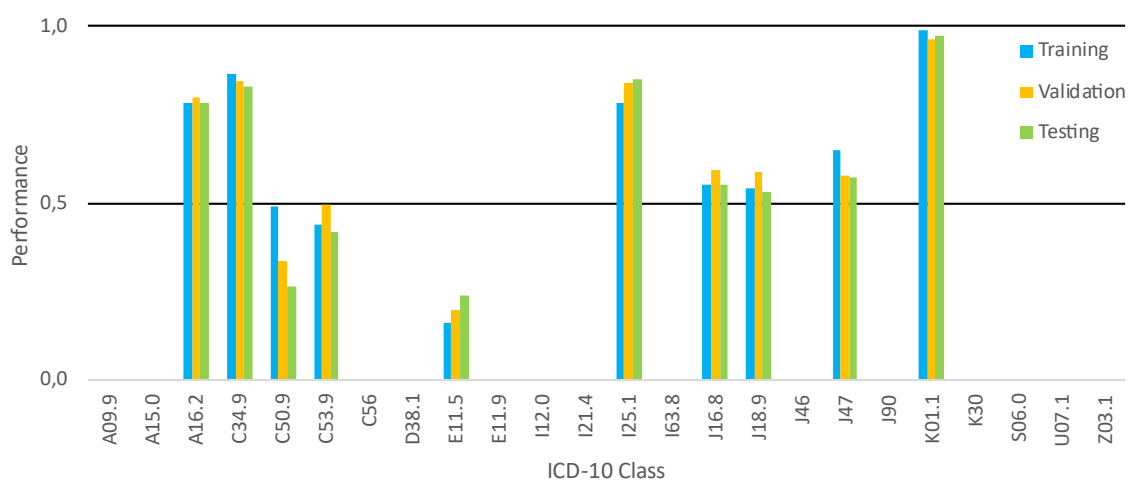
ICD-10	All Diagnosis	Unique Diagnosis	Label Map
C34.9	1792	1692	C34.9
A16.2	1088	785	A16.2
I25.1	1026	411	I25.1
J16.8	978	617	J16.8
K01.1	972	513	K01.1
J18.9	830	273	J18.9
J47	527	270	J47
C53.9	489	278	NN
C50.9	194	187	NN
A09.9	194	157	NN
I12.0	193	181	NN
C56	184	180	NN
E11.5	183	174	NN
I63.8	166	154	NN
E11.9	165	162	NN
J90	153	147	NN
I21.4	141	78	NN
A15.0	137	132	NN
K30	123	114	NN
J46	122	101	NN
Z03.1	71	69	NN
D38.1	65	64	NN
U07.1	47	44	NN
S06.0	42	36	NN
Total	9882	6819	24

Berdasarkan hasil evaluasi pada eksperimen awal, pemilihan kelas yang efektif, yaitu kelas dengan kinerja F1-score di atas 50%, akan menjadi strategi yang tepat untuk meningkatkan kinerja model secara keseluruhan. Tabel 4.2 secara rinci menampilkan pengukuran performa masing-masing kelas selama eksperimen awal. Kelas-kelas yang melewati ambang batas ini berpotensi menghasilkan prediksi yang lebih baik, sehingga dapat menjadi kandidat utama dalam proses pengembangan model lebih lanjut. Dari Gambar 4.4 diketahui bahwa terdapat 7 (tujuh) kelas yang memenuhi kriteria tersebut, sehingga ketujuh kelas ini menjadi kandidat yang dapat dipertahankan dalam model. Sementara itu, kelas-kelas dengan kinerja di bawah ambang batas akan digabungkan ke dalam satu kategori baru ("NN"). Langkah ini bertujuan untuk menyeimbangkan distribusi dataset serta menyederhanakan klasifikasi, sehingga model dapat bekerja lebih optimal tanpa terdistraksi oleh kelas dengan kinerja yang sangat rendah (Chawla et al., 2002).

Tabel 4.2 Performa setiap kelas dari hasil evaluasi pada eksperimen awalan

Class	Training			Valid			Test		
	prec	recall	f1	prec	recall	f1	prec	recall	f1
A09.9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
A15.0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
A16.2	0,717	0,863	0,783	0,712	0,913	0,800	0,720	0,855	0,781
C34.9	0,790	0,962	0,867	0,752	0,970	0,847	0,711	1,000	0,831
C50.9	0,451	0,541	0,492	0,278	0,417	0,333	0,208	0,357	0,263
C53.9	0,317	0,722	0,440	0,375	0,729	0,496	0,319	0,603	0,417
C56	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
D38.1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
E11.5	0,209	0,129	0,160	0,267	0,157	0,198	0,294	0,200	0,238
E11.9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
I12.0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
I21.4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
I25.1	0,681	0,923	0,783	0,766	0,935	0,842	0,860	0,841	0,851
I63.8	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
J16.8	0,413	0,823	0,550	0,474	0,796	0,594	0,429	0,776	0,552
J18.9	0,422	0,763	0,544	0,500	0,718	0,589	0,406	0,765	0,531
J46	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
J47	0,561	0,773	0,650	0,507	0,667	0,576	0,457	0,762	0,571
J90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
K01.1	0,986	0,989	0,988	0,938	0,989	0,963	0,964	0,981	0,972
K30	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
S06.0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
U07.1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Z03.1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Model performance

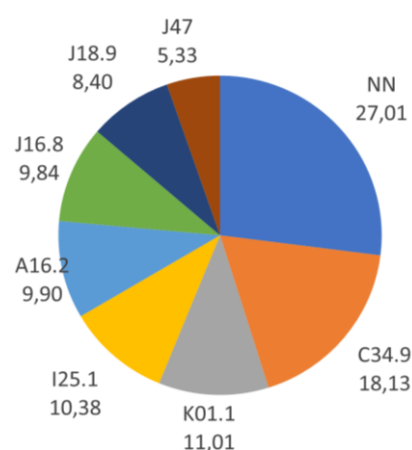


Gambar 4.4 Visualisasi F1-score tiap kelas pada eksperimen awal menunjukkan kelas dengan performa di atas 50%, untuk dijadikan target untuk pelatihan model utama.

Tabel 4.3 Distribusi frekuensi kelas ICD-10 paska pemetaan label

New ICD-10	All Data	Unique data
NN	2669	2258
C34.9	1792	1692
A16.2	1088	785
I25.1	1026	411
J16.8	978	617
K01.1	972	513
J18.9	830	273
J47	527	270
Total	9882	6819

Penelitian ini kemudian dapat berfokus pada kelompok kelas ICD-10 yang memiliki minimal performa 50%, yang digunakan sebagai label kelas untuk keperluan klasifikasi. Kelompok data lainnya dipetakan ke dalam kode khusus yang diberi label "NN" untuk mensimulasikan kode yang tidak dikenali. Total label kelas yang akan menjadi target pelatihan adalah 7 kelas ICD-10 dan 1 kelas NN sebagai gabungan. Tabel 4.3 menampilkan populasi data hasil pemetaan kode ICD-10 ke dalam label kelas. Dari tabel tersebut terlihat bahwa label NN memiliki jumlah tertinggi karena menggabungkan berbagai kode ICD-10 ke dalam satu kategori. Rasio distribusi frekuensi label baru ini dapat dilihat pada Gambar 4.5.

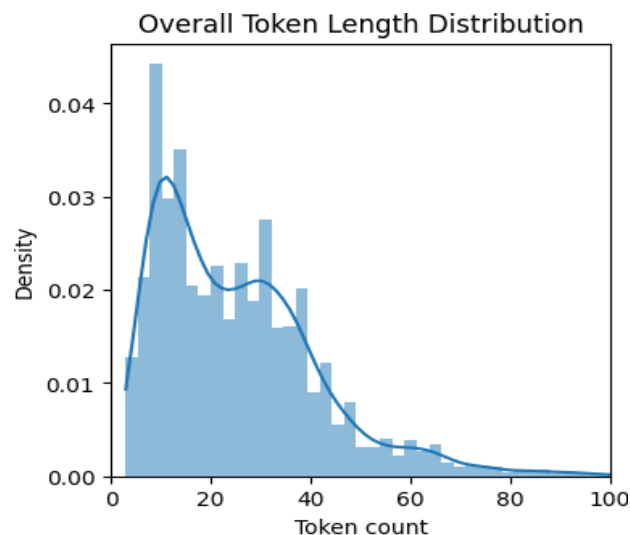


Gambar 4.5 Distribusi frekuensi kelas ICD-10 paska pemetaan

4.4 Distribusi Token

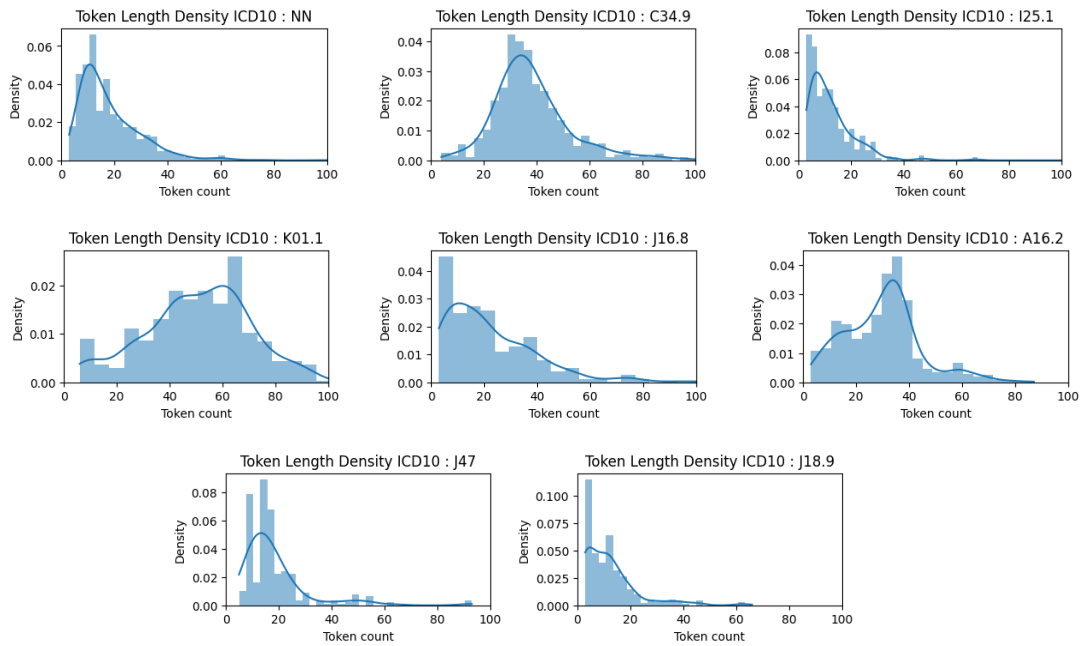
Sebuah kalimat dapat diuraikan menjadi daftar kata, yang kemudian dapat digunakan untuk merekonstruksi kalimat aslinya (Sarkar, 2019). Proses ini dikenal sebagai tokenisasi. Dalam BERT, setiap token dikonversi menjadi representasi numerik yang disebut embedding.

Dalam konteks klasifikasi teks, distribusi panjang token pada data dapat memberikan wawasan awal yang penting terhadap karakteristik dan kualitas representasi data (Yehezkel & Pinter, 2022). Panjang token yang dimaksud merujuk pada jumlah unit token hasil proses tokenisasi yang dihasilkan oleh *tokenizer* (Xu et al., 2019). Distribusi ini merefleksikan seberapa bervariasi struktur kalimat dalam data teks diagnosis yang digunakan. Untuk tujuan tersebut, penelitian ini menggunakan *Visual exploratory analysis* menggunakan *histogram plotting* terhadap panjang token untuk keseluruhan data dan setiap kelas sebagai pendekatan visual guna mengevaluasi keseimbangan dan penyebaran data, dan juga memberikan indikasi awal suatu kelas memiliki kecenderungan tertentu.



Gambar 4.6 Histogram distribusi frekuensi panjang token untuk seluruh dataset

Histogram pada Gambar 4.6 menunjukkan distribusi frekuensi panjang token, yang menunjukkan persentase data dengan jumlah token yang sama. Teks diagnosis pada dataset didominasi oleh kalimat pendek, yang ditunjukkan dengan rendahnya frekuensi token yang lebih tinggi. Variasi panjang kalimat, atau inkonsistensi panjang data, menjadi faktor penting dalam menganalisis performa dari model seperti presisi dan recall pada tahap evaluasi.



Gambar 4.7 Distribusi frekuensi panjang token untuk setiap label kelas ICD-10 pada dataset sebelum proses split

Sebagai dasar analisis model lebih lanjut, informasi terperinci tentang distribusi token untuk setiap kelompok kode ICD-10 pada dataset awal sebelum pembagian training, validation, dan testing, juga penting untuk divisualkan. Gambar 4.7 menampilkan histogram frekuensi token untuk setiap label kelas ICD-10 pada dataset, memberikan wawasan mengenai bagaimana token didistribusikan dalam setiap kelas. Beberapa kelas menunjukkan distribusi yang lebih merata dibandingkan yang lain, yang dapat mengindikasikan penyebaran panjang kalimat yang seimbang. Secara khusus, kelas C34.9 dan K01.1 menunjukkan distribusi yang menyerupai distribusi normal tanpa *skew*, yang menunjukkan keseragaman dalam frekuensi token yang dapat berdampak berbeda terhadap prediksi model. Pada analisa lebih lanjut terhadap performa model diamati bahwa kelas dengan distribusi panjang token yang lebih seimbang cenderung memiliki performa yang lebih baik dalam klasifikasi. Observasi ini lebih jauh didukung oleh hasil yang ditampilkan dalam *confusion matrix* pada Gambar 11 dan kinerja model pada Tabel 7, yang akan dibahas pada bagian berikutnya.

4.5 Pelatihan Model

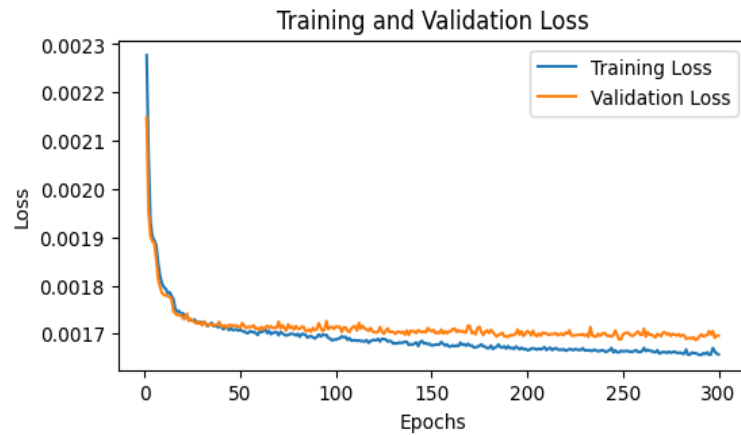
Model telah dilatih menggunakan arsitektur *Stacked* BERT-BiGRU dengan menggunakan parameter sebagaimana tercantum dalam Tabel 4.4. Parameter ini telah melalui beberapa

iterasi penyesuaian dan pengujian untuk mencapai kinerja optimal selama pelaksanaan penelitian. Proses pelatihan juga melibatkan evaluasi performa model dengan menggunakan perhitungan loss. Perhitungan ini mengukur perbedaan antara nilai prediksi model dan nilai sebenarnya. Untuk model klasifikasi teks, perhitungan training loss umumnya dilakukan menggunakan *Cross Entropy Loss* (Zhang et al., 2023). Selama lima *epoch* pertama, terjadi penurunan kesalahan yang signifikan, diikuti oleh penurunan yang berkelanjutan pada *training loss* maupun *validation loss*. Seperti yang ditunjukkan dalam Gambar 4.8, setelah *epoch* ke-36, nilai loss menjadi minimal, dengan nilai mendekati 10^{-6} per *epoch* untuk *training* dan *validasi* dataset.

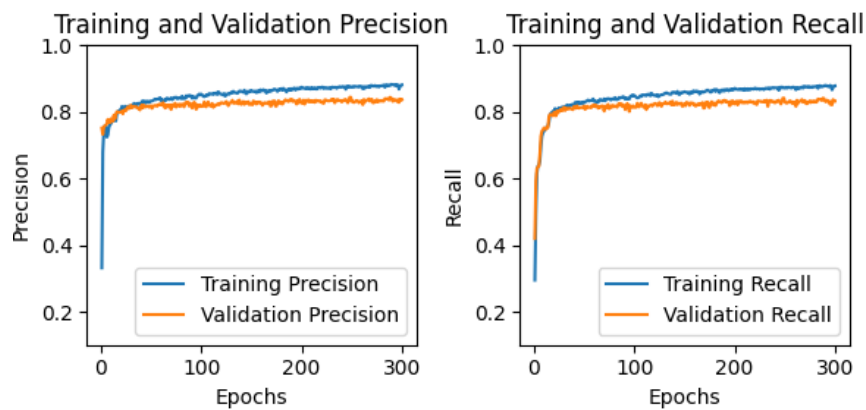
Tabel 4.4 Hyperparameter yang digunakan untuk pelatihan model

Parameter	Nilai
BERT model	BERT Base
Learning rate	0.0001
Epoch	300
Dropout rate	0.2
Layer	2
Hidden layer dimension	128
Loss function	Cross Entropy
Optimizer	Adam

Setelah *epoch* ke-36, peningkatan kinerja model menjadi kurang seimbang. Gambar 4.9 menunjukkan bahwa peningkatan performa pada *validation dataset* berlangsung lebih lambat dibandingkan dengan *training dataset*. Pada *epoch* ke-35, perbedaan nilai *precision* antara kedua dataset hanya 0,0003, tetapi pada *epoch* ke-300, selisihnya meningkat menjadi 0,0457. Meskipun tanda-tanda *overfitting* belum sepenuhnya muncul, perbedaan performa antara kedua dataset mulai terlihat jelas.



Gambar 4.8 Penurunan nilai loss selama proses pelatihan 300 epoch



Gambar 4.9 Komparasi peningkatan performa pada *training dataset* dan *validation dataset*

4.6 Optimisasi Hyperparameter

Menyesuaikan berbagai parameter saat membangun model merupakan praktik umum untuk mencapai kinerja optimal (Parwez et al., 2023). Penelitian sebelumnya oleh Yu et al. (2021) mengidentifikasi bahwa ukuran batch (*batch size*) 4, ukuran *hidden layer* 768, dan *learning rate* 1E-4 adalah parameter terbaik untuk klasifikasi BERT-BiGRU.

Dalam penelitian ini, berbagai kombinasi *hyperparameter* diuji untuk menemukan konfigurasi optimal dalam model klasifikasi teks. Secara khusus, eksperimen dilakukan dengan variasi *dropout rate* antara 0.2 hingga 0.4 serta jumlah layer yang berbeda pada BiGRU. Hasil eksperimen ini ditampilkan dalam Tabel 4.5, menunjukkan bagaimana pengaruh perubahan *hyperparameter* terhadap peforma model. Dipastikan hingga epoch ke-15 ditemukan bahwa konfigurasi yang optimal, sebagaimana ditunjukkan dalam Tabel 4.5, adalah *dropout rate* 0.2 dengan 2 *n-layer*.

Tabel 4.5 Evaluasi performa model menggunakan BERT-base, *learning rate* 1E-4 dan beberapa variasi *drop rate* dan *n-layer*

Epoch	Bert model = Base, Learning rate = 1E-04						
	Evaluation	Dropout = 0.2		Dropout = 0.3		Dropout = 0.4	
		Layer = 2	Layer = 3	Layer = 2	Layer = 3	Layer = 2	Layer = 3
5	Precision	0,773	0,7583	0,7206	0,7377	0,7596	0,7514
	Recall	0,7113	0,672	0,6676	0,6631	0,6559	0,6863
	F1 score	0,6827	0,6105	0,6056	0,608	0,5914	0,6331
10	Precision	0,7761	0,7748	0,7636	0,7679	0,7669	0,7751
	Recall	0,7525	0,7498	0,7391	0,7462	0,7489	0,7551
	F1 score	0,7312	0,7308	0,7172	0,729	0,7296	0,7314
15	Precision	0,8031	0,8159	0,767	0,8146	0,7936	0,7822
	Recall	0,8034	0,798	0,7426	0,7971	0,7554	0,7819
	F1 score	0,8022	0,7969	0,7217	0,7776	0,7568	0,7793

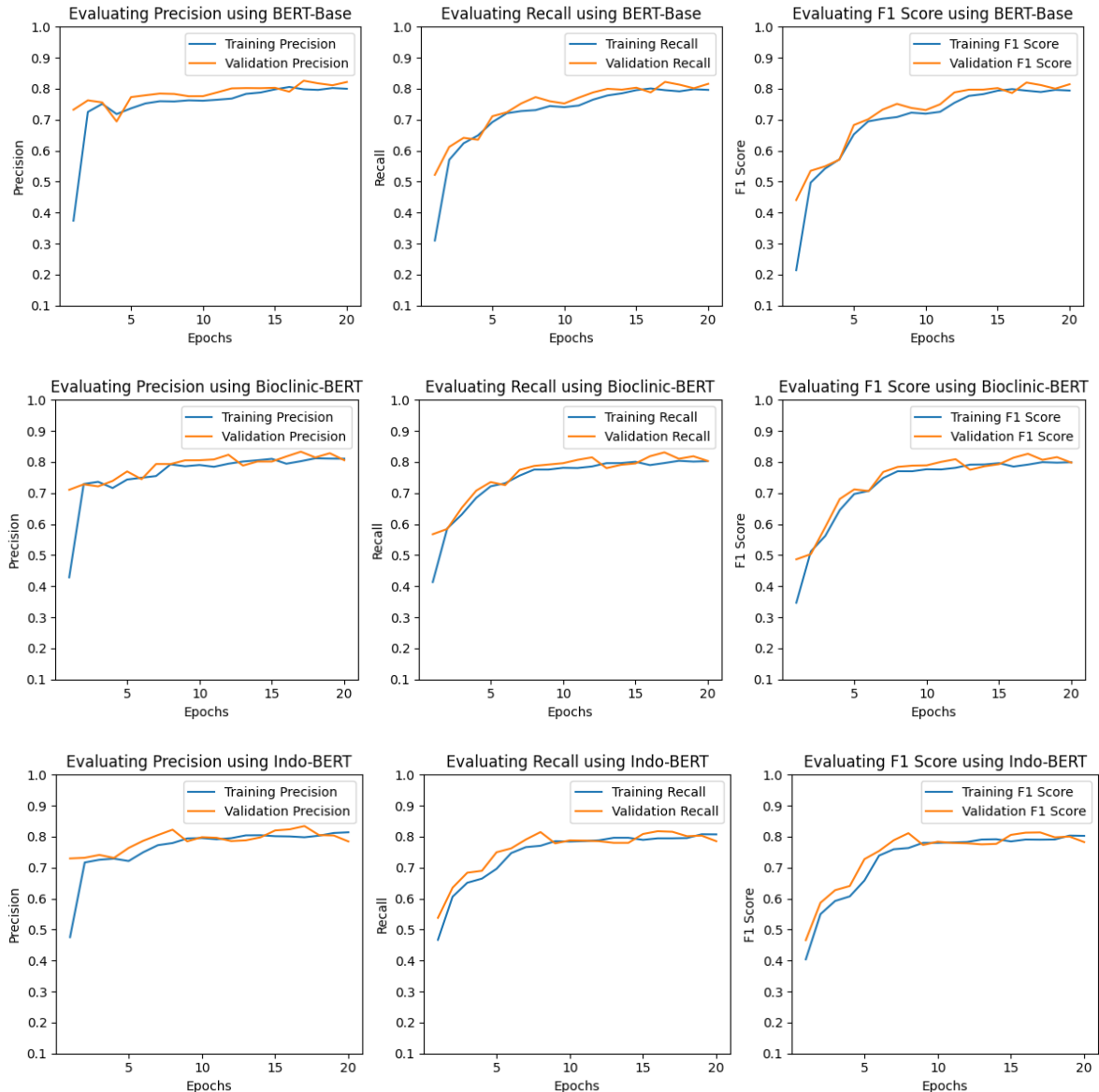
4.7 Vector Embedding

Penelitian ini, perlu untuk mengevaluasi penggunaan beberapa *word embedding vectors* yang digunakan. Variasi ini untuk mengoptimalkan kinerja model dalam klasifikasi teks (Duarte et al., 2018). Seperti yang ditunjukkan dalam penelitian sebelumnya (Ni et al., 2021), beberapa *word embeddings* diuji untuk melihat pengaruhnya terhadap kinerja model. Dalam konteks ini, eksperimen dilakukan dengan menggunakan tiga model BERT *pre-training* sebagai *vector embedding*, yaitu BERT-Base, Indo-BERT, dan BioClinicalBERT. Performa ketiga model BERT ini dibandingkan untuk menentukan pendekatan yang paling efektif.

Pemilihan ketiga embedding ini tidak dilakukan secara acak, tetapi didasarkan pada asumsi awal mengenai karakteristik data. Mengingat teks diagnosis sebagian besar ditulis dalam bahasa Indonesia, IndoBERT diperkirakan akan memberikan performa terbaik karena model ini telah dilatih pada korpus berbahasa Indonesia yang luas, sehingga diasumsikan memiliki tingkat *out-of-vocabulary* yang lebih rendah. Di sisi lain, penggunaan BioClinicalBERT dipertimbangkan karena banyaknya terminologi medis terkandung dalam data. Oleh karena itu, BioClinicalBERT diharapkan dapat menangkap konteks domain secara lebih presisi.

Namun, hasil eksperimen, seperti yang divisualkan pada Gambar 4.10, justru menunjukkan tren peningkatan kinerja yang mirip baik pada *training* dan *validation* di

semua *vector embedding*. Meskipun tren-nya mirip, Tabel 4.6 secara detail menunjukkan bahwa BERT-Base memberikan dampak terbaik dalam eksperimen ini, dengan *precision* sebesar 82,2% dan *recall* sebesar 81,6% pada *epoch* ke-20.



Gambar 4.10 Evaluasi performa model menggunakan BERT-base, Indo-BERT, dan BioClinical-BERT

Temuan ini bertentangan dengan ekspektasi awal dan menunjukkan bahwa asumsi tentang kecocokan *embedding* berbasis bahasa atau domain belum tentu selalu menghasilkan peningkatan performa. Dalam hal ini, karakteristik data yang merupakan kombinasi antara bahasa Indonesia, istilah medis, dan bahkan potongan bahasa Inggris, tampaknya lebih sesuai dengan BERT-Base yang memiliki cakupan tokenisasi yang lebih luas dan generalis.

Dengan demikian, meskipun IndoBERT dan BioClinicalBERT memiliki keunggulan pada domain teks tertentu, keduanya tidak sepenuhnya optimal dalam menangani kompleksitas campuran teks diagnosis yang digunakan dalam penelitian ini.

Tabel 4.6 Evaluasi performa model dengan variasi penggunaan *vector embedding*

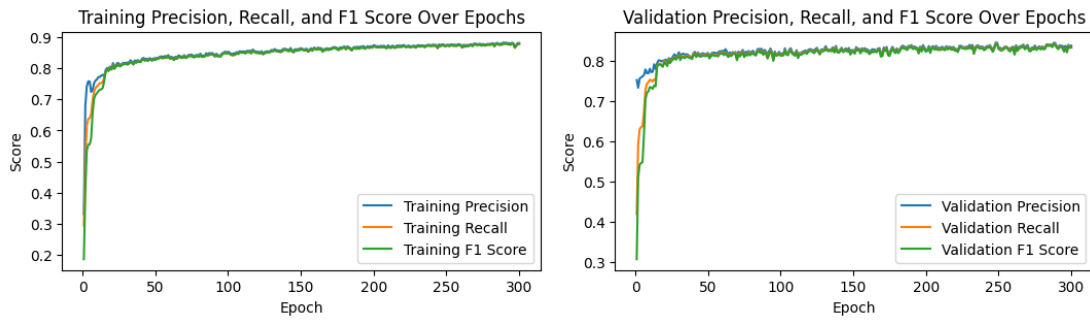
Epoch	BERT Based		Indo-BERT		BioClinical BERT	
	Presisi	Recall	Presisi	Recall	Presisi	Recall
5	0,771	0,711	0,764	0,750	0,770	0,736
10	0,776	0,753	0,798	0,788	0,806	0,796
15	0,803	0,803	0,820	0,809	0,801	0,795
20	0,822	0,816	0,785	0,786	0,806	0,803

4.8 Evaluasi Model

Sepanjang proses pelatihan, evaluasi dilakukan pada *training* dan *validation dataset* untuk memantau peningkatan kinerja di setiap epoch, sebagaimana diterapkan dalam (Batu et al., 2024). Evaluasi yang berkelanjutan ini memungkinkan penyesuaian secara real-time serta memberikan *feedback* yang berguna terhadap pengembangan model.

Setelah melatih model selama 300 *epoch*, evaluasi final pada *testing dataset* menunjukkan performa yang baik, dengan *precision* sebesar 83,67%, *recall* 82,74%, *F1-score* 82,66%, dan akurasi 81,92%. *Precision* yang tinggi menunjukkan bahwa model mampu mengidentifikasi kasus positif dengan baik, di mana 83,04% dari prediksi positifnya benar. Sementara itu, *recall* sebesar 82,74% mencerminkan kemampuan model dalam mendeteksi sebagian besar kasus positif yang sebenarnya. Meskipun metrik ini cukup kuat, akurasi sebesar 81,92% menunjukkan bahwa masih ada ruang untuk meningkatkan akurasi keseluruhan model

Performa pada *training dataset* dan *validation dataset* menunjukkan peningkatan yang konsisten sepanjang 300 *epoch*, seperti yang ditunjukkan pada Gambar 4.11. Tren peningkatan yang stabil ini menegaskan kemampuan model dalam belajar secara efektif. Namun, perlu diperhatikan kemungkinan munculnya tanda-tanda *overfitting*.



Gambar 4.11 Peningkatan performa pada training dan validation dataset. Meliputi precision, recall, dan F1-score

Seperti yang terlihat pada Gambar 4.9, terdapat perbedaan yang cukup jelas antara performa *training dataset* dan *validation dataset*. Meskipun *validation dataset* tidak menunjukkan tanda-tanda penurunan yang signifikan, akan tetapi peningkatan terus-menerus pada performa *training dataset* menunjukkan bahwa model mungkin *overfit* terhadap *training dataset*. Untuk menghadapi masalah ini, maka proses pelatihan dibatasi hingga pada *epoch* dimana tanda-tanda kecurigaan *overfitting* mulai muncul, guna tetap mempertahankan peningkatan performa model.

Selanjutnya *Confusion matrix* untuk *testing dataset* dapat memberikan gambaran menyeluruh tentang akurasi klasifikasi model pada berbagai kelas (Berutu et al., 2023; Desiani et al., 2024). *Confusion matrix* dalam Gambar 4.12 menunjukkan secara detail kemampuan model dalam memprediksi setiap kelas serta membantu mengidentifikasi kelas tertentu dimana model mungkin bermasalah. Evaluasi ini menyoroti kinerja secara keseluruhan dan mengungkap variasi performanya di tiap kelas, dengan menunjukkan kelas mana yang memiliki performa baik dan di mana diperlukan perbaikan lebih lanjut untuk meningkatkan performa spesifik per kelas.

Dengan menganalisis matrix seperti yang ditunjukkan dalam Gambar 4.12, kinerja setiap kelas/label dapat dievaluasi untuk mengidentifikasi keunggulan dan kelemahan model. Secara umum, dataset yang lebih besar dapat meningkatkan performa klasifikasi dengan menyediakan sampel yang lebih representatif, mengurangi *overfitting*, dan meningkatkan kemampuan model dalam melakukan generalisasi terhadap data baru, sebagaimana disebutkan juga dalam penelitian (Althnian et al., 2021; Arhami et al., 2024). Korelasi yang kuat antara jumlah populasi data per kelas dan performa model memprediksi setiap kelas dapat dikonfirmasi dengan mengacu kembali pada Tabel 4.3 dan Tabel 4.7. Kelas dengan jumlah populasi data terbanyak, kecuali kelas NN, menunjukkan performa

tertinggi. Namun, penting untuk mempertimbangkan bahwa populasi data yang dimaksud tidak hanya dinilai berdasarkan jumlah data mentah, tetapi juga memperhatikan jumlah data yang unik, yaitu data yang benar-benar berbeda. Dalam proses analisis ditemukan bahwa banyak entri dalam suatu kelas ternyata merupakan duplikasi yang hanya berbeda pada penggunaan spasi atau format huruf sehingga secara substansi tidak menambah keragaman informasi. Data yang berulang dalam suatu kelas dapat membuat ukuran dataset tampak lebih besar, tetapi berpotensi menyebabkan perhitungan yang berlebihan terhadap input pelatihan model (Mohammed et al., 2024). Oleh karena itu, dapat menjadi sebuah argumentasi bahwa dampak nyata terhadap performa model juga mempertimbangkan keberagaman data dalam setiap kelas, karena data yang berulang tidak selalu memberikan kontribusi yang efektif dalam proses pembelajaran model.

Confusion Matrix - Testing dataset

True Labels \ Predicted Labels	A16.2	C34.9	I25.1	J16.8	J18.9	J47	K01.1	NN
A16.2	47	2	0	3	4	4	0	3
C34.9	1	80	0	1	0	0	0	1
I25.1	0	0	58	0	0	0	0	8
J16.8	0	3	1	29	4	2	0	17
J18.9	2	0	0	1	39	2	0	8
J47	2	1	0	2	0	23	0	2
K01.1	0	0	0	0	0	0	69	1
NN	9	12	3	3	1	3	0	109

Gambar 4.12 *Confusion Matrix* menunjukkan performa model memprediksi setiap kelas

Namun, kinerja model tidak hanya dipengaruhi oleh ukuran populasi dataset. Penelitian Mohammed et al. (2024) menyoroti bahwa enam dimensi kualitas data yaitu *consistency*, *completeness*, *feature accuracy*, *uniqueness*, *target accuracy*, dan *target class balance* memainkan peran penting dalam performa model. Dalam penelitian ini, beberapa kelas dengan populasi kecil dapat menunjukkan performa yang lebih baik dibandingkan kelas dengan populasi lebih besar. Perbedaan ini bisa jadi disebabkan oleh kualitas data yang lebih tinggi pada kelas dengan populasi kecil. Sebagaimana terlihat dalam Tabel 4.7, kelas

NN dan J16.8 menunjukkan performa lebih rendah dibandingkan kelas lain dengan ukuran dataset yang lebih kecil.

Tabel 4.7 Evaluasi performa model untuk tiap kelas

Class	Precision (%)	Recall (%)	F1 Score (%)
K01.1	1.0000	0.9857	0.9928
I25.1	0.9206	0.8788	0.8992
C34.9	0.8163	0.9639	0.8831
NN	0.7315	0.7786	0.7543
A16.2	0.7705	0.7460	0.7580
J47	0.7188	0.7667	0.7420
J18.9	0.7500	0.7222	0.7358
J16.8	0.7632	0.5179	0.6170

Completeness mengacu pada keberadaan semua data yang relevan dan diperlukan, sedangkan *feature accuracy* menunjukkan sejauh mana nilai dalam dataset mencerminkan nilai sebenarnya, dan *consistency* adalah representasi teks yang tidak beragam untuk menunjukkan terminology semantik yang sama. Analisis terhadap dimensi *completeness*, *consistency* dan *feature accuracy* membutuhkan pengetahuan spesifik dalam domain tertentu, dalam kasus ini bidang medis. Namun demikian, setelah meninjau kembali Gambar 3.5 pada sub-bab 3.3 Dataset, terlihat bahwa kedua kelas tersebut mengalami masalah terkait *target accuracy*. Kelas J16.8, khususnya, banyak mengalami kerancuan pelabelan. Beberapa label yang berbeda diberikan pada sampel data yang sangat mirip. Ketidakkonsistenan dalam pelabelan ini berkontribusi pada ketidakakuratan model dalam mengidentifikasi kelas dengan benar (Shah et al., 2024), sehingga berdampak pada kinerja keseluruhan. fitur dengan kelas J16.8 yang menunjukkan kerancuan ini, yang dapat menghambat model mempelajari data secara efektif.

Di sisi lain, kelas NN merupakan kelas gabungan yang terdiri dari 17 *sub-class* yang lebih kecil, dan Gambar 3.5 menunjukkan bahwa kelas NN ini memiliki banyak data yang secara semantik menggambarkan penyakit yang serupa akan tetapi diberi label berbeda. Pada gambar tersebut terlihat data penyakit “CAP”, “Gea”, dan “Gagal napas” bercampur antara label NN, J16.8, A16.2 dan I25.1. Masalah ini menghasilkan noise yang signifikan, yang semakin mempersulit proses pembelajaran model klasifikasi. Menurut Mohammed et al. (2024), masalah kualitas data seperti konsistensi, akurasi target, dan keseimbangan kelas

dapat menurunkan kinerja model. Meskipun class NN memiliki jumlah data unik yang lebih besar, kurangnya kualitas dalam hal konsistensi pelabelan data dan keunikan fitur menyebabkan akurasi prediksi yang menurun. Hal ini juga mengonfirmasi bahwa dimensi kualitas data lebih berpengaruh dibandingkan sekadar ukuran dataset.

Pengamatan lebih lanjut mengungkap adanya potensi hubungan antara kinerja kelas dan distribusi ukuran token per kelas. Disebutkan pada penelitian sebelumnya, seperti Goldman et al. (2024), menunjukkan bahwa frekuensi panjang token dapat memengaruhi kinerja model klasifikasi, dan penelitian Xu et al. (2019) membahas dampak panjang token terhadap hasil model, penelitian (Yehezkel & Pinter, 2022) menunjukkan bahwa dataset dengan distribusi menyerupai bentuk distribusi normal pada plot histogram frekuensi panjang tokennya dan *skew* yang lebih rendah cenderung berkorelasi dengan performa yang lebih baik. Penelitian ini mengamati kecenderungan serupa berdasarkan data yang dianalisis, sebagaimana ditunjukkan dalam Gambar 4.7, di mana histogram distribusi panjang token pada beberapa kelas yang secara visual tampak menyerupai bentuk distribusi normal, cenderung memiliki oleh performa model yang relatif lebih baik. Namun, pengamatan visual ini masih bersifat eksploratif dan memerlukan analisis statistik lebih lanjut dalam studi mendatang untuk mengkonfirmasi hubungan antara distribusi panjang token dan performa model.

Temuan ini mengindikasikan bahwa tidak hanya ukuran dataset, tetapi juga representasi token yang seimbang dalam setiap kelas dapat berkontribusi positif terhadap performa model dalam memprediksi kelas tersebut. Kelas seperti K01.1, I25.1, dan C34.9 memiliki distribusi token yang mendekati distribusi normal, yang terlihat memiliki kinerja yang lebih tinggi. Sebaliknya, meskipun kelas NN menunjukkan distribusi frekuensi token yang lebih baik, seperti telah disebutkan sebelumnya, kelas ini mengalami noise yang signifikan serta inkonsistensi dalam fiturnya, yang pada akhirnya menghambat kinerjanya meskipun memiliki distribusi token yang lebih baik.

4.9 Implikasi dan Kontribusi Penelitian

Penelitian ini memiliki sejumlah implikasi praktis dan akademik yang signifikan. Dari sisi praktis, model klasifikasi berbasis BERT-BiGRU yang dikembangkan mampu menangani teks diagnosis yang ditulis secara bebas dalam bahasa Indonesia, sebagaimana umum dijumpai di RSUP Persahabatan dan rumah sakit lain yang sejenis. Hal ini menunjukkan bahwa penerapan model klasifikasi teks berbasis *deep learning* memiliki potensi untuk

meningkatkan akurasi proses pengkodean ICD-10, serta memperkuat validitas dalam pengajuan klaim BPJS, terutama pada rumah sakit yang menghadapi tantangan serupa.

Pendekatan pelabelan "NN" yang diterapkan dalam penelitian ini berperan sebagai solusi untuk menggabungkan kelas-kelas kecil yang memiliki jumlah data rendah ke dalam satu kelas agregat. Pendekatan ini memungkinkan model untuk tetap mempertahankan performa pada kelas-kelas utama, tanpa harus menghapus data dari kelas kecil ataupun memaksa pemetaan yang tidak akurat. Strategi ini mempertahankan keseimbangan antara representasi cakupan data yang luas dan tetap menjaga performa model.

Pendekatan ini berbeda dari yang digunakan oleh Parjono & Kusumadewi (2023), yang hanya memilih lima kelas ICD-10 utama dan mengecualikan data kelas lainnya. Meskipun pendekatan tersebut berhasil mencapai performa tinggi mencapai 86.8%, ia tidak mampu menangani prediksi untuk kelas-kelas di luar lima yang dipilih. Sementara itu, studi oleh Amin et al. (2021) justru menggunakan semua kelas tanpa agregasi, sehingga modelnya mengalami penurunan performa yang signifikan hingga akurasi 52% akibat penyebaran data yang tidak seimbang. Dengan mengadopsi kelas "NN", model dalam penelitian ini tetap dapat menangani data yang tidak termasuk dalam kelas-kelas utama secara lebih fleksibel, sekaligus mempertahankan kestabilan performa dan memberikan ruang untuk pengembangan model bertahap di masa depan.

Dari sisi akademik, penelitian ini memperluas bukti empiris mengenai efektivitas arsitektur BERT-BiGRU dalam menangani teks medis otentik berbahasa Indonesia, yang selama ini masih sedikit eksplorasi dalam literatur. Selain itu, observasi mengenai hubungan antara distribusi panjang token dan performa model memberikan wawasan awal yang dapat mendorong studi lebih lanjut tentang pengaruh variasi ukuran teks pada kelas yang sama terhadap hasil klasifikasi. Oleh karena itu, penelitian ini tidak hanya menyajikan solusi yang dapat langsung diterapkan di dunia nyata, tetapi juga membangun dasar pengetahuan yang relevan untuk riset lanjutan dalam domain NLP medis.

BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

Penelitian ini berhasil membangun model klasifikasi teks untuk melakukan kodefikasi diagnosis ke dalam kode ICD-10 menggunakan arsitektur BERT-BiGRU. Kombinasi ini memanfaatkan kemampuan BERT dalam menghasilkan representasi kontekstual dari teks medis serta kemampuan BiGRU dalam menangkap urutan dan dinamika teks diagnosis secara lebih baik. Evaluasi performa model dilakukan dengan menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy* untuk memberikan gambaran yang menyeluruh mengenai efektivitas model. Berdasarkan hasil evaluasi, model yang dikembangkan menunjukkan kinerja yang menjanjikan dengan akurasi sebesar 81,92%, *precision* 82,18%, dan *recall* 81,59%. Hasil ini mengindikasikan bahwa model mampu mengenali teks diagnosis secara efektif, sehingga berpotensi meningkatkan akurasi proses pengkodean serta memperbaiki sistem klaim asuransi kesehatan. Namun demikian, implementasi model ini di lingkungan rumah sakit seperti RSUP Persahabatan tetap memerlukan validasi lanjutan guna memastikan kesesuaian dengan praktik, alur kerja, dan kebutuhan spesifik di lapangan.

Evaluasi model mengungkapkan secara umum, populasi data yang lebih besar meningkatkan kemampuan prediksi model. Selain itu, kelas dengan pola plot histogram panjang token yang menyerupai distribusi normal, teramati cenderung memiliki kinerja yang lebih baik. Hal ini mengindikasikan keseimbangan representasi panjang token dapat menjadi faktor penting yang layak untuk dieksplorasi lebih lanjut. Sebaliknya, beberapa kelas menunjukkan kinerja rendah meskipun jumlah data besar. Pelabelan yang tidak konsisten terhadap teks yang mirip juga ditemukan dalam kelas-kelas ini, yang kemungkinan besar berkontribusi terhadap penurunan kinerja model.

5.2 Saran

Dengan akurasi model yang saat ini mencapai 81%, penelitian ini membuka peluang untuk pengembangan lebih lanjut. Berdasarkan temuan yang diperoleh, perhatian lebih harus diberikan pada peningkatan jumlah data, perbaikan akurasi pelabelan, serta pemeliharaan fitur yang jelas, konsisten, dan khas dalam dataset. Ketiga faktor ini terbukti memainkan peran penting dalam meningkatkan performa prediksi model klasifikasi.

Daftar Pustaka

- Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), Article 2. <https://doi.org/10.3390/app11020796>
- Amin, Z. A., Cholil, W., Herdiansyah, M. I., & Negara, E. S. (2021). Analisa Rekam Medis Elektronik Untuk Menentukan Diagnosa Medis Dalam Kategori Bab ICD 10 Menggunakan Machine Learning [Analysis of Electronic Medical Records to Determine Medical Diagnoses in ICD-10 Categories Using Machine Learning]. *POSITIF: Jurnal Sistem dan Teknologi Informasi*, *7*(2), 127–132. <https://doi.org/10.31961/positif.v7i2.1140>
- Anyaprita, D., Siregar, K. N., Hartono, B., Fachri, M., & Ariyanti, F. (2020). *Dampak Keterlambatan Pembayaran Klaim Bpjs Kesehatan Terhadap Mutu Pelayanan Rumah Sakit Islam Jakarta Sukapura [The Impact of BPJS Health Claim Payment Delays on Service Quality at Jakarta Sukapura Islamic Hospital]*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/62996>
- Arhami, M., F, F. Y. R., Hendrawaty, H., & Adriana, A. (2024). A Semantic Segmentation of Nucleus and Cytoplasm in Pap-smear Images using Modified U-Net Architecture. *JURNAL INFOTEL*, *16*(2), Article 2. <https://doi.org/10.20895/infotel.v16i2.1006>
- Ariyanti, F., & Gifari, M. T. (2019). Analisis Persetujuan Klaim BPJS Kesehatan pada Pasien Rawat Inap [Analysis of BPJS Health Claim Approval for Inpatient Patients]. *Jurnal Ilmu Kesehatan Masyarakat*, *8*(04), 156–166. <https://doi.org/10.33221/jikm.v8i04.415>
- Azam, S. S., Raju, M., Pagidimarri, V., & Kasivajjala, V. C. (2020). CASCADENET: An LSTM Based Deep Learning Model for Automated ICD-10 Coding. In K. Arai & R. Bhatia (Eds.), *Advances in Information and Communication* (pp. 55–74). Springer International Publishing. https://doi.org/10.1007/978-3-030-12385-7_6
- Batu, B. E. L. B. L., Saputra, W. A., & Sa'adah, A. (2024). Banana and Orange Classification Detection Using Convolutional Neural Network. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, *13*(3), Article 3. <https://doi.org/10.23887/janapati.v13i3.80032>

- Berutu, S. S., Budiati, H., Jatmika, J., & Gulo, F. (2023). Data preprocessing approach for machine learning-based sentiment classification. *JURNAL INFOTEL*, 15(4), Article 4. <https://doi.org/10.20895/infotel.v15i4.1030>
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, C., Zhang, P., Zhang, H., Dai, J., Yi, Y., Zhang, H., & Zhang, Y. (2020). Deep Learning on Computational-Resource-Limited Platforms: A Survey. *Mobile Information Systems*, 2020, e8454327. <https://doi.org/10.1155/2020/8454327>
- Chen, W., Fang, F., Wang, P., Kan, J., Li, W., & Wu, W. (2022). Research on Medical Text Classification based on BioBERT-GRU-Attention. *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, 213–219. <https://doi.org/10.1109/AEECA55500.2022.9919061>
- Desiani, A., Primartha, R., Hanum, H., Dewi, S. R. P., Suprihatin, B., Al-Filambany, M. G., & Suedarmin, M. (2024). Weighted Voting Ensemble Learning of CNN Architectures for Diabetic Retinopathy Classification. *JURNAL INFOTEL*, 16(1), Article 1. <https://doi.org/10.20895/infotel.v16i1.999>
- Duarte, F., Martins, B., Pinto, C. S., & Silva, M. J. (2018). Deep neural models for ICD-10 coding of death certificates and autopsy reports in free-text. *Journal of Biomedical Informatics*, 80, 64–77. <https://doi.org/10.1016/j.jbi.2018.02.011>
- Goldman, O., Caciularu, A., Eyal, M., Cao, K., Szpektor, I., & Tsarfaty, R. (2024). *Unpacking Tokenization: Evaluating Text Compression and its Correlation with Model Performance* (arXiv:2403.06265). arXiv. <https://doi.org/10.48550/arXiv.2403.06265>
- Guo, B., Zhang, C., Liu, J., & Ma, X. (2019). Improving text classification with weighted word embeddings via a multi-channel TextCNN model. *Neurocomputing*, 363, 366–374. <https://doi.org/10.1016/j.neucom.2019.07.052>
- Huang, Y., Dai, X., Yu, J., & Huang, Z. (2023). SA-SGRU: Combining Improved Self-Attention and Skip-GRU for Text Classification. *Applied Sciences*, 13(3), Article 3. <https://doi.org/10.3390/app13031296>

- Jiang, T., & Wang, Z. (2022). Text Classification Using BiGRU with Directional Self-Attention. *2022 11th International Conference of Information and Communication Technology (ICTech)*, 394–397. <https://doi.org/10.1109/ICTech55460.2022.00085>
- Kemkes, K. K. R. (2016). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 52 Tahun 2016 Tentang Standar Tarif Pelayanan Kesehatan Dalam Penyelenggaraan Program Jaminan Kesehatan [Regulation of the Minister of Health of the Republic of Indonesia Number 52 of 2016 on Standard Tariffs for Health Services in the Implementation of the Health Insurance Program]*. Jakarta.
- Li, X., Zhang, Y., Jin, J., Sun, F., Li, N., & Liang, S. (2023). A model of integrating convolution and BiGRU dual-channel mechanism for Chinese medical text classifications. *PLOS ONE*, *18*(3), e0282824. <https://doi.org/10.1371/journal.pone.0282824>
- Maimun, N. (n.d.). Factors Delayed of Insurance Claim Service Process (BPJS) at Annisa Pekanbaru Maternity Hospital. *Jurnal Kesehatan Komunitas (Journal Of Community Health)*.
- Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2024). *The Effects of Data Quality on Machine Learning Performance* (arXiv:2207.14529). arXiv. <https://doi.org/10.48550/arXiv.2207.14529>
- Moons, E., Khanna, A., Akkasi, A., & Moens, M.-F. (2020). A Comparison of Deep Learning Methods for ICD Coding of Clinical Records. *Applied Sciences*, *10*(15), Article 15. <https://doi.org/10.3390/app10155262>
- Nabila, S. F., Santi, M. W., Tabrani, Y., & Deharja, A. (2020). Analisis Faktor Penyebab Pending Klaim Akibat Koding Berkas Rekam Medis Pasien Rawat Inap di RSUPN Dr. Cipto Mangunkusumo [Analysis of Factors Causing Pending Claims Due to Medical Record Coding for Inpatient Cases at Dr. Cipto Mangunkusumo National General Hospital]. *J-REMI: Jurnal Rekam Medik dan Informasi Kesehatan*, *1*(4), 519–528. <https://doi.org/10.25047/j-remi.v1i4.2157>
- Ni, P., Li, G., Hung, P. C. K., & Chang, V. (2021). StaResGRU-CNN with CMedLMs: A stacked residual GRU-CNN with pre-trained biomedical language models for predictive intelligence. *Applied Soft Computing*, *113*, 107975. <https://doi.org/10.1016/j.asoc.2021.107975>

- Ouyang, S., Shao, Y., Du, J., & Li, A. (2022). *Scientific and Technological Text Knowledge Extraction Method of based on Word Mixing and GRU* (arXiv:2203.17079). arXiv. <https://doi.org/10.48550/arXiv.2203.17079>
- Parjono, P., & Kusumadewi, S. (2023). Pemodelan Text Mining dalam Pengkodean Penyakit Pasien Berdasar Kode ICD 10 [Text Mining Modeling in Disease Coding Based on ICD-10 Codes]. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(2), 200–207. <https://doi.org/10.25077/TEKNOSI.v9i2.2023.200-207>
- Parwez, Md. A., Fazil, Mohd., Arif, M., Nafis, M. T., & Auwul, Md. R. (2023). Biomedical Text Classification Using Augmented Word Representation Based on Distributional and Relational Contexts. *Computational Intelligence and Neuroscience*, 2023, 1–22. <https://doi.org/10.1155/2023/2989791>
- Prabhakar, S. K., & Won, D.-O. (2021). Medical Text Classification Using Hybrid Deep Learning Models with Multihead Attention. *Computational Intelligence and Neuroscience*, 2021, e9425655. <https://doi.org/10.1155/2021/9425655>
- Pradani, E. A., Lelonowati, D., & Sujianto, S. (2017). Keterlambatan Pengumpulan Berkas Verifikasi Klaim BPJS Di RS X: Apa Akar Masalah Dan Solusinya? [Delay in Collecting BPJS Claim Verification Documents at Hospital X: What Are the Root Causes and Solutions?]. *Jurnal Medicoeticolegal dan Manajemen Rumah Sakit*, 6(2), 112–121.
- Qin, Q., Zhao, S., & Liu, C. (2021). A BERT-BiGRU-CRF Model for Entity Recognition of Chinese Electronic Medical Records. *Complexity*, 2021, e6631837. <https://doi.org/10.1155/2021/6631837>
- Rifandi, M., & Shofiani, P. (2019). Pengaruh Sistem Pembayaran Asuransi Kesehatan BPJS Terhadap Akuntansi Pendapatan Rumah Sakit (Studi Pada RS PKU Muhammadiyah Gamping) [The Impact of the BPJS Health Insurance Payment System on Hospital Revenue Accounting (Study at PKU Muhammadiyah Gamping Hospital)]. *RELASI : JURNAL EKONOMI*, 15(1), 51–68. <https://doi.org/10.31967/relasi.v15i1.301>
- Rios, A., & Kavuluru, R. (2015). Convolutional neural networks for biomedical text classification: Application in indexing biomedical articles. *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, 258–267. <https://doi.org/10.1145/2808719.2808746>
- Rohman, H., Wintolo, A., & Susilowati, E. (2021). Analisis Penundaan Pembayaran Klaim pada Sistem Vedika Badan Penyelenggara Jaminan Sosial Kesehatan di Rumah Sakit Nur Hidayah Yogyakarta [Analysis of Claim Payment Delays in the Vedika System

- of the Social Security Administering Agency for Health at Nur Hidayah Hospital, Yogyakarta]. *Jurnal Wiyata: Penelitian Sains dan Kesehatan*, 8(1), 72–83. <https://doi.org/10.56710/wiyata.v8i1.254>
- Sakinah, A., & Yunengsih, Y. (2021). Analisis Kelengkapan Resume Medis Rawat Inap BPJS Terhadap Kelancaran Klaim BPJS Kesehatan di Rumah Sakit Melania Bogor [Analysis of the Completeness of BPJS Inpatient Medical Resumes on the Smoothness of BPJS Health Claims at Melania Hospital, Bogor]. *Jurnal Kesehatan Tambusai*, 2(4), 70–77. <https://doi.org/10.31004/jkt.v2i4.2665>
- Sari, I. P. P., Yulianis, F., & Aini, R. (2023). Analisis Pengakuan Pendapatan Klaim Asuransi Kesehatan Berdasarkan Psak 23 Pada Rumah Sakit Umum (RSU) ‘Aisyiyah Padang [Analysis of Revenue Recognition for Health Insurance Claims Based on PSAK 23 at ‘Aisyiyah Padang General Hospital]. *Jurnal Menara Ekonomi : Penelitian dan Kajian Ilmiah Bidang Ekonomi*, 9(2). <https://doi.org/10.31869/me.v9i2.4855>
- Sarkar, D. (2019). *Text Analytics with Python: A Practitioner’s Guide to Natural Language Processing*. Apress. <https://doi.org/10.1007/978-1-4842-4354-1>
- Shah, V., Parashos, T., & Kumar, A. (2024). How Do Categorical Duplicates Affect ML? A New Benchmark and Empirical Analyses. *Proc. VLDB Endow.*, 17(6), 1391–1404. <https://doi.org/10.14778/3648160.3648178>
- Shreyashree, S., Sunagar, P., Rajarajeswari, S., & Kanavalli, A. (2022). BERT-Based Hybrid RNN Model for Multi-class Text Classification to Study the Effect of Pre-trained Word Embeddings. *International Journal of Advanced Computer Science and Applications*, 13(9). <https://doi.org/10.14569/IJACSA.2022.0130979>
- Sourav, M. S. U., Wang, H., Mahmud, M. S., & Zheng, H. (2023). *Transformer-based Text Classification on Unified Bangla Multi-class Emotion Corpus* (arXiv:2210.06405). arXiv. <https://doi.org/10.48550/arXiv.2210.06405>
- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? In M. Sun, X. Huang, H. Ji, Z. Liu, & Y. Liu (Eds.), *Chinese Computational Linguistics* (pp. 194–206). Springer International Publishing. https://doi.org/10.1007/978-3-030-32381-3_16
- Tama, A. I., & Kusmiyatno, K. (2019). Kajian Kemampuan Keuangan Rumah Sakit Umum Daerah Sebagai Badan Layanan Umum Dan Pengaruhnya Terhadap Kualitas Pelayanan [Study on the Financial Capability of Regional General Hospitals as Public Service Agencies and Its Impact on Service Quality]. *Jurnal Penelitian Teori & Terapan Akuntansi (PETA)*, 4(1), 1–15. <https://doi.org/10.51289/peta.v4i1.374>

- Tian, Y. (2023). Multi-label Text Classification Combining BERT and Bi-GRU Based on the Attention Mechanism. *Journal of Network Intelligence*.
- Trisetyawan, R., & Eryani, S. (2020). Hubungan Ketepatan Kode Diagnose Dyspepsia Dengan Klaim BPJS di Rumah Sakit Rafflesia Bengkulu [The Relationship Between the Accuracy of Dyspepsia Diagnosis Codes and BPJS Claims at Rafflesia Hospital, Bengkulu]. *Mitra Raflesia (Journal of Health Science)*, 10(2). <https://doi.org/10.51712/mitrarafflesia.v10i2.10>
- Udin, I. A. (2022). Factors Causing Delayed Claims At The Hospital In Collaboration With Health Social Security Agency Branch Office Of Tasikmalaya. *Proceeding of International Conference Sustainable Competitive Advantage*, 3(0), Article 0.
- Wang, S., Huang, M., & Deng, Z. (2018). Densely Connected CNN with Multi-scale Feature Attention for Text Classification. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 4468–4474. <https://doi.org/10.24963/ijcai.2018/621>
- Wiyanto, A. (2023, Desember). *Business Process Model and Notation* [Personal Communication/Interview]. Student Castle Yogyakarta.
- Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access*, 7, 51522–51532. IEEE Access. <https://doi.org/10.1109/ACCESS.2019.2909919>
- Yastori, Mrs. (2022). Cases of Dispute and Pending Claims in Hospitals in the Era of National Health Insurance. *Proceeding International Conference on Medical Record*, 2(1), 32–38. <https://doi.org/10.47387/icmr.v2i1.152>
- Yehezkel, S., & Pinter, Y. (2022). *Incorporating Context into Subword Vocabularies*. <https://doi.org/10.48550/arXiv.2210.07095>
- Yu, Q., Wang, Z., & Jiang, K. (2021). Research on Text Classification Based on BERT-BiGRU Model. *Journal of Physics: Conference Series*, 1746, 012019. <https://doi.org/10.1088/1742-6596/1746/1/012019>
- Yussianto, A. (2023, September 4). *Business Process Model and Notation* [Personal Communication/Interview]. zoom.
- Yussianto, A. (2024, January 5). *Hubungan Causal* [Personal Communication/Interview]. Student Castle Yogyakarta.
- Zhang, X., Wu, Z., Liu, K., Zhao, Z., Wang, J., & Wu, C. (2023). Text Sentiment Classification Based on BERT Embedding and Sliced Multi-Head Self-Attention Bi-GRU. *Sensors*, 23(3), Article 3. <https://doi.org/10.3390/s23031481>.

LAMPIRAN A

Wawancara Model Bisnis RSUP Persahabatan dalam konteks klaim BPJS

Interview Business Process Model and Notation	
Penelitian	Pembuatan Model Klasifikasi Text Diagnosa Utama Dengan Deep Learning BERT dan BiGRU
Narasumber	dr. Andi Yussianto, S.H., M.Epid
Expertise	Direktur Pelayanan Medis RS Persahabatan (2022 s.d sekarang) Kepala Balai Besar Laboratorium Kesehatan Palembang (2020 sd 2022) Kepala Bagian Pendidikan dan Pelatihan Diklat RS Cipto Mangunkusumo (2014 sd 2018) Kepala Bagian SDM RS Cipto Mangunkusumo (2018 sd 2019) Kepala Bidang Penunjang Medik RSUP Cipto Mangun Kusumo (2019 s.d 2020)
Lokasi	RSUP Persahabatan Jakarta
Tanggal	4 September 2023
Topik	Alur/proses bisnis pencatatan dan penagihan BPJS

Transkrip Intevieiw

Q : Assalamualaikum Bapak Direktur Pelayanan Medik Dokter Andi Yussianto, Selamat siang. Terima kasih banyak Dokter atas kesediaan dan waktunya, ditengah kesibukan tugasnya.

Dokter Andi, saya izin untuk mengulang kembali tentang isu yang dulu pernah kita singgung, yaitu potensi kendala pada penagihan klaim BPJS yang konon sangat mungkin menimpa semua rumah sakit yang bekerja sama dengan BPJS Kesehatan.

ER : Iya pak sama-sama. Saya juga berterima kasih ni, pak Aqge jadi berkunjung ke RSUP, sangat sempat terfikir jadi tidak ini penelitiannya. Alhamdulillah hari ini sudah di RSUP. Nanti kita perkenalkan juga ya dengan tim di RSUP, jam 10 saya mengundang beberapa tim saya untuk mengawali ini.

Jadi bagaimana ini rencananya? Jadi kan mengangkat isu ini?

Q : Waduh dok, sampai memperkenalkan dengan tim, saya jadi sungkan.

Tapi iya betul Pak Dokter Andi, saya tertarik untuk mengangkatnya sebagai topik penelitian. Justru karena nya saya ingin mengawali untuk untuk mendapatkan informasi awal tentang seperti apa proses penagihan BPJS ini.

ER : Tidak apa-apa mas, silahkan.

Q : Iya baik pak, Jika tidak salah dokter menyebutkan, pihak rumah sakit sering dirugikan karena klaim BPJS itu tertunda. Masalah lainnya penagihan itu kadang berada dibawah nilai yang seharusnya kita terima.

Sebagai awalan ini saya berharap bisa memahami proses yang terjadi di rumah sakit, yang mempengaruhi penagihan klaim tersebut.

ER : Jadi begini, BPJS itu menyiapkan mekanisme penagihan klaim pasien peserta BPJS yang namanya **INA CBG's**. Regulasinya ada, panduannya ada, dan aplikasinya juga sudah disiapkan.

Kami sebagai rumah sakit yang bekerja sama, setiap periode penagihan melengkapi data penagihan itu lewat INA CBG's. Dan itu ada kode-kodenya. Ada kelompok pembiayaannya. Kelompok itu yang menentukan BPJS, nah kita tinggal melengkapi datanya. Untuk kasus RSUP Persahabatan, kita sudah didukung oleh Sistem informasi sehingga kita tinggal mengumpulkan data yang dibutuhkan dari SIM RS tersebut.

Nah kerugiannya itu yang terjadi adalah ketika kita salah dalam memberikan data penagihan. Kode yang kita pilih itu keliru sehingga pembayaran yang kita terima lebih rendah dari tindakan atau pemeriksaan yang diterima pasien.

Q : Baik pak, Penagihan dan pengumpulan data itu, dilakukan oleh bagian apa Pak Andi?

ER : Penagihannya sih tanggung jawabnya bagian keuangan, ada sendiri divisi untuk penagihan klaim, tetapi itu sub dari **Direktorat Perencanaan dan keuangan**. Untuk melengkapi data-datanya itu dilakukan oleh **Tim Koding** namanya pak. Teman-teman dari **rekam medis**.

Q : Bagian keuangan saya sudah sering mendengarnya ini Pak Andi, tetapi tim koding ini apa ya? Apakah mereka melakukan Koding pemrograman begitu pak?

ER : Ya itu! Mereka yang mengisikan data kodefikasi tindakan dan diagnosanya pak.

Begini mas, dari **dokter** maupun dari **penunjang medis**, data yang mereka masukkan kedalam SIM RS dalam bentuk isian text.

Q : Free text maksudnya pak?

ER : Sebentar saya jelaskan dulu.

Dokter, lab, dan radiologi menginputkan data kedalam SIM RS dalam bentuk text. Lab dan Radiologi bisa jadi sudah dibakukan dalam bentuk pilihan dan isian ya, tetapi jika dokter setahu saya itu free text. Nah kemudian data itu perlu kita kodekan agar bisa dikelompokkan padap INA CBG's.

Q : Ok pak, sepertinya saya sudah mulai dapat gambaran. Ada data yang bersumber dari dokter, pada poliklinik ya pak, dan ada data yang bersumber dari lab, ataupun radiologi. Keduanya menjadi acuan untuk Tim Koding ketika akan memberikan kode, begitu ya pak.

ER : Sebetulnya yang dikodekan itu adalah kode untuk diagnosa penyakitnya pak, namanya kode ICD-10. Ada dua nih kode diagnosa yang diterima oleh INA CBG's kode **diagnosa utama**, dan kode **diagnosa sekunder** keduanya menggunakan kode ICD-10. Dan yang ketiga adalah Kodefikasi untuk **tindakan medis** yang diterima pasien. Kode ICD-9 kalo tidak salah, yang hafal teman-teman rekam medis. Nah data Lab, radiologi ataupun penunjang lainnya ini mendukung diagnosa dari dokter.

Q : Jadi data hasil lab maupun radiologi itu tidak perlu dikodekan oleh teman-teman dari rekam medis ya pak?

ER : Tidak, justru data pemeriksaan dari penunjang itu menjadi pertimbangan dokter untuk menuliskan diagnosa utamanya maupun diagnosa sekundernya. Ya logikanya mestinya juga menjadi pendukung pilihan tindakannya juga.

Tetapi begini mas, terkadang dokter itu fokus pada pencatatan gejala dan kecurigaan diagnosa, sedangkan diagnosa sekundernya itu kurang diperhatikan, padahal kode penagihan itu bisa kita naikan jika berbeda diagnosa sekundernya. Disinilah **verifikator** itu perlu memeriksa juga data hasil labnya atau profil pasiennya.

Q : Verifikator pak?

ER : Iya, Instalasinya namanya **Instalasi Verifikasi Penjamin Pasien** disingkat IVPP, ini adanya di direktorat keuangan. Mereka memverifikasi diagnosa dan treatment isian dokter kemudian menyesuaikan penyebutan diagnosanya. Tindakan biasanya sudah tidak masalah.

Q : Verifikator ini kompeten pak untuk melakukan itu

ER : Ada dokter yang melakukan itu dan memang kompeten untuk menuliskan diagnosa. Kemudian ini yang dibaca oleh tim koder untuk menginputkan kodefikasinya.

Q : Bisa dianggap text isian dari IVPP ini sudah lebih standar dari isian asli dokter dan pelayanan lainnya ya pak.

ER : Iya, tetapi tetap masih masalah karena pengisian-pengisian ini masih dilakukan manual. Termasuk Koding ny aitu juga manual.

Q : Oh Ok pak. Output dari tim koding ini, apakah sudah cukup untuk menjadi sumber penagihan oleh teman di bagian keuangan?

ER : Iya, diagnosa dan tindakan yang sudah dilengkapi koding nya itu kembali lagi ke IVPP untuk disubmit, tetapi kadang ada beberapa dokumen yang perlu disertakan sebelum penagihan. Datanya sudah ada di rekam medis SIM RS, hanya saja misalnya hasil scannya perlu dilengkapi juga.

(Diskusi tambahan tentang komposisi Tim Koding...)

ER : Ini jika sudah selesai programnya, bisa kita terapkan pada Rumah sakit langsung atau bagaimana ya nanti.

Q : Mm.. dalam penelitian ini output yang terdekat bukan program pak, ya semacam library mungkin pak, kita biasanya menyebut sebagai model. SIM RS yang berjalan saat ini perlu melakukan penyesuaian untuk bisa memanfaatkan model tersebut.

ER : Tapi bisa kan

Q : Bisa pak, jika Tim siap saya bisa mengarahkan dan mendampingi. Tapi ini tidak semua Code ICD-10 mampu ditebak oleh modelnya. Biasanya sangat terpengaruh oleh jumlah data yang tersedia.

ER : Perlu data banyak ya mas?

Q : Setahu saya perlu banyak Pak Andi, bahkan kecenderungannya adalah semakin banyak jumlah populasi datanya, akan semakin bagus kemampuan model tersebut.

ER : Kita berharapnya, meskipun ini penelitian ya, tetapi Rumah sakit bisa dapat manfaatnya, jika memang tidak semua ICD bisa ditebak, maka bisa kita ambil yang bagus-bagus saja dulu. Nanti jika itu bisa terbukti baik bisa kita lanjutkan untuk memperbaiki dengan kode ICD sisanya.

Untuk penelitian ini kira-kira bisa berapa ya? 20 an ICD mungkin ya?.

Q : Wah dapat sepuluh saja sangat beruntung Pak Andi, dan itu juga akurasi jarang bisa 100%, antara 80 - 95 persen sudah sangat baik. Pembuatan model atau model training ini ada banyak metode Pak Andi, dan metode yang saya pilih ini dari beberapa referensi menghasilkan performa model yang memuaskan, tetapi hasilnya memang harus kita riset dulu.

ER : Oo Begitu ya..., tapi ini makin sering kita pakai, tingkat akurasi akan semakin baik kan ya?

(Diskusi tambahan tentang komposisi AI supervised, non-supervised, dan reinforcement ...)

Q : Ok Pak Andi, Izin Kembali ke SIM RS tadi Pak, apakah semua pasien itu datanya diinputkan kedalam SIM RS pak, maksud saya mungkin ada perlakuan khusus untuk pasien peserta BPJS dan bukan peserta.

ER : Lho iya dong, semua pasien harus masuk kedalam SIM RS, sudah standar operasional pelayanannya seperti itu. Lha kita bisa tahu profil pasien atau riwayat penyakitnya bagaimana jika tidak semua kita rekam. Hanya saja dari awal pasien masuk itu sudah bisa dikenali apakah ini peserta BPJS atau bukan peserta BPJS.

Tapi begini, RSUP persahabatan itu kan rumah sakit rujukan nasional, sehingga hampir sebagian besar pasiennya itu adalah peserta BPJS yang dirujuk dari rumah sakit lainnya.

Q : SIM RS ini sudah beroperasi berapa lama ya Pak Andi?

ER : Sudah lama ya, sebelum saya dipindahkan ke RSUP persahabatan ini SIM RS nya sudah ada.

Q : iya pak, maksud pertanyaan saya sebetulnya ingin mengukur kumulasi data yang tersimpan pada database nya apakah sudah cukup banyak Pak Andi. Agar bisa didekati dengan solusi data science.

ER : Banyak kok mas, belum jutaan sih tapi banyak kok mungkin ribuan atau bahkan ratusan ribu ya. Memang butuh data berapa banyak untuk AI data science itu.

(Diskusi tambahan tentang data science dan AI...)

Q : Baik Pak Dir Yanmed Dokter Andi, saya coba susun alur pengumpulan data yang berjalan dalam rangka data pendukung penagihan klaim BPJS ini. Kemudian saya coba konfirmasi lagi benar atau tidak nya.

Ini penting bagi kami pak, karena untuk menguatkan pemilihan solusi data science itu relevan atau tidak untuk menyelesaikan kasus penagihan BPJS tersebut.

ER : Iya pak tidak masalah, dibuat saja dan silahkan dikonfirmasi. Ke saya langsung bisa, ke tim saya juga bisa. Oh ya ini kembali ke jogjanya hari ini? Jam berapa? Saya perkenalkan dulu ke tim saya agar yang lain juga tahu kegiatan penelitian kita ini.

Q : Dari Halim jam 13.30 pak, insyaallah bisa Pak Andi.

(Diskusi tambahan tentang topik lain ..)

(Dan Perkenalan dengan tim RSUP persahabatan tentang isu data science ..)

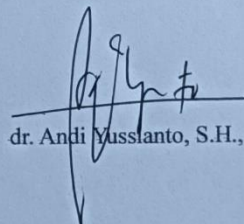
Yogyakarta, 5 Januari 2024

Peneliti

Narasumber



Aqge Priwibowo, ST



dr. Andi Yusslanto, S.H., M.Epid

Wawancara Model Bisnis RSUP Persahabatan dalam konteks klaim BPJS

Interview Business Notification and Model	
Penelitian	Pembuatan Model Klasifikasi Text Diagnosa Utama Dengan Deep Learning BERT dan BiGRU.
Nara Sumber	Agus Wiyanto, S.Kom.
Expertise	Kepala Sub Pelayanan Instalasi Sistem Informasi Manajemen RSUP Persahabatan (2020 s.d -)
Lokasi	Zoom meeting
Tanggal	20 Desember 2023
Topik	Alur/proses bisnis pencatatan dan penagihan BPJS
Notasi	Q : Question, ER : Expert Response

Transkrip Interview

Q : Assalamualaikum Pak Agus, Selamat pagi. Terima kasih banyak Pak Agus atas kesediaan dan waktunya, ditengah kesibukan tugasnya.

ER : Tidak apa Pak, ini memang sudah kami agendakan untuk pertemuannya. Kita menunggu dokter Andi dulu ya Pak mungkin 10 menit kedepan.

Q : Baik Pak, kita tunggu sebentar. Coba sambil saya kirim WA.

(Setelah menunggu beberapa waktu...)

ER : Pak kita mulai saja Pak, sepertinya Bapak Diryanmed belum bisa bergabung dengan kita. Silahkan Pak jika ada yang akan disampaikan dari Pak Aqge.

Q : Baik pak Agus, terima kasih atas kesempatannya. Sepertinya ini singkat saja Pak Agus. Saya coba menyusun urutan proses untuk penagihan ini, dalam rangkat memahami alur terbentuknya data untuk pengkodean diagnose ini.

ER : Baik Pak Silahkan.

Q : Setahu saya yang kita kodekan itu adalah data rekam medis yang sdh tersimpan di database SIM RSUP ya pak Agus.

ER : Iya Pak, itu adalah dalam dari SIM RS.

Q : Itu artinya mulai dari awal pendaftaran pasien sudah diidentifikasi sebagai pasien BPJS ya?

ER : Iya betul Pak, dari pendaftaran oleh petugas rekam medis di bagian pendaftaran data pasien sudah yang baru terdaftar ataupun pasien lama, sudah bisa dikenali sebagai pasien BPJS atau non BPJS.

Selanjutnya, pemeriksaan di poliklinik ditambahkan oleh dokter biasanya dibantu perawat di poliklinik.

Q : Apakah itu termasuk data diagnosa dan resep dokternya pak?

ER : Iya, Setahu saya data diagnosa penyakit dan terapi atau tindakan medis, dan obat juga, itu masuk ke dalam SIM RS saat pemeriksaan dokter di poliklinik.

Q : Eem.. ini Pak Agus, jika pasien itu harus periksa darah atau misalnya perlu rontgen, setahu saya juga bagian dari rekam medis ya. Nah itu masuk ke sebagai data SIM RS-nya seperti apa Pak?

ER : Oh itu tergantung Pak, tidak semua pemeriksaan terjadi seperti itu. Tergantung dokternya memerintahkan untuk periksa lab atau radiologi.

Q : Iya pak, berarti datanya bukan dikelola oleh dokter ya Pak ?

ER : Bukan, petugas Laboratorium dan Petugas Radiologi sudah punya akses sendiri Pak ke dalam SIM RS. Sehingga mereka bisa langsung menginputkannya. Justru dokter sudah bisa melihat inputan dari lab ataupun radiologi, sebelum berkasnya dibawa ke poliklinik.

(Diskusi tambahan tentang teknologi SIM RS, dan Database...)

Q : Kita kembali lagi ya pak Agus, Proses koding atau pemberian kode diagnosa itu sendiri terjadinya kapan Pak Agus?

ER : Tim memberikan kode diagnosa itu setelah semua data diagnosa primer, sekunder dan data tindakan diverifikasi oleh verifikator pak. IVPP akan memverifikasi diagnosa dan tindakan dari dokter, dengan melihat informasi tambahan seperti data lab, radiologi, profil pasien bahkan mungkin riwayat medis pasien. Nanti diagnosa yang sudah diverifikasi ini yang akan di kodekan oleh tim koder.

Q : Apakah IVPP sebagai verifikator ini bisa saja mengganti atau menambahkan isian diagnosa nya pak Agus?

ER : Iya pak, tugas verifikatornya memang seperti itu, kadang diagnosa isian dari dokter itu kurang spesifik untuk dikodekan, atau hanya mengisikan gejala saja. IVPP ini lah yang melengkapkannya agar bisa dikodekan oleh tim.

Q : Jadi proses koding itu memang setelah semua kegiatan entry data rekam medis selesai pak.

ER : Untuk selanjutnya, Tim Koder menginputkan Kode ICD-10 dan ICD-9 sesuai dengan diagnosa yang sudah diverifikasi oleh IVPP tadi pak. Informasi Kode ini kemudian kembali lagi ke IVPP tetapi sudah tinggal submit. Atau akhirnya lah, dikembalikan lagi ke IVPP.

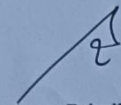
Q : Aplikasi INA CBG's ini bagian dari SIM RS?

- ER : Bukan itu sudah disiapkan oleh BPJS, kita tinggal melengkapinya.
- Q : baik Pak Agus terima kasih banyak, mohon maaf ini bagian dari persyaratan penelitiannya untuk mendapatkan alur proses bisnisnya, sehingga saya harus merepotkan Pak Agus. Matur suwun ya pak.
- ER : Iya pak tidak masalah, tapi nanti saya juga minta diagram alur yang sudah disusun itu untuk kami juga.
- Q : Siap pak segera saya kirim. Via WA tidak masalah ya pak.
- ER : Ok pak.

Jakarta, 30 Mei 2024

Peneliti

Narasumber



Aqge Priwibowo, ST



Agus Wiyanto, S.Kom

Wawancara Hubungan sebab-akibat tentang Klaim BPJS

Interview hubungan sebab akibat	
Penelitian	Pembuatan Model Klasifikasi Text Diagnosa Utama Dengan Deep Learning BERT dan BiGRU
Nara Sumber	Dr Andi Yussianto, S.H., M.Epid
Expertise	Direktur Pelayanan Medis RS Persahabatan (2022 s.d sekarang) Kepala Balai Besar Laboratorium Kesehatan Palembang (2020 sd 2022) Kepala Bagian Pendidikan dan Pelatihan Diklat RS Cipto Mangunkusumo (2014 sd 2018) Kepala Bagian SDM RS Cipto Mangunkusumo (2018 sd 2019) Kepala Bidang Penunjang Medik RSUP Cipto Mangun Kusumo (2019 s.d 2020)
Lokasi	Student Park Hotel Yogyakarta
Tanggal	5 Januari 2024
Topik	Hubungan sebab akibat kodifikasi diagnosa pada penagihan BPJS
Notasi	Q : Question, ER : Expert Response

Transkrip Intevieiw

Q : Assalamualaikum Dokter Andi, dokter ini saya ada beberapa hal yang perlu saya diskusikan pak, terkait dengan sebab dan akibat sekitar kesalahan pada koding **INA CBG's**. Apakah bisa kita lanjutkan lagi.

ER : Sebab akibat maksudnya bagaimana ya?

Q : Dari diskusi awal yang lalu pak, tentang dampak dari kesalahan **mengkodekan diagnosa** pada proses penagihan BPJS tersebut, mungkin bisa kita identifikasi lebih jauh dan rinci

ER : Iya bagaimana

Q : Seperti ini pak, kemaren pak Andi menyebutkan kesalahan dalam coding diagnosa itu kadang bisa karena up-claim, atau yang dikhawatirkan justru bisa juga down-claim. Keduanya berpotensi untuk memicu masalah.

ER : Jika koding **down-claim**, rumah sakit akan dirugikan karena pembayaran yang diperoleh akan berkurang dari yang seharusnya, dimana hal ini biasanya dapat menimbulkan defisit bagi Rumah Sakit dengan biaya operasional yang dikeluarkan Rumah Sakit untuk menangani pasien dengan diagnosa tersebut . Nah jika **up-claim**, bisa menimbulkan fraud. Dan resiko yang akan dihadapi bisa lebih jauh dari sekedar pengembalian dokumen oleh BPJS (dispute atau pending).

Q : BPJS menolak penagihannya ya pak? Berarti salah pengkodean yang bukan up-claim maupun down-claim itu juga masuk ya pak?

ER : Bisa tapi itu kasus jarang jika sampai salah mengkodekan, yang sering ya up-claim atau down-claim itu.

Q : Itu apakah bisa dimunculkan dalam bentuk persentasi pak? Maksudnya kasus yang sering terjadi itu up-claim itu seberapa banyak, yang down-claim seberapa banyak, dan juga yang salah tadi.

ER : Yang sering terjadi itu down-claim, ya bisa dikatakan kesalahan dalam koding itu **80%** adalah kasus down-claim, untuk up-claim itu jarang muncul hanya sekitar **20%**-an kasus up-claim.

Q : Yang **miss-claim** nya bagaimana pak Andi, seberapa banyak terjadinya?

ER : Tidak sampai 3%, kalo miss claim itu tidak sampai 3% mungkin lebih sedikit.

Q : Apakah bisa disimpulkan untuk down-claim itu **80%**, up-claim **17%**, dan miss nya itu **3%** ya pak?

ER : Ya secara umum bisa dikatakan demikian

(Diskusi lainnya yang tidak berhubungan...)

Q : Baik pak andi, izin ini pak kita kembali lagi ke permasalahan koding diagnosa untuk INA-CBG's tadi pak andi. Begini pak, Apakah kesalahan koding itu merupakan **faktor utama** dalam masalah penagihan BPJS.

ER : Oh tidak, adalah masalah kelengkapan dokumen juga. Menagihkan BPJS itu selain input kodefikasi pada INA-CBG's juga perlu menyertakan **dokumen pendukungnya**.

Q : Oh baik pak, saya hanya coba memastikan saja, maksud saya pada kegiatan kodefikasi diagnosa di INA CBG's itu, apakah selama ini masalahnya hanya seputar kesalahan pengkodean?

ER : Iya pak, jika kodenya benar kadang perlu waktu lama. Nah ini, jika kita ngomongin efisiensi, maka untuk waktu mengkodekan itu perlu kita minimalkan. Ini kan banyak ya mas ya sedangkan penagihan BPJS itu ada periodenya.

Q : Seberapa besar masalah **penundaan** atau **penambahan** waktu untuk koding ini pak? Maksud saya untuk menggambarkan porsinya dibandingkan dengan masalah akibat kesalahan koding.

ER : Hampir **seperti tiga** masalah yang terjadi pada penagihan BPJS ini karena masalah waktunya, penginputnya kodingnya karena perlu menterjemah dan sebagaimana menjadi tertunda.

Q : Koding itu oleh satu petugas atau banyak pak?

ER : **Tim** rekam medis. Ada sendiri **tim koding** dibawah Rekam medis.

Q : Mungkin tidak pak Andi, sudah waktunya lama, ternyata kodingnya juga masih belum tepat?

ER : Ya Adalah.. Nah harapannya dengan tool alat bantu ini, dapat jadi salah satu pilihan yang bisa membantu.

Q : Kemudian pak, jika ini tim, bukankah tidak sama skill nya maupun **basis pengetahuannya** tentang diagnose? Bagaimana ini pak?

ER : Ada ini mas, ada **validatornya**. Dan itu dokter. Tapi memang iya, beda petugas kadang muncul perbedaan. Ini juga salah satu alasan saya mendorong penelitian untuk tool AI nya.

Q : Apapun output penelitiannya saya sangat terbuka pak untuk pihak Rumah sakit. Insyaallah..

(Diskusi lainnya yang tidak berhubungan...)

Q : Pada beberapa tulisan yang saya baca pak, penundaan klaim pembayaran itu katanya berdampak merugikan untuk rumah sakit dan seperti sebuah kerugian besar yang dikhawatirkan begitu. Tetapi seperti apa dampaknya saya belum bisa menangkapnya dengan jelas boleh mohon penjelasan dari pak andi?

ER : Lha iya kan pak, jika pembayaran klaim tertunda, itu kan **income** ya bagi rumah sakit, sehingga **operasionalnya** akan terpengaruh. Rumah sakit kan perlu membayar operasional, nah jika operasional itu terganggu maka **pelayanan** juga terpengaruh. Kita ini kan rumah sakit pemerintah ya, jadi tidak fokus pada profit melainkan pada pelayanan. Sedangkan pelayanan itu membutuhkan biaya operasional.

Q : Jika pelayanan itu berhubungan dengan anggaran operasional, apakah berarti jika anggarannya rendah maka pelayanannya juga akan berkurang pak?

ER : Secara umum iya, tetapi ada **regulasi** yang mengatur. Mungkin rumah sakit swasta akan terpengaruh langsung, tetapi untuk rumah sakit pemerintah harus sesuai dengan regulasinya.

Q : Mana yang lebih dominan pak, regulasi atau dampak anggaran di kualitas pelayanan.

ER : Keduanya mas. Mungkin bisa dikatakan lebih dari **setengah** kualitas pelayanan itu ditentukan oleh regulasi, sisanya oleh anggaran, apalagi jika sudah BLU.

Q : Pendapatan rumah sakit ini, apakah hanya bersumber dari klaim BPJS dan asuransi lainnya pak?

ER : Ya tidak, ada **anggaran** dari **pemerintah** juga karena kita rumah sakit milik pemerintah. Pendapatan pasien yang **non BP JS** juga ada, tetapi tidak banyak. Sebagian besar pasien adalah peserta BPJS.

Q : Kalau yang non pasien pak. Seperti koperasi atau apa begitu hehe..

ER : Ada juga, tapi bukan koperasi. Misalnya **Kerjasama** dengan Yayasan, ada juga beberapa klinik yang tidak punya incinerator, kemudian bekerjasama, dan kerja sama lainnya.

(Diskusi lainnya yang tidak berhubungan...)

Q : Pak Andi, maaf ini pak. Tanggal 20 Desember kemarin, saya sempat diskusi dengan pak Agus via zoom pak. Untuk mendapatkan informasi seperti apa alur kegiatan terkait penagihan BPJS ini. Penjelasan dari beliau cukup detail sebenarnya, tetapi saya coba menangkap alur utamanya saja untuk tujuan pemahaman umum.

(menunjukkan diagram model swimlane..)

ER : Iya untuk alur proses, apalagi terkait koding ini memang pak Agus lebih paham dari saya. Iya ini sepertinya bisa menggambarkan proses yang terjadi. Saya juga memahaminya seperti ini.

Q : Saya bisa menggunakan diagram alur ini sebagai referensi awal ya pak.

ER : Iya ini tidak masalah.

Q : Baik pak Andi, terima kasih banyak waktunya. Mohon maaf untuk kurang kurangnya saya pak.


ER : Iya mas, kita tunggu saja progressnya. Syukur-syukur bisa kita implementasi di rumah sakit.

Q : insyaallah pak, terima kasih banyak.

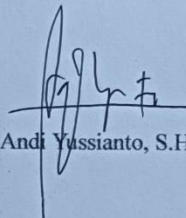
Yogyakarta, 5 Januari 2024

Peneliti

Narasumber



Aqge Priwibowo, ST



Dr Andi Yussianto, S.H., M.Epid