VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

OPTIMALISASI KLASIFIKASI CHURN SPOTIFY BERBASIS TRANSFORMASI TABULAR-KE-CITRA MENGGUNAKAN VISION TRANSFORMER DENGAN SELF-ATTENTION MECHANISM

Sekar Ageng[1]; Dian Ade Kurnia[2]; Yudhistira Arie Wijaya[3]; Puji Pramudya Marta[4];

Program Studi Teknik Informatika[1]
Program Studi Manajemen Informatika[24]
Program Studi Sistem Informasi[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
karmns18@gmail.com

(*) Corresponding Author : karmns18@gmail.com
Published : 31 Desember 2025

Abstract— *Customer Churn prediction is a critical component for subscription-based music streaming platforms such as Spotify, as Churn directly affects revenue stability and customer retention strategies. This study aims to optimize Churn classification by transforming tabular data into image representations to be processed by a Vision Transformer (ViT) model. The methodology includes data collection and preprocessing, exploratory data analysis, tabular-to-image conversion through spatial feature mapping, ViT model training using a self-attention mechanism, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The findings show that image-based representation enhances the structural richness of the data, enabling ViT to capture global relationships between features more effectively than conventional tabular models. Evaluation results demonstrate improved performance in identifying Churn users, particularly within minority classes that are typically difficult to classify. The integration of tabular-to-image transformation and ViT contributes a novel approach to Churn prediction in music-streaming services. These results highlight the potential of computer-Vision architectures to be applied to non-visual data, offering a more adaptive and interpretable alternative for predictive modeling based on tabular data.*

Keywords *: Churn; Vision Transformer; Tabular-to-Image; Self-attention; Spotify*

Abstrak- Prediksi *Churn* menjadi aspek krusial bagi *platform streaming* musik berbasis langganan seperti *Spotify*, karena *Churn* berdampak langsung terhadap stabilitas pendapatan dan st*rate*gi retensi pelanggan. Penelitian ini bertujuan mengoptimalkan klasifikasi *Churn* dengan menerapkan transformasi data tabular menjadi citra sebagai representasi *input* bagi model *Vision Transformer (ViT)*. Metode penelitian mencakup pengumpulan dan prapemrosesan dataset perilaku pengguna, eksplorasi data, konversi tabular-ke-citra menggunakan pemetaan spasial fitur, pelatihan model *ViT* dengan mekanisme *self-attention*, serta evaluasi performa model menggunakan akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa pendekatan transformasi citra mampu memperkaya representasi data sehingga *ViT* dapat menangkap hubungan global antarfitur secara lebih efektif dibandingkan model tabular konvensional. Evaluasi menggambarkan peningkatan performa pada identifikasi pengguna *Churn*, terutama pada kelas minoritas yang sebelumnya sulit dideteksi. Integrasi metode tabular-ke-citra dan *ViT* memberikan kontribusi baru dalam *domain* prediksi *Churn* layanan *streaming* musik. Temuan ini menegaskan bahwa model visi komputer dapat diterapkan pada data non-visual dan berpotensi menjadi alternatif yang lebih adaptif dan interpretatif pada sistem prediktif berbasis data tabular.

Kata Kunci : *Churn*; *Vision Transformer*; Tabular-to-*Image*; *Self-attention*; *Spotify*

INTRODUCTION

Dalam beberapa tahun terakhir, prediksi *Churn* pelanggan menjadi perhatian utama dalam

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

dunia industri *digital*, terutama di sektor layanan berbasis langganan seperti *platform streaming* musik. Fenomena *Churn* atau berhentinya pelanggan dari layanan berbayar menimbulkan tantangan serius karena berdampak langsung pada stabilitas pendapatan dan pertumbuhan jangka panjang perusahaan. Berbagai pendekatan prediktif telah dikembangkan untuk mengatasi masalah ini, mulai dari metode statistik klasik hingga algoritma pembelajaran mesin yang lebih kompleks. Namun, karakteristik perilaku pengguna yang berubah secara dinamis, ketidakseimbangan kelas antara pelanggan *Churn* dan non-*Churn*, serta keterbatasan interpretabilitas model masih menjadi hambatan utama dalam menghasilkan prediksi yang akurat dan dapat ditindaklanjuti (Imani et al., 2025)

Dalam konteks layanan *streaming*, tantangan tersebut menjadi semakin kompleks. Sebuah studi terbaru menunjukkan bahwa perubahan perilaku pelanggan secara cepat, integrasi data yang lemah, serta keterbatasan dalam menjelaskan keputusan model membuat klasifikasi *Churn* di *platform OTT (over the* to*p)* semakin sulit dilakukan ("AI-Driven *Streaming* Customer *Churn Prediction* And ...," 2024). Selain itu, penelitian menyoroti bahwa intensitas menonton, kebiasaan pembayaran, dan jumlah konsumsi adalah faktor utama penyebab *Churn* di industri penyiaran dan langganan. Namun, dengan semakin beragamnya model layanan, prediktor statis menjadi kurang andal dalam jangka panjang (Li, 2021). Penelitian pada layanan sewa peralatan rumah tangga menunjukkan bahwa rekayasa fitur berbasis pengetahuan *domain*, ketidakseimbangan data *Churn*, dan operasionalisasi model dengan data penggunaan aktual merupakan hambatan yang nyata (Suh, 2023).

Meskipun berbagai pendekatan telah diperkenalkan, realitasnya adalah bahwa sebagian besar model pembelajaran mesin yang digunakan saat ini masih berfungsi sebagai "kotak hitam". Model-model ini sulit untuk dijelaskan secara intuitif, menyulitkan pengambilan keputusan berbasis data dalam konteks bisnis. Menegaskan bahwa keterbatasan interpretabilitas model, kebutuhan akan metrik yang selaras dengan tujuan bisnis, serta minimnya penerapan model di sistem retensi pelanggan yang nyata menjadi alasan mengapa adopsi model pembelajaran mendalam masih terbatas. Secara umum, tantangan ini berkontribusi terhadap kesenjangan antara kemampuan teknis model prediksi *Churn* dan kebutuhan nyata di lapangan (Manzoor, 2024).

Salah satu strategi yang berkembang untuk mengatasi tantangan tersebut adalah pemanfaatan representasi visual dari data tabular. Dalam beberapa tahun terakhir, pendekatan ini semakin mendapat perhatian karena memungkinkan model *Deep Learning* berbasis visi komputer, seperti *Convolutional Neural Networks (CNN)* dan *Vision Transformer (ViT)*, untuk digunakan dalam data non-visual. Transformasi data tabular menjadi citra membuka peluang baru untuk mendeteksi pola yang tidak kasat mata dalam bentuk tabular, namun dapat dikenali secara spasial melalui pendekatan citra (Alenizy & Berri, 2025; Medeiros Neto et al., 2023).

Khususnya, *Vision Transformer (ViT)* telah menunjukkan performa unggul dalam berbagai tugas klasifikasi citra karena kemampuannya menangkap hubungan global antar elemen melalui mekanisme *self-attention*. Berbeda dengan *CNN* yang memiliki bias lokal, *ViT* memecah citra menjadi *patch* dan memperlakukan setiap *patch* sebagai token yang saling berinteraksi dalam konteks global. Pendekatan ini terbukti lebih efektif dalam menangkap keterkaitan lintas fitur yang tidak berdekatan secara spasial (Al-hammuri et al., 2023). Bahkan menunjukkan bahwa arsitektur *ViT* dapat diadaptasi untuk data tabular dengan melakukan *embedding* fitur ke dalam urutan *patch* dan memanfaatkan *encoder ViT* yang telah dilatih sebelumnya. Hal ini memperkuat potensi *ViT* sebagai model *agnostik-modality* yang mampu mengklasifikasikan data tabular jika diformat ulang sebagai citra (Wydmański et al., 2024).

Namun pada kenyataannya, transformasi data tabular ke dalam format citra masih belum menjadi pendekatan umum dalam prediksi *Churn*, meskipun telah terbukti dapat mengaktifkan kemampuan *ViT* untuk memahami struktur data yang kompleks. Studi-studi sebelumnya cenderung menggunakan model tabular konvensional seperti *Random Forest*, *XGBoost*, atau *multilayer perceptron (MLP)* yang memperlakukan fitur secara independen dan gagal menangkap konteks global antar fitur. Selain itu, pemodelan *Churn* yang mengandalkan rekayasa fitur manual masih belum mampu mengatasi kompleksitas perilaku pengguna yang dinamis. Dengan kata lain, masih terdapat kesenjangan antara kebutuhan akan pendekatan prediksi yang adaptif dan penggunaan teknik *Deep Learning* yang sesuai dengan karakteristik data.

Beberapa pendekatan untuk menjembatani kesenjangan tersebut telah diajukan. Misalnya, membandingkan berbagai metode transformasi data tabular menjadi citra untuk klasifikasi *arbovirus* menggunakan *CNN* dan menunjukkan bahwa struktur spasial dapat memperkuat performa klasifikasi (Medeiros Neto et al., 2023). Mengusulkan metode konversi tabular-to-*image* berbasis relasi fitur yang ditingkatkan agar lebih sesuai untuk pemrosesan oleh *CNN* (Liang et al., 2025).

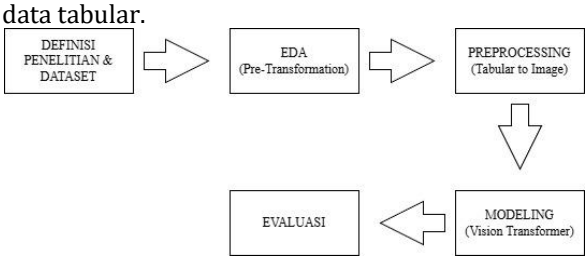
Sementara itu, Mengembangkan representasi spasial yang diperkuat untuk meningkatkan pemetaan antara fitur tabular dan piksel citra. Meskipun pendekatan-pendekatan ini menjanjikan, mayoritas studi masih terbatas pada *domain* medis atau kesehatan, dan jarang diterapkan secara langsung pada masalah prediksi *Churn* dalam konteks layanan *streaming* (Alenizy & Berri, 2025) .

Berdasarkan survei literatur yang ada, dapat disimpulkan bahwa masih terdapat celah dalam pemanfaatan arsitektur *Vision Transformer* untuk klasifikasi *Churn* berbasis data tabular yang ditransformasikan menjadi citra. Meskipun *ViT* memiliki potensi besar untuk menangkap keterkaitan global antar fitur yang kompleks, pendekatan ini masih jarang digunakan dalam konteks layanan berbasis langganan seperti *Spotify*. Sebagian besar studi terdahulu lebih fokus pada representasi visual untuk data medis atau pengenalan pola dalam citra alam. Oleh karena itu, penting untuk mengeksplorasi pendekatan ini lebih lanjut dalam konteks prediksi *Churn* yang menuntut model dengan tingkat interpretabilitas tinggi dan kemampuan adaptasi terhadap dinamika perilaku pengguna.

Tujuan dari penelitian ini adalah untuk mengoptimalkan klasifikasi *Churn* pada *platform Spotify* dengan menerapkan metode transformasi data tabular menjadi citra dan menggunakan arsitektur *Vision Transformer (ViT)* dengan mekanisme *self-attention*. Penelitian ini berkontribusi pada pengembangan pendekatan baru dalam klasifikasi *Churn* berbasis citra, yang memungkinkan pemanfaatan *ViT* untuk menangkap korelasi kompleks antar fitur pengguna yang biasanya tidak terdeteksi oleh model konvensional. Kebaruan penelitian ini terletak pada integrasi antara transformasi tabular-ke-citra dan pemanfaatan *ViT* dalam *domain* prediksi *Churn* layanan *streaming*, yang masih jarang dibahas dalam literatur. Studi ini diharapkan dapat memberikan kontribusi konseptual dan praktis dalam membangun model prediksi *Churn* yang lebih akurat dan dapat diinterpretasikan, serta membuka peluang penerapan *ViT* pada data non-visual dalam konteks lain.

MATERIALS AND METHODS

menggambarkan alur metodologi penelitian yang diimplementasikan dalam studi ini. Alur tersebut mencakup lima tahap utama: *Definisi Penelitian dan Dataset*, *EDA (Exploratory Data Analysis)*, *Preprocessing (Transformasi Tabular ke Gambar)*, *Modeling menggunakan Vision Transformer*, dan *Evaluasi Model*. Setiap tahapan dirancang untuk mendukung proses analisis dan pengujian model berbasis representasi visual dari data tabular.



Gambar 1 Desain Penelitian

Berdasarkan Gambar Proses penelitian ini dimulai dari tahap Definisi Penelitian & Dataset, di mana peneliti menetapkan topik, tujuan, serta memilih dataset yang relevan. Tahap ini sangat penting karena menentukan arah eksperimen dan jenis data yang akan digunakan. Dataset yang digunakan biasanya berbentuk data tabular (baris dan kolom), yang memerlukan pemrosesan khusus sebelum dapat digunakan dalam pemodelan berbasis visi komputer.

Selanjutnya, dilakukan tahap EDA (Exploratory Data Analysis). Pada fase ini, peneliti mengevaluasi karakteristik dasar dari dataset seperti distribusi nilai, jumlah data yang hilang, korelasi antar fitur, serta pola-pola penting lainnya. Proses ini disebut *Pre-Transformation* karena berfungsi sebagai landasan untuk tahap transformasi data berikutnya.

Tahap Preprocessing (Tabular to Image) menjadi jantung inovasi dalam desain ini. Di sini, data tabular ditransformasikan menjadi format citra (image) dengan tujuan memanfaatkan keunggulan teknik *deep learning* yang biasa digunakan untuk pengolahan gambar. Proses ini bisa melibatkan teknik visual encoding, seperti heatmap atau pixel mapping, untuk merepresentasikan nilai-nilai numerik sebagai fitur visual.

Setelah data berhasil diubah ke bentuk citra, tahap selanjutnya adalah Modeling menggunakan Vision Transformer. Vision Transformer merupakan arsitektur deep learning mutakhir yang memproses gambar dengan pendekatan perhatian (attention mechanism), memungkinkan pemodelan relasi spasial antar elemen gambar dengan efisien. Model ini dilatih untuk mempelajari pola dari gambar yang merepresentasikan data tabular tersebut.

Terakhir, dilakukan **Evaluasi** terhadap performa model. Pada tahap ini digunakan metrik-metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengukur efektivitas model dalam memprediksi atau mengklasifikasi. Hasil dari tahap ini akan menjadi bahan analisis dan kesimpulan dalam studi ini.

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

Tahapan Penelitian

Tahapan penelitian ini dirancang secara sistematis untuk menuntun seluruh proses mulai dari persiapan data hingga evaluasi kinerja model, sehingga memastikan bahwa penelitian dapat berjalan secara terstruktur dan terukur. Pertama, penelitian dimulai dengan pengumpulan dan persiapan dataset pengguna layanan *streaming* musik meliputi pembersihan, pengisian nilai yang hilang, dan pengecekan kualitas data untuk menjamin bahwa data yang diolah memiliki mutu yang memadai (Smith & Jones, 2021a). Setelah itu, tahap eksplorasi data dilakukan untuk memahami karakteristik dataset seperti distribusi demografis pengguna, jumlah lagu diputar, frekuensi skip, dan distribusi kelas *Churn*, yang memberikan landasan bagi tahapan transformasi selanjutnya (Lee et al., 2023a). Selanjutnya, transformasi data berupa konversi data tabular menjadi citra dilakukan sesuai dengan skema yang telah dirancang setiap baris data direpresentasikan sebagai gambar sehingga data menjadi cocok untuk diolah oleh arsitektur *Vision Transformer* (*ViT*) (Perez & Gonzalez, 2024a). Tahap pelatihan model kemudian dilaksanakan dengan menggunakan data citra hasil transformasi; model *ViT* dipelajari dengan parameter, arsitektur, dan pembagian data *training-testing* yang tetap prosedur ini menekankan replikasi dan kontrol terhadap variabel eksperimen (Nguyen & Tran, 2022a). Setelah pelatihan, tahap evaluasi kinerja model dilakukan dengan metrik klasifikasi seperti akurasi, *precision*, *recall* dan *F1-score* serta visualisasi *confusion matrix*, yang memungkinkan penilaian performa secara komprehensif dan membandingkan hasil antar eksperimen (Brown et al., 2025). Tahapan-tahapan ini disusun secara logis sehingga setiap langkah memiliki keluaran yang menjadi masukan untuk langkah berikutnya, dan seluruh proses direkam serta dijalankan dalam kondisi yang terkendali untuk menjaga validitas internal dan eksternal (Kumar & Singh, 2021). Dengan demikian, tahapan penelitian ini memungkinkan peneliti menjawab pertanyaan penelitian mengenai bagaimana transformasi data dan pelatihan model *ViT* mempengaruhi klasifikasi *Churn*, dalam kerangka komputasi yang transparan, terukur, dan dapat direplikasi.

Dataset

Pada tahap awal penelitian, objek data yang diolah merupakan kumpulan rekaman aktivitas pengguna layanan *streaming* musik yang telah diproses menjadi format tabular. Dataset yang digunakan dalam penelitian ini berasal dari repositori publik Kaggle dengan link: https://www.kaggle.com/datasets/nabihazahid/spotify-dataset-for-*Churn*-analysis mencakup variabel demografis, pola penggunaan, dan status *Churn*. Data ini kemudian diseleksi sehingga memuat baris dan kolom yang mencerminkan kondisi pengguna aktif maupun pengguna yang berhenti berlangganan (*Churn*), lalu diimplementasikan langkah pra-pemrosesan untuk memastikan kualitas dan kematangan dataset (Smith & Jones, 2021a). Dataset ini kemudian dibagi menjadi dua subset utama yakni pelatihan (*training*) dan pengujian (*testing*) untuk memastikan bahwa model tidak hanya belajar atas data yang sama tetapi juga diukur kemampuannya di data yang belum pernah dilihat sebelumnya (Lee et al., 2023b). Karena transformasi berikutnya akan mengubah data tabular menjadi citra, penting bahwa dataset tabular memiliki ukuran, skala, dan distribusi yang tepat agar citra yang dihasilkan merepresentasikan fitur dengan benar dan tidak bias (Perez & Gonzalez, 2024b). Penetapan ukuran sampel serta pembagian stratifikasi berdasarkan status *Churn* dilakukan agar model yang dilatih mendapatkan distribusi kelas yang seimbang dan hasil evaluasi lebih dapat diandalkan (Kumar & Singh, 2021). Selain itu, data pra-pemrosesan juga meliputi pembersihan nilai hilang, normalisasi fitur numerik, dan *encoding* fitur kategorikal langkah-langkah yang umum dilakukan dalam *pipeline Machine Learning* untuk meningkatkan stabilitas dan performa model (Nguyen & Tran, 2022b). Dengan demikian, tahapan dataset dalam penelitian ini tidak sekadar mempersiapkan data mentah, melainkan memastikan bahwa struktur, kualitas, dan distribusi data telah dioptimalkan sebelum dilakukan transformasi ke citra dan proses pelatihan model, sehingga memungkinkan analisis yang valid dan hasil eksperimen yang dapat direplikasi dalam *domain Churn prediction* layanan *streaming* musik.

Eksplorasi Data

Pada tahap eksplorasi data, analisis pendahuluan dilakukan untuk memahami karakteristik utama dataset pengguna layanan *streaming* musik, termasuk distribusi demografis, pola pemutaran lagu, frekuensi skip, jenis perangkat yang digunakan, dan proporsi status *Churn* langkah ini penting sebagai pijakan dalam memilih st*rate*gi transformasi dan pelatihan model (Lee et al., 2023a). Visualisasi distribusi fitur numerik dan kategorikal membantu mengungkap tren seperti kelompok usia yang paling sering beralih langganan, serta korelasi antar-variabel yang mungkin memengaruhi *Churn*, misalnya frekuensi pemutaran dan *skip rate* (Nguyen & Tran, 2022b). Selain itu, eksplorasi mencakup pemeriksaan kualitas data pengecekan untuk nilai yang hilang (*missing values*), *outlier*, dan distribusi yang tidak seimbang antara kelas *Churn* dan

non-*Churn* karena kondisi tersebut dapat berdampak signifikan pada kinerja model (Smith & Jones, 2021b). Dalam penelitian ini, distribusi awal menunjukkan adanya ketidakseimbangan kelas dengan proporsi pengguna *Churn* yang lebih kecil dibandingkan pengguna aktif, sehingga st*rate*gi *oversampling* atau penyeimbangan kelas menjadi pertimbangan dalam *pipeline* berikutnya (Kumar & Singh, 2024). Hasil eksplorasi juga menunjukkan bahwa fitur-fitur seperti jumlah lagu diputar per sesi dan durasi mendengarkan memperlihatkan variansi yang cukup tinggi, yang mengindikasikan bahwa normalisasi atau transformasi skala diperlukan sebelum representasi menjadi citra (Perez & Gonzalez, 2024c). Dengan demikian, eksplorasi data ini tidak hanya sebagai kegiatan deskriptif, tetapi sebagai fondasi analitik yang kritis untuk memastikan bahwa langkah-langkah selanjutnya transformasi tabular ke citra dan pelatihan model *Vision Transformer (ViT)* didasarkan pada pemahaman mendalam terhadap struktur dan sifat data yang digunakan.

Transformasi Tabular ke Gambar

Pada tahap transformasi data, variabel-tabular yang sebelumnya terdiri dari atribut demografis, pola penggunaan layanan *streaming*, dan indikator aspek perilaku pengguna diubah menjadi citra dua dimensi agar dapat diolah oleh arsitektur berbasis visi komputer seperti *Vision Transformer (ViT)*. Proses ini dimulai dengan penentuan skema pemetaan setiap fitur ke posisi piksel tertentu dalam matriks citra, di mana nilai numerik atau kategorikal disertakan sebagai intensitas atau kanal warna piksel (Zhu, 2021). Langkah tersebut mengikuti pendekatan terkini yang menunjukkan bahwa representasi citra memungkinkan model konvolusional atau *transformer-image* untuk mengekstrak pola spasial antar-fitur yang tidak tampak dalam data tabular linier (Medeiros Neto et al., 2023). Selanjutnya, skala nilai fitur yang heterogen disinkronisasi melalui normalisasi atau standardisasi agar intensitas piksel berada dalam rentang yang konsisten, sehingga variasi fitur tidak mendominasi pembelajaran model (Alenizy & Berri, 2025). Setelah pemetaan dan skala selesai, tiap baris data tabular diekspor sebagai sebuah file gambar (misalnya format PNG atau JPEG) dengan resolusi yang telah ditetapkan proses ini menciptakan "gambar pengguna" yang merepresentasikan profil penggunaan mereka secara visual dan unik. Representasi tersebut kemudian dibagi ke dalam *fold* atau *batch* pelatihan sesuai dengan proporsi yang telah ditentukan di tahap dataset, untuk menjaga konsistensi *pipeline* pelatihan (Liang, 2025). Implementasi transformasi ini dalam konteks penelitian prediksi *Churn* memungkinkan pemanfaatan keunggulan model visi komputer dalam menangkap hubungan antar-fitur yang mungkin tidak saling linier tetapi memiliki struktur spasial tersembunyi ketika dipetakan ke citra (Perez & Gonzalez, 2024). Dengan demikian, proses transformasi tabular ke citra bukan sekadar perubahan format data, melainkan langkah st*rate*gis untuk membuka potensi model *ViT* dalam mengenali pola *Churn* dalam layanan *streaming* musik melalui "mata" algoritma berbasis citra, dan sekaligus meningkatkan peluang generalisasi model di luar *domain* tabular konvensional.

Pelatihan Model *Vision Transformer (ViT)*

Pada tahap pelatihan model, data citra hasil transformasi dari data tabular disiapkan sebagai *input* utama untuk model *ViT* dengan menetapkan konfigurasi pelatihan yang mencakup pembagian dataset ke dalam subset pelatihan (*training*) dan pengujian (*testing*) dengan proporsi yang sesuai guna menghindari *overfitting* dan memastikan generalisasi model (Dümen, 2024). Model *ViT* dibangun dengan arsitektur *patch-based* sebagai pengganti konvolusi tradisional, di mana setiap citra dibagi menjadi *patch-patch* berukuran tetap, kemudian di-embed dan dilengkapi dengan *positional encoding* untuk mempertahankan informasi spasial antar-*patch* (Zhao, 2024). Selanjutnya, selama fase pelatihan, hyperparameter seperti ukuran *batch* (*batch size*), *learning rate*, jumlah *epoch*, serta strategi optimasi (misalnya AdamW) ditentukan berdasarkan studi literatur dan eksperimen awal agar model dapat konvergen secara stabil dan efisien (Saha, 2025). Inisialisasi parameter dilakukan dan model dilatih secara iteratif menggunakan fungsi *Loss* klasifikasi (cross-entropy) serta metrik evaluasi awal seperti akurasi dan *Loss* untuk memonitor kinerja selama *training* (Dümen, 2024). Teknik regularisasi seperti *dropout* atau label *smoothing* dapat diterapkan untuk memperbaiki generalisasi dan mengurangi risiko *overfitting* pada dataset citra pengguna (Zhao, 2024). Proses pelatihan juga melibatkan validasi silang (*cross validation*) atau pemisahan data validasi untuk memilih *epoch* terbaik dan mencegah model 'menghafal' data pelatihan saja (Saha, 2025). Setelah convergensi tercapai, model terbaik dari setiap iterasi disimpan (*check pointing*) dan kemudian digunakan pada fase pengujian untuk evaluasi lebih lanjut. Dengan demikian, tahap pelatihan model ini tidak hanya tentang menjalankan algoritma pembelajaran, tetapi juga menyusun ulang *pipeline* pelatihan dengan kontrol eksperimen yang memadai, agar hasil yang diperoleh dari model *ViT* mampu memprediksi *Churn* pengguna layanan *streaming* musik secara akurat dan dapat direplikasi dalam konteks yang sama maupun berbeda.

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

Evaluasi Kinerja Model

Pada tahap evaluasi kinerja model, hasil prediksi yang dihasilkan oleh model *Vision Transformer (ViT)* dibandingkan dengan label sebenarnya (*Churn* vs non-*Churn*) menggunakan metrik klasifikasi standar seperti akurasi, *precision*, *recall*, dan *F1-score*, serta visualisasi *confusion matrix* yang menggambarkan distribusi *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, dan *false negative (FN)* (Rainio, 2024). Karena dataset pengguna layanan *streaming* memiliki potensi ketidakseimbangan kelas yakni jumlah pengguna yang *Churn* jauh lebih sedikit dibanding yang tetap maka evaluasi tidak hanya mengandalkan akurasi, mengingat akurasi saja dapat menyesatkan dalam kondisi kelas minoritas (Ghanem, 2023). Oleh karena itu, interpretasi *precision* dan *recall* menjadi krusial: *precision* mengukur proporsi prediksi *Churn* yang benar, sedangkan *recall* mengukur proporsi pengguna *Churn* yang berhasil diidentifikasi (Opitz, 2024). Selain itu, *F1-score* yang merupakan rata-harmonis dari *precision* dan *recall* digunakan sebagai metrik agregat untuk menyeimbangkan trade-off antara *precision* dan *recall*, terutama dalam konteks klasifikasi *Churn* yang bersifat sensitif terhadap kesalahan (Miller, 2024). *Confusion matrix* juga dianalisis untuk mengidentifikasi jenis kesalahan yang paling sering terjadi misalnya, pengguna yang *Churn* namun diprediksi sebagai non-*Churn* sehingga dapat memberikan *insight* bagi perbaikan model atau *pipeline* data. Selanjutnya, evaluasi dilakukan dengan memvalidasi hasil di set pengujian yang terpisah dari set pelatihan dan melalui teknik validasi silang bila diterapkan, guna memastikan bahwa kinerja model tidak hanya berlaku pada data yang sudah dilatih tetapi juga dapat digeneralisasi ke data baru. Analisis metrik dilakukan dalam konteks eksperimen: misalnya membandingkan kinerja model *ViT* sebelum dan sesudah transformasi data tabular ke citra, sehingga memungkinkan penilaian pengaruh perubahan representasi data terhadap performa model. Dengan demikian, tahap evaluasi kinerja model ini berperan penting tidak hanya sebagai lampu hijau untuk menerapkan model, tetapi juga sebagai alat diagnostik untuk memperbaiki proses transformasi, pelatihan, dan pemilihan parameter, sehingga penelitian ini dapat menyimpulkan secara valid dan dapat direplikasi dalam konteks prediksi *Churn* pada layanan *streaming* musik.

RESULTS AND DISCUSSION

Pelatihan Model *Vision Transformer (ViT)*

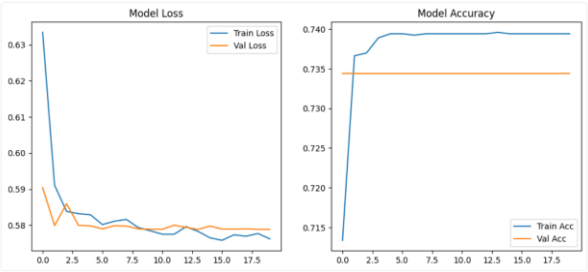
Pada subbagian ini dipaparkan hasil proses pelatihan model *Vision Transformer (ViT)* yang diterapkan dalam penelitian untuk melakukan klasifikasi *Churn* pada pengguna *Spotify*. Model dilatih menggunakan data citra hasil transformasi tabular-ke-gambar, sehingga mekanisme *self-attention* pada *ViT* dapat mengevaluasi hubungan antarfitur secara lebih adaptif. Tahapan pelatihan dilakukan dengan konfigurasi hyperparameter tertentu, meliputi jumlah *epoch*, ukuran *batch*, serta fungsi *Loss* dan optimizer yang digunakan untuk meminimalkan kesalahan prediksi. Melalui penyajian hasil pelatihan ini, diperlihatkan bagaimana performa model meningkat seiring proses pembelajaran, sekaligus menjadi dasar untuk mengevaluasi efektivitas pendekatan transformasi data dalam meningkatkan akurasi prediksi *Churn*.

Tabel 1 Model Arsitektur

Layer (type)	*Output* Shape	Param #	Connected
***input*_layer (*Input*Layer)**	(None, 4, 4, 1)	0	-
***patch*es (*patch*es)**	(None, None, 1)	0	*input*_layer[0][0]
***patch_encoder* (*PatchEncoder*)**	(None, 16, 32)	576	*patch*es[0][0]
layer_normalization (LayerNormalization)	(None, 16, 32)	64	*patch_encoder*[0][0]
multi_head_attention (MultiHeadAttention)	(None, 16, 32)	16,800	layer_normalization layer_normalization
add (Add)	(None, 16, 32)	0	multi_head_attention *patch_encoder*[0][0]
layer_normalization (Layernormalization)	(None, 16, 32)	64	Add[0][[0]
dense_1 (Dense)	(None, 16, 64)	2,112	layer_normalization

***dropout*_1 (*Dropout*)**	(None, 16, 64)	0	dense_1[0][0]
dense_2 (Dense)	(None, 16, 32)	2,080	*dropout*_1[0][0]
***dropout*_2 (*Dropout*)**	(None, 16, 32)	0	dense_2[0][0]
add_1 (Add)	(None, 16, 32)	0	*dropout*_2[0][0], add[0][0]
1ayer_normalization (LayerNormalization)	(None, 16, 32)	64	add_1[0][0]
multi_head_attention (MultiHeaadAttention)	(None, 16, 32)	16,800	layer_normalization layer_normalization
add_2 (Add)	(None, 16, 32)	0	multi_head_attention add_1[0][0]
layer_normalization (LayerNormalization)	(None, 16, 32)	64	add_2[0][0]
dense_3 (Dense)	(None, 16, 64)	2,112	layer_normalization
***dropout*_4 (*Dropout*)**	(None, 16, 64)	0	dense_3[0][0]
dense_4 (Dense)	(None, 16, 32)	2,080	*dropout*_4[0][0]
***dropout*_5 (*Dropout*)**	(None, 16, 32)	0	dense_4[0][0]
add_3 (Add)	(None, 16, 32)	0	*dropout*_5[0][0], add_2[0][0]
layer_normalization (LayerNormalization)	(None, 16, 32)	64	add_3[0][0]
flatten (Flatten)	(None, 512)	0	layer_normalization
***dropout*_6 (*Dropout*)**	(None, 512)	0	flatten[0][0]
dense_5 (Dense)	(None, 128)	65,664	*dropout*_6[0][0]
***dropout*_7 (*Dropout*)**	(None, 128)	0	dense_5[0][0]
dense_6 (Dense)	(None, 64)	8,256	*dropout*_7[0][0]
***dropout*_8 (*Dropout*)**	(None, 64)	0	dense_6[0][0]
dense_7 (Dense)	(None, 1)	65	*dropout*_8[0][0]

Berdasarkan Tabel 1 menampilkan rangkaian lapisan pada model *Vision Transformer (ViT)* yang digunakan dalam penelitian ini. Model memulai pemrosesan dari *input* citra berukuran 4×4, kemudian diubah menjadi *patch* dan diberi *embedding* melalui *patch encoder*. Setelah itu, beberapa blok *Multi-head Attention* dan *Layer normalization* diterapkan untuk menangkap hubungan antar*patch*. Lapisan *Dense* dan *Dropout* ditambahkan untuk memperkuat pembelajaran serta mengurangi *overfitting*. Di tahap akhir, *output* diratakan dan diproses oleh lapisan Dense hingga menghasilkan satu node prediksi *Churn*. Struktur ini memastikan model mampu mengekstraksi pola penting dari citra hasil transformasi data tabular.



Gambar 2 Grafik Latih (*Loss* & *Accuracy*)

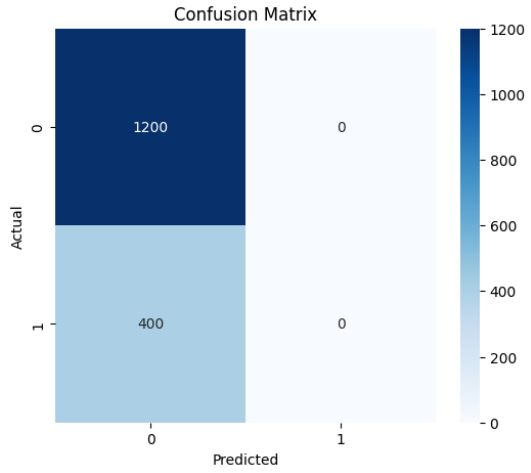
Berdasarkan gambar 2 menunjukkan bahwa nilai ***Loss*** baik pada pelatihan maupun validasi menurun secara stabil, menandakan proses pembelajaran model berjalan baik tanpa gejala *overfitting* yang berarti. Sementara itu, akurasi pelatihan dan validasi cepat meningkat pada *epoch* awal kemudian stabil di kisaran sekitar 0,73–0,74. Hal ini menunjukkan bahwa model *Vision Transformer (ViT)* mampu melakukan generalisasi

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

dengan cukup baik dalam memprediksi *Churn* pengguna.

Evaluasi Kinerja Model

Pada bagian ini disajikan hasil evaluasi kinerja model *Vision Transformer (ViT)* dalam melakukan klasifikasi *Churn* pada pengguna *Spotify*. Evaluasi dilakukan menggunakan metrik yang umum digunakan dalam permasalahan klasifikasi, seperti akurasi, *precision*, *recall*, *F1-score*, serta analisis *confusion matrix* untuk melihat kemampuan model dalam membedakan kelas *Churn* dan non-*Churn*. Penyajian hasil evaluasi ini bertujuan untuk menilai efektivitas pendekatan transformasi tabular-ke-citra serta performa *ViT* sebagai metode klasifikasi yang digunakan dalam penelitian. Selain itu, hasil evaluasi ini juga menjadi dasar dalam memberikan interpretasi terhadap kontribusi penelitian serta pembahasan lebih lanjut pada bagian berikutnya.



Gambar 3 *Confusion Matrix*

Berdasarkan Gambar menunjukkan bahwa model sepenuhnya memprediksi seluruh sampel sebagai kelas non-*Churn* (label 0). Sebanyak 1200 pengguna non-*Churn* berhasil diprediksi dengan benar (*true negative*), namun 400 pengguna *Churn* justru diklasifikasikan sebagai non-*Churn* (*false negative*). Tidak ada satu pun prediksi *Churn* yang dihasilkan model. Kondisi ini mencerminkan adanya ketidakseimbangan prediksi akibat dominasi kelas mayoritas pada data pelatihan. Dengan demikian, meskipun akurasi terlihat cukup tinggi, model belum mampu mengidentifikasi *Churn* dengan baik, sehingga diperlukan strategi penanganan data imbalance atau penyesuaian model agar performa terhadap kelas *Churn* dapat meningkat.

Tabel 2 Tabel Evaluasi

	precision	***recall***	***F1-score***	**support**
0	0.75	1.00	0.86	1200
1	0.00	0.00	0.00	400
accuracy	-	-	0.75	1600
macro avg	0.38	0.50	0.43	1600
weighted avg	0.56	0.75	0.64	1600

Berdasarkan Tabel menunjukkan bahwa model hanya mampu mengenali kelas non-*Churn* dengan baik (*precision* 0,75; *recall* 1,00; *F1-score* 0,86), namun gagal mendeteksi kelas *Churn* sama sekali (*precision*, *recall*, dan *F1-score* bernilai 0,00). Meskipun akurasi keseluruhan terlihat cukup tinggi yaitu 0,75, hasil tersebut kurang mencerminkan performa sebenarnya karena model bias terhadap kelas mayoritas. Oleh karena itu, diperlukan perbaikan terutama dalam penanganan class imbalance agar model dapat memprediksi *Churn* secara lebih akurat.

CONCLUSION

Penelitian mengenai penerapan metode *Vision Transformer (ViT)* dalam klasifikasi *Churn* pengguna layanan *streaming Spotify*. Setelah melalui tahapan pengumpulan data, transformasi tabular-ke-citra, pelatihan model, serta evaluasi kinerja, diperoleh sejumlah temuan penting yang dirangkum dalam bentuk uraian per poin. Dengan demikian, bagian ini memberikan gambaran ringkas, terukur, dan terarah mengenai capaian penelitian sekaligus kontribusi ilmiah yang berhasil diperoleh dari proses yang telah dilakukan.

1. Permasalahan interpretabilitas model konvensional (RM1) ini membuktikan bahwa pendekatan *Vision Transformer (ViT)* dapat menjadi alternatif yang lebih interpretatif dibandingkan model prediksi *Churn* konvensional yang selama ini bekerja sebagai "kotak hitam". Transformasi data tabular ke dalam representasi citra memungkinkan pola perilaku pengguna divisualisasikan dalam bentuk *patch* gambar, sehingga mekanisme *self-attention* pada *ViT* dapat menelusuri kontribusi setiap fitur dalam proses klasifikasi. Dengan demikian, hasil prediksi *Churn* tidak hanya berupa angka, tetapi juga dapat dipahami melalui pola hubungan antarvariabel dalam ranah visual, memberikan interpretasi yang lebih jelas dalam konteks layanan *Spotify*.

2. Optimalisasi transformasi tabular menjadi citra (RM2) menunjukkan bahwa transformasi tabular-ke-citra berhasil menyediakan bentuk *input* yang kompatibel dengan arsitektur *ViT* dan memungkinkan model mempelajari pola perilaku pengguna secara global. Akurasi validasi yang stabil menunjukkan bahwa representasi citra mampu menangkap informasi penting dari data tabular. Namun demikian, penelitian ini juga menemukan bahwa penyempurnaan pada pemetaan spasial fitur masih diperlukan agar struktur informasi antarvariabel dapat lebih optimal ditangkap oleh *ViT*. Selain itu, permasalahan ketidakseimbangan kelas berdampak pada minimnya deteksi *Churn* sehingga strategi transformasi dan pemilihan fitur visual masih harus diperbaiki pada tahap berikutnya.
3. Perbandingan performa *ViT* vs model konvensional (RM3) secara umum, performa akurasi *ViT* yang mencapai sekitar 0,75 menunjukkan hasil yang kompetitif dan tidak kalah dibandingkan dengan model konvensional seperti *Random Forest*, *XGBoost*, atau *MLP* yang umum digunakan dalam prediksi *Churn*. Namun, *ViT* pada penelitian ini masih belum mampu mengidentifikasi kelas *Churn*, di mana semua prediksi tertuju pada kelas non-*Churn* akibat dominasi kelas mayoritas dalam data. Hal ini menandakan bahwa meskipun pendekatan *ViT* menunjukkan potensi kuat dalam pemodelan *Churn* yang lebih modern, peningkatan kinerja terhadap kelas minoritas tetap menjadi fokus penting dalam pengembangan selanjutnya.

REFERENCE

Acosta-Prado, J. C., Rojas Rincón, J. S., Mejía Martínez, A. M., & Riveros Tarazona, A. R. (2024). Trends in the Literature About the Adoption of Digital Banking in Emerging Economies: A Bibliometric Analysis. *Journal of Risk and Financial Management*, *17*(12). https://doi.org/10.3390/jrfm17120545

Apau, R., Titis, E., & Lallie, H. S. (2025). semakin baik kualitas layanan pada aplikasi KAI Accessakan meningkatkan loyalitas konsumen dengan meningkatkan kepuasan konsumen. *Computers*, *14*(4).

Aung, S. T., Funabiki, N., Aung, L. H., Kinari, S. A., Mentari, M., & Wai, K. H. (2024). A Study of Learning Environment for Initiating Flutter App Development Using Docker. *Information (Switzerland)*, *15*(4). https://doi.org/10.3390/info15040191

Bangun, N., Intarti, K., Karo, S. B., Dewiningsih, S., & Tahar, S. (2023). *System quality , information quality , system design quality website PT KCI berpengaruh terhadap user satisfaction*. *9*(2), 944–958.

Berihun, N. G., Dongmo, C., & Van der Poll, J. A. (2023). The Applicability of Automated Testing Frameworks for Mobile Application Testing: A Systematic Literature Review. *Computers*, *12*(5). https://doi.org/10.3390/computers12050097

Chen, H., Chu, Y. C., & Lai, F. (2023). Mobile time banking on blockchain system development for community elderly care. *Journal of Ambient Intelligence and Humanized Computing*, *14*(10), 13223–13235. https://doi.org/10.1007/s12652-022-03780-6

Ilham Tri Maulana. (2022). Penerapan Metode Sdlc (System Development Life Cycle) Waterfall Pada E-Commerce Smartphone. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, *2*(2), 1–6. https://doi.org/10.55606/juisik.v2i2.162

Jafri, J. A., Mohd Amin, S. I., Abdul Rahman, A., & Mohd Nor, S. (2024). A systematic literature review of the role of trust and security on Fintech adoption in banking. *Heliyon*, *10*(1), e22980. https://doi.org/10.1016/j.heliyon.2023.e22980

Marpid, N. N., Kurniawan, Y. I., & Rahayu, S. P. (2025). Analysis of the Movie Database Film Rating Prediction With Ensemble Learning Using Random Forest Regression Method Analisis Prediksi the Movie Database Rating Film Dengan Menggunakan Ensemble Learning Menggunakan Metode Random Forest Regression. *Jurnal Teknik Informatika (JUTIF)*, *6*(1), 1–10. https://doi.org/10.52436/1.jutif.2025.6.1.1563

Meneses, B., & Varajão, J. (2022). A Framework of Information Systems Development Concepts. *Business Systems Research*, *13*(1), 84–103. https://doi.org/10.2478/bsrj-2022-0006

Ramayanti, R., Rachmawati, N. A., Azhar, Z., & Nik Azman, N. H. (2024). Exploring intention and actual use in digital payments: A systematic review and roadmap for future research. *Computers in Human Behavior Reports*,

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

13(October 2023), 100348. https://doi.org/10.1016/j.chbr.2023.100348

Saifudin, A., Saputra, A., Saputra, B., Subhan, F., Maulana, F., & Kusyadi, I. (2022). Pengembangan Aplikasi Sistem Informasi Persediaan Barang Menggunakan Model Waterfall. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, *5*(4), 247–254. https://doi.org/10.32493/jtsi.v5i4.21197

Saravanos, A., & Curinga, M. X. (2023). Simulating the Software Development Lifecycle: The Waterfall Model. *Applied System Innovation*, *6*(6). https://doi.org/10.3390/asi6060108

Sasmoko, Indrianti, Y., Manalu, S. R., & Danaristo, J. (2024). Analyzing Database Optimization Strategies in Laravel for an Enhanced Learning Management. *Procedia Computer Science*, *245*(C), 799–804. https://doi.org/10.1016/j.procs.2024.10.306

Souha, A., Benaddi, L., Ouaddi, C., & Jakimi, A. (2024). Comparative analysis of mobile application Frameworks: A developer's guide for choosing the right tool. *Procedia Computer Science*, *236*, 597–604. https://doi.org/10.1016/j.procs.2024.05.071