BAB 2
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Konsep Dasar Kualitas Pelayanan

Dalam teori, kualitas layanan merupakan aspek fundamental dalam manajemen jasa karena menentukan tingkat kepuasan dan loyalitas pelanggan. Model awal kualitas layanan diperkenalkan oleh Grönroos (1984) yang membedakan antara *technical quality* (hasil layanan yang diterima pelanggan) dan *functional quality* (cara layanan diberikan). Model ini menekankan bahwa layanan tidak hanya dinilai dari hasil akhir, melainkan juga dari proses interaksi yang terjadi selama penyampaian layanan.

2.1.2 Model SERVQUAL

Perkembangan penting selanjutnya adalah hadirnya model *SERVQUAL* oleh Parasuraman dkk. (1988) yang menjadi kerangka teoritis paling banyak digunakan dalam mengukur kualitas layanan. *SERVQUAL* memperkenalkan lima dimensi utama, yaitu *Tangible, reliability, responsiveness, assurance, dan empathy.* Model ini berlandaskan pada konsep kesenjangan (*gap model*) antara harapan pelanggan dengan persepsi mereka terhadap kinerja layanan aktual. Pendekatan ini menjadikan *SERVQUAL* sebagai instrumen yang komprehensif karena mampu menggambarkan bagaimana pelanggan menilai kualitas secara menyeluruh. Meskipun dikembangkan sejak akhir 1980-an, SERVQUAL terus digunakan secara luas di berbagai industri dan negara, termasuk perbankan di berbagai belahan dunia

Adapun Lima Dimensi SERVQUAL sebagai berikut :

Tabel 2.1 Lima Dimensi SERVQUAL
Sumber: Diadaptasi dari Parasuraman, Zeithaml, & Berry (1988)

Dimensi	Definisi	Indikator Perbankan
Tangible (Bukti Fisik)	Bukti fisik fasilitas, peralatan, dan penampilan personel	Kondisi gedung, ruang tunggu, kebersihan, kelengkapan ATM
Reliability (Keandalan)	Kemampuan memberikan layanan yang dijanjikan secara andal dan akurat	Kecepatan transaksi, akurasi data nasabah
Responsiveness (Daya Tanggap)	Kemauan membantu nasabah dan memberikan layanan cepat dan tepat	Waktu respons teller, kesigapan satpam
Assurance (Jaminan)	Pengetahuan, kesopanan, dan kemampuan untuk menumbuhkan kepercayaan dan keyakinan nasabah	Kompetensi CS, keramahan pegawai *frontliner*
Empathy (Empati)	Kepedulian dan perhatian personal yang tulus dan individual diberikan perusahaan kepada nasabah	Kepedulian *frontliner*, pemahaman kebutuhan nasabah

Seiring waktu, model *SERVQUAL* banyak diaplikasikan dalam berbagai sektor, termasuk perbankan. Namun, dalam konteks perbankan modern, penelitian oleh Pakurár dkk. (2019) menemukan bahwa dimensi *Tangible* dan *People* (yang mencakup *responsiveness, assurance, dan empathy*) merupakan faktor dominan yang memengaruhi kepuasan nasabah. Hal ini sejalan dengan kenyataan bahwa layanan perbankan tidak hanya dinilai dari kecanggihan fasilitas fisik, tetapi juga dari kualitas interaksi antara pegawai dan nasabah. Selanjutnya, Donthu & Gustafsson (2020) menunjukkan bahwa pandemi COVID-19 turut mengubah pola evaluasi kualitas layanan, di mana interaksi langsung menurun drastis sementara perhatian pada fasilitas fisik dan kebersihan meningkat signifikan. Secara khusus, dalam konteks BRK Syariah, dimensi *People* pada instrumen IKP merupakan representasi operasional dari tiga dimensi *intangible* SERVQUAL, yaitu *responsiveness* (ketanggapan), *assurance* (jaminan), dan *empathy* (empati). Sementara itu, dimensi *Tangible* merepresentasikan aspek fisik layanan sebagaimana definisi aslinya dalam SERVQUAL.

1) Dimensi *Tangible*

Dimensi *Tangible* adalah salah satu unsur utama dalam model kualitas layanan *SERVQUAL* yang diperkenalkan oleh (Parasuraman dkk., 1988). *Tangible* merujuk pada aspek fisik layanan yang dapat diamati secara langsung oleh pelanggan, seperti fasilitas gedung, ketersediaan sarana dan peralatan, serta kebersihan lingkungan. Pakurár dkk. (2019) menyatakan bahwa *Tangible* berperan signifikan terhadap kualitas pelayanan, khususnya pada sektor jasa seperti perbankan. Pelanggan menilai kualitas layanan dari kondisi fasilitas dan sarana yang teramati langsung olehnya.

Pada PT Bank Riau Kepri Syariah menetapkan Indeks Kualitas Pelayanan (IKP) sebagai instrumen evaluasi kualitas pelayanan kantor unit melalui BPP Kualitas Pelayanan Nomor 041 Tahun 2022. Tim *Service Quality* menggunakan instrumen tersebut untuk menilai dimensi *People* dan *Tangible*. Dimensi *People* mencakup keramahan, ketanggapan, dan kesigapan petugas layanan. Dimensi *Tangible* mencakup keberadaan dan kondisi fasilitas fisik kantor unit. Sistem kemudian mengolah hasil penilaian menjadi skor IKP dan menetapkan kategori layanan berdasarkan rentang nilai yang berlaku.

2) Dimensi *People*

Dimensi *People* dalam konteks kualitas layanan merujuk pada aspek interaksi manusia dalam proses pelayanan, yang mencakup sikap, perilaku, dan kemampuan pegawai dalam memberikan layanan kepada pelanggan. Dalam kerangka SERVQUAL, dimensi ini merupakan gabungan dari tiga unsur: responsiveness (ketanggapan pegawai dalam membantu nasabah), *assurance* (kemampuan memberi jaminan, rasa aman, serta pengetahuan yang memadai), dan empathy (kepedulian dan perhatian personal terhadap nasabah). Adapun Parasuraman dkk. (1988) menekankan bahwa kualitas interaksi pegawai merupakan faktor kunci dalam pembentukan persepsi layanan, sementara Pakurár dkk. (2019) menambahkan bahwa faktor manusia berperan sebagai pembeda signifikan dalam menciptakan loyalitas jangka panjang pelanggan. Adapun indikator *People* di PT. Bank Riau Kepri Syariah di dalam BPP Layanan *Service Quality* seperti:

keramahan *customer service*, ketanggapan *teller* dalam melayani transaksi, serta kesigapan satpam dalam membantu nasabah.

2.1.3 Indeks Kualitas Pelayanan (IKP)

Indeks Kualitas Pelayanan (IKP) merupakan target evaluasi yang dikembangkan secara internal oleh PT Bank Riau Kepri Syariah melalui Buku Pedoman Pelaksanaan (BPP) Kualitas Pelayanan *Service Quality* BRK Syariah Nomor 041 (2022). IKP disusun berdasarkan hasil penilaian tahunan dengan menggunakan *checklist* indikator layanan yang mencakup dimensi *Tangible* (fasilitas fisik, sarana penunjang, restroom dan ruangan ATM) serta dimensi *People* (keramahan, ketanggapan, dan kesigapan pegawai *frontliner* seperti *customer service*, *teller*, dan satpam). Meskipun berbasis pada praktik manajerial perusahaan, penyusunan indikator IKP sejatinya mengacu pada kerangka konseptual *SERVQUAL* yang diperkenalkan oleh Parasuraman dkk. (1988) yang menekankan pentingnya lima dimensi kualitas layanan, di mana *Tangible* dan *People* terbukti paling relevan dalam konteks perbankan (Pakurár dkk., 2019). Dengan demikian, IKP dapat dipandang sebagai adaptasi dari konsep kualitas layanan yang telah teruji secara akademik, sekaligus sebagai instrumen dalam mengevaluasi kinerja unit layanan kantor bank. Berikut ini Tabel Indeks Kualitas Pelayanan dari PT.Bank Riau Kepri Syariah :

Tabel 2.2 Indeks Kualitas Pelayanan
Sumber: BPP Kualitas Layanan BRK Syariah No. 41 Tahun 2022

Rentang IKP	**Label**
85 – 100	Sangat Baik
80 – 84	Baik
75 – 79	Cukup
70 – 74	Kurang
< 70	Buruk

Berdasarkan tabel di atas menjelaskan untuk penetapan ambang batas tersebut didasarkan pada standar mutu layanan dan kebijakan yang berlaku pada PT. Bank Riau Kepri Syariah, sehingga dapat memastikan adanya konsistensi dalam penilaian antar kantor unit. Kategorisasi IKP ini memiliki beberapa tujuan strategis, antara lain:

1) Memberikan gambaran yang jelas dan terstandarisasi mengenai kualitas pelayanan setiap kantor unit.
2) Memudahkan manajemen dalam memonitor serta membandingkan capaian kualitas secara berkala.
3) Menjadi dasar pengambilan keputusan strategis, seperti pemberian penghargaan, penetapan program pembinaan, atau implementasi upaya perbaikan layanan.
4) Mendukung pelaporan dan komunikasi hasil evaluasi kepada seluruh pemangku kepentingan secara sistematis dan mudah dipahami.

2.1.4 Self Assessment

Self-assessment merupakan pendekatan evaluasi di mana suatu unit organisasi melakukan penilaian kinerja secara internal berdasarkan instrumen yang telah ditetapkan. Pendekatan ini banyak digunakan dalam bidang pendidikan (Boud & Falchikov, 2006) karena memberikan otonomi kepada individu atau organisasi untuk menilai kualitas kinerjanya sendiri secara sistematis. Dalam konteks organisasi, *self-assessment* sering dikaitkan dengan kerangka kerja manajemen mutu seperti *European Foundation for Quality Management* (EFQM) yang menekankan peran evaluasi internal untuk meningkatkan kinerja secara berkelanjutan (Calvo-Mora dkk., 2015).

Namun, *self-assessment* juga memiliki kelemahan yang signifikan. Penilaian internal cenderung dipengaruhi oleh bias responden, sehingga rentan menghasilkan hasil yang tidak objektif dan tidak konsisten (Kahneman dkk., 2021). Selain itu, hasil evaluasi berbasis *checklist* sederhana sering kali hanya bersifat deskriptif dan belum memberikan wawasan analitis untuk mendukung pengambilan keputusan strategis. Oleh karena itu, meskipun *self-assessment* penting sebagai mekanisme awal evaluasi, dibutuhkan pendekatan lanjutan berbasis data, seperti *machine learning*, untuk meningkatkan objektivitas, konsistensi, dan prediktabilitas hasil evaluasi.

2.1.5 Konsep Rule-based System

Rule-based system adalah sistem yang membuat keputusan berdasarkan sekumpulan aturan logis yang telah ditetapkan secara eksplisit dalam format kondisi-aksi (*if-then*). Sistem ini sangat sesuai untuk domain di mana aturan keputusan sudah terdefinisi dengan jelas, konsisten, dan harus dapat diaudit secara

transparan. Berdasarkan penelitian Manulak dkk. (2024) mengimplementasikan sistem DSS berbasis *rule-based* untuk evaluasi kinerja pegawai berbasis *web* dan menemukan bahwa sistem tersebut mencapai skor kegunaan (*usability*) rata-rata 80,4%, mengonfirmasi efektivitasnya sebagai alat evaluasi yang konsisten dan efisien.

2.1.6 Machine Learning

Machine learning (ML) merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data secara otomatis tanpa perlu pemrograman khusus untuk setiap tugas (Geron, 2022). Dalam penelitian ini, implementasi model *machine learning* akan dilakukan menggunakan bahasa pemrograman *Python*, yang dikenal luas karena ekosistem *library* yang kaya dan kemudahan sintaksisnya untuk pengembangan *Machine Learning*. Dengan menggunakan model matematis, *machine learning* mampu mengenali pola-pola kompleks dalam data dan mengaplikasikannya untuk memproses serta menganalisis data baru. Dalam konteks evaluasi kualitas layanan, penerapan *machine learning* memberikan keunggulan signifikan, seperti peningkatan akurasi dan efisiensi dalam proses penilaian, deteksi dini terhadap penurunan kualitas layanan, serta identifikasi pola-pola tersembunyi yang sulit ditemukan melalui analisis manual. *Machine learning* dapat dimanfaatkan untuk mengklasifikasikan tingkat kualitas layanan dengan memanfaatkan data survei pelanggan, log operasional layanan, maupun umpan balik digital. Hal ini memungkinkan manajemen melakukan perbaikan layanan secara lebih cepat dan berbasis data (Ashmore dkk., 2021). Selain itu, penerapan *machine learning* juga membuka peluang bagi organisasi untuk melakukan pemantauan kualitas secara *real-time*, mengurangi bias subjektif dalam evaluasi, serta meningkatkan kualitas pelayanan secara keseluruhan.

2.1.7 Tahapan Machine Learning

Tahapan dalam pemodelan *machine learning* untuk evaluasi kualitas layanan merupakan proses sistematis yang bertujuan membangun model klasifikasi berdasarkan data penilaian layanan. Setiap tahap dalam pipeline machine learning, mulai dari pengumpulan dan pemrosesan data, pemilihan fitur, pemodelan, hingga evaluasi dan interpretasi hasil, harus dilakukan secara terstruktur agar hasil

klasifikasi kualitas layanan yang dihasilkan valid, akurat, dan dapat diandalkan. Menurut Ashmore dkk. (2021), keterkaitan setiap tahapan dalam proses ini sangat menentukan keberhasilan pemodelan *machine learning*, terutama dalam konteks evaluasi kualitas layanan. Implementasi pipeline yang sistematis memastikan bahwa model klasifikasi mampu mengidentifikasi kategori kualitas layanan dengan tepat, sehingga dapat mendukung pengambilan keputusan yang berbasis data.

2.1.8 Algoritma Machine Learning

1) Random Forest

   Algoritma *Random Forest* bekerja dengan membangun sejumlah pohon keputusan secara acak, lalu menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat dan tahan terhadap *overfitting* (Breiman, 2001). Dalam konteks klasifikasi kualitas layanan, *Random Forest* sangat relevan karena kemampuannya menangani data tabular dengan banyak fitur, serta memberikan performa yang *robust* terhadap *noise* dan *outlier* yang mungkin ada dalam data penilaian. Setiap pohon keputusan dalam *Random Forest* dibangun menggunakan subset acak dari data latih dan fitur, sehingga korelasi antar pohon dapat diminimalkan dan akurasi generalisasi model meningkat (Breiman, 2001).

2) *Gradient Boosting*

   Gradient Boosting merupakan salah satu metode *ensemble* berbasis *boosting* yang bekerja dengan membangun model prediktif secara bertahap dan sekuensial. Berbeda dengan metode *bagging* seperti *Random Forest* yang membangun pohon-pohon keputusan secara paralel dan independen, *Gradient Boosting* membangun setiap pohon baru dengan tujuan mengoreksi kesalahan prediksi (*residuals*) dari model sebelumnya. Pendekatan ini pertama kali diperkenalkan secara formal oleh Friedman (2001) yang memodelkan *boosting* sebagai proses optimasi fungsi kerugian (*loss function*) dalam ruang fungsional, sehingga setiap iterasi pada dasarnya merupakan langkah gradient descent untuk meminimalkan sisa kesalahan model.

3) *Support Vector Machine* (SVM)

Support Vector Machine (SVM) pertama kali dikembangkan oleh Cortes & Vapnik (1995) sebagai algoritma klasifikasi yang bekerja dengan membangun *hyperplane* optimal pada ruang fitur berdimensi tinggi untuk memisahkan kelas secara maksimal melalui margin terbesar antar kelas (*maximum margin classifier*). Penggunaan fungsi kernel seperti *Radial Basis Function* (RBF), *linear*, dan *polynomial* memungkinkan SVM menangani data yang tidak terpisah secara linear dengan memetakan data ke ruang berdimensi lebih tinggi, sehingga cocok untuk data kompleks dengan banyak fitur *biner* dan numerik seperti indikator *People* dan *Tangible* dalam dataset IKP. Dalam konteks evaluasi kualitas layanan, Indrawan dkk. (2022) menerapkan SVM *multiclass* pada data survei kepuasan layanan kesehatan berbahasa Indonesia dan menunjukkan bahwa SVM mampu menghasilkan klasifikasi yang akurat meskipun ukuran dataset relatif terbatas.

4) *Decision Tree*

Decision Tree didefinisikan sebagai model yang menggambarkan serangkaian keputusan, klasifikasi dan konsistensi berdasarkan fitur *input* (Setiawan dkk., 2023). Model ini bekerja dengan membagi data menjadi lebih kecil dengan menggunakan atribut tertentu, akan menghasilkan struktur pohon keputusan yang terdiri dari *node,*cabang dan *leaf* (daun). Dalam pembangunan model *decision tree*, pemilihan atribut terbaik pada setiap node merupakan tahap yang sangat krusial untuk memastikan struktur pohon yang dihasilkan akurat dan efisien. Terdapat beberapa ukuran yang umum digunakan untuk menentukan atribut pemisah, yaitu *entropy*, *information gain*, dan *gini index*. *Entropy* digunakan untuk mengukur tingkat ketidakpastian atau ketidakteraturan dalam suatu dataset, sehingga semakin tinggi nilai entropy menandakan distribusi kelas yang semakin acak. *Information gain* memanfaatkan konsep entropy dengan cara menghitung penurunan ketidakpastian setelah data dibagi berdasarkan suatu atribut; atribut dengan nilai *information gain* terbesar dipilih sebagai pemisah terbaik (Witten dkk., 2011).

2.1.9 Kerangka CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah metodologi standar yang mendefinisikan siklus pengembangan proyek data mining dan *machine learning* ke dalam enam fase yang saling terhubung: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan penerapan (*deployment*) (Wirth & Hipp, 2000). CRISP-DM dikenal luas karena sifatnya yang iteratif dan *domain-agnostic*, sehingga dapat diterapkan pada berbagai jenis proyek analitik data termasuk klasifikasi kualitas layanan perbankan.

2.1.10 Class Imbalance dalam Klasifikasi

Ketidakseimbangan kelas (*class imbalance*) terjadi ketika distribusi kategori dalam dataset tidak proporsional, sehingga kelas mayoritas mendominasi secara signifikan dibanding kelas minoritas. Kondisi ini menyebabkan model ML cenderung menghasilkan akurasi keseluruhan yang tampak tinggi namun gagal mengidentifikasi kelas minoritas secara akurat, kondisi yang berbahaya dalam konteks manajerial di mana justru kelas minoritas adalah yang paling membutuhkan perhatian. Alamsyah dkk. (2024) dalam *TELKOMNIKA* mengonfirmasi bahwa imbalance data merupakan tantangan utama dalam klasifikasi data perbankan dan menunjukkan bahwa SMOTE efektif meningkatkan performa model secara signifikan.

2.1.11 Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) merupakan salah satu teknik penanganan ketidakseimbangan kelas (*class imbalance*) yang banyak digunakan pada tugas klasifikasi. Ketidakseimbangan kelas dapat menyebabkan model lebih cenderung memprediksi kelas mayoritas sehingga performa pada kelas minoritas menurun, meskipun nilai akurasi terlihat tinggi. SMOTE diperkenalkan oleh Chawla dkk. (2002) dengan prinsip menghasilkan sampel baru secara sintetis pada kelas minoritas melalui interpolasi di antara data minoritas dan *k-nearest neighbors*, sehingga berbeda dengan *oversampling* sederhana yang hanya menduplikasi data dan berisiko meningkatkan *overfitting*.

Dalam implementasi praktis, SMOTE tersedia pada pustaka *imbalanced-learn* dan umum diintegrasikan dalam *pipeline* agar proses resampling dilakukan hanya pada data latih di setiap proses validasi silang, sehingga mengurangi risiko *data leakage* dan menghasilkan evaluasi model yang lebih valid.

2.1.12 Metrik Evaluasi Model Klasifikasi Tidak Seimbang

1) *Accuracy* pada Data Tidak Seimbang

   Accuracy adalah metrik yang menyesatkan pada dataset tidak seimbang karena model yang selalu memprediksi kelas mayoritas pun dapat mencapai akurasi tinggi secara artifisial. Sokolova & Lapalme (2009) dalam kajian sistematis terhadap 24 metrik performa klasifikasi yang diterbitkan di Information Processing & Management menegaskan bahwa penggunaan metrik tunggal tidak cukup untuk mengevaluasi model pada data tidak seimbang.

2) *F1-Score Macro* dan *Weighted*

   F1-*score* adalah rata-rata harmonik antara *precision* dan *recall*, yang memberikan bobot seimbang pada kedua aspek tanpa dipengaruhi dominasi kelas mayoritas. F1-*score macro* menghitung rata-rata F1 masing-masing kelas tanpa pembobotan frekuensi, sehingga kelas minoritas berkontribusi setara dengan kelas mayoritas. metrik ini ideal untuk menilai performa model yang harus adil terhadap semua kelas. F1-*score weighted* menghitung rata-rata F1 dengan pembobotan berdasarkan jumlah sampel di setiap kelas, mencerminkan performa pada distribusi asli data.

3) *Balanced Accuracy* dan *Confusion Matrix*

   Balanced accuracy adalah rata-rata *recall* (sensitivitas) dari setiap kelas, yang secara efektif mengukur kemampuan model mengenali setiap kelas secara proporsional tanpa dipengaruhi distribusi kelas. Nilai *baseline* model acak untuk tiga kelas adalah 0,33, sehingga model yang bermakna harus menghasilkan *balanced accuracy* secara substansial di atas nilai tersebut. *Confusion matrix* berukuran 3×3 dalam penelitian ini memungkinkan identifikasi pola kesalahan klasifikasi spesifik per kelas, yang krusial untuk memahami implikasi praktis dari kesalahan model dalam konteks pengambilan keputusan manajerial

4) Uji Kesepakatan Cohen's Kappa

Cohen's Kappa, yang dilambangkan dengan κ, merupakan ukuran statistik yang diperkenalkan oleh Cohen (1960) untuk mengukur tingkat kesesuaian *(agreement)* antara dua penilai dalam mengklasifikasikan data ke dalam kategori nominal yang saling lepas (*mutually exclusive*) (Cohen, 1960). Keunggulan utama Cohen's Kappa dibandingkan dengan persentase kesepakatan sederhana adalah kemampuannya dalam memperhitungkan kemungkinan kesepakatan yang terjadi secara kebetulan (*chance agreement*). Dengan demikian, Penelitian ini menggunakan Cohen's Kappa untuk membandingkan kategori yang dihasilkan oleh mekanisme *rule-based* dan model *machine learning* pada objek penilaian yang sama.

Koefisien Cohen's Kappa dihitung menggunakan persamaan:

$$\kappa = (P_o - P_e) / (1 - P_e)$$

P_o merupakan proporsi kesepakatan yang teramati, sedangkan P_e merupakan proporsi kesepakatan yang diperkirakan terjadi secara kebetulan. Nilai κ berada pada rentang -1 hingga +1. Nilai -1 menunjukkan ketidaksepakatan sempurna. Nilai 0 menunjukkan kesepakatan yang tidak lebih baik dari pada peluang acak. Nilai +1 menunjukkan kesepakatan sempurna. Berdasarkan skala tersebut, nilai $\kappa < 0{,}00$ menunjukkan kesepakatan yang buruk (*poor agreement*), nilai 0,00–0,20 menunjukkan kesepakatan sangat lemah (*slight agreement*), nilai 0,21–0,40 menunjukkan kesepakatan lemah (*fair agreement*), nilai 0,41–0,60 menunjukkan kesepakatan sedang (*moderate agreement*), nilai 0,61–0,80 menunjukkan kesepakatan kuat (*substantial agreement*), dan nilai 0,81–1,00 menunjukkan kesepakatan hampir sempurna (*almost perfect agreement*).

2.1.13 Python

Python merupakan bahasa pemrograman tingkat tinggi yang bersifat *open source* dan saat ini menjadi salah satu bahasa yang paling banyak digunakan dalam pengembangan sistem kecerdasan buatan dan machine learning (Pedregosa dkk., 2011). *Python* memiliki ekosistem *library* yang komprehensif dan terstruktur untuk mendukung setiap tahapan dalam alur kerja ilmu data (*data science pipeline*), mulai dari pra-pemrosesan data, eksplorasi, hingga pembangunan dan evaluasi model. Beberapa *library* utama yang relevan dalam penelitian ini antara lain NumPy dan

Pandas untuk manipulasi dan analisis data terstruktur (McKinney, 2010), *Scikit-learn* untuk implementasi algoritma klasifikasi seperti *Support Vector Machine* dan *Random Forest* , serta *Imbalanced-learn* untuk penanganan ketidakseimbangan kelas menggunakan metode SMOTE (Chawla dkk., 2002). Selain kemampuan teknis tersebut, *Python* didukung oleh komunitas pengembang yang besar dan aktif secara global, sehingga memungkinkan reprodusibilitas penelitian yang baik serta kemudahan dalam integrasi dan pengembangan sistem berbasis data lebih lanjut.

2.1.14 Streamlit

Streamlit adalah *framework* Python *open-source* yang dirancang khusus untuk memudahkan pengembangan aplikasi web interaktif berbasis data dan *machine learning* (ML) tanpa memerlukan pengetahuan mendalam tentang pengembangan web seperti HTML, CSS, maupun JavaScript. *Framework* ini bekerja dengan mengonversi skrip Python secara langsung menjadi antarmuka web yang responsif dan dilengkapi komponen interaktif seperti *form input*, tabel, grafik, dan *widget* lainnya hanya dengan beberapa baris kode (Streamlit Inc, 2024). Karakteristik ini menjadikan Streamlit pilihan yang ideal bagi ilmuwan data dan peneliti yang ingin membangun prototipe sistem analitik berbasis *Machine Learning* yang dapat diakses oleh pengguna non-teknis secara cepat.

Dalam konteks penelitian, Streamlit telah digunakan secara luas sebagai sarana *deployment* model *Machine Learning* sebagai aplikasi *web* yang dapat dibagikan dan dioperasikan dengan mudah. Studi oleh Wallaard (2023) menunjukkan bahwa Streamlit mampu menjembatani kesenjangan antara pengembangan model *Machine Learning* prediktif dan implementasinya sebagai alat bantu keputusan berbasis *web*, dengan membangun *framework* standar menggunakan *Python* dan Streamlit untuk *deployment* model kesehatan secara lokal maupun daring.

2.1.15 SQLite

SQLite adalah sistem manajemen basis data relasional yang bersifat *serverless, file-based*, dan tidak memerlukan konfigurasi server terpisah, menjadikannya pilihan tepat untuk aplikasi penelitian *prototipe*. Seluruh basis data

SQLite tersimpan dalam satu file .db yang portabel, sehingga mudah dikelola, dibackup, dan dimigrasi. Integrasi Streamlit dengan SQLite memungkinkan penyimpanan histori *self-assessment*, log prediksi, dan data pengguna secara lokal tanpa infrastruktur tambahan, sekaligus menyediakan antarmuka SQL standar untuk ekspor data dan analisis lebih lanjut.

Dalam ekosistem penelitian berbasis *Python*, integrasi antara Streamlit dan SQLite menawarkan solusi yang lengkap dan mandiri untuk membangun sistem analitik prototipe. SQLite dipilih sebagai sistem manajemen basis data lokal karena bersifat *serverless*, *zero-configuration*, dan *transactional*, sehingga cocok digunakan pada aplikasi yang memerlukan akses data yang cepat dan andal tanpa konfigurasi server yang rumit (Arsyanda dkk., 2024). Dalam penelitian ini, SQLite digunakan untuk menyimpan histori *self-assessment*, log prediksi model, dan data pengguna secara lokal, sekaligus menyediakan antarmuka SQL standar yang memungkinkan ekspor data dan analisis lebih lanjut.

2.1.16 Tahapan Pengembangan Sistem

Pada tahapan ini adalah proses merancang dan membangun sistem yang dirancang untuk mengelola data informasi menggunakan *Software Development Life Cycle* (SDLC). *Software Development Life Cycle (SDLC)* adalah kerangka kerja terstruktur untuk mengembangkan perangkat lunak mulai dari perencanaan hingga pemeliharaan. SDLC membantu tim memastikan proses pengembangan berjalan sistematis, menghasilkan perangkat lunak yang sesuai kebutuhan pengguna, dan menjaga kualitas melalui tahapan yang jelas, seperti perencanaan, analisis kebutuhan, desain, implementasi, pengujian, hingga pemeliharaan. (Murdiani & Sobirin, 2022).

SDLC memiliki beberapa model pengembangan, misalnya *Waterfall, V-Model, Agile, Iterative, Spiral, dan Prototype.* Pada penelitian ini digunakan model Waterfall karena pendekatannya bersifat linear dan terdokumentasi, sehingga sesuai untuk pengembangan sistem yang kebutuhannya relatif jelas serta perubahan selama pengembangan cenderung minimal. *Waterfall* menuntut setiap tahap diselesaikan terlebih dahulu sebelum masuk ke tahap berikutnya.

Tahapan model *Waterfall* yang digunakan dalam pengembangan sistem adalah sebagai berikut (Sommerville, 2016) :

1) Analisis kebutuhan: mengidentifikasi kebutuhan pengguna dan fungsi sistem yang harus disediakan.
2) Perancangan (*desain*): menyusun rancangan arsitektur, alur proses, basis data, dan rancangan antarmuka sesuai kebutuhan yang telah ditetapkan.
3) Implementasi: merealisasikan rancangan ke dalam kode program dan membangun komponen sistem.
4) Pengujian: memeriksa apakah sistem berjalan sesuai kebutuhan dan menemukan kesalahan sebelum sistem digunakan.
5) Pemeliharaan: melakukan perbaikan, penyesuaian, dan peningkatan setelah sistem digunakan agar tetap berjalan stabil.

2.2 Penelitian Terdahulu

Penelitian terdahulu pada topik yang berkaitan dengan evaluasi kualitas layanan, klasifikasi berbasis *machine learning*, dan implementasi sistem pendukung keputusan menunjukkan bahwa ketiga aspek tersebut telah banyak dikaji, namun umumnya masih dilakukan secara terpisah. Pada konteks evaluasi kualitas layanan, Altuntaş dkk. (2022) meneliti penggunaan instrumen SERVQUAL yang dipadukan dengan algoritma *machine learning* untuk mengevaluasi kualitas layanan di sektor kesehatan. Hasil penelitiannya menunjukkan bahwa pendekatan *machine learning* dapat membantu mengidentifikasi dimensi layanan yang dominan serta mendukung klasifikasi mutu layanan secara lebih sistematis.

Dalam konteks data tidak seimbang pada sektor perbankan, Alamsyah dkk. (2024) mengembangkan model *stacking ensemble* yang dikombinasikan dengan SMOTE untuk klasifikasi data kredit. Penelitian tersebut menunjukkan bahwa penerapan SMOTE mampu meningkatkan *precision*, *recall*, dan F1-*score* secara signifikan pada data yang mengalami ketidakseimbangan kelas. Hasil ini mendukung penggunaan teknik penanganan *class imbalance* dalam penelitian ini, mengingat distribusi kategori IKP pada data BRK Syariah juga didominasi oleh kelas mayoritas Sangat Baik, sedangkan kelas Cukup muncul dalam jumlah yang

relatif kecil.

Penelitian lain oleh Nanda dkk. (2025) juga menunjukkan efektivitas SMOTE dalam meningkatkan performa *Random Forest* untuk klasifikasi risiko kredit pada data perbankan. Setelah dilakukan penyeimbangan kelas, model menunjukkan peningkatan pada metrik *accuracy*, *precision*, dan *recall*. Temuan tersebut memperkuat argumen bahwa penanganan ketidakseimbangan kelas merupakan aspek metodologis penting dalam klasifikasi data operasional perbankan, khususnya ketika kelas minoritas memiliki makna manajerial yang lebih kritis dibandingkan kelas mayoritas.

Pada ranah komparasi algoritma klasifikasi untuk data layanan, Banjanin dkk. (2022) membandingkan beberapa algoritma *machine learning* untuk klasifikasi kualitas pengalaman layanan telekomunikasi. Hasil penelitian menunjukkan bahwa *Gradient Boosting* memberikan performa tertinggi pada data yang digunakan. Sementara itu, Indrawan dkk. (2022) meneliti penerapan SVM *multiclass* pada data survei kepuasan layanan kesehatan berbahasa Indonesia dan menemukan bahwa SVM mampu memberikan hasil klasifikasi yang baik meskipun ukuran dataset terbatas. Kedua penelitian tersebut relevan karena memberikan landasan bahwa algoritma seperti *Gradient Boosting* dan SVM layak dipertimbangkan sebagai kandidat model klasifikasi pada data layanan yang bersifat *multiclass*.

Dari sisi konseptual kualitas layanan, Pakurár dkk. (2019) menunjukkan bahwa dimensi *Tangible* dan aspek interaksi manusia merupakan faktor yang berpengaruh terhadap kepuasan nasabah pada sektor perbankan. Temuan ini mendukung fokus penelitian ini yang menggunakan dua kelompok indikator utama, yaitu *People* dan *Tangible*, sebagai representasi operasional dari kualitas pelayanan kantor unit BRK Syariah. Dengan demikian, pemilihan variabel penelitian tidak hanya didasarkan pada instrumen internal organisasi, tetapi juga memiliki justifikasi teoretis dari studi sebelumnya.

Sementara itu, dari sisi pengembangan sistem berbasis aturan, Manulak dkk. (2024) mengembangkan sistem pendukung keputusan *rule-based* berbasis *web* untuk evaluasi kinerja dan kepuasan layanan. Penelitian tersebut menunjukkan bahwa pendekatan *rule-based* efektif digunakan ketika aturan keputusan telah terdefinisi dengan jelas dan perlu diterapkan secara konsisten. Studi ini relevan

dengan penelitian yang dilakukan karena evaluasi IKP di BRK Syariah juga mengacu pada aturan internal organisasi yang telah baku, sehingga mekanisme *rule-based* menjadi dasar yang tepat untuk perhitungan dan kategorisasi hasil penilaian.

Berdasarkan telaah terhadap penelitian-penelitian terdahulu tersebut, dapat diketahui bahwa setiap studi memberikan kontribusi yang penting, tetapi masih memiliki fokus yang parsial. Sebagian penelitian menitikberatkan pada analisis kualitas layanan berbasis instrumen SERVQUAL, sebagian lain menekankan performa algoritma *machine learning* pada data layanan atau perbankan, sedangkan penelitian lainnya lebih berfokus pada pengembangan sistem evaluasi berbasis aturan. Dengan demikian, masih terdapat ruang penelitian pada integrasi ketiga aspek tersebut dalam satu kerangka penelitian terapan.

Penelitian ini memiliki perbedaan sekaligus kontribusi dibandingkan penelitian sebelumnya. Pertama, penelitian ini tetap mempertahankan mekanisme *rule-based* sebagai dasar resmi penetapan skor dan kategori IKP sesuai kebijakan internal BRK Syariah. Kedua, penelitian ini memanfaatkan data historis penilaian untuk membangun model klasifikasi *multiclass* berbasis *machine learning* dengan perhatian khusus pada masalah ketidakseimbangan kelas. Ketiga, keluaran *rule-based* dan *machine learning* diintegrasikan ke dalam satu *prototipe self-assessment* berbasis *web* yang mendukung kebutuhan penilai, administrator, dan Tim *Service Quality*. Dengan demikian, penelitian ini tidak hanya menguji performa model klasifikasi, tetapi juga menghasilkan sistem terapan yang dapat mendukung *monitoring* kualitas pelayanan secara lebih cepat, terdokumentasi, dan analitis.

Tabel 2.3 Perbandingan Penelitian Terdahulu

Peneliti	Fokus Penelitian	Metode/Algoritma	Data/Objek	Temuan Utama	Relevansi dengan Penelitian
Altuntaş dkk. (2022)	Evaluasi kualitas layanan	SERVQUAL + ML	Layanan kesehatan	ML membantu identifikasi dimensi layanan dominan	Dasar penerapan ML pada evaluasi layanan
Alamsyah dkk. (2024)	Klasifikasi data tidak seimbang	Stacking + SMOTE	Data kredit perbankan	SMOTE meningkatkan	Dasar penanganan

Peneliti	Fokus Penelitian	Metode/Algoritma	Data/Objek	Temuan Utama	Relevansi dengan Penelitian
				performa klasifikasi	*class imbalance*
Nanda dkk. (2025)	Klasifikasi risiko kredit	RF + SMOTE	Data perbankan	SMOTE meningkatkan accuracy, precision, recall	Dukungan metodologis pada data perbankan
Banjanin dkk. (2022)	Klasifikasi kualitas layanan	DT, GB, SVM	Layanan telekomunikasi	Gradient Boosting unggul	Dasar komparasi algoritma
Indrawan dkk. (2022)	Klasifikasi *multiclass*	SVM *multiclass*	Survei layanan kesehatan	SVM efektif pada data layanan	Dukungan pemilihan SVM
Pakurár dkk. (2019)	Dimensi kualitas layanan	SERVQUAL	Perbankan	*Tangible* dan aspek *People* berpengaruh	Justifikasi variabel penelitian
Manulak dkk. (2024)	Sistem evaluasi berbasis aturan	*Rule-based* DSS	Evaluasi layanan/kinerja	*Rule-based* efektif dan konsisten	Dasar mekanisme *rule-based* dalam sistem

Berdasarkan Tabel 2.3, bahwa penelitian terdahulu yang telah diuraikan, dapat diidentifikasi keterkaitan antara landasan teori, kesenjangan (gap) penelitian yang ada, dan implementasi dalam penelitian ini. Tabel 2.4 menyajikan matriks sinkronisasi teori, gap, dan model secara eksplisit untuk mempertegas koherensi kerangka penelitian ini.

Tabel 2.4 Matriks Sinkronisasi Teori, Gap, dan Implementasi Penelitian

No	Teori / Konsep	Gap yang Diatasi	Implementasi dalam Penelitian
1	SERVQUAL — dimensi *Tangible* dan *People* (Parasuraman dkk., 1988; Pakurár dkk., 2019)	Belum ada sistem evaluasi terstruktur berbasis dua dimensi People–Tangible di konteks BRK Syariah	Basis 142 indikator biner: 70 *People* dan 72 *Tangible* sebagai variabel masukan sistem

2	*Class Imbalance* dan SMOTE (Chawla dkk., 2002; Alamsyah dkk., 2024)	Dataset IKP didominasi kelas Sangat Baik (77,6%), kelas Cukup hanya 4,1% — menyebabkan *accuracy* semu jika tidak ditangani	*Pipeline* SMOTE + GridSearchCV dengan F1-*macro* sebagai kriteria seleksi utama
3	Cohen's Kappa (*inter-rater agreement*) (Cohen, 1960; Landis & Koch, 1977)	Belum ada metrik validasi kesepakatan antara sistem *rule-based* dan ML pada konteks IKP perbankan	Pengujian eksternal 21 data operasional 2025–2026 menggunakan Cohen's Kappa
4	*Rule-based System* (Manulak dkk., 2024)	Proses manual tidak konsisten dan tidak terdokumentasi secara terpusat	Mesin *rule-based* deterministik sesuai BPP No. 041 sebagai keluaran resmi IKP
5	CRISP-DM (Wirth & Hipp, 2000)	Belum ada kerangka sistematis pengembangan ML untuk IKP di konteks perbankan syariah	Kerangka tahapan penelitian mengikuti fase CRISP-DM: *Business Understanding* → *Data Understanding* → *Data Preparation* → *Modeling* → *Evaluation* → *Deployment*
6	SVM, Random Forest, Decision Tree, Gradient Boosting (Cortes & Vapnik, 1995; Breiman, 2001; Friedman, 2001)	Belum ada studi komparasi empat algoritma pada data IKP multiclass perbankan syariah	Eksperimen komparatif empat algoritma dengan *pipeline* terstandarisasi dan evaluasi multi-metrik