

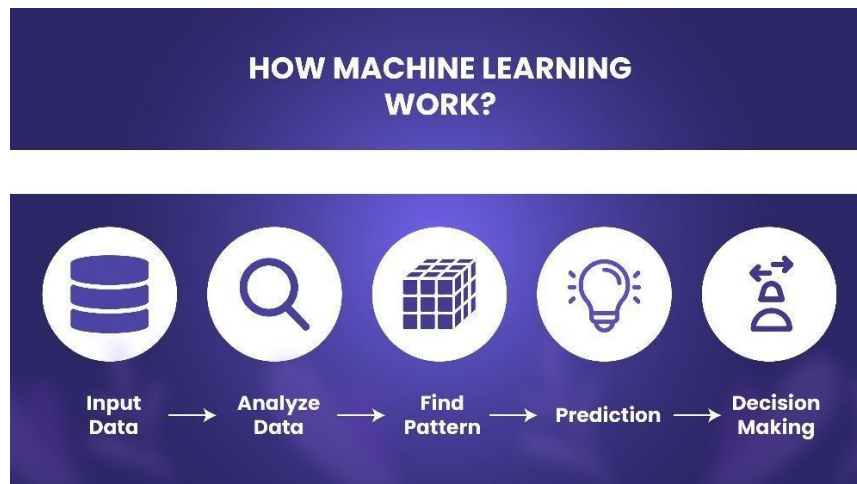
BAB II

KERANGKA TEORITIS

A. Landasan Teori

1. Machine Learning

Machine learning adalah pendekatan berbasis data yang memungkinkan komputer untuk belajar dari data historis dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. Dalam konteks validasi data, algoritma ini dapat digunakan untuk memprediksi nilai yang hilang atau tidak valid berdasarkan pola yang ditemukan dalam dataset.



Gambar 2. 1 Cara Kerja Machine Learning

(Sumber : (Fauzan et al., 2024))

Cara kerja machine learning yaitu Input Data sebagai masukan awal, diikuti oleh *Analyze Data* untuk memahami pola dan struktur. Kemudian, model dilatih pada tahap *Find Pattern* untuk menemukan hubungan dalam data. Setelah itu, model digunakan untuk *Prediction* menghasilkan output berdasarkan data baru, dan hasilnya digunakan dalam *Decision Making* untuk mendukung atau mengotomatiskan pengambilan keputusan.

2. Form Digital

Formulir digital merupakan alat berbasis web yang memungkinkan pengguna untuk membuat dan mengisi survei atau kuesioner secara online dengan mudah. Alat ini sering digunakan untuk mengumpulkan data dari responden dalam berbagai bentuk, seperti teks, pilihan ganda, kotak centang, dan jenis input lainnya. Salah satu fitur utama dari formulir digital adalah kemampuannya untuk mengumpulkan data secara otomatis dan menyimpannya dalam format yang terstruktur pada basis data atau lembar kerja daring.

Dalam konteks otomatisasi, formulir digital dapat diintegrasikan dengan berbagai aplikasi lain melalui API atau layanan middleware untuk mempermudah proses pengisian dan pengumpulan data tanpa memerlukan interaksi manual. Hal ini memungkinkan penggunaan yang lebih efisien dalam pengumpulan data berskala besar, sekaligus mengurangi potensi kesalahan manusia dan meningkatkan kecepatan serta akurasi proses pengolahan data.

3. Agile

Agile adalah metodologi manajemen proyek yang fokus pada pengembangan iteratif dan inkremental, dengan tujuan untuk memberikan hasil yang lebih cepat dan fleksibel dalam menghadapi perubahan kebutuhan proyek. Pendekatan Agile mengutamakan kolaborasi tim, komunikasi yang terbuka, dan adaptasi berkelanjutan terhadap perubahan, dibandingkan dengan pendekatan manajemen proyek tradisional yang lebih terstruktur dan kaku.

Dalam konteks *Agile Project Management*, proses iteratif dilakukan dalam bentuk sprint atau siklus waktu yang pendek, di mana setiap iterasi menghasilkan produk yang dapat digunakan dan diuji. Hal ini memungkinkan tim untuk merespons perubahan dengan cepat, baik itu perubahan dalam kebutuhan pengguna, teknologi, maupun prioritas bisnis. Agile mengedepankan nilai-nilai seperti keterbukaan, keberlanjutan, dan penyesuaian diri terhadap perubahan, yang semua itu sangat penting untuk keberhasilan proyek dalam lingkungan yang dinamis.



Gambar 2. 2 Alur Agile

(Sumber : (Castillo et al., 2024))

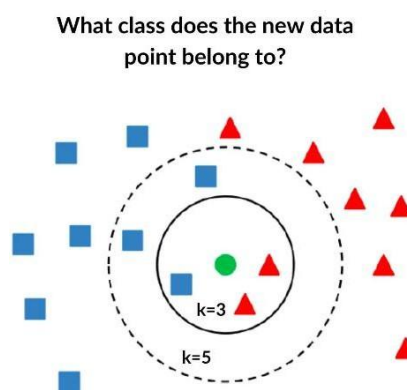
Metode Agile yang digambarkan dalam diagram tersebut adalah sebuah siklus iteratif yang terdiri dari beberapa langkah yang saling terkait. Berikut adalah penjelasan cara kerja setiap langkah dalam siklus Agile yang digambarkan:

1. **Plan (Perencanaan):** Proses dimulai dengan perencanaan. Pada tahap ini, tim menentukan apa yang ingin dicapai dalam siklus pengembangan dan menetapkan tujuan serta prioritas untuk fitur yang akan dikembangkan. Hal ini mencakup pengumpulan kebutuhan, pemahaman tentang masalah yang akan diselesaikan, serta penetapan batasan-batasan proyek.
2. **Design (Desain):** Setelah perencanaan, tim melanjutkan dengan merancang solusi untuk masalah yang telah diidentifikasi. Desain mencakup aspek teknis dan visual dari sistem yang akan dibangun, serta pembuatan rencana arsitektur.
3. **Develop (Pengembangan):** Pada tahap ini, tim mulai mengembangkan dan membangun produk sesuai dengan desain yang telah disepakati. Pengembangan dilakukan secara iteratif, dengan menyelesaikan sebagian dari fitur atau komponen produk pada setiap siklus.
4. **Test (Pengujian):** Setelah pengembangan, produk diuji untuk memastikan bahwa fungsionalitas dan fitur yang dikembangkan bekerja dengan baik. Pengujian ini dilakukan secara menyeluruh untuk menemukan bug atau masalah yang perlu diperbaiki sebelum rilis.

5. Release (Rilis): Setelah produk diuji dan dianggap siap, fitur atau bagian dari produk yang telah selesai akan dirilis kepada pengguna. Proses rilis ini bisa melibatkan deployment ke lingkungan produksi atau memberikan akses kepada pengguna untuk mencoba fitur baru.
6. Feedback (Umpan Balik): Setelah rilis, tim mengumpulkan umpan balik dari pengguna atau stakeholder untuk mengevaluasi hasil yang dicapai. Feedback ini penting untuk mengetahui apakah produk memenuhi harapan dan kebutuhan pengguna. Feedback tersebut akan digunakan untuk perbaikan dalam siklus berikutnya.

4. K-Nearest Neighbors (KKN)

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran supervised yang digunakan terutama untuk tugas klasifikasi (dan dalam beberapa kasus regresi), di mana prediksi dibuat berdasarkan k tetangga terdekat dari data yang diuji. Algoritma ini digolongkan sebagai instance-based learning atau lazy learning, karena tidak membangun model eksplisit selama pelatihan, melainkan menyimpan data latih lengkap dan melakukan komputasi saat melakukan prediksi.



Gambar 2. 3 K-Nearest Neighbors

(Sumber : (Wang et al., 2023))

Cara kerja KNN dimulai dengan memilih parameter k , yaitu jumlah tetangga terdekat yang akan dipertimbangkan. Kemudian, untuk setiap data baru yang akan diklasifikasi, algoritma menghitung jarak antara data tersebut dengan semua data dalam dataset latih. Jarak yang umum digunakan adalah Euclidean distance,

meskipun metrik lain seperti Manhattan atau Minkowski juga bisa digunakan tergantung jenis data dan fitur.

Setelah menghitung jarak, algoritma memilih k data latih dengan jarak paling dekat. Dari kelompok k tetangga ini, kelas terbanyak (mayoritas) digunakan sebagai prediksi kelas untuk data baru tersebut. Dalam klasifikasi biner, hal ini berarti mayoritas tetangga menentukan apakah data baru dikategorikan sebagai “kosong” atau “tidak kosong” dalam konteks validasi kolom formulir.

Pemilihan nilai k sangat penting karena mempengaruhi bias dan varians model. Jika k terlalu kecil, model menjadi sangat sensitif terhadap noise (overfitting); jika terlalu besar, model menjadi terlalu umum dan bisa kehilangan detail penting (underfitting). Salah satu metode umum untuk memilih k adalah dengan cross-validation, di mana beberapa nilai k diuji lalu dibandingkan performanya berdasarkan metrik seperti akurasi.

Sebelum klasifikasi dilakukan, penting untuk melakukan pra-pemrosesan data—termasuk normalisasi fitur agar skala antar fitur yang berbeda tidak mendominasi perhitungan jarak. Misalnya, kolom umur atau numerik lainnya harus dinormalisasi (misalnya min-max scaling atau standardization), sedangkan fitur kategorikal harus dikonversi ke bentuk numerik (contoh: one-hot encoding). Jika tidak, fitur dengan rentang nilai besar dapat mendistorsi jarak Euclidean dan menurunkan akurasi model.

Kelebihan KNN antara lain kemudahannya dalam implementasi, efektif untuk dataset yang relatif kecil hingga menengah, dan fleksibilitas karena tidak membuat asumsi distribusi data. Namun, terdapat juga beberapa kelemahan: waktu prediksi bisa menjadi lambat ketika dataset latih sangat besar karena perhitungan jarak harus dilakukan terhadap semua data latih; sensitivitas terhadap fitur yang tidak relevan atau skala yang berbeda; serta kemungkinan kesalahan akibat bias class imbalance jika satu kelas jauh lebih banyak dari kelas lainnya.

5. TensorFlow Framework

TensorFlow adalah sebuah framework open-source yang dikembangkan oleh Google untuk membangun dan melatih model machine learning serta deep learning. Framework ini menggunakan konsep grafik komputasi, di mana operasi matematis diwakili sebagai grafik, dengan node yang mewakili operasi dan tepi yang mewakili tensor, yaitu struktur data dasar dalam TensorFlow yang berbentuk array multidimensi. TensorFlow dirancang untuk efisiensi dan kinerja tinggi, memungkinkan pengguna

untuk memanfaatkan CPU dan GPU guna mempercepat proses pelatihan model. Selain itu, TensorFlow dapat dijalankan di berbagai platform, termasuk desktop, server, dan perangkat mobile, sehingga memudahkan pengembang dalam mengimplementasikan aplikasi machine learning di berbagai lingkungan. Framework ini juga mendukung Keras, sebuah API tingkat tinggi yang menyediakan antarmuka yang lebih sederhana dan intuitif untuk membangun dan melatih model neural network. Dengan berbagai fitur dan kemudahan yang ditawarkannya, TensorFlow telah menjadi salah satu alat utama dalam penelitian dan pengembangan machine learning, digunakan dalam berbagai aplikasi seperti pengenalan gambar, pemrosesan bahasa alami, dan analisis data.

Menurut (Suvra Das, 2024) Menyatakan bahwa TensorFlow sebagai sistem machine learning berskala besar. TensorFlow menggunakan grafik komputasi untuk mendefinisikan dan menjalankan model machine learning. Jurnal ini juga membahas berbagai aplikasi TensorFlow dalam pengolahan data, pelatihan model, dan penerapan dalam berbagai domain, termasuk visi komputer dan pemrosesan bahasa alami. TensorFlow memungkinkan pengguna untuk membangun model yang kompleks dengan efisiensi tinggi dan kemudahan dalam pengelolaan data besar.

B. Tinjauan Studi (Penelitian Rujukan)

Penelitian ini merujuk pada beberapa penelitian terdahulu yang relevan:

1. Penelitian ini dilakukan oleh (Belgacem et al., 2022) dengan judul Skripsi “***A Machine Learning Approach for Automated Filling of Categorical Fields in Data Entry Forms***” mengusulkan penggunaan machine learning untuk mengisi kolom data kategorikal secara otomatis dalam formulir entri data. Pendekatan ini dapat diterapkan pada Form digital untuk validasi dan pengisian data kosong, di mana algoritma machine learning mempelajari pola dari data yang ada untuk mengisi kolom yang kosong berdasarkan informasi yang relevan.
2. Penelitian ini dilakukan oleh (Shankar et al., 2023) dengan judul Skripsi “***Automatic and Precise Data Validation for Machine Learning***” menjelaskan penggunaan metode validasi otomatis untuk meningkatkan kualitas data yang digunakan dalam model machine learning. Penelitian ini menawarkan metode yang dapat digunakan untuk mendeteksi dan memperbaiki masalah data yang tidak valid atau hilang, yang sangat relevan dengan pengaplikasian machine learning dalam validasi data pada Form digital untuk memastikan akurasi dan konsistensi data.

3. Penelitian ini dilakukan oleh (Fauzan et al., 2024) dengan judul Skripsi **“Penerapan Machine Learning pada Analisis Sentimen Aplikasi MyTelkomsel Menggunakan Data Ulasan Google Playstore** “Penelitian ini menunjukkan penerapan machine learning untuk menganalisis dan mengisi data kosong dalam konteks analisis sentimen. menggunakan machine learning untuk mendeteksi pola dan melengkapi data yang hilang dalam dataset ulasan aplikasi. Teknik serupa dapat diterapkan pada Form digital untuk memvalidasi dan mengisi data yang kosong secara otomatis, memastikan integritas data yang terkumpul dan mempercepat proses pengumpulan informasi.
4. Penelitian ini dilakukan oleh (Kaur et al., 2022) dengan judul Skripsi **“An Application to Automate the Form digital Submission”** Penelitian ini mengusulkan aplikasi yang dapat mengotomatisasi pengisian formulir Form digital. Dengan otomatisasi ini, data dapat dikumpulkan dan diproses secara realtime, mengurangi kesalahan manusia dan meningkatkan efisiensi. Integrasi machine learning dengan sistem ini dapat memvalidasi data secara otomatis saat pengisian formulir, memperbaiki data yang kosong atau tidak valid pada saat formulir diisi.
5. Penelitian ini dilakukan oleh (Lin et al., 2024) dengan judul Skripsi **“Automatic String Data Validation with Pattern Discovery”** mengembangkan algoritma yang menggunakan penemuan pola untuk validasi data string secara otomatis. Dalam konteks Form digital, pendekatan ini dapat digunakan untuk memverifikasi dan memperbaiki data berbentuk string yang hilang atau tidak konsisten, sehingga meningkatkan akurasi dan efisiensi pengolahan data.
6. Penelitian ini dilakukan oleh (Yining et al., 2022) dengan judul Skripsi **“Verification and Validation Methods for Decision-Making and Planning of Automated Vehicles: A Review”** Meskipun penelitian ini berfokus pada kendaraan otonom, metode verifikasi dan validasi yang dijelaskan dalam studi ini dapat diterapkan pada sistem pengumpulan data. Metode verifikasi untuk memastikan kualitas dan keandalan data sangat relevan dalam konteks validasi data pada Form digital, di mana kesalahan atau data yang hilang harus diperbaiki untuk memastikan akurasi dan ketepatan keputusan berbasis data.

7. Penelitian ini dilakukan oleh (Triyono et al., 2021) dengan judul Skripsi **Perancangan Arsitektur Aplikasi Web Menggunakan Data Real Time Dari Form digital**. Penelitian ini membahas perancangan aplikasi web yang memanfaatkan data real-time dari Form digital. Dalam konteks ini, data yang dikumpulkan melalui Form digital harus melalui proses validasi untuk memastikan bahwa data yang dikirimkan akurat dan lengkap. Mesin pembelajaran dapat diintegrasikan untuk memvalidasi dan melengkapi data kosong secara otomatis, yang meningkatkan keandalan aplikasi berbasis web.
8. Penelitian ini dilakukan oleh (Triyono et al., 2021) dengan judul Skripsi ***“Using Form digitals as an Online Learning Assessment Tool for Non-EFL Students”***. Penelitian ini membahas penggunaan Form digitals sebagai alat penilaian pembelajaran daring untuk siswa non-EFL. Dalam konteks ini, data yang dikumpulkan melalui Form digitals perlu melalui proses validasi otomatis untuk memastikan bahwa informasi yang terkumpul akurat dan lengkap. Dengan menggunakan teknologi otomatisasi, proses penilaian dapat lebih efisien, mengurangi kesalahan manusia, dan memberikan umpan balik yang lebih cepat kepada siswa, sehingga meningkatkan kualitas dan keandalan sistem penilaian pembelajaran.
9. Penelitian ini dilakukan oleh (Castillo et al., 2024) dengan judul Skripsi ***“A Systematic Literature Review on Machine Learning Applications for Agile Project Management”***. mengkaji penerapan machine learning dalam manajemen proyek Agile, khususnya dalam estimasi proyek, pengelolaan risiko, kolaborasi tim, dan pengalokasian sumber daya. Machine learning dapat meningkatkan akurasi estimasi dengan menganalisis data historis proyek, memprediksi risiko, serta mengoptimalkan pengalokasian sumber daya. Selain itu, teknik ini juga dapat menganalisis dinamika tim untuk meningkatkan kolaborasi dan memberikan rekomendasi berbasis data untuk perencanaan sprint yang lebih efisien, meningkatkan efisiensi dan hasil proyek dalam metodologi Agile.
10. Penelitian ini dilakukan oleh (Castillo et al., 2024) dengan judul Skripsi ***Data Science and Its Importance***. membahas pentingnya Data Science dalam pengolahan dan analisis data. Data Science memiliki peran penting dalam menerapkan algoritma machine learning untuk validasi dan pembersihan data. Pendekatan ini relevan untuk meningkatkan akurasi dan reliabilitas data yang dikumpulkan melalui Form digital, terutama dalam konteks validasi data yang kosong.

Tabel 2. 1 Kontribusi Jurnal

No	Author(s)	Judul	Jurnal	Kontribusi
1	Belgacem Hichem, Li Xiaochen, Bianculi D, Briand L, 2022	A Machine Learning Approach for Automated Filling of Categorical Fields in Data Entry Forms	<u>ACM Transactions on Software Engineering and Methodology, Volume 32, I ssue 2</u> Article No.: 47, Pages 1 - 40	Kontribusi jurnal ini mengusulkan penggunaan machine learning untuk mengisi kolom data kategorikal secara otomatis, yang relevan untuk validasi dan pengisian data kosong di Google Form
2	Shankar Shreya, Fawaz Labib, Gyllstrom Karl, Parameswara n Aditya, 2023	Automatic and Precise Data Validation for Machine Learning	<u>CIKM '23: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management</u> Pages 2198 - 220	Kontribusi jurnal ini menjelaskan metode validasi otomatis untuk meningkatkan kualitas data

No	Author(s)	Judul	Jurnal	Kontribusi
3	Fauzan Farin, Afdal M, Novita Rice, Mustakim, 2024	Penerapan Machine Learning pada Analisis Sentimen Aplikasi MyTelkomsel Menggunakan Data Ulasan Google Playstore	<u>Indonesian Journal of Computer Science 13(3)</u>	Kontribusi jurnal ini menunjukkan penggunaan machine learning menganalisis dan mengisi data kosong, yang dapat diterapkan pada Form digital
4	Kaur P, K D, Vinjamuri L, 2022	An Application to Automate the Form digital Submission	<u>2022 International Conference on Cyber Resilience (ICCR)</u>	Kontribusi jurnal ini mengusulkan aplikasi untuk mengotomatisasi pengisian formulir Form digital
5	Lin Xinwei, Zhao Jing, Di Peng, Xiao Chuan, Mao Rui, Ji Yan, Onizuka Makoto, Ding Zishuo, Shang Weiyi, Qin Jianbin, 2024	Automatic String Data Validation with Pattern Discovery	arXiv:2408.03005	Kontribusi jurnal ini, yang digunakan untuk memperbaiki data string yang hilang atau tidak konsisten dalam Form digital.

No	Author(s)	Judul	Jurnal	Kontribusi
6	Yining, Sun Chen, Chen Junyi, Cao Dongpu, Xiong Lu, 2022	Verification and Validation Methods for Decision Making and Planning of Automated Vehicles: A Review	<u>IEEE Transactions on Intelligent Vehicles</u> (Volume: 7, <u>Issue: 3</u> , September 2022)	Kontribusi jurnal ini mengulas metode verifikasi dan validasi yang dapat diterapkan pada sistem pengumpulan data
7	Triyono J, Fatkhiyah Erfanti, Ramadhan Haidar Ikram, Fadlan Nizar Izzuddin Yatim, 2021	Perancangan Arsitektur Aplikasi Web Menggunakan Data Real Time Dari Form digital	<u>Vol. 15 No. 1 (2022): Jurnal Teknologi</u>	Kontribusi jurnal ini data harus divalidasi untuk memastikan akurasi dan kelengkapan sebelum diproses lebih lanjut, relevan dengan penerapan validasi otomatis di Form digital.
8	Fitria Tira Nur, 2023	Using Form digitals as an Online Learning Assessment Tool for Non-EFL Students	Rumah Jurnal IAIN Manado, <u>Vol 3, No 2</u> (2023), Fitria	Kontribusi jurnal ini membahas penggunaan Form digitals dengan penekanan proses validasi otomatis untuk memastikan data yang terkumpul akurat dan lengkap. Meningkatkan efisiensi penilaian dan kualitas data.

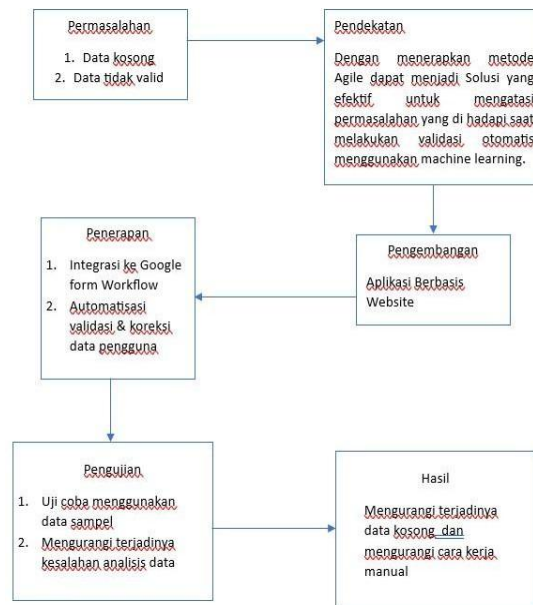
No	Author(s)	Judul	Jurnal	Kontribusi
9	Castillo Yadira Jazmín Pérez, Jiménez, Sandra Dinora Orantes, Torres, Patricio Orlando Letelier, 2024	A Systematic Literature Review on Machine Learning Applications for Agile Project Management	TEKNIK RISET DAN TEKNOLOGI, Vol. XXVI, No.1, Januari-Maret 2025,	Kontribusi jurnal ini mengulas penerapan machine learning dalam manajemen proyek Agile.
10	Galkaduwa Chethiya, Ranasinghe Nethmini, 2024	Data Science and Its Importance	<u>Biomedical Science and Clinical Research</u> 3(1)	Kontribusi jurnal ini menekankan peran Data Science dalam pengolahan dan analisis data, serta penerapan machine learning untuk validasi.

Secara keseluruhan, kontribusi dari penelitian-penelitian ini menunjukkan bahwapenerapan metode machine learning dalam pengembangan sistem berbasis data, baik untuk aplikasi web maupun mobile, dapat secara signifikan meningkatkan akurasi, efisiensi, dan efektivitas dalam pengolahan informasi serta pengambilan keputusan. Penelitian-penelitian ini juga menyoroti pentingnya integrasi machine learning dengan pendekatan pengembangan perangkat lunak modern untuk memenuhi kebutuhan pengguna dan memastikan sistem yang adaptif terhadap data yang terus berkembang.

Dengan demikian, penelitian-penelitian ini memberikan bukti yang kuat bahwapenggunaan machine learning tidak hanya meningkatkan kualitas dan kinerja sistem, tetapi juga memastikan bahwa produk akhir mampu merespons dinamika kebutuhan pengguna secara optimal. Pendekatan ini mendorong terciptanya solusi digital yang cerdas, fleksibel, dan berkualitas tinggi di era transformasi digital saat ini.

C. Kerangka Berfikir

Kerangka pemikiran ini dibuat mewakili konsep pemecahan masalah penelitian tentang penerapan Machine Learning merupakan solusi yang efektif untuk meningkatkan kualitas. Berikut gambaran flow kerangka pemikiran :



Gambar 2. 4 Kerangka Berfikir

D. Hipotesis

Telah dilakukan penelitian oleh Belgacem et al. (2022) dengan judul “**A Machine Learning Approach for Automated Filling of Categorical Fields in Data Entry Forms**” yang menunjukkan bahwa *machine learning* dapat membantu proses validasi otomatis data kosong. Berdasarkan hal tersebut, hipotesis dalam penelitian ini adalah bahwa algoritma *machine learning* K-Nearest Neighbors (KNN) yang dilatih menggunakan data historis mampu memprediksi nilai kategori dengan akurasi tinggi, sehingga dapat meminimalkan kesalahan pengisian dan mempercepat proses validasi data. Selain itu, semakin besar dan representatif data pelatihan yang digunakan, semakin baik pula kemampuan KNN dalam mengenali pola dan menangani variasi proses pengisian data kategori.