

BAB II

TINJAUAN PUSTAKA

Penelitian ini berfokus pada pengembangan sistem pencatatan dan penyortiran otomatis *data entry* berbasis *rule-based matching* di PT Cipta Sedaya Digital Indonesia (Berijalan). Oleh karena itu, diperlukan kajian literatur yang relevan terkait sistem informasi, pencatatan data otomatis, serta penerapan *rule-based matching* dalam pengembangan aplikasi. Peninjauan ini bertujuan untuk membangun landasan teoritis yang kuat dalam merancang dan mengimplementasikan sistem pencatatan dan penyortiran data yang efektif dan efisien..

Penelitian pertama oleh Navlani, Fandango, dan Idris (2021) dalam buku berjudul *Python Data Analysis* menyoroti kekuatan Python sebagai bahasa pemrograman yang unggul dalam seluruh tahapan *pipeline* pengolahan data, mulai dari pengumpulan, pembersihan (*wrangling*), pemrosesan, hingga visualisasi. Buku ini menegaskan bahwa Python, yang didukung oleh ekosistem *library* yang kaya seperti *pandas*, *NumPy*, dan kemampuan *regular expression* (*re*), memberikan fondasi yang ideal untuk membangun alat-alat pemrosesan teks dan data yang efisien dan fleksibel. Kemampuan *scripting*-nya yang cepat dan ringan sangat cocok untuk mengembangkan utilitas otomatisasi, sementara dukungan untuk komputasi paralel (misalnya dengan *Dask*) membuka peluang skalabilitas. Oleh karena itu, buku ini menjadi landasan teoritis yang komprehensif bagi pemilihan Python sebagai bahasa utama dalam penelitian ini, yang memanfaatkan kekuatan-kekuatan tersebut untuk mengimplementasikan modul *datasort* (pemrosesan teks dan *pattern matching*), *merge* (penggabungan file), *main_GUI* (pembangunan antarmuka utilitas), dan integrasi dengan *database*.

Penelitian kedua oleh Reddy et al. (2025) menginvestigasi implementasi *keylogger* sebagai instrumen untuk pemantauan aktivitas pengguna yang etis dan terkelola dalam konteks keamanan siber. Penelitian ini membuktikan bahwa pencatatan setiap penekanan tombol (*keystroke logging*) merupakan metode yang efektif untuk merekam input data pengguna secara lengkap dan akurat. Hasil penelitian menunjukkan bahwa pendekatan ini dapat secara signifikan meningkatkan transparansi dan auditabilitas proses input data. Oleh karena itu, penelitian ini menjadi dasar akademis yang kuat bagi pengembangan modul pencatatan dalam penelitian ini, yang mengadopsi konsep serupa untuk keperluan audit *data entry* yang etis dan terukur

Penelitian ketiga yang dilakukan oleh Ma, J., Liu, Y., Wan, H., dan Sun, G. (2023) memberikan tinjauan sistematis yang komprehensif mengenai teknik-teknik dalam *parsing log* dan ekstraksi fitur, yang merupakan fondasi kritis untuk mengubah *data log* tidak terstruktur menjadi format terstruktur yang siap dianalisis. Makalah ini tidak hanya mengklasifikasikan berbagai metode parsing berdasarkan prinsip kerjanya, tetapi juga menegaskan bahwa pendekatan berbasis aturan (*rule-based*) dan pencocokan pola (*pattern matching*) merupakan salah satu metode yang robust dan efektif untuk mengekstraksi serta mengklasifikasikan entitas data dari teks yang acak. Temuan dalam survei ini secara langsung relevan dan menjadi landasan teoritis yang kuat bagi pengembangan modul *datasort* dalam penelitian ini. Prinsip-prinsip yang diulas diterapkan untuk memproses data mentah hasil pencatatan *keylogger*, di mana algoritma dirancang untuk mengidentifikasi pola spesifik (seperti format NIK, nama, atau kata kunci) dan mengelompokkannya secara sistematis ke dalam kategori terstruktur seperti BOR, COBOR, dan Lainnya, sehingga memudahkan proses penyusunan data secara rapi dan konsisten.

Penelitian yang dilakukan oleh Qiaolin Qin, Roozbeh Aghili, Heng Li, dan Ettore Merlo (2025) pada konferensi SANER (IEEE) memperkenalkan sebuah *framework preprocessing* umum yang dirancang untuk meningkatkan kinerja *log parser* berbasis statistik seperti Drain. Hasil penelitian menunjukkan bahwa tahap *preprocessing* yang kerap dianggap hanya sebatas langkah sederhana seperti masking angka sebenarnya memiliki peran penting. Dengan penerapan *preprocessing* secara terstruktur, performa *parser* pada metrik seperti FTA/F1 *template accuracy* mengalami peningkatan signifikan hingga 108,9%, disertai kenaikan akurasi pengelompokan (GA, FGA) lebih dari 28%. Temuan ini relevan sebagai dasar teoretis dalam perancangan *pipeline* pengolahan data hasil *keylogger*, karena *preprocessing* menjadi tahap krusial untuk menyiapkan data agar dapat diurutkan (*datasort*) dan digabung (*merge*) secara optimal.

Penelitian kelima oleh Abreu (2020) dalam disertasinya "*Development of a Centralized Log Management System*" membahas tantangan pengelolaan data *log* yang terdistribusi dari berbagai sumber, serta menawarkan solusi arsitektural berupa pemusatan *log* ke dalam satu repositori terintegrasi. Walaupun konteks penelitian tersebut berfokus pada *log* sistem dan keamanan jaringan dalam skala enterprise, prinsip inti yang diangkat yakni urgensi konsolidasi data *log* agar lebih mudah dipantau dan dianalisis masih relevan secara konseptual. Prinsip ini

diadaptasi dalam penelitian ini melalui modul *merge*, yang berfungsi menggabungkan berbagai file teks (.txt) hasil modul *datasort* ke dalam satu file terpadu. Konsolidasi tersebut menjadi tahapan penting untuk memungkinkan semua hasil modul *datasort* menjadi tergabung dalam 1 file sehingga memudahkan efisiensi QC

Penelitian keenam merupakan buku oleh Moore, A. D. (2018) yang membahas pengembangan aplikasi *GUI* menggunakan Python dan *Tkinter*. Buku ini menjelaskan pendekatan berbasis OOP dan pola arsitektur MVC, serta menyertakan tutorial membangun aplikasi dari formulir input sederhana hingga alat manajemen data yang lebih kompleks. Selain itu, buku ini juga membahas teknik optimasi seperti pemeriksaan input, *unit testing*, integrasi dengan *database* SQL, penggunaan *multithreading* untuk menjaga responsivitas aplikasi, serta visualisasi data dan pengemasan aplikasi agar dapat dijalankan di berbagai *platform*.

Penelitian ketujuh yang ditulis oleh Xanthidis, Manolas, dan Alhousary (2022) dalam bab berjudul "*Database Programming with Python*" yang termuat dalam buku *Advances in Data Science and Artificial Intelligence*, menyajikan pembahasan mendalam mengenai teknik dan pemanfaatan library untuk menghubungkan Python dengan sistem manajemen basis data relasional. Bab ini tidak hanya mencakup langkah-langkah dasar koneksi tetapi juga menekankan praktik terbaik dalam operasi *CRUD*, manajemen transaksi, dan efisiensi koneksi. Prinsip-prinsip tersebut diadaptasi dalam pengembangan fungsi yang menghubungkan aplikasi Python dengan MySQL pada XAMPP, yang berfungsi sebagai repositori penyimpanan data terstruktur hasil proses modul *datasort*.

Melalui referensi dari penelitian-penelitian tersebut, penelitian ini diharapkan mampu merancang dan mengembangkan sistem informasi pencatatan dan penyortiran otomatis yang efektif, dengan pendekatan *rule-based matching* yang dapat mengurangi *human error* dan mempercepat proses operasional di lingkungan PT Cipta Sedaya Digital Indonesia (Berjalan).

Tabel 2.1 Ringkasan Tinjauan Pustaka

Penulis	Judul Penelitian	Tujuan Penelitian	Metode Pengembangan	Objek Penelitian	Bahasa Pemrograman	Konteks Aplikasi	Relevansi dengan Penelitian Ini
Abreu (2020)	<i>Development of a centralized log management system</i>	Merancang sistem terpusat untuk pengelolaan <i>log</i> yang lebih efisien	Perancangan sistem (disertasi)	Sistem manajemen <i>log</i>	-	Manajemen <i>log</i> terpusat	Landasan untuk modul <i>merge</i> (konsep konsolidasi data terpusat)
Ma et al. (2023)	<i>Automatic parsing and utilization of system log features in log analysis: A survey</i>	Meninjau teknik parsing <i>log</i> dan pemanfaatan fitur <i>log</i> dalam analisis	Studi survei/literatur	Fitur <i>log</i> sistem	-	Analisis <i>log</i> & pemrosesan data	Landasan untuk modul <i>datasort</i> (pemrosesan dan parsing <i>log</i>)
Moore (2018)	<i>Python GUI programming with Tkinter</i>	Membuat antarmuka grafis pengguna menggunakan <i>Tkinter</i>	Pengembangan perangkat lunak	Aplikasi <i>GUI</i>	Python (<i>Tkinter</i>)	Aplikasi <i>GUI</i> desktop	Landasan untuk pembangunan <i>main_GUI</i>
Navlani et al. (2021)	<i>Python data analysis (3rd ed.)</i>	Mengolah, membersihkan, dan menganalisis data menggunakan Python	Studi praktis / penerapan	Data analisis	Python (Pandas, NumPy)	Analisis data	Landasan pemilihan Python untuk pemrosesan data

Qin et al. (2025)	<i>Preprocessing is all you need: Boosting the performance of log parsers with a general preprocessing framework</i>	Meningkatkan akurasi parser log dengan kerangka preprocessing umum	Eksperimen / framework preprocessing	Parser log (Drain, dll.)	Python	Analisis log sistem	Landasan untuk algoritma <i>datasort</i> (pentingnya preprocessing)
Reddy et al. (2025)	<i>Keylogger and screenlogger tools for robust cybersecurity and ethical user activity monitoring</i>	Mengembangkan <i>keylogger/screenlogger</i> untuk keamanan siber & pemantauan etis	Studi konseptual / pengembangan alat	<i>Keylogger & screenlogger</i>	-	Keamanan siber & monitoring aktivitas	Justifikasi etis & teknis untuk modul pencatatan input
Xanthidis et al. (2022)	<i>Database programming with Python</i>	Mengajarkan pemrograman <i>database</i> menggunakan Python	Studi teknis (bab buku)	<i>Database relasional</i>	Python (SQL libraries)	Pemrograman basis data	Landasan untuk integrasi <i>database</i> Python-MySQL

Berdasarkan tinjauan pustaka yang dilakukan, dapat disimpulkan bahwa pemrosesan data berbasis aturan (*rule-based*) serta penerapan otomatisasi alur kerja data terbukti efektif untuk meningkatkan efisiensi dan kerapian pengelolaan data. Sejumlah penelitian sebelumnya menunjukkan bahwa pencatatan aktivitas secara otomatis (misalnya melalui *keylogger* yang dikelola secara etis), tahapan pra-pemrosesan teks, *parsing log* berbasis aturan, serta penggabungan data secara terpusat mampu mengubah data mentah yang tidak teratur menjadi informasi yang lebih terstruktur dan siap dianalisis.

Dari sisi teknis, pemanfaatan bahasa pemrograman Python dengan dukungan pustaka seperti *pandas* dan *NumPy* untuk pengolahan data, *Tkinter* untuk antarmuka grafis, serta konektor *MySQL* untuk pengelolaan basis data menjadi fondasi yang fleksibel sekaligus andal dalam pengembangan sistem. Selain itu, konsep manajemen *log* terpusat dan pra-pemrosesan yang sistematis berperan penting dalam menangani data dalam jumlah besar dan kompleks.

Berdasarkan landasan teori tersebut, penelitian ini diarahkan pada pengembangan program pencatatan dan penyortiran *data entry* berbasis *rule-based matching*. Program ini bertujuan membantu *Quality Control* (QC) dalam proses pemeriksaan manual dengan menyediakan data input karyawan yang sudah terdokumentasi dan tersusun rapi, sehingga proses verifikasi oleh QC dapat dilakukan secara lebih cepat, konsisten, dan dapat dipertanggungjawabkan.