

**ANALISIS DAN VISUALISASI METADATA
DOKUMEN PDF DAN REPOSITORY GIT UNTUK
PENDEKATAN PENILAIAN KUANTITATIF DALAM
PEMBELAJARAN**

PROPOSAL SKRIPSI

Oleh:

OKTAVIAN VANDI TRI SHAKTI

NIM. 2041720208



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG**

2023

HALAMAN PENGESAHAN
ANALISIS DAN VISUALISASI METADATA
DOKUMEN PDF DAN REPOSITORY GIT UNTUK
PENDEKATAN PENILAIAN KUANTITATIF
DALAM PEMBELAJARAN

Disusun oleh:

OKTAVIAN VANDI TRI SHAKTI. NIM. 2041720208

Proposal Skripsi ini telah diuji pada tanggal 6 Desember 2023

Disetujui oleh:

- | | | | | |
|----|---------------------|---|------------------------------------------------------------------|-------|
| 1. | Pembimbing
Utama | : | <u>Habibie Ed Dien, S.Kom., M.T.</u>
NIP. 199204122019031013 | |
| 2. | Pembahas I | : | <u>Yoppy Yunhasnawa, S.ST., M.Sc.</u>
NIP. 198906212019031000 | |
| 3. | Pembahas II | : | <u>Noprianto, S.Kom., M.Eng</u>
NIP. 198911082019031020 | |

Mengetahui,

Ketua Jurusan
Teknologi Informasi

Ketua Program Studi
D4 Teknik Informatika

Rosa Andrie Asmara, ST., MT., Dr. Eng
NIP. 19801010 200501 1 001

Dr. Ely Setyo Astuti, ST., MT.
NIP. 197605152009122001

DAFTAR ISI

DAFTAR TABEL	6
1.1. Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
2.1 Studi Literatur.....	4
2.1 Dasar Teori	6
2.1.1 Metadata.....	6
2.1.2 Dokumen PDF.....	7
2.1.3 Git.....	7
2.1.4 Repository	7
2.1.5 Metode Parsing	8
2.1.6 Regular Expression	8
2.1.7 Crawling.....	8
BAB III. METODOLOGI PENELITIAN	10
3.1. Waktu dan Tempat Penelitian	10
3.2. Teknik Pengumpulan Data	10
3.3 Teknik Pengolahan Data	16
3.3.1 Teknik Pengolahan Data Dokumen PDF	16
3.3.2 Teknik Pengolahan Data Repository Git	16
3.4 Desain Sistem	17
3.4.1 Wireframe	17

3.4.2	Use Case.....	22
3.4.3	Arsitektur Sistem.....	23
3.5	Uji Coba Sistem.....	24

DAFTAR GAMBAR

Gambar 1 Dashboard Metadata PDF Wireframe	18
Gambar 2 Dashboard Metadata Git Wireframe	18
Gambar 3 Metadata Git Wireframe.....	19
Gambar 4 Metadata PDF Wireframe	20
Gambar 5 Wireframe Data Master	21
Gambar 6 Wireframe Visualisasi Data	22
Gambar 7 Use Case.....	23
Gambar 8 Arsitektur Sistem.....	24

DAFTAR TABEL

Tabel 1 Uji Coba Sistem	25
Tabel 2 Jadwal Penelitian.....	26

BAB I. PENDAHULUAN

1.1. Latar Belakang

Dalam era perkembangan teknologi dan transformasi digital, pendidikan memainkan peran sentral dalam pembangunan sumber daya manusia yang berkualitas. Perubahan ini telah membawa dampak signifikan pada sektor pendidikan, yang semakin mengadopsi teknologi informasi dan komunikasi sebagai alat utama dalam memberikan materi pembelajaran, akses ke informasi, dan kolaborasi antara dosen dan mahasiswa. Dokumen PDF sebagai format dokumen elektronik yang umum digunakan, telah menjadi salah satu alat utama dalam penyediaan sumber daya pendidikan.

Sementara itu, penggunaan Git, sistem kontrol versi distribusi, telah berkembang pesat dalam konteks pengembangan perangkat lunak. Git memungkinkan kolaborasi, pencatatan manajemen kode sumber yang efisien. Namun, penggunaan Git dalam pengelolaan sumber daya pendidikan belum sepenuhnya dijelajahi.

Dalam konteks ini, penelitian ini bertujuan untuk menganalisis dan visualisasi metadata dokumen PDF dan repository Git sebagai pendekatan penilaian kuantitatif dalam pembelajaran. Metadata, sebagai informasi yang mendeskripsikan, menjelaskan, atau memfasilitasi penggunaan data lainnya, memiliki potensi untuk merefleksikan aktivitas, kinerja, dan kemampuan mahasiswa dalam pembelajaran.

Penelitian ini menggunakan pendekatan kuantitatif dengan mengumpulkan data metadata dari dokumen PDF dan repository Git yang dihasilkan oleh mahasiswa dalam mata kuliah tertentu. Data metadata tersebut akan dianalisis menggunakan teknik-teknik statistik, data mining, dan machine learning untuk mendapatkan informasi-informasi yang relevan dengan pembelajaran. Informasi tersebut mencakup frekuensi, durasi, kualitas, kesalahan, kerjasama, dan prestasi mahasiswa dalam pembelajaran.

Penelitian ini terletak pada perpotongan beberapa bidang penelitian yang relevan, seperti penggunaan data mining dan analisis repository dalam meningkatkan proses pembangunan perangkat lunak dan penggunaan teknik analisis dokumen PDF. Selain itu, latar belakang literatur yang kuat dalam *learning analytics* menunjukkan pentingnya penggunaan data dan analisis dalam pemahaman dan perbaikan proses dan hasil pembelajaran.

Pada penelitian ini, metadata dan repository memiliki hubungan erat yang menciptakan dasar analisis untuk pendekatan penilaian kuantitatif terhadap proses dan hasil pembelajaran mahasiswa dalam konteks pengembangan perangkat lunak. Dalam

hal ini, merujuk pada informasi tambahan yang terkandung dalam suatu dokumen atau proyek, sedangkan repository adalah tempat penyimpanan digital yang menyimpan riwayat perubahan atau versi suatu proyek.

Penelitian ini memiliki potensi untuk memberikan kontribusi signifikan bagi pengembangan metode pendekatan penilaian kuantitatif pembelajaran yang lebih objektif, akurat, dan efisien. Dengan menganalisis metadata dokumen PDF dan repository Git, penelitian ini berusaha memahami dampak perubahan teknologi terhadap pendidikan, dengan harapan dapat meningkatkan kualitas pembelajaran dan mendukung pengambilan keputusan yang akurat.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil adalah sebagai berikut:

1. Bagaimana metode analisis metadata dokumen PDF dan repository Git dapat dikembangkan untuk mengukur kualitas, relevansi, dan efektivitas dokumen serta sumber daya pengembangan perangkat lunak dalam konteks pembelajaran mahasiswa dengan pendekatan kuantitatif?
2. Bagaimana penerapan metode analisis metadata repository Git dapat memberikan wawasan tentang kerjasama, kontribusi, dan sejarah perubahan sumber daya dalam konteks tugas individu dan kolaboratif antar kelompok mahasiswa dengan pendekatan kuantitatif?
3. Apa dampak penggunaan metode analisis metadata dokumen PDF dan repository Git terhadap pemahaman proses dan hasil pembelajaran mahasiswa dalam pengembangan perangkat lunak secara kuantitatif?

1.3 Batasan Masalah

Adapun batas permasalahan dalam penelitian ini adalah:

1. Penelitian ini terbatas pada analisis metadata dokumen PDF.
2. Penelitian ini membatasi penggunaan sistem kontrol versi Git dalam konteks tugas individu dan kelompok.
3. Penelitian ini menggunakan framework Flask dalam pengembangan aplikasi web.

1.4 Tujuan Penelitian

Tujuan dari dilakukannya skripsi dengan judul “**ANALISIS METADATA DAN VISUALISASI DOKUMEN PDF DAN REPOSITORY GIT UNTUK PENDEKATAN PENILAIAN KUANTITATIF DALAM PEMBELAJARAN**” adalah sebagai berikut:

1. Penelitian ini bertujuan untuk menyelidiki metode analisis metadata dokumen PDF dan metadata repository Git yang dapat diterapkan untuk memahami secara kuantitatif proses dan hasil pembelajaran mahasiswa.
2. Metode analisis metadata dokumen PDF akan difokuskan pada pengukuran kualitas, relevansi, dan efektivitas dokumen sebagai bagian dari proses pembelajaran.
3. Metode analisis metadata repository Git akan digunakan untuk memahami kerjasama dan kontribusi pada perubahan sumber daya pembelajaran dalam konteks pengembangan perangkat lunak.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan metode penilaian pembelajaran yang lebih efektif dan efisien dengan menggunakan pendekatan kuantitatif. Manfaat penelitian ini juga diharapkan akan melibatkan pemahaman yang lebih mendalam tentang pemanfaatan metadata dokumen PDF dan repository Git dalam proses dan hasil pembelajaran. Pendekatan kuantitatif akan memperkuat analisis data dan memberikan dasar yang lebih kokoh untuk menghasilkan temuan yang dapat diukur dan dianalisis secara statistik.

BAB II. LANDASAN TEORI

2.1 Studi Literatur

Dalam penelitian ini penulis memaparkan lima penelitian terdahulu yang relevan dengan permasalahan yang akan diteliti tentang “Analisis Metadata Dokumen PDF dan Repository Git Untuk Pembelajaran”.

1. (Zhu & Cole, 2022) dalam jurnalnya yang berjudul *PDFDataExtractor: A Tool for Reading Scientific Text and Interpreting Metadata from the Typeset Literature in the Portable Document Format*, dalam penelitiannya telah menunjukkan bahwa penelitian ini fokus pada pengembangan alat ekstraksi teks dari dokumen PDF dalam domain ilmiah, terutama dalam bidang kimia. Beberapa penelitian telah menghasilkan alat ekstraksi teks yang menggunakan pendekatan berbasis template dan NLP untuk mengekstrak metadata seperti judul artikel, penulis, abstrak, kata kunci, jurnal, tahun, DOI, dan referensi. Namun, hasil penelitian sebelumnya menunjukkan bahwa alat-alat tersebut memiliki keterbatasan dalam menangani rumus, tabel, dan informasi penulis. Penelitian ini akan memperkenalkan *PDFDataExtractor*, sebuah alat ekstraksi teks yang dikembangkan dengan fokus pada presisi dalam mengekstrak informasi terkait kimia dari artikel ilmiah dalam format PDF. *PDFDataExtractor* juga dapat diintegrasikan dengan paket perangkat lunak penambahan teks *ChemDataExtractor*.
2. (Nurandini et al., n.d.) dalam jurnalnya yang berjudul *Clustering Dokumen dengan Menambahkan Metadata Menggunakan Algoritma COATES*, dalam penelitiannya algoritma COATES untuk mengklasterkan dokumen dengan memanfaatkan metadata sebagai "*side information*". Penelitian ini menemukan bahwa penambahan metadata dapat mempengaruhi hasil klaster, terutama dalam perpindahan dokumen antar klaster. Selain itu, inisialisasi *cluster* juga mempengaruhi proses pengelompokan data. Namun, penelitian sebelumnya belum secara khusus menggunakan COATES algorithm dengan inisialisasi k-means berdasarkan perhitungan jarak Euclidean pada dataset dokumen sebanyak 50 dari berbagai kategori. Oleh karena itu, penelitian ini akan memperluas pemahaman tentang penggunaan COATES algorithm dalam mengklasterkan dokumen dengan memanfaatkan metadata, khususnya dengan pendekatan inisialisasi k-means dan perhitungan jarak Euclidean.

3. (Alamoudi et al., 2021) dalam jurnalnya yang berjudul *A rule-based information extraction approach for extracting metadata from pdf books*, dalam penelitiannya melakukan penerapan teknik dan pendekatan yang berbeda dalam ekstraksi informasi dari dokumen digital, seperti PDF. Beberapa penelitian telah menggunakan teknik *machine learning* dan *data science pipeline* untuk ekstraksi informasi terstruktur dari berbagai dokumen, sementara yang lain mengusulkan sistem *a rule based information extraction* untuk dokumen PDF. Hasil ini menunjukkan bahwa pendekatan *rule based* dapat mencapai tingkat akurasi yang tinggi dalam ekstraksi metadata dari dokumen PDF. Pada penelitian ini, difokuskan pada pengembangan pendekatan *rule based* untuk mengekstrak metadata dari buku PDF dengan tujuan meningkatkan akurasi dan efisiensi. Penelitian ini akan memperhatikan tantangan khusus dalam mengekstrak nama penulis dan data lainnya dari buku PDF, serta akan mencari cara untuk meningkatkan kinerja ekstraksi metadata tersebut.
4. (Scott et al., n.d.) dalam jurnalnya yang berjudul *Metagit: Decentralised Metadata Management With Git*, dalam penelitiannya dilakukan penelitian mengenai arsitektur penyimpanan metadata dalam sistem kontrol Git. Penelitian tersebut membahas penggunaan fitur-fitur Git seperti *blobs*, *trees*, *references*, dan *commits* untuk menyimpan metadata dan memperkenalkan alat manajemen metadata berbasis *command line*. Selain itu, penelitian ini juga menyoroti konflik metadata, pencarian file berdasarkan metadata, dan pendekatan penyimpanan metadata menggunakan jalur objek atau ID commit data. Namun, penelitian yang akan ini akan memfokuskan pada pengembangan model metadata yang dapat melacak perubahan metadata, serta mengeksplorasi integrasi model tersebut ke dalam sistem lain seperti *Mercurial* dan *Subversion*.
5. (Kim et al., 2021) dalam jurnalnya yang berjudul *Githru: Visual analytics for understanding software development history through git metadata analysis*, dalam penelitiannya telah dilakukan berbagai studi terkait visualisasi dan analisis metadata Git. Githru, sebuah sistem visual analytics untuk metadata Git, dikembangkan untuk mengatasi kompleksitas dan skalabilitas dari kumpulan metadata Git yang besar. Sistem ini memberikan desain visualisasi baru untuk eksplorasi interaktif dan telah terbukti efektif dalam memahami sejarah pengembangan perangkat lunak. Penelitian ini juga membahas teknik rekonstruksi data dan visualisasi yang digunakan dalam Githru, seperti

pengelompokan, visualisasi grafik batang, dan kontrol cakupan analisis. Namun, penelitian ini akan memperluas analisis metadata dan mengatasi isu-isu skalabilitas yang ada.

6. (Atkinson et al., 2014) dalam jurnalnya yang berjudul *Web metadata extraction and semantic indexing for learning objects extraction*, dalam penelitiannya menyajikan pendekatan multi-strategi untuk ekstraksi metadata dan pengindeksan semantik objek pembelajaran dari web. Ini menggabungkan teknik seperti *Latent Semantic Analysis* (LSA), analisis konsep berbasis *Lattice*, ekstraksi informasi, kategorisasi teks, dan pembelajaran mesin untuk meningkatkan akses dan penggunaan materi digital pendidikan bagi guru. Pendekatan ini dievaluasi oleh guru sekolah menengah dan menunjukkan harapan dalam menyediakan metadata berkualitas tinggi untuk dokumen yang diambil dari web. Efektivitas pendekatan dalam ekstraksi metadata pendidikan dan pengindeksan semantik dievaluasi melalui eksperimen dan hasil, menunjukkan kinerja yang baik dalam akurasi klasifikasi dan kebenaran ekstraksi.

2.1 Dasar Teori

2.1.1 Metadata

Metadata adalah data yang digunakan untuk mendeskripsikan data lain, misalnya pemilik dan tanggal pembuatan file. Ini berguna untuk mengkategorikan data untuk membantu pengorganisasian dan pengelolaannya. Metadata terkadang disimpan di dalam file data untuk memberikan informasi yang relevan tentang datanya metadata tersemat. Misalnya, dokumen Microsoft Word menyimpan nama penulis, perusahaan, dan tanggal pembuatan/modifikasi dokumen. File Portable Document Format (PDF) menggunakan dua jenis metadata: kamus key-value dan aliran metadata yang sesuai dengan XMP yang disimpan sebagai XML. Tidak semua format file mendukung metadata, dan modifikasi metadata pada format yang mendukung metadata mengubah file data itu sendiri sehingga sulit untuk mengidentifikasi apakah itu data atau metadata yang telah berubah (Alghamdi et al., 2022).

Metadata digunakan untuk mencatat produk data yang dibuat dan menjawab pertanyaan dasar tentang siapa, apa, kapan, di mana, dan untuk tujuan apa data tersebut dibuat atau disiapkan. Metadata dihasilkan berdasarkan tipe data dan konteks

penggunaan. Sebagai contoh sederhana, dalam konteks perpustakaan yang berisi banyak koleksi buku berbeda, pustakawan dapat membuat metadata tentang judul buku. Metadata ini berisi uraian isi buku, nomor halaman, pengarang, tahun terbit, nomor ISSN, dan lokasi fisik buku tersebut berada. Contoh lain dimana metadata dapat dibuat adalah dalam konteks membuat peta, lembaga penerbit, metode memperoleh peta, dan harga peta (Dwiyani, 2017)

2.1.2 Dokumen PDF

Dokumen PDF adalah dokumen digital yang dapat dibaca dan dicetak tanpa kehilangan format aslinya. Dokumen PDF dapat berisi teks, gambar, dan grafik, dan dapat dilindungi dengan kata sandi untuk mencegah penyuntingan atau duplikasi ilegal. Dokumen PDF sering digunakan untuk berbagi dokumen secara elektronik, karena formatnya yang universal dan mudah dibaca di berbagai perangkat. Dokumen PDF juga dapat digunakan untuk membuat buku elektronik, dokumen akademik, dan dokumen hukum (Alghamdi et al., 2022)

Dokumen PDF memiliki beberapa karakteristik berikut:

- Dapat dibaca dan dicetak tanpa kehilangan format aslinya
- Dapat berisi teks, gambar, dan grafik
- Universal dan mudah dibaca di berbagai perangkat

2.1.3 Git

Git adalah sebuah sistem kontrol yang dapat digunakan untuk melacak perubahan pada metadata. Git menyediakan kemampuan untuk melacak perubahan dan memiliki kemampuan berbagi yang kuat, memungkinkan perubahan pada metadata untuk dipertukarkan dengan repositori pusat dan pengguna lain. *Branches* pada git memungkinkan untuk memisahkan data dan metadata sehingga komit metadata dapat dibuat secara independen dari komit data (Alghamdi et al., 2022)

2.1.4 Repository

Repository adalah suatu tempat penyimpanan yang dapat menampung berbagai aplikasi atau berkas dan dapat diakses melalui internet. Penyimpanan dalam repository memiliki beragam jenis, termasuk pengumpulan tugas mata kuliah tertentu yang diperlukan untuk keperluan tugas mata kuliah (Ambriani & Nurhidayat, n.d.)

2.1.5 Metode Parsing

Metode parsing merujuk pada teknik atau pendekatan yang digunakan untuk mengurai atau mengekstraksi informasi tertentu dari suatu teks. Dalam konteks parsing berdasarkan kutipan jurnal, metode ini dirancang khusus untuk mengidentifikasi dan mengekstraksi metadata penting, seperti nama penulis, tanggal, judul, atau elemen-elemen lainnya dari kutipan jurnal atau dokumen serupa. Metode parsing dapat melibatkan penggunaan ekspresi reguler, pengetahuan domain, atau teknik pembelajaran mesin seperti CRF, SVM, atau HMM untuk mengotomatiskan proses ekstraksi informasi dari teks yang kompleks. (Tkaczyk et al., 2015)

2.1.6 Regular Expression

Regular Expression (RegEx) adalah formula yang digunakan untuk mencari dan memanipulasi pola dalam kalimat atau string. RegEx sangat kuat dan dapat digunakan untuk mencari fragmen kata serta mengontrol manajemen data. Banyak bahasa pemrograman mendukung regular expressions seperti PHP, Perl, VB, Java, dan Python. Dalam konteks jurnal tersebut, RegEx digunakan untuk mengidentifikasi kesalahan penulisan dalam referensi bibliografi dan memperbaikinya sesuai dengan aturan, dengan tujuan meningkatkan kualitas metadata (Indrawat ariani et al., 2020)

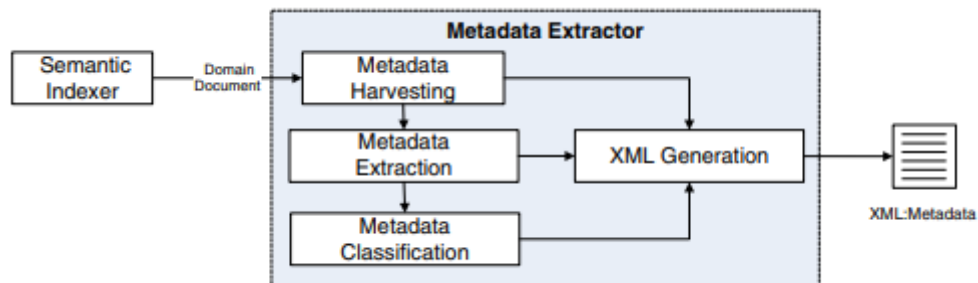
2.1.7 Crawling

Dalam Jurnal yang dikutip dari (Barbaresi, n.d.).Crawling dalam konteks jurnal tersebut merujuk pada proses pengumpulan data dari internet, khususnya dalam hal ini adalah pengumpulan teks dari blog dan situs web. Proses ini melibatkan penggunaan program komputer untuk mengakses dan menyalin informasi dari berbagai sumber online, seperti blog dan situs web, untuk tujuan analisis dan penelitian. Metode crawling ini dapat digunakan untuk mengumpulkan teks dari berbagai sumber online yang kemudian dapat dianalisis untuk memahami tren, topik, atau pola tertentu yang terkait dengan blog dan situs web tersebut.

2.1.8 Ekstraksi Metadata

Ekstraksi Metadata melibatkan proses mengenali informasi *key implicit* dan *explicit information* yang ada dalam dokumen yang diekstraksi untuk mengidentifikasi dan mengindeks *learning objects* (LOs) dari berbagai sumber seperti halaman web, file,

teks, file audio, dan grafik. Proses ekstraksi melibatkan metode seperti *tokenizing string*, *counting token occurrences*, normalisasi dan pembobotan frekuensi token, dan penggunaan teknik pengumpulan, ekstraksi, dan klasifikasi informasi untuk mengisi properti metadata. (Atkinson et al., 2014)



sumber (Atkinson et al., 2014)

BAB III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

1. Waktu Penelitian

Waktu yang digunakan peneliti untuk melakukan penelitian ini adalah dalam jangka waktu 3 bulan antara bulan Desember 2023 sampai Februari 2023 ijin penelitian dalam kurun waktu kurang lebih 2 (dua) bulan, 2 bulan pengumpulan data dan 1 bulan pengolahan data yang meliputi penyajian dalam bentuk skripsi dan proses bimbingan berlangsung.

2. Tempat Penelitian

Tempat pelaksanaan penelitian ini adalah di Lingkungan Politeknik Negeri Malang di Jurusan Teknologi Informasi pada laboratorium Jurusan Teknologi Informasi

3.2. Teknik Pengumpulan Data

1. Pengumpulan data Metadata Dokumen PDF diperoleh dari beberapa hasil tugas prodi Sistem Informasi Bisnis dan prodi Teknik Informatika yang diperoleh pada Google Drive pengumpulan tugas. Data metadata dokumen PDF yang akan dikumpulkan meliputi:

No	Data Metadata PDF	Penjelasan	Contoh Metadata	Justifikasi
a	Nama dokumen	Nama dokumen menjadi identifikasi untuk setiap tugas, membantu dalam pemisahan dan pelacakan dokumen yang diunggah oleh mahasiswa.	File name : "Tugas 2.pdf"	Nama dokumen menjadi kunci utama untuk mengidentifikasi dan membedakan tugas yang diunggah oleh mahasiswa.
b	Tanggal pembuatan dokumen	Tanggal pembuatan dokumen memberikan gambaran waktu saat tugas dibuat, memungkinkan mengetahui durasi dan pola pengerjaan tugas oleh mahasiswa	Created : 17/04/2022 18:00:01	Informasi ini berguna untuk mengetahui kualitas tugas dan pola waktu pengerjaan mahasiswa, memberikan perspektif waktu pada proses

				pembuatan dokumen.
c	Tanggal terakhir diubah berdasarkan batas waktu pengumpulan tugas	Mengetahui tanggal terakhir dokumen diubah menjadi penting untuk menilai ketepatan waktu pengumpulan tugas dan pola revisi yang dilakukan oleh mahasiswa.	Modified 17/04/2022 19:05:20	Informasi ini memberikan wawasan tentang revisi yang dilakukan oleh mahasiswa dan membantu menilai ketepatan waktu pengumpulan tugas.
d	Jumlah halaman dokumen	Jumlah halaman dokumen memberikan gambaran tentang kompleksitas tugas. Metadata ini dapat membantu memahami sejauh mana mahasiswa dapat menyajikan informasi dengan detail dan kedalaman.	Numbers of Page : 20	Jumlah halaman dapat memberikan indikasi tingkat kedalaman analisis atau eksplorasi topik dalam tugas mahasiswa.
e	Jumlah kata dokumen	Jumlah kata adalah parameter penting yang dapat memberikan informasi tentang kedalaman analisis atau eksplorasi topik yang dilakukan oleh mahasiswa dalam tugasnya.	Word count : 100	Jumlah kata membantu dalam mengevaluasi sejauh mana mahasiswa menjelajahi dan menganalisis topik yang diberikan dalam tugas yang sudah diberikan.
f	Judul Dokumen	Judul dokumen memberikan gambaran tentang fokus atau topik tugas, membantu dalam mengidentifikasi konten dan tujuan dari setiap dokumen yang diajukan.	Title : "Laporan Praktikum"	Judul dokumen menjadi petunjuk awal untuk memahami topik atau tujuan dari tugas yang diajukan oleh mahasiswa.

g	Author Dokumen	Informasi penulis memberikan identitas unik untuk mahasiswa yang menyerahkan tugas. Pemantauan kemiripan nama author dapat membantu mendeteksi potensi kecurangan akademik, seperti salinan atau pekerjaan yang diserahkan oleh orang lain.	Author : vian	Identifikasi penulis penting untuk memonitor kemiripan nama author dan mendeteksi potensi kecurangan akademik, seperti pengumpulan tugas yang identik oleh mahasiswa lain.
---	----------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Data ini akan diolah melalui analisis metadata untuk mendapatkan informasi yang dapat memberikan pemahaman lebih lanjut tentang pola pengerjaan, distribusi waktu, dan karakteristik unik dari setiap tugas. Data metadata dokumen PDF akan dikumpulkan menggunakan Python. Python menyediakan library untuk mengakses metadata dokumen PDF, seperti PyPDF2.

PyPDF2 adalah *library* Python yang memungkinkan untuk bekerja dengan file PDF. Dalam konteks ini, PyPDF2 digunakan untuk mengekstrak metadata seperti nama dokumen, tanggal pembuatan, tanggal terakhir diubah, jumlah halaman, jumlah kata, judul dokumen, dan author dokumen dari setiap file PDF. (PyPDF2, n.d.)

Berikut adalah contoh penggunaan *library* Python PyPDF2 untuk membaca metadata dari dokumen PDF:

```
from PyPDF2 import PdfReader

reader = PdfReader("example.pdf")

meta = reader.metadata

print(len(reader.pages))

# All of the following could be None!
print(meta.author)
print(meta.creator)
print(meta.producer)
print(meta.subject)
```

```
print(meta.title)
```

sumber (PyPDF2, n.d.)

2. Pengumpulan data Metadata Repository Git diperoleh dari hasil tugas mata kuliah kelas 3 prodi Teknik Informatika yang diperoleh pada link API Github.

Data metadata repository Git yang akan dikumpulkan meliputi:

No	Data Metadata Git	Penjelasan	Contoh Metadata	Justifikasi
a	Nama repository	Nama repository menjadi identifikasi untuk proyek pengembangan perangkat lunak yang dihasilkan oleh mahasiswa. Ini membantu dalam melacak dan membedakan antara berbagai proyek yang ada.	Repository Name: "Project_Repository"	Nama repository menjadi kunci utama untuk mengidentifikasi dan membedakan antara berbagai proyek pengembangan perangkat lunak yang dilakukan oleh mahasiswa.
b	Nama Kontributor dan Jumlah Commit	Informasi kontributor dan commit memberikan wawasan tentang siapa saja yang terlibat dalam pengembangan proyek dan kolaborasi kontribusi kelompok.	Vian Vandi (5 commits)	Mengetahui nama kontributor dan jumlah commit memberikan informasi tentang kontribusi individu dan level keterlibatan dalam proyek, serta memberikan indikasi kolaborasi dalam tim.
c	Tanggal dibuatnya repository	Tanggal dibuatnya repository memberikan informasi waktu awal pembuatan proyek,	Repository Creation Date: January 10, 2023	Tanggal pembuatan repository penting untuk mengetahui sejak kapan proyek dimulai, memberikan konteks waktu yang penting untuk melihat progres dan

		memberikan informasi dengan konteks waktu untuk v progres dan perkembangan tugas.		perkembangan proyek.
d	Tanggal terakhir diubah repository	Mengetahui tanggal terakhir perubahan pada repository membantu dalam menilai kebaruan dan keaktifan proyek.	Last Modified Date: January 25, 2023	Tanggal terakhir perubahan membantu dalam menilai keaktifan proyek, memberikan informasi apakah proyek masih dalam pengembangan atau sudah stabil.
e	Jumlah commit di repository dengan dibandingkan tanggal dan waktu batas pengumpulan	Membandingkan jumlah commit dengan batas pengumpulan tugas dapat memberikan gambaran tentang keterlibatan dan intensitas kerja pada proyek, serta menilai kepatuhan terhadap tenggat waktu tugas.	Number of Commits by Deadline: 13	Membandingkan jumlah commit dengan batas pengumpulan tugas memberikan gambaran tentang tingkat keterlibatan dan kepatuhan terhadap tenggat waktu, indikasi kerja keras dan komitmen tim.
f	Pesan commit	Menganalisis pesan commit dengan hashtag tertentu dapat membantu dalam mengidentifikasi tujuan, atau kategori tertentu dari setiap kontribusi.	Commit Messages : "#bug_fix" (2 commits)	Analisis pesan commit membantu dalam memahami tujuan atau kategori kontribusi, memberikan konteks tambahan tentang perubahan yang dilakukan dalam proyek.

g	Tanggal dan Waktu masing-masing commit	Informasi ini memberikan kerangka waktu yang lebih rinci untuk setiap commit, memungkinkan untuk mengetahui pola kerja dan distribusi waktu kontributor.	Date and Time of Each Commit: <ul style="list-style-type: none"> Commit 1: January 12, 2023, 14:30 Commit 2: January 15, 2023, 10:45 	Informasi tanggal dan waktu commit membantu dalam melihat pola kerja kontributor, distribusi waktu pekerjaan, dan memberikan wawasan tentang intensitas kerja pada berbagai tahap proyek.
---	----------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Untuk mengimplementasikan pengumpulan metadata dari repository Git, penelitian ini akan menggunakan skrip Python yang memanfaatkan library PyGithub. PyGithub adalah sebuah library Python yang dirancang khusus untuk berinteraksi dengan API GitHub, memudahkan akses dan pengambilan data dari repository. (PyGithub, n.d.)

Berikut adalah contoh skrip Python sederhana yang menggunakan PyGithub untuk mengekstrak metadata dari repository Git:

a. Membuat Github *Instance*

```
from github import Github

# Authentication is defined via github.Auth
from github import Auth

# using an access token
auth = Auth.Token("access_token")

# Public Web Github
g = Github(auth=auth)

# Github Enterprise with custom hostname
g = Github(auth=auth, base_url="https://{hostname}/api/v3")
```

sumber (PyGithub, n.d.)

b. Membuat Github *Object*

```
for repo in g.get_user().get_repos():
    print(repo.name)
    repo.edit(has_wiki=False)
# to see all the available attributes and methods
print(dir(repo))
```

sumber (PyGithub, n.d.)

3.4 Teknik Pengolahan Data

Pada tahapan ini, data yang telah dikumpulkan dari metadata dokumen PDF dan repository Git akan mengalami serangkaian proses pengolahan data untuk mendapatkan informasi yang relevan untuk melakukan pendekatan kuantitatif dalam pembelajaran. Berikut adalah teknik pengolahan data yang akan diterapkan:

3.3.1 Teknik Pengolahan Data Dokumen PDF

Penelitian ini bertujuan untuk menyempurnakan teknik pengolahan data dokumen PDF dengan fokus pada ekstraksi metadata yang mencakup tanggal pembuatan dokumen, tanggal terakhir diubah, judul dokumen. Teknik yang digunakan untuk ekstraksi metadata tersebut adalah metode *parsing*, metode *parsing* digunakan sebagai pendekatan utama untuk mengurai dokumen PDF menjadi komponen-komponennya, termasuk metadata yang relevan.

Proses ekstraksi metadata terkait topik dokumen dilakukan menggunakan metode *regular expression* (Indrawat ariani et al., 2020). Dalam hal ini, *regular expression* digunakan untuk mengidentifikasi pola-pola khusus dalam string, seperti pola yang biasanya digunakan untuk menandai judul bab atau sub bab dalam dokumen. Hal ini bertujuan untuk mendapatkan informasi yang lebih rinci mengenai topik atau struktur dokumen.

Selain itu, penelitian ini mengimplementasikan metode *machine learning* untuk mengekstrak metadata terkait isi dokumen. Metode *machine learning* memungkinkan sistem untuk memahami dan mempelajari pola-pola kompleks dalam data teks, termasuk identifikasi kata kunci atau frasa penting yang dapat memberikan gambaran mendalam tentang isi dari dokumen.

3.3.2 Teknik Pengolahan Data Repository Git

Penelitian ini mengusulkan pengembangan metode *parsing* yang komprehensif untuk mengekstrak informasi dari Git. Pendekatan utama yang peneliti pilih adalah metode *parsing* yang dapat diterapkan untuk memecah *repository* Git menjadi komponen-komponennya yang mencakup ekstraksi daftar file dan folder.

Untuk mendapatkan informasi kode sumber dari *repository*, peneliti melakukan proses *parsing* pada file-file yang ada. Dengan demikian, penelitian ini tidak hanya memfokuskan pada identifikasi file, tetapi juga memungkinkan ekstraksi kode sumber untuk analisis lebih lanjut terkait evolusi dan perubahan kode dalam *repository* tersebut.

Proses ekstraksi *history commit* melibatkan penggunaan metode *parsing* untuk memecah *history commit* menjadi komponen. Informasi yang diambil mencakup tanggal dan waktu commit, nama committer, serta rincian perubahan yang dilakukan pada setiap commit.

Mengimplementasikan metode parsing dapat mengekstrak informasi terkait *issues* dari *repository* Git. Pendekatan ini memungkinkan pemecahan *issues* menjadi komponen-komponennya, termasuk judul *issue*, deskripsi *issue*, dan status *issue*, sehingga dapat memberikan pemahaman yang lebih mendalam terkait dengan permasalahan yang diidentifikasi.

Dalam konteks ekstraksi *pull requests*, peneliti mengaplikasikan metode *parsing* untuk memecah *pull requests* menjadi komponen-komponennya. Ini mencakup judul *pull request*, deskripsi *pull request*, dan status *pull request*, sehingga informasi terkait penggabungan kode dapat diakses dengan lebih sistematis.

Dengan menyeluruh mengintegrasikan metode parsing pada berbagai aspek *repository* Git, penelitian ini bertujuan untuk menyediakan landasan analisis yang lebih lengkap dan mendalam.

3.5 Desain Sistem

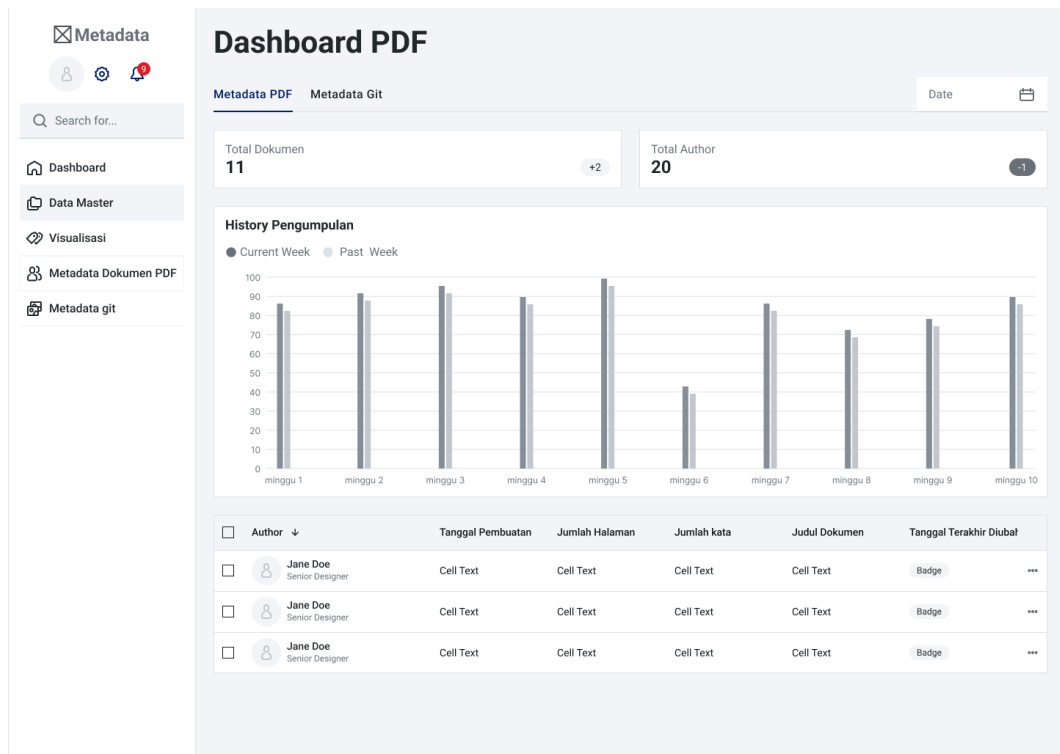
3.4.1 Wireframe

Dashboard merupakan antarmuka visual utama yang memberikan ringkasan hasil analisis metadata dokumen PDF dan repository Git. Dalam hal ini, elemen-elemen yang dapat ditampilkan di dashboard meliputi:

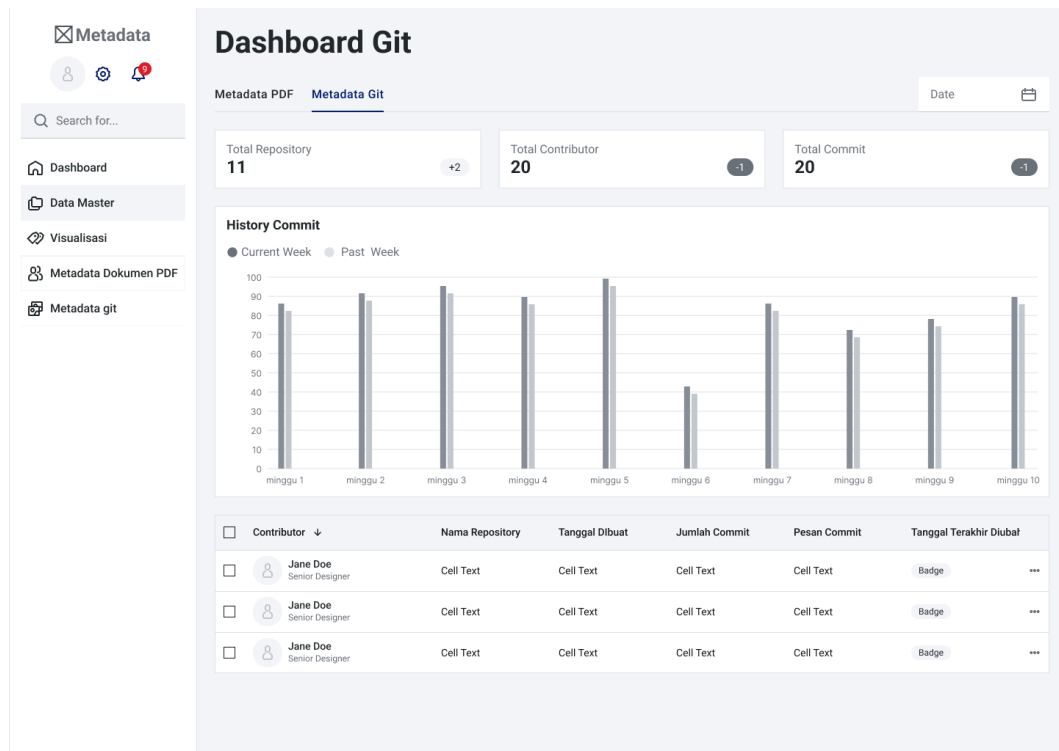
a. Dashboard

Grafik atau diagram yang menyajikan ringkasan keseluruhan analisis, seperti total dokumen PDF, jumlah commit, dan statistik lainnya. Informasi ini

memberikan gambaran singkat tentang volume dan distribusi data.



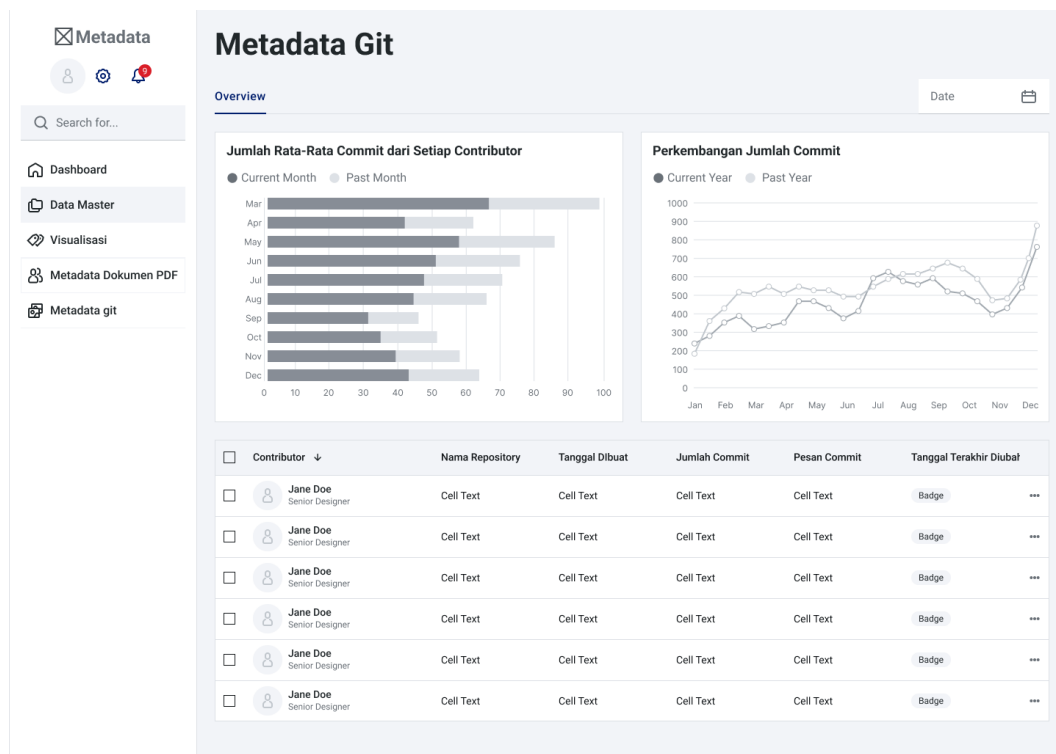
Gambar 1 Dashboard Metadata PDF Wireframe



Gambar 2 Dashboard Metadata Git Wireframe

b. Metadata Git

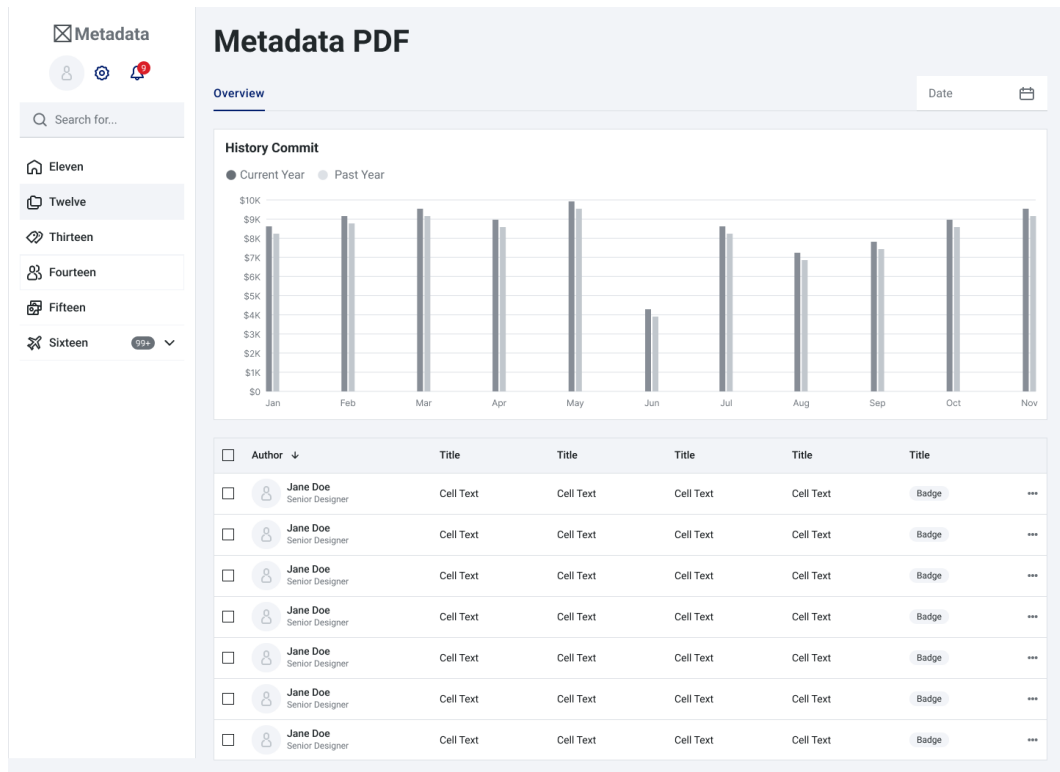
Grafik garis atau batang yang mencerminkan history commit pada repository Git. Periode waktu dapat disesuaikan, dan grafik dapat menyoroti titik-titik penting, seperti kolaborasi atau periode intensif pekerjaan.



Gambar 3 Metadata Git Wireframe

c. Metadata Dokumen PDF:

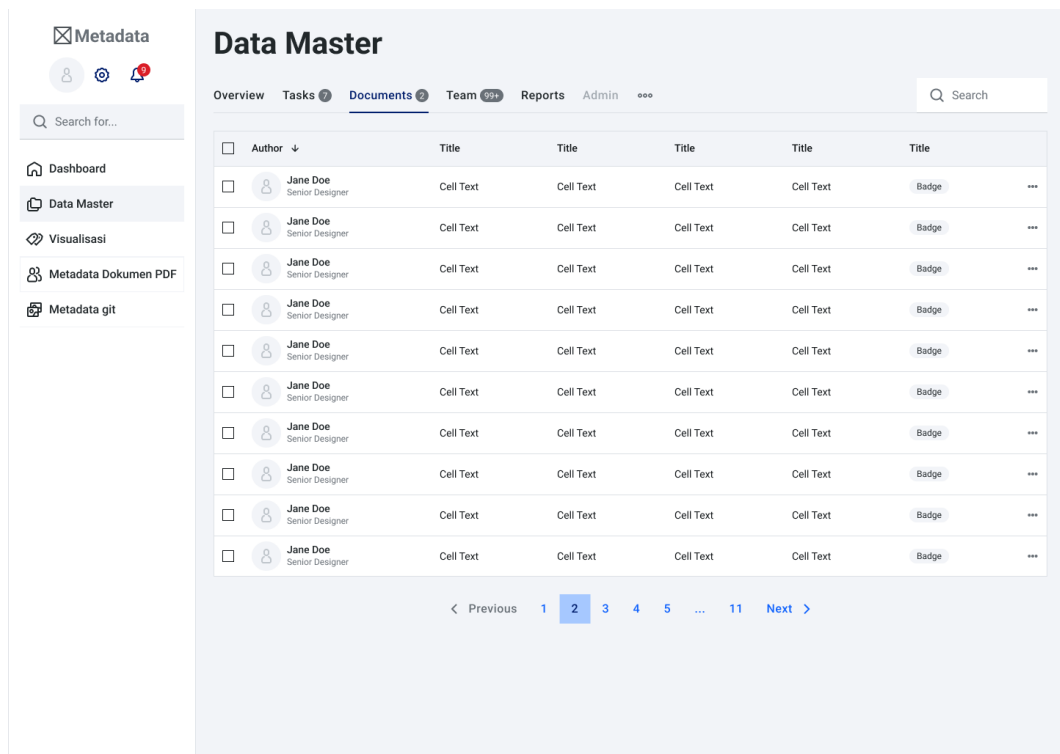
Tabel dan grafik yang menyajikan statistik metadata dokumen PDF, seperti rata-rata frekuensi penggunaan, distribusi penulis, atau jenis tugas yang paling umum. Informasi ini memberikan wawasan tentang karakteristik dokumen yang dihasilkan oleh mahasiswa.



Gambar 4 Metadata PDF Wireframe

d. Data Master

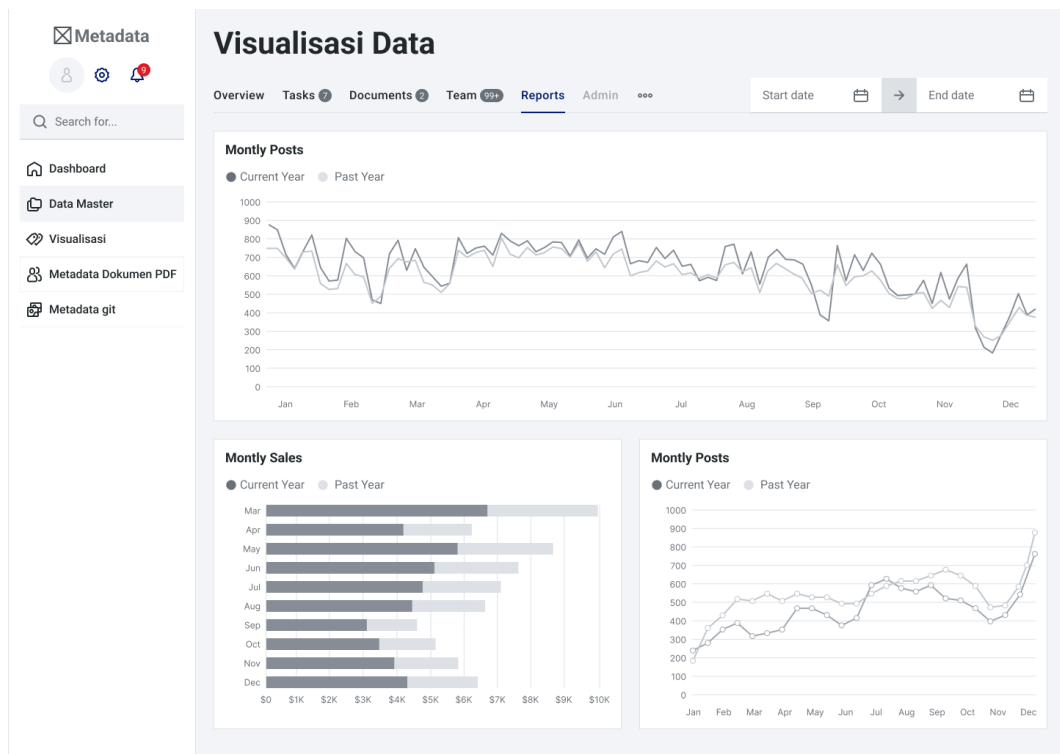
Data Master memberikan gambaran keseluruhan tentang bagaimana Dosen akan berinteraksi dengan fokus pada kemudahan navigasi, pencarian, dan manajemen data master terkait dokumen PDF dan repository Git.



Gambar 5 Wireframe Data Master

e. Data Visualisasi

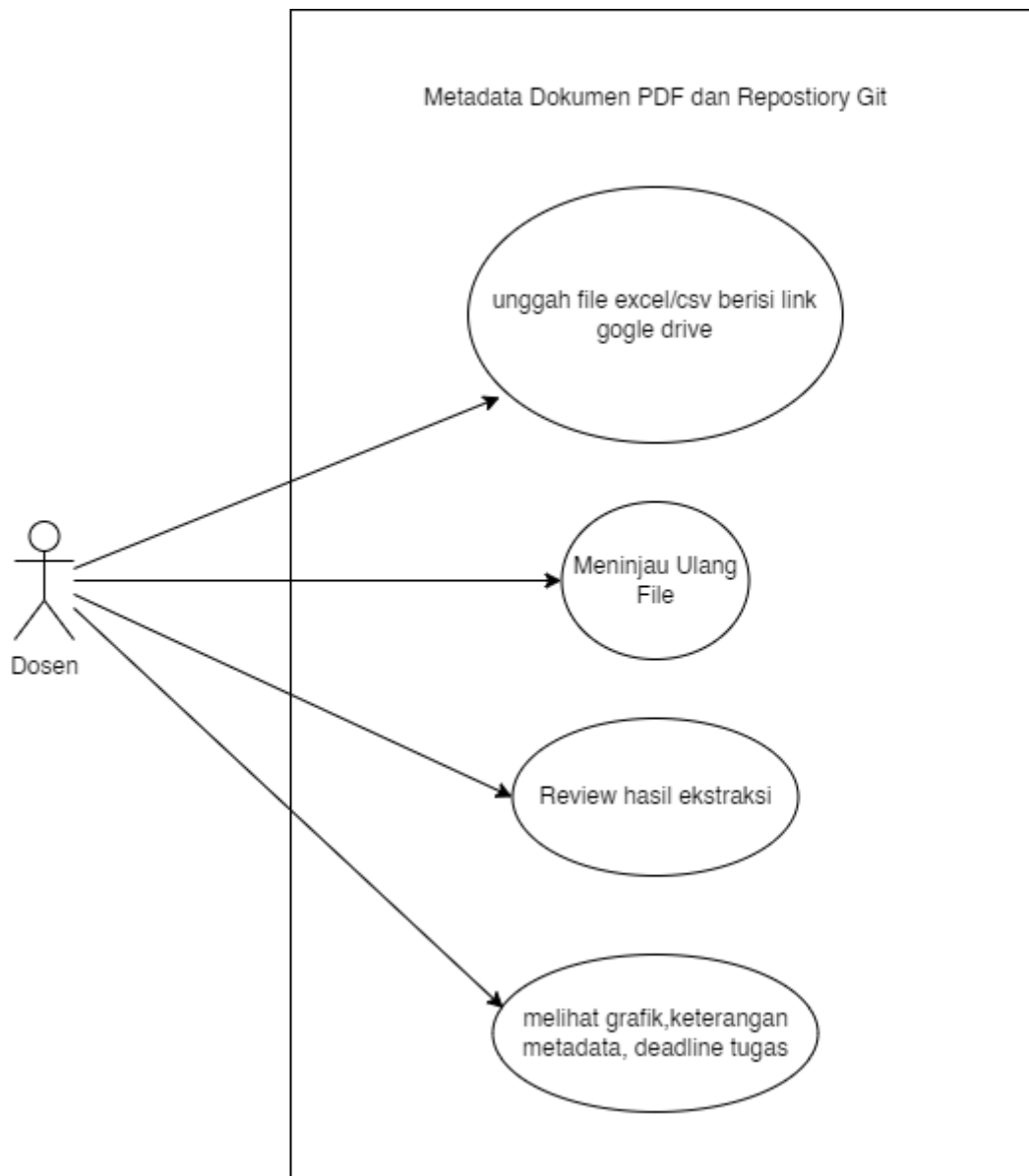
Halaman Visualisasi Data dirancang untuk memberikan gambaran yang jelas dan informatif mengenai hasil analisis metadata dokumen PDF dan repository Git serta memberikan gambaran keseluruhan tentang bagaimana pengguna akan melihat dan berinteraksi dengan visualisasi hasil analisis metadata dokumen PDF dan repository Git dalam halaman Visualisasi Data.



Gambar 6 Wireframe Visualisasi Data

3.4.2 Use Case

Dosen mengunggah file Excel atau CSV yang berisi link ke dokumen PDF. Ini adalah langkah awal untuk memulai proses ekstraksi metadata. Setelah file diunggah, sistem melakukan validasi untuk memastikan bahwa format file yang diunggah valid. Jika file tidak valid, langkah ini akan mengarahkan ke pesan atau tindakan yang sesuai. Jika file dinyatakan valid, sistem melakukan ekstraksi informasi dari dokumen PDF yang terkait. Ini bisa mencakup metadata seperti penulis, tanggal pembuatan, dan lainnya. Hasil ekstraksi dari dokumen PDF ditampilkan kepada Dosen. Ini dapat berupa ringkasan metadata atau informasi lain yang diperoleh dari dokumen. Dosen memiliki opsi untuk menampilkan grafik, metadata lebih rinci, atau informasi seputar kesesuaian deadline tugas mahasiswa. Ini memberikan fleksibilitas dalam melihat informasi sesuai kebutuhan.



Gambar 7 Use Case

3.4.3 Arsitektur Sistem

Sistem ini dirancang untuk analisis metadata dokumen PDF dan repository Git yang berasal dari tugas mahasiswa. Berikut adalah deskripsi alur kerja sistem:

1. Input Data:

Dosen, sebagai pengguna utama, mengunggah file PDF ke Google Drive. File dan menaruh repository Git yang terkait.

2. Ekstraksi Metadata:

Sistem melakukan dua tahap ekstraksi metadata terpisah:

Ekstraksi Metadata Dokumen PDF: Informasi metadata dari dokumen PDF diekstraksi, termasuk frekuensi penggunaan, penulis, dan jenis tugas.

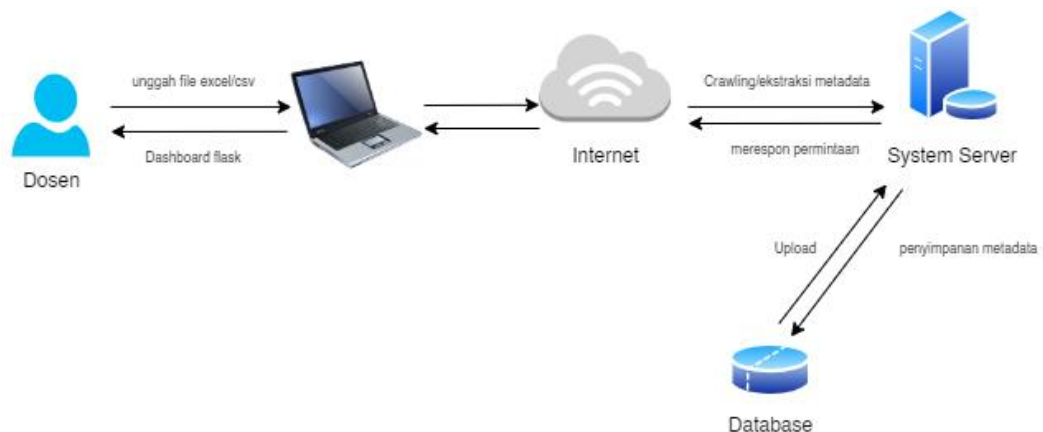
Ekstraksi Metadata Repository Git: Melibatkan proses crawling untuk mendapatkan informasi terkini dari repositori Git yang diakses melalui link-link yang diberikan.

3. Penyimpanan Metadata

Hasil ekstraksi metadata dari dokumen PDF dan repository Git disimpan dalam sebuah database di server atau cloud. Metadata ini mencakup berbagai informasi seperti aktivitas, kolaborasi, dan kualitas pekerjaan mahasiswa.

4. Dashboard Flask:

Metadata yang telah disimpan digunakan sebagai basis untuk menampilkan informasi melalui sebuah Dashboard Flask yang interaktif. Dosen dapat dengan mudah melihat ringkasan hasil analisis, grafik, dan detail metadata untuk mengetahui proses dan hasil pembelajaran mahasiswa.



Gambar 8 Arsitektur Sistem

Dengan demikian, arsitektur sistem ini memberikan pendekatan yang terstruktur dan efisien dalam menganalisis data tugas mahasiswa, serta memberikan pemahaman yang lebih mendalam tentang aktivitas dan capaian mahasiswa dalam pembelajaran.

3.6 Uji Coba Sistem

Pengujian sistem pendukung keputusan dilakukan dengan metode *Black Box Testing* yang bertujuan untuk mengevaluasi fungsionalitas sistem tanpa memperhatikan struktur internal atau implementasi kode.

Berikut adalah contoh tabel untuk uji coba sistem dengan fokus pada validasi dan kemungkinan kegagalan. Tabel ini mencakup beberapa skenario uji dan mencantumkan atribut kemungkinan gagal yang mungkin terjadi:

Tabel 1 Uji Coba Sistem

No	Skenario Uji	Langkah-langkah Uji	Ekspetasi Hasil	Kemungkinan Gagal	Validasi
1	Unggah File Excel/CSV Valid	Dosen unggah file Excel/CSV yang berisi link GDrive dokumen PDF dan repository Git.	File berhasil diunggah.	File tidak valid, link GDrive tidak sesuai format.	
2	Validasi File Excel/CSV	Sistem melakukan validasi file Excel/CSV.	File Excel/CSV valid.	File Excel/CSV tidak valid (bukan format Excel/CSV).	
3	Ekstraksi Metadata PDF	Dosen pilih opsi untuk mengekstraksi metadata dokumen PDF.	Metadata PDF berhasil diekstraksi.	Proses ekstraksi gagal atau tidak memberikan hasil.	
4	Ekstraksi Metadata Repository Git	Dosen pilih opsi untuk mengekstraksi metadata repository Git.	Metadata Repository Git berhasil diekstraksi.	Proses ekstraksi gagal atau tidak memberikan hasil.	
5	Tampilkan Hasil Ekstraksi	Dosen pilih opsi untuk melihat hasil ekstraksi.	Hasil ekstraksi ditampilkan dengan benar.	Tampilan hasil ekstraksi tidak sesuai.	
6	Tampilkan Grafik, Metadata, Kesesuaian Deadline	Dosen pilih opsi untuk melihat grafik, metadata, dan kesesuaian deadline.	Tampilan sesuai dengan ekspektasi.	Tampilan tidak sesuai atau tidak lengkap.	

DAFTAR PUSTAKA

- Alamoudi, A., Alomari, A., Alwarthan, S., & Atta-Ur-rahman. (2021). A rule-based information extraction approach for extracting metadata from pdf books. *ICIC Express Letters, Part B: Applications*, 12(2), 121–132. <https://doi.org/10.24507/icicelb.12.02.121>
- Alghamdi, H., Dawwas, W., Almutairi, T. H., & Atta-Ur-rahman. (2022). EXTRACTING TOC AND METADATA FROM PDF BOOKS: A RULE-BASED APPROACH. *ICIC Express Letters, Part B: Applications*, 13(2), 133–143. <https://doi.org/10.24507/icicelb.13.02.133>
- Ambriani, D., & Nurhidayat, A. I. (n.d.). RANCANG BANGUN REPOSITORY PUBLIKASI ILMIAH DOSEN BERBASIS WEB MENGGUNAKAN FRAMEWORK LARAVEL. In *Jurnal Manajemen Informatika* (Vol. 10, Issue 01).
- Atkinson, J., Gonzalez, A., Munoz, M., & Astudillo, H. (2014). Web metadata extraction and semantic indexing for learning objects extraction. *Applied Intelligence*, 41(2), 649–664. <https://doi.org/10.1007/s10489-014-0557-6>
- Barbarese, A. (n.d.). *Efficient construction of metadata-enhanced web corpora*. <http://w3techs.com/technologies/details/cm->
- Dwiyani, N. (2017). *SKENARIO METADATA UNTUK SISTEM E-LEARNING*.
- Indrawati ariani, Yoganingrum Ambar, & Subagyo Hendro. (2020). *Identifying and Correcting the Indonesian Bibliography Metadata using Regular Expression*.
- Kim, Y., Kim, J., Jeon, H., Kim, Y. H., Song, H., Kim, B., & Seo, J. (2021). Github: Visual analytics for understanding software development history through git metadata analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 656–666. <https://doi.org/10.1109/TVCG.2020.3030414>
- Nurandini, I., Fatchul Huda, A., Matematika, J., Sunan, U., & Djati, G. (n.d.). Klastering Dokumen dengan Menambahkan Metadata Menggunakan Algoritma COATES. *Jurnal Kubik*, 2(2).
- PyGithub. (n.d.). *PyGithub*. <https://Pygithub.Readthedocs.io/En/Latest/Introduction.Html>.

- PyPDF2. (n.d.). *PyPDF2*.
<https://Pypdf2.Readthedocs.Io/En/3.0.0/User/Metadata.Html>.
- Scott, M., Johnston, S. J., & Cox, S. J. (n.d.). *Metagit: Decentralised Metadata Management With Git*. <https://doi.org/10.5258/SOTON/393614>
- Tkaczyk, D., Szostek, P., Fedoryszak, M., Dendek, P. J., & Bolikowski, Ł. (2015). CERMINE: Automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition*, 18(4), 317–335. <https://doi.org/10.1007/s10032-015-0249-8>
- Zhu, M., & Cole, J. M. (2022). PDFDataExtractor: A Tool for Reading Scientific Text and Interpreting Metadata from the Typeset Literature in the Portable Document Format. *Journal of Chemical Information and Modeling*, 62(7), 1633–1643. <https://doi.org/10.1021/acs.jcim.1c01198>