

Evaluasi Kualitas Pendidikan Dasar di Sumatera Selatan Menggunakan Business Intelligence Model

Muhamad Meiko Triputra¹, Ahmad Rifai^{*2}, Ken Ditha Tania³

^{1,2,3}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia
Email: triputrameiko@gmail.com, ahmadrifai@unsri.ac.id, kenya.tania@gmail.com

Abstrak

Statistik Pendidikan 2023 yang dipublikasikan oleh Badan Pusat Statistik (BPS) menyatakan bahwa sarana, prasarana, dan hasil capaian proses pendidikan menjadi faktor-faktor yang memegang peranan penting dalam mendukung terwujudnya keberhasilan siswa saat belajar di sekolah. Tujuan dari penelitian ini yaitu membangun sebuah alat pendukung keputusan berdasarkan data pendidikan dasar di provinsi Sumatera Selatan dengan mengimplementasikan model *business intelligence*. Data pendidikan dasar yang dikumpulkan akan diintegrasikan dan disimpan ke dalam *data warehouse*. Selanjutnya dilakukan analisis *data mining* untuk mengelompokkan kabupaten atau kota menggunakan algoritma *k-means clustering*. Terdapat tiga pola *clustering* yang dianalisis dan dievaluasi menggunakan matriks evaluasi *Davies-Bouldin Index (DBI)*, yakni pengelompokan berdasarkan keseimbangan infrastruktur dan sumber daya pendidikan dengan nilai evaluasi 0.519, pengelompokan berdasarkan angka siswa mengulang dengan nilai evaluasi 0.41, dan pengelompokan berdasarkan angka siswa putus sekolah dengan nilai evaluasi 0.596. Data yang tersimpan di *data warehouse* dan hasil analisis *clustering* akan divisualisasikan dalam bentuk *dashboard* interaktif yang dirancang untuk meningkatkan pemahaman terkait kondisi pendidikan dasar secara menyeluruh. *Dashboard* yang dihasilkan berkontribusi secara signifikan dalam mengidentifikasi kabupaten atau kota yang membutuhkan perhatian khusus sehingga memudahkan pemangku kebijakan untuk mengambil keputusan berbasis data dalam mengevaluasi maupun meningkatkan kualitas pendidikan dasar.

Kata kunci: *business intelligence, clustering, dashboard, data warehouse, pendidikan dasar*

Evaluating the Quality of Basic Education in South Sumatera Using the Business Intelligence Model

Abstract

Education Statistics 2023 published by the Central Bureau of Statistics (BPS) states that facilities, infrastructure, and the results of the educational process are several factors that play an important role in supporting the realization of student success while studying at school. The purpose of this research is to build a decision support tool based on basic education data in South Sumatera Province by implementing a business intelligence model. The collected basic education data will be integrated and stored in a data warehouse. In addition, data mining analysis is performed to cluster districts or cities using the k-means clustering algorithm. Three clustering patterns are analyzed and evaluated using the Davies-Bouldin Index (DBI) scoring matrix, namely clustering based on the balance of infrastructure and educational resources with a score of 0.519, clustering based on the number of repeating students with a score of 0.41, and clustering based on dropout rates with a score of 0.596. The data stored in the data warehouse and the results of the clustering analysis are visualized in the form of an interactive dashboard designed to improve understanding of the overall state of basic education. The resulting dashboard is instrumental in identifying districts or cities that need special attention, making it easier for policymakers to make data-driven decisions to evaluate and improve the quality of basic education.

Keywords: *business intelligence, clustering, dashboard, data warehouse, basic education*

1. PENDAHULUAN

Sumatera Selatan menjadi salah satu provinsi yang ada di Indonesia dengan jumlah penduduk muda terbanyak. Sebanyak 2,9 juta atau 34% dari semua penduduk di provinsi ini adalah anak-anak. Anak-anak yang berusia enam hingga dua belas tahun biasanya melakukan tugas perkembangan seperti kemampuan yang harus

dikuasai anak Sekolah Dasar (SD) [1]. Pendidikan di tingkat SD bagi anak-anak memegang peranan penting dalam pembentukan karakter dan perkembangan intelektualnya. Sekolah juga merupakan lembaga pendidikan yang bertanggung jawab sepenuhnya atas karakter generasi masa depan [2]. Lingkungan sekolah menjadi salah satu aspek yang dapat mempengaruhi kenyamanan dan motivasi belajar anak. Lingkungan sekolah yang mendukung dapat menciptakan rasa nyaman dan meningkatkan motivasi belajar sehingga berkontribusi positif terhadap kualitas pendidikan mereka, begitupun sebaliknya.

Pada kasus provinsi Sumatera Selatan, kualitas pendidikan masih menjadi masalah utama. Berdasarkan profil singkat mengenai *SDG* untuk anak-anak yang dipublikasi oleh *UNICEF*, dalam hal akses ke sekolah dasar, provinsi ini memiliki nilai yang sama dengan rata-rata nasional. Namun, hanya 50% dari anak-anak SD yang mampu mencapai ambang batas minimum nasional untuk kemampuan membaca dan 20% untuk kemampuan matematika. Hal ini menjadi perhatian yang serius bagi setiap Dinas Pendidikan daerah di provinsi Sumatera Selatan untuk mengambil keputusan atau kebijakan yang dapat mengevaluasi kualitas pendidikan anak-anak SD.

Berdasarkan UU No. 20 Tahun 2003 Bab IV Pasal 10 dan 11, Pemerintah dan pemerintah daerah dalam hal ini yaitu Dinas Pendidikan daerah memiliki hak untuk memutuskan bagaimana penyelenggaraan pendidikan sesuai peraturan perundang-undangan yang berlaku serta memiliki kewajiban untuk memberikan layanan dalam menjamin terselenggarakannya pendidikan yang bermutu bagi setiap warganya. Keputusan tentang kebijakan pendidikan tidak boleh didasarkan pada otoritas atau kepentingan para pemegang keputusan (*authority based*) sebaliknya, harus didasarkan pada kondisi dan kebutuhan yang nyata di wilayah tersebut (*research based*) [3]. Dalam pengambilan kebijakan yang bersifat *research based*, Dinas Pendidikan di setiap daerah provinsi Sumatera Selatan memerlukan sebuah alat yang dapat mengekstraksi pengetahuan sebagai dasar pengambilan kebijakan yang akan dibuat.

Business Intelligence (BI) didefinisikan sebagai kombinasi proses, kebijakan, dan teknologi untuk mengumpulkan, memanipulasi, menyimpan, dan menganalisis kumpulan data dalam jumlah besar yang berasal dari sumber internal maupun eksternal, untuk mengkomunikasikan informasi, menciptakan pengetahuan dan pengambilan keputusan [4]. BI mendukung proses pengambilan keputusan dengan memungkinkan agregasi data, *warehousing* dan *processing* data, integrasi data dari berbagai sumber dan jenis data, penemuan pengetahuan baru dan kemampuan *forecasting* [5]. Tujuan utama BI yaitu mengubah data menjadi pengetahuan untuk meningkatkan keputusan. Alat-alat penting BI terdiri dari penyimpanan data, ekstraksi, transmisi, pemuatan, proses analisis *online*, penambangan data, dan pelaporan [6]. BI tidak hanya memberikan informasi baru dan pendukung keputusan, namun dapat juga memberikan umpan balik dan evaluasi terhadap keputusan sebelumnya. BI yang sukses dapat membantu organisasi membuat keputusan yang paling efektif pada waktu yang tepat melalui pengintegrasian dan analisis data dengan sistem pendukung keputusan [6]. Sistem BI dimulai dari pengumpulan data-data dari berbagai sumber internal maupun eksternal. Data-data yang telah dipilih untuk dikumpulkan tersebut secara kolektif lalu dilakukan proses *Extract, Transform, Load* (ETL) [7] sebelum dimasukkan ke dalam *data warehouse*. *Data warehouse* sebagai tempat penyimpanan data yang konsisten dapat memungkinkan pemanfaatan visualisasi, pelaporan, dan menyederhanakan penggunaan *library* saat dilakukan *data mining* dan *machine learning* [8].

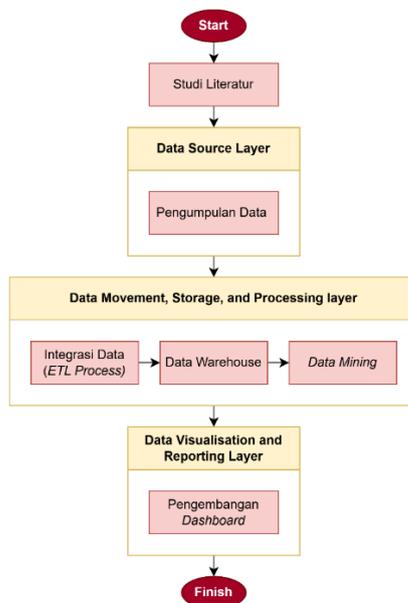
Pengimplementasian BI telah banyak dilakukan di berbagai proses bisnis baik itu dalam lingkup perusahaan, organisasi, bahkan lembaga pemerintahan. Beberapa penelitian sebelumnya terkait implementasi BI di sektor pendidikan yaitu penelitian mengenai pengembangan model BI untuk mengevaluasi pendidikan tinggi nasional di negara Iran [9]. Penelitian ini menerapkan model BI untuk mendukung *monitoring* indikator pendidikan tinggi menggunakan *OLAP* dan *data mining* pada peramalan tren masa depan (*forecasting*) melalui integrasi dengan sumber data internal dan eksternal yang disimpan di *Data Warehouse*. Sistem dalam bentuk *dashboard* menjadi hasil yang dikembangkan dari penelitian ini sebagai upaya memberikan pandangan terpadu tentang sistem pendidikan tinggi Iran dibandingkan dengan negara tetangga lainnya. Penelitian selanjutnya yaitu implementasi BI pada proses penerimaan mahasiswa baru di universitas Mulia [10]. Penelitian ini bertujuan untuk memudahkan dalam manajemen laporan terkait penerimaan mahasiswa, mendukung pengambilan keputusan dan perencanaan strategi di masa depan. Penelitian hanya dilakukan dengan dua tahap analisis yaitu pengumpulan data yang perancangan dashboard. Selanjutnya yaitu penelitian yang membahas solusi berbasis BI untuk mendukung pengelolaan urusan akademik di *Taibah University* [11]. Solusi ini mencakup pengumpulan data, pengembangan model multidimensional, dan visualisasi hasil. Penelitian ini menggunakan *SQL Server Data Tools* untuk menghasilkan indikator statistik dan prediktif yang dibutuhkan dalam kegiatan akademik.

Penelitian sebelumnya diatas telah berhasil diimplementasikan di sektor pendidikan terkhususnya pendidikan tinggi. Namun, penelitian tersebut belum mengkaji pengimplementasian BI pada sektor pendidikan dasar. Penelitian ini akan mengadaptasi dari beberapa penelitian diatas dengan fokus pada sektor pendidikan dasar, sehingga pengimplementasian BI pada penelitian ini akan dapat menjadi solusi pengembangan alat bagi Dinas Pendidikan daerah untuk mendapatkan pengetahuan yang digunakan dalam memutuskan suatu kebijakan yang bersifat *research based*. Dengan adanya penelitian ini, diharapkan dapat menjadi panduan yang strategis

dalam menyusun dan mengambil kebijakan yang tepat sasaran dan berorientasi pada peningkatan kualitas pendidikan dasar di provinsi Sumatera Selatan.

2. METODE PENELITIAN

Tahap-tahap pada penelitian ini dilakukan dengan arsitektur umum *Business Intelligence* berbasis *data warehouse* tradisional dengan pendekatan seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penelitian ini diawali dengan studi literatur yaitu mencari beberapa artikel terkait pengimplementasian BI dan sumber materi-materi yang terkait dengan BI. Dari hasil pencarian didapatkan bahwa BI tidak hanya di implementasikan oleh sebuah perusahaan saja, namun juga dapat diimplementasikan terhadap setiap organisasi dengan proses bisnisnya masing-masing termasuk pada instansi pemerintahan seperti dinas pendidikan. Berdasarkan tahapan penelitian pada Gambar 1 usulan arsitektur BI mencakup tiga *layer* yaitu *data source layer*, *data movement, storage, and processing layer*, dan *data visualization and reporting layer* [12][13].

2.1. Data Source Layer

Data source layer mencakup semua sumber data yang digunakan oleh sistem [13]. Pada lapisan ini akan diawali dengan pengumpulan data-data sebaran siswa SD yang bersumber dari dataset pendidikan di laman portal data Kemendikbudristek. Data yang dikumpulkan tersebut merupakan data dari sumber eksternal yang dapat diakses oleh semua orang (*open source*) yang semuanya dikumpulkan dari rentang tahun 2021 sampai 2023, sehingga total data yang dikumpulkan sebanyak 48 data dalam format .csv. Kumpulan data tersebut selanjutnya akan diolah pada *data movement, storage, dan processing layer*.

2.2. ETL Process

ETL (*Extract, Transform, Load*) merupakan sekumpulan proses seperti mengekstrak data dari berbagai sumber data internal dan eksternal, melakukan operasi transaksional yang mencakup pembersihan, penyatuan, perubahan format, dan berbagai operasi transaksional lainnya, dan memuatnya ke *data warehouse* [14]. Proses ETL dilakukan untuk memastikan bahwa data yang akan dimasukkan ke dalam *data warehouse* sesuai dengan kebutuhan pengembangan. Proses ini dilakukan menggunakan alat *Pentaho Data Integration* versi 9.4.0.0.

2.3. Data Warehouse

Data Warehouse (DW) adalah lokasi penyimpanan data non-volatil yang berfokus pada domain, terintegrasi, bervariasi waktu, dan digunakan untuk mendukung pengambilan keputusan. DW menggunakan model multidimensi untuk mengatur datanya. DW didefinisikan oleh dimensi dan fakta. Dimensi menunjukkan

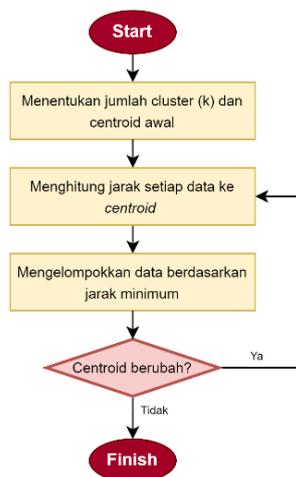
perspektif dimana fakta-fakta direpresentasikan. Fakta berkaitan dengan peristiwa, biasanya dikaitkan dengan nilai numerik yaitu ukuran, dan direferensikan menggunakan elemen dimensi [9][15][16]. Sebelum disimpan kedalam *data warehouse*, data harus sudah terlebih dahulu dilakukan integrasi data dengan proses ETL agar memudahkan proses analisis dan visualisasi pada tahap selanjutnya.

Saat melakukan pemodelan skema DW terdapat tiga skema desain yang dapat digunakan yaitu *star schema*, *snowflake schema*, dan *fact constellation schema*. Pemilihan masing-masing skema dilakukan berdasarkan identifikasi data yang telah dikumpulkan sebelumnya. Data dengan satu tabel fakta terhadap beberapa dimensi cocok menggunakan *star schema* atau *snowflake schema*. Sedangkan *fact constellation schema* cocok dipilih apabila terdapat lebih dari satu tabel fakta yang berbagi dimensi.

2.4. Clustering

Beberapa data yang telah disimpan di *data warehouse* dapat dilakukan proses analisis data dengan berbagai metode, pada penelitian ini teknik *data mining* merupakan metode analisis yang dipilih. *Data mining* dikembangkan untuk mengurangi tenaga kerja, menghasilkan kecerdasan manusia dengan bekerja lebih efisien daripada manusia. *Data mining* harus dapat belajar secara mandiri. *Data mining* juga dapat dimanfaatkan dalam konsep eksplorasi data untuk mengintegrasikan sejumlah besar data yang tidak saling berhubungan, menemukan korelasi dan pola yang berguna, dan memulihkan beberapa informasi yang berharga dari data [17]. Teknologi *Data mining* terdiri dari statistik, klasifikasi, *clustering*, regresi, dan asosiasi [17][18][19]. *Clustering* menjadi teknologi *data mining* yang digunakan untuk menganalisis data yang tidak memiliki label sehingga cocok dengan data sebaran siswa SD yang ada.

K-Means Clustering menjadi satu dari semua metode *clustering* data non-hierarki populer yang dipilih untuk mengelompokkan beberapa pola yang didapatkan pada data. *K-Means Clustering* dipilih karena kompleksitas komputasionalnya yang rendah dan sederhana sehingga dapat digunakan dengan mudah dalam memecahkan masalah pengelompokan [20]. Selain itu, algoritma ini mampu menangani data yang berukuran besar, fleksibel dalam melakukan *tuning* terhadap paramaterya dibanding dengan algoritma pengelompokan yang lain seperti DBSCAN, hierarchical clustering. Beberapa pola pengelompokan yang akan dilakukan menggunakan *K-Means Clustering* yaitu pengelompokan berdasarkan keseimbangan infrastruktur dan sumber daya pendidikan, pengelompokan berdasarkan angka siswa mengulang, dan pengelompokan berdasarkan angka siswa putus sekolah. Langkah kerja *K-Means Clustering* dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Algoritma *K-Means Clustering*

Langkah dimulai dengan menentukan jumlah *cluster* dan *centroid* awal. Untuk menentukan jumlah *cluster* dapat dilakukan dengan beberapa pendekatan seperti menggunakan metode *Elbow*, *Davies-Bouldin Index*, dan pendekatan lainnya sedangkan untuk menentukan *centroid* awal dapat dipilih secara acak. Langkah berikutnya adalah menghitung jarak masing-masing data terhadap *centroid* dan mengelompokkan data menurut jarak terkecil dengan *centroid*. Untuk menghitung jarak tersebut biasanya menggunakan *Euclidean Distance* dengan persamaannya sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

Pada persamaan (1), x_1, x_2, \dots, x_n dinyatakan sebagai titik data sedangkan y_1, y_2, \dots, y_n dinyatakan sebagai titik *centroid*. Langkah tersebut akan diulang secara iteratif hingga tercapai kondisi *konvergen*, yaitu ketika *centroid* tidak mengalami perubahan signifikan lagi, atau ketika jumlah iterasi maksimum yang ditentukan telah tercapai. Jika telah mencapai kondisi tersebut dapat dilakukan identifikasi perbedaan ciri tiap kluster dari setiap fitur dasar yang dimiliki setiap data dalam suatu kluster tersebut.

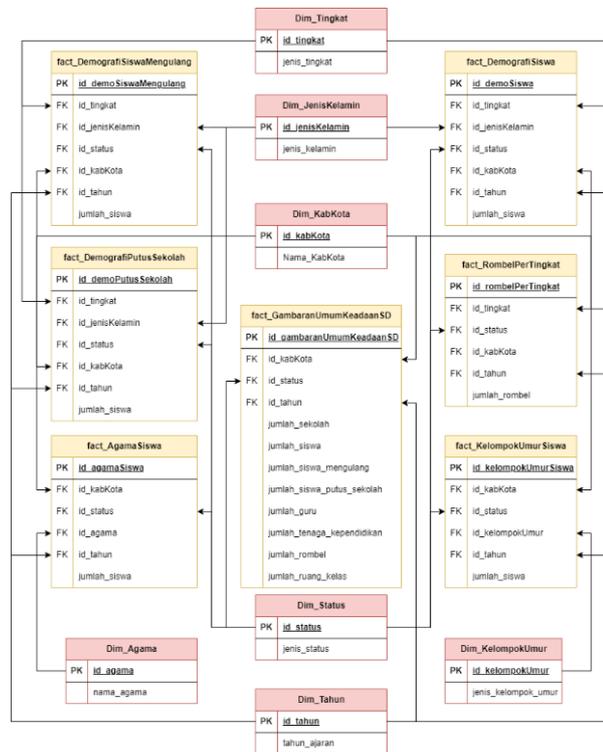
2.5. Data Visualization dan Reporting Layer

Terakhir pada *data visualization and reporting layer* menampilkan data yang telah dianalisis kepada para pemangku kebijakan dalam beberapa bentuk visualisasi seperti *dashboard, report, spreadsheet* dan *ad hoc queries* [13]. Pada penelitian ini, pemaparan hasil dari *data mining* dan data pada *data warehouse* berupa visualisasi dan laporan yang dikembangkan dalam bentuk *dashboard*. *Dashboard* yang disajikan akan memiliki kemampuan visual seperti kontrol filter yang mudah digunakan, laporan-laporan yang ada pada dashboard akan dihubungkan satu sama lain dengan menggunakan id dari tabel dimensi sehingga ketika filter diterapkan pada sebuah laporan, semua laporan yang terhubung akan diperbarui. *Dashboard* juga memungkinkan pengguna untuk melakukan berbagai analisis untuk mempelajari sebaran data siswa Sekolah Dasar di Provinsi Sumatera Selatan. Perancangan *dashboard* dilakukan dengan alat *Microsoft Power BI* versi 2.138.1452.0.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari tahapan penelitian yang telah dilakukan mencakup pemodelan desain *data warehouse*, proses ETL yang dilakukan pada data yang telah dikumpulkan, dan analisis *data mining* dengan algoritma *K-means clustering* pada *data movement, storage, and processing layer*. Sedangkan pada *data visualization and reporting layer* akan menghasilkan *dashboard* yang interaktif berdasarkan data yang telah disimpan di *data warehouse*.

3.1. Pemodelan Desain Data Warehouse



Gambar 3. Fact Constellation Schema

Pada penelitian ini, model skema yang dipilih yaitu *fact constellation schema* atau *galaxy schema*. Pada skema ini terdapat lebih dari satu tabel fakta dimana setiap tabel fakta memungkinkan berbagi tabel dimensi yang sama. Pemilihan skema ini juga dapat mendukung analisis multidimensional yang lebih kompleks dari berbagai perspektif seperti demografi siswa, rombongan belajar, dan tabel fakta lainnya. Pemodelan desain *data warehouse* dengan skema ini dapat dilihat pada Gambar 3.

Fact Constellation Schema diatas digunakan sebagai model desain *data warehouse* yang akan dirancang. Skema memiliki 7 tabel dimensi yaitu tingkat, jenis kelamin, kabupaten atau kota, status, agama, tahun, dan kelompok umur dan 7 tabel fakta yaitu demografi siswa, demografi siswa mengulang, demografi siswa putus sekolah, gambaran umum keadaan SD, rombel per tingkat, agama siswa, dan kelompok umur siswa. Tabel fakta demografi siswa, demografi siswa mengulang dan demografi siswa putus sekolah menggambarkan karakteristik dari setiap jumlah siswa secara umum, jumlah siswa yang mengulang, dan jumlah siswa yang putus sekolah dengan mengacu pada dimensi yang sama yaitu tingkat, jenis kelamin, status, kabupaten atau kota dan tahun. Tabel fakta gambaran keadaan umum SD menggambarkan karakteristik dari setiap jumlah sekolah, jumlah siswa, jumlah siswa mengulang, jumlah siswa putus sekolah, jumlah guru, jumlah tenaga kependidikan, jumlah rombel dan jumlah ruang kelas dengan mengacu pada dimensi kabupaten atau kota, status dan tahun. Tabel fakta rombel per tingkat menggambarkan karakteristik dari setiap jumlah rombel dengan mengacu pada dimensi tingkat, status, kabupaten atau kota dan tahun. Tabel fakta agama siswa menggambarkan karakteristik dari setiap jumlah siswa dengan mengacu pada dimensi kabupaten atau kota, status, agama dan tahun. Tabel fakta kelompok umur siswa menggambarkan karakteristik setiap jumlah siswa dengan mengacu pada dimensi kabupaten atau kota, status, kelompok umur dan tahun.

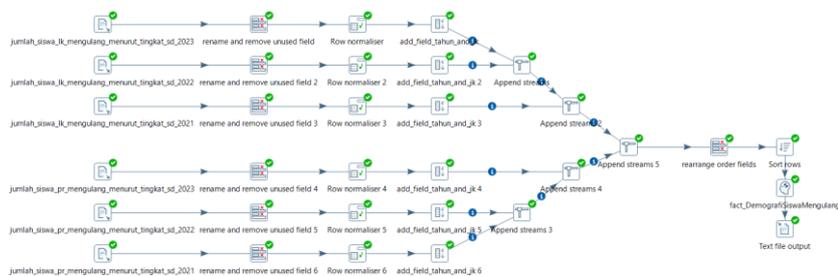
3.2. Hasil Proses ETL

Beberapa tabel fakta dan dimensi yang dimodelkan pada desain *data warehouse* dikembangkan berdasarkan integrasi data dengan proses ETL dari data yang telah dikumpulkan. Beberapa proses *transformation* seperti menginput data, menghapus fitur-fitur yang tidak diperlukan, melakukan penamaan ulang fitur, menambahkan beberapa fitur, menormalisasikan nilai pada fitur tertentu, menggabungkan data input, hingga mengeksport output dilakukan di dalam lembar kerja *Pentaho Data Integration*. Output dari proses ETL selanjutnya dilakukan *import* kedalam *database*.

Pada penelitian ini, proses ETL dilakukan untuk menghasilkan semua tabel fakta. Proses ETL dari semua tabel fakta yang dilakukan di *pentaho data integration* dapat dilihat pada gambar 4-10 berikut.



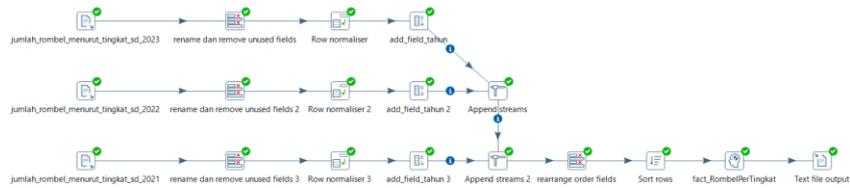
Gambar 4. Proses ETL untuk Membuat Tabel Fakta Demografi Siswa



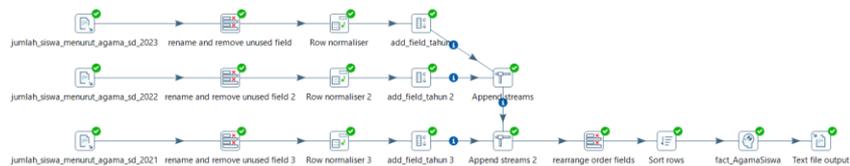
Gambar 5. Proses ETL untuk Menghasilkan Tabel Fakta Demografi Siswa Mengulang



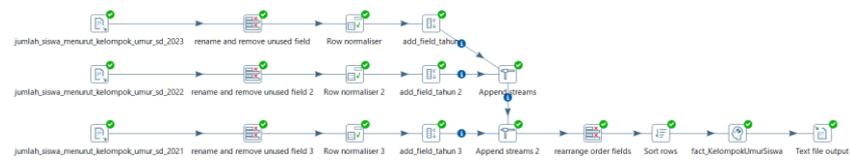
Gambar 6. Proses ETL untuk Menghasilkan Tabel Fakta Demografi Siswa yang Putus Sekolah



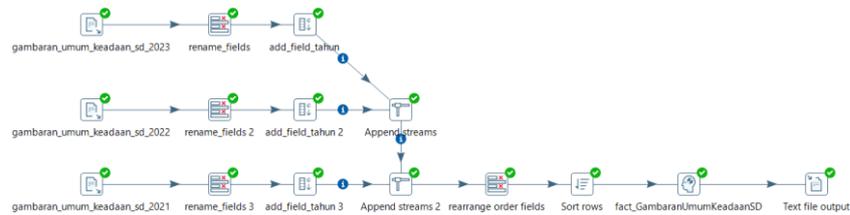
Gambar 7. Proses ETL untuk Menghasilkan Tabel Fakta Rombel per Tingkat



Gambar 8. Proses ETL untuk Menghasilkan Tabel Fakta Agama Siswa



Gambar 9. Proses ETL untuk Menghasilkan Tabel Fakta Kelompok Umur Siswa



Gambar 10. Proses ETL untuk Menghasilkan Tabel Fakta Gambaran Umum Keadaan SD

3.3. Hasil Clustering Data

Berbeda dengan penelitian terdahulu [9] yang menggunakan metode *forecasting* dengan model analisis *time series* untuk mengevaluasi pendidikan tinggi nasional negara Iran. Penelitian ini memanfaatkan metode *clustering* dengan algoritma *k-means clustering* untuk mengelompokkan kabupaten atau kota di Sumatera Selatan berdasarkan beberapa faktor seperti sarana dan prasarana pendidikan serta pencapaian hasil akademik siswa. Pendekatan ini memetakan kondisi yang berbasis data sehingga memungkinkan pengambilan keputusan yang tepat oleh pemangku kebijakan. Terdapat tiga pola pengelompokan yang dilakukan dengan algoritma *k-means clustering* yaitu pengelompokan berdasarkan keseimbangan infrastruktur dan sumber daya pendidikan, berdasarkan angka siswa mengulang, dan berdasarkan angka siswa putus sekolah. Hasil *clustering* dari beberapa pola diatas yaitu sebagai berikut.

Pada *clustering* berdasarkan keseimbangan infrastruktur dan sumber daya pendidikan, fitur yang akan digunakan dilakukan *feature engineering* terhadap beberapa fitur dasar dalam mengetahui keseimbangan jumlah siswa dengan semua fitur yang ada. Fitur-fitur tersebut didapatkan dari data dasar pada tabel fakta gambaran umum keadaan SD. Fitur yang digunakan yaitu fitur rasio siswa per sekolah, rasio siswa per guru, rasio siswa per ruang kelas, rasio siswa per rombel dan rasio siswa per tenaga kependidikan. Fitur-fitur tersebut direduksi

dimensinya dengan *Principal Component Analysis* (PCA) dan didapatkan tiga fitur dengan total variansi data mencapai 95% sehingga siap untuk dilakukan proses *clustering*. Dalam menentukan nilai k optimal menggunakan *Elbow Method*, didapatkan nilai k=4. Hasil *clustering* dievaluasi menggunakan *Davies-Bouldin Index* (DBI) sebagai *metric* evaluasinya dan didapatkan nilainya 0,519. Hasil akhir dari proses *clustering* berdasarkan keseimbangan infrastruktur dan sumber daya pendidikan dapat dilihat pada Tabel 1

Tabel 1. Hasil *Clustering* dan Nilai *Euclidean Distance* pada Iterasi Terakhir Berdasarkan Pola Keseimbangan Infrastruktur dan Sumber Daya Pendidikan

Kabupaten / Kota	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster
Kab. Banyuasin	0.955074	4.421403	3.748414	2.844608	0
Kab. Empat Lawang	2.757748	7.936432	0.376798	2.507745	2
Kab. Lahat	3.188237	8.366621	0.366224	3.199307	2
Kab. Muara Enim	0.476030	5.444892	3.003084	3.045886	0
Kab. Musi Banyuasin	0.927826	5.560672	3.245443	2.987611	0
Kab. Musi Rawas	2.483938	7.122377	2.385887	1.227613	3
Kab. Musi Rawas Utara	1.161943	6.264589	2.287197	1.971201	0
Kab. Ogan Ilir	2.638638	6.251687	3.231370	0.449224	3
Kab. Ogan Komering Ilir	0.320661	5.327445	2.954369	2.414734	0
Kab. Ogan Komering Ulu	1.080032	5.700117	2.490072	1.684140	0
Kab. Ogan Komering Ulu Selatan	2.791028	7.968547	0.156861	2.851465	2
Kab. Ogan Komering Ulu Timur	1.283501	6.609008	1.727434	2.362386	0
Kab. Penukal Abab Lematang Ilir	0.935461	5.955896	2.347434	2.978488	0
Kota Lubuk Linggau	0.927199	4.421663	3.765689	2.951526	0
Kota Pagar Alam	3.223954	6.634743	3.307994	1.006822	3
Kota Palembang	5.339619	0.000000	8.086704	6.610828	1
Kota Prabumulih	1.969858	4.249616	4.431906	4.383552	0

Berdasarkan hasil *clustering* diatas, terdapat empat kluster yang tiap klusternya memiliki perbedaan ciri masing-masing. Untuk mengetahui perbedaan antara tiap kluster, dilakukan identifikasi nilai tiap rasio-rasio terhadap hasil kluster. Perbedaan karakteristik tiap kluster dapat dilihat pada Tabel 2.

Tabel 2. Perbedaan Karakteristik Tiap Kluster dari Hasil *Clustering* Pola Keseimbangan Infrastruktur dan Sumber Daya Pendidikan

Cluster	Ciri Cluster
0	Kota atau kabupaten di cluster ini memiliki rasio siswa pada setiap fitur yang tidak terlalu tinggi atau rendah dibandingkan cluster lain (seimbang). Meskipun demikian, masih diperlukan pengawasan dan peningkatan kecil antara jumlah siswa dengan infrastruktur dan sumber daya pendidikan. Cluster ini dapat dimasukkan ke dalam kategori Dalam Pengawasan .
1	Kota atau kabupaten di cluster ini memiliki rasio siswa pada setiap fitur yang paling tinggi dibandingkan cluster lain. Kondisi ini menciptakan tekanan terhadap infrastruktur dan sumber daya pendidikan untuk seimbang dengan jumlah siswa yang tinggi, sehingga cluster ini dapat dimasukkan kedalam kategori Perlu Penanganan Khusus untuk mencapai keseimbangan tersebut.
2	Kota atau kabupaten di cluster ini memiliki rasio siswa pada setiap fitur yang paling rendah dibandingkan cluster lain. Namun, kondisi ini menciptakan keseimbangan ideal antara jumlah siswa terhadap semua fitur infrastruktur dan sumber daya pendidikan, sehingga tidak memerlukan penanganan segera. Cluster ini dapat dimasukkan kedalam kategori Ideal .
3	Kota atau kabupaten di cluster ini memiliki rasio siswa pada setiap fitur yang hampir sama dengan cluster dua. Namun, terdapat ketidakseimbangan yang signifikan pada rasio siswa terhadap guru dan tenaga kependidikan yang lebih tinggi. Hal ini menciptakan kondisi yang tidak ideal sehingga cluster ini dapat dimasukkan dalam kategori Perlu Tindakan Segera agar dapat menambah guru dan tenaga kependidikan untuk mencapai rasio yang seimbang terhadap jumlah siswa.

Pada *clustering* selanjutnya yaitu berdasarkan angka siswa mengulang, fitur yang digunakan untuk dilakukan proses *clustering* adalah fitur jumlah siswa dan jumlah siswa mengulang. Dari hasil *Elbow method* didapatkan bahwa nilai k=3 merupakan jumlah kluster yang optimal. Hasil *clustering* dievaluasi menggunakan

Davies-Bouldin Index (DBI) sebagai *metric* evaluasinya dan didapatkan nilainya 0,41. Hasil akhir dari proses *clustering* berdasarkan angka siswa mengulang dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Clustering* dan Nilai *Euclidean Distance* pada Iterasi Terakhir Berdasarkan Pola Angka Siswa Mengulang

Kabupaten / Kota	Cluster 0	Cluster 1	Cluster 2	Cluster
Kab. Banyuasin	0.483394	2.601501	1.767512	0
Kab. Empat Lawang	2.622016	0.241880	3.810732	1
Kab. Lahat	2.119679	0.429950	3.302689	1
Kab. Muara Enim	0.688834	1.870249	2.525609	0
Kab. Musi Banyuasin	0.485144	2.161085	2.081131	0
Kab. Musi Rawas	2.009714	0.541005	3.234398	1
Kab. Musi Rawas Utara	2.805624	0.326470	3.940881	1
Kab. Ogan Ilir	2.050408	0.502637	3.196487	1
Kab. Ogan Komering Ilir	1.259342	3.685236	2.980732	0
Kab. Ogan Komering Ulu	2.484653	0.132080	3.492164	1
Kab. Ogan Komering Ulu Selatan	2.683419	0.249802	3.58939	1
Kab. Ogan Komering Ulu Timur	1.971435	0.825232	2.789469	1
Kab. Penukal Abab Lematang Ilir	2.678956	0.246108	3.843357	1
Kota Lubuk Linggau	2.936997	0.388112	3.926183	1
Kota Pagar Alam	3.354990	0.811535	4.334074	1
Kota Palembang	2.246620	3.614660	0.000000	2
Kota Prabumulih	2.979276	0.435951	4.004465	1

Berdasarkan hasil *clustering* diatas, terdapat tiga kluster yang tiap klusternya memiliki perbedaan ciri masing-masing. Untuk mengetahui perbedaan antara tiap kluster, dilakukan identifikasi nilai tiap rasio-rasio terhadap hasil kluster. Perbedaan karakteristik tiap kluster dapat dilihat pada Tabel 4.

Tabel 4. Perbedaan Karakteristik Tiap Kluster dari Hasil *Clustering* Pola Angka Siswa Mengulang

Cluster	Ciri Cluster
0	Kota atau kabupaten pada cluster ini memiliki fitur jumlah siswa mengulang yang lebih tinggi dibanding cluster lain, sehingga cluster ini dapat dimasukkan kedalam kategori Perlu Perhatian agar dapat mengurangi angka pengulangan siswa.
1	Kota atau kabupaten pada cluster ini memiliki fitur jumlah siswa mengulang yang lebih rendah dibanding cluster lain, sehingga cluster ini dapat dimasukkan kedalam kategori Aman dan tidak memerlukan perhatian yang khusus dalam mengurangi jumlah angka siswa yang mengulang.
2	Kota atau kabupaten pada cluster ini memiliki fitur jumlah siswa mengulang yang cukup tinggi dengan jumlah siswa yang juga tinggi, sehingga cluster ini dapat dimasukkan kedalam kategori Dalam Pengawasan agar dapat menahan bahkan menurunkan jumlah siswa yang mengulang.

Terakhir yaitu pada *clustering* berdasarkan angka siswa putus sekolah, fitur yang digunakan untuk dilakukan proses *clustering* adalah fitur jumlah siswa dan jumlah siswa putus sekolah. Dari hasil *Elbow method* didapatkan bahwa nilai $k=2$ merupakan jumlah kluster yang optimal. Hasil *clustering* dievaluasi menggunakan *Davies-Bouldin Index* (DBI) sebagai *metric* evaluasinya dan didapatkan nilainya 0.596. Hasil akhir *clustering* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Clustering* dan Nilai *Euclidean Distance* pada Iterasi Terakhir Berdasarkan Pola Angka Siswa Putus Sekolah

Kabupaten / Kota	Cluster 0	Cluster 1	Cluster
Kab. Banyuasin	2.960132	0.551378	1
Kab. Empat Lawang	0.470401	3.096981	0
Kab. Lahat	0.073323	2.883843	0
Kab. Muara Enim	1.232456	1.732070	0
Kab. Musi Banyuasin	1.303742	1.929890	0
Kab. Musi Rawas	0.102949	2.906781	0
Kab. Musi Rawas Utara	0.783658	3.736669	0
Kab. Ogan Ilir	0.609736	2.447812	0

Kab. Ogan Komering Ilir	3.171258	1.092730	1
Kab. Ogan Komering Ulu	0.394636	3.272341	0
Kab. Ogan Komering Ulu Selatan	0.827078	2.497337	0
Kab. Ogan Komering Ulu Timur	0.628199	2.463658	0
Kab. Penukal Abab Lematang Ilir	0.449534	3.375937	0
Kota Lubuk Linggau	0.514990	3.471225	0
Kota Pagar Alam	1.212424	4.159036	0
Kota Palembang	3.407279	1.642041	1
Kota Prabumulih	0.766330	3.719337	0

Berdasarkan hasil *clustering* diatas, terdapat dua kluster yang tiap klusternya memiliki perbedaan ciri masing-masing. Untuk mengetahui perbedaan antara tiap kluster, dilakukan identifikasi nilai tiap rasio-rasio terhadap hasil kluster. Perbedaan karakteristik tiap kluster dapat dilihat pada Tabel 6.

Tabel 6. Perbedaan Karakteristik Tiap Kluster dari Hasil *Clustering* Pola Angka Siswa Putus Sekolah

Cluster	Ciri Cluster
0	Kota atau kabupaten pada cluster ini memiliki jumlah fitur siswa putus sekolah yang lebih rendah dibanding cluster satu, sehingga cluster ini dapat dimasukkan kedalam kategori Aman dan tidak memerlukan perhatian khusus dalam meningkatkan tingkat kelulusan siswa.
1	Kota atau kabupaten pada cluster ini memiliki jumlah fitur siswa putus sekolah yang lebih tinggi dibanding cluset nol, sehingga cluster ini dapat dimasukkan kedalam kategori Perlu Perhatian yang khusus agar dapat mengurangi jumlah putus sekolah siswa.

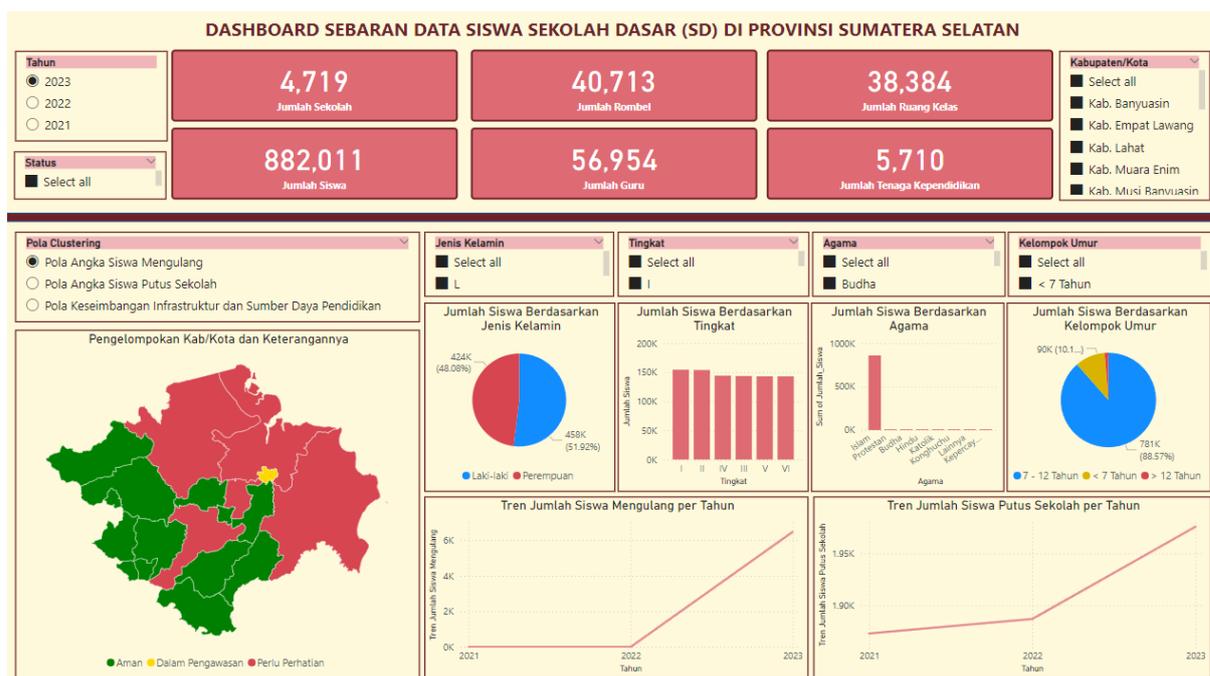
Dari hasil *clustering* ketiga pola diatas, nilai evaluasi DBI belum mencapai hasil yang optimal. Jumlah data yang terbatas (17 kota/kabupaten) dan jumlah dimensi atau fitur yang tinggi (*curse of dimensionality*) menjadi beberapa faktor penyebab nilai evaluasinya tidak mendekati 0. Nilai evaluasi dapat ditingkatkan dengan melakukan PCA untuk mereduksi apabila terdapat banyak fitur yang akan ditambahkan dalam pengelompokkan, menentukan nilai k yang lebih optimal dan tuning paramater lain seperti jumlah iterasi dan lain sebagainya.

3.4. Pengembangan *Dashboard*

Dari hasil *clustering* yang dilakukan dan beberapa data yang telah disimpan pada *data warehouse*. Langkah terakhir dalam implementasi BI yaitu melakukan visualisasi terhadap data yang ada dengan mengembangkan *dashboard*. *Dashboard* memiliki fungsionalitas untuk membantu Dinas Pendidikan daerah di Provinsi Sumatera Selatan untuk mempelajari setiap data yang ada sebelum membuat kebijakan yang bersifat *research based*. Pengembangan *dashboard* dapat dilihat pada Gambar 11.

Dashboard dirancang menggunakan alat Power BI dengan menghubungkan basis data yang telah dibuat sebagai sumber data visualisasinya. Pada *Dashboard* tersebut terdapat kontrol filter berdasarkan setiap dimensi yang ada pada basis data yaitu filter kabupaten atau kota, filter status, filter tahun, filter jenis kelamin, filter tingkat, filter agama, dan filer kelompok umur. Setiap kontrol filter akan responsive terhadap tiap visualisasi yang memiliki hubungan dengan dimensi tersebut. Beberapa laporan yang ditampilkan yaitu indikator seperti jumlah sekolah, jumlah siswa, jumlah rombel, jumlah ruang kelas, jumlah guru, dan jumlah tenaga kependidikan menggunakan modal card, visualisasi jumlah siswa berdasarkan jenis kelamin dan kelompok umur menggunakan pie chart, visualisasi jumlah siswa berdasarkan tingkat dan agama menggunakan bar chart, visualisasi tren jumlah siswa yang mengulang per tahunnya dan tren jumlah siswa yang putus sekolah pertahunnya menggunakan line chart dan yang terakhir yaitu visualisasi hasil *clustering* menggunakan peta bentuk (*shape map*) provinsi Sumatera Selatan yang tiap warnanya menunjukkan kluster.

Dashboard yang telah dikembangkan ini dapat memberikan dampak yang signifikan pada pengambilan keputusan oleh Dinas Pendidikan daerah. Dengan menyajikan data secara visual yang interaktif, *dashboard* dapat membantu para pemangku kebijakan dalam memahami kondisi pendidikan di setiap kabupaten atau kota dengan lebih jelas. Ini dapat mencakup lebih banyak faktor, seperti distribusi siswa, tingkat keberagaman siswa, serta pola seperti angka siswa mengulang dan putus sekolah. Sebagai contoh visualisasi tren siswa yang putus sekolah dapat dioptimalkan untuk mengidentifikasi daerah yang memerlukan intervensi segera. Selain itu, peta kluster memberikan gambaran mengenai wilayah yang tergolong ideal, memerlukan pengawasan, atau yang butuh perhatian khusus. Dengan kata lain, *dashboard* ini tidak hanya mempermudah akses terhadap data, tetapi juga mendukung pengambilan keputusan berbasis data. Hal ini memungkinkan kebijakan yang dihasilkan menjadi lebih tepat sasaran dan efektif dalam meningkatkan kualitas pendidikan di Provinsi Sumatera Selatan.



Gambar 11. Dashboard Sebaran Data Siswa SD di Provinsi Sumatera Selatan

4. KESIMPULAN

Penelitian mengenai ini memberikan pandangan yang lebih luas dan mendalam mengenai kualitas pendidikan di berbagai kota atau kabupaten di provinsi Sumatera Selatan. Penelitian berhasil mengidentifikasi beberapa pola dengan metode *k-means clustering* seperti pola infrastruktur dan sumber daya pendidikan dengan empat cluster dan nilai evaluasi DBI 0.519, pola angka siswa mengulang dengan tiga cluster dan nilai evaluasi DBI 0.41, dan pola angka siswa putus sekolah dengan dua cluster dan nilai evaluasi DBI 0.596. Pengembangan *dashboard* yang interaktif berdasarkan hasil analisis data juga bermanfaat dalam mendukung pengambilan keputusan berbasis data oleh pemangku kebijakan dari beberapa visualisasi yang telah ditampilkan seperti visualisasi indikator pendidikan, visualisasi jumlah siswa berdasarkan beberapa fitur, visualisasi tren jumlah siswa yang mengulang dan putus sekolah serta visualisasi *map* dari hasil *clustering* dalam memetakan kondisi tiap kabupaten atau kota.

Penelitian ini juga memiliki implikasi yang besar dalam membantu pengambilan keputusan yang bersifat *research based* dan berorientasi data oleh Dinas Pendidikan daerah sebagai pemangku kebijakan. Dengan hasil pengembangan *dashboard*, Dinas pendidikan dapat memiliki pemahaman yang lebih baik dan mendalam untuk setiap kota atau kabupaten yang dapat diprioritaskan dalam mengevaluasi dan meningkatkan kualitas pendidikan di wilayah tersebut. Hasil *clustering* yang disertakan di *dashboard* juga dapat memberikan acuan dalam mengalokasikan infrastruktur atau sumber daya pendidikan yang lebih efisien untuk meminimalisir ketimpangan pendidikan antar kabupaten atau kota di Sumatera Selatan.

Penelitian lanjutan dapat difokuskan pada evaluasi jangka panjang implementasi BI dalam mengukur dampak dari kebijakan yang diterapkan berdasarkan *dashboard* yang dibuat terhadap kualitas pendidikan dasar. Penelitian juga dapat mengeksplorasi aspek eksternal seperti kesehatan, ekonomi, dan sosial untuk memberikan analisis yang komprehensif terhadap berbagai aspek yang berpengaruh terhadap kualitas pendidikan. Penelitian selanjutnya juga diharapkan dapat melakukan eksplorasi pendekatan analisis data lain selain *data mining* seperti OLAP, *text mining*, *enterprise search engine*, dan lain sebagainya. Dengan memperluas pendekatan analisis data dan aspek-aspek eksternal lainnya diharapkan dapat memberikan dampak yang lebih masif bagi pemangku kebijakan dalam mengambil keputusan yang tepat dari disisi waktu maupun sasaran.

DAFTAR PUSTAKA

- [1] F. Khaulani, S. Neviyarni, and I. Irdamurni, "Fase dan tugas perkembangan anak Sekolah Dasar," Jurnal Ilmiah Pendidikan Dasar, vol. 7, no. 1, pp. 51–59, 2020, doi: <http://dx.doi.org/10.30659/pendas.7.1.51-59>

-
- [2] A. Prihatmojo and B. Badawi, "Pendidikan karakter di sekolah dasar mencegah degradasi moral di era 4.0," *DWIJA CENDEKIA: Jurnal Riset Pedagogik*, vol. 4, no. 1, pp. 142–152, 2020, doi: <https://doi.org/10.20961/jdc.v4i1.41129>
- [3] S. Samar, R. Rudiyanasyah, and N. I. Saiful, "Pemanfaatan Data Pokok Pendidikan Jenjang SD Dalam Mempengaruhi Formulasi Kebijakan Pemerintah Daerah Pada Dinas Pendidikan di Kabupaten Biak Numfor," *Equilibrium: Jurnal Pendidikan*, vol. 9, no. 2, pp. 215–227, 2021, doi: <https://doi.org/10.26618/equilibrium.v9i2.5230>
- [4] L. J. Basile, N. Carbonara, R. Pellegrino, and U. Panniello, "Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making," *Technovation*, vol. 120, p. 102482, 2023, doi: <https://doi.org/10.1016/j.technovation.2022.102482>.
- [5] A. Obidat, Z. Alziyadat, and Z. Alabaddi, "Assessing the effect of business intelligence on supply chain agility. A perspective from the Jordanian manufacturing sector," *Uncertain Supply Chain Management*, vol. 11, no. 1, pp. 61–70, Dec. 2023, doi: [10.5267/j.uscm.2022.11.010](https://doi.org/10.5267/j.uscm.2022.11.010).
- [6] F. Zafary, "Implementation of business intelligence considering the role of information systems integration and enterprise resource planning," *Journal of Intelligence Studies in Business*, May 2020, doi: <https://doi.org/10.37380/jisib.v1i1.563>.
- [7] W. Villegas-Ch, X. Palacios-Pacheco, and S. Luján-Mora, "A business intelligence framework for analyzing educational data," *Sustainability (Switzerland)*, vol. 12, no. 14, pp. 1–21, Jul. 2020, doi: [10.3390/su12145745](https://doi.org/10.3390/su12145745).
- [8] E. Zdravevski, P. Lameski, C. Apanowicz, and D. Ślęzak, "From Big Data to business analytics: The case study of churn prediction," *Appl Soft Comput*, vol. 90, p. 106164, 2020, doi: <https://doi.org/10.1016/j.asoc.2020.106164>.
- [9] V. Khatibi, A. Keramati, and F. Shirazi, "Deployment of a business intelligence model to evaluate Iranian national higher education," *Social Sciences & Humanities Open*, vol. 2, no. 1, p. 100056, 2020, doi: <https://doi.org/10.1016/j.ssaho.2020.100056>.
- [10] D. Mirwansyah, Riyayatsyah, and D. W. Sari, "Implementation of Business Intelligence in Data and Information for Student Admission Process in Mulia University," *J Phys Conf Ser*, vol. 1807, no. 1, p. 012015, 2021, doi: [10.1088/1742-6596/1807/1/012015](https://doi.org/10.1088/1742-6596/1807/1/012015).
- [11] W. Boulila, M. Al-kmal, M. Farid, and H. Mugahed, "A business intelligence based solution to support academic affairs: case of Taibah University," *Wireless Networks*, vol. 29, no. 3, pp. 1051–1058, 2023, doi: [10.1007/s11276-018-1880-3](https://doi.org/10.1007/s11276-018-1880-3).
- [12] A. Sorour and A. S. Atkins, "Big data challenge for monitoring quality in higher education institutions using business intelligence dashboards," *Journal of Electronic Science and Technology*, vol. 22, no. 1, p. 100233, 2024, doi: <https://doi.org/10.1016/j.jnlest.2024.100233>.
- [13] A. Sorour, A. S. Atkins, C. F. Stanier, and F. D. Alharbi, "The Role of Business Intelligence and Analytics in Higher Education Quality: A Proposed Architecture," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, 2020, pp. 1–6. doi: [10.1109/AECT47998.2020.9194157](https://doi.org/10.1109/AECT47998.2020.9194157).
- [14] J. George and Dr.Jeyakumar, "A comparative analysis of data integration and business intelligence tools with an emphasis on healthcare data," *International journal of engineering trends and technology*, vol. 68, no. 9, pp. 5–9, 2020, doi: [10.14445/22315381/ijett-v68i9p202](https://doi.org/10.14445/22315381/ijett-v68i9p202).
- [15] W. H. Inmon, *Building the data warehouse*. John wiley & sons, 2005.
- [16] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
- [17] Y.-S. Su and S.-Y. Wu, "Applying data mining techniques to explore user behaviors and watching video patterns in converged IT environments," *J Ambient Intell Humaniz Comput*, 2021, doi: [10.1007/s12652-020-02712-6](https://doi.org/10.1007/s12652-020-02712-6).
- [18] S. Lee, D. Hooshyar, H. Ji, K. Nam, and H. Lim, "Mining biometric data to predict programmer expertise and task difficulty," *Cluster Comput*, vol. 21, no. 1, pp. 1097–1107, 2018, doi: [10.1007/s10586-017-0746-2](https://doi.org/10.1007/s10586-017-0746-2).
- [19] C.-H. Chou, Y.-S. Su, C.-J. Hsu, K.-C. Lee, and P.-H. Han, "Design of desktop audiovisual entertainment system with deep learning and haptic sensations," *Symmetry (Basel)*, vol. 12, no. 10, p. 1718, 2020, doi: [10.3390/sym12101718](https://doi.org/10.3390/sym12101718).

- [20] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.139>.