

Analisis Tingkat Pengetahuan Mahasiswa tentang Pencegahan BBLR Menggunakan Pendekatan Data Science

Quratih Adawiyah
Institut Teknologi dan Kesehatan Ika Bina
Rantauprapat, Indonesia

quratihadawiyah29@gmail.com

*Penulis Korespondensi

Diajukan : 29/07/2026

Diterima : 30/07/2026

Dipublikasi : 01/08/2026

ABSTRAK

Berat Badan Lahir Rendah (BBLR) merupakan salah satu permasalahan kesehatan yang memerlukan upaya pencegahan melalui peningkatan pengetahuan masyarakat, termasuk mahasiswa. Penelitian ini bertujuan untuk menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR menggunakan pendekatan Data Science. Penelitian menggunakan metode kuantitatif dengan data yang diperoleh melalui kuesioner terhadap 100 mahasiswa. Analisis dilakukan menggunakan kerangka Cross-Industry Standard Process for Data Mining (CRISP-DM) yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Hasil penelitian menunjukkan bahwa rata-rata tingkat pengetahuan mahasiswa sebesar 83,60, yang termasuk dalam kategori baik. Proses analisis melalui pendekatan Data Science memberikan gambaran mengenai karakteristik, distribusi, dan pola tingkat pengetahuan mahasiswa sehingga mendukung evaluasi hasil edukasi kesehatan secara lebih sistematis. Penelitian ini menunjukkan bahwa pendekatan Data Science dapat dimanfaatkan sebagai salah satu alternatif dalam analisis data edukasi kesehatan untuk mendukung penyusunan program edukasi yang lebih berbasis data.

Kata Kunci: Data Science, BBLR, CRISP-DM, Pengetahuan Mahasiswa, Analisis Data

I. PENDAHULUAN

Berat Badan Lahir Rendah (BBLR) masih menjadi salah satu tantangan utama dalam bidang kesehatan ibu dan anak karena berkaitan dengan tingginya risiko gangguan kesehatan pada masa neonatal maupun tahap pertumbuhan berikutnya. Bayi yang lahir dengan berat kurang dari 2.500 gram memiliki kemungkinan lebih besar mengalami komplikasi seperti gangguan pernapasan, infeksi, keterlambatan pertumbuhan, hingga peningkatan risiko kematian pada periode awal kehidupan. Oleh karena itu, berbagai upaya promotif dan preventif terus dilakukan untuk menekan angka kejadian BBLR, salah satunya melalui peningkatan pengetahuan masyarakat mengenai faktor risiko dan strategi pencegahannya. Edukasi kesehatan menjadi langkah penting dalam membangun kesadaran sehingga masyarakat mampu menerapkan perilaku yang mendukung kesehatan ibu dan bayi sejak masa kehamilan.

Perkembangan teknologi informasi telah membawa perubahan pada cara data dimanfaatkan dalam berbagai bidang, termasuk kesehatan. Pendekatan Data Science memungkinkan data yang diperoleh dari suatu kegiatan atau penelitian diolah melalui serangkaian proses mulai dari pengumpulan data, *data preprocessing*, eksplorasi data, visualisasi, hingga interpretasi hasil. Pendekatan ini tidak hanya menghasilkan informasi yang bersifat deskriptif, tetapi juga mampu mengidentifikasi pola dan hubungan antarvariabel sehingga mendukung

pengambilan keputusan berbasis data. Pemanfaatan Data Science dalam bidang kesehatan terus berkembang karena mampu mendukung evaluasi program kesehatan, analisis perilaku masyarakat, serta penyusunan kebijakan yang lebih tepat sasaran (Kaur et al., 2022; Sarker, 2021).

Mahasiswa merupakan salah satu kelompok yang memiliki akses luas terhadap teknologi digital serta sumber informasi kesehatan. Kondisi tersebut menjadikan mahasiswa sebagai kelompok yang potensial untuk berperan dalam penyebaran informasi kesehatan di lingkungan masyarakat. Namun, kemudahan memperoleh informasi tidak selalu diikuti dengan tingkat pemahaman yang memadai. Perbedaan kemampuan dalam memahami, menyeleksi, dan menginterpretasikan informasi kesehatan dapat menyebabkan variasi tingkat pengetahuan antarindividu. Oleh sebab itu, pengukuran tingkat pengetahuan mahasiswa mengenai pencegahan BBLR menjadi penting untuk mengetahui sejauh mana informasi kesehatan telah dipahami dan menjadi dasar dalam penyusunan strategi edukasi yang lebih efektif.

Berbagai penelitian telah mengevaluasi efektivitas edukasi kesehatan dalam meningkatkan pengetahuan mengenai pencegahan BBLR. Penelitian (Naafi Hidayanti et al., 2024) menunjukkan bahwa pemberian video edukasi mampu meningkatkan pengetahuan keluarga mengenai pencegahan BBLR. Selanjutnya, (Rachmani et al., 2022) menemukan bahwa tingkat *digital health literacy* mahasiswa masih bervariasi sehingga diperlukan evaluasi terhadap tingkat pengetahuan kesehatan. Meskipun demikian, penelitian-penelitian tersebut masih berfokus pada pengukuran peningkatan pengetahuan menggunakan pendekatan statistik konvensional dan belum memanfaatkan pendekatan Data Science untuk mengeksplorasi karakteristik maupun pola data hasil edukasi kesehatan. Kondisi tersebut menunjukkan adanya peluang untuk menerapkan pendekatan Data Science dalam menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR dengan menerapkan pendekatan Data Science. Analisis dilakukan melalui tahapan pengumpulan data, *data preprocessing*, eksplorasi data, visualisasi, dan interpretasi hasil sehingga diperoleh gambaran mengenai tingkat pengetahuan mahasiswa serta pola-pola yang terdapat pada data penelitian. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan penerapan Data Science dalam bidang kesehatan, khususnya sebagai pendekatan analitis untuk mengevaluasi tingkat pengetahuan mahasiswa mengenai pencegahan BBLR sekaligus menjadi dasar dalam penyusunan program edukasi kesehatan yang lebih efektif.

II. STUDI LITERATUR

Berat Badan Lahir Rendah (BBLR) merupakan kondisi bayi yang lahir dengan berat badan kurang dari 2.500 gram tanpa memperhatikan usia kehamilan. BBLR menjadi salah satu indikator penting kesehatan ibu dan anak karena berkaitan dengan meningkatnya risiko kesakitan, kematian neonatal, serta gangguan pertumbuhan dan perkembangan. Kejadian BBLR dipengaruhi oleh berbagai faktor, seperti status gizi ibu, kondisi kesehatan selama kehamilan, usia kehamilan, dan faktor sosial ekonomi. Oleh karena itu, peningkatan pengetahuan mengenai faktor risiko dan upaya pencegahan BBLR menjadi salah satu strategi penting dalam menekan angka kejadiannya. Tingkat pengetahuan merupakan indikator yang umum digunakan untuk mengevaluasi keberhasilan program edukasi kesehatan. Mahasiswa sebagai kelompok yang memiliki akses luas terhadap berbagai sumber informasi diharapkan memiliki pemahaman yang baik mengenai pencegahan BBLR. Namun, perbedaan kemampuan dalam memahami dan mengevaluasi informasi menyebabkan tingkat pengetahuan setiap individu berbeda. Oleh karena itu, diperlukan pendekatan analisis yang mampu menggambarkan karakteristik data secara lebih mendalam sehingga hasil evaluasi tidak hanya menunjukkan nilai rata-rata, tetapi juga pola dan distribusi tingkat pengetahuan responden.

Data Science merupakan bidang multidisiplin yang menggabungkan ilmu komputer, statistika, dan analisis data untuk menghasilkan informasi yang mendukung pengambilan keputusan. Pendekatan ini meliputi tahapan pengumpulan data, *data preprocessing*, eksplorasi data (*exploratory data analysis*), analisis, visualisasi, dan interpretasi hasil. Melalui tahapan tersebut,

data yang semula berupa kumpulan angka dapat diolah menjadi informasi yang lebih bermakna dan mudah dipahami.

Dalam bidang kesehatan, Data Science dimanfaatkan untuk menganalisis data, mengidentifikasi pola, mengevaluasi program kesehatan, serta mendukung pengambilan keputusan berbasis data (data-driven decision making). Dibandingkan analisis deskriptif konvensional, pendekatan ini mampu memberikan hasil yang lebih komprehensif melalui eksplorasi hubungan antarvariabel dan penyajian visual yang memudahkan interpretasi hasil.

Pada penelitian ini, pendekatan Data Science digunakan untuk menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR berdasarkan data hasil kuesioner. Analisis dilakukan melalui tahapan data preprocessing, eksplorasi data, dan visualisasi untuk memperoleh gambaran karakteristik responden serta distribusi tingkat pengetahuan mahasiswa. Hasil analisis diharapkan dapat menjadi dasar dalam mengevaluasi efektivitas edukasi kesehatan dan mendukung penyusunan strategi edukasi yang lebih tepat sasaran.

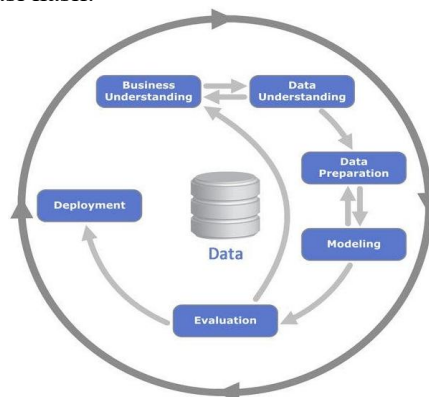
Berdasarkan penelitian oleh (Naafi Hidayanti et al., 2024)) menunjukkan bahwa pengaruh media edukasi terhadap pengetahuan keluarga mengenai pencegahan BBLR. Hasil penelitian menunjukkan bahwa pemberian edukasi mampu meningkatkan tingkat pengetahuan responden secara signifikan. Namun, analisis yang digunakan masih terbatas pada metode statistik konvensional dan belum menerapkan pendekatan Data Science untuk mengeksplorasi pola data secara lebih mendalam.

Penelitian lain oleh Lee et al. (2024) membahas penerapan Data Science berperan penting dalam pengolahan, analisis, dan visualisasi data kesehatan sehingga mendukung pengambilan keputusan berbasis data. Penelitian tersebut menunjukkan bahwa penerapan Data Science mampu menghasilkan informasi yang lebih komprehensif dibandingkan analisis deskriptif. Meskipun demikian, penelitian ini belum mengkaji analisis tingkat pengetahuan peserta pada kegiatan edukasi kesehatan.

Selanjutnya, penelitian oleh Sjamssuddin dan Anshari (2024) mengenai *digital health literacy* mahasiswa dan menemukan bahwa kemampuan mahasiswa dalam memahami informasi kesehatan masih bervariasi. Penelitian tersebut menekankan pentingnya evaluasi tingkat pengetahuan, namun belum memanfaatkan pendekatan Data Science dalam proses analisis data. Berdasarkan penelitian tersebut, penelitian ini menerapkan pendekatan Data Science untuk menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR sehingga diharapkan menghasilkan informasi yang lebih komprehensif.

III. METODE

Metode Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis Data Science untuk menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan Berat Badan Lahir Rendah (BBLR). Data yang digunakan merupakan data primer yang diperoleh melalui penyebaran kuesioner kepada mahasiswa yang mengikuti kegiatan edukasi kesehatan. Analisis dilakukan menggunakan tahapan Cross-Industry Standard Process for Data Mining (CRISP-DM) karena mampu memberikan alur analisis data yang sistematis, mulai dari identifikasi permasalahan hingga interpretasi hasil.



Gambar 1. Alur Penelitian Menggunakan CRISP-DM

Tahapan penelitian terdiri atas enam fase, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Pada tahap *Business Understanding*, ditentukan tujuan penelitian, yaitu menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan BBLR menggunakan pendekatan Data Science. Tahap *Data Understanding* dilakukan dengan mengumpulkan data hasil kuesioner serta mengidentifikasi karakteristik responden, seperti jenis kelamin, program studi, dan nilai jawaban.

Pada tahap *Data Preparation*, dilakukan proses *data preprocessing* yang meliputi pemeriksaan data, penghapusan data duplikat, penanganan data yang tidak lengkap, serta transformasi data agar siap dianalisis. Tahap *Modeling* dilakukan menggunakan analisis deskriptif yang didukung dengan eksplorasi data (*Exploratory Data Analysis/EDA*) dan visualisasi untuk mengetahui distribusi tingkat pengetahuan mahasiswa. Tahap *Evaluation* bertujuan menginterpretasikan hasil analisis serta mengevaluasi efektivitas penyampaian materi berdasarkan pola data yang diperoleh. Tahap terakhir, yaitu *Deployment*, dilakukan dengan menyajikan hasil analisis dalam bentuk grafik, tabel, dan rekomendasi yang dapat digunakan sebagai dasar pengambilan keputusan dalam penyusunan program edukasi kesehatan berikutnya.

Instrumen penelitian berupa kuesioner yang terdiri atas dua bagian, yaitu karakteristik responden dan pertanyaan mengenai pengetahuan pencegahan BBLR. Data yang telah melalui proses *preprocessing* dianalisis menggunakan perangkat lunak Python dengan bantuan pustaka Pandas untuk pengolahan data, NumPy untuk komputasi numerik, serta Matplotlib dan Seaborn untuk visualisasi data. Hasil analisis disajikan dalam bentuk tabel, grafik, dan interpretasi sehingga mampu memberikan gambaran mengenai tingkat pengetahuan mahasiswa secara lebih komprehensif.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja dalam menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan Berat Badan Lahir Rendah (BBLR). Penerapan CRISP-DM bertujuan agar proses analisis data dilakukan secara sistematis, mulai dari memahami permasalahan, menyiapkan data, melakukan eksplorasi data, hingga menghasilkan informasi yang dapat digunakan sebagai dasar pengambilan keputusan. Pada penelitian ini, tahapan yang diimplementasikan meliputi Data Understanding, Data Preparation, Exploratory Data Analysis (EDA), Evaluation, dan Deployment.

Pada tahap Data Understanding, dilakukan identifikasi terhadap karakteristik dataset yang diperoleh dari hasil kuesioner. Dataset terdiri atas 100 data responden dengan beberapa atribut, yaitu program studi, jenis kelamin, semester, serta skor pengetahuan mengenai pencegahan BBLR. Selain mengidentifikasi jumlah data dan jenis variabel, tahap ini juga bertujuan memastikan bahwa setiap atribut memiliki kualitas data yang memadai sebelum dilakukan proses analisis lebih lanjut.

Pada tahap ini dilakukan identifikasi terhadap jumlah data, jenis variabel, serta kelengkapan data yang diperoleh. Proses tersebut bertujuan untuk memastikan bahwa data yang dikumpulkan telah sesuai dengan kebutuhan penelitian dan siap diproses pada tahapan berikutnya.

Tahap berikutnya adalah Data Preparation, yaitu proses menyiapkan data agar memiliki kualitas yang baik sebelum dilakukan analisis. Tahapan ini merupakan bagian penting dalam Data Science karena kualitas data akan memengaruhi hasil analisis yang diperoleh. Data yang tidak lengkap, tidak konsisten, atau mengandung kesalahan dapat menghasilkan informasi yang kurang akurat.

Selain pemeriksaan missing value dan duplicate data, dilakukan proses transformasi data dengan mengubah jawaban kuesioner menjadi bentuk numerik menggunakan skema biner, yaitu skor 1 untuk jawaban benar dan 0 untuk jawaban salah. Selanjutnya dilakukan perhitungan skor total masing-masing responden sebagai variabel utama dalam analisis. Proses ini bertujuan menghasilkan dataset yang konsisten dan siap digunakan pada tahap eksplorasi data.

Tabel 1. Hasil Data Preparation

Tahapan	Hasil
Jumlah Data Awal	100
Missing Value	0
Data Duplikat	0
Data Valid	100

Berdasarkan Tabel 1, Hasil data preparation menunjukkan bahwa seluruh data dapat digunakan pada proses analisis karena tidak ditemukan data kosong maupun data ganda. Dengan demikian, data dinilai memiliki kualitas yang baik dan siap digunakan pada tahap eksplorasi data.

Setelah data dinyatakan bersih, dilakukan Exploratory Data Analysis (EDA) untuk memperoleh gambaran mengenai karakteristik data. EDA bertujuan mengeksplorasi distribusi data, melihat kecenderungan nilai, serta mengidentifikasi pola yang muncul pada tingkat pengetahuan mahasiswa.

Analisis dilakukan menggunakan statistik deskriptif yang meliputi nilai rata-rata (mean), median, nilai minimum, nilai maksimum, dan standar deviasi. Selain itu, data juga disajikan dalam bentuk visualisasi seperti diagram batang, histogram, atau boxplot sehingga pola distribusi data lebih mudah dipahami.

Tabel 2. Statistik Deskriptif Tingkat Pengetahuan Mahasiswa (n = 100)

Statistik	Nilai
Jumlah Responden	100
Mean	83,60
Median	84,00
Minimum	58,00
Maksimum	98,00
Standar Deviasi	7,45

Berdasarkan Tabel 2, diperoleh nilai rata-rata tingkat pengetahuan mahasiswa sebesar 83,60, yang menunjukkan bahwa secara umum responden memiliki pemahaman yang baik mengenai pencegahan BBLR. Nilai median sebesar 84,00 yang hampir sama dengan nilai rata-rata menunjukkan bahwa distribusi data relatif seimbang dan tidak dipengaruhi oleh nilai yang terlalu ekstrem. Nilai minimum sebesar 58,00 menunjukkan masih terdapat beberapa mahasiswa dengan tingkat pengetahuan yang perlu ditingkatkan, sedangkan nilai maksimum 98,00 menunjukkan bahwa sebagian responden telah menguasai materi dengan sangat baik. Nilai median yang hampir sama dengan nilai rata-rata mengindikasikan bahwa distribusi data relatif simetris dan tidak dipengaruhi secara signifikan oleh nilai ekstrem (*outlier*). Hal ini menunjukkan bahwa sebagian besar responden memiliki tingkat pengetahuan yang berada pada kisaran nilai rata-rata.

Standar deviasi sebesar 7,45 mengindikasikan bahwa penyebaran nilai responden relatif rendah sehingga tingkat pengetahuan mahasiswa cenderung homogen. Hal ini menunjukkan bahwa sebagian besar mahasiswa memperoleh pemahaman yang hampir seragam setelah mengikuti kegiatan edukasi. Hasil EDA ini memberikan gambaran awal bahwa materi yang disampaikan dapat dipahami oleh mayoritas responden, meskipun masih terdapat sebagian kecil mahasiswa yang memerlukan penguatan materi.

Hasil penelitian ini sejalan dengan penelitian Hidayanti dan Aryani (2024) yang melaporkan bahwa edukasi kesehatan mampu meningkatkan pengetahuan mengenai pencegahan BBLR. Demikian pula, penelitian (Haikal et al., 2023) menunjukkan bahwa pendidikan kesehatan memberikan pengaruh positif terhadap pemahaman wanita usia subur mengenai faktor risiko BBLR. Perbedaannya, penelitian ini tidak hanya menyajikan hasil statistik deskriptif, tetapi juga menerapkan tahapan Data Science melalui kerangka CRISP-DM sehingga proses pengolahan dan eksplorasi data dilakukan secara lebih sistematis.

Berdasarkan hasil visualisasi histogram, distribusi skor pengetahuan mahasiswa cenderung terkonsentrasi pada rentang nilai tinggi (80–90). Sebagian besar responden memiliki tingkat pengetahuan yang baik dengan sebaran data yang relatif homogen. Temuan ini diperkuat oleh nilai standar deviasi sebesar 7,45, yang menunjukkan variasi skor antarresponden relatif rendah.

Tabel 3. Distribusi Kategori Tingkat Pengetahuan

Kategori	Rentang Skor	Jumlah	Persentase
Tinggi	80-100	68	68%
Sedang	60-79	26	26%
Rendah	0-59	6	6%
Total	58,00	100	100%

Berdasarkan Tabel 3, pembaca dapat langsung melihat proporsi mahasiswa pada setiap kategori tingkat pengetahuan, bukan hanya nilai rata-ratanya. Ini juga sangat selaras dengan pendekatan Exploratory Data Analysis (EDA) dalam Data Science.

V. KESIMPULAN

Penelitian ini menerapkan pendekatan Data Science menggunakan metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM) untuk menganalisis tingkat pengetahuan mahasiswa mengenai pencegahan Berat Badan Lahir Rendah (BBLR). Hasil analisis terhadap 100 responden menunjukkan bahwa tingkat pengetahuan mahasiswa berada pada kategori baik dengan nilai rata-rata sebesar 83,60. Pendekatan Data Science memberikan gambaran mengenai distribusi dan karakteristik tingkat pengetahuan mahasiswa sehingga mendukung proses interpretasi data hasil edukasi kesehatan. Temuan penelitian ini menunjukkan bahwa pendekatan Data Science berpotensi mendukung penyusunan strategi edukasi kesehatan berbasis data sebagai bahan pertimbangan dalam pengambilan keputusan. Penelitian ini masih terbatas pada analisis deskriptif terhadap data hasil kuesioner sehingga belum mengidentifikasi faktor-faktor yang memengaruhi tingkat pengetahuan mahasiswa maupun melakukan pemodelan prediktif. Oleh karena itu, penelitian selanjutnya disarankan untuk melibatkan jumlah responden yang lebih besar serta menerapkan metode analisis Data Science yang lebih lanjut, seperti klasifikasi atau *clustering*, guna memperoleh informasi yang lebih mendalam.

VI. REFERENSI

- Haikal, H., Rachmani, E., Setyo Nugroho, B. Y., Iqbal, M., Prasetya, J., & Yudi Nugroho, S. (2023a). Digital Health Literacy Competencies of Students in Faculty of Health Science. *Disease Prevention and Public Health Journal*, 17(1), 39–46. <https://doi.org/10.12928/dpphj.v17i1.6448>
- Kaur, P., Singh, M., & Joshi, A. (2022). Data science approaches in healthcare analytics: A review of current applications and future perspectives. *Healthcare Analytics*, 2, 100048.
- Naafi Hidayanti, H., Aryani, T., & Kesehatan Hermina, I. (2024). PENGARUH VIDEO EDUKASI TERHADAP PENGETAHUAN KELUARGA PADA BAYI BERAT

- BADAN LAHIR RENDAH (BBLR) DI RUANG NICU. In *Indonesian Journal of Health Science* (Vol. 4, Number 3).
- Nabilla, A., Monayo, E. R., Hiola, D. S., Gorontalo, U. N., & Id, D. A. (2026). TINGKAT PENGETAHUAN IBU PRIMIGRAVIDA DALAM PENGASUHAN BAYI BARU LAHIR. *Journal of Telenursing (JOTING)*, 8(1). <https://doi.org/10.31539/jthtxy69>
- Rachmani, E., Haikal, H., & Rimawati, E. (2022). Development and validation of digital health literacy competencies for citizens (DHLC), an instrument for measuring digital health literacy in the community. *Computer Methods and Programs in Biomedicine Update*, 2. <https://doi.org/10.1016/j.cmpbup.2022.100082>
- Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in medicine. *New England Journal of Medicine*, 386, 1377–1380.
- Sarker, I. H. (2021a). Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2, 377.
- Situmorang, H., Dewi Br Siregar, S., Sitorus, N. Y., Tinggi, S., & Flora, I. K. (n.d.). *The Impact of Health Education Using Audio-Visual Media on the Family's Role in LBW Care at Mutia Sari Hospital, Duri, Riau*.