



Ensemble Tree untuk Memprediksi Level Resiko Maternal Mortality di Bangladesh

Aviolla Terza Damaliana¹, Trimono², Dwi Arman Prasetya³

^{1,2,3} Program Studi Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur

¹aviolla.terza.sada@upnjatim.ac.id

³arman.prasetya.sada@upnjatim.ac.id

Corresponding author email: trimono.stat@upnjatim.ac.id

Abstract: *Pregnancy complications are common in developing countries. According to WHO, more than 99% of women die every day in low and middle-income countries due to pregnancy and birth complications. One country that has an average population of low and middle income is Bangladesh. One way to monitor the health of the mother and fetus is to know the risk factors for maternal mortality which are then used to predict the level of risk. Based on this, the level of risk of maternal mortality needs to be investigated further. In this study using the Machine Learning approach with the ensemble tree method to predict the level of risk. The methods used include Random Forest, Bagging Decision Tree, Adaptive Booster, GBM, and XGBoost. To get the best method, the model is evaluated using accuracy, precision and recall. Furthermore, the four methods are compared, the result is that the Random Forest method has the highest accuracy value of 86.03% with the most important attribute that influences predictions, namely Blood Sugar.*

Keywords: *Ensemble Tree, Machine Learning, Maternal Mortality*

Abstrak: *Komplikasi kehamilan banyak terjadi di negara yang berkembang. Menurut WHO, Lebih dari 99% wanita meninggal setiap harinya pada negara berpenghasilan rendah dan menengah dikarenakan adanya kehamilan dan komplikasi kelahiran. Salah satu negara yang memiliki rata-rata penduduk berpenghasilan rendah dan menengah adalah Bangladesh. Salah satu cara untuk memantau Kesehatan ibu dan janin adalah dengan mengetahui faktor resiko terjadinya maternal mortality yang selanjutnya digunakan untuk memprediksi tingkat risikonya. Berdasarkan hal itu tingkat resiko terjadinya Maternal Mortality perlu diteliti lebih lanjut. Pada penelitian ini menggunakan pendekatan Machine Learning dengan metode ensemble Tree untuk memprediksi tingkat risikonya. Metode yang digunakan antara lain Random Forest, Bagging Decision Tree, Adaptive Booster, GBM, dan XGBoost. Untuk mendapatkan metode terbaik, model dievaluasi menggunakan *accuracy*, *precision*, dan *recall*. Selanjutnya ke-empat metode tersebut dibandingkan, hasilnya adalah metode *Random Forest* memiliki nilai *accuracy* tertinggi yaitu sebesar 86,03% dengan atribut paling penting yang mempengaruhi prediksi yaitu Gula Darah.*

Kata kunci: *Ensemble Tree, Machine Learning, Maternal Mortality*

I. PENDAHULUAN

Komplikasi kehamilan banyak terjadi di negara yang berkembang. Hal ini dikarenakan banyak ibu hamil yang tidak memiliki informasi yang cukup mengenai perawatan selama kehamilan dan pasca kehamilan. [1]. Menurut WHO, Lebih dari 99% wanita meninggal setiap harinya pada negara berpenghasilan rendah dan menengah dikarenakan adanya kehamilan dan komplikasi kelahiran [2]. Salah satu negara yang memiliki rata-rata penduduk berpenghasilan rendah dan menengah adalah Bangladesh. Indikator Maternal Mortality sangat sensitive seperti halnya Indikator Kesehatan lainnya di Bangladesh, hal ini dikarenakan tingkat komplikasi selama hamil dan melahirkan 80 hingga 600 kali lebih tinggi dibandingkan wanita di negara berkembang lainnya [3]. Oleh karena itu kesehatan ibu dan janin harus dipantau untuk memastikan keselamatan selama hamil hingga pasca kelahiran [4].

Salah satu cara untuk memantau Kesehatan ibu dan janin adalah dengan mengetahui faktor resiko terjadinya maternal mortality. Pada penelitian terdahulu yang dilakukan oleh [5], selama kehamilan faktor resiko dari maternal mortality nya diukur berdasarkan Usia, *Body Mass Index* (BMI), Oksigen Darah, Tekanan Darah, Suhu Tubuh, dan aktivitas fisik. Sedangkan penelitian yang dilakukan oleh [6], faktor risikonya diukur berdasarkan Maternal ECG, Keputihan pada awal trisemester, Mual pada awal trisemester, kontraksi pada trisemester ke-tiga, Denyut jantung janin, protein janin abnormal, electrical uterine activity, dan aktivitas Gerakan janin. Faktor-faktor resiko terjadinya maternal mortality ini sangat perlu untuk diidentifikasi dan dimodifikasi untuk menentukan strategi yang efektif untuk memastikan keselamatan ibu hamil selama kehamilan hingga pasca kelahiran. Setelah



mengetahui faktor-faktornya, diperlukan suatu pendekatan untuk memprediksi tingkat resiko maternal mortality-nya. Adapun pendekatan yang saat ini populer untuk memprediksi adalah pendekatan machine learning. Pada penggunaan machine learning ini diharapkan maternal mortality dapat dideteksi sedini mungkin sehingga akan ditemukan strategi yang lebih baik untuk menekan maternal mortality di Negara Bangladesh tersebut.

Pada penelitian ini menggunakan data yang dikumpulkan oleh [4] dimana faktor-faktor resiko terjadinya Maternal Mortality antara lain Usia, Tekanan Darah, Gula Darah, Suhu Tubuh, dan Denyut Jantung Janin. Data yang dikumpulkan oleh [4] ini diambil langsung dari daerah pedesaan di Bangladesh sehingga faktor-faktor resiko lainnya tidak dapat diukur. Pada penelitian yang dilakukan oleh [4] menggunakan pendekatan machine learning untuk memprediksi tingkat resiko Maternal Mortality-nya. Metode yang dibandingkan adalah Decision Tree, Random Forest, Support Vector Machine, Sequential Minimal Optimizasion, Logistik Regression, Naïve Bayes, dan Logistic Model Tree. Dikarenakan penelitian terdahulu yang memiliki akurasi tertinggi adalah Decision Tree maka pada penelitian ini akan menggunakan metode ensemble trees untuk menambah akurasinya [7]. Penambahan nilai akurasi diperlukan karena Maternal Mortality merupakan masalah kesehatan yang serius sehingga harus terus dilakukan penelitian agar didapatkan akurasi yang mendekati 100%.

II. METODE PENELITIAN

Metode ensemble berdasarkan Decision Trees yang akan digunakan pada penelitian ini antara lain *Random Forest* [8], *Bagging Decision Tree* [9] *Adaptive Booster* [10], *Gradient Boost* [11], dan *XGBoost* [12]. Metode-metode tersebut kemudian akan dibandingkan nilai akurasinya sehingga akan diketahui metode mana yang akan menghasilkan nilai akurasi tertinggi untuk memprediksi tingkat resiko Maternal Mortality di Bangladesh.

2.1. *Random Forest*

Random Forest merupakan metode gabungan dari classification tree yang saling independen dan hasil prediksi diperoleh dari proses voting dari beberapa clasification tree yang terbentuk. Saat memprediksi, metode random forest ini menggunakan proses pengacakan untuk membentuk classification tree pada *data training* dan *data testing* sehingga akan menghasilkan kumpulan classification tree dengan ukuran dan bentuk yang berbeda-beda. Dampak yang diharapkan adalah antar classification memiliki korelasi yang kecil sehingga akan meningkatkan akurasi prediksi Random Forest [8].

2.2. *Bagging Decision Tree*

Metode *Bagging Decision Tree* merupakan algoritma ensemble machine learning yang juga dikenal sebagai *bootstrap aggregating*. Metode Bagging ini diciptakan oleh [9] karena pada metode Decision Tree memiliki model yang bervarians tinggi sehingga biasanya terjadi overfitting. Metode ini memiliki dua langkah pengerjaan yaitu *Bootstrapping* dan *Aggregation*. Pada tahap *bootstrapping* akan dibangun banyak model *tree* pada sampel bootsrap dan kemudian secara acak memilih data berulang dari data training dan membuat beberapa subset. Kemudian pada tahap *aggregation*, setiap model yang berasal dari data testing dipilih berdasarkan mekanisme *majority voting*.

2.3. *Adaptive Booster*

Metode *Adaptive Booster* merupakan salah satu ensemble tree yang algoritmanya hampir sama dengan metode *Random Forest* karena keduanya menghitung prediksi yang dibuat setiap decision kemudian dipilih berdasarkan *majority voting*-nya. Adapun perbedaannya adalah menggabungkan model klasifikasi berkinerja buruk sehingga akan mendapatkan model yang akurasinya tinggi. Konsep dasarnya adalah mengatur bobot pengklasifikasi dan melatih data training setiap iterasi sedemikian rupa hingga prediksi yang dihasilkan akurat dibandingkan dengan hanya menggunakan metode decision tree [10]

2.4. *Gradient Boost (GB)*

Metode *Gradient Boost* atau dapat disingkat sebagai GB merupakan salah satu metode ensemble machine learning yang memiliki kemampuan untuk meningkatkan akurasi dari prediksi model. Hal ini

karena metode ini meminimalkan kesalahan prediksi secara keseluruhan. Metode ini mampu mencegah decision tree mengalami terjadinya *overfitting* yang akan mempengaruhi hasil prediksinya. Konsep dari metode ini adalah menyesuaikan model data training yang lemah dengan gradient dari *loss function* dan meningkatkan performa *trees* yang dihasilkan dengan parameter yang mewakili variabel split yang dipasang pada setiap *node* terminal *tree*-nya [11].

2.5. XGBoost

Metode *XGBoost* atau dikenal sebagai *Xtreme Gradient Boosting* merupakan metode machine learning kombinasi dari metode *Adaptive booster* dan *Gradient Boost*. Model dibangun dengan membuat model baru untuk memprediksi *error* dari model sebelumnya dengan menggunakan *gradient* untuk memperkecil *error*. Metode ini lebih efisien daripada GB karena mampu menyelesaikan berbagai fungsi termasuk klasifikasi [12].

2.6. Performance Measure

Performance measure digunakan untuk mengevaluasi kualitas dari model klasifikasi dengan menggunakan *Confusion Matriks*. Dalam *Confusion Matriks* mengandung nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang kemudian dicari nilai *accuracy*, *precision*, dan *recall* dengan rumus sebagai berikut.

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$recall = \frac{TP}{(TP + FN)} \quad (3)$$

III. PENGUMPULAN DATA DAN LANGKAH PENELITIAN

Penelitian ini menggunakan dataset yang dikumpulkan oleh [4]. Dataset ini diperoleh dari UCI Machine Learning Repository. Dataset ini adalah data untuk memprediksi tingkat resiko Maternal Mortality pada ibu hamil di daerah pedesaan di Negara Bangladesh yang memiliki tujuh atribut antara lain Usia, Tekanan Darah Systolic, Tekanan Darah Diastolic, Gula Darah, Suhu Tubuh, Detak Jantung Janin, dan Level Resiko terjadinya Maternal Mortality. Jumlah pengamatan yang digunakan ada sebanyak 1014 responden dari berbagai klinik di Negara Bangladesh. Berikut adalah penjelasan lebih detail mengenai dataset yang digunakan.

Tabel 1. Dataset Resiko Maternal Mortality

Atribut	Nama Atribut	Jenis Data Atribut
y	Level Resiko Maternal Mortality dimana terdiri dari 3 kategori yaitu Level Tinggi, Level Sedang, dan Level Rendah	Kategorik
x_1	Usia	Numerik
x_2	Tekanan Darah Systolic (mmHg)	Numerik
x_3	Tekanan Darah Diastolic (mmHg)	Numerik
x_4	Gula Darah (milligram/desiliter)	Numerik
x_5	Suhu Tubuh (Fahrenheit)	Numerik
x_6	Detak Jantung Janin	Numerik

Adapun langkah penelitian menggunakan pendekatan *machine learning* untuk memprediksi tingkat resiko Maternal Mortality di Bangladesh. Langkah pertama yang dilakukan adalah dengan melakukan eksplorasi data untuk melihat deskripsi permasalahan pada dataset ini. Langkah kedua adalah melakukan *pre-processing* data. Langkah ketiga adalah mencari nilai *accuracy*, *precision*, dan *recall* pada masing-masing metode *ensemble tree*-nya yaitu *Random Forest*, *Bagging Decision Tree*, *Adaptive Booster*, *Gradient Boost*, dan *XGBoost* yang kemudian dibandingkan nilai *performance*-nya sehingga memberikan kesimpulan hasil analisis.

IV. HASIL DAN PEMBAHASAN

Prediksi tingkat resiko maternal mortality di daerah pedesan di Negara Bangladesh menggunakan pendekatan Machine Learning, sehingga diperlukan Langkah-langkah sebagai berikut.

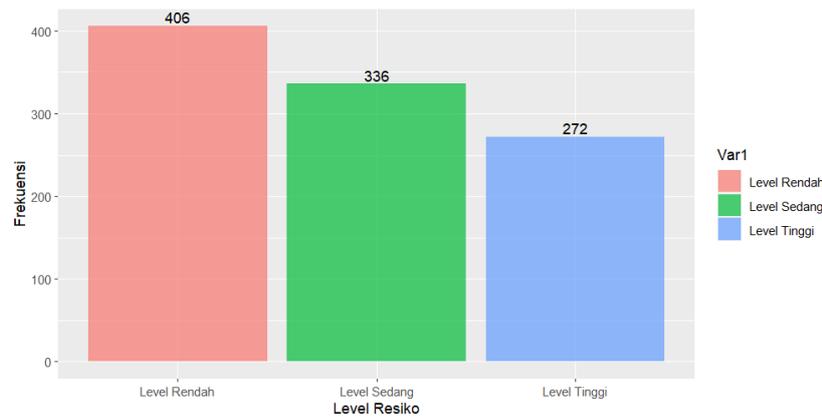
4.1. Eksplorasi Data

Eksplorasi Data digunakan untuk melihat permasalahan dari dataset yang akan diteliti. Berikut adalah eksplorasi dataset hasil kesehatan ibu hamil di Bangladesh dari 1014 pengamatan.

Tabel 2. Statistika Deskripsi Dataset Kesehatan Maternal

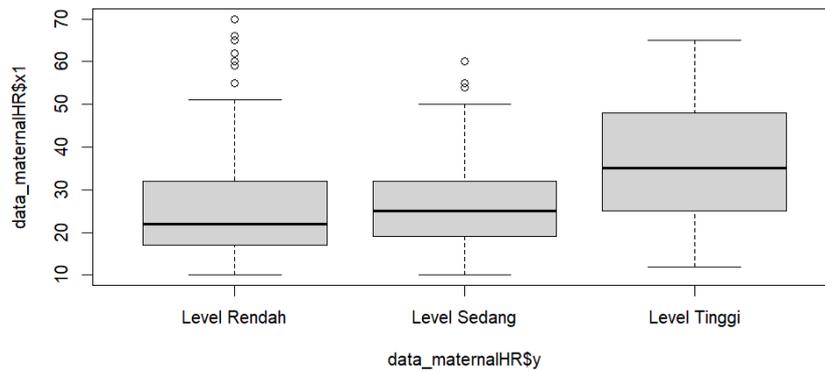
No	Feature	Minimum	Mean	Maksimum
1.	Usia	10,00	29,87	70,00
2.	Tekanan Darah Systolic (mmHg)	70,00	98,67	103,00
3.	Tekanan Darah Diastolic (mmHg)	49,00	76,46	100,00
4.	Gula Darah (milligram/desiliter)	60,00	87,26	190,00
5.	Suhu Tubuh (Fahrenheit)	98,00	98,67	103,00
6.	Detak Jantung Janin (kali/menit)	7,00	74,30	90,00

Tabel 2 menunjukkan bahwa usia Wanita hamil di pedesan Bangladesh sudah ideal karena rata-rata usia saat hamil yaitu 29 tahun, tetapi masih ada juga yang hamil di usia 10 tahun dan 70 tahun dimana hal tersebut sangatlah berisiko tinggi untuk keselamatan ibu dan janinnya. Selain itu rata-rata tekanan darah systolic dan diastolic yang dimiliki Wanita hamil menunjukkan rata-rata tekanan darahnya 98,67/76 mmHg yang artinya tekanan darah tersebut normal. Untuk rata-rata Gula Darah dan Suhu Tubuh Wanita hamil juga dapat dikatakan normal yaitu sebesar 87,26 miligram/desiliter dan 98,67 Fahrenheit. Namun keadaan normal pada detak jantung janin Wanita hamil yaitu berkisar antara 120 – 160 x/menit tidak terjadi karena rata-rata Wanita hamil tersebut memiliki detak jantung janin sebesar 74,30 x/menit . Sehingga janin di dalam Wanita hamil memiliki resiko tinggi untuk tidak terselamatkan.



Gambar 1. Level Resiko Maternal Mortality

Adapun jika dilihat dari Gambar 1 menunjukkan bahwa ada 272 wanita hamil yang memiliki level maternal mortality yang tinggi dan 336 wanita hamil yang memiliki level maternal mortality yang sedang, hal ini berarti hanya 40% Wanita hamil yang memiliki level resiko rendah untuk terjadinya maternal mortality.



Gambar 2. Boxplot Level Resiko Maternal Mortality Berdasarkan Usia

Gambar 2 menunjukkan bahwa pada kategori level rendah dan sedang, distribusi rata-rata usia Wanita hamil pada usia 20 tahunan. Sedangkan untuk kategori tinggi, distribusi rata-rata usia Wanita hamil diatas 30 tahunan. Kondisi tersebut sesuai dengan level resiko usia wanita untuk hamil yang telah ditetapkan WHO. Selain itu pada Gambar 2 box plot level rendah dan level sedang memiliki data outlier, sehingga agar penelitian ini menghasilkan akurasi prediksi yang lebih tinggi maka data outlier tersebut dihilangkan dalam dataset.

4.2. Pre-Processing Data

Langkah pertama dalam tahapan pre-processing data adalah mengetahui apakah terjadi *missing value* pada dataset Kesehatan Maternal di pedesaan di Negara Bangladesh. Setelah dilakukan pengecekan missing value diketahui bahwa dataset ini tidak mengalami missing value. Langkah kedua adalah dengan menghilangkan data outlier pada feature atribut usia, sehingga sampel pengamatannya menjadi 893 wanita hamil. Dikarenakan jenis atribut yang kategorik hanya satu yaitu atribut targetnya maka atribut target yaitu level resiko maternal mortality diubah dari data teks ke data multinomial. Data dari 893 wanita hamil tersebut dibagi menjadi 80% data training dan 20% data testing. Sehingga jumlah data training dan data testingnya sebanyak 714 dan 179 wanita hamil.

4.3. Model Performance

Setelah dilakukan pre-processing data, Langkah selanjutnya menghitung nilai *accuracy*, *precision*, dan *recall* pada masing-masing metode *ensemble tree*-nya yaitu *Random Forest*, *Bagging Decision Tree*, *Adaptive Booster*, *Gradient Boost*, dan *XGBoost*. Hasil yang didapat adalah sebagai berikut.

Tabel 3. Accuracy, Precision, dan Recall Metode Ensemble Tree

No	Metode	Accuracy (%)	Precision (%)	Recall (%)
1.	<i>Random Forest</i>	86,03	86,00	85,78
2.	<i>Bagging Decision Tree</i>	84,90	84,98	84,63
3.	<i>Adaptive Booster</i>	84,36	84,00	83,85
4.	<i>Gradient Boost</i>	77,09	77,70	76,53
5.	<i>XGBoost</i>	72,62	75,97	72,64

Tabel 3 menunjukkan bahwa nilai *accuracy*, *precision*, dan *recall* tertinggi adalah metode *Random Forest* yaitu sebesar 86,03% , 86%, dan 85,78% yang disusul oleh metode *Bagging Decision Tree*, *Adaptive Booster*, *Gradient Boost*, dan *XGBoost*. Sehingga berdasarkan Tabel 2 menunjukkan bahwa Metode *Random Forest* terpilih untuk memprediksi tingkat resiko terjadinya Maternal Mortality di daerah pedesaan di Negara Bangladesh.

Berikut adalah hasil prediksi tingkat resiko terjadinya Maternal Mortality menggunakan 5 pengamatan awal pada *data testing*.

Tabel 4. Hasil Prediksi Metode *Random Forest*

No	x_1	x_2	x_3	x_4	x_5	x_6	Hasil Prediksi
1.	42	130	80	180	98	70	Level Tinggi
2.	40	140	100	180	98	90	Level Tinggi
3.	21	90	65	75	98	76	Level Rendah



4.	21	120	80	75	98	76	Level Rendah
5.	26	85	60	61	102	60	Level Sedang

Berdasarkan Tabel 4 diketahui bahwa Level Tinggi diprediksi oleh wanita hamil pada usia 42 tahun dengan tekanan darah 130/80 mmHg, Gula darah sebesar 80 miligram/desiliter, suhu tubuh sebesar 98 Fahrenheit, dan detak jantung janin sebesar 70 kali/Menit. Level Rendah diprediksi oleh wanita hamil pada usia 21 tahun dengan tekanan darah 120/80 mmHg, Gula darah sebesar 75 miligram/desiliter, suhu tubuh sebesar 98 Fahrenheit, dan detak jantung janin sebesar 76 kali/Menit. Level Sedang diprediksi oleh wanita hamil pada usia 26 tahun dengan tekanan darah 85/60 mmHg, Gula darah sebesar 61 miligram/desiliter, suhu tubuh sebesar 102 Fahrenheit, dan detak jantung janin sebesar 60 kali/Menit.

Selanjutnya dari Metode *Random Forest* kita dapat melihat atribut feature penting yang mempengaruhi prediksi tingkat resiko terjadinya Maternal Mortality yaitu sebagai berikut.

Tabel 5. Atribut Feature Penting

Atribut	Mean Decrease Gini
x_1	77,68
x_2	80,22
x_3	44,62
x_4	131,87
x_5	28,18
x_6	38,49

Berdasarkan Tabel 5 diketahui bahwa urutan atribut yang paling mempengaruhi prediksi terjadinya *Maternal Mortality* di daerah pedesaan di Negara Bangladesh antara lain Gula darah, Tekanan Darah *Systolic*, Usia, Tekanan Darah *Diastolic*, Detak Jantung Janin, dan yang terakhir adalah Suhu Tubuh wanita hamil.

V. KESIMPULAN

Berdasarkan hasil penelitian yang sudah kita dapatkan untuk memprediksi tingkat resiko terjadinya *Maternal Mortality* di daerah pedesaan di Negara Bangladesh disimpulkan bahwa metode *ensemble trees* yang menghasilkan nilai akurasi tertinggi adalah metode *Random Forest* dengan atribut yang paling mempengaruhi hasil prediksi adalah atribut Gula Darah pada wanita hamil. Metode ini masih menghasilkan nilai akurasi dibawah 90%, sehingga untuk penelitian selanjutnya menggunakan dataset ini diharapkan mampu menambah nilai akurasi prediksi hingga mendekati 100%.

UCAPAN TERIMA KASIH

Terimakasih saya ucapkan kepada [4] karena telah menyumbangkan dataset Kesehatan Maternal di daerah pedesaan di Negara Bangladesh kepada UCI Machine Learning Dataset sehingga saya bisa melakukan penelitian untuk memprediksi tingkat resiko terjadinya Maternal Mortality-nya.

REFERENSI

1. M. Ahmed, M. Kashem, M. Rahman and S. Khatun, "Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)," in *In InECCE2019 (pp. 357-365)*, Springer, Singapore, 2020.
2. A. S. Shakil and M. R. Rashedur, "Fuzzy predictive model for estimating risk level of maternal mortality while child birth," in *International Conference Of Intelligent System (IS)*, Bangladesh, 2018.
3. C. Shena and J. B. Williamson, "Maternal mortality, women's status, and economic dependency in less developed countries: a cross-nasional analysis," *Social Science & Medicine*, vol. 49, pp. 197-214, 1999.
4. M. Ahmed and M. A. Kashem, "IoT Based Risk Level Prediction Model For Maternal Health Care In The Context of Bangladesh," in *2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, 2020.
5. Mustafa, Reem, S. Ahmed, A. Gupta and R. C. Venuto, "A comprehensive review of hypertension in pregnancy," *Journal of pregnancy*, p. 2012, 2012.
6. V. Santhi, K. Ramya, A. Tarana and G. Vinitha, "IoT based wearble health monitoring system for preganat ladies using cc3200," *International Journal of Advanced Research Methodology in Engineering & Technology 1*, vol. 3, pp. 56-60, 2017.



7. P. P.K., S. S. and A. D. L., "An Enhanced Decision Tree Ensemble Technique For Obesity Prediction," *International Journal of Scientific & Technology Research*, vol. 9, no. 02, pp. 1770 - 1775, 2020.
8. Y. Freund and E. Robert, "Experiments with a New Boosting Algorithm," in *Procidings of ICML'96*, 1996.
9. J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189 - 1232, 2001.
10. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2016.
11. P. Geurts, D. Ernst and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3 - 42, 2006.
12. L. Breiman, "Random Forest," *Machine Learning*, vol. 45, no. 1, pp. 5 - 32, 2001.