PREDIKSI STUNTING PADA BALITA DENGAN ALGORITMA *RANDOM FOREST*

Aditya Yudha Perdana[1], Roswan Latuconsina[2], Ashri Dinimaharawati[3]

[1,2,3] Universitas Telkom, Bandung

adityayudha@student.telkomuniversity.ac.id[1], roswan@telkomuniversity.ac.id[2], ashridini@telkomuniversity.ac.id[3]

ABSTRACT

Stunting is a condition where children experience growth failure due to malnutrition for a long time, this condition can be clearly seen from the height of the child which looks different from his peers. In this study, the author discusses stunting data taken from January to October 2020 in Pitu District, Ngawi Regency. The study was conducted using a random forest algorithm to predict stunting in toddlers and developing a system in the form of website and android-based software that is able to measure stunting in toddlers. The results obtained are learning from the data obtained and training is carried out so that the machine learning model is able to work and predict results for different data cases with the same parameters. This study was able to produce an average accuracy value of 97.87% using 10-fold cross validation.

Keywords: Random forest, stunting, machine learning

ABSTRAK

Stunting adalah suatu kondisi dimana anak mengalami gagal pertumbuhan akibat dari kekurangan gizi dalam waktu yang cukup lama, kondisi ini dapat terlihat jelas dari tinggi badan anak yang terlihat berbeda dari teman-teman seusianya. Dalam sstudi ini penulis membahas data *stunting* yang diambil pada bulan Januari hingga Oktober 2020 di Kecamatan Pitu, Kabupaten Ngawi. Penelitian dilakukan dengan menggunakan algoritma *random forest* untuk memprediksi keadaan *stunting* pada balita dan melakukan pengembangan sebuah sistem berupa perangkat lunak yang berbasis website dan android yang mampu melakukan pengukuran keadaan *Stunting* pada balita. Hasil yang didapat merupakan pembelajaran dari data yang diperoleh dan dilakukan pelatihan sehingga model *machine learning* mampu bekerja dan memprediksi hasil untuk kasus data yang berbeda dengan parameter yang sama. Penelitian ini mampu menghasilkan nilai akurasi rata-rata sebesar 97.87% dengan menggunakan 10-*fold cross validation.*

Kata Kunci: *Random forest, stunting, machine learnin*

1.Pendahuluan

Kemajuan teknologi informasi yang berkembang pesat di segala bidang kehidupan berbanding lurus dengan data yang dihasilkan. Mulai dari bidang industri, kesehatan dan berbagai bidang lainya. Dengan penerapan teknologi informasi di dunia kesehatan dan medis dapat menghasilkan data yang berlimpah. Data-data tersebut dapat berupa data tentang suatu penyakit maupun suatu kondisi medis tertentu. Seperti halnya pada puskesmas Pitu yang setiap bulan melakukan kegiatan posyandu guna memberikan kemudahan kepada masyarakat khususnya ibu dan balita untuk memperoleh pelayanan kesehatan dasar, disamping itu posyandu juga merupakan kegiatan untuk memonitoring dan pengambilan data untuk mengetahui ambang status gizi pada anak.

Dikutip dari website resmi Kementerian Kesehatan Repulik Indonesia di tahun 2018 persentase stunting di Indonesia berada pada angka 30,8% yang dimana persentase Indonesia berada jauh diatas negara-negara ASEAN lain yang berkisar antara 4%-17%. Sedangkan di Kabupaten Ngawi sendiri pada tahun 2019 dikutip dari portal suara.ngawikab.go.id persentase stunting berada pada angka 25,51% dimana angka tersebut jauh lebih tinggi dibanding Provinsi Jawa Timur itu sendiri yang berada pada angka 25%. Masalah kekurangan gizi ini antara lain disebabkan karena konsumsi yang tidak adekuat dipandang sebagai suatu permasalahan ekologis yang tidak saja disebabkan oleh ketidakcukupan ketersediaan pangan dan zat-zat gizi tertentu tetapi juga dipengaruhi oleh kemiskinan, sanitasi lingkungan yang kurang baik dan ketidaktahuan tentang gizi[1]. Puskesmas Pitu melakukan pengambilan dan penginputan data secara manual menggunakan Microsoft excel. Dimana ukuran yang dihasilkan

menjadi besar dan proses komputasi menjadi lebih berat dan masyarakat hanya bisa melakukan pengecekan dalam kurun waktu sebulan sekali pada saat kegiatan posyandu dilakukan. Beradarkan permasalahan yang terjadi perlu adanya perancangan sistem untuk memantau keadaan balita khususnya stunting dengan beberapa indikator guna mendukung inovasi dan pemahaman manyarakat akan pentinganya kecukupan gizi terhadap balita. Dengan adanya inovasi ini diharapkan orang tua dapat melakukan pengukuran indikator gizi balita khusunya yang berkaitan dengan kondisi stunting secara berkala dan tidak perlu menunggu satu bulan sekali untuk dilakukan pengukuran oleh petugas puskesmas setempat.

2.Dasar Teori

2.1 Stunting

Stunting merupakan suatu keadaan tinggi badan (TB) seseorang yang tidak sesuai dengan umur, yang dimana dapat diketahui dengan menghitung skor Z-indeks Tinggi badan menurut Umur (TB/U). Seseorang dikatakan stunting bila skor Z-indeks TB/U- nya di bawah -2 SD (standar deviasi)[2]. Pengukuran parameter biasa di lakukan setiap bulan sekali melalui posyandu yang diadakan oleh pihak puskesmas Pitu. Dari posyandu tersebut akan dilakukan pengukuran pada balita yang nantinya akan didapatkan parameter nama desa, nama posyandu, nama balita, jenis kelamin, umur, berat badan, dan tinggi badan.

Di Indonesia, pengukuran status gizi balita lebih banyak menerapkan *z-socre* atau Z-indeks. Z *score* adalah hasil yang menunjukkan pengukuran dari median[3].
Rumus *Z score* adalah :

$$\text{Z } Score = \frac{x\ hitung - MBR}{SBR} \quad \ldots(1)$$

Dimana :
x = Parameter yang di ukur
MBR = Median buku rujukan
SBR = Simpangan baku rujukan
Dari hasil Z *score* digunakan untuk menentukan status gizi pada anak, dan pada penelitian yang penulis lakukan, penulis berfokus pada *stunting* dimana indeks yang digunakan adalah Panjang/tinggi badan menurut umur. Untuk penjelasannya dapat dilihat pada tabel 2.1 berikut:

Tabel 1. Status gizi anak berdasar indeks[3]

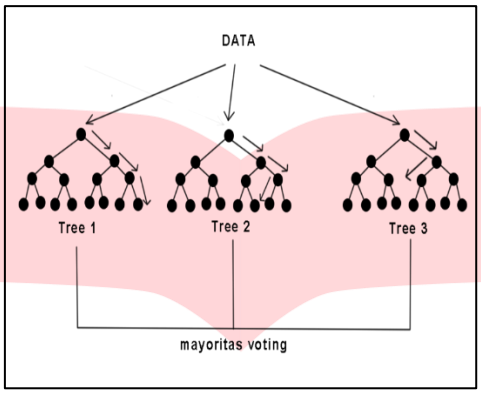
Nomor	Indeks	Kategori status gizi	Z *score*
1.	Berat badan menurut umur (BB/U)	Gizi buruk Gizi kurang Gizi baik Gizi lebih	<-3SD -3SD sampai dengan <-2SD -2SD sampai dengan 2SD >2SD
2.	Panjang badan menurut umur (TB/U)	Sangat pendek Pendek Normal Tinggi	<-3SD -3SD sampai dengan <-2SD -2SD sampai dengan 2SD >2SD
3.	Berat badan menurut Panjang badan (BB/TB)	Sangat pendek Pendek Normal Gemuk	<-3SD -3SD sampai dengan <-2SD -2SD sampai dengan 2SD >2SD

Kabupaten Ngawi sendiri pada tahun 2019 memiliki persentase balita dengan keadaan stunting yang cukup tinggi bahkan jika dibandingkan dengan provinsi Jawa Timur itu sendiri. Faktor yang mempengaruhi terjadinya stunting pada anak balita yang berada di wilayah pedesaan antara lain adalah pendidikan ibu, pendapatan keluarga,

pengetahuan ibu mengenai gizi, pemberian ASI eksklusif, umur pemberian MP-ASI, tingkat kecukupan zink dan zat besi, riwayat penyakit infeksi serta faktor genetik[1].

2.2 Random forest

Random forest didefinisikan sebagai kelompok klasifikasi dari pohon regresi, dilatih dari data pelatihan menggunakan pilihan fitur acak dalam proses *generate tree*. Setelah sejumlah besar *tree* telah di- *generate*, setiap *tree* divoting untuk mendapatkan kelas yang paling populer. Prosedur voting *tree* secara kolektif ini didevinisikan sebagai *random forest*. Untuk teknik klasifikasi *random forest* ini membutuhkan dua parameter yaitu jumlah *tree* dan jumlah atribut yang digunakan[4]. Banyak pohon ditumbuhkan sehingga terbentuk hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut. Respons suatu diprediksi dengan menggabungkan hasil prediksi k pohon. Pada masalah klasifikasi dilakukan berdasarkan *majority vote* (suara terbanyak).



Gambar 1 Gambaran umum dari algoritma *random forest*

Pada algoritma *random forest* tidak terlepas dari *decision tree* karena sejatinya algoritma *random forest* adalah kumpulan dari mayoritas voting beberapa *decision tree. Decision tree* sendiri merupakan algoritma kumpulan kondisional bersarang untuk menemukan *splitting* dengan nilai terbaik berdasarkan nilai *entropy* (ukuran informasi dalam sebuah *state*) untuk mencari nilai maksimal dari *information gain*.

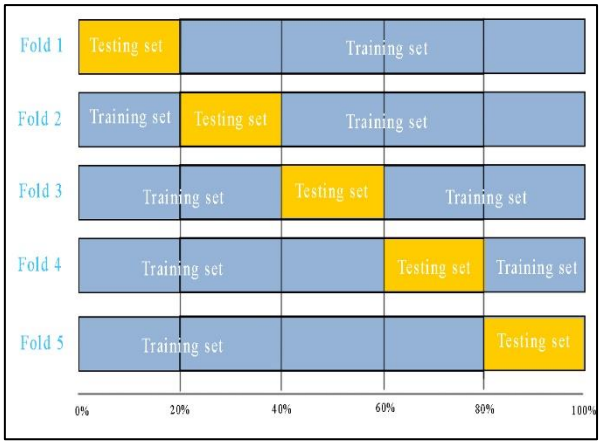
$$Entrophy = \sum -pi \log(pi) \qquad \ldots(2)$$

$$IG = E(parent) - \sum wi\ E(child\ i) \ldots(3)$$

Pi pada rumus entrophy merupakan kemungkinan dari *class* I sedangkan IG dalah *information gain* yang merupakan hasil pengurangan dari entrophy *parent* dengan entrophy *leaf node* (*child*). Dimana pada *random forest* akan dibentuk beberapa *decision tree*, dengan setiap hasilnya akan diambil untuk dijadikan voting (hasil dengan jumlah terbanyak) akan dijadikan keluaran dari algoritma *random forest*.

2.3 K-Fold Cross Validation

Pengujian *K-Fold Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dimana pada k yang digunakan pada penelitian ini sebanyak 3, 5, dan 10. Tiap percobaan akan menggunakan satu *data testing* dan k-1 bagian akan menjadi *data training*, kemudian *data testing* itu akan ditukar dengan satu buah *data training* sehingga untuk tiap percobaan akan didapatkan *data testing* yang berbeda-beda[5].



Gambar 2. 5-*fold cross validation*

3. Pembahasan

3.1 Metode Penelitian

Metode penelitian yang digunakan pada *project* kali ini mengarah pada penelitian absolut yang berpegangan pada dampak yang dihasilkan dari eksperimen[5].*Project* kali ini menggunakan algoritma *random forest* untuk membuat aplikasi prediksi kondisi stunting pada balita dimana sebelumnya pengukuran kondisi stunting pada balita yang ada di puskesmas pitu mesih menggunakan penghitungan manual. Beberapa tahapan yang dilakukan dalam project ini antara lain sebagai berikut:

1. Tahap *business understanding.* Tahap ini merupakan tahap pemahaman penelitian dan hasil yang akan didapat.
2. Tahapan data *understanding*. Tahap untuk mengumpulkan data
3. Tahap data *prepocessing* agar data yang didapat dapat diproses dengan algoritma yang digunakan dalam kasus kali ini menggunakan algoritma *random forest.*
   a. Data *cleaning*
      Merupakan tahapan dimana data akan dibersihkan dari data kosong
   b. Data labeling
      Pelabelan data berfungsi untuk menentukan target klasifikasi yang berupa class label yang telah dibuat.
   c. Feature Selection
      Digunakan untuk memilih fitur utama yang akan digunakan dalam model *machine learning* yang dibuat
4. Tahap modeling. Pada tahap ini dilakukan “fitting” data dengan algoritma yang digunakan dimana pada awalnya data telah dilakukan splitting dataset menjadi *train* set dan *test* set untuk dapat dihitung performanya dengan *confusion matix.*
5. Tahap evaluasi. Pada tahap ini dilakukan beberapa macam cara untuk mengukur evaluasi diantaranya dengan *train-test split* dan juga dengan menggunkan *cross validation.*
6. Tahap *Deployment.* Pada tahap ini merupakan tahapan pengaplikasian model *machine learning* yang telah dibuat kedalam aplikasi yang akan digunakan dalam project ini aplikasi dibuat dalam bentuk website dan android.

3.2 Hasil dan Pembahasan

1. Tahapan *Business Understanding*

   Berdasarkan data stunting puskesmas Pitu dan kegiatan posyandu yang setiap bulan dilakukan. Pendataan masih dilakukan dengan penghitungan manual dengan menggunakan median rujukan yang memakan waktu lama. Maka dengan itu dikembangkan model klasifikasi dengan menggunakan algoritma *random forest* untuk memprediksi apakah balita mengalami kondisi stunting atau tidak.
2. Tahap Data *Understanding*

   Data stunting diambil dari hasil pengukuran posyandu di 10 Desa di Kecamatan Pitu Kabupaten Ngawi. Data yang dipergunakan ialah data posyandu bulan Januari hingga Oktober 2020.
3. Tahapan Data Prerpocessing

Data yang telah terkumpul selanjutnya akan dilakukan tahap *preprocessing* data sebelum data dapat digunakan dalam proses *learning* menggunakan algoritma *random forest*. Data mentah yang terkumpul sebanyak 22855 baris dengan masih ada data kosong ataupun *missing value.* Perlu dilakukan pembersihan data dan pelabelan.

	SEX	UMUR	TB	TB/U	sangat_pendek	pendek	label
0	2.0	60.0	101.5	-1.665738	0	0	0
1	1.0	56.0	110.0	0.461114	0	0	0
2	1.0	55.0	97.0	-2.323074	0	1	1
3	1.0	52.0	106.6	0.200176	0	0	0
4	1.0	51.0	96.1	-2.096405	0	1	1
...	...	...	...	...	...	...	...
22850	1.0	22.0	77.3	-3.190909	1	0	1
22851	2.0	54.0	88.0	-4.069443	1	0	1
22852	1.0	46.0	82.8	-4.781951	1	0	1
22853	1.0	32.0	82.7	-3.153154	1	0	1
22854	1.0	16.0	70.0	-4.035724	1	0	1

18004 rows × 7 columns

Gambar 3. Data setelah pelabelan dan pembersihan data

Pada gambar 3 dapat dilihat bahwa jumlah data berkurang. Namun data masih belum baik digunakan dalam permasalahan klasifikasi dikarenakan data masih belum seimbang ini bisa dilihat dari jumlah masing-masing kelas label yang tersedia.

```
0    14102
1     3902
Name: label, dtype: int64
```

Gambar 4. Data Tidak Seimbang

Pada gambar 4 dapat dilihat bahwa jumlah perbandingan kelas yang ada sangat jauh. Dimana ini akan berpengaruh terhadap hasil klasifikasi yang didapat. Hasil akurasi yang didapat mungkin saja sangat tinggi namun hasil prediksi yang didapat akan tidak benar jika memprediksi kelas yang jumlahnya sedikit (minoritas). Maka dari itu perlu dilakukan penyeimbangan data dengan tekik *under sampling* dimana data dengan julah kelas terbanyak akan disesuaikan dengan jumlah jumlah kelas minoritas

```
1    3901
0    3901
Name: label, dtype: int64
```

Gambar 5. Data seimbang

4. Tahapan *modelling*

Pada tahapan ini data yang sudah melalui proses *preprocessing* selanjutanya dilakukan “fit” dengan menggunakan algoritma *random forest* dimana disini data akan dilakukan pemisahan menjadi sebesar 70:30 dimana 70% data digunakan sebagai *train* set dan 30% data digunakan sebagai *test* set. Dan Langkah akhirnya model akan simpan dan akan diaplikasikan kedalam aplikasi yang akan dibuat.

5. Tahapan Evaluasi

Evaluasi yang dilakukan pada *project* ini menggunakan du acara yaitu *Train-test split confusion matix* dengan pembagian data sebnyak tiga kali yaitu 50:50, 70:30, dan 90:10 dan k-*fold cross validation* dengan pembagian 3-fold, 5-fold, dan 10-fold.

a. Evaluasi model dengan *confusion matix* pembagian data 50:50

```
classification:
              precision    recall  f1-score   support

           0       0.98      0.97      0.98      1945
           1       0.97      0.98      0.98      1956

    accuracy                           0.98      3901
   macro avg       0.98      0.98      0.98      3901
weighted avg       0.98      0.98      0.98      3901

confusion:
[[1890   55]
 [  38 1918]]
Accuracy:
97.61599589848757
Precision:
97.21236695387735
Recall:
98.05725971370143
F1-Score:
97.63298549249174
```

Gambar 6. *Confusion matrix* splitting dataset 50:50

b. Evaluasi model dengan *confusion matix* pembagian data 70:30

```
classification:
              precision    recall  f1-score   support

           0       0.99      0.97      0.98      1176
           1       0.97      0.99      0.98      1165

    accuracy                           0.98      2341
   macro avg       0.98      0.98      0.98      2341
weighted avg       0.98      0.98      0.98      2341

confusion:
[[1145   31]
 [  17 1148]]
Accuracy:
0.9794959419051688
Precision:
0.9737065309584394
Recall:
0.9854077253218884
F1-Score:
0.9795221843003413
```

Gambar 7. *Confusion matrix* splitting dataset 70:30

c. Evaluasi model dengan *confusion matix* pembagian data 90:10

```
classification:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       381
           1       0.98      0.98      0.98       400

    accuracy                           0.98       781
   macro avg       0.98      0.98      0.98       781
weighted avg       0.98      0.98      0.98       781

confusion:
[[374   7]
 [  9 391]]
Accuracy:
97.95134443021767
Precision:
98.24120603015075
Recall:
97.75
F1-Score:
97.99498746867167
```

Gambar 8. *Confusion matrix* splitting dataset 90:10

d. Evalusai model dengan 3-*fold cross validation*

	Rata-rata nilai pengujian
3-*fold cross validation*	97.70576405524501

e. Evalusai model dengan 5-*fold cross validation*

	Rata-rata nilai pengujian
5-*fold cross validation*	97.88529706466926

f. Evalusai model dengan 10-*fold cross validation*

	Rata-rata nilai pengujian
10-*fold cross validation*	97.87236941462292

6. Tahapan Deplyment

Pada Tahapan ini model yang telah dibuat dan telah melalui proses pengujian dan hasil yang diperoleh dianggap layak maka model akan digunakan dalam aplikasi denganbantuan flask framework. Dimana flask digunakan sebagai jembatan untuk aplikasi dapat mengakses model yang telah dibuat sehingga website maupaun aplikasi android dapat mengguanakan model yang telah dibuat.

4. Kesimpulan

Algoritma *random forest* memiliki rata-rata akurasi pengujian menggunakan *confusion matix*sebesar 97.83% dan apada pengujian evaluasi dengan menggunakan *k-fold cross validation* rata-rata akurasi yang diperoleh sebesar 97.821% dimana hasil pada setiap pengujian evalusai akurasi yang dihasilkan algoritma *random forest* stabil pada *range* 97% diamana dapat disimpulkan bahwa algoritma random forest dapat digunakan dalam memprediksi kondisi stunting pada balita.

REFERENSI

[1] H. Mizobe *et al.*, “Structures and Binary Mixing Characteristics of Enantiomers of 1-Oleoyl-2,3-dipalmitoyl-sn-glycerol (S-OPP) and 1,2-Dipalmitoyl-3-oleoyl-sn- glycerol (R-PPO),” *JAOCS, J. Am. Oil Chem. Soc.*, vol. 90, no. 12, pp. 1809–1817, 2013, doi: 10.1007/s11746-013-2339-4.

[2] A. Boucot and G. Poinar Jr., “Stunting,” *Foss. Behav. Compend.*, vol. 5, pp. 243–243, 2010, doi: 10.1201/9781439810590-c34.

[3] clarita tiffany, “Stunting Balita,” vol. 1, pp. 367–373, 2016.

[4] R. Annisa, “Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung,” *J. Tek. Inform. Kaputama, vol. 3, no. 1*, pp. 22–28, 2019, [Online]. Available: available: http://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141.

[5] N. N. P. Galih. Surono, “Journal of technology information,” *Http://Jurnal.Kampuswiduri.Ac.Id/*, vol. 5, no. 1, pp. 25–30, 2020, [Online]. Available: http://jurnal.kampuswiduri.ac.id/index.php/infoteh/article/view/79/67.