

metabolisme, dan pengaruh zat-zat yang tidak digunakan untuk mempertahankan kehidupan, pertumbuhan dan fungsi normal dari organ-organ serta menghasilkan energi (Supariasa,dkk 2002).

Penilaian status gizi dapat digunakan sebagai pemantauan orang tua terhadap status gizi untuk pertumbuhan anak yang sangat dibutuhkan bagi perkembangan anak, juga dapat digunakan untuk memberikan penilaian status gizi terhadap seseorang yang berfungsi untuk rujukan dari masyarakat atau puskesmas. Banyak cara untuk menilai status gizi salah satunya adalah dengan cara pengukuran tubuh manusia yang dikenal dengan istilah “*Anthropometri*”.

Dalam menentukan status gizi balita memiliki ukuran baku. Ukuran baku yang di gunakan di Indonesia sekarang adalah standar baku *World Health Organization-National Center for Health Statistics (WHO-NCHS)*. Penilaian status gizi balita dipisahkan antara laki-laki dan perempuan.

Menurut WHO, (2005) yang dikutip dari (Febrealty, 2011) penilaian status gizi berdasarkan Indeks BB/U (Berat Badan menurut Umur), TB/U (Tinggi Badan menurut Umur), BB/TB (Berat Badan menurut Tinggi Badan) dengan standar baku *Anthropometri WHO-NCHS* dapat digolongkan menjadi:

Tabel 2.3 Penilaian Status Gizi (Febrealty, 2011)

No	Indeks yang dipakai	Batas Pengelompokan	Sebutan Status gizi
1.	BB/U	<-3 SD -2 s/d <-2 SD -2 s/d +2 SD > +2 SD	Gizi Buruk Gizi Kurang Gizi Baik Gizi Lebih
2.	TB/U	<-3 SD -2 s/d <-2 SD -2 s/d +2 SD > +2 SD	Sangat Pendek Pendek Normal Tinggi
3.	BB/TB	<-3 SD -2 s/d <-2 SD -2 s/d +2 SD > +2 SD	Sangat Kurus Kurus Normal Gemuk

Dimana SD adalah Skor Simpangan Baku (Standar Deviation = Z)

Cara menghitung sttus gizi dengan *Z-score* dapat ditentukan dengan menggunakan rumus:

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Mean Acuan})}{\text{Nilai Sim pang Baku Rujukan}} \quad (11)$$

Dimana terdapat dua kategori dalam menghitung status gizi balita menggunakan *Z-score*, yaitu:

Bila “Nilai Riel Perorangan” hasil pengukuran \geq “Nilai Median Acuan” BB/U, TB/U, BB/TB, maka rumusnya

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Mean Acuan})}{SD Upper}$$

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Mean Acuan})}{+1 SD - Median}$$

Bila “Nilai Riel Perorangan” hasil pengukuran \leq “Nilai Median Acuan” BB/U, TB/U, BB/TB, maka rumusnya

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Mean Acuan})}{SD Lower}$$

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Mean Acuan})}{Median - -1 SD}$$

Keterangan: Nilai Riel itu berat badan sebenarnya (aktual)

Nilai Median itu diambil dari nilai tabel Baku Rujukan WHO-NCHS

Nilai (-1SD) itu juga dapat dilihat pada tabel WHO-NCHS

- Jika nilai riel lebih kecil dari pada nilai median berarti yang digunakan sebagai pembagi adalah nilai -1SD
- Jika nilai riel lebih besar dari pada nilai median berarti yang digunakan sebagai pembagi adalah nilai +1SD

Menurut Arisman, (2008) yang dikutip dari (Febrealty, 2011) interpretasi status gizi berdasarka tiga indeks *Anthropometri* (BB/U, TB/U, BB/TB) adalah sebagai berikut:

Tabel 2.4 Interpretasi Status Gizi (Febrealty, 2011)

No	Indeks yang digunakan			Interpretasi
	BB/U	TB/U	BB/TB	
1.	Rendah Rendah Rendah	Rendah Tinggi Normal	Normal Rendah Rendah	Normal, dulu kurang gizi Sekarang kurang ++ Sekarang kurang +
2.	Normal Normal Normal	Normal Tinggi Rendah	Normal Rendah Tinggi	Normal Sekarang kurang Sekarang lebih, dulu kurang
3.	Tinggi Tinggi Tinggi	Tinggi Rendah Normal	Normal Tinggi Tinggi	Tinggi, normal Obesitas Sekarang lebih, belum obesitas

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti lain yang susah untuk dibaca dan dipahami, *python* lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat *python* sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain.

Bahasa ini muncul pertama kali pada tahun 1991, dirancang oleh seorang bernama Guido van Rossum. Sampai saat ini *pyhton* masih dikembangkan oleh *Python Software Foundation*. Bahasa *python* mendukung hampir semua sistem operasi, bahkan untuk sistem operasi *Linux*, hampir semua distronya sudah menyertakan *python* di dalamnya.

Menurut Luts, (2010) yang dikutip dari (Harismawan, dkk. 2018) Python adalah bahasa pemrograman yag bersifat open source. Bahasa pemrograman ini dioptimalkan untuk *software quality, developer productivity, program portability, dan component integration*.

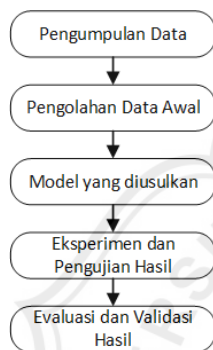
Menurut Luts, (2010) yang dikutip dari (Harismawan, dkk. 2018)Python telah digunakan untuk mengembangkan bahasa berbagai macam perangkat lunak, seperti *internet scripting, sytems programming, user interfaces, product customization, numeric programming* dll. Python saat ini telah menduduki posisi 4 atau 5 bahasa pemrograman paling sering digunakan diseluruh dunia.

III METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif. Metode kuantitatif bertujuan untuk mencapai pemahaman tentang bagaimana suatu dikonstruksi, bagaimana dibangun, atau bagaimana cara kerjanya menurut Berndtsson, Hansson, Olsson, & Lundell, (2008) yang dikutip dari (Sabaruddin,

2017). Penelitian kuantitatif umumnya didorong oleh hipotesis, yang dirumuskan dan diuji secara ketat, dengan tujuan menunjukkan bahwa hipotesisnya salah. Oleh karena itu, salah satu upaya adalah untuk menyalahkan hipotesis, dan jika hipotesis tahan uji, maka akan dianggap benar setelah terbukti. Aspek kuantitatif adalah untuk menekan bahwa pengukuran merupakan dasarnya karena memberikan hubungan antara observasi dan formalisasi model, teori, dan hipotesis.

Menurut Dawson, (2009) yang dikutip dari (Sabaruddin, 2017) Metode yang digunakan pada penelitian ini adalah eksperimen. Penelitian eksperimen mencakup investigasi hubungan sebab-akibat menggunakan pengujian yang dikontrol sendiri. Penelitian ini bertujuan untuk menerapkan pendekatan level data untuk mengurangi pengaruh ketidakseimbangan data. Karena penelitian yang diakui/diterima harus mengikuti aturan yang diakui, maka pada penelitian ini dilakukan dengan mengikuti tahapan seperti Gambar 3.1.

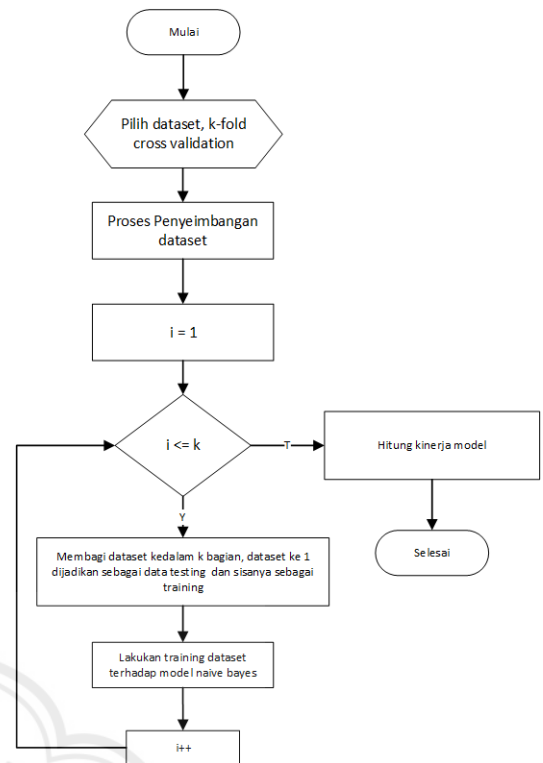


Gambar 3.1. Diagram Tahapan Penelitian

Pada penelitian ini digunakan data sekunder yaitu dari Puskesmas Tegaldlimo dengan proses pengambilan melalui Dinas terkait yaitu Dinas Kesehatan Kabupaten Banyuwangi terlebih dahulu, setelah mendapat persetujuan dari dinas kesehatan barulah dilanjutkan ke puskesmas Tegaldlimo. Adapun jumlah data yang diperoleh dari puskesmas tegaldlimo yaitu sebanyak 1186 dengan rincian balita gizi baik berjumlah 1101, balita gizi kurang berjumlah 39, balita gizi lebih berjumlah 39 dan balita gizi buruk berjumlah 7.

Pada pengolahan data awal ini akan dilakukan beberapa tahapan untuk memperoleh data yang diinginkan yaitu penggantian Status Gizi dengan gizi baik sama dengan 0, gizi kurang sama dengan 1, gizi lebih sama dengan 2, gizi buruk sama dengan 3.

Untuk mencari solusi masalah ketidakseimbangan kelas pada dataset, diusulkan model dengan pendekatan level data. Perancangan model yang diusulkan meliputi perancangan pendekatan level data ROS, dan algoritma pengklasifikasian berbasis *Naive Bayes*. Perancangan algoritma ditunjukkan melalui flowchart pada gambar 3.2.



Gambar 3.2 Flowchart Model yang Diusulkan.

Proses pertama yang dilakukan adalah memilih dataset, pendekatan level data, disini menggunakan pendekatan level data *Random Over Sampling* dan jumlah *k-fold* yang ingin diujikan, misal 10.

Proses kedua adalah proses penyeimbangan data, karena disini menggunakan *Random Over Sampling* jadi mencari selisih data antara data mayoritas dan data minoritas, setelah ketemu data minoritas ditambahkan secara *random* oleh sistem sehingga data minoritas dan data mayoritas seimbang.

Proses ketiga adalah membagi dataset kedalam *k* bagian, yang 1 dijadikan sebagai data *training* dan sisanya dijadikan data *testing*, lakukan training terhadap *Naive Bayes* dan ukur kinerjanya. Lakukan *looping* sesuai dengan *k* yang diinginkan.

Eksperimen dan hasil pengujian penelitian ini menggunakan *resampling* untuk menangani ketidakseimbangan kelas berbasis *Naive Bayes*. Gambar 3.3 menggambarkan algoritma model yang diusulkan pada penelitian ini.

Dataset yang digunakan dalam penelitian ini berasal dari data balita Puskesmas. Data ini diambil langsung dari Puskesmas dengan beberapa proses pengambilan yang sudah dijelaskan diawal.

Langkah pertama pada pengujian ini adalah menghitung terlebih dahulu selisih antara data gizi baik dan gizi kurang untuk menentukan jarak masing-masing data pada kelas mayoritas dan kelas minoritas. Algoritma yang diusulkan adalah *over-sampling*, yaitu menambah populasi kelas minoritas sedemikian hingga seimbang dalam jumlah kelas mayoritas.

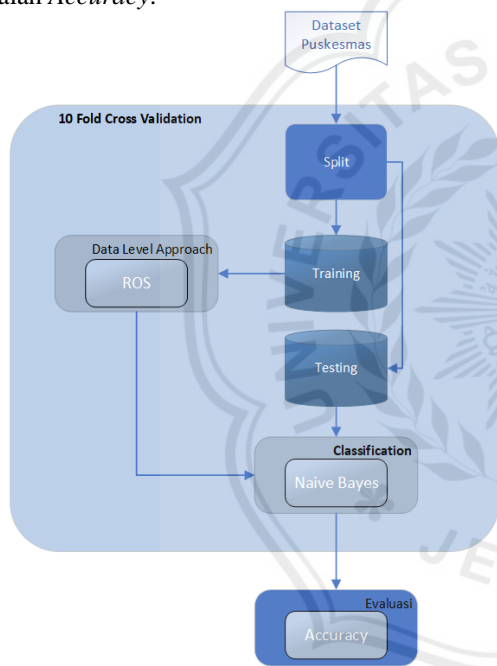
Langkah kedua pada pengujian ini adalah menggunakan hasil *resampling* yang didapat pada langkah sebelumnya sebagai data masukan untuk model prediksi berbasis *Naive Bayes*. Kinerja dari model prediksi kemudian dievaluasi menggunakan *2, 4, 5, 10-fold cross validation*.

Hasil validasi digunakan untuk mengukur kinerja masing-masing model, dan dilakukan perbandingan untuk mencari model yang memiliki kinerja terbaik.

Metode resampling yang digunakan adalah *random over sampling* (ROS). Perangkat lunak yang digunakan adalah Anaconda 3 dengan bahasa pemrograman Python dan Microsoft Excel.

Menurut Han, Kamber, & Pei, (2011, p.365) yang dikutip dari (Utomo Pujianto, 2016) Untuk mengukur kinerja model digunakan *confusion matrix*, karena *confusion matrix* merupakan alat yang berguna untuk menganalisa seberapa baik pengklasifikasi dapat mengenali tupel/ fitur dari kelas yang berbeda menurut yang dikutip dari. *Confusion matrix* dapat membantu menunjukkan rincian kinerja pengklasifikasi dengan memberikan informasi jumlah fitur suatu kelas yang diklasifikasikan dengan cepat dan tidak tepat menurut Bramer, (2007, p. 89) yang dikutip dari (Utomo Pujianto, 2016). *Confusion matrix* merupakan matrik 2 dimensi yang menggambarkan perbandingan antara hasil prediksi dengan kenyataan.

Untuk data tidak seimbang, akurasi lebih didominasi oleh ketepatan pada data kelas minoritas, maka metrik yang tepat adalah *Accuracy*.



Gambar 3.3 Kerangka kinerja yang diusulkan

Pengukuran kinerja model dilakukan dengan menggunakan confusion matrix. Kinerja yang diukur termasuk akurasi secara umum, akurasi dalam memprediksi kelas minoritas. Confusion matrix diperoleh dari validasi menggunakan 2, 4, 5, 10-fold cross validation, sehingga model yang terbentuk dapat langsung diuji dengan melakukan 2, 4, 5, 10 kali pengujian.

Kinerja model yang diperoleh digunakan untuk membandingkan antara model dasar dengan algoritma *Naive Bayes* dengan model yang dibentuk menggunakan algoritma *Naive bayes* dengan pendekatan level data ROS. Untuk melihat model yang terbentuk, dimana nilai Accuracy digunakan untuk menentukan klasifikasi keakuratan pengujian diagnostik.

Sebuah pedoman umum untuk mengklasifikasikan keakuratan pengujian diagnostik menggunakan Accuracy

dapat dilihat pada sistem tradisional (Gorunescu, 2011) disajikan pada tabel 3.3.

Nilai	Klasifikasi
0.9 – 1	Excellent classification
0.8 – 0.9	Good classification
0.7 – 0.8	Fair classification
0.6 – 0.7	Poor classification
< 0.6	Failure

IV HASIL DAN PEMBAHASAN

Sumber data yang dipakai dalam penelitian ini berasal dari dataset Puskesmas Tegaldlimo. Sedangkan aplikasi pendukung yang digunakan adalah Anaconda 3. Model yang diuji adalah model pengklasifikasian *Naive Bayes* dan *Random Over Sampling*.

Jumlah data yang diperoleh dari dataset Puskesmas adalah sebanyak 1186 data, dengan kategori gizi baik sebanyak 1101 balita, gizi kurang sebanyak 39 balita gizi lebih sebanyak 39 balita dan gizi buruk sebanyak 7 balita. Karena hasil dari gizi baik, gizi kurang, gizi lebih dan gizi buruk bertipe *String* sehingga tidak bisa digunakan dan harus dirubah dalam bentuk *integer*, dengan ketentuan gizi baik adalah 0, gizi kurang adalah 1, gizi lebih adalah 2 dan gizi buruk adalah 3.

Tabel 4.1 Dataset Puskesmas

	Sex	Umur	BB	TB	Status_Gizi
0	2	8	7,3	65	0
1	1	7	8	65	0
2	2	7	8,3	68	0
3	1	42	14,6	95	0
4	2	41	14	93	0
...
1182	2	34	10	86	3
1183	2	33	10	86	3
1184	1	44	12	88,5	3
1185	2	44	11	100	3
1186	1	38	10,8	79	3

Dataset diatas disimpan menggunakan ekstensi *xlsx*. Format *xlsx* merupakan salah satu format ekstensi dari Microsoft Office yaitu Microsoft Office Excel. Berikut adalah *source code* yang digunakan dalam penggunaan file ekstensi *xlsx* di Python.

```
import pandas as pd
df = pd.read_excel('C:\databalita.xlsx')
```

Import pandas as pd merupakan library untuk penggunaan file ekstensi *xlsx*, *df* bisa di ibaratkan dengan *class* yang menampung semua kolom yang ada di file tersebut. *pd.read_excel* merupakan perintah untuk membaca format *xlsx* atau *excel*. *C:\databalita.xlsx* merupakan lokasi penyimpanan dataset yang akan digunakan.

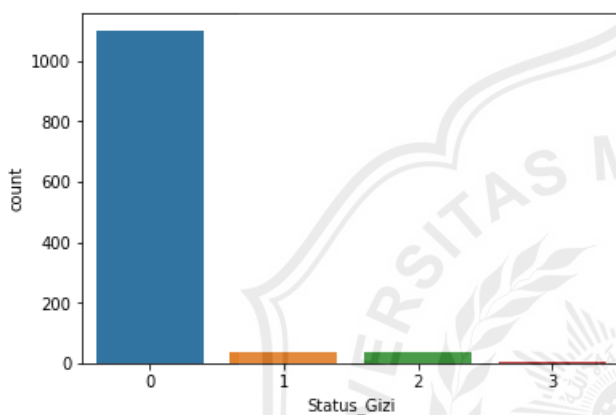
Teknik validasi yang digunakan adalah 2, 4, 5, 10-fold cross validation, sehingga dataset dibagi menjadi 2, 4, 5, 10 bagian, kemudian diambil 2 bagian secara random oleh sistem untuk dijadikan sebagai data uji (*test*) dan yang lainnya dijadikan data latihan (*training*).

Untuk tahap selanjutnya yaitu penanganan kelas data yang tidak seimbang dengan menggunakan *OverSampling*:

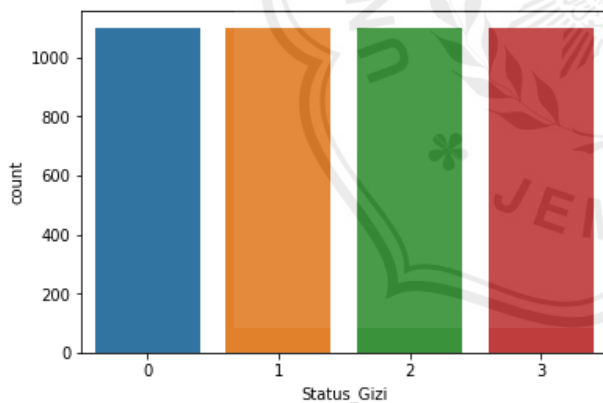
- i. Menentukan nilai Fold dengan k=2, 4, 5, 10 dengan 70% dari data dijadikan data training dan 30% data dijadikan data testing.

- ii. Menghitung jumlah data mayoritas dan data minoritas, yang selanjutnya dicari selisih antara ke 2 data tersebut.
- iii. Melakukan perhitungan untuk mencari data minoritas untuk ditambahkan secara random.
- iv. Menerapkan metode *Naive Bayes* untuk mengklasifikasi data uji.
- v. Membandingkan kinerja klasifikasi tanpa dan dengan diterapkannya ROS (*Random Over Sampling*).

Penerapan ROS untuk meningkatkan data minoritas yang jumlahnya jauh dibandingkan dengan data mayoritas pada dataset Puskesmas, dari data awal yang berjumlah 1186 data dengan Gizi baik 1101 data, gizi lebih 39 data, gizi kurang 39 data dan gizi buruk 7 data, sehingga total data keseluruhan setelah dilakukan *Random Over Sampling* menjadi 4404 yang terdiri dari 1101 Gizi Baik (label 0) , 1101 Gizi Kurang (label 1) 1101 Gizi Kurang(label 2), 1101 Gizi Buruk (label 3).



Gambar 4.1. Grafik ketidak seimbangan awal



Gambar 4.2. Grafik setelah diterapkannya Random Over Sampling

Berdasarkan gambar diatas dapat dipahami bahwa penerapan *Random Over Sampling* untuk data Gizi kurang (1) yaitu sebesar 1062 data, Gizi lebih (2) yaitu sebesar 1062 data dan Gizi buruk (3) yaitu sebesar 1094 data.

Pada tabel berikut ditampilkan performa klasifikasi dengan *Naive Bayes* dan dengan *Naive Bayes + Random Over Sampling*. Pada setiap percobaan menerapkan nilai k yang bervariasi yaitu 2, 4, 5, 10.

Tabel 4.2. Hasil dari NB + ROS

Fold	Akurasi Ros cv 2	Akurasi Ros cv 4	Akurasi Ros cv 5	Akurasi Ros cv 10
1	0,486405	0,510511	0,503759	0,544776
2	0,512121	0,504532	0,548872	0,537313

3		0,507599	0,483019	0,541353
4		0,516717	0,555133	0,462121
5			0,492366	0,5
6				0,484848
7				0,530303
8				0,606061
9				0,549618
10				0,530769
Rata-Rata	0,499263	0,50983975	0,5166298	0,5287162

Tabel 4.3. Hasil dari NB

Fold	Akurasi NB cv2	Akurasi NB cv4	Akurasi NB cv5	Akurasi NB cv10
1	0,928934	0,929293	0,930556	0,918919
2	0,948718	0,938776	0,985915	0,972222
3		0,938776	0,971831	0,972222
4		0,948454	0,957746	1
5			0,929577	0,972222
6				0,971429
7				0,942857
8				0,942857
9				0,914286
10				0,971429
Rata-Rata	0,938826	0,93882475	0,955125	0,9578443

Pada bagian ini akan membahas hasil pengukuran kinerja model, yaitu Bagaimana tingkat akurasi algoritma *Random Over Sampling* untuk mengatasi terjadinya ketidakseimbangan kelas? Untuk menjawab ditunjukkan perbandingan kinerja model *Naive Bayes* dan model *Random Over Sampling + Naive Bayes* pada tabel 4.4 dan tabel 4.5 menunjukkan tingkat *accuracy* pada dataset Puskesmas.

Tabel 4.4. Perbandingan *Accuracy* model dengan *Naive Bayes*

Fold	NB	Keterangan
Fold 1	0,918919	Excellent
Fold 2	0,972222	Excellent
Fold 3	0,972222	Excellent
Fold 4	1	Excellent
Fold 5	0,972222	Excellent
Fold 6	0,971429	Excellent
Fold 7	0,942857	Excellent
Fold 8	0,942857	Excellent
Fold 9	0,914286	Excellent
Fold 10	0,971429	Excellent
Rata-Rata	0,957844	Excellent

Tabel 4.5. Perbandingan *Accuracy* model dengan *Naive Bayes + ROS*

Fold	ROS	Keterangan
Fold 1	0,544776	Failure
Fold 2	0,537313	Failure
Fold 3	0,541353	Failure
Fold 4	0,462121	Failure
Fold 5	0,5	Failure
Fold 6	0,484848	Failure
Fold 7	0,530303	Failure
Fold 8	0,606061	Poor
Fold 9	0,549618	Failure
Fold 10	0,530769	Failure
Rata-Rata	0,528716	Failure

Tabel perbandingan akurasi pada tabel 4.4 dan tabel 4.5 dari hasil beberapa fold yang digunakan adalah 2, 4, 5, 10 dan dari hasil yang terbaik adalah 10 fold, maka 10 fold yang diambil, menunjukkan bahwa model *Naive Bayes* menghasilkan akurasi *Excellent* (Sangat Baik) dengan rata – rata akurasi yang dihasilkan adalah 0,957844. Sedangkan akurasi dengan model ROS + NB cenderung lebih rendah atau dibisa dikatakan belum maksimal untuk meningkatkan nilai akurasi dengan memiliki rata – rata yaitu 0,528716 *Fail* (Gagal) dalam mengklasifikasi *dataset* Puskesmas, hal ini dikarenakan jumlah data yang sangat signifikan.

V KESIMPULAN

Data Puskesmas merupakan sebuah dataset yang memiliki ketidak seimbangan kelas. Penelitian ini menerapkan *Random Over Sampling* untuk menangani ketidak seimbangan kelas pada dataset Puskesmas, serta metode *Naive Bayes* (NB) untuk melaksanakan fungsi klasifikasi. Hasil eksperimen dengan menerapkan nilai k yang bervariasi yaitu 2, 4, 5, 10 dan mengambil dari nilai k yang terbaik yaitu 10 fold diperoleh rata – rata nilai performa *Accuracy* sebesar 53% untuk skema ROS + NB dan 96% untuk skema NB, hal ini menunjukkan bahwa ROS + NB kurang maksimal .

Berdasarkan hasil eksperimen di atas diperoleh kesimpulan bahwa penerapan *Naive Bayes* + *Random Over Sampling* kurang maksimal untuk mengatasi masalah *imbalance* data pada klasifikasi gizi balita dikarenakan jumlah data pada gizi baik yang signifikan lebih banyak dari pada status gizi yang lainnya, sehingga akan mempengaruhi hasil dari proses *random over sampling* pada klasifikasi balita.

DAFTAR PUSTAKA

- Amelia, Rizka. (2013). *Penyapihan Dini dengan Status Balita Usia 0-24 Bulan Di Posyandu Dusun Kedungbendo Desa Gemekan Sooko Mojokerto*. Hospital Majapahit. Vol 5 No.1 Pebruari 2013.
- Bustami. (2013). *Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi*. Jurnal Informatika. Universitas Malikussaleh. Vol.8, No.1 Januari 2014.
- Eka, Rahmanurul Febrealti. (2011). *Sistem Penentuan Status Gizi Balita Menggunakan Metode K-NN (K-Nearest Neighbor)* [Skripsi]. Riau (ID). Universitas Islam Negeri Sultan Syarif Kasim.
- Febrealty, Eka. *Klasifikasi Status Gizi balita menggunakan metode k-Nearest Neighbor*[Skripsi].2011.
- Harismawan, F.A., Kharisma, P.A., Afirianto, T. (2018). *Analisis Perbandingan Performa Web Service Menggunakan Bahasa Pemrograman Python, PHP, dan Perl pada Client Berbasis Android*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Universitas Brawijaya. Vol. 2, No.1, Januari 2018.
- Prasetyo, Eko. (2012). *Data Mining-Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta :Andi Offset.
- Proverawati, Asfuah S., 2009. *Buku Ajar Gizi untuk Kebidanan*. Yogyakarta: Nuha Medika.
- Pujianto, Utomo. (2016). *Strategi Resampling berbasis Centroid untuk Menangani Ketidakseimbangan Kelas pada Prediksi cacat Perangkat Lunak*. Malang (ID). Universitas Negeri Malang. Tekno, Vol 25 Maret 2016.
- Riswanto Ricky, dkk (2007). *Metode Sampling dalam Menyelesaikan Data text Imbalance untuk Klasifikasi Multi-Label*. Bandung [ID]. Universitas Telkom.
- Sabaruddin, Raja. (2017). *Prediksi Cacat Software Menggunakan Resampling Berbasis C5.0 untuk Menangani Ketidakseimbangan Kelas* [tesis]. Jakarta (ID). STMIK Nusa Mandiri.
- Saifudin, Aries. (2014). *Pendekatan Level Data dan Algoritma untuk Penanganan Ketidakseimbangan pada Prediksi Cacat Software Berbasis Naive Bayes* [tesis]. Jakarta (ID). Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.
- Saifudin, Aries, Wahono. (2015). *Pendekatan Level Data Untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software*. Jurnal Of Software Engineering, Vol.1:2.

- Supariasa, D. dkk. (2002). *Penilaian Status Gizi*. Jakarta: EGC.
- Suyanto. 2009. *Ilmu Pengantar Status Gizi*. Penerbit Parana Ilmu. Bandung.
- Venny Lovina Gumiri, Diyah Puspitaningrum, Ernawati. (2015). *Sistem Pakar Klasifikasi Status Perkembangan Anak Usia Dini dengan Metode Naive Bayes Classifier Berbasis DDST Rules*. Jurnal Rekursif, Vol. 3:108-111.
- ZK. Abdurahman Baizal, Moch. Arif Bijaksana, Angelina Sagita Sastrawan. (2009). *Analisis Pengaruh Metode Over Sampling Dalam Churn Prediction untuk Perusahaan Telekomunikasi*. Seminar Nasional Aplikasi Teknologi Informasi 2009 (SNATI 2009). G61