



### PROTEKSI ISI PROPOSAL

Dilarang menyalin, menyimpan, memperbanyak sebagian atau seluruh isi proposal ini dalam bentuk apapun kecuali oleh pengusul dan pengelola administrasi pengabdian kepada masyarakat

### PROPOSAL PENELITIAN 2023

Rencana Pelaksanaan Penelitian: tahun 2023 s.d. tahun 2023

#### 1. JUDUL PENELITIAN

Optimasi Algoritma SVM dengan Chi-Square dan SMOTE untuk High Dimension Data Stunting Kota Samarinda

Bidang Fokus RIRN / Bidang Unggulan Perguruan Tinggi	Tema	Topik (jika ada)	Rumpun Bidang Ilmu
Teknologi Informasi dan Komunikasi	Pengembangan sistem berbasis Kecerdasan buatan	Pengembangan aplikasi sistem cerdas	Teknik Informatika

Kategori (Kompetitif Nasional/ Desentralisasi/ Penugasan)	Skema Penelitian	Strata (Dasar/ Terapan/ Pengembangan)	SBK (Dasar, Terapan, Pengembangan)	Target Akhir TKT	Lama Penelitian (Tahun)
Penelitian Kompetitif Nasional	Penelitian Dosen Pemula	Riset Dasar	SBK Riset Pembinaan/ Kapasitas	2	1

#### 2. IDENTITAS PENGUSUL

Nama, Peran	Perguruan Tinggi/ Institusi	Program Studi/ Bagian	Bidang Tugas	ID Sinta
TAGHFIRUL AZHIMA YOGA SISWA Ketua Pengusul	Universitas Muhammadiyah Kalimantan Timur	Teknik Informatika	Merancang ide dan pelaksanaan penelitian	<a href="#">6629575</a>
NAUFAL AZMI VERDIKHA Anggota Pengusul	Universitas Muhammadiyah Kalimantan Timur	Teknik Informatika	Membantu penyusunan proposal, pengumpulan data, analisis data, penulisan artikel ilmiah, dan penyusunan laporan akhir	<a href="#">6700346</a>
PUTRA SAYMEN Mahasiswa Bimbingan	Universitas Muhammadiyah Kalimantan Timur	Teknik Informatika	Mendukung Pengumpulan Data dan Data Preprocessing	-

#### 3. MITRA KERJASAMA PENELITIAN (JIKA ADA)

Pelaksanaan penelitian dapat melibatkan mitra kerjasama yaitu mitra kerjasama dalam melaksanakan penelitian, mitra sebagai calon pengguna hasil penelitian, atau mitra investor

Mitra	Nama Mitra	Dana
-------	------------	------

#### 4. LUARAN DAN TARGET CAPAIAN

##### Luaran Wajib

Tahun Luaran	Jenis Luaran	Status target capaian	Keterangan
1	Artikel di Jurnal	accepted/published	Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika (Sinta 2)

#### 5. ANGGARAN

Rencana Anggaran Biaya penelitian mengacu pada PMK dan buku Panduan Penelitian dan Pengabdian kepada Masyarakat yang berlaku.

**Total RAB 1 Tahun Rp. 15.400.000,00**

**Tahun 1 Total Rp15.400.000,00**

Jenis Pembelanjaan	Komponen	Item	Satuan	Vol.	Biaya Satuan	Total
Pengumpulan Data	Transport	Transport Perizinan Penelitian	OK (kali)	2	100.000	200.000
Pelaporan, Luaran Wajib, dan Luaran Tambahan	Biaya Publikasi artikel di Jurnal Nasional	Biaya Publikasi artikel di Jurnal Nasional (Sinta 1-3)	Paket	1	1.500.000	1.500.000
Bahan	ATK	(Tinta Printer, Kertas A4, Pulpen, Map, Materai 10.000, Spidol, Buku Catatan)	Paket	3	500.000	1.500.000
Pengumpulan Data	Biaya konsumsi	Konsumsi	OH	20	50.000	1.000.000
Pengumpulan Data	Transport	Pengambilan Data Penelitian	OK (kali)	10	20.000	200.000
Analisis Data	Honorarium narasumber	Honor Narasumber	OJ	1	1.000.000	1.000.000
Analisis Data	HR Pengolah Data	Honor Pengolahan Data	P (penelitian)	2	1.250.000	2.500.000
Pelaporan, Luaran Wajib, dan Luaran Tambahan	Biaya konsumsi rapat	Konsumsi	OH	10	50.000	500.000
Sewa Peralatan	Peralatan penelitian	Printer dan Scanner	Unit	1	1.500.000	1.500.000
Analisis Data	Biaya konsumsi rapat	Konsumsi	OH	40	50.000	2.000.000
Sewa Peralatan	Peralatan penelitian	Proyektor	Unit	1	1.250.000	1.250.000
Analisis Data	Transport Lokal	Transport	OK (kali)	2	300.000	600.000
Pengumpulan Data	HR Pembantu Peneliti	Hari	OJ	10	150.000	1.500.000
Pengumpulan Data	Biaya konsumsi	Konsumsi Perizinan Penelitian	OH	1	150.000	150.000



## Isian Substansi Proposal SKEMA PENELITIAN DASAR

Petunjuk: Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

### JUDUL

Tuliskan Judul Usulan

Optimasi Algoritma SVM dengan Chi-Square dan SMOTE untuk High Dimension Data Stunting Kota Samarinda

### RINGKASAN

Ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, dan luaran yang ditargetkan.

Stunting masih dianggap masalah serius dan menjadi fokus perhatian pemerintah dalam prioritas riset nasional tahun 2020-2024 dengan prevalensi stunting 21,6% pada tahun 2022. Kota Samarinda berada pada posisi tertinggi kedua setelah kabupaten Kutai Kartanegara di Provinsi Kalimantan Timur di tahun 2022 dengan persentase 25,3%. Diperlukan upaya untuk melakukan prediksi status stunting pada anak balita menggunakan metode analisis data yang canggih seperti data mining. Melihat trend penelitian sebelumnya, penggunaan metode klasifikasi seperti KNN, *Naïve Bayes*, *Random Forest*, *Neural Network*, CART, dan SVM pada dasarnya sudah menghasilkan akurasi yang cukup tinggi namun masih berfokus pada dimensi data yang rendah yang dapat memberikan potensi kehilangan informasi yang signifikan, kemungkinan terjadi *overfitting*, dan sulit diinterpretasikan. Sedangkan pada penelitian terkait data stunting berdimensi tinggi mayoritas menghasilkan akurasi yang masih rendah. Hal ini diperkuat dengan masih banyak ditemukan ketidakseimbangan kelas pada penelitian lainnya yang dapat berpengaruh terhadap nilai akurasi maupun *recall* pada performa model yang dibangun. **Tujuan Penelitian** yang diusulkan adalah penerapan algoritma SVM dengan seleksi fitur *Chi-Square* dan teknik SMOTE dalam mengatasi data stunting berdimensi tinggi yang dapat digunakan sebagai dasar pemerintah Kota Samarinda dalam membangun sistem cerdas yang dapat menghasilkan informasi dan pengetahuan sebagai alat dalam mengambil keputusan dan pertimbangan bagi pendeteksian stunting sejak dini. Data yang digunakan dalam penelitian ini bersumber dari Dinas Kesehatan Kota Samarinda yang terdiri dari 26 puskesmas. **Luaran** wajib yang dihasilkan dalam penelitian ini adalah hasil optimalisasi nilai akurasi dan atribut yang dominan dari evaluasi algoritma SVM dengan penerapan seleksi fitur *Chi-Square* dan SMOTE dalam menangani data stunting berdimensi tinggi yang dipublikasikan pada jurnal nasional terakreditasi sinta 2 di jurnal Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika.

### KATA KUNCI

Kata kunci maksimal 5 kata

Stunting;SVM;Chi-Square;SMOTE;Data Dimensi Tinggi.

## PENDAHULUAN

Penelitian Dasar merupakan riset yang memuat temuan baru atau pengembangan ilmu pengetahuan dari kegiatan riset yang terdiri dari tahapan penentuan asumsi dan dasar hukum yang akan digunakan, formulasi konsep dan/ atau aplikasi formulasi dan pembuktian konsep fungsi dan/ atau karakteristik penting secara analitis dan eksperimental.

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

- A. Latar belakang dan rumusan permasalahan yang akan diteliti
- B. Pendekatan pemecahan masalah
- C. *State of the art* dan kebaruan
- D. Peta jalan (*road map*) penelitian 5 tahun kedepan (jika dalam bentuk konsorsium harus dilengkapi dengan roadmap penelitian konsorsium)
- E. Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver

Stunting masih dianggap masalah serius dan menjadi fokus perhatian pemerintah dalam prioritas riset nasional tahun 2020-2024. Menurut data Kementerian Kesehatan Indonesia pada tahun 2022, prevalensi stunting pada anak balita di Indonesia masih tergolong tinggi yakni 21,6% (1). Dengan kata lain, stunting mempengaruhi 22 dari setiap 100 anak dibawah usia lima tahun. Meskipun prevalensi stunting telah menurun dari tahun ke tahun, prevalensi stunting di Indonesia juga masih termasuk tinggi jika dibandingkan dengan negara-negara ASEAN lainnya. Sebagai salah satu calon kota penyangga IKN, Samarinda masih memiliki prevalensi stunting tertinggi kedua setelah kabupaten Kutai Kartanegara di Provinsi Kalimantan Timur di tahun 2022 dengan persentase sebesar 25,3% (2). Stunting pada anak balita dapat berdampak buruk pada kesehatan dan perkembangan anak, bahkan dapat mempengaruhi kualitas hidupnya di masa depan. Oleh sebab itu, dibutuhkan upaya untuk melakukan prediksi status stunting pada anak balita menggunakan analisis data yang canggih seperti data mining dengan metode klasifikasi. Sehingga diharapkan dapat menghasilkan informasi dan pengetahuan sebagai alat dalam mengambil keputusan dan pertimbangan bagi pendeteksian stunting sejak dini.

Beberapa penelitian bidang data mining terkait prediksi status stunting sudah banyak dilakukan dengan berbagai pendekatan, mulai dari algoritma KNN (3,4), *Naïve Bayes* (5–7), *Random Forest* (8), *Neural Network* (9), CART (10), SVM (11) dan lain-lain. Melihat trend penelitian sebelumnya dengan penggunaan metode klasifikasi seperti algoritma KNN (3,4), *Naïve Bayes* (5,6), dan *Random Forest* (8), pada dasarnya sudah didapatkan akurasi yang cukup tinggi yakni dengan rata-rata persentase akurasi diatas 90%. Namun pada penelitian-penelitian tersebut masih berfokus pada dimensi data yang rendah. Data berdimensi rendah merupakan data yang terdiri dari jumlah variabel atau atribut yang sedikit (minim) yang digunakan untuk analisis. Sebagai contoh, jika suatu data hanya terdiri dari dua atau tiga variabel, maka data tersebut dapat dianggap sebagai data berdimensi rendah. Data berdimensi rendah juga memiliki kelemahan seperti kehilangan potensi informasi yang signifikan, kemungkinan terjadi *overfitting*, dan sulit diinterpretasikan (12).

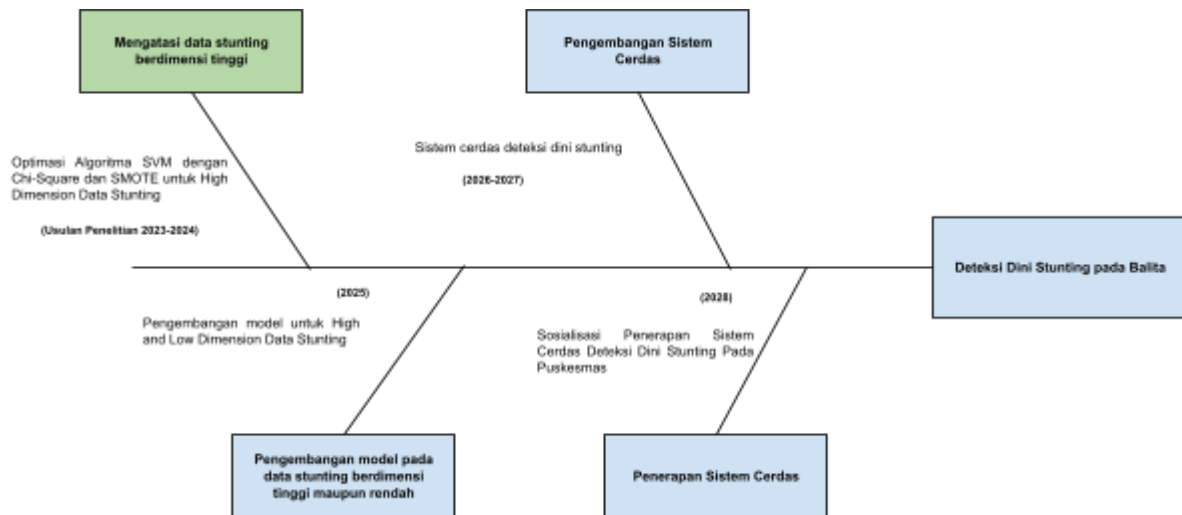
Data berdimensi tinggi merupakan data yang terdiri dari banyak variabel atau atribut yang digunakan untuk analisis. Sebagai contoh, jika suatu data memiliki puluhan atau bahkan ratusan variabel, maka data tersebut dapat dianggap sebagai data berdimensi tinggi. Data berdimensi tinggi memiliki beberapa tantangan dalam analisis data, seperti kompleksitas perhitungan, kecenderungan *overfitting*, dan kesulitan dalam visualisasi data (12). Oleh karena itu, teknik-teknik analisis data dengan dimensi tinggi telah dikembangkan untuk mengatasi masalah tersebut, seperti teknik reduksi dimensi dan teknik seleksi fitur. Beberapa penelitian terkait stunting data berdimensi tinggi (9,13,14) mayoritas masih menghasilkan akurasi yang lebih rendah dibandingkan dengan penelitian yang menggunakan data berdimensi rendah. Misalnya penelitian Jahidur dkk (14), hanya menghasilkan akurasi tertinggi sebesar 68.47%.

Penelitian ini akan mengumpulkan atribut-atribut yang pernah digunakan pada penelitian-penelitian sebelumnya sebagai referensi kelengkapan data yang kompleks, ditunjang data dari penelitian (13,14) seperti umur (dalam bulan), jenis kelamin, status pemberian ASI pada anak, pemberian vitamin A (dalam 6 bulan terakhir), nomor urutan kelahiran, diare pada anak (dalam 2 pekan terakhir), demam anak (dalam 2 pekan terakhir), usia ibu, dan usia ayah. Sedangkan pada penelitian lain (9) seperti berat anak (saat lahir), tinggi anak (saat lahir), polio 1, polio 2, polio 3, berat badan menurut umur pada anak, dan panjang badan menurut umur pada anak dapat dijadikan bahan kajian lain yang bisa dipertimbangkan. Kumpulan atribut tersebut dihimpun untuk membentuk data dengan dimensi yang tinggi. Data yang digunakan dalam penelitian ini bersumber dari Dinas Kesehatan Kota Samarinda yang terdiri dari 26 puskesmas.

Algoritma klasifikasi yang digunakan dalam penelitian ini yakni *Support Vector Machine* (SVM), hal ini mengacu pada penelitian Maftahul dkk. (11) dimana SVM memiliki performa terbaik jika dibandingkan dengan algoritma *Naïve Bayes* dan KNN terkait klasifikasi data stunting, ditunjang pendapat Widiarni dkk. (10) yang menyarankan penggunaan algoritma klasifikasi selain CART agar mendapatkan performa terbaik dalam menangani data stunting yang berdimensi tinggi. Penggunaan seleksi fitur Chi-Square juga akan digunakan untuk menemukan atribut-atribut yang relevan dalam proses klasifikasi. Berdasarkan penelitian Maftahul dkk. (11) *Chi-Square* dianggap memberikan peningkatan akurasi lebih tinggi terhadap algoritma klasifikasi jika dibandingkan dengan algoritma seleksi fitur ANOVA, dan mRMR. Dalam beberapa kasus penelitian sebelumnya juga masih banyak ditemukan ketidakseimbangan kelas seperti pada penelitian (3,5,6), hal itu akan berpengaruh terhadap nilai akurasi maupun *recall* pada performa model yang dibangun. Oleh sebab itu penelitian ini juga akan menggunakan teknik SMOTE dalam mengatasi ketidakseimbangan kelas agar performa model yang didapatkan dapat optimal.

Dari beberapa permasalahan tersebut maka, dalam penelitian ini akan mengangkat judul “Optimasi Algoritma SVM dengan Chi-Square dan SMOTE untuk High Dimension Data Stunting”. Hal ini bertujuan untuk mengoptimasi algoritma SVM dalam mengatasi data berdimensi tinggi khususnya pada data stunting di Kota Samarinda. Penelitian ini nantinya akan menghasilkan model terbaik yang dapat digunakan sebagai dasar pemerintah Kota Samarinda dalam membangun sistem cerdas pencegahan dini stunting pada balita. Dengan melakukan penelitian ini diharapkan juga akan membantu upaya pemerintah Indonesia dalam mencegah dan mengurangi tingkat prevalensi stunting pada balita di Indonesia.

## Road Map Penelitian



Gambar 1. Road Map Penelitian

Gambar 1 menunjukkan roadmap penelitian atau posisi penelitian yang diusulkan saat ini. Usulan penelitian tahun ini merupakan penelitian awal dari rencana penelitian dalam 5 tahun kedepan. Adapun penelitian tahun ini dan tahun selanjutnya dapat dikategorikan sebagai 4 objek penelitian:

1. Penanganan data stunting berdimensi tinggi.
2. Kajian pengembangan model pada data stunting berdimensi tinggi.
3. Pengembangan sistem cerdas untuk mendeteksi stunting secara dini.
4. Penerapan sistem cerdas untuk mendeteksi stunting secara dini.

Perjalanan penelitian yang diusulkan merupakan rangkaian pelaksanaan penelitian selama 5 tahun kedepan, penelitian yang diusulkan saat ini akan membahas mengenai tahapan awal dari proses pengembangan sistem deteksi dini penyakit stunting pada balita kedepannya. Penelitian tahun awal akan membahas mengenai optimalisasi algoritma SVM dengan tambahan seleksi fitur *Chi-Square* dan SMOTE untuk menangani data stunting berdimensi tinggi atau High Dimension Data Stunting. Pada penelitian tahun kedua, fokus penelitian akan dilanjutkan pada pengembangan pemodelan untuk menangani data stunting dengan dimensi yang lebih bervariasi lagi dari tahun sebelumnya, yaitu pada data stunting berdimensi tinggi maupun berdimensi rendah. Selanjutnya, pada penelitian tahun ketiga dan keempat akan dilakukan pengembangan sistem cerdas berbasis kerangka pemodelan prediktif terhadap data penyakit stunting yang telah dilakukan pada penelitian tahun sebelumnya. Terakhir, pada tahun penelitian kelima akan dilakukan penerapan sistem cerdas untuk mendeteksi penyakit stunting pada instansi kesehatan yang ada di kota Samarinda. Usulan penelitian ini diharapkan dapat terus dilanjutkan hingga terbangun sebuah sistem cerdas yang bermanfaat untuk membantu instansi kesehatan khususnya di kota Samarinda dalam mendeteksi penyakit stunting pada balita sedini mungkin.

## METODA

Metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan yang tercermin dalam Rencana Anggaran Biaya (RAB).

### 1. Metode Penelitian

Terdapat beberapa tahapan yang perlu dilakukan dalam penelitian ini dalam menyelesaikan permasalahan yang ada, meliputi : identifikasi masalah, studi pustaka, pengumpulan dataset, *data preparation*, seleksi fitur, pemodelan, evaluasi performa dan kesimpulan. Tahapan penelitian tergambar pada gambar 2.

#### Identifikasi Masalah dan Studi Pustaka

Tahap awal dalam melakukan penelitian, identifikasi masalah perlu dilakukan terlebih dahulu karena perumusan masalah akan membimbing seluruh proses penelitian. Dalam penelitian prediksi stunting pada balita di Kota Samarinda, masalah yang diangkat adalah mencari metode terbaik untuk memprediksi stunting. Selain itu, studi pustaka juga dilakukan untuk mencari informasi terkait gap penelitian pada penelitian-penelitian sebelumnya yang berkaitan dengan prediksi stunting pada balita.

#### Pengumpulan data (dataset)

Data yang digunakan berasal dari Dinas Kesehatan Kota Samarinda yang mencakup faktor-faktor seperti umur, jenis kelamin, status gizi, dan faktor lingkungan seperti ketersediaan air bersih dan sanitasi. Dalam penelitian ini, terdapat total 20 parameter atau atribut yang digunakan untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap stunting pada balita di Kota Samarinda.

Tabel 1. Atribut data stunting

No	Atribut	Tipe data	Keterangan
1	NIK	<i>string</i>	Nomor Induk Kependudukan
2	Nama	<i>string</i>	Nama
3	JK	<i>string</i>	Jenis Kelamin
4	Tgl Lahir	<i>date</i>	Tanggal Lahir
5	Nama Ortu	<i>string</i>	Nama Orang Tua
6	Provinsi	<i>string</i>	Provinsi
7	Kab/Kota	<i>string</i>	Kabupaten atau Kota
8	Kec	<i>string</i>	Kecamatan
9	Puskesmas	<i>string</i>	Lokasi Puskesmas
10	Posyandu	<i>string</i>	Lokasi Posyandu
11	Total Pengukuran	<i>integer</i>	Total Pengukuran
12	Tanggal Pengukuran	<i>integer</i>	Tanggal Pengukuran
13	Berat	<i>integer</i>	Berat Badan
14	Tinggi	<i>integer</i>	Tinggi Badan
15	BB/U	<i>numeric</i>	Berat Badan menurut Umur
16	ZS BB/U	<i>numeric</i>	Z Score Berat Badan menurut Umur
17	TB/U	<i>numeric</i>	Tinggi Badan menurut Umur
18	ZS TB/U	<i>numeric</i>	Z Score Tinggi Badan menurut Umur
19	BB/TB	<i>numeric</i>	Berat Badan menurut Tinggi Badan
20	ZS BB/TB	<i>numeric</i>	Z Score Berat Badan menurut Tinggi Badan

## Data Preparation

Tahap ini meliputi beberapa proses yang dilakukan seperti data *selection*, data *transformation*, data *cleaning*. Beberapa tahapan tersebut dilakukan agar data yang digunakan saat proses analisis data memiliki kualitas yang baik.

### Imbalance Data

Teknik SMOTE akan digunakan dalam mengatasi ketidakseimbangan kelas yang terdapat pada data stunting. Dalam teknik SMOTE sendiri ada beberapa metode yang digunakan seperti *Undersampling*, dan *Oversampling*. Pada penelitian ini akan menggunakan teknik *Oversampling*.

### Seleksi Fitur

Tahap seleksi fitur dilakukan untuk mengoptimalkan atribut-atribut yang digunakan dalam pemodelan, pada tahap ini juga dilakukan perangkingan terhadap atribut, mulai dari atribut yang memiliki pengaruh tertinggi terhadap hasil prediksi dan atribut yang memiliki pengaruh rendah atau bahkan tidak memiliki pengaruh terhadap hasil prediksi. Pada penelitian ini akan menggunakan seleksi fitur *Chi-Square* untuk optimasi algoritma SVM dalam prediksi status stunting balita.

### Pembagian data training dan testing

Dalam tahap ini dataset akan dibagi menjadi dua bagian yakni data *training* dan *data testing*. Data training digunakan untuk melatih model klasifikasi agar memiliki pengetahuan terhadap dataset yang akan digunakan, sedangkan data testing digunakan untuk memvalidasi model klasifikasi yang telah dibangun dengan melihat akurasi yang didapatkan dalam penerapan model tersebut. Pada penelitian ini akan menggunakan teknik *K-Fold Cross Validation* dalam pembagian data, sedangkan penggunaan nilai *k* akan dilakukan iterasi untuk mendapatkan performa terbaik.

### Pemodelan

Algoritma yang digunakan dalam tahap pemodelan yakni SVM (*Support Vector Machine*). Implementasi algoritma SVM menggunakan bantuan *library* sklearn yang terdapat pada bahasa pemrograman python.

### Evaluasi Performa

Untuk menguji dan menghitung tingkat akurasi model algoritma SVM, adalah dengan menggunakan teknik perhitungan *confusion matrix*.

Tabel 2. *Confusion matrix*

		<i>True values</i>	
		<i>True</i>	<i>False</i>
<i>Prediction</i>	<i>True</i>	TP <i>Correct result</i>	FP <i>Unexpected result</i>
	<i>False</i>	FN <i>Missing result</i>	TN <i>Correct absence of result</i>

Sumber: Pulungan (15)



Dalam *confusion matrix*, terdapat beberapa istilah yang digunakan dalam kasus klasifikasi seperti yang dijelaskan oleh Pulungan (15):

1. *True Positive* (TP): Merupakan data positif pada *confusion matrix* yang terdeteksi dengan benar.
2. *False Positive* (FP): Merupakan data negatif pada *confusion matrix* yang salah terdeteksi sebagai data positif.
3. *False Negative* (FN): Merupakan data positif pada *confusion matrix* yang salah terdeteksi sebagai data negatif.
4. *True Negative* (TN): Merupakan data negatif pada *confusion matrix* yang terdeteksi dengan benar.

Dalam teknik ini performa yang dihasilkan adalah nilai *accuracy*, *recall*, *precision*, dan *f-measure*. Berikut adalah perhitungan *accuracy*, *recall*, *precision*, dan *f-measure* yang akan digunakan pada penelitian ini:

1. *Accuracy*

$$accuracy = \frac{TP + TN}{n}$$

2. *Recall*

$$recall = \frac{TP}{TP + FN}$$

3. *Precision*

$$precision = \frac{TP}{TP + FP}$$

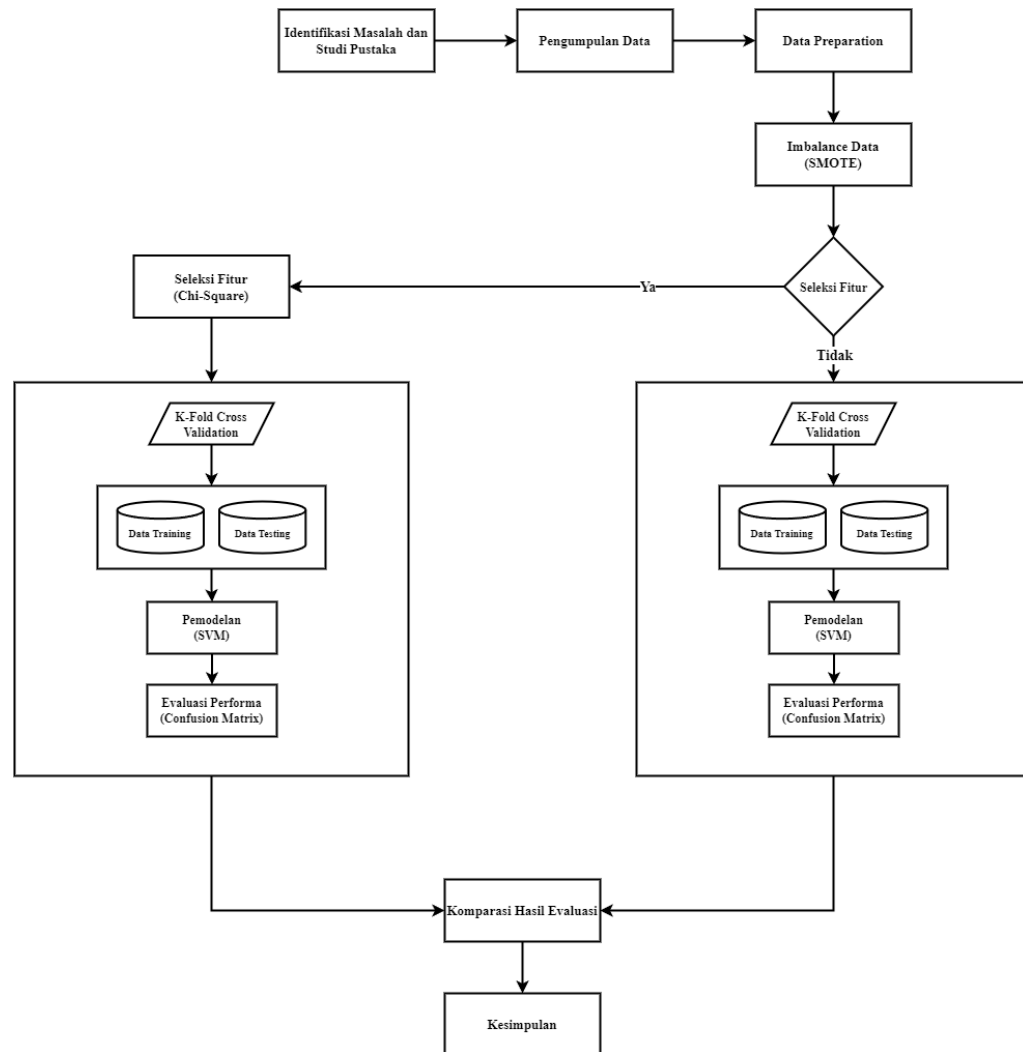
4. *F-Measure* ( $F_1$ )

$$F_1 = \frac{precision \times recall}{precision + recall}$$

## Kesimpulan

Tahap kesimpulan dalam sebuah penelitian adalah tahap akhir di mana peneliti menyajikan hasil penelitian dan menyimpulkan temuan-temuan yang telah ditemukan berdasarkan analisis data yang telah dilakukan.

## 2. Diagram Alir Penelitian



Gambar 2. Diagram Alir Penelitian

Setiap anggota tim bekerja sama dalam menjalankan tugas-tugas umum dan membagi tugas khusus kepada anggota tertentu. Tugas umum usulan penelitian ini meliputi menyusun proposal penelitian, mengumpulkan data, menyusun laporan kemajuan dan laporan akhir, serta membuat artikel untuk publikasi. Sementara itu, setiap anggota tim memiliki tanggung jawab khusus yang terdiri dari hal-hal berikut:

1. **Ketua Pengusul:** Menganalisis dan melakukan studi literatur permasalahan penelitian pada kasus data stunting berdimensi tinggi di kota Samarinda secara komprehensif, menganalisis dan mengevaluasi perbandingan hasil kinerja algoritma SVM dengan dan tanpa menggunakan seleksi fitur, menyusun laporan akhir (100%), serta melakukan koordinasi kegiatan penelitian.
2. **Anggota Pengusul:** Melakukan pengumpulan data stunting kota Samarinda, melakukan pengamatan terhadap dimensi data stunting yang diperoleh, menganalisis penerapan dan pengujian teknik seleksi fitur *Chi-Square* dan SMOTE untuk mengoptimasi algoritma SVM, menyusun laporan pelaksanaan (70%).
3. **Mahasiswa Bimbingan:** Membantu dalam pengamatan dimensi data stunting yang diperoleh, membantu proses pembersihan data stunting, membantu penyusunan laporan pelaksanaan (70%) dan laporan akhir (100%).

## JADWAL PENELITIAN

Jadwal penelitian disusun berdasarkan pelaksanaan penelitian, harap disesuaikan berdasarkan lama tahun pelaksanaan penelitian

No	Nama Kegiatan	Bulan											
		1	2	3	4	5	6	7	8	9	10	11	12
1	Pengumuman, Kontrak												
2	Administrasi surat menyurat dan perizinan												
3	Pengumpulan data												
4	Preprocessing Data												
5	Pembuatan dan Pengujian Model												
6	Penyusunan Hasil Penelitian												
7	Evaluasi dan Diskusi hasil penelitian												
8	Publikasi Jurnal Nasional												
9	Penyelesaian Laporan Penelitian												
10	Monitoring dan Evaluasi												

## DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

- [1] Annur CM. Daftar prevalensi balita stunting di indonesia pada 2022, provinsi mana Teratas? [Internet]. Katadata. 2023 [cited 2023 Apr 9]. Available from: <https://databoks.katadata.co.id/datapublish/2023/02/02/daftar-prevalensi-balita-stunting-di-indonesia-pada-2022-provinsi-mana-teratas>
- [2] Annur CM. Calon Ibu Kota Baru, Bagaimana Angka Balita Stunting di Wilayah di Kalimantan Timur? [Internet]. Katadata. 2023 [cited 2023 Apr 11]. Available from: <https://databoks.katadata.co.id/datapublish/2023/02/27/calon-ibu-kota-baru-bagaimana-angka-balita-stunting-di-wilayah-di-kalimantan-timur>
- [3] Lonang S, Normawati D. Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. Jurnal Media Informatika Budidarma. 2022;6(1):49.
- [4] Prasetya T, Ali I, Rohmat CL, Nurdiawan O. Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor. INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics. 2020;5(1):93.
- [5] Wahyudin WC, Hana FM, Prihandono A, Kudus UM, No JG, Email I, et al. Prediksi Stunting Pada Balita di Rumah Sakit Kota Semarang Menggunakan Naive Bayes. Jurnal Ilmu Komputer dan Matematika. 2023;4(1):32–6.
- [6] Arisandi RRR, Warsito B, Hakim AR. Aplikasi Naïve Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation. Jurnal Gaussian. 2022;11(1):130–9.

- [7] Sutarno HH, Latuconsina R, Dinimaharawati A. Prediksi Stunting Pada Balita Dengan Menggunakan Algoritma Klasifikasi Naive Bayes. e-Proceeding of Engineering [Internet]. 2021;8(5):6657–61. Available from: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16456/16167>
- [8] Perdana AY, Latuconsina R, Dinimaharawati A. Prediksi Stunting Pada Balita Dengan Menggunakan Algoritma Random Forest. e-Proceeding of Engineering. 2021;8(5):6650–6.
- [9] Purbasari A, Rinawan FR, Zulianto A, Susanti AI, Komara H. CRISP-DM for Data Quality Improvement to Support Machine Learning of Stunting Prediction in Infants and Toddlers. Proceedings - 2021 8th International Conference on Advanced Informatics: Concepts, Theory, and Application, ICAICTA 2021. 2021;(MI).
- [10] Sasmita WG, Nugroho WH, Astuti AB. Cart Method Approach and High Dimension Simulation Data Selection and Random Under-Sampling Method in Stunting Case. J Theor Appl Inf Technol. 2022;100(24):4810–7.
- [11] Hakimah M, Prabiantissa CN, Rozi NF, Yamani LN, Puspitasari I. Determination of Relevant Feature Combinations For Detection Stunting Status of Toddlers. In: International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) Determination. IEEE; 2022. p. 324–9.
- [12] Velliangiri S, Alagumuthukrishnan S, Thankumar Joseph SI. A Review of Dimensionality Reduction Techniques for Efficient Computation. In: Procedia Computer Science. Elsevier B.V.; 2019. p. 104–11.
- [13] Ndagijimana S, Kabano IH, Masabo E, Ntaganda JM. Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques. J Prev Med Public Health. 2023 Jan 1;56(1):41–9.
- [14] Khan JR, Tomal JH, Raheem E. Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh. Inform Health Soc Care. 2021;46(4):425–42.
- [15] Pulungan AF. Analisis Kinerja Bray Curtis Distance, Canberra Distance dan Euclidean Distance pada Algoritma K-Nearest Neighbour. [Internet]. 2019. Tersedia dari: <https://repositori.usu.ac.id/handle/123456789/15051> [Diakses 28 Januari 2022].