TUGAS AKHIR - KS184822

ANALISIS PERILAKU REMAJA MELAKUKAN SEKS PRA-NIKAH DI JAWA TIMUR MENGGUNAKAN CART DENGAN SMOTE-N-ENN DAN ADASYN-N (Analisis Lanjut SKAP Jawa Timur 2018)

KINANTHI SUKMA WENING
NRP 062116 4000 0044

Dosen Pembimbing
Dr. Dra. Kartika Fithriasari, M.Si.
Dr. Dra. Iswari Hariastuti, M.Kes.

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

TUGAS AKHIR – KS184822

ANALISIS PERILAKU REMAJA MELAKUKAN SEKS PRA-NIKAH DI JAWA TIMUR MENGGUNAKAN CART DENGAN SMOTE-N-ENN DAN ADASYN-N (Analisis Lanjut SKAP Jawa Timur 2018)

KINANTHI SUKMA WENING
NRP 062116 4000 0044

Dosen Pembimbing
Dr. Dra. Kartika Fithriasari, M.Si.
Dr. Dra. Iswari Hariastuti, M.Kes.

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

FINAL PROJECT – KS184822

ANALYSIS OF PREMARITAL SEX AMONG ADOLESCENT IN EAST JAVA USING CART WITH SMOTE-N-ENN AND ADASYN-N (Analysis of SKAP Jawa Timur 2018)

KINANTHI SUKMA WENING
SN 062116 4000 0044

Supervisors
Dr. Dra. Kartika Fithriasari, M.Si.
Dr. Dra. Iswari Hariastuti, M.Kes.

UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE AND DATA ANALYTICS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

LEMBAR PENGESAHAN

ANALISIS PERILAKU REMAJA MELAKUKAN SEKS PRA-NIKAH DI JAWA TIMUR MENGGUNAKAN CART DENGAN SMOTE-N-ENN DAN ADASYN-N
(Analisis Lanjut SKAP Jawa Timur 2018)

TUGAS AKHIR
Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada
Program Studi Sarjana Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh :
Kinanthi Sukma Wening
NRP. 062116 4000 0044

Disetujui oleh Pembimbing:

Dr. Dra. Kartika Fithriasari, M.Si.
NIP. 19691212 199303 2 002 ( )

Dr. Dra. Iswari Hariastuti, M.Kes.
NIP. 19661111 199203 2 008 ( )

Mengetahui,
Kepala Departemen Statistika

Dr. Dra. Kartika Fithriasari, M.Si.
NIP. 19691212 199303 2 002

SURABAYA, JANUARI 2020

ANALISIS PERILAKU REMAJA MELAKUKAN SEKS PRA-NIKAH DI JAWA TIMUR MENGGUNAKAN CART DENGAN SMOTE-N-ENN DAN ADASYN-N (Analisis Lanjut SKAP Jawa Timur 2018)

Identitas Mahasiswa	**:**	**Kinanthi Sukma Wening**
NRP	**:**	**062116 4000 0044**
Departemen	**:**	**Statistika**
Dosen Pembimbing I	**:**	**Dr. Dra. Kartika Fithriasari, M.Si.**
Dosen Pembimbing II	**:**	**Dr. Dra. Iswari Hariastuti, M.Kes.**

Abstrak

Pembahasan mengenai seks pra-nikah sering dianggap tabu, padahal efek yang ditimbulkan sangat besar. Di Jawa Timur, angka seks pra-nikah pada 2018 naik 1,4%, yakni dari 0,2% menjadi 1,6% (Badan Kependudukan dan Keluarga Berencana Nasional, 2018). Untuk menekan angka tersebut, perlu diketahui faktor yang melatarbelakanginya. Salah satu metode yang dapat digunakan adalah analisis klasifikasi. Namun, persentase yang kecil dan tidak seimbang akan membuat klasifikasi hanya menuju kelas mayoritas. Oleh karena itu, perlu dilakukan penanganan class imbalance. Salah satu metode yang sering digunakan pada data nominal adalah SMOTE-N, kemudian dikembangkan menjadi hybrid SMOTE-N hingga ADASYN-N, yang memberikan kinerja klasifikasi paling baik menggunakan SVM (Peters, 2018). Pada penelitian ini akan dibandingkan SMOTE-N, SMOTE-N-ENN dan ADASYN-N dalam mengatasi class imbalance pada klasifikasi remaja melakukan seks pra-nikah di Jawa Timur menggunakan base classifier CART. Data yang digunakan merupakan hasil SKAP BKKBN Jawa Timur 2018. Hasil analisis memberikan kesimpulan bahwa penanganan class imbalance terbaik adalah ADASYN-N, yang memberikan AUC tertinggi dibandingkan kedua metode lainnya. Selain itu, faktor yang dapat memprediksi remaja melakukan hubungan seksual pra-nikah adalah jenis kelamin, pendidikan terakhir, daerah tempat tinggal, gaya pacaran, pengetahuan mengenai masa subur, dan pengetahuan resiko menikah muda.

Kata kunci: ADASYN-N, CART, Premarital Sex, Remaja, SMOTE-N-ENN

(Halaman ini sengaja dikosongkan)

ANALYSIS OF PREMARITAS SEX AMONG ADOLESCENT IN EAST JAVA USING CART WITH SMOTE-N-ENN AND ASAYN-N (Analysis of SKAP Jawa Timur 2018)

Name	**:**	**Kinanthi Sukma Wening**
Student Number	**:**	**062116 4000 0044**
Department	**:**	**Statistics**
Supervisor I	**:**	**Dr. Dra. Kartika Fithriasari, M.Si.**
Supervisor II	**:**	**Dr. Dra. Iswari Hariastuti, M.Kes.**

Abstract

Premarital sex often considered taboo, eventhough the effects are very large. In East Java, the number of premarital sex in 2018 rose 1,4%, from 0.2% to 1,6% (Badan Kependudukan dan Keluarga Berencana Nasional, 2018). To ensure pre-marital sex rates among adolescents in East Java, it is important to consider the underlying factors. Classification analysis can be used to find out the factors that can predict premarital sex among adolescent. However, a small and unbalanced percentage will make the classification only predict into majority class. Therefore, it is necessary to deal with class imbalance by resampling (under-sampling and oversampling). One of popular method to deal with imbalance in nominal predictor is SMOTE-N, then developed into hybrid SMOTE-N until ADASYN-N, the best method using SVM (Peters, 2018). This study will compare SMOTE-N, SMOTE-N-ENN and ADASYN-N to deal with class imbalances in the classification of adolescent premarital sex using the base classifier CART. Data in this study based on SKAP BKKBN Jawa Timur 2018. The result say that ADASYN-N is the best method to deal with imbalance due to the highest AUC, compare with two other methods. The best decision tree provides information about factors that can predict premarital sex adolescents, they are sex, last education, residential area, dating style, knowledge about the fertility period, and knowledge about risk of getting married early.

Keywords: ADASYN-N, CART, Premarital Sex, Remaja, SMOTE-N-ENN

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan atas rahmat dan hidayah yang diberikan Allah SWT sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul "Analisis Perilaku Remaja Melakukan Seks Pra-Nikah di Jawa Timur Menggunakan CART dengan SMOTE-N-ENN dan ADASYN-N" dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Dr. Dra. Kartika Fithriasari, M.Si. selaku Ketua Departemen Statistika dan Dr. Santi Wulan Purnami, S.Si., M.Si. selaku Sekretaris Departemen Bidang Akademik yang telah memberikan fasilitas, sarana, dan prasarana.
2. Dr. Purhadi, M.Sc. selaku dosen wali yang telah memberikan saran, dukungan, arahan, nasihat selama proses belajar di Departemen Statistika ITS.
3. Dr. Dra. Kartika Fithriasari, M.Si. dan Dr. Dra. Iswari Hariasturi, M.Kes. selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan, serta motivasi selama penyusunan Tugas Akhir.
4. Irhamah, M.Si., Ph.D. dan Dr. Santi Wulan Purnami, S.Si., M.Si. selaku dosen penguji yang selalu sabar dalam mengomentari serta memberikan masukan dan saran dalam penyelesaian Tugas Akhir.
5. Seluruh dosen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS, khususnya Bapak Pendi dan Bapak Umam yang selalu siap siaga membantu dalam administrasi.
6. Kedua orang tua, Bapak dan Ibu, Ajeng serta Kakung dan Uti atas segala do'a, nasehat, kasih sayang, dan dukungan

yang diberikan kepada penulis demi kesuksesan dan kebahagiaan penulis.

7. Sahabat-sahabat Bunga Matahari : Inan, Frans, Cahya, Erika, Riris, dan Thalia yang selalu memberikan dukungan kepada penulis.
8. Geng Kosan Cantik Lantai 1 : Rivi, Marita, dan Rezki yang selalu menjadi pendengar dan penyemangat dalam mengerjakan Tugas Akhir.
9. *Statistics Computer Course* (SCC) 17/18 dan 18/19, khususnya *Training Development*, dan Kementerian Perkonomian BEM ITS Gelora Aksi dan Kolaborapi yang telah menjadi keluarga serta memberikan pembelajaran berorganisasi.
10. Teman-teman Statistika ITS ∑27 angkatan 2016, yang selalu memberikan dukungan kepada penulis selama ini.
11. Semua teman, relasi dan berbagai pihak yang tidak bisa penulis sebutkan identitasnya satu persatu yang telah membantu dalam penulisan laporan ini.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Januari 2020

Penulis

DAFTAR ISI

Halaman

DAFTAR GAMBAR

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

(Halaman ini sengaja dikosongkan)

DAFTAR SINGKATAN

SKAP	: Survei Kinerja dan Akuntabilitas KKBPK
KKBPK	: Kependudukan Keluarga Berencana dan Pembangunan Keluarga
BKKBN	: Badan Kependudukan dan Keluarga Berencana Nasional
IMS	: Infeksi Menular Seksual
HIV/AIDS	: Human Immunodeficiency Virus/ Acquired Immuno Deficiency Syndrome

(Halaman ini sengaja dikosongkan)

DAFTAR NOTASI

$g(t)$	: indeks gini simpul *t*
$p(i \mid t)$	: proporsi kelas *i* pada sumpul *t*
$\phi(s,t)$	: nilai *goodness of split s* pada simpul *t*
$R(T)$	: ukuran kesalahan klasifikasi pohon T
$\tilde{T}$	: banyak simpul terminal pada pohon T
N_{ij}	: jumlah data kelas *j* yang diprediksi menjadi kelas *i*
$\delta(V_1,V_2)$	: jarak antar kategori V_1 dan V_2
$\delta(x_b,y_b)$	: jarak antar kategori variabel *b* pada observasi X dan Y
$\Delta(X,Y)$	: jarak antar observasi X dan Y
d	: *degree of imbalance*
m_l	: jumlah data kelas mayoritas
m_s	: jumlah data kelas minoritas
H_i	: jumlah data terdekat yang termasuk kelas minoritas pada data minoritas ke-*i*
r_i	: dominasi kelas mayoritas pada data minoritas ke-*i*
$\hat{r}_i$	: normalisasi dominasi kelas mayoritas pada data minoritas ke-*i*

(Halaman ini sengaja dikosongkan)

BAB I
PENDAHULUAN

1.1 Latar Belakang

Menurut WHO, masa remaja adalah masa transisi dari kanak-kanak menjadi dewasa, yakni antara usia 12-13 tahun hingga sekitar 20 tahun. Menurut Peraturan Menteri Kesehatan RI Nomor 2005 tahun 2014, remaja adalah penduduk dalam rentang usia 10-18 tahun. Pada masa remaja, terjadi perubahan drastis pada semua aspek perkembangan, meliputi perkembangan fisik, kognitif, kepribadian, dan sosial. Salah satu perubahannya adalah pada gaya hidup dan pergaulan. Kombinasi antara usia perkembangan dan dinamisasi lingkungan sering kali membuat remaja masuk dalam lingkungan yang tidak sesuai.

Ada berbagai lingkungan yang mungkin dimasuki remaja pada masanya. Jika berada pada lingkungan yang baik, maka remaja akan terarahkan kepada hal-hal positif yang menunjang untuk masa depan. Namun sebaliknya, remaja juga dapat masuk ke dalam lingkungan yang negatif dan beresiko bagi masa depannya. Lingkungan tersebut merupakan lingkungan yang banyak ditemui dan memiliki daya tarik yang kuat terhadap remaja. Akibatnya remaja akan terjebak dalam pergaulan tidak sehat, seperti mengkonsumsi obat-obatan terlarang, pergaulan bebas, dan lain-lain, yang disebut kenakalan remaja.

Definisi kenakalan remaja ialah perilaku jahat, atau kenakalan anak-anak muda yang merupakan gejala sakit (patologis) secara sosial pada anak-anak dan remaja yang disebabkan oleh satu bentuk pengabaian sosial, sehingga mengembangkan bentuk tingkah laku yang menyimpang. Kenakalan remaja disebabkan kegagalan mereka dalam memperoleh penghargaan dari masyarakat tempat mereka tinggal (Willis & Sofyan, 2005). Bentuk kenakalan remaja sangat bervariasi, mulai pada tingkatan yang dapat ditoleransi hingga yang di luar nalar. Kenakalan remaja pun berubah dari waktu ke waktu. Mulai dari tawuran pelajar dan balap motor hingga saat ini

menjadi konsumsi narkoba. Perubahan kultur dan gaya hidup menyebabkan pergeseran pola kenakalan remaja. Namun, salah satu isu dan bentuk kenakalan yang selalu ada dan menjadi makin parah adalah kenakalan seksual. Kenakan ini muncul karena rasa ingin tahu remaja mengenai permasalahan seksual namun tidak diimbangi dengan pengetahuan yang sesuai.

Pada pertengahan November 2018 beredar sebuah video mesum pelajar SMA di Karawang (Merdeka.com, 2018). Adanya video yang tidak selayaknya beredar menjadi tanda meluasnya kenakalan remaja yang sangat serius, yakni hubungan seks pra-nikah. Berdasarkan survei yang dilakukan oleh Komite Perlindungan Anak Indonesia (KPAI), dan Kementrian Kesehatan, (Kemenkes) pada Oktober 2013, tercatat bahwa 62,7% remaja di Indonesia telah melakukan hubungan seks pra-nikah. Survei serupa juga pernah dilakukan pada remaja 14-18 tahun di kota-kota besar (Jakarta, Surabaya, Bandung), dan tercatat 32% remaja pernah melakukan hubungan seks pra nikah. Di Provinsi Jawa Timur, angka remaja melakukan seks pra-nikah pada tahun 2018 juga naik 1,4% dari tahun sebelumnya (Badan Kependudukan dan Keluarga Berencana Nasional, 2018). Persentase tersebut naik dari 0,2% menjadi 1,6%. Bertambahnya persentase remaja yang melakukan seks-pra nikah menandakan kenakalan seksual di Jawa Timur tidak dapat diabaikan. Namun permasalahan seksual pada remaja sering kali dianggap tabu, padahal dampak yang ada sangat besar. Dari sisi kesehatan misalnya, hubungan seks yang dilakukan sebelum usia 17 tahun memberikan risiko terkena penyakit mencapai empat hingga lima kali lipat (Kasim, 2014). Efek yang dapat ditimbulkan adalah terjangkitnya HIV/AIDS, *sexually transmitted infection* (STIs), maupun kehamilan tidak diinginkan (Naroozi, dkk, 2014)

Penelitian mengenai *premarital sex* pada remaja telah dilakukan sebelumnya. Berdasarkan penelitian yang dilakukan oleh Dave, *et al*. (2013), didapatkan bahwa sebagian besar partner dalam melakukan penyimpangan adalah pacar. Selain itu, remaja dengan tingkat ekonomi menengah ke bawah cenderung melakukan hubungan seks dan sering kali berganti pasangan.

Faktor eksternal yakni pengaruh teman dan paparan pornografi juga berpengaruh signifikan terhadap perilaku remaja melakukan hubungan seks pra-nikah (Harnani, dkk, 2018). Tidak jauh berbeda dengan kedua penelitian di atas, Teferra, dkk. (2015) juga pernah melakukan penelitian mengenai faktor-faktor mempengaruhi mahasiswa ilmu kesehatan Universitas Madawalabu, Bale Goba, Ethiopia Tenggara. Hasil yang diperoleh adalah kemungkinan remaja melakukan hubungan seksual pra-nikah semakin tinggi jika tinggal jauh dari orang tua, merokok, dan terpapar pornografi.

Beberapa penelitian mengenai perilaku seks pra-nikah pada remaja juga pernah dilakukan di Indonesia. Triningsih, dkk (2015) pernah melakukan penelitian mengenai faktor-faktor yang berpengaruh terhadap praktik seks pra-nikah remaja SMA daerah eks lokalisasi di Kabupaten Malang. Faktor yang didapatkan berpengaruh adalah *self esteem*, praktik religius, dan pengetahuan mengenai IMS dan HIV/AIDS. Sebagian besar penelitian tersebut merupakan penelitian dalam bidang kesehatan, dan analisis yang digunakan adalah metode kualitatif serta *cross sectional study*. Selain metode kualitatif, metode klasifikasi juga dapat digunakan untuk mengetahui faktor-faktor yang berpengaruh terhadap perilaku remaja melakukan seks pra-nikah. Salah satu penelitian yang menggunakan analisis klasifikasi adalah Rosdarni, dkk. (2015). Pada penelitiannya, Rosdarni, dkk. menggunakan regresi logistik dan mengatakan bahwa faktor yang mempengaruhi remaja melakukan seks pra-nikah adalah pengetahuan mengenai kesehatan seksual, pengetahuan IMS dan HIV / AIDS, sikap permisif terhadap seksualitas, harga diri, dan efikasi diri.

Selain regresi logistik, metode klasifikasi yang sering digunakan dan mudah dipahami adalah *decision tree* algoritma CART (*Classification and Regression Trees*) dan Naïve Bayes. Adiansyah (2017) melakukan penelitian dan memberikan kesimpulan bahwa berdasarkan kurva *Receiver Operating Characteristic* (ROC), metode CART memiliki kinerja yang lebih baik dibanding model regresi logistik. Penelitian serupa juga pernah dilakukan oleh Waluyo, dkk. (2014). Pada penelitiannya,

Waluyo membandingkan regresi logistik dan CART pada klasifikasi nasabah kredit yang mana CART memberikan ukuran klasifikasi lebih baik daripada regresi logistik. Kemudian jika dibandingkan antara naïve bayes dan CART, keduanya memberikan hasil klasifikasi yang hampir sama, namun CART lebih cocok untuk data dengan skala besar (Jeyarani, dkk., 2013). Oleh karena itu, pada penelitian kali ini digunakan metode klasifikasi CART. CART merupakan salah satu algoritma *decision tree* yang diperkenalkan oleh Breiman, *et al*. (1993). Konsep CART adalah sebuah pohon keputusan yang berawal dari simpul induk, memecah menjadi simpul anak, hingga dilabelkan pada suatu kelas. CART mempunyai beberapa kelebihan dibandingkan metode klasifikasi lainnya, yaitu hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya, selain itu CART bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner (Lewis, 2000).

Salah satu masalah dalam klasifikasi adalah kasus *class imbalance*. Pada data SKAP 2018, persentase remaja yang melakukan hubungan seksual pra-nikah di Jawa Timur sebesar 1,6%, sedangkan yang tidak melakukan adalah 98,4%. Perbedaan persentase antar kelas tersebut menandakan bahwa adanya kasus *class imbalance*. Ketidakseimbangan kelas berpengaruh terhadap klasifikasi yang dihasilkan, yakni setiap data akan cenderung diklasifikasikan pada kelas mayoritas. Salah satu metode untuk mengatasi *class imbalance* adalah metode *resampling* berupa *under-sampling* maupun *over-sampling*. Konsep *resampling* adalah menambah atau mengurangi data sehingga jumlah masing-masing kelas sama. Salah satu metode *over-sampling* yang banyak digunakan adalah SMOTE (*Synthetic Minority Over-sampling Technique*). Metode *resampling* pun dapat digunakan secara bersamaan, yang disebut *hybrid approach*. Misalnya gabungan SMOTE dan ENN (*Edited Nearest Neighbor*) memberikan hasil paling baik dalam klasifikasi menggunakan SVM (Vluymans,

2014). Chawla, *et al.* (2002) mengembangkan SMOTE yang hanya digunakan pada data numerik menjadi SMOTE-N bagi data nominal. Kurniawati pada tahun 2017 mengembangkan ADASYN menjadi ADASYN-N dan ADASYN-KNN untuk data berskala nominal. Rahayu, dkk (2017) melakukan pengujian beberapa metode tersebut menggunakan *Random Forest* pada beberapa dataset dengan kategori *multiclass*. Hasil yang didapatkan adalah ADASYN-N dapat meningkatkan akurasi lebih baik daripada SMOTE-N dan ADASYN-KNN (Rahayu, dkk., 2017).

Berdasarkan penelitian sebelumnya, variabel yang digunakan dalam analisis lebih mengarah pada pengaruh eksternal remaja. Selain itu pada metode yang digunakan juga belum ada justifikasi mengenai penanganan *class imbalance* paling baik dalam kasus seks pra-nikah remaja dan skala data nominal. Oleh karena itu pada penelitian kali ini akan dilakukan klasifikasi remaja melakukan seks pra-nikah menggunakan ADASYN-N dan *hybrid approach* SMOTE-N dengan metode klasifikasi CART. *Hybrid approach* SMOTE-N yang digunakan adalah perpaduan SMOTE-N dan *edited nearest neighbor* (SMOTE-N-ENN). Variabel independen yang digunakan dalam penelitian ini mencakup demografi sosial, pengetahuan, dan perilaku beresiko. CART merupakan metode yang rentan terhadap *overfitting*, sehingga digunakan metode validasi *k-fold cross validation*.

1.2 Rumusan Masalah

Perilaku seks pra-nikah pada remaja di Jawa Timur semakin tinggi, untuk menekan angka persentase tersebut dapat dilakukan prediksi menggunakan demografi sosial, pengetahuan, dan perilaku beresiko. Selain itu, persentase remaja yang melakukan seks pra-nikah di Jawa Timur adalah 1,6%, sehingga dapat dikatakan *imbalance*. Oleh karena itu, perlu suatu justifikasi mengenai metode terbaik untuk mengatasi kasus *imbalance* pada klasifikasi remaja yang melakukan hubungan seks pra-nikah di Jawa Timur.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan utama yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mengetahui metode terbaik dalam mengatasi *class imbalace* pada klasifikasi remaja yang melakukan seks pra-nikah di Jawa Timur menggunakan *base classifier* CART.
2. Mengetahui apakah demografi sosial, pengetahuan, dan perilaku beresiko dapat memprediksi perilaku seks pra-nikah pada remaja di Jawa Timur.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat sebagai informasi bagi masyarakat mengenai faktor yang dapat memprediksi remaja melakukan seks pra-nikah. Selain itu juga menjadi informasi bagi BKKBN sebagai evaluasi terkait *sex education* dan cara menekan angka seks pra-nikah di Jawa Timur. Bagi penelitian salanjutnya, dapat menjadi tambahan informasi terkait metode mengatasi *class imbalance* terbaik.

1.5 Batasan Masalah

Batasan masalah pada penelitian kali ini adalah responden yang digunakan merupakan remaja yang sudah pernah berpacaran berdasarkan hasil Survei Kinerja Akuntabilitas Program KKBPK (SKAP) Jawa Timur 2018

BAB II
TINJAUAN PUSTAKA

Bab ini membahas mengenai *Classification and Regression Trees* (CART), SMOTE-N, ENN, ADASYN-N, k-*fold cross validation*, kinerja klasifikasi, dan perilaku seksual pra-nikah.

2.1 CART

CART (*Classification and Regression Trees*) adalah salah satu algoritma dalam metode klasifikasi *decision tree*. Konsep dari CART adalah *binary recursive partitioning* (Breiman, *et al.*, 1993). Maksud dari *partitioning* adalah membagi *dataset* menjadi beberapa bagian. Istilah *binary* memberikan arti bahwa setiap kelompok diwakili oleh simpul (*node*) dalam pohon keputusan, yang hanya dapat dibagi menjadi dua kelompok. Kemudian, masing-masing simpul tersebut disebut simpul induk dan dapat dibagi menjadi dua simpul anak. Maksud *recursive* menyatakan bahwa *binary partitioning process* dapat berulang terus menerus. Maka dari itu, setiap simpul induk akan menghasilkan dua simpul anak, dan masing-masing simpul anak akan menjadi simpul induk dan menghasilkan simpul anak, begitu seterusnya.

Proses pembentukan pohon keputusan pada CART melalui tiga tahap utama. Pertama adalah pembentukan pohon klasifikasi. Pada tahap ini, terdapat proses pemecahan simpul induk mejadi dua simpul anak melalui aturan pemecahan tertentu dan dilakukan secara berulang. Selanjutnya adalah pemberhentian dan pemangkasan pohon klasifikasi. Tahap terakhir adalah penentuan pohon klasifikasi optimal.

2.1.1 Pembentukan Pohon Klasifikasi

Proses pembentukan pohon klasifikasi terdiri dari tahap pemilihan pemilah, penentuan simpul terminal, serta penandaan label kelas (*class assignment*).

a. Pemilihan Pemilah

Aturan pemilahan simpul induk menjadi dua simpul anak didasarkan pada nilai yang berasal dari satu variabel independen. Satu pemilahan hanya bergantung pada nilai yang berasal dari satu

variabel independen (Breiman, *et al.*, 1993). Jika variabel independen kontinu, X_j dengan ruang sampel *n* dan terdapat *n* nilai amatan sampel berbeda, maka terdapat *n*-1 pemilahan yang berbeda. Jika variabel independen kategorik nominal bertaraf *L*, maka akan diperoleh pemilahan sebanyak $2^{L-1}-1$. Namun jika variabel independen adalah kategorik ordinal, maka akan diperoleh L-1 pemilahan (Hartati, dkk, 2012).

Hasil dari proses pemilahan harus lebih homogen daripada simpul induk. Fungsi heterogenitas yang sering digunakan adalah indeks gini dengan fungsi sebagai berikut.

$$g(t)=\sum_{i\neq j} p(j|t)\,p(i|t) \tag{2.1}$$

Keterangan :

$g(t)$ = fungsi heterogenan (indeks gini) pada simpul *t*

$p(i|t)$ = proporsi kelas *i* pada simpul *t*

$p(j|t)$ = proporsi kelas *j* pada simpul *t*

Selanjutnya menentukan kriteria *goodness of split* yang menjadi suatu evaluasi pemihalan pemilah *s* pada simpul *t*. Rumus untuk mencari nilai *goodness of split* ditulis pada persamaan 2.2 sebagai berikut.

$$\phi(s,t)=\Delta i(s,t)=g(t)-p_L g(t_L)-p_R g(t_R) \tag{2.2}$$

Keterangan :

$\phi(s,t)$ = nilai *goodness of split*

$g(t)$ = fungsi heterogenitas pasa simpul *t*

p_L = proporsi pengamatan simpul kiri

p_R = proporsi pengamatan simpul kanan

$g(t_L)$ = fugsi heterogenitas simpul kiri

$g(t_R)$ = fungsi hetergenitas simpul kanan

Pemilah yang menghasilkan nilai *goodness of split* paling tinggi merupakan pemilah terbaik, karena mampu menurunkan nilai heterogenitas lebih tinggi. Langkah tersebut diulang untuk

menentukan variabel yang digunakan sebagai pemilah *node*, mulai dari *root node* hingga *internal node*.

b. Penentuan Simpul Terminal

Suatu simpul dikatakan sebagai simpul terminal ketika terdapat minimum *m* pengamatan pada suatu simpul anak yang dihasilkan. Dapat dikatakan juga bahwa suatu simpul terminal merupakan simpul ketika tidak terdapat penurunan heterogenitas (Breiman, *et al.*, 1993)

c. Penandaan Label Kelas

Proses penandaan label kelas merupakan proses identifikasi simpul terminal pada suatu kelas tertentu. Pendaaan kelas simpul terminal didasarkan pada aturan jumlah terbanyak.

$$p(j_0 \mid t) = \max_j p(j \mid t) = \max_j \frac{N_j(t)}{N(t)} \tag{2.3}$$

Label kelas untuk simpul terminal *t* adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian paling kecil sebesar $r(t) = 1 - \max_j p(j \mid t)$.

2.1.2 Pemangkasan Pohon Klasifikasi

Pemangkasan pohon klasifikasi atau yang disebut *tree pruning* dimaksudkan untuk menghindari *overfitting* akibat semakin kecilnya kesalahan prediksi karena banyaknya pemilahan. Metode yang digunakan dalam proses pemangkasan pohon adalah *minimal cost complexity* (Breiman, *et al.*, 1993).

$$R_\alpha(T) = R(T) + \alpha \left| \tilde{T} \right| \tag{2.4}$$

Keterangan :

$R_\alpha(T)$ = ukuran kompleksitas suatu pohon T pada kompleksitas α

$R(T)$ = ukuran kesalahan klasifikasi pohon T

α = parameter *cost complexity* bagi penambah satu simpul terminal pada pohon T

$\tilde{T}$ = banyaknya simpul terminal pada pohon T

Dimana ukuran kesalahan pengklasifikasian pohon T dihitung berdasarkan penjumlahan kesalahan pada setiap simpul terminal seperti pada persamaan 2.5.

$$R(T) = \sum_{t \in \tilde{T}} R(t) \tag{2.5}$$

Cost complexity pruning menentukan pohon bagian $T(\alpha)$ yang meminimumkan $R_\alpha(T)$ untuk setiap nilai α. Nilai kompleksitas α akan meningkat seiring proses pemangkasan. Kemudian, dilakukan pencarian pohon bagian $T(\alpha) < T_{\max}$ yang meminimumkan $R_\alpha(T)$.

$$R_\alpha\left(T\left(\alpha\right)\right) = \min_{T<T_{\max}} R_\alpha\left(T\right) \tag{2.6}$$

Titik awal dari *pruning* bukanlah $T_{\max}$, tetapi $T_1 = T(0)$. Dimana T_1 merupakan sub pohon terkecil dari $T_{\max}$ yang memenuhi $R(T_1) = R(T_{\max})$. Untuk mendapatkan T_1 dari $T_{\max}$, diambil t_L dan t_R yang merupakan dua simpul terminal dari $T_{\max}$ yang berasal dari pemecahan simpul dalam (*internal node*) *t*. Jika $R(t) = R(t_L) + R(t_R)$ maka potong t_L dan t_R. Proses tersebut dilanjutkan sampai tidak ada pemotongan yang mungkin.

Dimulai dari T_1, inti dari *minimal cost complexity pruning* terletak pada pemahaman bahwa hal tersebut bekerja berdasarkan *weakest-link cutting*. Untuk mencari nilai kritis dari α, digunakan persamaan sebagai berikut.

$$g_1(t) = \frac{R(t) - R(T_t)}{\left|\tilde{T}_t\right| - 1}, t \notin \tilde{T}_1 \tag{2.7}$$

kemudian untuk menentukan *weakest-link* $\bar{t}_1$ pada T_1 sebagai suatu simpul, yang mana

$$g_1(\bar{t}_1) = \min_{t \in T_1} g_1(t) \tag{2.8}$$

dan

$$\alpha_2 = g_1(\bar{t}_1) \tag{2.9}$$

2.1.3 Penentuan Pohon Klasifikasi Optimum

Setelah proses pemangkasan, maka didapatkan pohon klasifikasi yang berukuran sederhana. Selanjutnya, dapat dipilih salah satu dari beberapa pohon hasil *pruning* yang akan dijadikan sebagai pohon optimal. Pemilihan pohon dapat langsung menggunakan pohon yang meminimumkan $R_\alpha(T)$ atau dengan membentuk estimasi tak bias dari *missclassification cost* $R^*(T_t)$. Metode yang dapat digunakan adalah *test sample estimate* dan *cross validation*. Pada *test sample estimate*, data *training* dibagi dalam dua bagian, *training* (L_1) dan *validasi* (L_2). Pengamatan *training* digunakan untuk membentuk pohon T, dan *validasi* untuk menduga $R(T)$. Jika $N_j^{(2)}$ adalah jumlah data kelas *j* pada data *validasi*. Untuk setiap pohon $T_1, T_2, ..., N_{ij}^{(2)}$ adalah jumlah data kelas *j* dalam data *validasi* yang diprediksi menjadi kelas *i*. Kemudian *expexted missclassification cost* untuk *classifier d* atau total kesalahan klasifikasi *test sample* oleh T dinyatakan oleh persamaan 2.10.

$$R^{ts}(T_t) = \frac{1}{N^{(2)}} \sum_{i,j} c(i \mid j) N_{ij}^{(2)} \qquad (2.10)$$

Dalam hal pendugaan proporsi kesalahan yang dihasilkan dalam pembentukan pohon klasifikasi, pohon klasifikasi optimal T_t yang memiliki estimasi *missclassification cost* minimum atau $R^{ts}(T_t) = \min_t R^{ts}(T_t)$.

2.2 SMOTE-N

Synthetic Minority Over-sampling Technique (SMOTE) merupakan salah satu cara mengatasi *class imbalance* yang diusulkan oleh Chawla *et al.* (2002). Konsep dari SMOTE adalah melakukan *oversampling* pada *minority class* dengan membuat contoh atau data *synthetic* dibanding melakukannya dengan perulangan (Chawla, *et al.*, 2002). SMOTE menambah data buatan dengan *k-nearest neighbor*, sehingga jumlah kelas minor setara

dengan kelas mayor. SMOTE-N merupakan pengembangan dari SMOTE yang awalnya hanya dapat digunakan pada data numerik. SMOTE-N digunakan untuk melakukan *oversampling* pada data dengan kategori nominal.

Jika pada SMOTE, untuk menentukan *k*-data terdekat digunakan jarak *euclidean*, sedangkan pada SMOTE-N jarak terdekat dihitung menggunakan versi modifikasi dari *Value Difference Metric* yang disebut MVDM (Cost & Salzberg, 1993).

Jarak antar kategori dalam suatu variabel independen dijelaskan oleh persamaan berikut.

$$\delta(V_1, V_2) = \sum_{i=1}^{h} \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k \qquad (2.11)$$

dimana :

$\delta(V_1, V_2)$ = jarak antara kategori V_1 dan V_2

C_1 = banyaknya V_1 terjadi

C_2 = banyaknya V_2 terjadi

C_{1i} = banyaknya V_1 yang masuk kelas respon *i*

C_{2i} = banyaknya V_2 yang masuk kelas respon *i*

k = konstan (digunakan 1)

h = jumlah kelas pada variabel respon

Jarak antar data/observasi dihitung menggunakan persamaan berikut.

$$\Delta(X, Y) = w_x w_y \sum_{b=1}^{p} \delta(x_b, y_b)^r \qquad (2.12)$$

dimana :

$\Delta(X, Y)$ = jarak antara observasi X dan Y

w_x, w_y = bobot (dapat diabaikan)

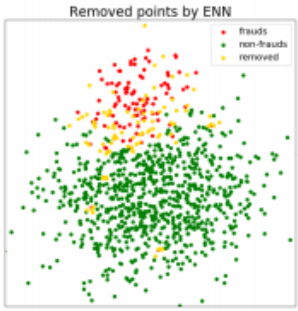
p = banyaknya variabel independen

$\delta(x_b, y_b)$ = jarak antara kategori *x* dan *y* pada variabel independen ke-*b*

r = 1 (Manhattan) atau 2 (Euclidean)

2.3 *Edited Nearest Neighbor* (ENN)

ENN merupakan salah satu metode *under-sampling* yang menggunakan *nearest neighbor*. *Edited Nearest Neighbors* mempertimbangkan *k* tetangga terdekat dalam menentukan titik mana yang harus disimpan dan dihilangkan (Peters, 2018). Suatu data dapat dihilangkan jika hasil klasifikasi (*Y*) tidak sesuai dengan sebagian besar *k* tetangga terdekatnya.



Gambar 2.1 Ilustrasi *Edited Nearest Neighbor*
Sumber : (Peters, 2018)

Pada data kategorik, jarak antar dua kategori dihitung menggunakan jarak *Overlap* sebagai berikut.

$$\delta(V_1, V_2) = \begin{cases} 0, \text{ jika } V_1 = V_2 \\ 1, \text{ jika lainnnya} \end{cases} \tag{2.13}$$

Kemudian jarak antara nilai dihitung sebagai berikut.

$$\Delta(X, Y) = \sum_{b=1}^{p} \delta(x_b, y_b) \tag{2.14}$$

2.4 *Adaptive Synthetic Nominal* (ADASYN-N)

ADASYN atau *Adaptive Synthetic* merupakan salah satu metode dalam mengatasi *class imbalance* yang diajukan oleh He, *et al*. Konsep dari ADASYN adalah memberikan bobot pada data dalam kelas minoritas. Data sintesis yang dihasilkan dari kelas minoritas yang susah untuk belajar atau susah membentuk model akan lebih banyak jika dibandingkan dengan data minoritas yang

lebih mudah untuk belajar. ADASYN meningkatkan hasil pembelajaran dengan mengurangi bias akibat *class imbalance* dan secara adaptif menggeser batas keputusan klasifikasi. Sebagai pengembangan dasi ADASYN, terdapat ADASYN-N (*Adaptive Synthetic Nominal*) untuk data dengan skala nominal, dimana perhitungan jarak pada ADASYN-N hampir sama dengan SMOTE-N, yakni menggunakan modifikasi *Value Difference Metric* (VDM) (Rahayu, dkk, 2017).

Pada ADASYN-N, *training dataset* dengan *m* sampel $\{x_i, y_i\}$, $i = 1, 2, ..., m$, dimana x_i adalah data dalam *n* dimensional *feature space* $\mathbf{X}$ dan $y_s \in Y = \{1, ..., C\}$ adalah label identitias kelas dengan jumlah data terbanyak. Kemudian data dibagi menjadi m_s dan m_l, dimana merupakan jumlah data kelas minoritas dan mayoritas. Oleh karena itu $m_s \le m_l$ dan $m_s + m_l = m$. Selanjutnya dilakukan perhitungan untuk *degree of class imbalance* sebagai berikut.

$$d = \frac{m_s}{m_l} \tag{2.15}$$

dimana $d \in [0,1]$

Perhitungan jumlah data data sintesis yang perlu di-*generate* untuk kelas minoritas digunakan persamaan sebagai berikut.

$$G = (m_l - m_s) \times \beta \tag{2.16}$$

dimana $\beta \in [0,1]$ merupakan parameter yang digunakan dalam penetapan *level balance* yang diinginkan setelah generalisasi data sintetis. Untuk setiap $x_i \in$ *minority class*, kemudian ditentukan k-data terdekat pada *n dimensional space*, dan kalkulasi rasio r_i sebagai berikut.

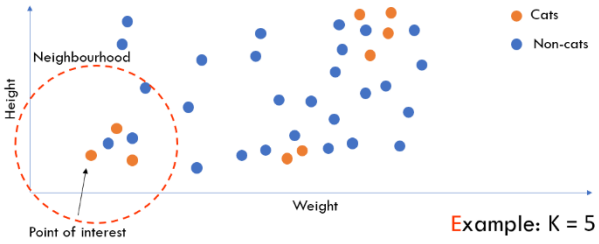
$$r_i = \frac{H_i}{k}, i = 1, ..., m_s \tag{2.17}$$

Keterangan :

r_i = dominasi kelas mayoritas pada masing-masing tetangga

H_i = jumlah data terdekat yang termasuk kelas mayoritas

Semakin tinggi nilai r_i, maka sebagian besar data terdekat merupakan kelas mayoritas dan data akan lebih sulit dipelajari. Berikut merupakan visualisasi nilai r_i dengan *k*=5.



Gambar 2.2 Ilustrasi ADASYN-N
Sumber : (Nian, 2018)

Selanjutnya dilakukan normaliasasi r_i dengan persamaan berikut.

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \tag{2.18}$$

dimana

$$\sum_{i=1}^{m_s} \hat{r}_i = 1 \tag{2.19}$$

Jumlah data sintetis yang perlu dihasilkan bagi setiap data minoritas mengikuti persamaan 2.16 berikut.

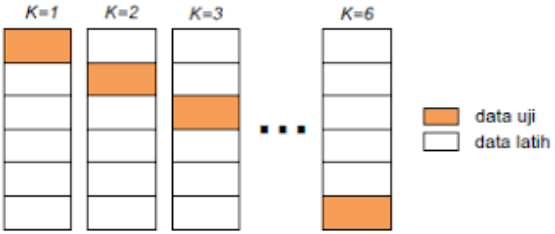
$$g_i = \hat{r}_i \times G \tag{2.20}$$

Kemudian setelah didapatkan banyaknya data yang harus dibangkitkan bagi setiap data pada kelas minotitas, masing-masing data tersebut direplikasi sebanyak g_i kali.

2.5 K-*fold Cross Validation*

K-fold cross validation merupakan salah satu metode yang dapat mempartisi data menjadi *training* dan *testing*. Banyak peneliti yang menerapkan metode ini dikarenakan dapat mengurangi bias yang terjadi dalam pengambilan sampel. *K-fold cross validation* bekerja dengan cara membagi data *training* dan

testing secara berulang, dimana setiap data memiliki kesempatan untuk menjadi data *training* dan *testing*. Dalam metode ini, *K* didefinisikan sebagai angka partisi data yang digunakan untuk pembagian data *training-testing* (Raju, *et al.*, 2018).



Gambar 2.3 Ilustrasi Pembagian *Training Testing*
Sumber : (Pattipeilohy, *et al.*, 2017)

2.6 Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi dilakukan untuk mengetahui seberapa besar kemampuan suatu metode dalam mengklasifikasikan data ke dalam kelas yang tepat. Beberapa cara yang sering digunakan adalah akurasi, sensitifitas, dan spesifisitas. Akurasi merupakan ukuran yang paling umum digunakan menilai kebaikan klasifikasi yang menilai efektivitas algoritma secara keseluruhan dengan memperkirakan probabilitas sebenarnya dari label kelas. Di sisi lain, sensitifitas mengukur kelengkapan atau keakuratan data positif (yang diklasifikasikan benar) (Bekkar, *et al*., 2013). Pengukuran ketepatan klasifikasi dirumuskan berdasarkan tabel *confusion matrix* di bawah ini.

Tabel 2.1 Ketepatan Klasifikasi

Kelas Aktual	**Kelas Prediksi**	
	Positif	Negatif
Positif	*True Positive* (TP)	*False Negative* (FN)
Negatif	*False Positive* (FP)	*True Negative* (TN)

Di bawah ini merupakan rumus untuk mengukur ketepatan klasifikasi.

$$akurasi = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.21}$$

Setelah memperoleh nilai akurasi, sensitifitas, dan spesifisitas, maka akan dilakukan perhitungan ketepatan klasifikasi yang lain, yaitu *Area Under Curve* (AUC). AUC merupakan indikator performasi kurva *Receiver Operating Characteristic* (ROC) yang dapat meringkas kinerja sebuah *classifier* menjadi satu nilai. Nilai AUC sangat bermanfaat untuk menilai kinerja klasifikasi pada data *imbalanced.*

$$sensitivitas = \frac{TP}{TP + FN} \tag{2.22}$$

$$spesifisitas = \frac{TN}{TN + FP} \tag{2.23}$$

$$AUC = \frac{sensitifitas + spesisifitas}{2} \tag{2.24}$$

Adapun interval ketepatan klasifikasi yang baik dengan perhitungan AUC adalah sebagai berikut (Bekkar, *et al*., 2013).

Tabel 2.2 Kriteria Nilai AUC

Nilai AUC	Keterangan
0,5-0,6	Kurang
0,6-0,7	Cukup
0,7-0,8	Baik
0,8-0,9	Sangat Baik
0,9-1,0	Sempurna

2.7 Perilaku Seks Pra-Nikah

Perilaku adalah bentuk respon atau reaksi terhadap gangguan dari luar. Perilaku seksual adalah segala tingkah laku manusia yang didorong oleh hasrat seksual, baik dengan lawan jenisnya maupun dengan sesama jenis (Sarwono, 2011). Perilaku seks pra-nikah berarti perilaku seksual yang dilakukan sebelum adanya proses pernikahan resmi secara agama maupun kepercayaan tertentu. Perilaku seks pra-nikah pada remaja merupakan tindakan seksual yang dilakukan remaja sebelum pernikahan. Pada penelitian ini, perilaku seksual dikhususkan pada perilaku berhubungan badan.

(Halaman ini sengaja dikosongkan)

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari SKAP (Survei Kinerja dan Akuntabilitas Program KKBPK) BKKBN Jawa Timur 2018, modul remaja. Data yang digunakan terdiri atas 695 sampel remaja usia 15-24 tahun yang pernah berpacaran dan belum pernah menikah. Maka, karakteristik inklusi pada penelitian ini adalah remaja pernah berpacaran, sehingga remaja yang belum pernah berpacaran tidak dimasukkan dalam penelitian.

3.2 Deskripsi Data

Survei Kinerja dan Akuntabilitas Program (SKAP) BKKBN 2018 secara keseluruhan populasinya adalah remaja usia 15-24 tahun. Sampel SKAP 2018 tersebar di 34 provinsi, sebanyak 67.725 rumah tangga, di 514 kab/kota, 1935 desa/kelurahan/klaster yang mana setiap klaster terdiri dari 35 ruta. Pada SKAP Provinsi Jawa Timur, sampel diambil dari 38 kab/kota dengan jumlah 3570 ruta yang tersebar di 102 klaster. Jumlah remaja yang berhasil diwawancarai adalah 1027 dari 1032 yang memenuhi persyaratan.

Sampel remaja diambil dari keluarga. Pada rumah tangga yang terpilih dan telah diwawancara, diidentifikasi semua keluarga yang terdapat pada daftar anggota rumah tangga terpilih, termasuk keluarga yang sedang bertamu (menginap) pada suatu rumah tangga terpilih. Selanjutnya, semua anak remaja pria maupun wanita usia 15-24 tahun yang belum menikah (dapat berupa anak kandung, anak tiri, maupun anak asuh) dan tercatat sebagai anggota keluarga, menjadi tanggung jawab serta tinggal bersama keluarga yang menjadi responden, merupakan responden remaja pada SKAP 2018 (Badan Kependudukan dan Keluarga Berencana Nasional, 2018).

Dalam rangka mendapatkan angka estimasi populasi, terlebih dahulu harus dihitung *design weight* dari rancangan sampling yang sudah dibuat. *Design weight* adalah invers dari

fraksi sampling dari setiap tahap penarikan sampel yang dilakukan. Data perlu dilakukan penimbangan dengan benar untuk memastikan bahwa hasilnya tidak terjadi bias estimasi. *Design weight* adalah penimbangan (*weighting*) rumah tangga. Setelah didapatkan *design weight* ini, selanjutnya dilakukan *normalized weighting* untuk mendapatkan normal *weight* rumah tangga, normal *weight* keluarga, normal *weight* WUS, dan normal *weight* remaja. Penggunaan normalisasi untuk menghindari penyajian angka yang besar dalam laporan kegiatan, dan me-retrive nilai *w* (*weight*) yang cenderung besar ke nilai *n* (jumlah sampel) (Badan Kependudukan dan Keluarga Berencana Nasional, 2018). *Weight* pada setiap data/observasi merepresentasikan jumlah setiap data, sehingga sebelum dilakukan analisis, setiap data direpliasi sesuai *weight*-nya masing-masing.

3.3 Kerangka Konsep

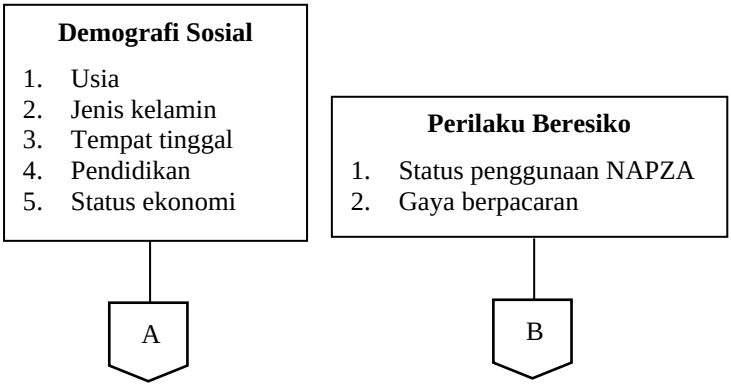
Pada penelitian ini dilakukan peninjauan mengenai perilaku seks pra-nikah pada remaja di Jawa Timur. Jenis kelamin berpengaruh secara langsung terhadap perilaku seksual pra-nikah. Secara praktis, remaja yang berjenis kelamin laki-laki memberikan peluang sebesar 1,4 kali lebih berisiko untuk melakukan perilaku seksual pra-nikah (Rosdarni, dkk, 2015).

1. Usia, berdasarkan SDKI tahun 2017 diketahui bahwa persentase remaja laki-laki yang melakukan hubungan seks pra-nikah lebih banyak pada usia 20-24 tahun dibandingkan 15-19 tahun (Wahyuni & Fahmi, 2019). Terdapat berbagai cara dalam menentukan *treshold* usia, pada penelitian kali ini digunakan batas umur perkawinan sesuai UU No. 16 Tahun 2019 tentang perkawinan. Pembagian usia remaja laki-laki adalah <19 tahun dan ≥ 19 tahun.
2. Pendidikan, tingkat pendidikan terakhir berhubungan negatif dengan perilaku seks remaja. Artinya, jika seorang remaja memiliki pendidikan yang tinggi, maka memiliki kecenderungan untuk tidak melakukan hubungan seks pra-nikah (Wahyuni & Fahmi, 2019).

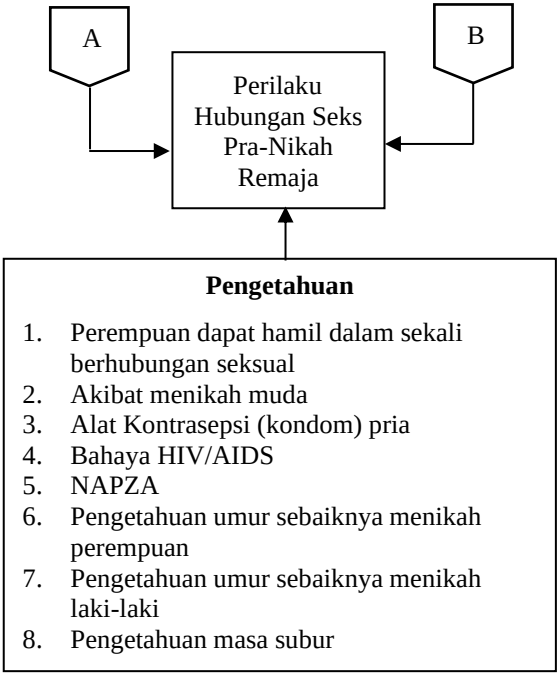
3. Tempat tinggal, karakteristik remaja yang tinggal di perkotaan dan pedesaan sangatlah berbeda. Prevalensi akses pornografi pada remaja yang tinggal di pedesaan lebih tinggi dibandingkan perkotaan (Hastuti, dkk, 2013). Menurut Harnani, dkk (2018), paparan pornografi berperngaruh terhadap perilaku seks pra-nikah pada remaja, sehingga tempat tinggal juga merupakan salah satu faktor pemicu.
4. Status ekonomi, menurut Dave, *et al.* (2013) remaja dengan tingkat ekonomi menengah ke bawah cenderung melakukan hubungan seks dan sering kali berganti pasangan.
5. Pengetahuan kondom pria. Berdasarkan penelitian Harnani, dkk (2018), faktor eksternal berupa paparan pornografi berpengaruh signifikan terhadap perilaku remaja melakukan hubungan seks pra-nikah, yang mana dalam film tersebut berisi penyalahgunaan alat kontrasepsi.
6. Kehamilan. Kurangnya *sex education* mejadi salah satu hal yang memicu remaja melakukan hubungan seksual pra-nikah. Salah satu tanda kurangnya *sex education* adalah masih terdapat remaja yang percaya akan mitos bahwa hubungan seks pertama kali tidak menyebabkan kehamilan.
7. Pengetahuan NAPZA, disebutkan bahwa semakin tinggi pengetahuan tentang NAPZA maka semakin rendah perilaku seks pranikah remaja (Hardiansyah, 2018)
8. Konsumsi NAPZA, residen penyalahgunaan jenih shabu akan meningkatkan kinerja seksual sehingga lebih lama dan agresit (Harbia, dkk, 2018). Penelitian yang sama juga pernah dilakukan Sitorus (2014) dan menyebutkan bahwa 61,3% pecandu narkotika telah melakukan hubungan seksual pra-nikah
9. Pengetahuan bahaya HIV/AIDS, remaja yang memiliki pengetahuan yang rendah berpeluang lebih dari 1,5 kali untuk melakukan perilaku seksual pranikah yang berisiko dibandingkan remaja yang memiliki pengetahuan yang tinggi (Rosdarni, dkk, 2015).

10. Gaya berpacaran, sebagian besar partner dalam melakukan penyimpangan adalah pacar (Dave, et al., 2013). Hal tersebut berarti interaksi maupun gaya pacaran dengan pacar mungkin memicu perilaku seks pra-nikah pada remaja.
11. Pengetahuan mengenai akibat menikah muda, salah satu akibat dari hubungan seks pra-nikah adalah kehamilan tidak diinginkan (KTD), yang mana mungkin memicu pernikahan dini. Pengetahuan mengenai akibat menikah muda/dini merupakan hal yang penting.
12. Pengetahuan umur menikah pertama, pengetahuan tentang umur sebaiknya menikah dan melahirkan merupakan bagian dari KRR yang harus diketahui remaja. Jawaban responden dikategorikan menjadi ≥ 19 tahun dan <19 tahun, sesuai batas usia menikah.
13. Pengetahuan masa subur, pengetahuan kesehatan reproduksi remaja yang berpengaruh secara individu terhadap pengalaman melakukan hubungan seksual pranikah adalah pengetahuan masa subur dan pengetahuan tentang NAPZA (Nasution, 2012).

Agar lebih jelas mengenai gambaran teori dan konsep pada perilaku seksual remaja, maka disajikan kerangkan konsep pada Gambar 3.1



Gambar 3.1 Kerangka Konsep



Gambar 3.1 Kerangka Konsep (Lanjutan)

3.4 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas variabel dependen dan variabel independen. Keterangan lebih lanjut mengenai variabel penelitian yang digunakan ditunjukkan oleh Tabel 3.1 berikut.

Tabel 3.1 Variabel Penelitian

No	Nama Variabel	Definisi	Kode Kuesioner	Keterangan
Variabel Dependen				
1	Pernah melaku-kan hubungan seks (Y)	Apakah responden pernah melakukan seks	YQ44	Kategori : 0. Tidak 1. Ya

Tabel 3.1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Definisi	Kode Kuesioner	Keterangan
Variabel Independen				
Demografi Sosial				
1	Usia (X1)	Usia responden	YQ1	Kategori : 0. <19 tahun 1. ≥ 19 tahun
2	Jenis Kelamin (X2)	Jenis kelamin responden	YQL	Kategori: 0. Laki-laki 1. Perempuan
3	Pendidikan (X3)	Pendidikan responden terakhir yang dinyatakan lulus	HQ2a	Kategori: 0. ≤ SLTP 1. > SLTP
4	Tempat Tinggal (X4)	Responden ditanya berkaitan dengan tempat tinggal saat ini yang dilihat berdasarkan perkotaan atau pedesaan.	Kota_desa	Kategori: 0. Perkotaan 1. Pedesaan
5	Status Ekonomi (X5)	Kuintil kekayaan remaja	wealthquintile	Kategori: 0. ≤Menengah ke bawah 1. >Menengah ke bawah
6	Kondom (X6)	Pengetahuan responden tentang alat kontrasepsi (kondom) pria	YQ3H	Kategori : 0. Tidak 1. Ya

Tabel 3.1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Definisi	Kode Kuesioner	Keterangan
Pengetahuan				
7	Kehamilan (X7)	Pengetahuan responden tentang apakah seorang remaja perempuan yang telah haid dapat hamil meskipun hanya sekali melakukan hubungan seksual	YQ5	Kategori: 0. Tidak tahu 1. Tidak dapat 2. Dapat
8	NAPZA (X8)	Apakah responden pernah mendengar mengenai NAPZA atau tidak	YQ13	Kategori : 0. Tidak 1. Ya
9	Bahaya HIV/ AIDS (X9)	Pengetahuan responden tentang bahaya HIV/AIDS	YQ17	Kategori : 0. Tidak 1. Ya
10	Akibat Menikah Muda (X10)	Mengetahui akibat menikah muda	YQ12	Kategori : 0. Tidak 1. Ya
11	Menikah_ P (X11)	Pengetahuan umur perempuan sebaiknya menikah pertama	YQ6	Kategori : 0. Tidak tahu 1. <19 tahun 2. ≥19 tahun

Tabel 3.1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Definisi	Kode Kuesioner	Keterangan
Pengetahuan				
12	Menikah_L (X12)	Pengetahuan umur laki-laki sebaiknya menikah pertama	YQ7	Kategori : 0. Tidak tahu 1. <19 tahun 2. ≥19 tahun
13	Subur (X13)	Pengetahuan remaja mengenai masa subur	YQ4	Kategori : 0. Tidak tahu/Tidak pernah mendengar 1. Tahu namun salah 2. Tahu benar
Perilaku				
14	Konsumsi NAPZA (X14)	Apakah responden pernah mengkonsumsi NAPZA atau tidak	YQ15	Kategori : 0. Tidak 1. Ya
15	Gaya Pacaran (X15)	Gaya berpacaran	YQ43a – YQ43d	Kategori : 0. Tidak melakukan apa-apa 1. Berpegangan tangan 2. Berpelukan 3. Ciuman bibir 4. Meraba/merang-sang

3.5 Struktur Data

Struktur data dari variabel-variabel yang digunakan dalam penelitian adalah sebagai berikut.

Tabel 3.2 Struktur Data

No	X_1	X_2	X_3	…	X_k	Y
1	X_{11}	X_{21}	X_{31}	…	X_{k1}	Y_1
2	X_{12}	X_{22}	X_{32}	…	X_{k2}	Y_2
⋮	⋮	⋮	⋮		⋮	⋮
n	X_{1n}	X_{2n}	X_{3n}	…	X_{kn}	Y_n

Keterangan :

X_k = variabel ke-*k*

Y_n = kelas pada data ke-*n*

3.6 Langkah Analisis

Langkah analisis digunakan untuk menggambarkan langkah-langkah penelitian yang akan dilakukan secara urut. Langkah analisis yang digunakan adalah sebagai berikut.

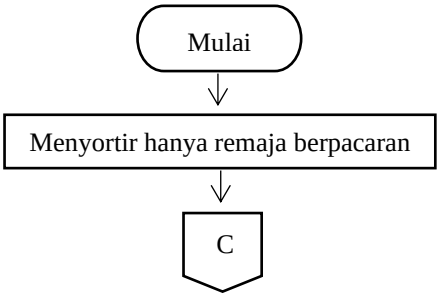
1. Melakukan penyortiran (pemilahan) pada data remaja yang pernah berpacaran.
2. Melakukan replikasi masing-masing data sesuai *weight.*
3. Melakukan eksplorasi data variabel dependen dan independen.
4. Melakukan *preprocessing* data.
5. Membagi data *training* dan *testing* menggunakan 5-*fold cross validation*.
6. Melakukan analisis klasifikasi remaja melakukan seks pra-nikah menggunakan CART pada data *imbalance*.
   a. Melakukan pembentukan pohon klasifikasi optimal
   b. Melakukan pemangkasan pohon klasifikasi
   c. Melakukan pemilihan pohon terbaik
   d. Menghitung kinerja klasifikasi
7. Mengatasi *imbalance* menggunakan SMOTE-N hanya pada data *training*.
   a. Menghitung jarak antar amatan pada kelas minor menggunakan rumus VDM.

b. Menentukan nilai *k* yaitu 10.
c. Memilih salah satu data pada kelas minor secara acak.
d. Menentukan *k* tetangga terdekat dengan mengurutkan. jarak data terpilih dengan semua amatan pada kelas minor.
e. Data sintetis dibuat dengan menentukan nilai masing-masing variabel independen. Nilai tersebut diperoleh dari mayoritas *k* tetangga terdekat.
f. Ulangi langkah c hingga e sampai mencapai banyaknya *oversampling* yang dibutuhkan

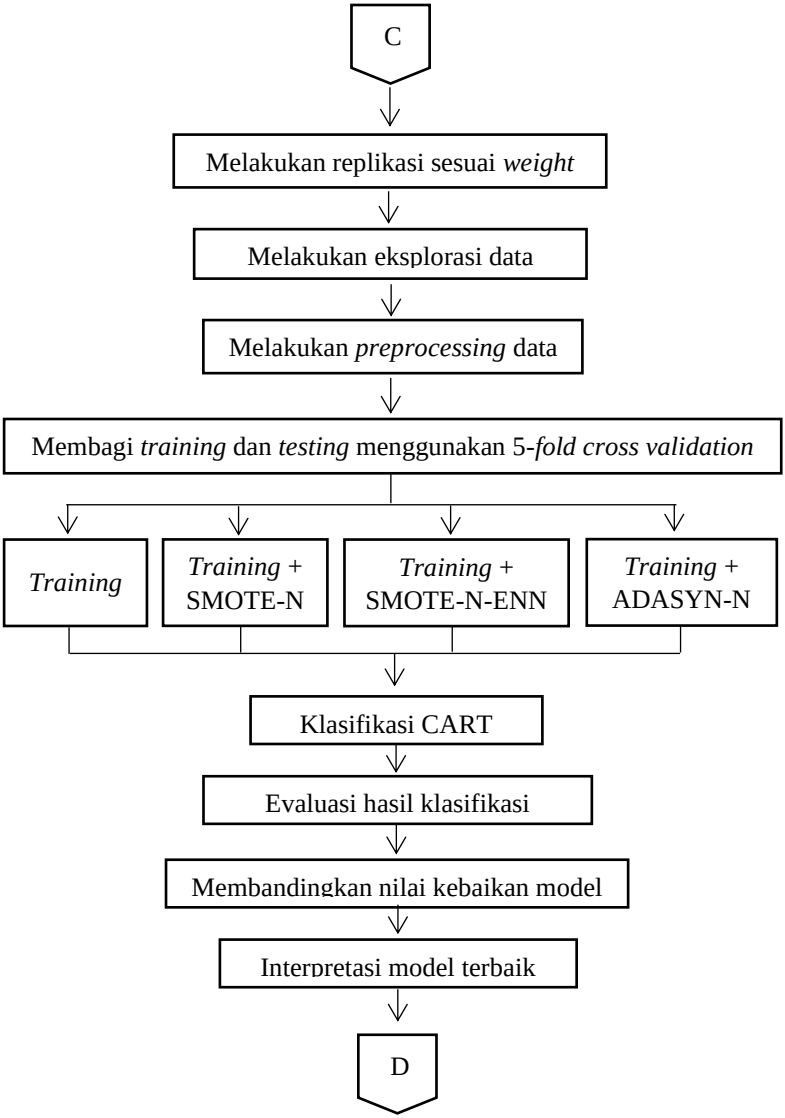
8. Melakukan analisis klasifikasi remaja yang melakukan seks pra-nikah menggunakan *classification and regression tree* (CART) pada data *balance*.
9. Mengulangi tahap 7 menggunakan metode SMOTE-N-ENN, dan ADASYN-N
10. Membandingkan rata-rata akurasi dan AUC dari hasil klasifikasi menggunakan berbagai metode *imbalance*.
11. Menginterpretasikan hasil klasifikasi dengan kinerja klasifikasi terbaik.
12. Menarik kesimpulan dan saran.

3.7 Diagram Alir

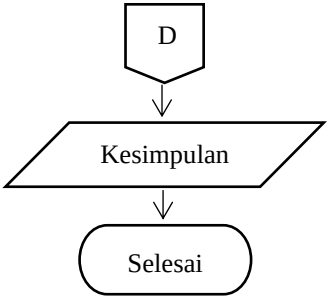
Diagram alir dari langkah analisis pada penelitian ini adalah sebagai berikut.



Gambar 3.2 Diagram Alir



Gambar 3.2 Diagram Alir (Lanjutan)



Gambar 3.2 Diagram Alir (Lanjutan)

BAB IV
ANALISIS DAN PEMBAHASAN

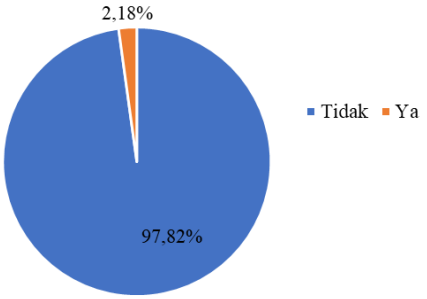
Pada penelitian ini dilakukan klasifikasi remaja yang merupakan prediksi apakah remaja melakukan seks pra-nikah atau tidak. Penelitian ini membandingkan tiga metode penanganan *imbalance* yakni SMOTE-N, SMOTE-N-ENN, dan ADASYN-N dengan *classifier Classification and Regression Trees* (CART). Kebaikan hasil klasifikasi tersebut didapatkan dari hasil nilai akurasi dan AUC.

4.1 *Preprocesssing* dan Analisis Karakteristik Data

Pada sub-bab ini akan dibahas mengenai *preprocessing* dan eksplorasi atau karakteristik data. Data yang dieksplorasi merupakan data yang telah direplikasi berdasarkan *weight* masing-masing data/observasi.

4.1.1 Eksplorasi Data

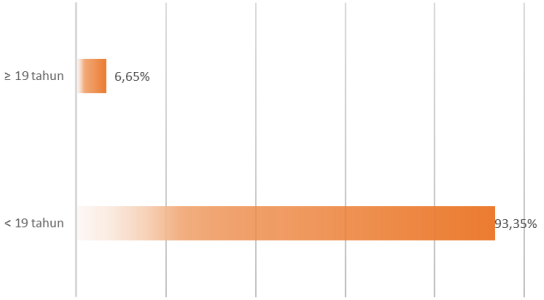
Analisis karakteristik data dilakukan untuk melihat bagaimana gambaran data. Eksplorasi data tahap awal dilakukan pada beberapa variabel awal seperti variabel respon dan demografi sosial.



Gambar 4.1 Persentase Seks Pra-Nikah Remaja

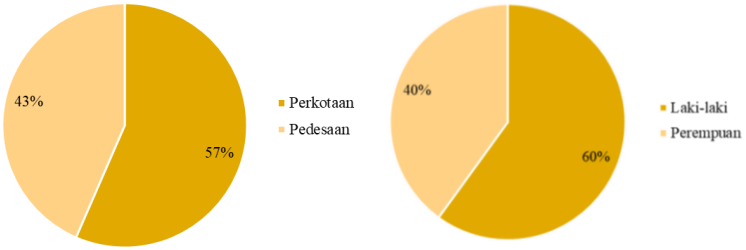
Pada SKAP Jawa Timur 2018, responden atau remaja yang pernah berpacaran ditanya tentang pengalaman seksual yang digambarkan pada Gambar 4.1 Dari gambar tersebut dapat dikatakan bahwa secara umum remaja di Jawa Timur yang melakukan hubungan seksual pra-nikah lebih kecil daripada yang

tidak melakukan, yakni 2,18% dibanding 97,82% Walaupun secara umum persentase remaja yang melakukan hubungan seksual pra-nikah masih terlihat sedikit, namun jika kita kalikan dengan total jumlah remaja yang terdapat di Jawa Timur maka hal ini dapat menjadi masalah dan tugas yang besar bagi pemerintah Jawa Timur dalam penanggulangan permasalahan remaja. Jumlah total data setelah replikasi adalah 2015, hal tersebut berarti jumlah remaja melakukan seks pra-nikah adalah 44 remaja.



Gambar 4.2 *Bar Chart* Kelompok Usia Remaja Responden SKAP 2018

Berdasarkan Gambar 4.2, terlihat bahwa responden remaja dengan usia kurang dari 19 tahun lebih banyak daripada 19 tahun atau lebih. Hal tersebut memberikan informasi bahwa sebagian besar responden berada pada usia belum seharusnya menikah. Remaja pada usia tersebut seharusnya sudah mendapat pendidikan seks maupun informasi yang berkaitan dengan kesehatan reproduksi remaja.



(a) (b)

Gambar 4.3 Persentase (a) Daerah (b) Jenis Kelamin Remaja yang Melakukan Hubungan Seks Pra-Nikah

Gambar 4.3 menunjukkan bahwa sebagian besar remaja yang melakukan hubungan seksual pra-nikah di Jawa Timur bertempat tinggal di perkotaan. Perbedaan persentase tempat tinggal tidak terlalu jauh, hal tersebut dapat diartikan bahwa permasalahan remaja tidak hanya berada di kota besar saja, yang biasa dianggap sebagai sarang penyebab kenakalan remaja dan pergaulan bebas, namun juga sudah terjadi di daerah pedesaan. Selain itu, remaja dengan jenis kelamin laki-laki lebih banyak melakukan hubungan seksual dibandingkan dengan perempuan.

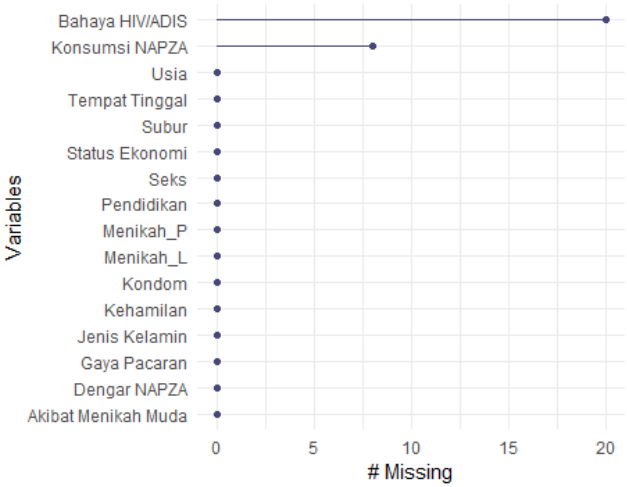
Gambar 4.4 *Bar Chart* Pendidikan Terakhir Remaja yang Melakukan Hubungan Seksual Pra-Nikah

Berdasarkan Gambar 4.4, terlihat bahwa remaja yang melakukan hubungan seksual pra-nikah sebagian besar memiliki pendidikan terakhir lebih dari SLTP, atau dapat dikatakan minimal telah lulus SLTA. Hal tersebut memberikan informasi bahwa remaja yang melakukan hubungan seksual memiliki tingkat pendidikan menengah ke atas, atau telah menempuh masa wajib belajar 12 tahun.

4.1.2 *Preprocessing Data*

Setelah melihat bagaimana gambaran data awal, selanjutnya dilakukan *preprocessing* sebagai persiapan data untuk analisis selanjutnya. *Preprocessing data* yang digunakan pada penelitian ini adalah *missing value*. Pada analisis *missing value*, data yang seharusnya digunakan adalah data setelah replikasi, namun untuk mempermudah imputasi, pada sub bab ini ditampilkan *missing*

value pada data awal sebelum replikasi. Berikut merupakan gambaran *missing value* masing-masing variabel sebelum dilakukan replikasi.

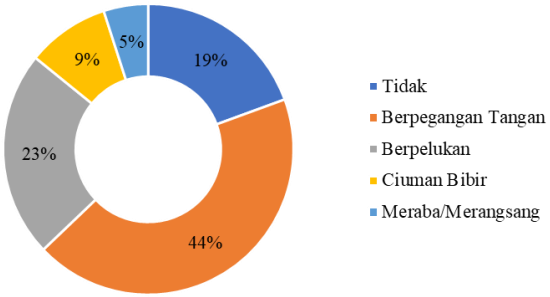


Gambar 4.5 *Missing Value Plot*

Gambar 4.5 menunjukkan bahwa terdapat *missing value* pada variabel pengetahuan bahaya HIV/AIDS dan status konsumsi NAPZA. Karena variabel yang digunakan merupakan variabel kategorik, sehingga dilakukan imputasi dengan nilai modus masing-masing variabel. Variabel pengetahuan HIV/AIDS diimputasi dengan nilai kategori "1" yang berarti mengetahui bahaya HIV/AIDS, sedangkan *missing value* variabel konsumsi NAPZA diisi dengan kategori "0" yang memiliki makna tidak mengkonsumsi NAPZA. Setelah dilakukan imputasi, kemudian data direplikasi kembali sesuai *weight* masing-masing data/observasi.

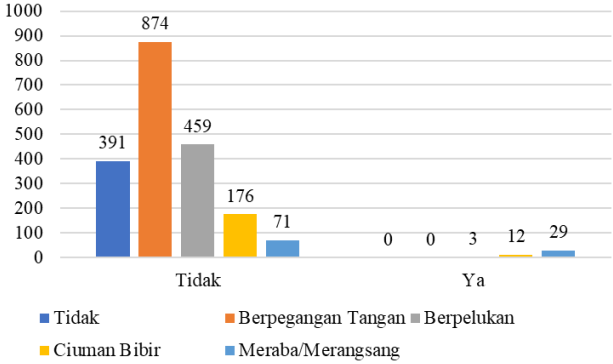
4.1.3 Analisis Karakteristik Data Setelah *Preprocessing*

Setelah dilakukan *preprocessing data*, selanjutnya dilakukan analisis karakteristik data kembali. Berikut merupakan analisis karakteristik data pada data replikasi setelah imputasi.



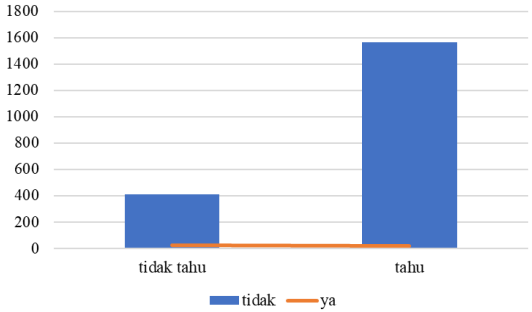
Gambar 4.6 Persentase Gaya Pacaran Remaja di Jawa Timur

Berdasarkan Gambar 4.6, terlihat bahwa sebagian besar remaja yang berpacaran di Jawa Timur minimal berpegangan tangan. Persentase remaja yang berpelukan lebih besar daripada tidak melakukan apapun saat berpacaran. Jadi dapat dikatakan bahwa gaya pacaran yang mulai intens sudah biasa di kalangan remaja di Jawa Timur. Kemudian, berikut merupakan *crosstabulation* gaya pacaran dan perilaku seks pra-nikah.



Gambar 4.7 Karakteristik Gaya Pacaran Remaja di Jawa Timur

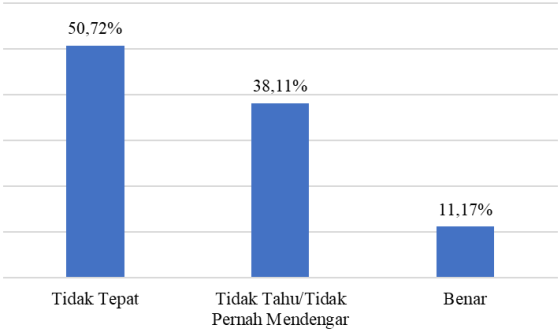
Gambar 4.7 menunjukkan bahwa pada remaja yang melakukan hubungan seks pra-nikah, gaya pacaran yang muncul hanya berpelukan, ciuman bibir, dan meraba/merangsang. Hal tersebut menunjukkan bahwa remaja dengan gaya pacaran mengarah ke fisik cenderung melakukan hubungan seksual pra-nikah.



Gambar 4.8 Karakteristik Pengetahuan Resiko Menikah Muda

Gambar 4.8 memberikan informasi bahwa pada remaja yang tidak melakukan hubungan seksual pra-nikah, jumlah yang mengetahui akibat menikah muda jauh lebih banyak daripada yang tidak mengetahui. Namun, pada remaja yang melakukan hubungan seksual pra-nikah, jumlah yang mengetahui dan tidak mengetahui bahaya HIV/AIDS tidak terpaut jauh.

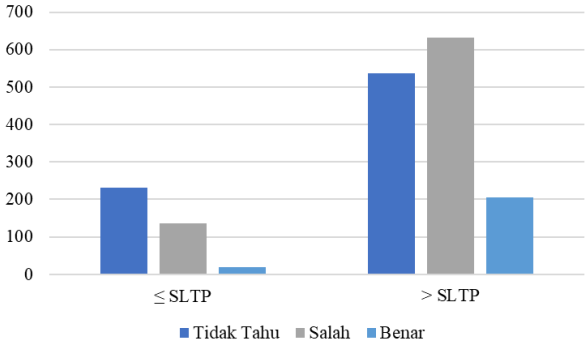
Berkaitan dengan *sex education*, berikut merupakan gambaran pengetahuan mengenai masa subur seorang wanita. Pilihan jawaban pada SKAP, terdapat tidak tahu hingga tidak pernah mendengar.



Gambar 4.9 *Bar Chart* Pengetahuan Masa Subur

Berdasarkan Gambar 4.9, terlihat bahwa pengetahuan remaja di Jawa Timur mengenai masa subur sangat rendah. Hal terebut terlihat dari banyaknya remaja yang masih belum mengetahui

kapan masa subur seorang wanita. Lebih dari 50% remaja salah mendefinisikan masa subur. Jawaban yang muncul dalam salah definisi tersebut adalah segera setelah haid, menjelang haid, bahkan saat haid berlangsung, padahal masa subur yang tepat adalah di antara dua haid. Persentase remaja di Jawa Timur yang mengetahui masa subur secara benar adalah 11,17%. Nilai tersebut hampir sama dengan persentase pengetahuan masa subur secara benar di Indonesia, yakni 11,11%. Informasi tersebut menunjukkan bahwa pengetahuan mengenai masa subur di Jawa Timur maupun Indonesia tergolong rendah.



Gambar 4.10 Karakteristik Masa Subur dan Pendidikan

Gambar 4.10 menunjukkan pada bahwa remaja dengan pendidikan maksimal SLTP, sebagian besar tidak memiliki pengetahuan mengenai masa subur seorang wanita. Kemudian pada remaja dengan pendidikan SLTA atau lebih, pengetahuan mengenai masa subur sudah bergeser dan didominasi oleh pengetahuan yang salah. Hal tersebut memberikan informasi bahwa pengetahuan mengenai masa subur belum sampai pada remaja dengan pendidikan maksimal SLTP dan menjadi salah pada remaja pendidikan SLTA. Bergesernya pengetahuan seiring naiknya tingkat pendidikan juga mengindikasikan adanya pengabaian mengenai informasi masa subur, padahal pengetahuan tersebut erat kaitannya dengan kehamilan tidak diinginkan (KTD).

4.2 Analisis Klasifikasi Remaja Melakukan Hubungan Seksual Pra-Nikah Menggunakan CART

Setelah dilakukan *preprocessing* data, selanjutnya akan dilakukan analisis klasifikasi perilaku remaja melakukan hubungan seksual pra-nikah. Pertama, digunakan data asli yang merupakan *imbalanced data*. Terdapat tiga tahapan dalam metode klasifikasi CART, yaitu pembentukan pohon, pemangkasan pohon, hingga penentuan pohon klasifikasi yang optimal. Berikut merupakan ilustrasi penggunaan metode CART pada klasifikasi dengan data *imbalance* pada *fold* 1.

1. Pembentukan Pohon

Pada pembentukan pohon klasifikasi, diperlukan variabel-variabel yang berperan sebagai pemilah. Jika variabel berskala nominal bertaraf *L*, maka akan diperoleh pemilahan sebanyak $2^{L-1}-1$. Jumlah kemungkinan pemilah untuk membentuk pohon yang berbeda pada pohon klasifikasi remaja yang melakukan hubungan seksual pra-nikah ditampilkan pada Tabel 4.1.

Tabel 4.1 Jumlah Kemungkinan Pemilah pada Variabel Independen

Variabel	Jumlah Kategori	Kemungkinan Pemilah
Usia	2	1
Tempat Tinggal	2	1
Jenis Kelamin	2	1
Pendidikan	2	1
Kondom	2	1
Kehamilan	3	3
Dengar NAPZA	2	1
Konsumi NAPZA	2	1
Bahaya HIV/ADIS	2	1
Akibat Menikah Muda	2	1
Gaya Pacaran	5	15
Status Ekonomi	2	1
Subur	3	3
Menikah_W	3	3
Menikah_L	3	3

Setelah dilakukan perhitungan jumlah kemungkinan pemilah dalam pembentukan pohon klasifikasi, selanjutnya adalah pemilihan pemilah menggunakan indeks gini. Indeks gini merupakan karakteristik dari CART. Berikut merupakan contoh perhitungan indeks gini pada variabel usia data *train*.

Tabel 4.2 Ilustrasi Pemilahan pada Simpul Usia

Usia	Seks		Total
	Tidak	Ya	
< 19 tahun	1470	36	1506
≥ 19 tahun	107	0	107

Selanjutnya melakukan perhitungan untuk nilai indeks gini sesuai Persamaan 2.1 pada masing-masing simpul kanan dan kiri sebagai berikut.

$$g\left(t_L\right)=2\times\left(\frac{1470}{1506}\times\frac{36}{1506}\right)=0,04666$$

$$g\left(t_R\right)=2\times\left(\frac{107}{107}\times\frac{0}{107}\right)=0$$

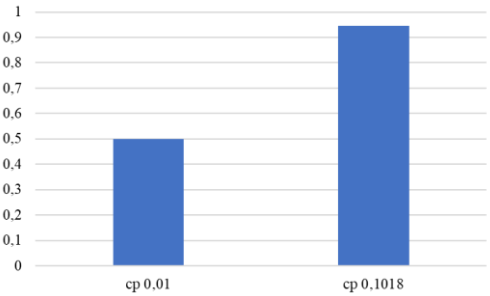
Kemudian menentukan kriteria *goodness of split* untuk evaluasi pemilahan yang telah dilakukan oleh pemilah *s* pada simpul *t* sesuai Persamaan 2.2. Karena hanya ada satu kemungkinan pemilahan, maka untuk simpul Usia hanya ada satu kriteria *goodness of split*.

$$\phi\left(s,t\right)=0,04361-\left(\frac{1506}{1613}\right)\times 0,04666-\left(\frac{107}{1613}\right)\times 0=3,97\times10^{-5}$$

Goodness of split pada variabel usia adalah $3,97\times10^{-5}$. Selanjutnya indeks gini dan *goodness of split* variabel lainnya dihitung dengan perhitungan serupa variabel usia. Variabel yang menjadi simpul akar adalah variabel dengan *goodness of split* paling tinggi. Hal tersebut berulang terus menerus hingga didapatkan simpul terminal. Simpul *t* dikatakan sebagai simpul terminal jika tidak terdapat penurunan heterogenitas atau dengan kata lain hanya terdapat satu kelas pada simpul anak.

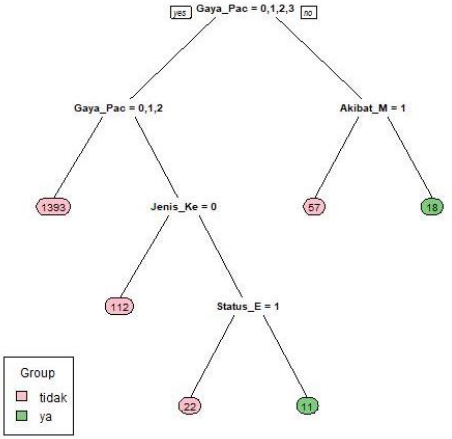
2. Pemangkasan Pohon

Complexity parameter beserta besar error *cross validation* terdapat pada Gambar 4.11, yang merupakan salah satu langkah dalam tahap metode CART, yakni *tree pruning* atau pemangkasan pohon klasifikasi. *Tree pruning* dilakukan untuk menghindari adanya *overfitting* yang diakibatkan oleh jumlah pemilahan yang terlalu banyak. Gambar 4.11 menunjukkan bahwa pada klasifikasi remaja melakukan hubungan seksual pra-nikah, parameter *cp* paling optimum sebesar 0,01.



Gambar 4.11 Parameter *cp* Data *Imbalance Fold 1*

Selanjutnya dibetuk pohon keputusan dengan menggunakan *cp* optimum sebagai berikut.



Gambar 4.12 CART Data *Imbalance Fold 1*

Pohon klasifikasi pada Gambar 4.12 merupakan pohon keputusan menggunakan data yang masih *imbalance*. Seanjutnya dibentuk *confusion matrix* sebagai berikut.

Tabel 4.3 *Confusion Matrix Training Imbalance Fold* 1

Aktual	Prediksi	
	Tidak	**Ya**
Tidak	1574	3
Ya	10	26

Berdasarkan *confusion matrix*, terlihat bahwa masih terdapat *missclassification* yang cukup banyak, terutama pada remaja yang seharusnya melakukan seks pra-nikah namun diprediksi tidak melakukan. Selanjutnya dibentuk pohon lain sesuai *fold* yang telah ditentukan. Berikut merupakan nilai akurasi dan AUC masing-masing *fold* pada data *imbalance* dengan nilai *cp* optimum.

Tabel 4.4 Ketepatan Klasifikasi 5 *Fold* SMOTE-N

Fold	**Akurasi *Training***	**AUC *Training***	**Akurasi *Testing***	**AUC *Testing***
1	0,991941	0,86016	0,99005	0,75
2	0,993176	0,9127	0,985112	0,77524
3	0,985112	0,68508	0,987593	0,776509
4	0,985723	0,699366	0,985149	0,720956
5	0,986973	0,769843	0,980149	0,664129
Rata-rata	0,988585	0,78543	0,98561	0,737367

Tabel 4.4 menunjukkan bahwa klasifikasi remaja melakukan hubungan seksual pra-nikah di Jawa Timur menggunakan CART memberikan hasil rata-rata akurasi *testing* sebesar 98,56% dengan rata-rata nilai AUC *testing* sebesar 73,73%. Jika dilihat pada masing-masing *fold*, CART dapat mengklasifikasikan data *imbalance testing* dengan akurasi minimum 98,01% dan maksimal 99% dengan minumum AUC 66,41% dan maksimum 77,65%. Tingginya nilai akurasi namun AUC rendah mengindikasikan masing-masing kelas tidak diklasifikasikan secara tepat, atau kelas minoritas diklasifikasikan pada kelas mayoritas. Sebagian besar

data akan diklasifikasikan pada kelas mayoritas karena kelas yang tidak seimbang. Oleh karena itu, data yang *imbalance* tidak dapat dibiarkan, karena klasifikasi yang dihasilkan akan menjadi salah.

4.3 Penanganan *Imbalance*

Pada karakteristik data telah dijelaskan bahwa persentase remaja yang melakukan dan tidak melakukan hubungan seksual pra-nikah berbeda jauh, atau dapat dikatakan tidak seimbang (*imbalance*). Selain itu, pada sub bab 4.2 juga terlihat bahwa penggunaan data *imbalance* pada klasifikasi CART tidak memberikan hasil yang baik, karena semua kelas diklasifikan pada kelas mayoritas. Oleh karena itu, perlu dilakukan penanganan kasus *imbalance* sebelum melakukan analisis klasifikasi. Pada penelitian ini akan dilakukan replikasi (pembuatan data sintetis) untuk mengatasi *imbalanced data*. Penanganan *imbalance* hanya dilakukan pada data *training*, sedangkan *testing* merupakan dataset asli. Metode penanganan *imbalance* yang digunakan adalah SMOTE-N, SMOTE-N-ENN, dan ADASYN-N.

4.3.1 SMOTE-N

Prinsip *Synthetic Minority Oversamling Technique Nominal* (SMOTE-N) adalah membuat data sintetis (*oversampling*) sehingga jumlah remaja yang melakukan dan tidak melakukan hubungan seksual sama. Data baru akan dibangkikan pada sekitar *k*-tetangga terdekat atau *k*-data terdekat yang memiliki kemiripan berdasarkan jarah VDM. Berikut merupakan ilustrasi proses SMOTE-N untuk menangani kasus *imbalance* pada klasifikasi remaja melakukan hubungan seksual pra-nikah di Jawa Timur.

1. Mengambil data kelas minoritas secara random

Langkah pertama dalam SMOTE-N adalah menghitung jarak VDM antar kelas minoritas. Sebagai ilustrasi, misal pada *fold* 1 terdapat 36 data yang merupakan kelas minor (remaja melakukan hubungan seksual pra-nikah), kemudian diambil salah satu observasi secara random. Misalkan data terpilih adalah data ke-5 (x_5).

2. Menghitung jarak VDM

Selanjutnya adalah menghitung jarak VDM x_5 dengan 35 data lain yang termasuk dalam kelas minoritas. Misal jarak x_5 dan x_{15}. Berikut merupakan gambaran masing-masing variabel x_5 dan x_{15}.

Tabel 4.5 Data Ilustasi SMOTE-N

Variabel	Data	
	x_5	x_{15}
Usia	0	0
Tempat Tinggal	**0**	**1**
Jenis Kelamin	0	0
Pendidikan	**1**	**0**
Kondom	1	1
Kehamilan	2	2
Dengar NAPZA	1	1
Konsumi NAPZA	**1**	**0**
Bahaya HIV/ADIS	1	1
Akibat Menikah Muda	**1**	**0**
Gaya Pacaran	4	4
Status Ekonomi	1	1
Subur	1	1
Menikah_W	2	2
Menikah_L	2	2

Tabel 4.5 menunjukkan nilai atau kategori pada masing-masing variabel prediktor pada x_5 dan x_{15}. Terlihat bahwa beberapa variabel memberikan nilai/kategori yang sama, seperti variabel usia, jenis kelamin, pengetahuan kondom pria, pengetahuan wanita hamil, pengetahuan NAPZA, pengetahuan bahaya HIV/AIDS, gaya pacaran, status ekonomi, pengetahuan masa subur, serta pengetahuan umur wanita dan pria menikah. Sesuai pada persamaan 2.11, maka jarak antara kategori yang sama adalah $\delta(x_b, y_b) = 0$.

Selanjutnya, pada variabel dengan kelas berbeda, selanjutnya dihitung nilai jarak VDM. Misalkan pada variabel

tempat tinggal, terlebih dahulu dibuat tabel *crosstabulation* seperti berikut.

Tabel 4.6 Ilustrasi SMOTE-N Tempat Tinggal

Tempat Tinggal	Seks	
	Tidak	Ya
Perkotaan	891	18
Pedesaan	686	18

Kemudian dihitung jarak antara dua kategori tempat tinggal sesuai dengan persamaan 2.11 sebagai berikut.

$$\delta(\text{desa},\text{kota}) = \left|\frac{891}{909} - \frac{686}{704}\right| + \left|\frac{18}{909} - \frac{18}{704}\right|$$
$$= 0,01153$$

Setelah didapatkan jarak antar kategori pada variabel tempat tinggal, selanjutnya dilakukan perhitungan jarak antar kategori pada variabel lainnya (pendidikan, konsumsi NAPZA, dan akibat menikah muda) dengan cara yang sama seperti variabel tempat tinggal. Berikut merupakan jarak antar kategori masing-masing-masing variabel.

Tabel 4.7 Ilustrasi SMOTE-N VDM

	$\delta(x_b, y_b)$
Usia	0
Tempat Tinggal	0,01153
Jenis Kelamin	0
Pendidikan	0,064404
Kondom	0
Kehamilan	0
Dengar NAPZA	0
Konsumi NAPZA	0,22481
Bahaya HIV/ADIS	0
Akibat Menikah Muda	0,088949
Gaya Pacaran	0
Status Ekonomi	0
Subur	0
Menikah_W	0
Menikah_L	0

Untuk menghitung jarak VDM antara x_5 dan x_{15}, digunakan persamaan 2.12 sebagai berikut.

$$\Delta(x_5, x_{15}) = \sum_{b=1}^{15} \delta\left(x_{5,b}, x_{15,b}\right)^2 = 0,250464$$

Jarak VDM antara x_5 dan x_{15} adalah 0,250464. Perhitungan serupa juga dilakukan antara x_5 dengan 34 data minor lainnya.

3. Menentukan *k*-tetangga terdekat atau *k*-data yang memiliki kemiripan.

Setelah dilakukan perhitungan jarak VDM antara data x_5 dengan seluruh data minoritas, selanjutnya jarak tersebut diurutkan sehingga didapatkan jarak paling kecil hingga besar. Berikut merupakan jarak x_5 dengan 15 data minor lainnya.

Tabel 4.8 Ilustrasi SMOTE-N Jarak VDM x_5

Data ke-	Jarak VDM	Data ke-	Jarak VDM
241	0.01153240	733	0.1581820
548	0.08912230	734	0.1581820
1298	0.1112093	735	0.1581820
1299	0.1112093	15	0.25046407
1300	0.1112093	16	0.25046407
1301	0.1112093	17	0.25046407
731	0.1581820	18	0.25046407
732	0.1581820		

Tabel 4.8 menunjukkan jarak VDM antara x_5 dengan 15 data minor lain secara urut dari yang terkecil. Semakin kecil jarak, maka suatu data dikatakan semakin dekat atau semnakin mirip. Dengan mengambil *k*=10, atau dengan kata lain dicari 10 data terdekat, maka data yang paling dekat dengan x_5 adalah data ke-241, ke-548, ke-1298, ke-1299, ke-1300, ke-1301, ke-731, ke-732, ke-733, dan ke-734.

4. Membentuk data sintetis

Selanjutnya, dibentuk satu data sintetis dengan mempertimbangkan kategori mayoritas masing-masing variabel *k*-data terdekat. Daimbil contoh tiga variabel pada 10 data terdekat.

Tabel 4.9 Ilustrasi SMOTE-N Data *k*-Tetangga Terdekat

Data ke-	Variabel		
	Usia	Tempat Tinggal	Jenis Kelamin
241	0	1	0
548	0	0	0
1298	0	1	0
1299	0	1	0
1300	0	1	0
1301	0	1	0
731	0	0	1
732	0	0	1
733	0	0	1
734	0	0	1

Masing-masing variabel data sintetis dibentuk berdasarkan mayoritas masing-masing variabel seperti pada Tabel 4.9. Misal pada variabel usia, terlihat bahwa 10 data terdekat bernilai “0” yang berarti berusia <19 tahun. Kemudian pada variabel tempat tinggal, 5 data terdekat memiliki kategori perkotaan dan 5 memiliki kategori pedessan. Selanjutnya variabel jenis kelamin, 6 data memiliki kategori “0” yang berarti berjenis kelamin laki-laki, dan 4 sisanya memiliki kategori “1” atau jenis kelamin perempuan.

Berdasarkan informasi tersebut, maka mayoritas pada variabel usia dan jenis kelamin adalah kategori “0”. Proprosi yang seimbang pada variabel tempat tinggal memberikan kebebasan untuk menentukan nilai/kategori, misal dipilih kategori “1”. Data sintetis baru yang terbentuk dari x_5 memiliki nilai sebagai berikut.

Tabel 4.10 Ilustasi SMOTE-N Data Sintetis

Variabel	Kategeori Data Sintetis
Usia	0 (< 19 tahun)
Tempat Tinggal	1 (Perkotaan)
Jenis Kelamin	0 (Laki-laki)

Ilustrasi pembentukan data sintetis di atas hanya memasukkan 3 variabel, pada 12 variabel lainnya dilakukan cara serupa sehingga sehingga didapatkan 1 data sintetis baru lengkap dengan 15 variabel yang terisi dan memiliki kelas respon "1" (melakukan seks pra-nikah).

Proses 1 sampai 4 diulang hingga didapatkan jumlah data *training* kelas minoritas seimbang seimbang dengan kelas mayoritas. Pada *fold* 1, jumlah awal dataset adalah 1613, dimana kelas mayoritas adalah 1577 dan kelas minoritas 36. Kemudian setelah SMOTE-N 36 data minoritas berubah menjadi 1577, sehingga setelah SMOTE-N dataset *training* menjadi 3154.

Menggunakan k = 5, maka proses SMOTE-N diulang sebanyak 5 kali pada data *training* masing-masing *fold*. Selanjutnya, dibentuk pohon dengan tahapan CART untuk setiap *fold*. Selanjutnya dihitung kebaikan model klasifikasi, yakni akurasi dan AUC pada *training* dan *testing*. Berikut merupakan kinerja klasifikasi remaja melakukan hubungan seks pra-nikah menggunakan CART dengan SMOTE-N.

Tabel 4.11 Kinerja Klasifikasi SMOTE-N

Fold	**Akurasi *Training***	**AUC *Training***	**Akurasi *Testing***	**AUC *Testing***
1	0,932784	0,932784	0,883085	0,940355
2	0,93086	0,930872	0,933003	0,91145
3	0,928639	0,928648	0,8933	0,945432
4	0,932741	0,932741	0,908416	0,898875
5	0,937837	0,937848	0,888338	0,888607
Rata-rata	0,932572	0,932578	0,901228	0,916944

Berdasarkan Tabel 4.11, terlihat bahwa akurasi dan AUC pada *training* maupun *testing* data hasil SMOTE-N sudah baik karena memiliki nilai lebih dari 90%. Nilai AUC pada SMOTE-N sudah naik jauh dibanding pada data *imbalance*, yakni dari 73,73% menjadi 91,69%. Hal tersebut menandakan bahwa menggunakan SMOTE-N, *imbalance* sudah mulai teratasi.

4.3.2 SMOTE-N-ENN

Metode penanganan *imbalance* SMOTE-N dapat digabung dengan metode *undersampling*, yang disebut dengan *hybrid* SMOTE-N. Pada penelitian kali ini akan dilakukan penggabungan SMOTE-N dan *Edited Nearest Neighbors* (SMOTE-N-ENN). Pada ENN, data yang dianggap sebagai *noise* akan dihapus. Secara singkat, data *imbalance* akan dilakukan *oversampling* SMOTE-N lalu *undersampling* menggunakan ENN untuk mengurangi *noise*.

Sebagai ilustrasi, data hasil SMOTE-N pada sub-bab 4.3.1 akan dilakukan *undersampling* menggunakan ENN.

1. Menghitung jarak *Overlap*

Tahap awal yang dilakukan adalah mencari 10 tetangga terdekat masing-masing data menggunakan jarak *Overlap* pada data SMOTE-N. Menggunakan *fold* 1, data awal dalam ENN bukanlah data *training* 1613, melainkan data hasil SMOTE-N yang berjumlah 3154. Sebagai ilustrasi perhitungan jarak *Overlap*, diambil observasi pertama dan ke-5, berikut merupakan jarak *Overlap*-nya.

Tabel 4.12 Data Ilustrasi SMOTE-N-ENN

Variabel	Data	
	x_1	x_5
Usia	0	0
Tempat Tinggal	1	0
Jenis Kelamin	0	0
Pendidikan	1	1
Kondom	1	1
Kehamilan	2	2
Dengar NAPZA	1	1
Konsumi NAPZA	0	1
Bahaya HIV/ADIS	1	1
Akibat Menikah Muda	1	1
Gaya Pacaran	3	4
Status Ekonomi	1	1
Subur	1	1
Menikah_W	2	2
Menikah_L	2	2

Tabel 4.12 menunjukkan bahwa nilai/kategori pada variabel tempat tinggal, konsumsi NAPZA, dan gaya berpacaran berbeda. Maka jarak *overlap* ketiga variabel tersebut masing-masing adalah 1, sedangkan lainnya 0. Oleh karena itu, jarak *overlap* x_1 dan x_5 sesuai dengan Persamaan 2.14 adalah sebagai berikut.

$$\Delta(x_1, x_5) = \sum_{i=1}^{14} overlap(V_1, V_2) = 3$$

Perhitungan jarak *overlap* dilakukan pada semua data, yakni x_1 dengan 3153 data lainnya.

2. Menentukan *k*-tetangga terdekat atau *k*-data yang memiliki kemiripan

Setelah dilakukan perhitungan jarak *overlap,* berikut merupakan jarak x_1 dengan 15 data lainnya.

Tabel 4.13 Ilustrasi SMOTE-N-ENN *k*-Tetangga Terdekat

Data ke-	Jarak	Data ke-	Jarak
2	0	363	1
3	0	364	1
4	0	365	1
981	0	366	1
1276	0	368	1
1277	0	382	1
276	1	446	1
277	1		

Jarak *overlap* pada Tabel 4.13 sudah terurut dari yang paling dekat, dimana $\Delta(X, Y) = 0$ berarti data identik. Menggunakan k=10, diambil 10 data terdekat dengan x_1, yaitu data ke-2, ke-3, ke-4, ke-981, ke-1276, ke-1277, ke-276, ke-277, ke-363, dan ke-364.

3. Menentukan data *noise*

Data ke-1 atau x_1 memiliki kelas respon "0" atau tidak melakukan seks pra-nikah. Kemudian berdasarkan 10 tetangga/data terdekat, semua data masuk dalam kategori tidak melakukan

hubungan seksual. Informasi tersebut menunjukkan bahwa kelas respon x_1 sama dengan kelas respon *k*-tetangga terdekatnya. Karena memiliki kelas respon yang sama dengan mayoritas 10 tetangga terdekat, maka x_1 tidak dianggap *noise* dan tidak dihapus.

Tahap 1 sampai 3 diulang pada semua data hingga sudah tidak ada data yang mungkin terhapus. Setelah itu, dilakukan analisis klasifikasi menggunakan data hasil SMOTE-N-ENN. Berikut merupakan akurasi dan AUC yang menggambarkan kinerja klasifikasi.

Tabel 4.14 Kinerja Klasifikasi SMOTE-N-ENN

Fold	**Akurasi *Training***	**AUC *Training***	**Akurasi *Testing***	**AUC *Testing***
1	0,962335	0,9623	0,945274	0,972081
2	0,960665	0,961595	0,960298	0,925409
3	0,954003	0,954365	0,925558	0,961929
4	0,960283	0,96047	0,94802	0,919128
5	0,961185	0,9604	0,947891	0,810491
Rata-rata	0,959694	0,959826	0,945408	0,917808

Tabel 4.14 menunjukkan bahwa kinerja klasifikasi remaja melakukan hubungan seksual pra-nikah, terutama AUC, mengalami kenaikan daripada data asli maupun SMOTE-N. AUC *testing* naik dari 73,73% menjadi 91,78% Kombinasi SMOTE-N dan ENN dapat menghilangkan *noise* pada data, atau data yang tidak seragam dengan *k* tetangga terdekatnya. Namun, nilai akurasi yang lebih tinggi menandakan masih terdapatnya *imbalance* pada data.

4.3.3 ADASYN-N

Hampir sama dengan kedua metode sebelumnya, prinsip dari ADASYN-N adalah membuat data sintetis berdasarkan *k*-data terdekat menggunakan jarak VDM. Namun terlebih dulu dilakukan perhitungan jumlah data baru yang akan dibentuk. Berikut merupakan ilustrasi ADASYN-N pada *fold* 1.

1. Menentukan *G*

Pada *fold* 1 jumlah kelas mayoritas adalah 1577 dengan kelas minoritas 36. Selanjutnya digunakan $\beta = 1$, yang berarti data *training* nantinya seimbang sempurna. Oleh karena itu, jumlah data sintetis yang terbentuk sesuai persamaan 2.16 adalah sebagai berikut.

$$G = (1577 - 36) \times 1$$
$$= 1541$$

2. Menentukan *k*-tetangga terdekat

Perhitungan jarak dalam ADASYN-N sama dengan SMOTE-N, yakni menggunakan jarak VDM. Namun jika pada SMOTE-N, tetangga terdekat dihitung hanya antar kelas minoritas, pada ADASYN-N tetangga terdekat dihitung dari keseluruhan data *training*. Misal data terpilih atau data yang diambil adalah data ke-5. 10 tetangga terdekat data ke-5 adalah data ke-1, ke-2, ke-3, ke-4, ke-14, ke-15, ke-16, ke-17, ke-18, dan ke-29.

3. Menghitung r_i dan $\hat{r}_1$

Dari ke-10 tetangga terdekat, terdapat 4 yang merupakan kelas minoritas, yakni 15,16, 17, dan 18, sedangkan 6 sisanya adalah kelas mayoritas, sehingga kalkulasi rasio r_i adalah sebagai berikut.

$$r_1 = \frac{\Delta_1}{k} = \frac{6}{10} = 0,6$$

Proses tersebut diulang pada seluruh data minoritas, sehingga masing-masing data minoritas memiliki rasio r_i masing-masing. Selanjutnya dilakukan normalisasi r_1 menggunakan persamaan 2.18, dengan $\sum_{i=1}^{m_s} r_i$ = dan 18,5 sehingga didapatkan $\hat{r}_1$ = 0,03243.

4. Menghitung jumlah data sintetis masing-masing data minor

Setelah didapatkan jumlah total data baru dan $\hat{r}_i$ masing-masing data minor, maka selanjutnya dihitung jumlah data baru yang harus dibangkitkan untuk masing-masing data minoritas.

Menggunakan persamaan 2.20, jumlah *inscance* sintetis yang terbentuk dari data 5 adalah :

$$\hat{r}_1 \times G = 0,03243 \times 1541 \approx 50$$

Tahap terakhir, data 5 selanjutnya direplikasi sebanyak 50 kali. Proses tersebut diulang pada setiap data kelas minoritas hingga didapatkan dataset *training* baru yang seimbang. Berikut merupakan kinerja klasifikasi CART dengan ADASYN-N.

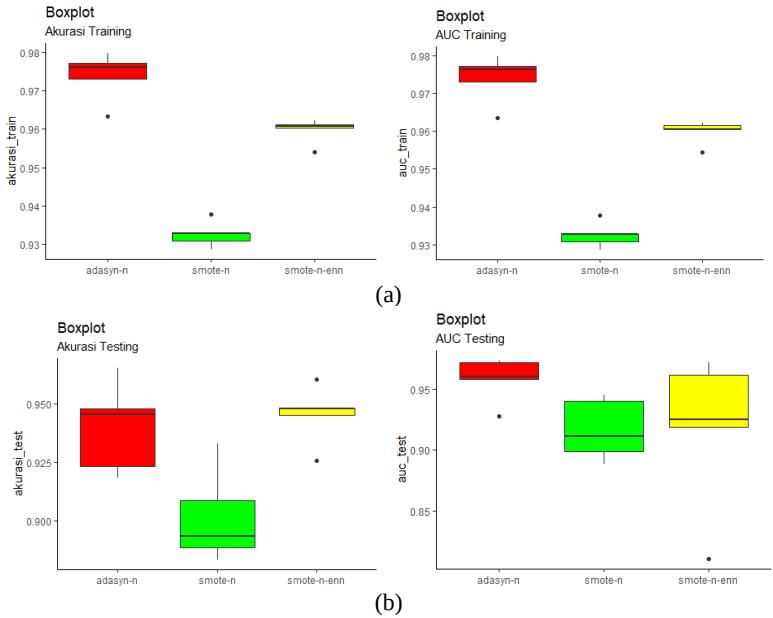
Tabel 4.15 Kinerja Klasifikasi CART dengan ADASYN-N

Fold	Akurasi *Training*	AUC *Training*	Akurasi *Testing*	AUC *Testing*
1	0,97978	0,97971	0,94776	0,97335
2	0,97723	0,977172	0,965261	0,927947
3	0,973076	0,97305	0,923077	0,96066
4	0,963422	0,963515	0,918317	0,958228
5	0,976153	0,976221	0,945409	0,972081
Rata-rata	0,973932	0,973933	0,939965	0,958453

Berdasarkan Tabel 4.15, terlihat bahwa AUC yang dihasilkan data *training* dan *testing* relatif tinggi, yakni 97,39% dan 95,84%. Jika dilihat pada masing-masing *fold*, terlihat bahwa dua *fold* dengan AUC *testing* tertinggi adalah *fold* 1 dan 5. Hal tersebut berkebalikan dengan kedua metode sebelumnya, dimana pada *fold* 5 menjadi *fold* dengan AUC *testing* terendah. Hal tersebut menunjukkan bahwa ADASYN-N mampu meningkatkan kinerja klasifikasi, meskipun pada *fold* dengan data *testing* yang tidak mudah.

4.4 Perbandingan Metode

Perbandingan kebaikan hasil klasifikasi perilaku remaja melakukan hubungan seksual pra-nikah di Jawa Timur menggunakan SMOTE-N, SMOTE-N-ENN, dan ADASYN-N dengan CART terdapat pada Gambar 4.13.

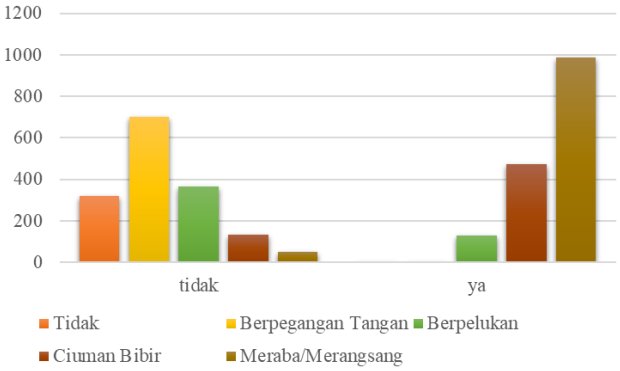


Gambar 4.13 Perbandingan Nilai Akurasi dan AUC (a) *Training* (b) *Testing*

Berdasarkan Gambar 4.13, terlihat bahwa metode terbaik dalam mengatasi adanya *imbalanced data* adalah *Adaptive Synthetic Nominal* (ADASYN-N). Hal tersebut dapat diketahui dari nilai akurasi dan AUC pada metode ADASYN-N yang lebih tinggi jika dibandingkan dua metode lainnya, baik dalam *training* maupun *testing*. Rata-rata AUC *testing* klasifikasi remaja melakukan hubungan seksual pra-nikah di Jawa Timur dengan penanganan *imbalance* ADASY-N adalah 95,84%. Jika dilihat dari nilai akurasi, SMOTE-N-ENN jauh lebih unggul dibanding SMOTE-N, namun berdasarkan nilai AUC *testing* selisih yang diberikan tidak terlalu besar. Selain itu, jika dibandingkan hasil klasifikasi antara data *training* dan *testing* baik akurasi maupun AUC, tampak terlihat adanaya *overfitting*, dimana hasil klasifikasi pada data *training* lebih baik. Namun pada klasifikasi menggunakan ADASYN-N, *overfitting* tidak lebih besar daripada klasifikasi menggunakan SMOTE-N dan SMOTE-N-ENN. Hal

tersebut dapat dilihat dari selisih nilai akurasi dan AUC data *training* dan *testing*.

Penggunakan *resampling* atau penambahan data sintetis dikhawatirkan dapat mengubah pola data. Berikut merupakan salah satu karakteristik data setelah ADASYN-N untuk menggambarkan bagaimana keadaan data setelah ditambahkan data sintetis.



Gambar 4.14 Karakteristik Gaya Pacaran ADASYN-N

Gambar 4.14 menunjukkan bahwa pada remaja yang melakukan seks pra-nikah, gaya pacaran yang muncul adalah berpelukan, ciuman bibir, dan meraba merangsang. Informasi yang dihasilkan Gambar 4.14 sama dengan karakteristik awal data *imbalance*. Hal tersebut berarti bahwa adanya penambahan data sintetis tidak mengubah pola data. Pada ADASYN-N, data minor yang dibangkitkan berasal dari data minor itu sendiri, sehingga pola data yang dihasilkan tidak mungkin berubah.

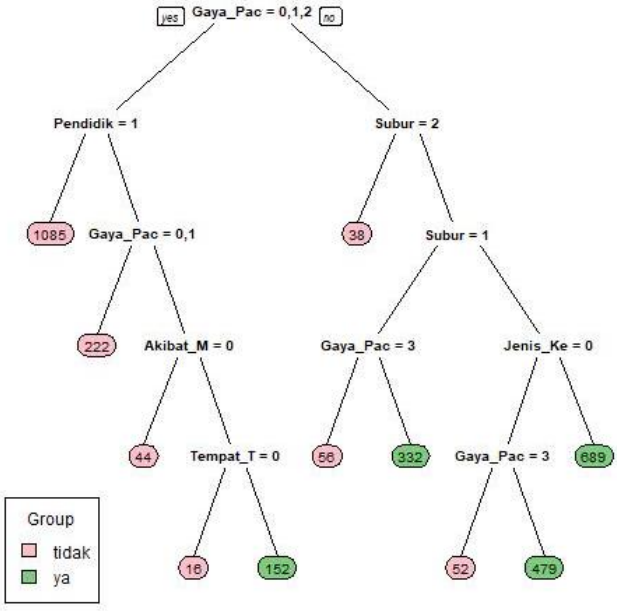
4.5 Metode Terbaik

Berdasarkan pembahasan pada sub-bab 4.4, didapatkan metode penanganan *imbalance* yang menghasilkan kinerja klasifikasi terbaik adalah ADASYN-N. Terdapat 5 pohon yang dihasilkan dalam klasifikasi CART dengan ADASYN-N karena digunakan 5-*fold cross validation*. Untuk memilih pohon mana yang merupakan pohon klasifikasi terbaik, dapat dilihat pada kinerja klasifikasi *testing* masing-masing *fold*.

Tabel 4.16 Ketepatan Klasifikasi *Testing* 5 *Fold* ADASYN-N

Fold	Akurasi	AUC
1	0,94776	**0,97335**
2	**0,965261**	0,927947
3	0,923077	0,96066
4	0,918317	0,958228
5	0,945409	0,972081

Tabel 4.16 menunjukkan bahwa klasifikasi pada *fold* 2 memberikan akurasi tertinggi, sedangkan AUC tertinggi terdapat pada *fold* 1. Namun, pada penelitian ini hasil terbaik dinilai dari AUC, sehingga hasil terbaik terdapat pada *fold* 1. Oleh karena itu, untuk interpretasi selanjutnya difokuskan pada pohon klasifikasi pada *fold* dengan nilai kinerja klasifikasi paling tinggi. Berikut merupakan pohon klasifikasi *fold* 1.



Gambar 4.15 Pohon Keputusan CART Metode Terbaik

Gambar 4.15 menunjukkan bahwa variabel yang pertama keluar dan menjadi simpul akar adalah gaya berpacaran. Hal tersebut memberikan informasi bahwa gaya berpacaran menjadi variabel yang paling dapat mengklasifikasikan perilaku seks pra-nikah remaja di Jawa Timur. Pada pohon di atas, terdapat variabel yang menjadi simpul lebih dari satu kali, yakni gaya berpacaran dan pengetahuan masa subur, karena masih terdapat penurunan heterogenitas pada simpul tersebut. Angka pada simpul terminal menunjukkan jumlah data yang tergolong pada kelas simpul terminal.

Berdasarkan pohon keputusan pada Gambar 4.15, berikut merupakan *gain nodes* atau *rule* dari pohon keputusan metode terbaik. *Rule* dapat digunakan untuk mengidentifikasi bagaimana karakteristik remaja hingga digolongkan pada kelas melakukan seks pra-nikah maupun tidak.

Tabel 4.17 *Gains Nodes* Remaja Melakukan Seks Pra-Nikah

No	Karateristik dan Perilaku
1.	a. Memiliki gaya pacaran ciuman bibir atau meraba/merangsang, dan b. Tidak memiliki pengetahuan mengenai masa subur wanita, dan c. Berjenis kelamin perempuan
2.	a. Memiliki gaya pacaran meraba/merangsang, dan b. Tidak memiliki pengetahuan mengenai masa subur wanita, dan c. Berjenis kelamin laki-laki.
3.	a. Memiliki gaya pacaran meraba/merangsang, dan b. Memiliki pengetahuan mengenai masa subur wanita yang kurang tepat.
4.	a. Memiliki gaya pacaran berpelukan, dan b. Memiliki pendidikan terakhir SLTP, dan c. Mengetahui akibat menikah muda, dan d. Bertempat tinggal di pedesaan.

*nb : masa subur yang tepat adalah di antara dua haid

Tabel 4.18 *Gains Nodes* Remaja Tidak Melakukan Seks Pra-Nikah

No	Karateristik dan Perilaku
1.	a. Memiliki gaya pacaran ciuman bibir atau meraba/merangsang, dan b. Memiliki pengetahuan masa subur wanita yang tepat.
2.	a. Memiliki gaya pacaran ciuman bibir, dan b. Tidak mengetahui masa subur wanita, dan c. Berjenis kelamin laki-laki.
3.	a. Memiliki gaya pacaran meraba/merangsang, dan b. Memiliki pengetahuan mengenai masa subur wanita yang kurang tepat.
4.	a. Memiliki gaya pacaran tidak mekukan apa-apa/berpegangan tangan/berpelukan, dan b. Berpendidikan terakhir minimal SLTA.
5.	a. Memiliki gaya pacaran tidak melakukan apa-apa/berpegangan tangan, dan b. Berpendidikan terakhir maksimal SLTP.
6.	a. Memiliki gaya pacaran berpelukan, dan b. Berpendidikan terakhir maksimal SLTP.Mengetahui bahaya HIV/AIDS, dan c. Tidak mengetahui resiko menikah muda.
7.	a. Memiliki gaya pacaran berpelukan, dan b. Berpendidikan terakhir maksimal SLTP.Mengetahui bahaya HIV/AIDS, dan c. Mengetahui resiko menikah muda, dan d. Tinggal di daerah perkotaan.

*nb : masa subur yang tepat adalah di antara dua haid

Pohon keputusan dan informasi *gain* atau *rule* memberikan informasi bahwa variabel yang dapat memprediksi remaja melakukan seks pra-nikah adalah gaya berpacaran, pengetahuan mengenai masa subur, pengetahuan mengenai resiko menikah muda, jenis kelamin, pendidikan terakhir, dan daerah tempat tinggal. Terlihat bahwa gaya pacaran yang sudah mengarah ke kontak fisik secara intens cenderung mengarah pada hubungan seksual pra-nikah. Hal tersebut terlihat dari jika gaya pacaran remaja adalah tidak melakukan apa-apa/berpegangan tangan maka cenderung tidak melakukan seks pra-nikah.

Berdasarkan Tabel 4.17, terlihat bahwa baik laki-laki maupun perempuan, jika memiliki gaya pacaran ciuman bibir atau meraba/merangsang dan tidak memiliki pengetahuan masa subur remaja secara tepat (atau tidak tahu sama sekali), maka akan cenderung melakukan seks pra-nikah. Gaya pacaran yang sudah yang sudah sangat intens namun tidak diiringi dengan informasi KRR yang tepat dapat menjerumuskan remaja.

Selain ciuman bibir dan meraba/merangsang, seorang remaja dengan gaya pacaran berpelukan juga akan cenderung melakukan seks pra-nikah. Remaja yang tinggal di pedesaan, tingkat pendidikan terakhir maksimal SLTP dengan gaya pacaran berpelukan dan mengetahui akibat menikah muda cenderung akan melakukan seks pra-nikah. Walaupun gaya pacaran yang belum terlalu parah, namun lingkungan tempat tinggal dan tingkat pendidikan juga dapat berpengaruh terhadap perilaku seksual remaja. Remaja yang tinggal di pedesaan dengan pendidikan yang rendah cenderung mengabaikan *sex education* baik berupa kesehatan maupun akibat menikah muda seperti KDRT dan perceraian.

Sex education merupakan informasi yang sangat penting bagi remaja, dan harus disampaikan secara tepat. Dengan adanya pengetahuan tersebut, remaja menjadi lebih waspada mengenai bahaya yang mungkin terjadi. Selain itu, wawasan mengenai keluarga juga tidak kalah penting. Jika remaja sudah mengetahui akibat dari menikah muda, selayaknya akan lebih waspada terhadap hal menyangkut seks. Namun ternyata terdapat pengabaian pengetahuan, yang mana remaja dengan pengetahuan resiko menikah muda ternyata cenderung melakukan seks pra-nikah.

Informasi mengenai pengetahuan seksual hendaknya diberikan pada remaja sebelum masa pubertas dan tidak hanya di lingkungan sekolah. Pada pohon keputusan terlihat bahwa remaja dengan pendidikan rendah lebih cenderung melakukan hubungan

seksual pra-nikah dibanding yang memiliki pendidikan tinggi. Hal tersebut mengindikasikan bahwa mungkin remaja yang melakukan seks belum mengenyam pendidikan formal sama sekali, sehingga sangat tepat jika *sex education* disampaikan secara langsung dalam lingkungan tempat tinggal.

(Halaman ini sengaja dikosongkan)

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, maka diperoleh kesimpulan bahwa metode penanganan *imbalance* yang memiliki kinerja klasifikasi paling baik dengan *base classifier* CART adalah ADASYN-N (*Adaptive Synthetic Nominal*), dengan rata-rata kinerja klasifikasi tertinggi yaitu akurasi sebesar 93,99% dan AUC sebesar 95,84%. Selain itu, faktor yang dapat memprediksi seorang remaja melakukan hubungan seksual pra-nikah adalah gaya pacaran, *sex education* (masa subur), informasi keluarga (resiko menikah muda), serta demografi sosial (pendidikan, jenis kelamin, dan daerah tempat tinggal).

Secara garis besar, karakteristik remaja cenderung melakukan seks pra-nikah terbagi berdasarkan tingkat gaya pacaran. Pada tingkat gaya pacaran ciuman bibir atau meraba/merangsang, baik laki-laki maupun perempuan dengan gaya pacaran tersebut dan tidak memiliki pengetahuan masa subur dengan tepat akan cenderung melakukan seks pra-nikah. Kemudian pada remaja yang tinggal di pedesaan, pendidikan terakhir maksimal SLTP, memiliki gaya pacaran berpelukan, dan mengetahui akibat menikah muda juga cenderung melakukan seks pra-nikah.

5.2 Saran

Berdasarkan kesimpulan yang telah diperolah, faktor yang pertama dapat memprediksi remaja melakukan hubungan seksual pra-nikah di Jawa Timur adalah gaya pacaran, sehingga saran yang dapat diberikan adalah meningkatkan pengawasan, terutama di lingkungan keluarga, terhadap gaya pacaran remaja. Jika gaya pacaran sudah mengarah pada kontak fisik, maka remaja tersebut perlu perhatian khusus agar tidak terjerumus ke dalam seks pra-nikah. Selain itu, faktor yang juga dapat memprediksi perilaku seksual pra-nikah adalah informasi terkait KRR (*sex education*)

dan resiko menikah muda. Oleh karena itu, sebaiknya informasi tersebut lebih disebar luaskan dan disampaikan secara tepat.

Sex education dapat ditekankan pada remaja dengan jenjang pendidikan SLTP atau kurang, khususnya pada daerah pedesaan. Penyampaian informasi juga dapat dilakukan langsung pada komunitas, tidak hanya terpaku pada lingkungan sekolah. Salah satu program BKKBN, Bina Keluarga Remaja merupakan wadah yang tepat, sehingga edukasi yang disampaikan tidak hanya pada remaja yang akan atau sedang pubertas, namun juga pada orang tua remaja. Kemudian, saran untuk penelitian selanjutnya adalah dapat ditambahkan variabel lain yang berhubungan dengan media sosial, sehingga terdapat tambahan informasi mengenai paparan media sosial, khususnya pornografi.

DAFTAR PUSTAKA

Adiansyah, M. C. N., 2017. Perbandingan Metode CART dan Analisis Regresi Logistik serta Penerapannya untuk Klasifikasi Ketertinggalan Kabupaten dan Kota di Indonesia. *Skripsi.*

Badan Kependudukan dan Keluarga Berencana Nasional, 2018. *Survei Kinerja dan Akuntabilitas Program KKBPK (SKAP) Remaja,* Jakarta: Badan Kependudukan dan Keluarga Berencana Nasional.

Bekkar, M., Djemaa, H. K. & Alitouche, T. A., 2013. Evaluation Measure for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Aplications.*

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J., 1993. *Classification and Regression Trees.* New York: Chapman Hall.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelegence Research,* Volume 16, pp. 321-257.

Cost, S. & Salzberg, S., 1993. A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. *Machine Learning,* Volume 10.

Dave, V. R., Makwana, N. R., Yadav, B. S. & Yadav, S., 2013. A Study on High Risk Premarital Sexual Behavior of College Going Male Students in Jamnagar City of Gujarat, India. *International Journal of High Risk Behaviors and Addiction.*

Gokgoz, E. & Subasi, A., 2015. Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT. *Biomedical Signal Processing and Control,* Volume 18, pp. 138-144.

Harbia, Multazam, M. & Asrina, A., 2018. Dampak Penyalahgunaan Narkotika, Psikotropika dan Zat Aditif

Lainnya (NAPZA) terhadap Perilaku Seks Pranikah. *Jurnal Kesehatan,* Volume 3.

Hardiansyah, M. A., 2018. *Efek Pengetahuan tentang NAPZA dan HIV/AIDS terhadap Perilaku Seks Pranikah (Siswa Kelas X SMA Negeri Se-Kotamadya Yogyakarta,* Yogyakarta: Universitas Negeri Yogyakarta.

Harnani, Y., Alamsyah, A. & A., 2018. Premarital Sex among Adolescent Street Children in Pekanbaru. *International Journal of Public Health Science (IJPHS),* Volume 7, pp. 22-26.

Hastuti, D., Agung, S. S. & A., 2013. *Kajian Karakteristik Remaja Desa-Kota, Sekolah serta Keluarga untuk Mengatasi Perilaku Anti Sosial Remaja SMK di Kota dan Kabupaten Bogor.* Bogor, Institut Pertanian Bogor.

Jeyarani, D. S., Anushya, G., Rajeswari, R. R. & Pethalaksmi, A., 2013. *A Comparative Study of Decision Tree and Naive Bayesian Classifiers on Medical Dataset.* s.l., International Conference on Computing and information Technology.

Kasim, F., 2014. Dampak Perilaku Seks Beresiko terhadap Kesehatan Reproduksi dan Penangannya. *Jurnal Studi Pemuda.*

Lewis, R. J., 2000. *An Introduction to Classification and Regression Tree (CART) Analysis.* s.l.:Presented at the 2000 Anual Meeting of Society for Academy Emergency Medicine in San Fransisco, California.

Merdeka.com, 2018. *Peristiwa.* [Online] Available at: https://www.merdeka.com/peristiwa/beredar-video-mesum-2-pelajar-di-karawang-berdurasi-3-menit.html

Naroozi, M. et al., 2014. Premarital sexual relationships: Explanation of the actions and functions of family. *Iranian Journal of Nursing and Midwifery Research.*

Nasution, S. L., 2012. Pengaruh Pengetahuan tentang Kesehatan Reroduksi Remaja terhadap Perilaku Seksual Pranikah di Indonesia. *Widyariset,* Volume 15.

Nian, R., 2018. *medium.com.* [Online] Available at: https://medium.com/@ruinian/an-introduction-to-adasyn-with-code-1383a5ece7aa [Accessed 21 10 2019].

Pattipeilohy, W. F., Wibowo, A. & Utari, D. R., 2017. *Pemodelan dan Prototipe Sistem Informasi untuk Prediksi Pembaharuan Polis Asuransi Mobil Menggunakan Algoritma C.45.* Jakarta Selatan, s.n., p. 793.

Peters, T., 2018. Binary Classification on Highly Imbalance Dataset. *Thesis.*

Rahayu, S., Adji, T. B. & Setiawan, N. A., 2017. Analisis Perbandingan Metode Over-Sampling Adaptive Synthetic Nominal (ADASYN-N). s.1. *Conference on Information Technology and Electrical Engineering*

Raju, K. S., Murty, M. R., Rao, M. V. & Satapathy, S. C., 2018. Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction. *International Journal of Pure and Applied Mathematics,* Volume 118.

Rosdarni, Dasuki, D. & Waluyo, S. D., 2015. Pengaruh Faktor Personal terhadap Perilaku Seksual. *Jurnal Kesehatan Masyarakat Nasional,* Volume 9.

Sarwono, 2011. *Psikologi Remaja.* Jakarta: Rajawali Pers.

Teferra, T. B., Erena, A. N. & Kabede, A., 2015. Prevalence of Premarital Sexual Practice and Associated Factor among Undergraduate Helath Science Students of Madawalabu Uniersity, Bale Goba, South East Ethiopia : Institution Based Cross Sectional Study.

Triningsih, R. W., Widjanarko, B. & Istiarti, V. T., 2015. Faktor-faktor yang Berpengaruh terhadap Praktik Seks Pranikah pada Remaja di SMA Dekat Lokalisasi di Wilayah

Kabupaten Malang. *Jurnal Promosi Kesehatan Indonesia*, Volume 10.

Vluymans, S., 2014. Instance Selection for Imbalanced Data. *Master Thesis.*

Wahyuni, S. & Fahmi, I., 2019. Determinan Perilaku Seksual Pra-Nikah Remaja Pria di Indonesia Hasil SDKI. *Euclid*, Volume 6.

Waluyo, A., Mukid, M. A. & Wuryandari, T., 2014. Perbandingan Klasifikasi Nasabah Kredit Menggunakan Regresi Logistik Biner dan CART (Classification and Regression Trees). *Media Statistika*, Volume 7.

Willis & Sofyan, S., 2005. *Remaja dan Masalahnya : Mengupas Berbagai Bentuk Kenakalan Remaja Seperti Narkoba, Free Sex, dan Pemecahannya.* Bandung: Alfabeta.

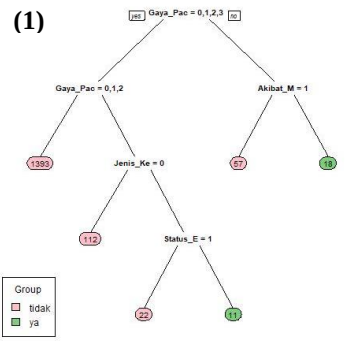
LAMPIRAN

Lampiran 1. Data yang Digunakan

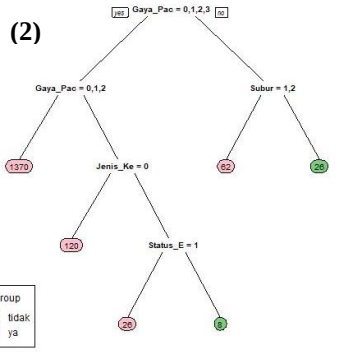
No	X1	X2	X3	X4	⋯	X13	X14	X15	Y	weight
1	0	0	1	1		3	0	3	0	4
2	0	0	1	0		1	1	4	1	1
3	0	1	1	0		4	0	1	0	1
4	1	1	1	0		3	0	1	0	5
5	0	0	1	1		0	1	0	0	4
6	0	1	1	0		0	0	1	0	1
7	0	1	1	0		1	0	2	0	1
8	0	0	0	1		3	0	4	1	4
9	0	0	0	0		0	1	1	0	1
10	0	1	1	0		4	0	2	0	1
11	0	1	1	0		1	0	1	0	1
12	0	1	1	0		4	0	0	0	2
13	0	1	1	1		3	0	1	0	3
14	0	1	1	0		0	0	1	0	3
15	0	1	0	1		3	0	0	0	3
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
682	0	0	1	0		3	0	1	0	1
683	0	0	1	1		3	0	2	0	3
684	0	0	1	0		3	0	1	0	1
685	0	1	0	0		3	0	1	0	7
686	0	1	1	0		3	0	2	0	7
687	0	0	0	0		2	0	1	0	3
688	0	1	1	0		2	0	1	0	3
689	0	1	1	0		3	0	1	0	3
690	0	0	1	0		1	0	2	0	1
691	0	1	1	0		3	0	0	0	1
692	0	0	1	0		3	0	4	0	1
693	0	1	1	0		4	0	4	0	1
694	1	1	1	1		6	0	0	0	8
695	0	0	1	1		0	0	1	0	2
696	0	1	0	0		3	0	1	0	1

Lampiran 2. Pohon Keputusan CART *imbalanced data*

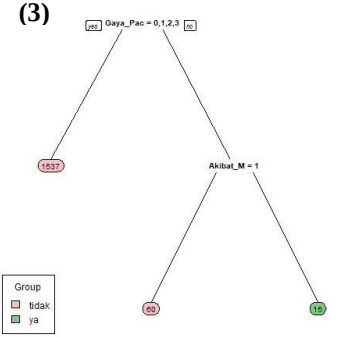
(1)



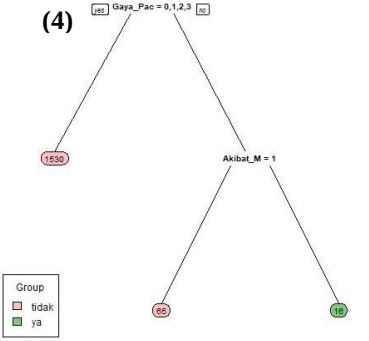
(2)



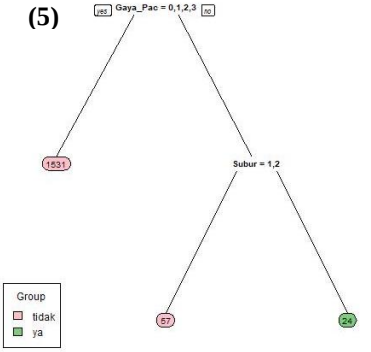
(3)



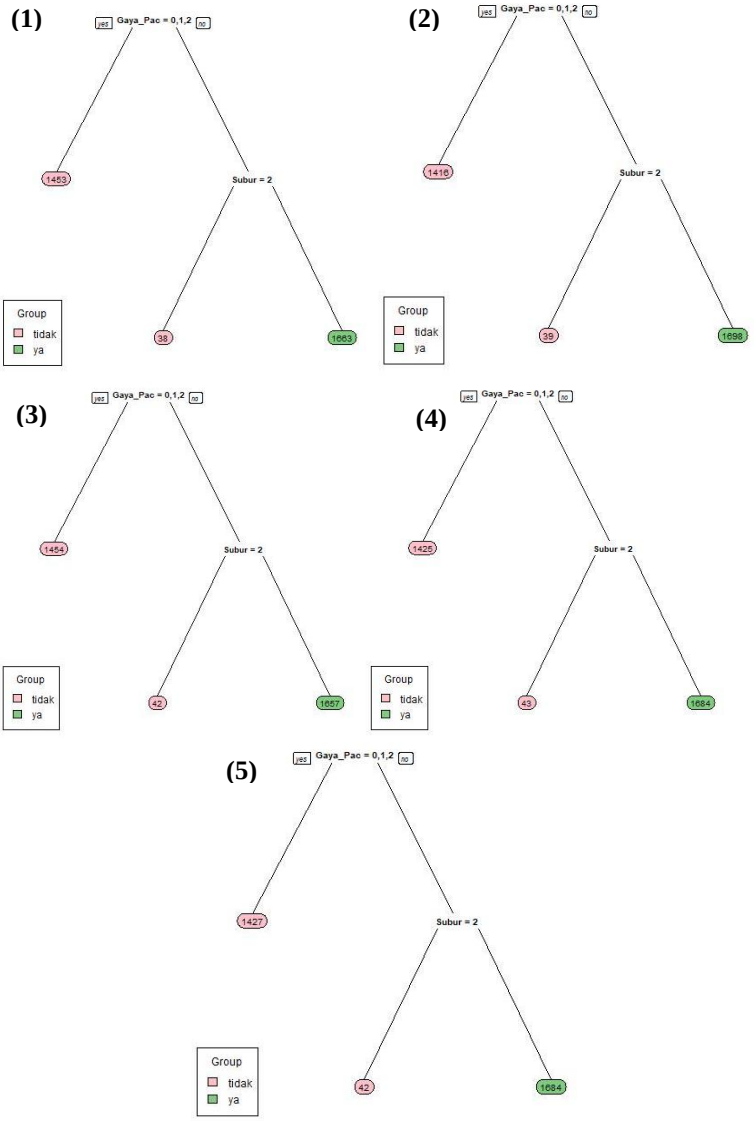
(4)



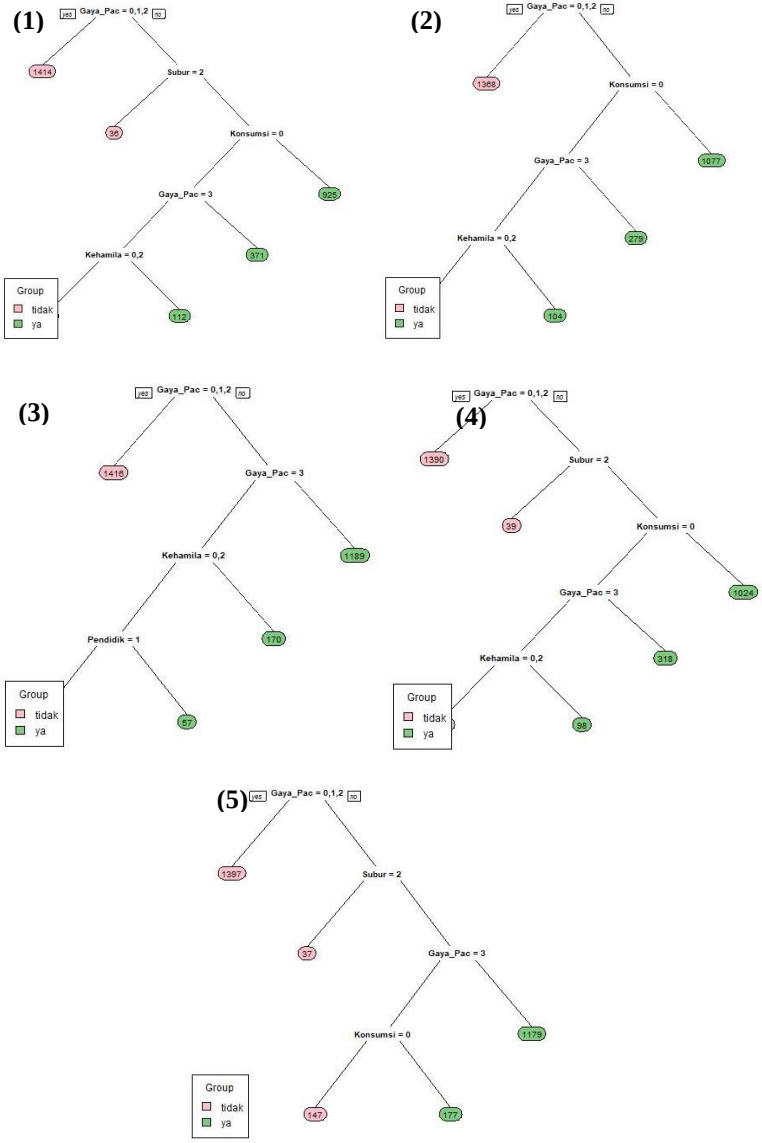
(5)



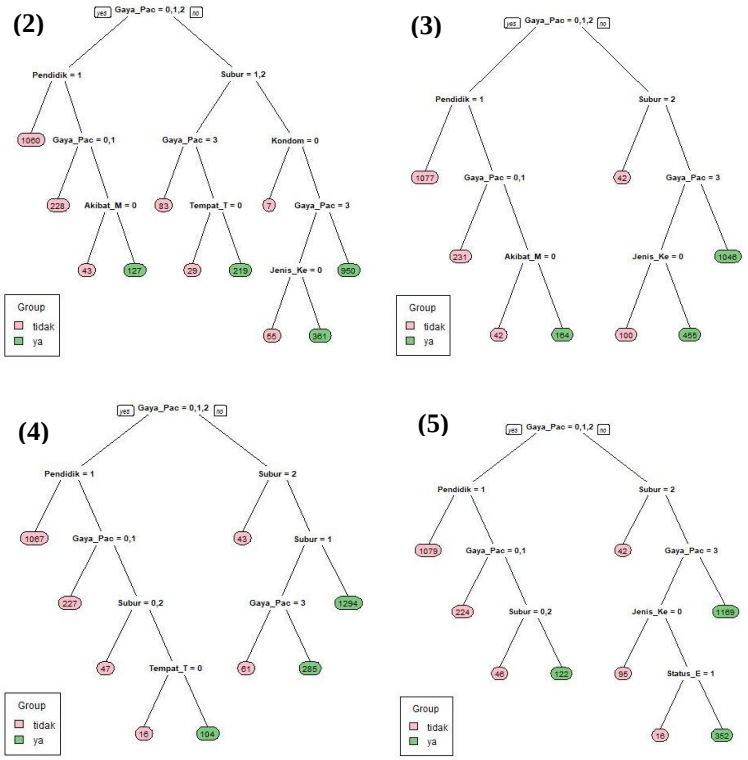
Lampiran 3. Pohon Keputusan CART dengan SMOTE-N



Lampiran 4. Pohon Keputusan CART dengan SMOTE-N-ENN



Lampiran 5. Pohon Keputusan CART dengan ADASYN-N



Lampiran 6. *Syntax* R *Function* ADASYN-N

```
adasynn <- function(data, neig)
{
  li <- class.freq(data, "Seks")
  tgt <- which(data$Seks == li[[1]][2])
  ms <- length(tgt)
  ml <- nrow(data) - ms
  d <- ms/ml
  q <- ncol(data)
  G <- (ml-ms)

  #menghitung jarak amatan kelas minor pada semua data
  kNNs <- neighbours("Seks", data, dist = "HVDM", k=neig)

  er <- array(0,c(length(tgt),1))

  for (i in 1:length(tgt)) {
    r <- 0
    for (j in 1:10) {
     r <- r+length(which(data[kNNs[tgt[i],j],16] ==
          li[[1]][1]))
    }
    er[i] <- r/10
  }

  newr <- array(0,c(length(tgt),1))
  g <- array(0,c(length(tgt),1))
  for (k in 1:length(tgt)) {
    newr[k] <- er[k]/sum(er[,1])
    g[k] <- round(newr[k]*G)
  }

  datamin <- array(0,c(length(tgt),q))
  datamin <- data[data$Seks == li[[1]][2],]

  datamin <- cbind(datamin,g)
  datamin$g <- as.numeric(datamin$g)

  #replicate
  datamin <- datamin[rep(seq(nrow(datamin)), datamin$g),]
  datamin <- datamin[ , !(names(datamin) %in% "g")]

  new <- rbind(trainData,datamin)
  new
}
```

Lampiran 7. *Syntax* R CART

```
setwd("E:/TA/plot 12345")
#data
data <- read.csv("E:/TA/data gaya pacar jadi
1.csv",sep=";")
#replicate
data <- data[rep(seq(nrow(data)), data$haha),]
data <- data[ , !(names(data) %in% "haha")]

set.seed(12345)
library(caret)
library(rpart)
library(rpart.plot)
library(dplyr)
library(UBL)

#factorized
for (i in 1:ncol(data)) {
  data[,i] <- as.factor(data[,i])
}
fold <- createFolds(factor(data$Seks), k=5,
         list=FALSE)
auc <- array(0,c(5,3))
akuras <- array(0,c(5,3))
auct <- array(0,c(5,3))
akurast <- array(0,c(5,3))

class.freq <- function (dat, tgt)
{
  names <- sort(unique(dat[, tgt]))
  li <- list(names, sapply(names, function(x)
length(which(dat[, tgt] == x))))
  li
}
```

```
for (o in 1:5) {
  listdat <- list()
  testIndex <- which(fold == o, arr.ind = TRUE)
  testData <- data[testIndex, ]
  trainData <- data[-testIndex, ]

  prop <- table(trainData$Seks)
  perc_ov <- prop[1]/prop[2]

  listdat[[1]] <- SmoteClassif(Seks ~ .,trainData,
                  C.perc = list("0"=1,
                  "1"=perc_ov),k=10, dist = "HVDM")
  a <- ENNClassif(Seks~.,listdat[[1]], k=10,
       dist="Overlap")
  listdat[[2]] <- a[[1]]
  listdat[[3]] <- adasynn(trainData, 10)

  for (t in 1:3) {
    jpeg(file = paste("plot baru ada train",t,' ke-
         ',o,'.jpeg', sep = ''))

    # grow tree
    fit <- rpart(Seks ~ .,method="class",
           data=listdat[[t]])
    #pruned tree
    bestcp <-fit$cptable[which.min(fit$cptable
             [,"xerror"]),"CP"]
    tree.pruned <- prune(fit, cp = bestcp)

    # plot tree
    plot(tree.pruned)
    text(tree.pruned, cex = 0.8, use.n = TRUE, xpd =
    TRUE)
    prp(tree.pruned, faclen = 0, cex = 0.8, extra = 1)
    tot_count <- function(x, labs, digits, varlen)
    {
      paste(labs, "\n\nn =", x$frame$n)
    }

    prp(tree.pruned, faclen = 0, cex = 0.8,
        node.fun=tot_count)
```

```
only_count <- function(x, labs, digits, varlen)
{
  paste(x$frame$n)
}

boxcols <- c("pink", "palegreen3")
            [tree.pruned$frame$yval]

par(xpd=TRUE)
prp(tree.pruned, faclen = 0, cex = 0.8,
     node.fun=only_count, box.col = boxcols)
legend("bottomleft", legend = c("tidak","ya"),
       fill = c("pink", "palegreen3"),
       title = "Group")

dev.off()

#predict test data
cartpredict <- tree.pruned %>% predict(testData,
                type="class")

# confusion matrix (training data) cart
conf.matrixt1 <- table(listdat[[t]]$Seks,
                  predict(tree.pruned,type="class"
                  ))
rownames(conf.matrixt1) <- paste("Actual",
                           rownames(conf.matrixt1
                           ), sep = ":")
colnames(conf.matrixt1) <- paste("Pred",
                           colnames(conf.matrixt1
                           ), sep = ":")
print(conf.matrixt1)
# confusion matrix (testing data) cart
conf.matrix <- table(testData$Seks, cartpredict)
rownames(conf.matrix) <- paste("Actual",
                         rownames(conf.matrix),
                         sep = ":")
 colnames(conf.matrix) <- paste("Pred",
                         colnames(conf.matrix),
                         sep = ":")
print(conf.matrix)
```

```
    # Compute model accuracy rate on train data cart
        akurast[o,t] <- mean(predict(tree.pruned,
                        type="class") ==
                        listdat[[t]]$Seks)
    #auc training
        auct[o,t] <- ((conf.matrixt1[2,2]/
                      (conf.matrixt1[2,1]+
                      conf.matrixt1[2,2]))+
                      (conf.matrixt1[1,1]/
                      (conf.matrixt1[1,2]+
                      conf.matrixt1[1,1])))/2

    # Compute model accuracy rate on test data cart
        akuras[o,t] <- mean(cartpredict ==
                            testData$Seks)
        #auc testing
        auc[o,t] <-((conf.matrix[2,2]/
                    (conf.matrix[2,1]+conf.matrix[2,2]
                    ))+(conf.matrix[1,1]/
                    (conf.matrix[1,2]+conf.matrix[1,1]
                    )))/2
      }
    }
```

Lampiran 8. Surat Keterangan

SURAT KETERANGAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

   Nama : Kinanthi Sukma Wening

   NRP : 06211640000044

   Telah mengambil data di instansi/perusahaan kami :

   Nama Instansi : Perwakilan BKKBN Provinsi Jawa Timur

   Divisi/ bagian : Bidang Pelatihan dan Pengembangan

   sejak tanggal _____ sampai dengan _____ untuk keperluan Tugas Akhir/ Thesis Semester Gasal/~~Genap~~* 2019/ 2020.

2. Tidak Keberatan/~~Keberatan~~* nama perusahaan dicantumkan dalam Tugas Akhir/Thesis mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.

3. Tidak Keberatan/~~Keberatan~~* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.

Surabaya, 06 Januari 2020
Pimpinan Perusahaan
Bidang Pelatihan dan Pengembangan
Perwakilan BKKBN Prov. Jawa Timur

Dr. Iswari Hariastuti, M. Kes
NIP 19661111 199203 2 008

*(coret yang tidak perlu)

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS

Penulis dilahirkan di Pati, 12 November 1999 dengan nama lengkap Kinanthi Sukma Wening, biasa dipanggil Kinanthi. Penulis menempuh pendidikan formal di SDN Kajar 01, SMPN 3 Pati, dan SMAN 1 Pati. Semasa SMP, penulis mengikuti program akselerasi, sehingga masa SMP ditempuh selama dua tahun. Kemudian penulis diterima sebagai mahasiswa Departemen Statistika ITS pada tahun 2016. Selama masa perkuliahan, penulis aktif di Divisi Statistics Computer Course (SCC) Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS) sebagai staff *Training Development* periode 2017-2018 dan staff ahli *Training Development* pada periode 2018-2019. Selain itu, penulis juga aktif di Badan Eksekutif Mahasiswa ITS sebagai staff pada periode 2018-2019 dan asisten dirjen Publikasi dan Riset Ekonomi periode 2019-2020 pada Kementerian Perekonomian. Selain kegiatan organisasi, penulis juga aktif dalam kegiatan keilmiahan berupa lomba statistika tingkat nasional maupun internasional. Penulis pernah menjadi juara III dalam *Data Analysis Competition* se-asia Tenggara pada tahun 2019. Kemudian di tahun yang sama, penulis juga pernah menjadi juara III dalam Olimpiade Nasional Statistika. Bagi pembaca yang ingin berdiskusi, memberikan saran, dan kritik mengenai Tugas Akhir ini dapat disampaikan melalui email kinanthi.sukmawening@gmail.com.