I PENDAHULUAN

1.1 Latar Belakang

Peralihan semester baru bagi mahasiswa ditandai dengan diakhirinya ujian akhir semester. Selama rentang libur, mahasiswa diwajibkan untuk mengisi evaluasi pembelajaran selama satu semester, membayar Uang Kuliah Tunggal (UKT), serta mengisi kartu rencana studi secara daring. Pembayaran UKT bagi mahasiswa sesuai dengan surat edaran yang dikeluarkan oleh Dirjen Dikti Nomor 97/E/KU/2013 yang menyatakan bahwa tiap perguruan tinggi negeri dihimbau untuk menerapkan UKT bagi mahasiswa baru mulai tahun 2013.

Menurut Permendikbud No. 55 tahun 2013, Uang Kuliah Tunggal (UKT) merupakan selisih antara Biaya Kuliah Tunggal (BKT) dengan biaya yang ditanggung oleh pemerintah. Sementara BKT didefinisikan sebagai keseluruhan biaya operasional mahasiswa pada program studi di perguruan tinggi negeri. Dengan kata lain UKT merupakan akumulasi biaya operasional per semester yang dibebankan kepada mahasiswa. Biaya ini sebagai bentuk kontribusi orangtua mahasiswa terhadap penyelenggaraan pendidikan. Besaran biaya UKT disesuaikan dengan kondisi sosial-ekonomi dan biaya kuliah tunggal tiap program studi.

Menurut data yang diperoleh dari Direktorat Keuangan dan Akutansi (DIT KEU) IPB, terjadi peningkatan jumlah mahasiswa yang mengalami keterlambatan dalam pembayaran UKT di setiap angkatan mulai tahun 2016 sampai dengan 2018 dengan banyak mahasiswa yang mengalami keterlambatan yaitu 54, 64, dan 117 mahasiswa. Terdapat beberapa faktor yang dapat menyebabkan keterlambatan dalam pembayaran sumbangan pembinaan pendidikan diantaranya pendapatan orang tua, pendidikan orang tua, jumlah tanggungan keluarga, dan usia orang tua (Muqorobin 2019).

Pendekatan klasifikasi menggunakan *Random Forest* dan *AdaBoost* dapat digunakan sebagai solusi permasalahan tersebut. Pendekatan ini dapat memprediksi mahasiswa yang berindikasi terlambat dalam pembayaran UKT. Sehingga dapat dilakukan tindakan lebih dini dengan memberikan fasilitas kepada mahasiswa yang diprediksi terlambat. *Random Forest* merupakan salah satu pendekatan klasifikasi yang berasal dari gabungan pohon keputusan tunggal. Penggunaan pohon keputusan gabungan merupakan solusi alternatif dari kelemahan pohon keputusan tunggal. Pohon keputusan tunggal dinilai memiliki sifat yang tidak stabil jika dilakukan pemodelan menggunakan gugus data lain meskipun berasal dari populasi yang sama (Sartono dan Syafitri 2010). Sedangkan *AdaBoost* merupakan salah satu pendekatan klasifikasi *boosting* yang umum digunakan. Prinsip dalam *boosting* yaitu mengkombinasikan pengklasifikasi-pengklasifikasi lemah menjadi pengklasifikasian kuat. *AdaBoost* memberikan bobot lebih kepada objek yang diprediksi tidak tepat (Hastie *et al*. 2008).

1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun model klasifikasi keterlambatan pembayaran UKT dengan membandingkan pendekatan *Random Forest* dan *AdaBoost* serta mengidentifikasi peubah penting yang berpengaruh terhadap keterlambatan pembayaran UKT.

II TINJAUAN PUSTAKA

2.1 *Classification and Regression Tree* (CART)

Klasifikasi merupakan proses membangun model dengan tujuan untuk memprediksi suatu kelas yang bersifat kategorik (Han dan Kamber 2001). Terdapat dua tahapan umum dalam klasifikasi yaitu tahap pelatihan dan tahap pengujian. Tahap pelatihan berisi mengenai proses membangun model menggunakan gugus data latih dengan kelas yang sudah diketahui sebelumnya. Sedangkan tahap pengujian digunakan untuk mengevaluasi kinerja model klasifikasi menggunakan gugus data uji.

Classification and Regression Tree (CART) merupakan salah satu model non parametrik yang menghasilkan sebuah pohon keputusan berupa *binary tree*. Pohon keputusan yang dihasilkan dapat berupa pohon klasifikasi jika kelas bersifat kategorik atau pohon regresi jika kelas bersifat numerik (Breiman *et al*. 1993).

Menurut Breiman *et al*. 1993, penyekat terbaik dipilih berdasarkan ukuran penurunan keheterogenan suatu simpul yang selanjutnya disebut sebagai impuritas tereduksi. Semakin besar nilai impuritas tereduksi, suatu penyekat dikatakan dapat memaksimalkan kehomogenan di dalam masing-masing simpul yang terbentuk. Fungsi impuritas yang digunakan dalam penelitian ini yaitu Indeks Gini. Nilai impuritas pada simpul ke-t didefinisikan sebagai berikut:

$$i(t) = 1 - \sum_{j=0}^{1} p^2(j|t)$$

$$p(j|t) = \frac{N_j(t)}{N(t)}$$

dengan $p(j|t)$ merupakan peluang amatan kelas ke-j pada simpul ke-t, $N_j(t)$ merupakan total amatan kelas ke-j pada simpul ke-t, dan $N(t)$ merupakan total amatan pada simpul ke-t.

Sedangkan impuritas tereduksi dari suatu penyekat s didefinisikan sebagai berikut:

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

dengan p_L merupakan peluang amatan pada simpul kiri, $i(t_L)$ merupakan nilai impuritas pada simpul kiri, p_R merupakan peluang amatan pada simpul kanan, $i(t_R)$ merupakan nilai impuritas pada simpul kanan. Kriteria pemberhentian penyekatan pada pembentukan pohon klasifikasi yaitu ketika minimum jumlah amatan dalam simpul sebanyak 5 (Breiman *et al.* 1993).

2.2 Random Forest

Random Forest merupakan pengembangan dari CART dengan menerapkan proses *bootstrap aggregating* (*bagging*) dan *random feature selection* (Breiman *et al.* 2001). Random Forest merupakan salah satu pemodelan klasifikasi yang berasal dari gabungan pohon klasifikasi tunggal. Penggunaan pohon klasifikasi tunggal cenderung menghasilkan tingkat akurasi cukup tinggi ketika melakukan prediksi pada suatu gugus data latih namun menghasilkan tingkat akurasi rendah pada gugus data lain meskipun diambil dari populasi yang sama (Sartono dan Syafitri 2010).

Prinsip pemodelan klasifikasi Random Forest yaitu mengkombinasikan banyak pohon klasifikasi dan menentukan prediksi berdasarkan suara terbanyak (*majority vote*). Setiap gugus data latih baru diambil dari gugus data latih dengan melakukan penarikan contoh acak dengan pemulihan. Kemudian sebuah pohon klasifikasi dibangun menggunakan gugus data latih baru dengan menerapkan pemilihan m peubah penjelas secara acak pada setiap simpul yang terbentuk. Proses ini berlangsung hingga ukuran minimum amatan dalam simpul tercapai. Menurut Breiman *et al.* 2001, prosedur dalam membangun pemodelan klasifikasi Random Forest pada gugus data latih dengan n amatan dan p peubah penjelas sebagai berikut:

1. Melakukan proses *bootstrap* yaitu penarikan contoh acak berukuran n dengan pemulihan pada gugus data latih.
2. Membangun pohon klasifikasi tunggal CART menggunakan gugus data latih baru yang terbentuk dari proses *bootstrap*. Pembangunan pohon klasifikasi dilakukan dengan menerapkan *random feature selection* yaitu pemilihan peubah penjelas secara acak dengan $m < p$. Kemudian dari m peubah penjelas dipilih peubah penjelas terbaik sebagai penyekat dan dilanjutkan dengan penyekatan menjadi dua simpul baru. Proses ini berlangsung hingga ukuran minimum amatan dalam simpul tercapai.
3. Mengulangi prosedur pada langkah 1 dan 2 sebanyak k kali sehingga diperoleh k buah pohon klasifikasi. Tiap pohon klasifikasi menghasilkan satu suara sehingga didapatkan k buah suara. Penentuan klasifikasi didasarkan pada suara terbanyak (*majority vote*).

Menurut Hastie *et al.* 2008, meskipun nilai m yang disarankan yaitu $\sqrt{p}$ namun nilai m optimal bergantung pada masalah yang ingin diselesaikan dan sebaiknya dilakukan optimasi parameter terhadap nilai m.

Ukuran yang digunakan dalam menentukan peubah penting pada Random Forest yaitu *mean decrease gini* (MDG). MDG merupakan rasio antara penjumlahan dari impuritas tereduksi yang disebabkan oleh peubah penjelas ke-s dengan banyak pohon yang terbentuk. Semakin besar nilai MDG menandakan suatu peubah penjelas berperan penting dalam pembangunan pohon klasifikasi. Perhitungan MDG menurut Zuccoloto dan Sandri 2006 sebagai berikut:

$$MDG_s = \frac{1}{k}\sum_t [\Delta(s,t)I(s,t)]$$

dengan k merupakan banyak pohon klasifikasi yang terbentuk, $\Delta(s,t)$ merupakan impuritas tereduksi yang disebabkan oleh peubah penjelas ke-s pada simpul ke-t, dan $I(s,t)$ merupakan fungsi indikator yang bernilai 1 apabila peubah penjelas ke-s digunakan dalam penyekatan dan selainnya bernilai 0.

2.3 AdaBoost

Adaptive Boosting atau AdaBoost merupakan salah satu metode *boosting* yang umum dan popular digunakan. AdaBoost pertama kali diperkenalkan oleh Freund dan Schapire 1996 yang disebut dengan "AdaBoost.M1". Prinsip dari metode *boosting* yaitu menghasilkan prediksi yang akurat dengan mengkombinasikan pengklasifikasi-pengklasifikasi lemah (Freund dan Schapire 1996). Pengklasifikasi lemah merupakan pengklasifikasi yang memiliki tingkat

kesalahan sedikit lebih baik dibandingkan dengan tebakan acak. AdaBoost memberikan bobot lebih besar pada amatan yang diklasifikasikan tidak tepat. Sehingga pada iterasi selanjutnya, amatan yang sulit diklasifikasi menerima pengaruh yang lebih besar (Hastie *et al.* 2008). Menurut Hastie *et al.* 2008, prosedur dalam membangun pemodelan klasifikasi AdaBoost sebagai berikut:

1. Menentukan bobot awal w_i setiap amatan pada gugus data latih (x_i, y_i) dengan $i = 1,2,3, \dots, N$
$$w_i = \frac{1}{N}$$
2. Untuk setiap iterasi h, dengan $h = 1,2,3, \dots, H$ lakukan hal berikut:
   a. Lakukan pendugaan klasifikasi $G_h(x)$ pada gugus data latih dengan menerapkan bobot w_i^h untuk setiap amatan
   b. Hitung kesalahan klasifikasi dengan persamaan berikut
$$err_h = \frac{\sum_{i=1}^{N} w_i^h I[G_h(x_i) \neq y_i]}{\sum_{i=1}^{N} w_i^h}$$
   dengan $I[G_h(x_i) \neq y_i]$ merupakan fungsi indikator yang apabila benar bernilai 1 dan selainnya bernilai 0
   c. Hitung koefisien α_h dengan persamaan berikut
$$\alpha_h = \log\left(\frac{1 - err_h}{err_h}\right)$$
   d. Perbarui bobot amatan yang diklasifikasikan tidak tepat dengan persamaan:
$$w_i^{h+1} = w_i \exp(\alpha_h I[G_h(x_i) \neq y_i]), i = 1,2,3, \dots, N$$
3. Dugaan akhir prediksi merupakan total terboboti dugaan prediksi tiap iterasi dengan persamaan:
$$G(x) = Sign\left[\sum_{h=1}^{H} \alpha_h G_h(x)\right]$$

Penentuan prediksi klasifikasi akan menghasilkan kelas 1 apabila $G(x) \geq \frac{1}{2}\sum_{h=1}^{H} \alpha_h$ dan kelas 0 untuk selainnya.

2.4 Ketidakseimbangan Data

Ketidakseimbangan data merupakan kondisi ketika kelas dari gugus data tidak merepresentasikan secara merata antara kelas mayoritas dan kelas minoritas (Chawla 2002). Terdapat tiga pendekatan penanganan ketidakseimbangan data yang umum digunakan yaitu *Random Oversampling* (ROS), *Random Undersampling* (RUS), dan *Synthetic Minority Oversampling Technique* (SMOTE) (Burnaev 2017).

ROS melakukan duplikasi amatan kelas minoritas sehingga didapatkan jumlah amatan kelas minoritas mendekati jumlah amatan kelas mayoritas, RUS melakukan pengambilan amatan secara acak dari kelas mayoritas sehingga didapatkan jumlah amatan kelas mayoritas mendekati jumlah amatan kelas minoritas, dan SMOTE melakukan duplikasi dengan melibatkan pembangkitan amatan sintesis dari gugus data pada kelas minoritas. SMOTE menggunakan prinsip k tetangga terdekat (*k-nearest neighbor*) dalam pembangkitan amatan sintesis

(Chawla 2002). Menurut Chawla 2002, prosedur pembangkitan amatan sintesis pada SMOTE untuk masing-masing amatan kelas minoritas adalah sebagai berikut:

1. Menentukan k amatan tetangga terdekat dari amatan yang menjadi perhatian
2. Penentuan k amatan terdekat dihitung berdasarkan jarak antara vektor suatu amatan dengan vektor amatan lainnya berdasarkan tipe peubah
   a. Peubah numerik menggunakan ukuran jarak Euclidean dengan persamaan berikut:

$$d(\boldsymbol{a},\boldsymbol{b}) = \sqrt{(\boldsymbol{a}-\boldsymbol{b})'(\boldsymbol{a}-\boldsymbol{b})} = \sqrt{\sum_{i=1}^{p_1}(a_i - b_i)^2}$$

   Keterangan:

$d(\boldsymbol{a},\boldsymbol{b})$	:	jarak Euclidean antara vektor amatan a dengan vektor amatan b
a_i	:	nilai amatan a pada peubah numerik ke-i
b_i	:	nilai amatan b pada peubah numerik ke-i
p_1	:	banyak peubah numerik

   b. Peubah kategorik menggunakan ukuran *value difference metric* (VDM) dengan persamaan berikut:

$$\Delta(\boldsymbol{A},\boldsymbol{B}) = \sum_{i=1}^{p_2}\delta(A_i, B_i)$$

   Keterangan:

$\Delta(\boldsymbol{A},\boldsymbol{B})$	:	jarak antara vektor amatan A dengan vektor amatan B
p_2	:	banyak peubah kategorik
$\delta(A_i, B_i)$	:	jarak antara kategori A dengan kategori B pada peubah kategorik ke-i

   Adapun jarak antara kedua kategori pada suatu peubah kategorik ke-i didefinisikan sebagai berikut:

$$\delta(A_i, B_i) = \sum_{j=1}^{k}\left|\frac{C_{Aj}}{C_A} - \frac{C_{Bj}}{C_B}\right|$$

   Keterangan:

C_{Aj}	:	Banyak kemunculan kategori A yang termasuk kedalam kelas ke-j
C_{Bj}	:	Banyak kemunculan kategori B yang termasuk kedalam kelas ke-j
C_A	:	Banyak kemunculan kategori A
C_B	:	Banyak kemunculan kategori B
k	:	Banyak kelas pada peubah respon

3. Membangkitkan amatan sintesis berdasarkan tipe peubah
   a. Peubah numerik yaitu dengan mengambil secara acak sebuah amatan dari vektor *k*-tetangga terdekat. Kemudian menghitung vektor amatan sintesis dengan persamaan berikut:

$$\boldsymbol{x_{baru}} = \boldsymbol{x} + (\hat{\boldsymbol{x}} - \boldsymbol{x}) * rand[0,1]$$

Keterangan:

$\boldsymbol{x_{baru}}$	:	vektor amatan sintesis dengan ukuran $p_1\ x\ 1$
$\boldsymbol{x}$	:	vektor amatan dengan ukuran $p_1\ x\ 1$
$\hat{\boldsymbol{x}}$	:	vektor *k*-tetangga terdekat dengan ukuran $p_1\ x\ 1$
$rand[0,1]$	:	bilangan acak seragam yang berkisar antara 0 dan 1

b. Peubah kategorik yaitu dengan menentukan kategori amatan yang sering muncul (modus) antara vektor amatan dan vektor *k*-tetangga terdekat.

2.5 *K-Fold Stratified Cross Validation*

K-Fold Cross Validation merupakan salah satu metode untuk memvalidasi akurasi dari suatu pengklasifikasi. Prinsip kerja dari *K-Fold Cross Validation* yaitu membagi suatu gugus data menjadi k himpunan bagian sama besar secara acak. Dalam *K-Fold Stratified Cross Validation*, tiap himpunan bagian memiliki proporsi kelas data yang seimbang. Himpunan bagian ke-k berperan sebagai gugus data uji dan sisanya k-1 himpunan bagian berperan sebagai gugus data latih. Pendugaan validasi akurasi merupakan rataan dari masing-masing akurasi di setiap himpunan bagian. Ukuran 10-*Fold* dalam *Stratified Cross Validation* merupakan ukuran metode validasi yang optimal (Kohavi 1995).

2.6 Ukuran Kinerja Model Klasifikasi

Ukuran kinerja model klasifikasi berfungsi untuk mengevaluasi seberapa baik pengklasifikasi dapat mengklasifikasikan amatan dengat tepat. Kurva *Receiver Operator Characteristic* (ROC) merupakan kurva dari kinerja pengklasifikasi berupa plot antara sensitivitas dan (1-spesifisitas) untuk semua kemungkinan *cutoff*. Kurva ROC lebih informatif dibandingkan dengan matriks konfusi karena dapat merangkum kinerja prediksi (Agresti 2002). Luas dibawah kurva ROC (*Area Under Curve* ROC) yang selanjutnya disebut sebagai ROC-AUC merupakan sebuah ukuran yang merepresentasikan kinerja suatu pengklasifikasi. Nilai ini berkisar antara 0 sampai dengan 1 dan digunakan sebagai ukuran untuk membandingkan kinerja beberapa pengklasifikasi (Fawcett 2006).

Berikut merupakan perhitungan nilai sensitivitas dan spesifisitas dengan kelas positif merupakan mahasiswa yang dikategorikan terlambat dalam pembayaran UKT dan kelas negatif merupakan mahasiswa yang dikategorikan tidak terlambat dalam pembayaran UKT sebagai berikut:

Tabel 1 Matriks konfusi

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

Keterangan:

TP	:	banyak amatan kelas positif yang diprediksi tepat ke kelas positif
FN	:	banyak amatan kelas positif yang diprediksi tidak tepat ke kelas negatif
FP	:	banyak amatan kelas negatif yang diprediksi tidak tepat ke kelas positif
TN	:	banyak amatan kelas negatif yang diprediksi tepat ke kelas negatif

Adapun perhitungan nilai sensitivitas dan spesifisitas suatu model didefinisikan sebagai berikut:

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

$$Spesifisitas = \frac{TN}{FP + TN}$$

III METODE

3.1 Data

Data yang digunakan dalam penelitian ini meliputi data diri mahasiswa aktif program sarjana IPB dengan jalur masuk reguler dimulai tahun masuk 2016 sampai dengan 2018. Data tersebut merupakan data sekunder yang diperoleh dari Direktorat Administrasi dan Penerimaan Mahasiswa Baru (DIT APPMB) IPB. Sedangkan data terkait dengan peubah respon yang menyatakan keterlambatan mahasiswa dalam melakukan pembayaran UKT diperoleh dari Direktorat Keuangan dan Akutansi (DIT KEU) IPB. Peubah respon yang digunakan dalam penelitian ini yaitu keterlambatan pembayaran UKT yang didefinisikan sebagai pembayaran UKT yang melewati rentang waktu pembayaran yang telah ditentukan. Terdapat 11 peubah yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2

Tabel 2 Daftar peubah yang digunakan

Nama Peubah	Tipe Peubah	Keterangan
Keterlambatan pembayaran UKT	Nominal	0 = Tidak pernah 1 = Pernah
Jenis kelamin	Nominal	1 = Laki-Laki 2 = Perempuan
Jalur masuk IPB	Nominal	1 = SNMPTN 2 = SBMPTN 3 = UTM 4 = BUD 5 = PIN 6 = Ketua Osis
Urutan anak dalam keluarga	Ordinal	1 = 1 2 = 2 3 = 3 4 = 4 5 = 5+
Jumlah anak dalam keluarga	Rasio	Numerik
Pendidikan ayah	Ordinal	1 = SD 2 = SMP 3 = SMA 4 = Diploma 5 = Sarjana
Pendidikan ibu	Ordinal	1 = SD 2 = SMP 3 = SMA 4 = Diploma 5 = Sarjana

Nama Peubah	Tipe Peubah	Keterangan
Penghasilan ayah (dalam ribu rupiah)	Ordinal	1 = < 500 2 = 500-1000 3 = 1000-2000 4 = 2000-3500 5 = 3500-5000 6 = 5000-7500 7 = 7500-10000 8 = > 10000
Penghasilan ibu (dalam ribu rupiah)	Ordinal	1 = < 500 2 = 500-1000 3 = 1000-2000 4 = 2000-3500 5 = 3500-5000 6 = 5000-7500 7 = 7500-10000 8 = > 10000
Asal daerah	Nominal	1 = Luar jawa 2 = Jawa
Besar daya listrik (VA)	Nominal	1 = 0 2 = 450 3 = 900 4 = 1300 5 = 2200 6 = > 2200

3.2 Prosedur Analisis Data

Tahapan prosedur analisis data yang dilakukan adalah sebagai berikut:

1. Melakukan integrasi data diri mahasiswa yang diperoleh dari pihak DIT APPMB IPB dan data keterlambatan pembayaran UKT yang diperoleh dari pihak DITKEU IPB
2. Melakukan praproses data dengan melakukan kodifikasi pada peubah skala kategorik dan melakukan penyeragaman format pada peubah dengan tipe kategorik
3. Melakukan eksplorasi data dengan statistika deskriptif untuk mengetahui karakteristik keterlambatan pembayaran UKT berdasarkan peubah-peubah penjelas
4. Melakukan pemodelan klasifikasi Random Forest dan AdaBoost dengan menentukan parameter optimal pada masing-masing pemodelan. Penentuan parameter optimal menggunakan *10-fold stratified cross validation* dan pendekatan *resampling* yaitu *Random Undersampling* (RUS), *Random Oversampling* (ROS), dan *Synthetic Minority Oversampling Technique* (SMOTE).
    a. Random Forest, penentuan parameter optimal pada Random Forest dengan menentukan banyak peubah acak dalam pembentukan pohon (m) dan banyak pohon yang dihasilkan (k) dari berbagai kemungkinan nilai. Nilai m yang digunakan berkisar antara 3

sampai dengan 7 peubah dan nilai k yang digunakan yaitu 25, 50, 100, 200, 500, dan 600.

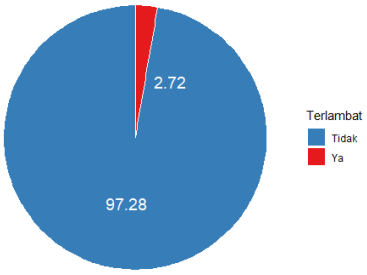
b. AdaBoost, penentuan parameter optimal pada AdaBoost dengan menentukan banyak iterasi (h). Nilai h yang digunakan yaitu 20, 40, 60, 80, dan 100.

5. Mengevaluasi kinerja pemodelan klasifikasi Random Forest dan AdaBoost menggunakan ukuran kinerja prediksi ROC-AUC untuk menentukan model dengan parameter optimal.
6. Menentukan model terbaik diantara Random Forest dan AdaBoost serta mengidentifikasi peubah penting yang memengaruhi keterlambatan pembayaran UKT.

IV HASIL DAN PEMBAHASAN

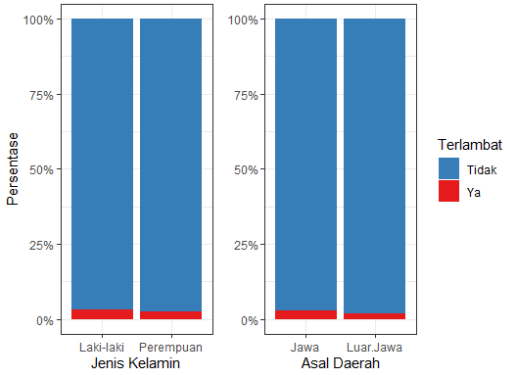
4.1 Gambaran Umum Karakteristik Mahasiswa

Setelah dilakukan integrasi dan praproses data didapatkan data berjumlah 8652 mahasiswa dengan 10 peubah penjelas dan 1 peubah respon. Sebanyak 235 mahasiswa mengalami keterlambatan pembayaran UKT dan sisanya 8417 mahasiswa tidak mengalami keterlambatan pembayaran UKT. Gambar 1 menunjukkan gambaran umum persentase mahasiswa yang mengalami keterlambatan dalam pembayaran UKT. Persentase mahasiswa yang mengalami keterlambatan pembayaran jauh lebih kecil dibandingkan dengan mahasiswa yang tidak mengalami keterlambatan. Persentase mahasiswa yang mengalami keterlambatan sebesar 2,72% dan mahasiswa yang tidak mengalami keterlambatan sebesar 97,28%. Hal ini mengindikasikan adanya ketidakseimbangan data antara kelas terlambat dan tidak terlambat.



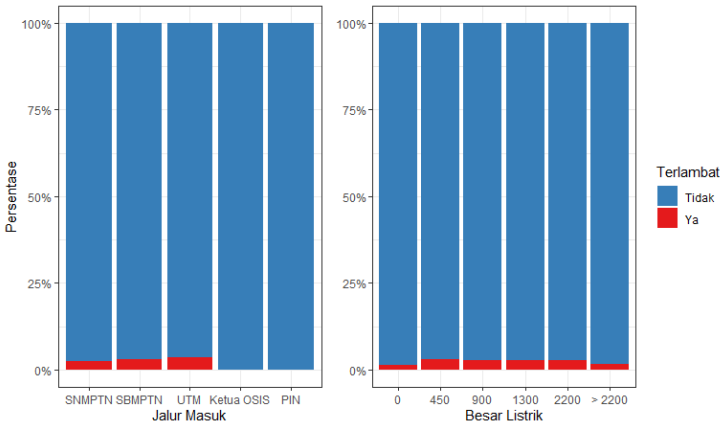
Gambar 1 Proporsi mahasiswa yang mengalami keterlambatan pembayaran UKT

Berdasarkan Gambar 2, persentase mahasiswa jenis kelamin laki-laki yang mengalami keterlambatan pembayaran UKT lebih besar dibandingkan dengan jenis kelamin perempuan. Adapun persentase mahasiswa jenis kelamin laki-laki sebesar 3,14% dan mahasiswa jenis kelamin perempuan sebesar 2,46%. Sedangkan pada peubah asal daerah, mahasiswa yang berasal dari pulau jawa yang mengalami keterlambatan pembayaran UKT memiliki persentase lebih besar dibandingkan dengan mahasiswa luar pulau jawa dengan persentase sebesar 2,94% dan 2%.



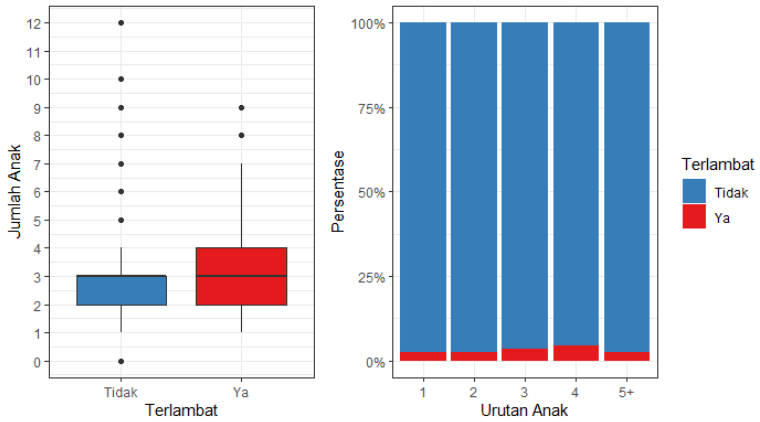
Gambar 2 Persentase jenis kelamin dan asal daerah mahasiswa terhadap keterlambatan pembayaran UKT

Berdasarkan Gambar 3, mahasiswa jalur penerimaan UTM yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan dengan mahasiswa dengan jalur penerimaan lain dengan persentase sebesar 3,65% diikuti oleh jalur SBMPTN sebesar 2,96%, dan jalur SNMPTN sebesar 2,51%. Adapun jalur penerimaan Ketua Osis dan PIN tidak memiliki mahasiswa yang terlambat melakukan pembayaran UKT. Pada peubah besar daya listrik, mahasiswa dengan besar daya listrik sebesar 450 VA yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar yaitu 3,08%.



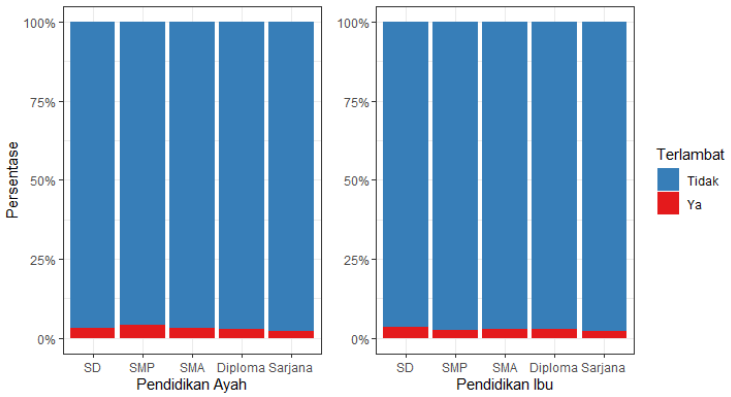
Gambar 3 Persentase jalur masuk penerimaan dan besar daya listrik terhadap keterlambatan pembayaran UKT

Berdasarkan Gambar 4, sebagian besar mahasiswa memiliki jumlah anak dalam keluarga sebanyak 3 orang anak. Namun, mahasiswa yang mengalami keterlambatan pembayaran UKT memiliki rentang jumlah anak yang lebih lebar dibandingkan dengan yang tidak mengalami keterlambatan pembayaran. Sementara itu, mahasiswa yang merupakan anak keempat dan mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan dengan kategori lainnya sebesar 4,57% kemudian disusul oleh anak ketiga sebesar 3,4% dan anak pertama sebesar 2,55%.



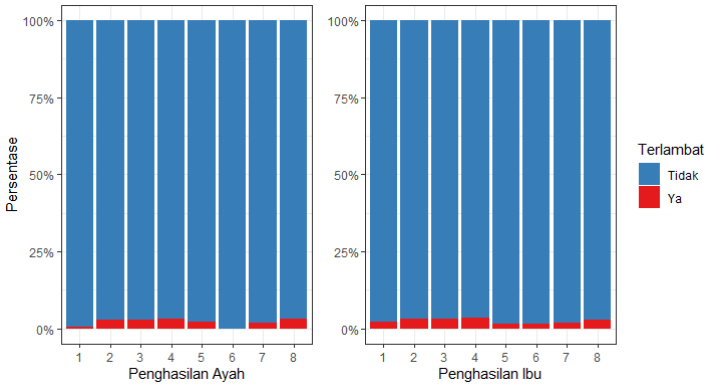
Gambar 4 Sebaran jumlah anak dan persentase urutan anak dalam keluarga terhadap keterlambatan pembayaran UKT

Berdasarkan Gambar 5, mahasiswa dengan jenjang pendidikan ayah SMP yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan dengan jenjang pendidikan lainnya sebesar 4,03%. Nilai ini diikuti oleh jenjang pendidikan SD sebesar 3,09% dan SMA sebesar 3,01%. Sementara itu, mahasiswa dengan pendidikan ibu SD yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan dengan jenjang pendidikan lainnya sebesar 3,61% diikuti oleh jenjang SMA sebesar 2,91% dan diploma sebesar 2,72%.



Gambar 5 Persentase pendidikan ayah dan pendidikan ibu terhadap keterlambatan pembayaran UKT

Berdasarkan Gambar 6, mahasiswa dengan penghasilan ayah kategori 8 (penghasilan lebih dari 10 juta) yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan kategori lainnya sebesar 3,23%. Sementara mahasiswa dengan penghasilan ibu kategori 4 (penghasilan 2 sampai 3,5 juta) yang mengalami keterlambatan pembayaran UKT memiliki persentase terbesar dibandingkan kategori lainnya sebesar 3,37%.



Gambar 6 Persentase penghasilan ayah dan penghasilan ibu terhadap keterlambatan pembayaran UKT

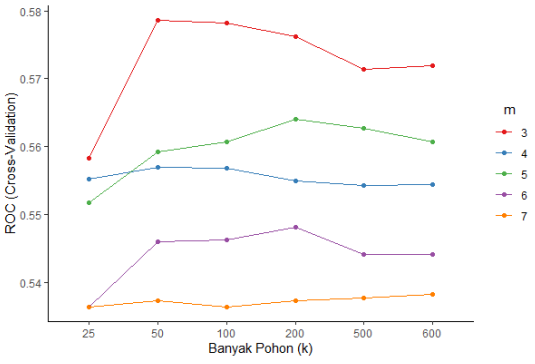
4.2 Pemodelan Klasifikasi Random Forest

Pemodelan klasifikasi menggunakan Random Forest dan AdaBoost diawali dengan menentukan parameter terbaik dari beberapa kemungkinan nilai parameter. Pemodelan dilakukan menggunakan 10-*Fold Cross Validation*. Hasil pemodelan kemudian dievaluasi menggunakan ukuran kinerja prediksi model yaitu *Receiver Operator Characteristic Area Under Curve* (ROC-AUC). Tiap pemodelan klasifikasi dilakukan penerapan *resampling* menggunakan *Random Undersampling* (RUS), *Random Oversampling* (ROS), dan *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan pada data.

Pemodelan Random Forest diawali dengan menentukan parameter optimal banyak peubah yang digunakan dalam pembentukan pohon klasifikasi (m) dan banyak pohon yang dihasilkan (k). Adapun nilai m yang digunakan berkisar antara 3 sampai dengan 7 sedangkan nilai k yang digunakan yaitu 25, 50, 100, 200, 500, dan 600.

4.2.1 Pemodelan Klasifikasi Random Forest dengan RUS

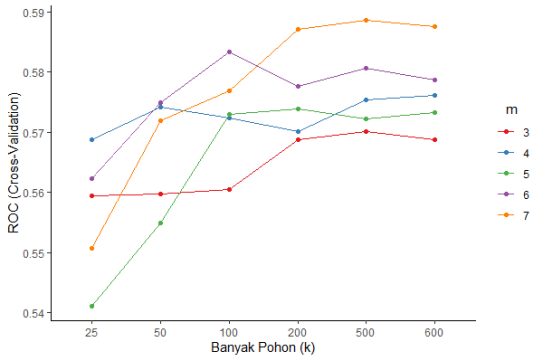
Hasil ROC-AUC dari beberapa banyak pohon (k) yang dicobakan berdasarkan tiap-tiap banyak peubah (m) menggunakan RUS disajikan pada Gambar 7. Dapat dilihat bahwa nilai ROC-AUC untuk setiap banyak peubah (m) memiliki hasil ROC-AUC yang bervariasi. Semakin besar nilai m yang digunakan, rentang nilai ROC-AUC yang didapatkan cenderung menurun. Hal ini terlihat pada m sebanyak 3 peubah penjelas merupakan banyak peubah yang memiliki rentang nilai ROC-AUC paling tinggi diantara m lainnya. Sedangkan pada m sebanyak 6 dan 7 peubah penjelas merupakan banyak peubah penjelas yang memiliki rentang nilai ROC-AUC terendah. Untuk setiap banyak peubah (m), semakin besar nilai banyak pohon (k) yang digunakan semakin besar pula nilai ROC-AUC yang dihasilkan. Hal ini terlihat bahwa terjadi peningkatan nilai ROC-AUC pada banyak pohon sebanyak 50 sampai dengan 600 pohon. Semakin besar banyak pohon yang digunakan nilai ROC-AUC cenderung semakin besar dan konstan. Nilai m sebanyak 3 peubah penjelas dan k sebanyak 50 pohon merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar 57,86%. Sehingga nilai m sebanyak 3 peubah penjelas dan k sebanyak 50 pohon merupakan parameter optimal menggunakan RUS.



Gambar 7 Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan *Random Undersampling*

4.2.2 Pemodelan Klasifikasi Random Forest dengan ROS

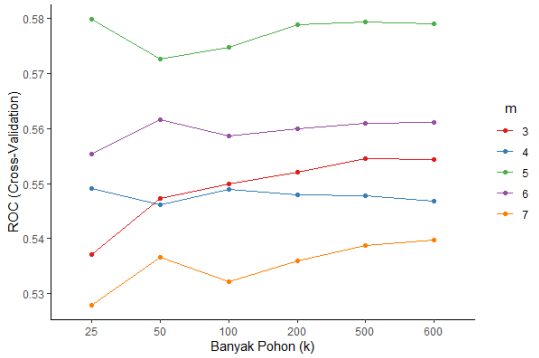
Hasil ROC-AUC dari beberapa banyak pohon (k) yang dicobakan berdasarkan tiap-tiap banyak peubah (m) menggunakan ROS disajikan pada Gambar 8. Dapat dilihat bahwa untuk setiap banyak peubah penjelas (m), semakin besar banyak pohon (k) yang digunakan semakin besar pula nilai ROC-AUC yang dihasilkan. Terjadi peningkatan nilai ROC-AUC sampai dengan banyak pohon (k) sebanyak 200 pohon kemudian nilai ROC-AUC yang dihasilkan cenderung konstan. Terlihat bahwa nilai m sebanyak 7 peubah memiliki rentang nilai ROC-AUC terbesar diantara yang lain. Sementara nilai m sebanyak 3 peubah penjelas memiliki rentang ROC-AUC terendah. Nilai ROC-AUC pada m sebanyak 7 peubah penjelas lebih besar dibandingkan dengan nilai m yang disarankan (*default*) yaitu sebanyak 3 peubah penjelas. Nilai m sebanyak 7 peubah penjelas dan k sebanyak 500 pohon merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar 58,70%. Sehingga nilai m sebanyak 7 peubah penjelas dan k sebanyak 500 pohon merupakan parameter optimal menggunakan RUS.



Gambar 8 Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan *Random Oversampling*

4.2.3 Pemodelan Klasifikasi Random Forest dengan SMOTE

Hasil ROC-AUC dari beberapa banyak pohon (k) yang dicobakan berdasarkan tiap-tiap banyak peubah (m) menggunakan SMOTE disajikan pada Gambar 9. Dapat dilihat bahwa untuk setiap banyak peubah penjelas (m) nilai ROC-AUC yang diihasilkan cenderung bervariasi. Nilai m sebanyak 4 peubah penjelas merupakan banyak peubah dengan rentang nilai ROC-AUC tertinggi dibandingkan yang lain. Sementara itu, nilai m sebanyak 7 peubah penjelas merupakan banyak peubah dengan rentang nilai ROC-AUC terendah. Nilai m sebanyak 4 peubah penjelas memiliki nilai ROC-AUC lebih besar dibandingkan dengan nilai m yang disarankan (*default*) yaitu sebanyak 3 peubah penjelas. Untuk setiap banyak peubah penjelas (m), terjadi peningkatan nilai ROC-AUC sampai dengan 100 pohon setelah itu nilai ROC-AUC cenderung konstan. Nilai m sebanyak 5 peubah penjelas dan k sebanyak 25 pohon merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar 57,99%. Sehingga nilai m sebanyak 5 peubah penjelas dan k sebanyak 25 pohon merupakan parameter optimal menggunakan SMOTE.



Gambar 9 Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan *Synthetic Minority Oversampling Technique*

4.2.4 Penentuan Nilai Parameter Optimal pada Random Forest

Berdasarkan pendekatan *resampling* yang digunakan yaitu *Random Undersampling*, *Random Oversampling*, dan *Synthetic Minority Oversampling Technique* dilakukan penentuan parameter optimal. Nilai parameter optimal Random Forest pada masing-masing pendekatan *resampling* disajikan pada Tabel 3. Pemodelan Random Forest memiliki nilai parameter optimal yaitu dengan banyak peubah (m) sebanyak 7 peubah penjelas dan banyak pohon (k) sebanyak 500 pohon dengan pendekatan *resampling Random Oversampling* (ROS) dengan ROC-AUC terbesar yaitu sebesar 58,70%.

Tabel 3 Nilai parameter optimal Random Forest pada masing-masing pendekatan *resampling*

Jenis *Resampling*	Banyak Peubah (m)	Banyak Pohon (k)	ROC-AUC (%)
RUS	3	50	57,86
ROS	7	500	58,70
SMOTE	5	25	57,99

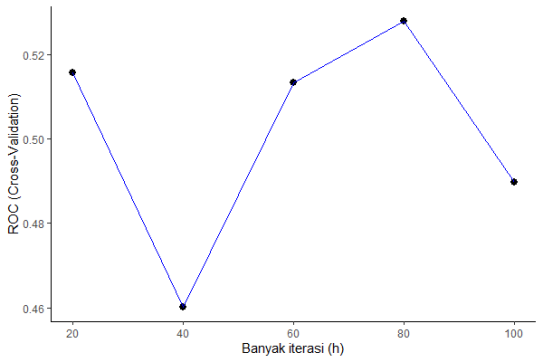
4.3 Pemodelan Klasifikasi dengan AdaBoost

Pemodelan klasifikasi menggunakan AdaBoost diawali dengan menentukan banyak iterasi (h) yang optimal. Adapun nilai h yang digunakan yaitu 20, 40, 60, 80, dan 100. Pemodelan dilakukan dengan menerapkan *resampling* yaitu *Random Undersampling* (RUS), *Random Oversampling* (ROS), dan *Synthetic Minority Oversampling Technique* (SMOTE).

4.3.1 Pemodelan AdaBoost dengan RUS

Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan RUS disajikan pada Gambar 10. Dapat dilihat bahwa terjadi penurunan nilai ROC-AUC yang cukup signifikan pada h sebanyak 40 iterasi, kemudian terjadi peningkatan nilai ROC-AUC sampai dengan 80 iterasi dan mengalami penurunan kembali pada

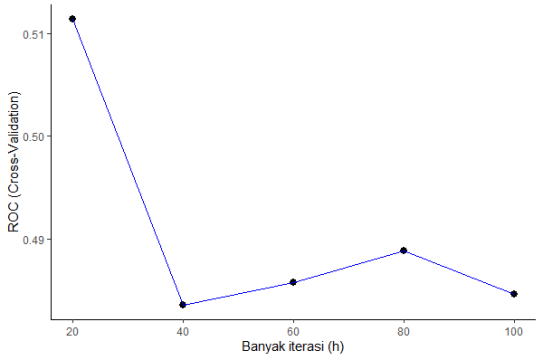
100 iterasi. Nilai h sebanyak 80 iterasi merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar 52,90%. Sehingga nilai h sebanyak 80 iterasi merupakan parameter banyak iterasi optimal menggunakan RUS.



Gambar 10 Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan *Random Undersampling*

4.3.2 Pemodelan AdaBoost dengan ROS

Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan ROS disajikan pada Gambar 11. Dapat dilihat bahwa terjadi penurunan nilai ROC-AUC yang cukup signifikan pada banyak iterasi (h) sebanyak 40 iterasi, kemudian terjadi sedikit peninkatan sampai dengan 100 iterasi. Nilai h sebanyak 20 iterasi merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar 51,14%. Sehingga nilai h sebanyak 20 iterasi merupakan parameter banyak iterasi optimal menggunakan ROS.

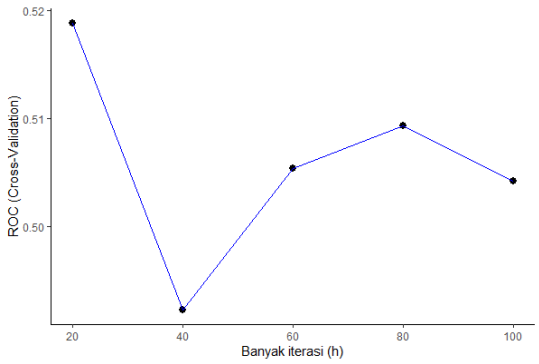


Gambar 11 Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan *Random Oversampling*

4.3.3 Pemodelan AdaBoost dengan SMOTE

Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan SMOTE disajikan pada Gambar 12. Dapat dilihat bahwa terjadi penurunan yang cukup signifikan pada 40 iterasi, kemudian terjadi peningkatan sampai dengan 80 iterasi, dan diikuti dengan penurunan kembali pada 100 iterasi. Nilai h sebanyak 20 iterasi merupakan parameter dengan nilai ROC-AUC terbesar yaitu sebesar

51,88%. Sehingga nilai h sebanyak 20 iterasi merupakan parameter banyak iterasi optimal menggunakan SMOTE.



Gambar 12 Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan *Synthetic Minority Oversampling Technique*

4.3.4 Penentuan Nilai Parameter Optimal pada AdaBoost

Berdasarkan pendekatan *resampling* yang digunakan yaitu *Random Undersampling*, *Random Oversampling*, dan *Synthetic Minority Oversampling Technique* dilakukan penentuan parameter optimal. Terlihat ketiga pendekatan *resampling* yang digunakan memiliki pola yang serupa yaitu terjadi penurunan nilai ROC-AUC pada 40 iterasi kemudian tejadi peningkatan sampai dengan 100 iterasi. Nilai parameter optimal AdaBoost pada masing-masing pendekatan *resampling* disajikan pada Tabel 4. Pemodelan AdaBoost memiliki nilai parameter optimal yaitu dengan banyak iterasi (h) sebanyak 80 iterasi dengan pendekatan *resampling Random Undersampling* (RUS) dengan ROC-AUC terbesar yaitu sebesar 52,90%.

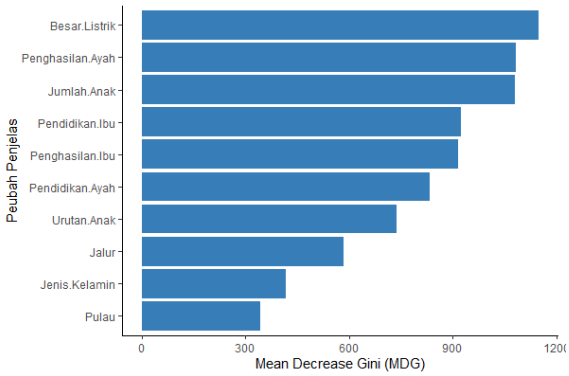
Tabel 4 Nilai parameter optimal AdaBoost pada masing-masing pendekatan *resampling*

Jenis *Resampling*	Banyak iterasi (h)	ROC-AUC (%)
RUS	80	52,90
ROS	20	51,14
SMOTE	20	51,88

4.4 Perbandingan Random Forest dan AdaBoost

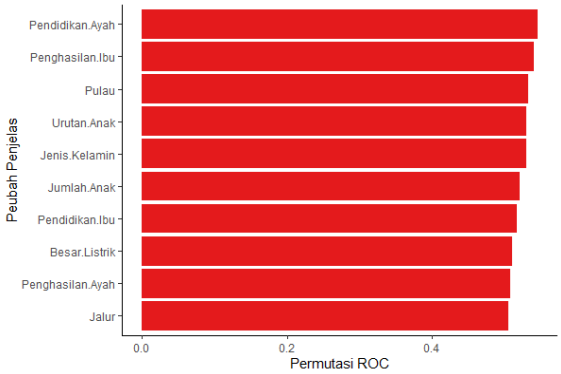
Pemodelan Random Forest menghasilkan banyak peubah penjelas (m) optimal yaitu sebanyak 7 peubah, banyak pohon (k) optimal sebanyak 500 pohon, serta menggunakan penanganan ketidakseimbangan data yaitu *Random Oversampling* (ROS). Sedangkan pemodelan AdaBoost menghasilkan banyak iterasi (h) yaitu 80 iterasi dengan penanganan ketidakseimbangan data yaitu *Random Undersampling* (RUS). Perbandingan antara model terbaik Random Forest dan AdaBoost diukur berdasarkan ukuran kinerja model klasifikasi ROC-AUC. Semakin besar nilainya menandakkan bahwa pemodelan memiliki kinerja yang lebih baik.

Pemodelan Random Forest menghasilkan tingkat kepentingan peubah yang diukur menggunakan ukuran *Mean Decrease Gini* (MDG). Nilai ini menunjukkan seberapa besar kontribusi suatu peubah penjelas dalam membentuk kumpulan pohon klasifikasi. Semakin besar nilainya, peubah tersebut semakin penting. Tingkat kepentingan peubah yang dihasilkan oleh pemodelan klasifikasi Random Forest ditunjukkan pada Gambar 13. Peubah besar listrik merupakan peubah yang memiliki tingkat kepentingan peubah tertinggi diantara peubah lainnya. Kemudian diikuti oleh peubah penghasilan ayah, dan jumlah anak dalam keluarga. Sementara itu, peubah jenis kelamin dan asal daerah merupakan peubah dengan tingkat kepentingan peubah terendah. Peubah-peubah ini memberikan kontribusi yang rendah dalam pemodelan klasifikasi Random Forest.



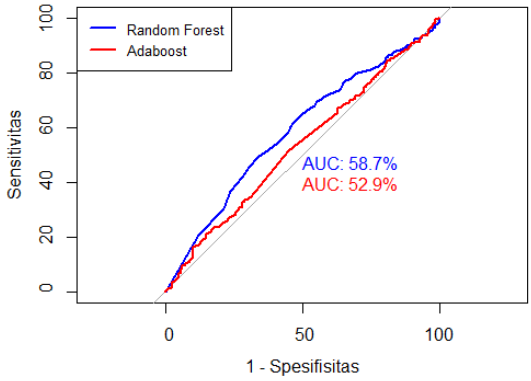
Gambar 13 Peubah penting Random Forest

Sedangkan untuk pemodelan AdaBoost menghasilkan tingkat kepentingan peubah menggunakan ukuran *permutation feature importance*. Tiap peubah dilakukan pengacakkan amatan secara bergantian dan dilakukan evaluasi model menggunakan pemodelan yang sudah didapatkan sebelumnya. Hasil akhir tingkat kepentingan tiap peubah dihitung berdasarkan selisih antara ukuran kinerja model klasifikasi ROC-AUC awal dengan rataan ukuran kinerja model klasifikasi ROC-AUC tiap ulangan. Semakin besar nilai ukuran *permutation feature importance* menandakan bahwa peubah tersebut memiliki tingkat kepentingan peubah yang tinggi. Hasil ukuran *permutation feature importance* tiap peubah ditunjukkan pada Gambar 14. Dapat dilihat bahwa tingkat kepentingan peubah yang dihasilkan tidak terlalu jauh berbeda antara satu peubah dengan peubah lainnya. Peubah pendidikan ayah, penghasilan ibu, dan pulau memiliki tingkat kepentingan peubah yang relatif sedikit lebih tinggi dibandingkan peubah lainnya.



Gambar 14 Peubah penting AdaBoost

Kurva perbandingan ROC Random Forest dan AdaBoost disajikan pada Gambar 15. Dapat dilihat bahwa pemodelan klasifikasi menggunakan Random Forest memiliki luas area dibawah kurva yang lebih besar dibandingkan dengan pemodelan klasifikasi menggunakan AdaBoost. Pemodelan Random Forest memiliki nilai AUC sebesar 58,70% sedangkan pemodelan AdaBoost memiliki nilai AUC sebesar 52,90%. Selain menggunakan ROC-AUC pemodelan juga dievaluasi menggunakan sensitivitas dan spesifisitas dengan *cut-off* optimal. Penentuan titik *cut-off* optimal didapatkan berdasarkan titik *cut-off* yang memaksimalkan nilai sensitivitas dan spesifisitas. Pemodelan terbaik Random Forest menghasilkan nilai sensitivitas sebesar 60,16% dan spesifitas sebesar 54,46% dengan titik *cut-off* optimal yaitu pada 0,97 sedangkan pemodelan terbaik AdaBoost menghasilkan nilai sensitivitas sebesar 51,45% dan spesifisitas sebesar 54,89% dengan titik *cut-off* optimal yaitu pada 0,5. Kemudian apabila dilihat berdasarkan ukuran tingkat kepentingan peubah pada Random Forest dan AdaBoost, tingkat kepentingan peubah pada Random Forest lebih terlihat perbedaan kepentingan peubah dibandingkan pada AdaBoost. Hal ini menunjukkan bahwa pemodelan Random Forest memiliki kinerja prediksi yang lebih baik dibandingkan dengan pemodelan AdaBoost. Setelah ditelusuri kembali, mahasiswa yang cenderung mengalami keterlambatan dalam pembayaran adalah mahasiswa dengan penghasilan ayah kategori menengah, namun memiliki daya listrik yang besar dan mahasiswa dengan penghasilan ayah kategori menengah, tetapi memiliki jumlah anak dalam keluarga yang cukup besar.



Gambar 15 Perbandingan Kurva ROC Random Forest dan AdaBoost

V SIMPULAN DAN SARAN

5.1 Simpulan

Pemodelan klasifikasi akhir Random Forest menggunakan parameter optimal banyak peubah penjelas (m) sebanyak 7 peubah dan banyak pohon (k) sebanyak 500 pohon dengan penanganan ketidakseimbangan data yaitu *Random Oversampling* (ROS). Sedangkan pemodelan klasifikasi akhir AdaBoost menggunakan parameter optimal banyak iterasi (h) sebanyak 80 iterasi dengan penanganan ketidakseimbangan data yaitu *Random Undersampling* (RUS). Pemodelan klasifikasi Random Forest dan AdaBoost memiliki ukuran kinerja prediksi ROC-AUC sebesar 58.70% dan 52.90%. Pemodelan klasifikasi Random Forest memiliki kinerja prediksi lebih baik dengan selisih AUC sebesar 5.8% dibandingkan dengan pemodelan klasifikasi AdaBoost dalam memprediksi mahasiswa yang berindikasi akan terlambat dalam pembayaran UKT. Adapun peubah penting pada pemodelan klasifikasi Random Forest yaitu besar daya listrik, penghasilan ayah, dan jumlah anak dalam keluarga.

5.2 Saran

Menambahkan peubah penjelas yang dapat menjelaskan keterlambatan pembayaran UKT lebih baik lagi sehingga dapat meningkatkan hasil ukuran kinerja prediksi model klasifikasi ROC-AUC seperti pengeluaran mahasiswa perbulan dan jumlah tanggungan dalam keluarga.

DAFTAR PUSTAKA

Agresti A. 2002. *Categorical Data Analysis*. Ed ke-2. New York (NY): John Wiley and Sons.

Breiman L, Friedman JH, Olshen RA, Stone CJ. 1993. *Classification and Regression Trees*. New York (NY): Chapman & Hall.

Breiman L. 2001. Random forests. *Machine Learning*. 45(1):11-13.

Burnaev E, Erofeev P, Papanov A. 2015. Influence of resampling on accuracy of imbalanced classification. *Eight International Conference on Machine Vision (ICMV)*. 9875.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: Synthetic Minority Over-sampling Techinque. *Journal of Artificial Intelligence Research*. 16(1):321-357.

[Dirjen Dikti] Direktorat Jendral Pendidikan Tinggi. 2013. Surat Edaran Nomor 97/E/KU/2013 tentang Uang Kuliah Tunggal. Jakarta: Dirjen Dikti.

Fawcet T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters*. 27:861-874.

Freund Y, Schapire RE. 1997. A Decision-Theoritic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. 55:119-139.

Han J, Kamber M. 2001. *Data Mining Concepts & Techniques*. San Fransisco (CA): Morgan Kaufmann.

Hastie T, Tibshirani R, Friedman J. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Ed ke-2. New York (NY): Springer-Verlag.

[Kemendikbud] Kementrian Pendidikan dan Kebudayaan. 2013. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 55 Tahun 2013 tentang Uang Kuliah Tunggal dan Biaya Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementrian Pendidikan dan Kebudayaan. Jakarta: Kemendikbud.

Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of 14th International Joint Conference on Artificial Intelligence (IJCAI)*. 2(1):1137-1143.

Muqorobin, Kusrini, Luthfi ET. 2019. Optimasi metode naïve bayes dengan feature selection information gain untuk prediksi keterlambatan pembayaran sumbangan pembinaan pendidikan sekolah. *Jurnal Ilmiah Sinus (JIS)*. 17(1):1693:1173.

Sandri M, Zuccolotto. 2006. Variable Selection Using Random Forest. Di dalam: Zani S, Cerioli A, Riani M, Vichi M. editor. *Data Analysis, Classification and the Forward Search. Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society;* 2005 jun 6-8; University of Parma, Italia. New York (NY): Springer-Verleg. hlm 263-270.

Sartono B, Syafitri UD. 2010. Metode pohon gabungan: solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal. *Forum Statistika dan Komputasi*. 15(1):1-7.

RIWAYAT HIDUP

Penulis merupakan anak pertama dari pasangan Bapak Firman dan Ibu Sukmawati yang dilahirkan di Kota Bukittinggi, Sumatra Barat pada tanggal 18 April 1998. Penulis menyelesaikan pendidikan di SD Negeri 3 Cilegon pada tahun 2010, SMP Negeri 5 Cilegon pada tahun 2013, dan SMA Negeri 1 Cilegon pada tahun 2016. Penulis diterima di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor melalui jalur SNMPTN pada tahun 2016.

Selama kuliah, penulis aktif mengikuti berbagai organisasi diantaranya yaitu Badan Eksekutif Mahasiswa (BEM) PPKU sebagai staff Departemen Pertanian dan Lingkungan Hidup (PLH) pada tahun 2016-2017, Himpunan Mahasiswa Profesi Gamma Sigma Beta (GSB) sebagai anggota Departemen Analisis Data pada tahun 2017-2018. Selain aktif mengikuti organisasi, penulis juga aktif berpartisipasi dalam berbagai kegiatan kepanitiaan diantaranya yaitu Anggota Divisi Tim Khusus Statistika Ria 12 (2017), Kepala Divisi Tim Khusus Statistika Ria 13 (2018), dan Kepala Sub Divisi Perlombaan STEM Pesta Sains Nasional (2019). Penulis juga mengikuti berbagai perlombaan diantaranya yaitu sebagai Finalis Kompetisi Data Mining JOINTS UGM 2018 dan Peringkat ke-4 Kompetisi Data Mining Arkavidia ITB 2019. Penulis melaksanakan kegiatan praktik lapang di MARS Digital sebagai Data Analyst pada tahun 2019.