JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Penerapan Metode *Bootstrap Aggregating K-Nearest Neighbor* dengan Analisis *Credit Scoring* Berdasarkan Status Pembayaran Kredit Motor

Ratna Ardita Sari[a,*], Elmanani Simamora [b]

[a,b] Universitas Negeri Medan, Deli Serdang, Indonesia
*email: ratna.arditas@gmail.com

Abstrak. Klasifikasi merupakan menggolongkan atau mengelompokan berdasarkan data yang ada serta dapat memprediksi data baru masuk ke dalam kelompok yang mana. Penelitian ini menerapkan gabungan algoritma klasifikasi dengan *Bootstrap Aggregating K-Nearest Neighbor* pada analisis *credit scoring*. Penelitian ini bertujuan agar dapat mengetahui klasifikasi hasil prediksi status pembayaran kredit motor dan agar dapat mengetahui tingkat akurasi terbaiknya. Data diambil dari data debitur PT. Mega Central Finance (MCF) pada Januari 2024 – Juni 2024 dengan klasifikasi status pembayaran kredit lancar dan tidak lancar terhadap 11 variabel bebas, yaitu usia, jumlah tanggungan, lama tinggal, masa kerja, pendapatan, pengeluaran, *On The Road* (OTR) motor, uang muka, besar pembayaran pinjaman, angsuran per bulan, dan jangka pembayaran. Hasil penelitian ini diperoleh 12 data *testing* dengan 11 debitur diklasifikasikan dengan benar dan 1 debitur diklasifikasikan dengan tidak benar, sehingga akurasi yang diperoleh dengan proporsi 90:10, K=7, *m*=80%, dan *C*=50 adalah sebesar 91,67%. Dengan implementasi metode *Bootstrap Aggregating* pada model klasifikasi *K-Nearest Neighbor* dengan hasil *Bagging K-NN* tidak meningkatkan ketepatan akurasi klasifikasi *K-NN* yang tetap sehingga tidak stabil.
Kata Kunci: Klasifikasi, *Bagging KNN,* Kredit*, Phyton*

PENDAHULUAN

Perkembangan dan kemajuan dunia pada saat ini di berbagai bidang mendorong semua masyarakat agar melakukan aktivitas dengan cepat. Masyarakat membutuhkan berbagai fasilitas sehingga dapat menunjang aktivitasnya, termasuk kendaraan motor yang saat ini dibutuhkan sebagian masyarakat (Saputro, 2011). Masyarakat mempunyai kesempatan agar mempunyai kendaraan motor dengan pembayaran baik secara tunai ataupun kredit dimana banyak pembayaran kredit yang ditawarkan khusus dari perusahaan jasa kredit dengan uang muka yang sesuai, salah satunya yaitu PT. Mega Central Finance (MCF)

Bentuk upaya pengolahan data sistem produk, konsumen, dan pelayanan dapat digunakan sebagai informasi yang lebih bermanfaat, maka ditemukan berbagai algoritma *data mining* (Siregar & Puspabhuana, 2017). Menurut Larose & Larose (2014) jumlah data yang cukup banyak dan bervariasi dapat diolah menggunakan *data mining* yaitu klasifikasi, dimana

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

menggolongkan atau mengelompokan berdasarkan data yang ada serta dapat memprediksi data baru masuk ke dalam kelompok yang mana.

Pemilihan algoritma disesuaikan dengan tujuan pengolahan data, seperti untuk klasifikasi (Siregar & Puspabhuana, 2017). Klasifikasi merupakan salah satu teknik pada *data mining* dengan mempertimbangkan atribut dari kelompok data yang telah dikategorikan sebelumnya. Atribut yang dimaksud adalah sebagai variabel yang digunakan untuk menentukan kelas suatu objek baru. Menentukan kelas dengan akurat bagi objek yang kelasnya belum diketahui adalah tujuan dari klasifikasi (Imanda, Hidayat, & Furqon, 2018).

Proses klasifikasi sering menggunakan kombinasi algoritma, yaitu *Bootstrap* dan *Aggregating* (*Bagging*) yang dikemukakan oleh Breiman pada tahun 1996 untuk meningkatkan akurasi hasil prediksi. Untuk mengurangi variansi dan menghindari *overfitting*, *Bagging* dirancang dengan mengombinasikan set data *training* secara acak guna meningkatkan hasil klasifikasi (Wahono & Suryana, 2013).

Alasan penggunakan *Bagging* yaitu mengurangi varians dalam prediksi dengan cara membangun beberapa model, setiap model dilatih dari beberapa data asli dan mengkombinasikan prediksi yang dihasilkan oleh setiap model (Serrant, 2023). Adapun peran bootstrap, yaitu membantu mencegah *overfitting*, mengurangi efek *noise* dalam *dataset*, dan menggabungkan hasil dari beberapa model agar menghasilkan prediksi yang lebih baik dari model individu.

Algoritma sederhana yang sering diterapkan pada klasifikasi yaitu *K-Nearest Neighbor* (*K-NN*). Tujuan *K-NN* yaitu mengkategorikan data *testing* sesuai sesuai dekat atau tidaknya jarak di antara data *training* terhadap data *testing*-nya (Han, Kambel & Pei, 2011). Pengklasifikasian data *testing* dilakukan untuk menemukan beberapa kategori dalam data *training* ke data *testing* menginformasi makin besarnya kemiripan yang kedua data miliki (Wahono & Suryana, 2013).

Bagging K-NN dapat diterapkan pada bidang pengkreditan. Mengingat banyaknya jumlah pengajuan kredit dengan dilengkapi berbagai karakter calon debitur, maka kreditur harus mempertimbangkan dengan tepat apakah seorang calon debitur bisa menyelesaikan pembayaran kredit dengan sangat lancar ataupun tidak (Astuti, Hayati & Goejantoro, 2021). Janji serta kesanggupan dalam membayar dari debitur kepada kreditur merupakan salah satu unsur penting dalam kredit (Firdaus, 2004).

Perusahaan akan melakukan evaluasi kredit ataupun disebut penilaian kredit (*credit scoring*) yang menjadi pertimbangan dalam pengambilan keputusan mengenai pengajuan kredit dari calon debitur. *Credit scoring* ialah tahap evaluasi pengajuan kredit dengan berbagai atribut yang diketauhi dari calon debitur. Model *credit scoring* diambil dari sampel waktu lampau kemudian terbagi 2 kelas, yaitu kredit baik (pembayaran lancar) dan kredit macet (pembayaran tidak lancar). Persetujuan kelayakan kredit pada *credit scoring* menggunakan *data mining* untuk memperoleh klasifikasi kredit lancar dan kredit macet yang dapat menjadi acuan pada pemberian kredit kepada calon debitur (Astuti, Hayati & Goejantoro, 2021).

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Dalam melakukan analisis kredit dilakukan dengan cara tahap penilaian 5C yaitu *character* (kepribadian), *capacity* (kemampuan), *capital* (modal), *condition* (kondisi ekonomi), dan *collateral* (angsuran). Penilaian 5C sering dilakukan manual sehingga masih sering terjadi permasalahan seperti adanya debitur yang terlambat membayar angsuran sehingga status pembayaran menjadi kredit macet (Kholifah & Insani, 2016).

Analisis pada permasalahan kredit dapat diselesaikan dengan analisis metode *data mining* untuk menentukan parameter kelayakan kredit dari debitur yang akan melakukan pinjaman (Melissa & Oetama, 2013). Dalam jumlah data yang besar, teknik *data mining* berfungsi untuk menggali pola dan informasi yang relevan (Han, et al, 2012: 8).

Data *training* adalah data asli yang ada sebelumnya berdasarkan dengan fakta, sedangkan *data testing* adalah data yang dipakai dalam mengukur sejauh mana pengklasifikasian berhasil dengan benar (Azmi, 2023). Teknik yang diterapkan dengan menggunakan data pelanggan sebelumnya yang sudah pernah mengajukan kredit dan dijadikan data *training* untuk menguji data baru (data *testing*). Sehingga data baru atau calon debitur baru berhak mendapat penyaluran kredit berdasarkan kemiripan atribut yang ada sesuai data lama dengan data baru (Jasmir, Sika & Amelia, 2022).

Pembagi data pada data *training* dan data *testing* dapat dilakukan menggunakan beberapa proporsi, misalnya 90:10, 80:20, 70:30, 60:10, 50:50, karena untuk mengevaluasi seberapa tepat suatu model klasifikasi memprediksi hasil pada setiap kategori. Dengan membandingkan proporsi hasil yang diprediksi dengan hasil data asli, sehingga dapat mengukur akurasi dan kinerja model secara keseluruhan. Proporsi yang optimal adalah proporsi yang menyajikan gambaran yang paling akurat dan terbaik.

Permasalahan yang sering terjadi di perusahaan pembiayaan konsumen dikarenakan ada banyak persyaratan yang perlu dilakukan pertimbangan sebelum pencairan peminjaman tersalurkan untuk calon debiturnya yang mengakibatkan kredit tidak lancar dari debitur maka berdampak adanya rugi dan resiko yang dialami lembaga pembiayaan (Amelia, Hayati & Prangga, 2022). Untuk menentukan status kredit lancar berdasarkan karakteristik debitur, mencakup usia, jumlah tanggungan, lama tinggal, masa kerja, pendapatan, pengeluaran, *On The Road* (OTR), uang muka, besar pembayaran pinjaman, angsuran per bulan, jangka pembayaran, dan status pembayaran kredit.

Menurut Kasanah, Muladi & Pujianto (2017) menyampaikan bahwa evaluasi klasifikasi sesuai pengujian pada objek yang benar dan salah guna melakukan penentuan jenis model yang paling baik. Evaluasi klasifikasi dalam penelitian ini menerapkan *confusion matrix* yang menjelaskan tentang hasil klasifikasi *actual* yang dapat diprediksi oleh sistem klasifikasi. Dari hasil *confusion matrix* dapat dihitung nilai *accuracy, precision, recall,* dan *F1-score*. Akurasi adalah ketetapan sistem dalam melakukan proses klasifikasi dengan benar, sehingga pemilihan tepat dengan perhitungan *accuracy*.

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Berdasarkan penelitian Putri Sri Astuti, dkk (2021) telah menerapkan model *credit scoring* untuk menganalisis status pembayaran barang elektronik dan *furniture* di PT. KB Finansia Multi *Finance* sehingga memperoleh model *credit scoring* dalam model regresi logistik. Maka dalam penelitian ini, penulis bertujuan untuk membuat model *credit scoring* dengan menerapkan metode *Bootstrap Aggregating K-Nearest Neighbor* sehingga menghasilkan model *credit scoring* pembayaran kredit kendaraan motor yang diinginkan.

METODE PENELITIAN

Jenis penelitian yang dipergunakan ialah metode kuantitatif, dimana bila menujuk pada (Aliaga & Gunderson, 2002) penelitian kuantitatif didefinisikan sebagai penelitian yang dimaksudkan untuk menguraikan suatu fenomena dengan melakukan pengumpulan data berupa angka kemudian dianalisis melalui metode dengan matematis atau statistik tertentu (Fauzy et al, 2022).

Penelitian ini dengan data sekunder, yaitu data yang didapat berdasarkan arsip perusahaan dimana telah dikumpulkan oleh PT. Mega Central *Finance* (MCF) Cabang Rantau Prapat selama 6 bulan pada Januari 2024 - Juni 2024. Penyelesaian metode *Bootstrap Aggregating K-Nearest Neighbor* pada penelitian ini dengan menggunakan bantuan *software* "*phyton*".

Dalam memudahkan peneliti agar memberikan hasil yang lebih baik, lebih teliti, lebih sistematis, dan lebih mudah diolah para peneliti menggunakan instrument penelitian yang merupakan fasilitas atau peralatan untuk mengumpulkan data (Hikmawati, 2020). Jenis instrumen penelitian dibagi menjadi tiga, sebagai berikut:

1. Wawancara; adalah skenario sosial Dimana kedua belah pihak harus bereaksi secara timbal balik untuk memenuhi sasaran penelitian (Hardani, et al, 2020). Wawancara dilakukan antara peneliti dengan yang bertanggung jawab dalam pengarsipan data perusahaan.
2. *Computer Assisted Personal Interviewing* (CAPI); peneliti memasukkan informasi secara langsung ke dalam basis data (Hardani, et al, 2020).
3. Observasi; merupakan pengamatan terjun langsung ke lapangan, diikuti dengan obsevasi terhadap gejala yang diteliti, setelah itu peniliti dapat menjelaskan permasalahan dan menghubungkannya dengan wawancara ataupun *Computer Assisted Personal Interviewing* (CAPI) (Sahir, 2022).

Jika disesuaikan pada jenis instrument penelitian di atas, maka dengan melakukan observasi, dimana peneliti melakukan wawancara terlebih dahulu, kemudian melakukan *Computer Assisted Personal Interviewing* (CAPI).

Teknik pengumpulan data pada penelitian ini adalah dengan dokumentasi, dimana berupa riwayat peristiwa yang telah berlalu yang bentuknya tulisan dari perusahaan (Hikmawati, 2020). Teknik pengambilan sampel yaitu dengan cluster sampling, dimana teknik sampling ini melalui dua tahap, yaitu tahap pertama menentukan sampel pada setiap bulan,

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume xx Nomor xx Bulan Tahun xx E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

tahap berikutnya menentukan status pembayaran nasabah yang ada pada bulan tertentu secara sampling (Hardani, et al, 2020).

Teknik analisis data yang dipergunakan yaitu statistik inferensial, dimana agar dapat melihat keeratan hubungan antara variabel sehingga dapat membentuk kesimpulan berdasarkan hasil penelitian (Sahir, 2022). Penelitian ini menggunakan analisis data untuk menginterpretasikan dan merumuskan kesimpulan dari data yang diperoleh (Ngaisah, 2019).

HASIL DAN PEMBAHASAN

Standardisasi Data

Pada data penelitian ini dapat ditinjau bahwa terdapat perbedaan satuan setiap variabel, maka dapat dilakukan standardisasi data. Standardisasi data adalah teknik untuk mengubah skala data. Tujuannya adalah untuk menampilkan data yang seragam dan dapat menurunkan risiko duplikasi data serta memudahkan analisis data. Prinsip standardisasi data yaitu memastikan keterbandingan. Standardisasi data cuma dilakukan pada semua variabel bebasnya (X). Langkah-langkah menghitung standardisasi data, yaitu mencari mean dan standard deviasi, lalu mencari Z-score dengan rumus $z = \frac{(x-\mu)}{\sigma}$, kemudian menginterpretasi Z-score. Menunjukkan bahwa standardisasi sedang memusatkan semua data di satu titik yang sama pada interval -2,68 sampai dengan 2,84.

Permutasi Data

Jika sudah melakukan standardisasi data, dapat dilanjutkan mengacak data dengan permutasi data bertujuan untuk memastikan bahwa setiap data set memiliki kesempatan peluang yang sama untuk digunakan sebagai data *training* dan data *testing,* bahwa hasil klasifikasi dapat dipercaya, dan bahwa temuan klasifikasi yang diperoleh dapat bertahan dari permutasi berikutnya. Adapun alasan dilakukan permutasi data yaitu model tidak memperhatikan dari urutan data yang mungkin memiliki pola tertentu dan permutasi membantu memastikan bahwa setiap subset memiliki representasi yang baik dari keseluruhan dataset, menunjukkan hasil permutasi data dengan random_state=42.

Pembagi Data

Dengan dilakukan permutasi data, dapat dilakukan pembagi data dengan beberapa proporsi bertujuan mengoptimalkan model dan diuji langsung. Penelitian ini menetapkan tiga proporsi dalam pembagian data *training* dan data *testing* yaitu 90%:10%, 70%:30%, dan 50%:50%. Hasil pembagi data dengan proporsi 90:10 maka diperoleh 108 data *training* dan 12 data *testing*.

Pembagi data dapat dilakukan juga pada proporsi yang lain yaitu proporsi 70:30 diperoleh 84 data *training* dan 36 data *testing* begitu juga proporsi 50:50 diperoleh 60 data *training* dan 60 data *testing*.

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume xx Nomor xx Bulan Tahun xx E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Analisis Data

Jarak Data

Jarak euclid dipergunakan untuk menghitung jarak data *training* ke setiap data *testing*, maka digunakan rumus euclid sebagai berikut.

$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$

Dimana $a = a_1, a_2, \ldots, a_n$ dan $b = b_1, b_2, \ldots, b_n$ mewakili n nlai atribut dari dua record. (Leidiyana, 2013)

Tabel 1. Hasil Jarak Euclid Proporsi 90:10

Sampel	Klasifikasi	Jarak Data				
		DB-99	DB-91	DB-27	DB-114	DB-16
DB-63	TL	3,11	2,69	4,90	2,94	7,16
DB-31	L	2,79	3,87	5,53	4,48	7,22
DB-58	TL	4,31	3,36	5,82	3,67	8,51
DB-33	TL	7,46	4,63	2,79	4,18	6,74
DB-87	L	6,64	4,06	1,44	4,25	5,05
⋮	⋮	⋮	⋮	⋮	⋮	⋮
DB-3	TL	4,02	3,15	4,15	3,48	5,89

Perhitungan jarak euclid proporsi 90:10 dilakukan hingga data *testing* 12, begitu juga untuk proporsi lainnya.

Klasifikasi Data

Jarak euclid pada tiap-tiap data *testing* disusun dari yang terdekat (nilai terkecil) hingga yang terjauh (nilai terbesar). Data *testing* kemudian akan dikategorikan merujuk pada nilai K yaitu K=1, K=3, K=5, K=7, dan K=9. Batas K-NN mengacu pada batas K, yaitu jumlah tertangga terdekatnya yang dipergunakan sebagai panduan untuk mengidentifikasi kategori variabel *dependent*. Hasil klasifikasi proporsi 90:10 pada data *testing* 1 bisa ditinjau pada.

Tabel 2. Hasil Klasifikasi Proporsi 90:10 Pada Data *Testing* 1

Rank	Data *Training*		Jarak Data	Batas K-NN	Hasil Klasifikasi Data *Testing*
	Sampel	Klasifikasi			
1	DB-41	L	2,69	1-NN	L
2	DB-31	L	2,79		
3	DB-26	L	2,89	3-NN	L
4	DB-63	TL	3,11		
5	DB-5	L	3,20	5-NN	L
6	DB-111	L	3,33		
7	DB-38	L	3,41	7-NN	L
8	DB-75	L	3,42		

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

9	DB-108	TL	3,75	9-NN	L
⋮	⋮	⋮	⋮		
108	DB-110	L	8,89		

Hasil klasifikasi proporsi 90:10 dilakukan hingga data *testing* 12, begitu juga untuk proporsi lainnya.

Proporsi Terbaik

Dalam menentukan nilai K optimal serta proporsi terbaik bisa diperoleh dengan minjau nilai akurasi dari data *testing* setiap proporsi. Perhitungan akurasi dimulai dari menghitung akurasi setiap data *testing* berdasarkan nilai K, kemudian menghitung rata-rata akurasi dari setiap proporsi yang ada.

Setelah didapat hasil akurasi setiap K-NN pada 3 (tiga) proporsi, kemudian memilih proporsi terbaik berdasarkan nilai akurasi tertinggi dan jarak terdekat. Adapun nilai akurasi hasil klasifikasi dapat ditinjau pada Tabel 4.8 sebagai berikut.

Tabel 3. Nilai Akurasi Hasil Klasifikasi

KNN	Proporsi Data *Training* dan Data *Testing*		
	90:10	70:30	50:50
1-NN	83,33 %	75,00 %	73,33 %
3-NN	83,33 %	72,22 %	76,67 %
5-NN	83,33 %	80,56 %	81,67 %
7-NN	91,67 %	83,33 %	81,67 %
9-NN	91,67 %	83,33 %	78,33 %

dapat ditinjau bahwa proporsi 90:10 pada batas 7 tetangga terdekat (7-NN) memiliki nilai akurasi tertinggi, yaitu sebesar 91,67%. Berdasarkan hal ini, K=7 dengan 108 data *training* dan 12 data *testing* akan menjadi jumlah *K* yang ideal untuk digunakan pada analisis berikutnya.

Penerapan Bootstrap

Pada pembahasan ini berlandaskan dengan pengambilan dari sampel yang dilakukan dengan berulang-ulang (*resampling*) serta pengembaliannya, maka sangat mungkin ada kelebihan satu kali sampel terambil. Langkah awal yang harus dikerjakan yakni mengumpulkan sampelnya secara *random* dari data *training* bertujuan agar berubah menjadi data *bootstrap* sejumlah *m*% dari jumlah data *training*. Dimana proporsi 90.10 sudah terpilih sebab mempunyai nilai akurasi yang tertinggi, maka jumlah data *boootstrap* yang diambil yakni (*m*% × 108) data. Selanjutnya data *bootstrap* akan dijadikan data *training* bertujuan agar mengkategorikan ulang 12 data *testing* melalui langkah yang tidak berbeda seperti pada langkah K-NN. Sesudahnya, langkah *resampling* sampai langkah klasifikasinya dilakukan pengulangan sebanyak *C* kali.

Pada penelitian ini dilakukan penggunaan *m*=50%, *m*=60%, *m*=70%, *m*=80%, *m*=90%, dan *m*=100%. Dengan nilai *C* dipilih secara *random* sebanyak 8 dari rentang 20 hingga 108.

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Bila merujuk pada hasil random didapat 8 nilai *C* yang dipilih, yaitu 21, 33, 39, 46, 50, 64, 72, dan 76. Proses klasifikasi menerapkan nilai K=1, K=3, K=5, K=7, dan K=9. Penerapan nilai *m, C*, dan K dilaksanakan satu demi satu dalam langkah klasifikasi. Berikutnya dianalisis dan dipilih yang terbaik berdasarkan akurasi yang didapat.

Untuk menguraikan tahapan *bootstrap* hingga perhitungan akurasi, peneliti menentukan penggunaan dari *m*=80%, *C*=50, dan K=7 sebagai contoh. Sedemikian sehingga, 86 data *bootstrap* atau (80% × 108) dari total datanya dipilih. Metode klasifikasinya dilakukan pengulangan sebanyak 50 kali dengan menggunakan batas 7-NN. Data *bootstrap* 1 dapat ditinjau pada Tabel 4.9 sebagai berikut.

Tabel 4. Data *Bootstrap* 1

Sampel	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
DB-15	L	-0,91	-0,45	-1,24	-0,86	0,45	-0,59	-0,71	0,90	-1,40	-0,48	-0,89
DB-37	L	-0,62	-1,00	1,68	-0,49	0,45	-1,02	-0,88	-0,88	-0,66	-0,22	-0,89
DB-64	L	-0,34	0,11	-0,77	-0,49	0,45	-0,16	-0,15	-0,54	0,10	0,03	0,00
DB-13	L	-0,14	-0,45	0,05	-0,86	0,01	-1,02	-0,88	-0,63	-0,80	-0,93	0,75
DB-13	L	-0,14	-0,45	0,05	-0,86	0,01	-1,02	-0,88	-0,63	-0,80	-0,93	0,75
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
DB-65	TL	-0,24	0,11	-1,24	-0,30	-1,31	-0,16	-0,72	-0,88	-0,45	-0,42	-0,15

Dapat dilanjutkan hingga perhitungan pada data *bootstrap* 50.

Ketepatan Akurasi

Dengan sudah didapat klasifikasi hasil prediksi untuk setiap data *testing*, berikutnya dilakukanlah perbandingan dari hasil klasifikasi prediksi dengan metode *bagging K*-NN dengan klasifikasi data asli yakni setiap 12 data *testing*.

Diketahui bahwasannya sel yang diberi tambahan tanda (*) menandakan adanya perbedaan pada klasifikasi hasil prediksi dengan klasifikasi data asli. Dengan kesimpulan 10 debitur dengan status pembayaran kredit lancar dengan benar, dan juga 1 debitur dengan status pembayaran kredit tidak lancar dengan benar. Namun diperoleh 1 debitur yang tidak benar diklasifikasikan, dimana harusnya debitur dengan status pembayaran kredit tidak lancar tetapi klasifikasi hasil prediksi membuktikan status pembayaran kredit lancar.

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Setelah didapat klasifikasi hasil prediksi bagi tiap-tiap data *testing* dengan langkah *aggregating*, kemudian melakukan perhitungan akurasi dengan *confusion matrix*.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$
$$accuracy = \frac{10+1}{10+1+1+0} \times 100\%$$
$$accuracy = \frac{11}{12} \times 100\%$$
$$accuracy = 0{,}9167 \times 100\%$$
$$accuracy = 91{,}67\%$$

Berdasarkan perhitungan akurasi diatas menerapkan metode *bagging K-NN* dengan proporsi 90:10, K=7, *m*=80%, dan *C*=50 diperoleh akurasi terbaik sebesar 91,67%.

Aplikasi *KNN* dalam kehidupan sehari-hari salah satunya dibidang keuangan dalam penilaian kredit. *KNN* menganalisis riwayat data dalam mengidentifikasi pola dengan bermanfaat dalam menentukan keputusan klasifikasi status pembayaran kredit yang lebih baik.

SIMPULAN DAN SARAN

Simpulan

Bila merujuk pada hasil dan pembahasan, bisa diambil suatu kesimpulan sebagai berikut:

1. Klasifikasi hasil prediksi yang didapat berdasarkan klasifikasi data asli pada status pembayaran kredit motor diperoleh 12 data *testing* dengan klasifikasi 10 debitur dengan status pembayaran kredit lancar dengan benar, dan juga 1 debitur dengan status pembayaran kredit tidak lancar dengan benar. Namun diperoleh 1 debitur yang tidak benar diklasifikasikan, dimana harusnya debitur dengan status pembayaran kredit tidak lancar tetapi klasifikasi hasil prediksi menunjukkan status pembayaran kredit lancar. Akurasi terbaik yang diperoleh dari algoritma *Bootstrap Aggregating K Nearest Neighbor* (*Bagging K-NN*) dari hasil analisis *credit scoring* dengan proporsi 90:10, K=7, *m*=80%, dan *C*=50 adalah sebesar 91,67%.

Implementasi metode *Bootstrap Aggregating* pada model klasifikasi *K-Nearest Neighbor* dengan hasil *Bagging K-NN* tidak meningkatkan ketepatan akurasi klasifikasi *K-NN* yang semula 91,67% tetap menjadi 91,67% tidak stabil.

Saran

Bila merujuk pada hasil dan pembahasan, beberapa saran yang dapat dilakukan sebagai berikut:

Untuk penelitian berikutnya dengan mengaplikasikan dataset yang memiliki lebih dari satu kategori pada variabel terikat (y) dapat dilakukan perbandingan algoritma klasifikasi dengan Decision Tree, Naive Bayes, Random Forest, Logistic Regression, Neural Network, Linear Discriminant Analaysis, atau Support Vector Machine (SVM).

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

DAFTAR PUSTAKA

Adrian, C. (2019). *Pengembangan Model Credit Scoring Perusahaan X.* (Skripsi Sarjana, Universitas Katolik Parahyangan Bandung).

Amelia, S., Hayati, M. N., & Prangga, S. (2022). Penerapan Metode Modified K-Nearest Neighbor Pada Pengklasifikasian Status Pembayaran Kredit Barang Elektronik dan Furniture. *Statistika*, 95-104.

Andhayani, D., Harianto, & Achsani, N. A. (2009). Pengembangan Model Credit Scoring Untuk Proses Analisis Kelayakan Fasilitas Kredit Pemilikan Rumah (Studi Kasus di Bank X). *Jurnal Manajemen & Agribisnis*, 65-73.

Antonio, M. S. (2001). *Bank Syariah: Dari Teori ke Praktik.* Jakarta: Gema Insani Press.

Arif, N. R. (2012). *Dasar -Dasar Pemasaran Bank Syariah.* Bandung: Alfabeta.

Astuti, P. S., Hayati, M. N., & Goejantoro, R. (2021). Analisis Credit Scoring Terhadap Status Pembayaran Barang Elektronik dan Furniture Menggunakan Bootstrap Aggregating K-Nearest Neighbor. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 735-744.

Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 281-290.

Buku Panduan Pengajuan Kredit. (n.d.).

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 118-127.

Diaprina, S. R., & Suhartono. (2014). Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner dan Radial Basis Function Network di Bank 'X' Cabang Kediri. *Jurnal Sains dan Seni Pomits*, 219-223.

Efron, B., & Tibshirani, R. (1993). *An Introduction to The Bootstrap.* New York: Chapman & Hall.

Enterprise, J. (2019). *Phyton Untuk Programmer Pemula: Buku Bergizi Untuk Programmer Pemula Yang Ingin Mempelajari Phyton (Cet. 2).* Jakarta: Elex Media Komputindo.

Fauzy, A., & dkk. (2022). *Metodologi Penelitian.* Banyumas: Pena Persada.

Han, J., Kambel, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques Third Edition.* San Francisco: Morgan Kaufmann Publisher.

Hardani, & dkk. (2020). *Metode Penelitian Kualitatif & Kuantitatif.* Yogyakarta: Pustaka Ilmu.

Hariyani, I. (2010). *Restrukturisasi dan Penghapusan Kredit Macet.* Jakarta: Elex Media Komputindo.

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Henderi, Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of Min-Max Normalization and Z-Score Normalization in the K-Nearest Neighbor (KNN) Algorithm to Test the Accuracy of Type s of Breast Cancer. *International Journal of Informatics and Information System*, 13-20.

Hikmawati, F. (2020). *Metodologi Penelitian.* Depok: Rajawali Pers.

https://tambahpinter.com/metode-penelitian-kuantitatif/#Desain_Penelitian. (n.d.).

https://www.gramedia.com/literasi/pengertian-bi-checking/. (n.d.).

https://www.rumah.com/panduan-properti/bi-checking-kol-83595 . (n.d.).

Jasmir, Sika, X., & Amelia, R. (2022). Klasifikasi Kelayakan Pemberian Kredit Pada Calon Debitur Menggunakan Naive Bayes. *JURIKOM: Jurnal Riset Komputer*, 1833-1839.

Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data Transformation Pada Data Mining. *Prosiding Konferensi Nasional "Inovasi dalam Desain dan Teknologi"* (pp. 93-99). Surabaya: IDeaTech.

Kasanah, A. N., Muladi, & Pujianto, U. (2019). Penerapan Teknik SMOTE Untuk Mengatasi Imbalance Class Dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI: Rekayasa Sistem dan Teknologi Informasi*, 196-201.

Khoirudin, W. (2023). *Penerapan Metode Bootstrap Aggregating Classification and Regression Trees Untuk Menentukan Klasifikasi Indeks Khusus Penangan Stunting.* (Skripsi Sarjana, Universitas Islam Negeri Maulana Malik Ibrahim Malang).

Kholifah, A. N., & Insani, N. (2016). Analisis Klasifikasi Pada Nasabah Kredit Koperasi X Menggunakan Decision Tree C4.5 dan Naive Bayes. *JKTM: Jurnal Kajian dan Terapan Matematika*, 1-8.

Kurniawan, C. C., Rostianingsih, S., & Santoso, L. W. (2022). Penerapan Metode KNN-Regresi dan Multiplicative Decomposition untuk Prediksi Data Penjualan pada Supermarket X. *Jurnal Infra*, 1-7.

Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining.* Yogyakarta: Andi Publishing.

Larose, D. T. (2005). *Discovering Knowledge in Data.* New Jersey: John Willey & Sons, Inc.

Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining, Second Edition.* New Jersey: John Wiley & Sons, Inc.

Leidiyana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, 65-76.

Melissa, I., & Oetama, R. (2013). Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining. *ULTIMA InfoSys*, 18-27.

Mukid, M. A., Wuryandari, T., Ratnaningrum, D., & Rahayu, R. S. (2015). Bagging Classification Trees Untuk Prediksi Risiko Preeklampsia (Studi Kasus: Ibu Hamil Kategori Penerima Jampersal di RSUD Dr. Moewardi Surakarta). *Media Statistika*, 111-120.

JEms Emasains
Jurnal Edukasi Matematika dan Sains
Volume xx Nomor xx Bulan Tahun xx
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v14i2.5019

Ngaisah, S. (2019). *Pengaruh Kualitas Produk, Harga dan Citra Merek Terhadap Keputusan Pembelian Ulang Produk Natasha Skin Care.* (Skripsi Sarjana, Universitas Pelita Bangsa).

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI: Jurnal Sains Komputer & Informatika*, 697-711.

Nuriksan, A., Pudjiantoro, T. H., & Sabrina, P. N. (2021). Klasifikasi Kelayakan Kredit Motor Menggunakan Metode Naive Bayes dan Model Credit Scoring. *SNIA: Seminar Nasional Informatika dan Aplikasinya*, 1-6.

Patience, A. K., & Lakshmi. (2020). A Study on Credit Card Fraud Detection Using Machine Learning. *IJTSRD: International Journal of Trend in Scientific Research and Development*, 801-804.

Prasetyo, E. (2014). *Data Mining: Mengolah Data Menjadi Infomrasi Menggunakan Matlab.* Yogyakarta: ANDI.

Pratiwi, B. P., Handayani, A. S., & Sarjana. (2020). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusin Matrix. *Jurnal Informatika Upgris*, 66-75.

Sahir, S. H. (2022). *Metodologi Penelitian.* Medan: KBM Indonesia.

Saputra, K., & Utama, A. P. (2016). Klasifikasi Data Minuman Wine Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Pembangunan Panca Budi*, 1-3.

Saputro, D. (2011). *Model Credit Scoring Untuk Proses Analisa Kelayakan Fasilitas Kredit Motor Menggunakan Metode Classification And Regression Tree (CART).* (Skripsi Sarjana, Universitas Islam Negeri Syarif Hidayatullah).

Serrant, D. (2023). *What Is The Reason That Bagging (Bootstrap Aggregation) Reduces Variance More Than Bias For Regression Models?* Quora.

Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner.* Surakarta: CV Kekata Group.

Sugiyono. (2007). *Statistika Untuk Penelitian.* Bandung: Alfabeta.