

Prediksi Penambahan Piutang Iuran Jaminan Sosial Ketenagakerjaan menggunakan Algoritma K-Nearest Neighbor

Devi Efriadi^{1,*}, Rahmaddeni¹, Agustin², Junadhi¹

¹ Program Studi Teknik Informatika, STMIK Amik Riau, Indonesia

² Program Studi Teknologi Informasi, STMIK Amik Riau, Indonesia

* Correspondence: 2010031802031@sar.ac.id

Copyright: © 2022 by the authors

Received: 4 Maret 2022 | Revised: 19 Maret 2022 | Accepted: 1 April 2022 | Published: 20 Juni 2022

Abstrak

Saat ini Badan Penyelenggara Jaminan Sosial (BPJS) Ketenagakerjaan memiliki beberapa masalah salah satunya adalah Piutang Iuran. Untuk mengurangi piutang iuran BPJS tersebut sudah berbagai cara dilakukan oleh pihak BPJS. Namun usaha yang dihasilkan belum cukup maksimal untuk menekan angka piutang yang ada di BPJS. Penelitian ini bertujuan untuk memberikan masukan dengan memprediksi penambahan piutang iuran jaminan sosial Ketenagakerjaan yang dilakukan oleh beberapa perusahaan atau organisasi. Penelitian ini menggunakan Algoritma K-Nearest Neighbor (KNN) dengan teknik *cross validation*. KNN merupakan metode klasifikasi yang sangat sederhana dalam mengklasifikasikan sebuah gambar berdasarkan jarak terdekat dengan tetangganya. Penelitian ini melakukan pengolahan data dari BPJS ketenagakerjaan yang berjumlah 1193 data. Data tersebut kemudian dilakukan preprocessing agar data yang diolah sudah bersih dari missing dan noise, data ini menggunakan splitting data 70:30. Setelah dilakukan proses preprocessing dan splitting data selanjutnya yaitu melakukan pemodelan menggunakan KNN, pada penelitian ini juga menggunakan cross validation untuk meningkatkan hasil akurasi yang didapat dari algoritma KNN. Hasil yang didapatkan dari penelitian ini mendapatkan akurasi tertinggi 92% dengan Nilai K Optimal adalah 6, kemudian kurva ROC mendapatkan hasil akurasi 94%. Dari hasil ini dapat dikatakan bahwa penggunaan cross validation bisa meningkatkan akurasi pada penelitian ini.

Kata kunci: BPJS ketenagakerjaan; cross validation; KNN; piutang iuran

Abstract

There are several issues with Social Security Organizing Agency (BPJS) employment at the moment, one of which is contribution receivable. To reduce the BPJS contribution receivables, BPJS has done various ways. However, the resulting effort is not maximal enough to reduce the number of receivables in BPJS. This study aims to provide input by predicting the addition of receivables from social security contributions made by several companies or organizations. This study used the K-Nearest Neighbor (KNN) Algorithm with a cross-validation technique. KNN is a very simple classification method in classifying an image based on the closest distance to its neighbors. This study conducted data processing from BPJS use, which amounted to 1193 data. The data is then preprocessed so that the processed data is clean from missing and noise, this data uses 70:30 data splitting. After the preprocessing and splitting of data were carried out, the next step was to do modeling using KNN, so the cross-validation to improve the accuracy of results obtained from the KNN algorithm. The results obtained from this research get the highest accuracy of 92% with the Optimal K value being 6, then the ROC curve gets 94% accuracy. From these results, it can be said that the use of cross-validation can increase the accuracy of this study.

Keywords: BPJS ketenagakerjaan; cross validation; dues receivable; KNN



PENDAHULUAN

Badan Penyelenggara Jaminan Sosial (BPJS) adalah badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial. BPJS terdiri dari BPJS Kesehatan dan BPJS Ketenagakerjaan (Widiastuti, 2017). Keduanya BPJS ini sangat berbeda, BPJS kesehatan diperuntukan untuk kesehatan sedangkan BPJS Ketenagakerjaan melayani jaminan kepada perusahaan. Penelitian ini akan berfokus kepada BPJS ketenagakerjaan yang dijadikan objek penelitian. BPJS ketenagakerjaan merupakan perkembangan dari PT Jamsostek (Persero) pada tahun 2011. Program BPJS Ketenagakerjaan tidak hanya memberikan manfaat kepada pekerja dan pengusaha saja, tetapi juga memberikan kontribusi penting bagi peningkatan pertumbuhan ekonomi bangsa dan kesejahteraan masyarakat Indonesia (BPJS Ketenagakerjaan, 2017). BPJS ketenagakerjaan telah diatur oleh undang-undang Republik Indonesia Nomor 11 Tahun 2020 Tentang Cipta kerja. BPJS Ketenagakerjaan menyelenggarakan beberapa program, yaitu Jaminan kecelakaan kerja, Jaminan hari tua, Jaminan pensiun, Jaminan kematian, dan Jaminan kehilangan pekerjaan.

BPJS Ketenagakerjaan telah diatur dalam peraturan menteri ketenagakerjaan tahun 2016 tentang penyelenggaraan program-program yang ada pada BPJS ketenagakerjaan seperti jaminan hari tua, jaminan kematian, jaminan kecelakaan kerja (Pambudi, 2019). Saat ini BPJS ketenagakerjaan memiliki permasalahan terhadap piutang iuran jaminan sosial ketenagakerjaan. Iuran wajib yang dikeluarkan oleh pihak pemberi kerja/badan usaha sering mengalami keterlambatan bahkan sampai tidak membayarkan iuran tersebut.

Saat ini BPJS Ketenagakerjaan belum memiliki sebuah sistem prediksi terhadap pemberi kerja/badan usaha baru, sehingga menyebabkan penambahan piutang iuran dan klaim jaminan sosial mengalami penundaan, serta BPJS Ketenagakerjaan belum dapat menentukan atau melakukan langkah-langkah strategis untuk menekan penambahan piutang iuran. Dari permasalahan tersebut perlu dilakukan perbaikan, salah satunya dengan proses klasifikasi data mining terhadap pemberi kerja atau badan usaha. Pada penelitian ini akan melakukan analisis data yang tersedia dari BPJS ketenagakerjaan. Dari data tersebut akan dilakukan proses data mining dengan melakukan klasifikasi data dan melakukan uji performance dari hasil dengan menggunakan bahasa pemrograman python.

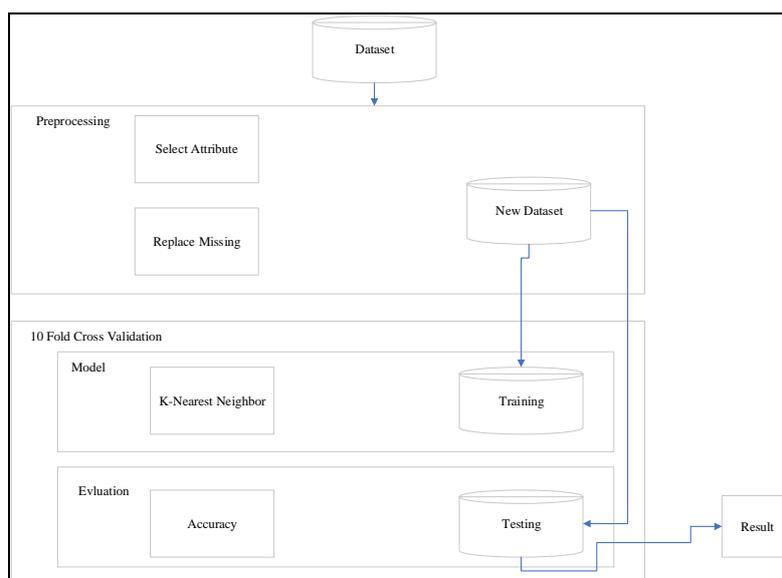
Klasifikasi data mining telah banyak dilakukan oleh peneliti terdahulu, diantaranya adalah Hendrian (2018) melakukan klasifikasi untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan. Kemudian Anggraini et al., (2019) melakukan klasifikasi terhadap data blogger, selanjutnya Parapat dan Sinaga (2018) klasifikasi kredit koperasi simpan pinjam. Dan yang terakhir adalah Iriadi & Nuraeni (2016) memprediksi kelayakan kredit pada bank mayapada jakarta. Dalam melakukan klasifikasi data mining ada beberapa metode yang dapat digunakan seperti algoritma C4.5 (Ardiansyah et al., 2021), Support Vector Machines (SVM) (Anam et al., 2022), Algoritma K-Nearest Neighbor (KNN) (Hidayat et al., 2021; Prasetyo et.al, 2020), dan lain sebagainya.

Pada penelitian ini metode yang digunakan adalah Algoritma KNN. Algoritma ini digunakan karena memiliki akurasi yang cukup tinggi seperti yang dihasilkan dari beberapa penelitian terdahulu. Penelitian (Fansyuri, 2020) menentukan nilai akurasi terhadap kepuasan pelanggan dan menghasilkan akurasi sebesar 83,33%, (Dinata et al., 2020) melakukan klasifikasi transportasi bus dan menghasilkan akurasi 84%, (Muniar et al., 2020) melakukan klasifikasi dokumen berita online dan menghasilkan akurasi 89,9%. Algoritma KNN merupakan metode klasifikasi yang sangat sederhana dalam mengklasifikasikan sebuah gambar berdasarkan jarak terdekat dengan tetangganya (Farokhah, 2019). Hasil yang didapatkan pada ketiga penelitian tersebut masih dibawah dengan metode yang lainnya menghasilkan akurasi diatas 90% (Rifai et al., 2019). Pada penelitian ini mencoba menggunakan KNN untuk meningkatkan menjadi diatas 90%. Selain itu, data penelitian ini menggunakan data piutang yang diambil dari nasabah BPJS ketenagakerjaan. Hasil yang

diharapkan dari penelitian ini adalah permasalahan yang dihadapi dapat terpecahkan. Sehingga tidak menambah piutang iuran dan pelayanan terhadap pekerja peserta program jaminan sosial ketenagakerjaan tidak mengalami penundaan. Penelitian ini juga menggunakan cross validation untuk meningkatkan akurasi yang lebih tinggi dibandingkan dengan penelitian sebelumnya sehingga hasil akurasi yang dihasilkan semakin baik.

METODE

Tahapan dalam penelitian ini di adopsi dari penelitian Purwaningsih & Nurelasari (2021), disajikan dalam gambar 1. Dataset yang digunakan dalam pengolahan adalah data pada BPJS ketenagakerjaan dumai mengenai piutang nasabah, dengan atribut yang dinilai adalah Skala Usaha, Rate Jkk, Kode Ilo, Umur Piutang, Keps Awal, Tenaga Kerja, Bulan Rekon, Nilai Rekon. Pada tahap *preprocessing* dilakukan *cleaning* untuk pemilihan fitur atribut yang dilakukan melalui *select attribute* dan *replace missing* untuk meminimalisir terjadinya kekosongan data dan *noise*. Tahap *cleaning* dilakukan untuk menghindari terjadinya masalah pada hasil yang didapatkan (Anam et al., 2021; Wang et.al, 2019). Kemudian langkah selanjutnya transformasi data merupakan suatu usaha yang ditujukan untuk mengubah skala pengukuran data asli menjadi bentuk yang lain (Sari et al., 2017). Dan yang terakhir adalah normalisasi dalam kegiatan data mining merupakan proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tersebut (Nasution et al., 2019).



Gambar 1. Metodologi penelitian

Selanjutnya tahap *cross validation* dilakukan pengujian data menggunakan metode k-Nearest Neighbor (KNN). *Nearest Neighbor* adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada (Mustafa & Simpen, 2019). Rumus untuk melakukan penghitungan kedekatan antara kedua dua kasus dapat dilihat pada persamaan 1 (Mustafa & Simpen, 2014). Untuk melakukan klasifikasi menggunakan Algoritma KNN, memiliki 2 tahapan yaitu: (1) Menentukan parameter K (jumlah tetangga paling dekat), (2) menghitung kuadrat jarak euclidian (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan dengan menggunakan persamaan 2, (3) Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil, (4) Mengumpulkan kategori Y (klasifikasi *nearest neighbor*), (5) Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung.

$$\frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i} \quad (1)$$

Dimana :

T : kasus baru

S : kasus dalam penyimpanan

n : jumlah atribut tiap kasus

i : atribut individu 1 s/d n

f : fungsi kesamaan atribut i antara kasus T dan S

w : bobot pada atribut yang ke i

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i, y_i)^2} \quad (2)$$

Penelitian ini menggunakan data *training* dan data *testing* yang diolah menggunakan cross validation. Cross validation dinilai menghasilkan performa yang baik untuk pembagian data (Pikir et al., 2021). Proses pengolahan data menggunakan KNN yang digunakan untuk menghasilkan akurasi klasifikasi piutang yang ada di BPJS ketenagakerjaan. Proses yang dilakukan menggunakan bahasa pemrograman Python. Sementara itu pada tahapan hasil adalah akurasi untuk memprediksi piutang terhadap BPJS ketenagakerjaan dapat tercapai.

HASIL DAN PEMBAHASAN

Hasil

Tahap pertama dari preprocessing adalah *cleaning*, gambar 2 merupakan tahap *cleaning* yang dihasilkan pada penelitian ini. Dari gambar 2 terlihat ada 1192 sampel yang digunakan pada penelitian ini. Kemudian setelah data tersebut dieksekusi masih terdapat data terindikasi *missing value*. Hasil output yang tertulis *False* dapat diartikan bahwa kolom tersebut tidak mengandung *missing value*. Kemudian pada hasil output *True* menandakan bahwa data tersebut ada *missing value*. Untuk itu perlunya langkah *cleaning* ini untuk menghilangkan *missing value* yang terdapat pada data yang akan diolah.

```

Number of features : 14
Number of sample : 1193

#cek kolom yang mengandung nilai null
df.isnull().any()
✓ 0.2s
NO                False
NPP               False
DIV              False
NAMA PK BU       False
ALAMAT           False
SKALA USAHA      False
RATE JIK         True
KODE ILO         False
UMUR PIUTANG     False
KEPS AMAL        True
TENAGA KERJA     True
BULAN REKON      True
NILAI REKON      True
KLASF            False
dtype: bool
    
```

Gambar 2. Tahapan *cleaning*

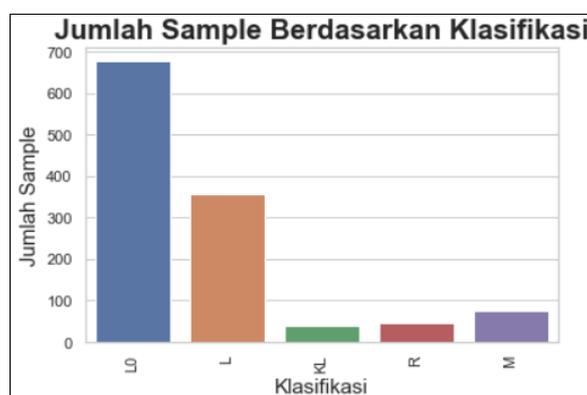
Setelah dilakukan proses *cleaning*, langkah selanjutnya ialah melakukan transformasi data yang terlihat pada gambar 3.

	SKALA USAHA	RATE JKK	UMUR PIUTANG	TENAGA KERJA	NILAI REKON	KLASF
0	1	0.24	0	1.0	208596.46	0
1	1	0.24	0	1.0	305760.00	0
2	0	1.27	3	7.0	1659705.03	1
3	1	0.24	0	1.0	208596.46	0
4	1	0.24	1	1.0	208596.40	1

Gambar 3. Transformasi data

Pada gambar 3, seluruh data dilakukan standarisasi dan normalisasi data, hal ini dilakukan agar data dapat dibaca oleh bahasa pemrograman python.

Pada tahapan *preprocessing* dilakukan klasifikasi data berdasarkan jumlah sampel yang digunakan. Hasil klasifikasi ini yang nampak pada gambar 4, dimana hasil tersebut terlihat bahwa masih terdapat beberapa perusahaan atau organisasi yang terindikasi melakukan piutang iuran, namun jumlah datanya tidak terlalu banyak dibandingkan dengan perusahaan atau organisasi yang melakukan pembayaran iuran dengan lancar.



Gambar 4. Visualisasi klasifikasi

```

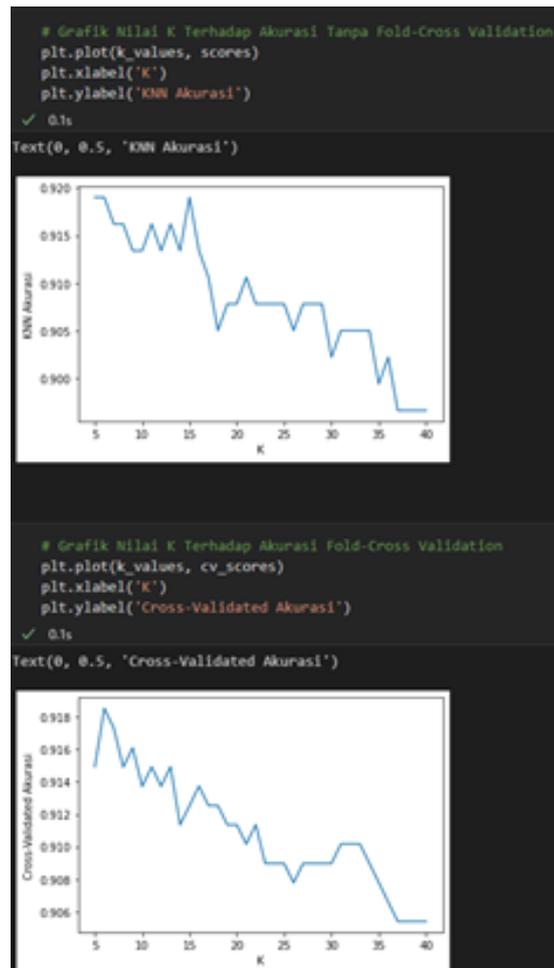
5
Akurasi = 0.91899
Akurasi K-Fold = 0.91493
6
Akurasi = 0.91899
Akurasi K-Fold = 0.91853

```

Gambar 5. Klasifikasi algoritma KNN

Sebelum data BPJS diolah menggunakan algoritma KNN, dilakukan *splitting* data yakni 70:30. Pada gambar 5 merupakan hasil klasifikasi yang dihasilkan. Dimana akurasi tertinggi didapat pada K=5. Pada gambar 4 terlihat bahwa nilai akurasi yang dihasilkan adalah 0,91899. Nilai ini merupakan nilai tertinggi dibandingkan dengan nilai K lainnya. Pada gambar 6 merupakan evaluasi akurasi baik menggunakan *cross validation* maupun tidak.

Selanjutnya hasil *cross validation* pada gambar 6, terlihat bahwa hasil ini sangat berpengaruh, akurasi tanpa menggunakan *cross validation* menghasilkan akurasi sebesar 0,920, ketika penelitian ini dilakukan menggunakan *cross validation*, angka akurasi yang dihasilkan meningkat menjadi 0,928. Pada gambar 7 merupakan akurasi yang dihasilkan dari akurasi model KNN dengan menggunakan *k-Fold validation*.

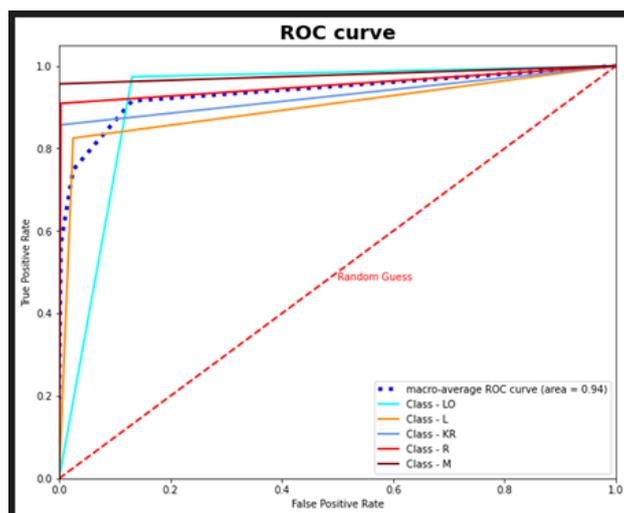


Gambar 6. Evaluasi akurasi terhadap prediksi dalam rentang nilai K=5 hingga K=40

	precision	recall	f1-score	support
0	0.89	0.97	0.93	190
1	0.94	0.82	0.88	120
2	0.92	0.86	0.89	14
3	0.91	0.91	0.91	11
4	1.00	0.96	0.98	23
accuracy			0.92	358
macro avg	0.93	0.90	0.92	358
weighted avg	0.92	0.92	0.92	358

Gambar 7. Accuracy model algoritma K-NN dengan k-Fold cross validation

Hasil akurasi yang terlihat pada gambar 8 menggunakan kurva ROC, nampak akurasi yang dihasilkan cukup tinggi yakni 92%. Hasil ini sudah cukup baik karena sudah melebihi dari 90%. Selanjutnya, Kurva ROC algoritma K-NN dengan k-Fold cross validation kurva ROC adalah tool 2 dimensi yang berfungsi sebagai penilaian kinerja klasifikasi yang menggunakan 2 class keputusan seperti positif dan negatif (Frastian, 2018).



Gambar 8. Kurva ROC algoritma K-NN dengan k-Fold *cross validation*

Dari hasil kurva ROC dapat dilihat bahwa penggunaan k-Fold Cross Validation juga mempengaruhi hasil yang didapat. Pada penelitian akurasi yang didapatkan dari ROC *score* adalah 94%. Tabel 1 merupakan hasil rekapitulasi yang dihasilkan dari uji coba yang dilakukan dengan splitting 70:30.

Tabel 1. Hasil rekapitulasi

No	Splitting Data	Tanpa Cross Validation K-NN	Nilai K Optimal	ROC Score	Dengan Cross Validation K-NN	Nilai K Optimal	ROC Score
1	70 : 30	0,911	5	0,93	0.916	6	0,94

Pembahasan

Hasil Penelitian yang telah dilakukan dengan menggunakan KNN cukup tinggi yakni diatas 90%. Kemudian penelitian ini menggunakan k-fold validation untuk meningkatkan akurasi yang telah didapatkan, dan hasilnya meningkat dari sebelumnya. Hal ini dapat disimpulkan bahwa penggunaan k-Fold *validation* pada KNN bisa meningkatkan akurasi. Akurasi yang dihasilkan dari Algoritma K-NN tanpa cross validation menghasilkan 91% dengan nilai K optimalnya adalah 5 dan kurva ROC menunjukkan hasil 93%. Hasil ini cukup tinggi bila dibandingkan dengan penelitian terdahulu seperti (Baharuddin et al., 2019) menggunakan KNN untuk identifikasi jenis kaca hanya menghasilkan akurasi 64%. Kemudian penelitian yang dilakukan (Fansyuri, 2020) menghasilkan akurasi 83,33%. Untuk meningkatkan akurasi 91% menjadi lebih tinggi lagi, penelitian ini menggunakan Cross Validation.

Cross validation digunakan oleh beberapa peneliti digunakan untuk meningkatkan akurasi. Seperti yang dilakukan oleh (Samponu & Kusri, 2018) menghasilkan akurasi 2% lebih tinggi dibandingkan tidak menggunakan cross validation. Kemudian hasil temuan yang dilakukan oleh peneliti sebelumnya menyatakan bahwa bahwa penggunaan *cross validation* dapat meningkatkan nilai akurasi (Azis et al., 2020; Rhomadhona & Permadi, 2019; Ulfah & Anam, 2020). Pada penelitian ini juga terjadi kenaikan akurasi menjadi 92% dengan Nilai K Optimal adalah 6. Kemudian kurva ROC menunjukkan hasil peningkatan yaitu 94%. Oleh karena itu, hasil ini dapat dijadikan referensi kepada BPJS ketenagakerjaan untuk menentukan atau mempertimbangkan iuran piutang jaminan sosial ketenagakerjaan kepada nasabah diberbagai perusahaan kaitannya dengan iuran jasa sosial

SIMPULAN

Hasil akurasi yang didapatkan pada penelitian ini yaitu 92% dengan nilai K Optimal yaitu 6 dan nilai ROC 94%. Hasil ini merupakan hasil peningkatan setelah dilakukan penambahan cross validation. Kemudian penelitian ini juga mendapatkan hasil yang lebih tinggi dibandingkan dengan penelitian lainnya. Hasil yang sebenarnya masih bisa ditingkatkan lagi dengan menggunakan splitting yang berbeda seperti 90:10, 80:20, dan 60:40. Dengan didapakkannya hasil dari splitting data tersebut dapat dibandingkan splitting data mana yang terbaik. Kemudian untuk meningkatkan akurasi juga bisa menggunakan feature selection. Selain itu bisa juga dibandingkan dengan menggunakan perbandingan algoritma seperti K-Means, C.45 dan lain sebagainya. Penelitian kedepannya masih banyak yang bisa dilakukan dengan menggunakan data yang sama, bisa mengikuti saran yang diberikan ataupun dengan cara yang diinginkan oleh peneliti selanjutnya.

REFERENSI

- Anam, M. K., Mahendra, M. I., Agustin, W., Rahmaddeni, & Nurjayadi. (2022). Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA). *Intensif*, 6(1), 2549–6824. <https://doi.org/10.29407/intensif.v6i1.15870>
- Anam, M. K., Pikir, B. N., Firdaus, M. B., Erlinda, S., & Agustin. (2021). Penerapan Naïve Bayes Classifier , K-Nearest Neighbor dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah. *Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, 21(1), 139–150. <https://doi.org/10.30812/matrik.v21i1.1092>
- Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 147–156. <https://doi.org/10.29408/edumatic.v5i2.3424>
- Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81–86. <https://doi.org/10.33096/ilkom.v12i2.507.81-86>
- Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269–274. <https://doi.org/10.33096/ilkom.v11i3.489.269-274>
- Fansyuri, M. (2020). Analisa algoritma klasifikasi k-nearest neighbor dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus pt. Trigatra komunikatama). *Humanika: Jurnal Ilmu Sosial, Pendidikan, Dan Humaniora*, 3(1), 29–33. <https://doi.org/https://doi.org/10.33050/tmj.v6i1.1531>
- Frastian, N. (2018). Implementasi Komparasi Algoritma Klasifikasi Menentukan Kelulusan Mata Kuliah Algoritma Universitas Budi Luhur. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 3(1), 1–8. <https://doi.org/10.30998/string.v3i1.2334>
- Hidayat, W., Utami, E., Iskandar, A. F., Hartanto, A. D., & Prasetyo, A. B. (2021). Perbandingan Performansi Model pada Algoritma K-NN terhadap Klasifikasi Berita Fakta Hoaks Tentang Covid-19. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 167–176. <https://doi.org/10.29408/edumatic.v5i2.3664>
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: ANDI.
- Mustafa, M. S., & Simpen, I. W. (2019). Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba. *Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 8(1), 1–10.
- Mustafa, M. S., & Simpen, I. W. (2014). Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar). *Creative Information Technology Journal*, 1(4), 270–281. <https://doi.org/10.24076/citec.2014v1i4.27>

- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Pikir, B. N., Anam, M. K., Asnal, H., Rahmaddeni, & Fitri, T. A. (2021). Sentiment Analysis of Technology Utilization by Pekanbaru City Government Based on Community Interaction in Social Media. *JAIA – Journal Of Artificial Intelligence And Applications*, 2(1), 32–40.
- Prasetio, A., Bijaksana, M. A., & Suryani, A. A. (2020). Name Disambiguation Analysis Using the Word Sense Disambiguation Method in Hadith. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 68-74. <https://doi.org/10.29408/edumatic.v4i2.2551>
- Rhomadhona, H., & Permadi, J. (2019). Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation. *Jurnal Sains Dan Informatika*, 5(2), 108–117. <https://doi.org/10.34128/jsi.v5i2.177>
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *Petir*, 12(2), 131–144. <https://doi.org/10.33322/petir.v12i2.471>
- Samponu, Y. B., & Kusriani, K. (2018). Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu. *Jurnal ELTIKOM*, 1(2), 56–63. <https://doi.org/10.31961/eltikom.v1i2.29>
- Sari, A. Q., Sukestiyarno, Y. L., & Agoestanto, A. (2017). Batasan Prasyarat Uji Normalitas Dan Uji Homogenitas Pada Model Regresi Linear. *Unnes Journal of Mathematics*, 6(2), 168–177. <https://doi.org/10.15294/ujm.v6i2.11887>
- Ulfah, A. N., & Anam, M. K. (2020). Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM). *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(1), 1–10. <https://doi.org/10.35957/jatisi.v7i1.196>
- Wang, T., Ke, H., Zheng, X., Wang, K., Sangaiah, A. K., & Liu, A. (2019). Big data cleaning based on mobile edge computing in industrial sensor-cloud. *IEEE Transactions on Industrial Informatics*, 16(2), 1321-1329. <https://doi.org/10.1109/TII.2019.2938861>
- Widiastuti, I. (2017). Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Jawa Barat. *Public Inspiration: Jurnal Administrasi Publik*, 2(2), 91–101. <https://doi.org/10.22225/pi.2.2.2017.91-101>