

## Implementasi Algoritma *Principal Component Analysis* (PCA) dan *K-Nearest Neighbor* (KNN) Untuk Memprediksi Kelayakan Kredit Pengguna *Smartphone* Di Indonesia Pada Masa Pandemi Covid-19

Rara Kartika Kusuma Winahyu<sup>1</sup>, Vian Ardiyansyah Saputro<sup>2</sup>

<sup>1,2</sup> Politeknik Astra, Jawa Barat, Indonesia

Corresponding : [raden.rara@polytechnic.astra.ac.id](mailto:raden.rara@polytechnic.astra.ac.id), [vian.saputro@polytechnic.astra.ac.id](mailto:vian.saputro@polytechnic.astra.ac.id)

### Abstrak

Dalam penelitian ini, kami bertujuan untuk memprediksi kelayakan kredit pengguna *smartphone* di Indonesia selama masa Pandemi COVID-19 menggunakan algoritma *machine learning*. Algoritma *Principal Component Analysis* (PCA) dan *K-means* digunakan untuk mereduksi dimensi dataset dan mengelompokkan kelayakan kredit dari sejumlah 1050 *data set* responden yang terdiri dari dua belas pertanyaan kepada pengguna *smartphone* di Indonesia selama pandemi COVID-19. Algoritma klasifikasi *K-Nearest Neighbor* (KNN) juga digunakan untuk mengklasifikasikan kelayakan kredit pengguna *smartphone* di Indonesia. Pengujian yang dilakukan meliputi pengujian akurasi, presisi, *recall*, dan F1-score. Hasil dari penelitian ini menunjukkan bahwa algoritma klasifikasi *K-Nearest Neighbor* memiliki akurasi sebesar 0.85, presisi sebesar 0.87, *recall* 0.84, dan F1 score sebesar 0.85 dalam mengklasifikasikan kelayakan kredit pengguna *smartphone* di Indonesia pada masa pandemi COVID-19.

**Kata kunci:** kelayakan kredit, *smartphone*, pandemi, *machine learning*

### Abstract

*In this research work, we aimed to predict the credit worthiness of smartphone users in Indonesia during the COVID-19 pandemic using machine learning algorithms. Principal Component Analysis (PCA) and K-means algorithms were used to reduce dataset dimensions and classify credit worthiness from a total of 1050 respondent dataset consisting of twelve questions to smartphone users in Indonesia during the COVID-19 pandemic. The K-Nearest Neighbor (KNN) classification algorithm was also used to classify the credit worthiness of smartphone users in Indonesia. The tests carried out included testing for accuracy, precision, recall, and F1-score. The results of this study indicated that the KNN classification algorithm has an accuracy of 0.85, a precision of 0.87, a recall of 0.84, and an F1 score of 0.85 in classifying the creditworthiness of smartphone users in Indonesia during the COVID-19 pandemic.*

**Keywords:** creditworthiness, *smartphone*, pandemic, *machine learning*

#### History:

Received : 16- Mei 2023

Revised : 07 Juni 2023

Accepted : 11 Juni 2023

Published: 12 Juni 2023

**Publisher:** LPPM Universitas Darma Agung

**Licensed:** This work is licensed under

[Attribution-NonCommercial-No](https://creativecommons.org/licenses/by-nc-nd/4.0/)

[Derivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)



## PENDAHULUAN

Bank adalah lembaga yang paling dikenal oleh kebanyakan orang dewasa ini. Kredit merupakan salah satu produk perbankan yang membantu masyarakat dalam memenuhi kebutuhan sehari-hari, namun masih ada sebagian masyarakat yang tidak dapat memperoleh kredit dikarenakan cara tradisional yang masih digunakan oleh lembaga keuangan khususnya perbankan (Alfat et al. 2019) (Shema 2019). Berdasarkan riset, pengguna *smartphone* di Indonesia dapat mencapai 269 juta orang pada tahun 2026 . Penggunaan data pengguna *smartphone* dapat digunakan untuk menilai skor kredit pengguna, mengurangi dokumen, mengurangi waktu pengiriman pinjaman, dan meningkatkan pemilihan pengguna yang memenuhi syarat (Martínez Sánchez and Pérez Lechuga 2016). Juga, pengeluaran orang dan kondisi keuangan mereka mungkin memiliki hubungan dengan penggunaan data di ponsel cerdas mereka (Alfat et al. 2019). Kemampuan membayar dan kemauan membayar merupakan variabel yang mempengaruhi kelayakan kredit. Faktor keuangan berkaitan dengan kemampuan membayar sedangkan faktor personal berkaitan dengan kemauan membayar (Kousayri 2020). Namun terjadi perubahan kondisi keuangan yang dialami oleh masyarakat Indonesia yang juga pengguna *smartphone* di masa pandemi COVID-19 karena menurut (Muhyiddin and Nugroho 2021), Indonesia mengalami krisis ekonomi dimana beberapa orang kehilangan pekerjaan dan lebih dari 80% perusahaan melihat penurunan yang signifikan dalam keuntungan. Produksi turun dan jalur distribusi/logistik terganggu akibat pembatasan mobilitas Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM).

Kondisi krisis ekonomi, ditambah dengan pembatasan mobilitas selama pelaksanaan Komunitas Pembatasan Kegiatan (PPKM), mengakibatkan penurunan pendapatan masyarakat dan minat belanja yang menyebabkan penurunan dalam ekspansi kredit dan menunjukkan likuiditas yang cukup di sektor perbankan (Muhyiddin and Nugroho 2021). Selain itu, bank dan lembaga keuangan lainnya di Indonesia juga mengalami peningkatan tingkat gagal bayar pinjaman akibat krisis ekonomi di masa pandemi (Rizwan, Ahmad, and Ashraf 2020).

Kontribusi dari makalah ini adalah mengelompokkan dan mengklasifikasikan kelayakan kredit pengguna *smartphone* di Indonesia berdasarkan dua belas variabel menggunakan algoritma *machine learning* yang dapat membantu lembaga keuangan di Indonesia, khususnya perbankan dalam menilai kelayakan kredit calon pelanggan selama masa pandemi.

Dua belas variabel kelayakan kredit yang paling berpengaruh bagi pengguna *smartphone* di Indonesia adalah produsen/ merk *smartphone*, frekuensi mengganti *smartphone*, pekerjaan, penggunaan data untuk game, penggunaan data untuk media sosial, biaya internet, alasan mengganti *smartphone*, pengeluaran uang untuk data, menjaga keinginan pribadi , usia, menghabiskan sisa uang, dan bayar sendiri (Alfat et al. 2019).

Setelah melakukan pengumpulan dataset pengguna *smartphone* di Indonesia pada masa pandemi COVID-19, Langkah selanjutnya yaitu menentukan metode untuk memprediksi kelayakan kredit pengguna *smartphone* di Indonesia pada masa

pandemi COVID-19. Hasil nya, algoritma *Principal Component Analysis* (PCA) dan algoritma K-means dipilih dalam penelitian ini. Algoritma PCA digunakan untuk mereduksi dimensi data untuk mendapatkan nilai K terbaik, yaitu pusat *cluster* awal dan jarak dengan beberapa fitur bobot (Mousavian Anaraki, Haeri, and Moslehi 2021). Sementara itu, algoritma *clustering* K-means digunakan untuk mengekstraksi kelompok secara efisien dari sekumpulan dataset yang besar. Proses *clustering* mencerminkan sejauh mana elemen data dan klien tertentu dihubungkan (Kavitha 2016). Setelah proses *clustering*, algoritma klasifikasi *machine learning* (ML) KNN digunakan untuk mengklasifikasikan pengguna *smartphone* berdasarkan *cluster* mereka. Alasan menggunakan algoritma KNN karena merupakan metode penilaian kredit yang sering digunakan (Mukid et al. 2018).

## METODE PENELITIAN

### a. Kuesioner

Kuesioner dirancang agar para peneliti dapat mengekstraksi informasi yang berguna. Metode pertanyaan tertutup digunakan untuk memungkinkan para peneliti memilih apa yang mereka yakini untuk mengungkapkan pendapat mereka. Metode Likert digunakan untuk menentukan lima tingkat jawaban.. Dari (1) hingga (5), rata-rata aritmetika berkisar dari (1,5 hingga 2,5) menunjukkan persetujuan rendah, bidang dari (2,5 hingga 3,5) menunjukkan persetujuan sedang, dan rentang dari (3,5 hingga 5) menunjukkan persetujuan tinggi [5].

Dua belas pertanyaan yang diajukan dalam kuesioner adalah diklasifikasikan ke dalam tiga kategori berbeda: latar belakang (2 pertanyaan), parameter penggunaan *smartphone* (8 pertanyaan), kepercayaan (2 pertanyaan) (Alfat et al. 2019) seperti yang ditunjukkan pada Tabel 1. Implementasi kedua belas variabel tersebut diwujudkan dalam bentuk kuesioner, pertanyaan ditunjukkan pada Tabel 2.

Tabel 1 Kedua belas variable dan kategorinya

Kategori	Variabel
Latar belakang	Usia
	Pekerjaan
Kepercayaan	Menghabiskan sisa uang
	Menjaga aset
	Produsen/ merk <i>smartphone</i>
	Penggunaan data internet untuk permainan ( <i>game</i> )
Penggunaan	Penggunaan data internet untuk media sosial
	Pengeluaran biaya untuk <i>smartphone</i>
	Pengeluaran biaya internet
	Frekuensi mengganti <i>smartphone</i>
	Alasan mengganti <i>smartphone</i>
	Membayar mandiri

Tabel 2. Kedua belas Variabel dan Implementasinya dalam bentuk pertanyaan kuesioner

Variabel	Pertanyaan Kuesioner
Usia	Berapakah usia Anda?
Pekerjaan	Apa pekerjaan Anda?
Menghabiskan sisa uang	Jika anda memiliki sisa uang sebesar Rp 100.000 (seratus ribu rupiah), maka Anda lebih memilih untuk menabungnya daripada menghabiskannya.
Menjaga aset	Saya lebih memilih menabung uang saya dari pada menghabiskannya
Produsen/ merk <i>smartphone</i>	Apa merk <i>smartphone</i> Anda?
Penggunaan data internet untuk permainan ( <i>game</i> )	Berapa penggunaan data untuk <i>game</i> yang Anda keluarkan dalam satu bulan?
Penggunaan data internet untuk media sosial	Berapa besar penggunaan data media sosial (facebook, twitter, instagram, tiktok) Anda dalam satu bulan?
Pengeluaran biaya untuk <i>smartphone</i>	Berapa biaya yang Anda keluarkan untuk <i>smartphone</i> (pulsa, <i>game</i> , dll) dalam satu bulan?
Pengeluaran biaya internet	Berapa biaya yang anda dikeluarkan untuk membeli paket data internet dalam satu bulan?
Frekuensi mengganti <i>smartphone</i>	Berapa kali Anda mengganti <i>smartphone</i> dalam jangka waktu satu tahun terakhir?
Alasan mengganti <i>smartphone</i>	Apa alasan Anda mengganti <i>smartphone</i> terakhir kali?
Membayar mandiri	Siapa yang membayar biaya penggunaan <i>smartphone</i> (pulsa) Anda?

## b. Responden

Survei dilakukan pada Juli-Agustus 2021 selama Pandemi COVID-19 dan diikuti oleh 1050 responden yang merupakan pengguna *smartphone* di Indonesia dari berbagai usia dan ekonomi dan latar belakang. Rentang usia responden antara 10 tahun hingga lebih dari 60 tahun. Pekerjaan responden terdiri dari 21% adalah karyawan swasta, 11% adalah PNS, 7% karyawan perusahaan BUMN, 6% adalah guru/dosen, 10% adalah pengusaha, 16% mahasiswa, dan 2% petani. Sisanya, sekitar 27% bekerja sebagai pekerja lepas dan profesi lainnya.

Jumlah responden yang terlibat dalam survei ini dengan memperhitungkan jumlah penduduk Indonesia pada tahun 2021 yaitu 272 juta jiwa [20]. Dengan memilih tingkat kepercayaan 99% dan margin of error 5% maka jumlah responden minimal adalah 663 orang [21]. Responden diberi tautan ke kuesioner online. Kemudian, berdasarkan penggunaan *smartphone* mereka, para responden secara sukarela menjawab kuesioner tersebut.

## c. Implementasi Algoritma PCA dan Kmeans

Proses penghilangan *outlier* diterapkan sebelum proses *Principal Component Analysis* (PCA) karena PCA mudah terpengaruh dengan adanya *outlier* (Jolliffe and Cadima 2016). Dari korelasi heatmap antar variabel yang ditunjukkan pada Gambar 1, dapat disimpulkan bahwa hanya 9 dari 12 variabel yang memiliki korelasi antar variabel. Tiga variabel yang tidak berpengaruh signifikan terhadap kondisi pengguna *smartphone* di Indonesia pada masa pandemi adalah frekuensi pergantian *smartphone*, alasan berganti *smartphone*, dan bayar mandiri. Hasil tersebut semakin

meyakinkan penulis bahwa kondisi keuangan pengguna *smartphone* di Indonesia telah berubah selama krisis ekonomi di masa pandemi.

Gambar. 1 menunjukkan diagram heatmap korelasi antar variabel



## HASIL Dan PEMBAHASAN

Setelah menghitung dataset yang terdiri dari dua belas pertanyaan yang disebutkan di atas dengan algoritma PCA dan K-means, tiga kluster kelayakan kredit pengguna *smartphone* di Indonesia dihasilkan selama pandemi COVID-19 seperti yang ditunjukkan pada Gambar 2 dan Tabel 3 di bawah ini. Selanjutnya, dataset baru dibagi menjadi dua bagian yaitu: pelatihan dan pengujian, dengan pembagian 80%:20%. Seperangkat pelatihan yang terdiri dari delapan puluh persen sampel yang dipilih secara acak dari kumpulan data digunakan untuk membuat model prediktif. Dua puluh persen sampel yang tersisa digunakan sebagai sampel buta untuk menilai kekokohan model prediktif.

Klaster 1 terdiri dari pengguna *smartphone* yang menghabiskan uang paling banyak untuk *smartphone* mereka dibandingkan dengan yang lain, tetapi kecenderungan mereka untuk menyimpan dan menjaga aset mereka tinggi. Sementara itu, klaster 2 terdiri dari pengguna *smartphone* yang paling sedikit membelanjakan uangnya untuk *smartphone* mereka, tetapi kecenderungan mereka untuk menyimpan dan menjaga asetnya tinggi. Klaster terakhir adalah klaster 3 yang terdiri dari pengguna *smartphone* yang membelanjakan uangnya cukup banyak untuk *smartphone*-nya, namun kecenderungan mereka untuk menyimpan dan menjaga asetnya paling rendah dibandingkan yang lain. Dari total 1050 responden diperoleh hasil sebagai berikut. Sebanyak 358 responden masuk dalam klaster 1. Sedangkan sebanyak 389 responden masuk dalam klaster 2 dan sisanya sebanyak 303 responden masuk dalam klaster 3.

Algoritma *machine learning* klasifikasi KNN yang diuji dalam penelitian ini dapat membangun model pembelajaran mesin yang baik dengan hasil uji akurasi, presisi, *recall*, dan skor F1 lebih dari 80% seperti yang ditunjukkan pada Tabel 4.

Gambar 2. Visualisasi kluster dataset pengguna *smartphone* di Indonesia dengan menggunakan algoritma PCA



Tabel 3. Hasil kluster berdasarkan variabel

Variabel	Klaster 1	Klaster 2	Klaster 3
Usia	Medium	Tua	Muda

Pengeluaran biaya untuk <i>smartphone</i>	Tertinggi	Terendah	Sedang
Pengeluaran biaya internet	Tertinggi	Terendah	Sedang
Penggunaan data internet untuk media sosial	Tertinggi	Terendah	Sedang
Menjaga aset	Tinggi	Tinggi	Sedang
Menghabiskan sisa uang	Tinggi	Tinggi	Rendah

Tabel 4. Hasil Uji Performa Algoritma KNN dalam memprediksi kelayakan pengguna *smartphone* pada masa pandemic COVID-19

Algoritma	Performa			
	Akurasi	Presisi	Recall	F1-score
KNN	0.85	0.87	0.84	0.85

## Pembahasan

### **Credit Scoring menggunakan data penggunaan *smartphone***

Sebelum adanya *credit scoring*, bank umum sering menilai risiko kredit pemohon pinjaman dengan sangat subyektif yaitu mengandalkan petugas kontrol risiko (Wang et al. 2020). Merujuk lima karakteristik model kredit (5C), lima kriteria harus diperiksa saat menetapkan kelayakan kredit peminjam. 5 C tersebut adalah Kapasitas (Capacity), Modal (Capital), Karakter (Character), Kondisi (Condition), dan Agunan (Collateral) yang merupakan sumber pembayaran alternatif tersebut tidak wajib menjadi prasyarat untuk persetujuan kredit.

Sayangnya, proses standar tradisional ini memiliki beberapa kelemahan seperti:

- 1) Keputusan kredit dipengaruhi oleh mood pengelola bank yang bisa berbeda-beda setiap harinya,
- 2) Keputusan kredit tidak dapat diganggu gugat karena keputusan manajer,
- 3) Tidak ada keputusan baku untuk diturunkan, serta
- 4) Dalam beberapa keadaan, kerugian dapat diperoleh dari persepsi manusia yang digunakan sebagai evaluasi (Alfat et al. 2019).

Dewasa kini, pengumpulan data dari penyedia telekomunikasi, platform media sosial, atau pengambilan data secara langsung dari handset seluler melalui aplikasi seluler dilakukan oleh beberapa perusahaan dan lembaga pemberi pinjaman yang menargetkan konsumen yang sebelumnya tidak memiliki rekening bank (Shema 2019), karena sumber informasi alternatif untuk *credit scoring* dapat diperoleh dari pengalaman dalam menggunakan *smartphone* (Ots, Liiv, and Tur 2020).

Sebuah studi yang dilakukan oleh Daniel Björkegren dan Darrell Grissen berhasil menemukan solusi atas masalah yang dihadapi penduduk negara berkembang di Amerika Selatan, khususnya terkait kesulitan dalam memperoleh persetujuan kredit dan perpanjangan kredit dari lembaga keuangan karena kurangnya sejarah keuangan formal. Sumber informasi alternatif, seperti riwayat penggunaan ponsel dapat memprediksi kelayakan kredit pada individu yang sebelumnya tidak memiliki rekening bank. Menurut penelitian tersebut, orang-orang di kuintil teratas hampir tiga kali lebih mungkin mengalami gagal bayar dibandingkan mereka yang berada di kuintil terbawah (Alfat et al. 2019).

### **Algoritma PCA**

Di bidang pengurangan dimensi kumpulan data besar, salah satu metode yang sering digunakan adalah PCA. PCA juga merupakan salah satu algoritma yang dapat mengkarakterisasi subruang. Transformasi ortonormal pada PCA menghasilkan variable linear independen dari kumpulan data yang memiliki potensi untuk dikorelasikan (Migenda, Möller, and Schenck 2019). PCA juga merupakan metode statistik yang menggantikan banyak elemen asli dengan beberapa elemen penting .

### **Algoritma K-means**

Di bidang pengurangan dimensi kumpulan data besar, salah satu metode yang sering digunakan adalah PCA. PCA juga merupakan salah satu algoritma yang dapat mengkarakterisasi subruang. Transformasi ortonormal pada PCA menghasilkan variable linear independen dari kumpulan data yang memiliki potensi untuk dikorelasikan (Migenda et al. 2019). PCA juga merupakan metode statistik yang menggantikan banyak elemen asli dengan beberapa elemen penting.

Langkah 1: Tentukan pusat kluster pertama dengan memilih jumlah cluster  $k$  dari dataset  $n$  objek data.

Langkah 2: Hitung seberapa jauh setiap objek data dalam dataset selain  $i$ , di mana  $i=0,1,2,\dots,k-1$  dan semua  $k$  pusat cluster  $c$  dengan  $j$ , di mana  $j=0,1,2,\dots,k-1$ . Alokasikan jarak setiap objek data dalam dataset dengan  $i$  ke cluster yang paling dekat dengannya.

Langkah 3: Evaluasi kembali nilai setiap pusat cluster  $j$ .

Langkah 4: Jika pusat kluster masih tidak berubah, ulangi Langkah 2 dan Langkah 3.

### **Algoritma K-Nearest Neighbor (KNN)**

KNN adalah algoritma *lazy learning* non-parametrik. Algoritma ini adalah teknik non-parametrik yaitu tidak membuat asumsi tentang distribusi data yang mendasarinya. Algoritma ini sangat berguna, karena di dunia nyata hampir semua data praktis tidak mengikuti asumsi teoritis yang dibuat. Sebagian besar algoritma *lazy learning* - terutama KNN - membuat keputusan berdasarkan seluruh dataset pelatihan (Mukid et al. 2018). KNN merupakan salah satu algoritma klasifikasi yang paling sederhana, yang sering digunakan sebagai tolak ukur untuk pengklasifikasi yang lebih kompleks. Fix dan Hodges memperkenalkan metode klasifikasi pola yang kemudian dikenal dengan aturan k-nearest neighbor (Beckmann, Ebecken, and Pires de Lima 2015). Klasifikasi K-nearest neighbor biasanya didasarkan pada jarak Euclidean antara sampel uji dan sampel yang ditentukan sampel pelatihan. Ide dasar dari algoritma K-NN adalah setiap kali titik prediksi baru muncul, k tetangga terdekatnya dipilih dari data pelatihan. Kemudian prediksi titik baru dapat berupa rata-rata nilai k tetangga terdekatnya (Zhang and Wang 2016).

#### **a. Penelitian yang Berkaitan**

Penelitian yang berkembang tentang keakuratan data *smartphone* dalam memprediksi kelayakan kredit pengguna telah menemukan pola yang tepat bahwa dari aktivitas keuangan *real-time* yang tercermin dalam data *smartphone* pintar dapat membantu meningkatkan pengetahuan kelayakan kredit dari pengguna *smartphone*

(Shema 2019). Contoh kasus mobiScore yang berpotensi mencetak dan memberikan kredit kepada jutaan orang yang saat ini ditolak (Pedro, Proserpio, and Oliver 2015). Di sisi lain, dengan memadukan catatan demografis dan riwayat detail panggilan (CDR) dengan catatan keuangan dari lembaga keuangan negara yang sama, sebuah organisasi kredit mikro di Negara Karibia mampu menurunkan gagal bayar hampir setengah dari total kredit dengan tetap mempertahankan tiga perempat debiturnya (Shema 2019). Sementara penelitian yang dilakukan oleh (Ruiz et al. 2017) menggunakan CDR dan data yang diperoleh dari *smartphone* konsumen digabungkan dengan data sistem penilaian kredit tradisional yang menghasilkan pemilihan prosedur penyaringan aplikasi pinjaman, meningkatkan tingkat persetujuan, dan menurunkan tingkat keterlambatan.

Di sisi lain, faktor jaringan sosial antar jaringan individu seperti usia, jenis kelamin, etnis, bahasa, faktor ekonomi, geografi, urbanisasi, dan epidemi dapat dipelajari melalui penilaian penilaian kredit menggunakan data CDR (Savero, Nasir, and Safitri 2017). Selanjutnya, penelitian tentang model *credit scoring* keuangan mikro berdasarkan informasi jaringan media sosial yang diambil dari akun Facebook yang dilakukan oleh De Cnudde et al. menemukan bahwa jaringan eksplisit orang yang berinteraksi lebih prediktif daripada jaringan eksplisit orang yang tidak, tetapi jaringan orang yang menunjukkan perilaku serupa mengungguli kedua jaringan eksplisit (Hazra 2017).

## **SIMPULAN**

Hasil penelitian ini menyajikan prediksi kelayakan kredit pengguna *smartphone* di Indonesia selama pandemi COVID-19 menggunakan algoritma *machine learning* KNN. Sebanyak 1050 responden berpartisipasi menjawab 12 pertanyaan yang terdiri dari tiga kategori kunci. Merujuk pada perhitungan PCA dan K-means, penelitian kami menyimpulkan bahwa ada tiga kluster pengguna *smartphone* di Indonesia selama ini krisis ekonomi di masa pandemi. Ketiga kluster yang dihasilkan diharapkan dapat membantu lembaga keuangan di Indonesia, khususnya perbankan untuk memprediksi kelayakan kredit calon nasabah sehingga bank dapat mengurangi risiko gagal pinjaman selama pandemi di masa yang akan datang

Penelitian ini juga menghasilkan empat uji coba performa algoritma KNN yang digunakan, yaitu uji akurasi, presisi, recall, dan F1-score. Hasil untuk uji akurasi yaitu 0.85, uji presisi sebesar 0.87, uji recall sebesar 0.84, uji F1-score sebesar 0.85. Di masa depan, penelitian untuk mengembangkan penilaian kelayakan kredit dari calon nasabah lembaga perbankan yang juga pengguna *smartphone* berdasarkan kinerja algoritma *machine learning* dan fitur-fitur penting yang berubah sebagai akibat pengaruh wilayah geografis dan kondisi ekonomi suatu negara merupakan hal yang menarik bagi topik penelitian untuk pembahasan lebih lanjut.

## **DAFTAR PUSTAKA**

Alfat, Lathifah, Mia Rizkinia, Riri Fitri Sari, and Daniela M. Romano. 2019. "Feature Selection of Credit Score Factor Based on Smartphone Usage Using MCFS." Pp.

1–5 in *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*. IEEE.

- Beckmann, Marcelo, Nelson F. F. Ebecken, and Beatriz S. L. Pires de Lima. 2015. "A KNN Undersampling Approach for Data Balancing." *Journal of Intelligent Learning Systems and Applications* 7(04):104–16.
- Hazra, Avijit. 2017. "Using the Confidence Interval Confidently." *Journal of Thoracic Disease* 9(10):4125.
- Jolliffe, Ian T., and Jorge Cadima. 2016. "Principal Component Analysis: A Review and Recent Developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065):20150202.
- Kavitha, K. 2016. "Clustering Loan Applicants Based on Risk Percentage Using K-Means Clustering Techniques." *International Journal of Advanced Research in Computer Science and Software Engineering* 6(2):162–66.
- Kousayri, Husain. 2020. "Creditworthiness Standards between Theory and Practice (a Field Study on the Syrian Commercial Bank in Lattakia)." *Available at SSRN* 3591460.
- Martínez Sánchez, José Francisco, and Gilberto Pérez Lechuga. 2016. "Assessment of a Credit Scoring System for Popular Bank Savings and Credit." *Contaduría y Administración* 61(2):391–417.
- Migenda, Nico, Ralf Möller, and Wolfram Schenck. 2019. "Adaptive Dimensionality Adjustment for Online 'Principal Component Analysis.'" Pp. 76–84 in *Intelligent Data Engineering and Automated Learning–IDEAL 2019: 20th International Conference, Manchester, UK, November 14–16, 2019, Proceedings, Part I* 20. Springer.
- Mousavian Anaraki, Seyed Alireza, Abdorrahman Haeri, and Fateme Moslehi. 2021. "A Hybrid Reciprocal Model of PCA and K-Means with an Innovative Approach of Considering Sub-Datasets for the Improvement of K-Means Initialization and Step-by-Step Labeling to Create Clusters with High Interpretability." *Pattern Analysis and Applications* 24(3):1387–1402.
- Muhyiddin, Muhyiddin, and Hanan Nugroho. 2021. "A Year of Covid-19: A Long Road to Recovery and Acceleration of Indonesia's Development." *Jurnal Perencanaan Pembangunan: The Indonesian Journal of Development Planning* 5(1):1–19.
- Mukid, M. A., T. Widiharih, A. Rusgiyono, and A. Prahutama. 2018. "Credit Scoring Analysis Using Weighted k Nearest Neighbor." P. 12114 in *Journal of Physics: Conference Series*. Vol. 1025. IOP Publishing.
- Ots, Henri, Innar Liiv, and Diana Tur. 2020. "Mobile Phone Usage Data for Credit Scoring." Pp. 82–95 in *Databases and Information Systems: 14th International Baltic*

*Conference, DB&IS 2020, Tallinn, Estonia, June 16–19, 2020, Proceedings 14.*  
Springer.

- Pedro, Jose San, Davide Proserpio, and Nuria Oliver. 2015. "MobiScore: Towards Universal Credit Scoring from Mobile Phone Data." Pp. 195–207 in *User Modeling, Adaptation and Personalization: 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29--July 3, 2015. Proceedings 23.* Springer.
- Rizwan, Muhammad Suhail, Ghufran Ahmad, and Dawood Ashraf. 2020. "Systemic Risk: The Impact of COVID-19." *Finance Research Letters* 36:101682.
- Ruiz, Saulo, Pedro Gomes, Luís Rodrigues, and João Gama. 2017. "Credit Scoring in Microfinance Using Non-Traditional Data." Pp. 447–58 in *Progress in Artificial Intelligence: 18th EPIA Conference on Artificial Intelligence, EPIA 2017, Porto, Portugal, September 5-8, 2017, Proceedings 18.* Springer.
- Savero, Daniel Oscar, Azwir Nasir, and Devi Safitri. 2017. "Pengaruh Komisaris Independen, Komite Audit, Kepemilikan Instutisional Dan Kepemilikan Manajerial Terhadap Integritas Laporan Keuangan (Studi Empiris Pada Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia Tahun 2012-2014)." *Jurnal Online Mahasiswa (JOM) Bidang Ilmu Ekonomi* 4(1):75–89.
- Shema, Alain. 2019. "Effective Credit Scoring Using Limited Mobile Phone Data." Pp. 1–11 in *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development.*
- Wang, Yuelin, Yihan Zhang, Yan Lu, and Xinran Yu. 2020. "A Comparative Assessment of Credit Risk Model Based on Machine Learning — — a Case Study of Bank Loan Data." *Procedia Computer Science* 174:141–49.
- Zhang, Yao, and Jianxue Wang. 2016. "K-Nearest Neighbors and a Kernel Density Estimator for GEFCom2014 Probabilistic Wind Power Forecasting." *International Journal of Forecasting* 32(3):1074–80.