

ANALISIS KELAYAKAN KREDIT PADA *PEER TO PEER LENDING* MENGUNAKAN DATA SOSIAL MEDIA (STUDI KASUS: LINKEDIN)

CREDIT WORTHINESS ANALYSIS IN PEER TO PEER LENDING USING SOCIAL MEDIA DATA (CASE STUDY: LINKEDIN)

Raynald Damara Eriyanto¹, Andry Alamsyah²

Prodi S1 Manajemen Bisnis Telekomunikasi dan Informatika, Fakultas Ekonomi Bisnis, Universitas Telkom
¹raynaldderiyanto@student.telkomuniversity.ac.id, ²andrya@telkomuniversity.ac.id

Abstrak

Credit scoring sangat membantu pemberi pinjaman dalam mengevaluasi kelayakan kredit calon peminjam. Ketersediaan data yang tidak mencukupi, memberikan tantangan kepada perusahaan *P2P Lending* untuk memanfaatkan data alternatif sebagai pengganti atau pelengkap indikator pemodelan penilaian kredit. *P2P Lending* menggunakan data sosial media untuk mengganti kurangnya data tersebut. Akan tetapi, data sosial media yang heterogen sangat susah digunakan untuk mengidentifikasi kelayakan kredit seseorang. Dengan adanya permasalahan tersebut penelitian ini bertujuan untuk menentukan model penilaian kelayakan kredit seseorang menggunakan data sosial media LinkedIn. Penelitian ini menggunakan proses *data mining* untuk melakukan *prediction analysis* berupa metode *classification* (klasifikasi) yang memanfaatkan algoritma *decision tree*, dan *random forest*. Hasil penelitian ini menunjukkan bahwa kombinasi data *user demographic attributes* dan data *user generated content* menghasilkan model penilaian kredit terbaik dengan nilai *accuracy* sebesar 87.12%, dengan hasil tersebut, *P2P Lending* memiliki kesempatan untuk membuka segmentasi baru konsumen atau nasabah yang tidak memiliki riwayat kredit secara lengkap. Temuan lainnya juga menjelaskan data *user generated content* lebih cocok untuk dijadikan data pelengkap model penilaian kredit.

Kata kunci: Penilaian Kredit, Klasifikasi, *P2P Lending*, Sosial Media

Abstract

Credit scoring is very helpful for lenders in evaluating the credit worthiness. Insufficient data availability, presents a challenge to P2P Lending companies to use alternative data as a substitute or complement to credit scoring modeling indicators. P2P Lending uses social media data to compensate for the lack of data. However, heterogeneous social media data is very difficult to identify the credit worthiness. This study aims to determine the credit worthiness of someone using LinkedIn social media data. This study use data mining methods to perform prediction analysis with classification method using decision tree and random forest algorithm. The results of this study indicate that the combination of user demographic attributes and user generated content data produces the best credit scoring model with an accuracy value of 87.12%, with these results, P2P Lending has the opportunity to open new segments of consumers who do not have a complete credit history. Other findings also explain the data user generated content is more suitable to be used as a complementary data credit scoring model.

Keywords: Credit Scoring, Classification, *P2P Lending*, Social Media

1. Pendahuluan

Menurut Campbell-Verduyn *et al.* (2016) [3] menyatakan bahwa *big data* menembus ke industri jasa keuangan melalui perusahaan jasa kredit dan *fintech* (*financial technology*) yang memanfaatkan *big data* sesuai algoritma yang dibutuhkan. Salah satunya, perusahaan American Express yang menggunakan metode *social media analytics and cloud* yang memanfaatkan sosial media Facebook dan Twitter untuk membuat metode pembayaran secara digital [16]. *Startup fintech* khususnya *P2P (peer to peer) Lending*, merupakan sebuah layanan keuangan individu yang memungkinkan pemberi pinjaman (kreditur/pendana/investor) untuk menawarkan pinjaman atau investasi secara langsung yang dibuat peminjam (debitur)[19]. *P2P Lending* memiliki banyak kelebihan seperti kecepatan pemrosesan pengajuan dana, akan tetapi juga memiliki kekurangan dalam pembuatan penilaian modelnya yang memberikan tantangan tersendiri di era sekarang ini. Ketersediaan data yang tidak mencukupi, memberikan tantangan kepada perusahaan *P2P Lending* untuk memanfaatkan data alternatif sebagai pengganti atau pelengkap indikator pemodelan penilaian kredit [8]. *P2P Lending* memanfaatkan data alternatif yang berasal dari berbagai macam sumber seperti akun sosial media [4]. Akan tetapi, data sosial media yang heterogen seperti unggahan artikel atau konten, koneksi pertemanan, dan sejenisnya tidak mengandung konten terkait kredit secara eksplisit dalam banyak kasus.

Sedikitnya relevansi antara data sosial media dengan konten terkait penilaian kredit, perlu adanya penyesuaian dan pengolahan data yang tepat. Pengolahan data sosial media perlu dilakukan dengan memanfaatkan metode statistika secara maksimal, sebagian literatur penilaian kredit sebelumnya merumuskannya sebagai masalah klasifikasi biner [8]. LinkedIn sebagai sosial media menyediakan berbagai macam data yang heterogen dalam jumlah yang besar, dengan melakukan pengolahan data menggunakan metode yang tepat mampu menjadi

indikator penentu kelayakan kredit, sehingga dapat membantu industri jasa keuangan dalam mengembangkan metode penilaian kelayakan kredit bagi konsumen.

2. Dasar Teori dan Metodologi

2.1 Dasar Teori

2.1.1 Kredit

Berdasarkan Pedoman Akuntansi Perbankan Indonesia (PAPI) tahun 2001 dalam Fahmi (2014) [6] mendefinisikan kredit sebagai penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dan pihak lain yang mewajibkan pihak peminjam (debitur) untuk melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga, imbalan atau pembagian hasil keuntungan.

2.1.2 Credit Scoring (Penilaian Kredit)

Setiap lembaga pembiayaan memiliki bentuk dan sistem analisis kredit yang berbeda-beda, namun setiap lembaga pembiayaan memiliki alasan yang sama dalam melakukan analisis kredit secara garis besar sebagai bentuk evaluasi dalam memutuskan pemberian atau peminjaman uang atau dana kepada pihak peminjam (Ikatan Bankir Indonesia, 2015) [10]. Penilaian kredit memiliki prinsip 5C yang meliputi *character*, *capacity*, *capital*, *condition*, dan *collateral* [6].

2.1.3 Sosial Media

Menurut Mulawarman dan Nurfitri (2017) [15] menjelaskan sosial media sebagai alat komunikasi yang digunakan oleh manusia untuk melakukan interaksi sosial. Pendapat tersebut diperjelas oleh (Shibuya, 2020) [17] yang menyatakan sosial media sebagai *platform online* yang dirancang untuk mendorong orang berbagi apa yang mereka pikirkan, rasakan, situasi yang dihadapi, dan berbagai informasi lainnya dengan topik tak terbatas.

2.1.4 Big Data

Big data merupakan kumpulan data yang memiliki *volume* (jumlah), *velocity* (kecepatan), dan *variety* (variasi) yang sangat besar sehingga sulit untuk mengolah dan menganalisa menggunakan cara tradisional. *Big data* memiliki karakteristik utama, yaitu *volume*, *velocity*, *variety*, *veracity*, dan *value* (Bahga & Medisetti, 2019) [1].

2.1.5 Data Mining

Data mining merupakan sebuah proses penemuan pola-pola baru dari kumpulan data yang sangat besar yang disimpan dalam suatu repositori, menggunakan Teknik pengenalan pola, dan Teknik statistika maupun matematika (Larose, 2005) [13].

2.1.6 Prediction Analysis

Prediction analysis didefinisikan sebagai model *data mining* yang merujuk proses membangun model representatif yang sesuai dengan data pengamatan, dimana model tersebut memiliki 2 tujuan yaitu memprediksi *output* berdasarkan variabel *input* dan tujuan lainnya digunakan untuk memahami hubungan antara variabel *output* dan semua variabel *input* (Kotu & Despande, 2015) [12].

2.1.7 Classification

Classification didefinisikan sebagai metode terapan yang paling banyak digunakan untuk menentukan keputusan atau ramalan berdasarkan informasi yang tersedia dari data yang dikelompokkan berdasarkan karakteristik yang memiliki kesamaan (Keramati & Yousefi, 2011) [11].

2.1.8 Decision Tree

Decision tree merupakan salah satu metode klasifikasi paling populer dan masih banyak digunakan. Metode ini juga memodelkan prediksi menggunakan struktur pohon atau struktur hierarki. *Decision tree* memiliki keunggulan sebagai langkah awal dalam proses pemodelan dan saat dijadikan model akhir dari beberapa teknik lainnya (Suyanto, 2019) [18].

2.1.9 Random Forest

Random forest dapat diartikan sebagai kombinasi dari predictor berbentuk pohon, dimana setiap pohon tergantung kepada nilai-nilai vektor random dengan distribusi yang sama pada semua pohon yang masing masing decision tree memiliki kedalaman yang maksimal (Breiman, 2001) [2].

2.1.10 Text Mining

Menurut Ignatow dan Mihalcea (2018) [9] *text mining* didefinisikan sebagai proses yang mencakup pencarian informasi dengan metode statistik canggih dan pemrosesan bahasa alami (*Natural Language Processing*) seperti *part of speech tagging* dan *syntactic parsing*.

2.1.11 User Demographic Attributes

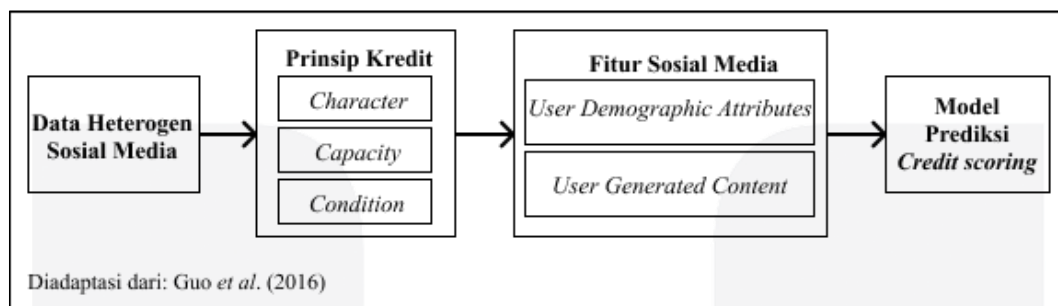
Menurut Guo *et al.* (2016) [8] *user demographic attributes* didefinisikan sebagai informasi pribadi yang ditambahkan sendiri oleh pengguna di sosial media yang dapat berupa informasi umur, jenis kelamin, pendidikan, dan pekerjaan.

2.1.12 User Generated Content

Menurut Moens *et al.* (2014) [14] merupakan data yang diunggah pengguna di web dimana unggahan tersebut tersedia untuk umum, berisi kreativitas pengguna, dan dibuat diluar kegiatan formal pengguna. Data berupa tulisan dalam sosial media menurut Guo *et al.* (2016) [8] dapat mewakili beberapa prinsip pokok kredit dengan pendekatan yang tepat.

2.2 Kerangka Pemikiran

Kerangka pemikiran yang digunakan pada penelitian ini mengadopsi model penelitian sebelumnya dari Guo *et al.*, (2016) [8].



Gambar 2.1 Kerangka Pemikiran

Sumber: Data yang telah diolah

2.3 Metodologi

Penelitian ini menggunakan data berupa pengguna sosial media LinkedIn sebanyak 132 akun yang berasal dari koneksi pertemanan salah satu dosen Fakultas Ekonomi Bisnis, Telkom University, dari setiap akun tersebut terdapat informasi untuk diolah menjadi fitur sosial media. Data tersebut terdiri dari data *user demographic attributes* dan data *user generated content*. Analisis kredit dilakukan menggunakan prinsip 3C, yaitu *character*, *capacity*, dan *conditions* [8]. Kriteria penilaian kelayakan kredit menggunakan data sosial media ditentukan berdasarkan tabel 2.1.

Tabel 2.1 Kredit Scorecard

No	Atribut	Nilai	Keterangan	No	Atribut	Nilai	Keterangan
1	Salary	1	0-5.000.000	7	Authenticity	1	<50
		2	5.000.0001-10.000.000			2	50-80
		3	10.000.001-20.000.000			3	>80
		4	20.000.001-40.000.000				
		5	>40.000.000 *dalam rupiah				
2	Age	1	<19	8	Emotional tone	1	<50
		2	19-34			2	50-80
		3	35-54			3	>80
		4	>54				
3	Gender	1	Wanita	9	Jumlah like	1	<100

		2	Pria			2 3	100-300 >300
4	<i>Education</i>	1 2 3 4 5	Kluster 5 Kluster 4 Kluster 3 Kluster 2 Kluster 1	10	Jumlah <i>comment</i>	1 2 3	<100 100-150 >150
5	<i>Analytical thinking</i>	1 2 3	<50 50-80 >80	11	Jumlah <i>post</i>	1 2	<15 ≥15
6	<i>Clout</i>	1 2 3	<50 50-80 >80				

Sumber: Data yang telah diolah

3. Hasil dan Pembahasan

3.1 Pembentukan Model Penilaian Kredit

Tabel 3.1 Desain Percobaan Model

Percobaan	Data Input
Model 1	Data user <i>demographic attributes</i>
Model 2	Data user <i>generated content</i>
Model 3	Data user <i>demographic attributes</i> + Data user <i>generated content</i>

Sumber: Data yang telah diolah

Penelitian ini membuat model menjadi 3 jenis seperti pada keterangan tabel 3.1. Setelah data melalui proses *preprocessing*, langkah selanjutnya menentukan *threshold* kelayakan kredit. Agar dapat dikatakan layak mendapatkan kredit akumulasi *threshold* dibandingkan dengan bobot skor setiap data yang memiliki ketentuan seperti tabel 2.1. Kemudian, data diolah menggunakan algoritma *decision tree* dan *random forest* untuk dievaluasi *performance metrics*-nya.

3.2 Hasil Evaluasi *Performance Metrics*

Tabel 3.2 Hasil *Performance Metrics*

Algoritma	Model Percobaan	<i>Performance Metrics</i>			
		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Decision tree</i>	Model 1	80.30%	83.33%	75.76%	79.37%
	Model 2	76.52%	78.69%	72.73%	75.59%
	Model 3	87.12%	85.29%	89.23%	87.22%
<i>Random forest</i>	Model 1	75.00%	74.63%	75.76%	75.19%
	Model 2	81.06%	83.61%	77.27%	80.31%
	Model 3	87.12%	87.50%	86.16%	86.82%

Sumber: Data yang telah diolah

. Dari hasil evaluasi *performance metrics*, model 3 yang diolah menggunakan kedua algoritma memiliki *accuracy* tertinggi diantara model lainnya yaitu sebesar 87.12%, akan tetapi algoritma *decision tree* memiliki kelebihan dengan memiliki *recall* dan *f-measure* lebih tinggi daripada algoritma *random forest*, dengan rincian *recall* sebesar 89.23%, dan *f-measure* sebesar 87.22%. Dari hasil tersebut dapat disimpulkan secara keseluruhan bahwa model 3 yang terdiri dari data *user demographic attributes* dan *user generated content* yang diolah menggunakan algoritma *decision tree* merupakan model terbaik dalam menentukan kelayakan kredit seseorang.

Data sosial media LinkedIn dapat dijadikan data penentu kelayakan kredit seseorang, data profil akun dapat dijadikan data alternatif pengganti data demografi dalam menentukan kelayakan kredit dan data unggahan atau post juga dapat dijadikan penentu kelayakan kredit [8], akan tetapi berdasarkan hasil kinerja akurasi pada tabel 3.2 dan proses pengolahannya, data unggahan akan lebih efektif dan efisien jika dijadikan data pelengkap karena data unggahan perlu melalui tahap pengolahan yang lebih panjang dan lama daripada data demografi sedangkan dalam dunia kerja fintech lending diperlukan adanya kecepatan pemrosesan data. Berdasarkan hasil tersebut, perusahaan *P2P Lending* akan mampu membuka segmentasi konsumen atau nasabah baru yang tidak memiliki riwayat kredit yang lengkap atau biasa disebut kelas *underbanked* [7]. Dengan adanya kesempatan membuka segmen baru konsumen, perusahaan *P2P Lending* dapat mengembangkan rencana bisnis yang dimilikinya, strategi pemasaran yang lebih terarah, dan menjadi daya tarik bagi konsumen [5], selain itu masyarakat yang tergolong kelas *underbanked* akan dimudahkan untuk mendapatkan akses pinjaman keuangan yang sebelumnya tidak dapat dijangkau jika tidak memiliki riwayat kredit yang lengkap.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

- a. Untuk menjadikan data sosial media menjadi indicator penentu kelayakan kredit seseorang memerlukan metode pengolahan data yang tepat dan pemahaman data yang lebih lanjut karena tidak dapat secara langsung menerjemahkan permasalahan kredit, terutama pada data user generated content. Sebagai contoh, dengan melihat data profil berupa atribut data job, gaji seseorang dapat diperkirakan besarnya secara langsung, akan tetapi berbeda dengan data unggahan yang perlu dilakukan analisis lebih lanjut untuk mengetahui sifat emosional seseorang sehingga bisa digunakan sebagai ukuran prinsip *character*.
- b. Proses pembuatan model penilaian kredit menggunakan data sosial media memanfaatkan data *User demographic attributes* dan *user generated content* yang nantinya data dipilih dengan melakukan pendekatan pada teori prinsip kredit yaitu *character*, *capacity*, dan *condition*. Model penentuan kelayakan kredit dibuat menjadi 3 model percobaan berdasarkan data input, yaitu model 1 menggunakan data *user demographic attribute*, model 2 menggunakan data *user generated content*, dan model 3 menggunakan data gabungan model 1 dan 2
- c. Dari hasil evaluasi performance metrics yang ada, model 3 yang diolah menggunakan algoritma decision tree dipilih menjadi model penilaian kredit menggunakan data sosial media karena memiliki 3 metric yang memiliki nilai tertinggi, yaitu accuracy sebesar 87.12%, nilai recall sebesar 89.23%, dan nilai f-measure sebesar 87.22%. Dengan membandingkan proses dan hasil performance metrics yang diperoleh model 1 dan model 2, dapat ditarik kesimpulan bahwa data user generated content pada model 2 cenderung lebih cocok bila dijadikan data pelengkap. Berdasarkan hasil tersebut, perusahaan *P2P Lending* akan mampu membuka segmentasi konsumen atau nasabah baru yang tidak memiliki riwayat kredit yang lengkap atau biasa disebut kelas *underbanked*.

4.2 Saran

4.2.1 Saran Teoritis

Pada penelitian selanjutnya diharapkan juga dapat mengidentifikasi data sosial media menggunakan tambahan prinsip capital dan collateral, pemahaman lebih jauh memang dibutuhkan untuk mengidentifikasi prinsip jaminan dan modal, akan tetapi bukan berarti hal tersebut tidak bisa dilakukan. Sebagai contoh, ada kemungkinan pada penelitian selanjutnya unggahan foto barang mewah dapat dijadikan data berdasarkan kedua prinsip tersebut atau jaringan pertemanan sosial media yang dapat dijadikan data berdasarkan prinsip collateral. Dengan terpenuhinya seluruh prinsip kredit 5C, diharapkan penilaian kelayakan kredit menggunakan data sosial media dapat ditingkatkan lebih baik lagi.

4.2.2 Saran Praktis

Penilaian kelayakan kredit menggunakan data sosial media LinkedIn dapat dijadikan contoh pembentukan penilaian kredit bagi perusahaan *P2P Lending* atau lembaga jasa keuangan lainnya yang memiliki permasalahan kekurangan data atau *data availability*. Dengan menggunakan data sosial media, lembaga jasa keuangan dapat menghemat waktu dan biaya dalam menentukan penilaian kelayakan kredit seseorang, sehingga keunggulan kompetitif perusahaan dapat ditingkatkan. Kedepannya, perusahaan *P2P Lending* dapat melakukan kerjasama dengan pihak penyedia layanan sosial media dalam mengumpulkan data konsumen atau nasabahnya untuk dijadikan data penentu kelayakan kredit. Selain itu, *P2P Lending* juga dapat menggunakan bantuan seorang psikolog untuk merancang model penilaian kredit menggunakan data user generated content agar mendapat hasil analisis kepribadian konsumen atau nasabah secara maksimal. *P2P Lending* juga dapat memberikan tes psikometri secara rutin kepada konsumen atau nasabah guna menjaga tingkat pengembalian kredit. Penggunaan data sosial

media juga dapat membantu meningkatkan penilaian kelayakan kredit, yang nantinya diharapkan dapat meningkatkan tingkat pengembalian kredit.

Daftar Pustaka:

- [1] Bahga, A., & Madiseti, V. (2019). *Big Data Analytics: A Hands-On Approach*. VPT.
- [2] Breiman, L. (2001). Random Forests. *Machine Learning*, 5-32.
- [3] Campbell-Verduyn, M., Goguen, M., & Porter, T. (2016). Big Data and algorithmic governance: the case of financial practices. *New Political Economy*.
- [4] DailySocial.id. (2019, Desember 4). *post: menggodok-data-membangun-credit-scoring*. Retrieved from DailySocial.id: <https://dailysocial.id/post/menggodok-data-membangun-credit-scoring>
- [5] Dharmmesta, B. S. (2017). *Azas-Azas Marketing*. Yogyakarta: Liberty
- [6] Fahmi, I. (2014). *Manajemen Perkreditan*. Bandung: Alfabeta.
- [7] Federal Reserve Bank of Kansas City. (2010). *A Study of the Unbanked and Underbanked Consumer in the Tenth Federal Reserve District May 2010*.
- [8] Guo, G., Zhu, F., Chen, E., Liu, Q., Wu, L., & Guan, C. (2016). From footprint to evidence: An exploratory study of mining social data for credit scoring. *ACM Transactions on the Web*, 10(4), 22:1-38. doi:10.1145/2996465
- [9] Ignatow, G., & Mihalcea, R. (2018). *An Introduction to Text Mining: Research Design, Data Collection, and Analysis*. California: SAGE Publications.
- [10] Ikatan Bankir Indonesia. (2015). *Bisnis Kredit Perbankan*. Jakarta Pusat: Gramedia Pustaka Utama.
- [11] Keramati, A., & Yousefi, N. (2011). A Proposed Classification of Data Mining Techniques in Credit Scoring. *Proc. 2011 Int. Conf. on Industrial Engineering and Operations Management*. Kuala Lumpur.
- [12] Kotu, V., & Deshpande, B. (2015). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Waltham: Elsevier.
- [13] Larose, D. T. (2005). *Data Mining Methods and Models*. John Wiley & Sons.
- [14] Moens, M.-F., Li, J., & Chua, T.-S. (2014). *Mining User Generated Content*. CRC Press.
- [15] Mulawarman, M., & Nurfitri, A. D. (2017). Perilaku Pengguna Media Sosial beserta Implikasinya Ditinjau dari Perspektif Psikologi Sosial Terapan. *Buletin Psikologi*, 36-44.
- [16] Raghunathan, B., Maiya, R. (2018). *SMACing the Bank: How to Use Social Media, Mobility, Analytics, and Cloud Technologies to Transform the Business Processes of Banks and the Banking Experience*. Boca Raton, FL : CRC Press, 2017.
- [17] Shibuya, Y. (2020). *Social Media Communication Data for Recovery*. Tokyo: Springer Singapore.
- [18] Suyanto. (2019). *Data Mining Untuk Klasifikasi dan Klusterisasi Data, Edisi Revisi*. Bandung: Informatika.
- [19] Zhao, H., Liu, Q., Wang, G., Ge, Y., & Chen, E. (2016). Portfolio Selections in P2P Lending: A Multi-Objective Perspective. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 2075–2084). New York: Association for Computing Machinery.