



PAPER – OPEN ACCESS

Disagregasi Preferensi Konsumen menggunakan Metode Bayesian Linear Model

Author : Chrisentiana Talithapurwa Jasindatama, dkk.
DOI : 10.32734/ee.v7i1.2170
Electronic ISSN : 2654-704X
Print ISSN : 2654-7031

Volume 7 Issue 1 – 2024 TALENTA Conference Series: Energy and Engineering (EE)



This work is licensed under a [Creative Commons Attribution-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nd/4.0/).

Published under licence by TALENTA Publisher, Universitas Sumatera Utara



Disagregasi Preferensi Konsumen menggunakan Metode *Bayesian Linear Model*

Chrisentiana Talithapurwa Jasindatama^{a*}, Andi Sudiarso^a, Shi-Woei Lin^b

^aDepartemen Teknik Mesin dan Industri, Fakultas Teknik, Universitas Gadjah Mada, Jl. Grafika No.2, Senolowo, Sinduadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281, Indonesia

^bDepartment of Industrial Management, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Rd, Da'an District, Taipei City, 106

✉ chrisentiana@mail.ugm.ac.id, a.sudiarso@ugm.ac.id, shiwoei@mail.ntust.edu.tw

Abstrak

Pemahaman terhadap preferensi dan perilaku konsumen sangat penting dalam berbagai rangkaian proses bisnis. Ulasan produk menjadi sumber informasi berharga untuk pengambilan keputusan dalam pemasaran, desain produk, dan segmentasi pasar. Pendekatan melalui *multiple attribute decision-aiding* (MADA) banyak diterapkan dalam konteks disagregasi preferensi. Khususnya dengan menggunakan regresi linear bayesian untuk mengekstraksi bobot dari aspek ulasan pelanggan. Penelitian ini menerapkan regresi linear bayesian yang berfokus pada ekstraksi bobot aspek yang memengaruhi penilaian peringkat konsumen untuk pengambilan keputusan bisnis dengan mempertimbangkan data yang hilang atau bernilai nol (0). Metode substitusi rerata yang digunakan untuk menyelesaikan masalah data yang hilang berhasil mengungguli model dengan data hilang dengan performa untuk menjelaskan keputusan penilaian peringkat produk. Bobot yang didapatkan dapat digunakan oleh pelaku bisnis dalam pengambilan keputusan seperti memfokuskan peningkatan kualitas produk dari aspek produk berdasarkan hasil matriks performa-kepentingan.

Kata Kunci: Bayesian Linear Model; Pengambilan Keputusan; Disagregasi Preferensi; Ulasan Konsumen

Abstract

Gaining insight into consumer preferences and behavior is crucial in multiple corporate operations. Product reviews are an invaluable resource for making decisions in marketing, product design, and market segmentation. Therefore, a robust analysis is required to effectively utilize the available information. The use of multiple attribute decision-aiding (MADA) is commonly employed in the field of preference disaggregation. Specifically, Bayesian linear regression has been employed in research to derive weights for individual aspects of customer reviews. This study examines the application of Bayesian linear regression to extract aspect weights that impact consumer ratings for corporate decision-making, taking into account missing data or zero values (0). The mean substitution strategy effectively outperforms the model with missing data in accurately explaining consumer product rating decisions. The derived weights can be utilized by business professionals to inform decision-making, specifically in prioritizing efforts to enhance product quality based on the outcomes of the performance-importance matrix.

Keywords: Bayesian Linear Model; Decision Making; Preference Disaggregation; Customer Review

1. Pendahuluan

Pemahaman terhadap preferensi pelanggan merupakan hal yang integral dalam bisnis, terlebih pada proses desain dan pengembangan produk. Dengan mengetahui preferensi pelanggan, bisnis dapat merancang produk dan layanan yang sesuai untuk pelanggan yang dituju, sehingga meningkatkan kepuasan pelanggan dan penjualan.

Perkembangan teknologi memudahkan akses terhadap pendapat dari konsumen secara langsung. Ulasan dari konsumen menjadi aset penting bagi bisnis. Informasi yang diberikan oleh konsumen terhadap suatu produk dapat menjadi masukan yang berharga dalam pengambilan keputusan pemasaran, desain produk, dan segmentasi pasar. Dengan mengetahui preferensi konsumen, bisnis dapat menciptakan produk yang sesuai dengan target pasar.

Dalam konteks pengambilan keputusan, penilaian dari ahli diperlukan untuk menentukan bobot kepentingan dari aspek-aspek yang menentukan keputusan, di antaranya adalah aspek produk. Pendapat ahli dapat menimbulkan bias dan berpeluang untuk kurang mewakili kondisi pasar yang sebenarnya. Oleh karena itu, penting untuk mendapatkan informasi tambahan tentang

preferensi pelanggan, yang dapat diperoleh dari ulasan pelanggan. Pada dasarnya, tiap produk biasanya memiliki ulasan peringkat keseluruhan dan juga ulasan lengkap dalam bentuk kalimat yang diberikan oleh pelanggan. Pendapat pelanggan mengenai baik atau buruknya suatu produk dapat dilihat sekilas dari peringkat keseluruhan yang diberikan. Walaupun demikian, [1] menemukan bahwa beberapa evaluasi mungkin memiliki peringkat keseluruhan yang sama namun menunjukkan skor aspek yang berbeda. Keterbatasan informasi dalam pemeringkatan ini mengakibatkan kesulitan untuk mendapatkan wawasan penuh tentang preferensi pelanggan. Peringkat tersebut tidak sepenuhnya mencerminkan preferensi pelanggan, namun hanya mencerminkan opini umum tentang produk. Hal ini mungkin dapat mengakibatkan bias yang mengakibatkan distorsi informasi dan penilaian kualitas produk yang tidak akurat karena penempatan bobot yang salah atau ada variabel lain yang mempengaruhi. Oleh karena itu, skor aspek harus dimanfaatkan untuk memastikan perbedaan kekuatan dan kelemahan masing-masing aspek produk dan signifikansinya terhadap peringkat keseluruhan. Namun, peringkat keseluruhan harus tetap dapat diakses sebagai dasar untuk menilai opini umum, sementara skor sentimen tingkat aspek harus digunakan untuk menganalisis fitur spesifik yang terkait dengan objek tersebut.

Penelitian mengenai disagregasi preferensi telah banyak dilakukan sebelumnya, terdapat tiga metode utama yang telah diteliti sebelumnya, pendekatan pemberi rekomendasi berbasis sistem [2], conjoint analysis [3], dan pendekatan multiple attribute decision-aiding (MADA) [4]. Penelitian mengenai disagregasi preferensi erat hubungannya dengan pengambilan keputusan dengan banyak atribut (MADA). Guo et al. dalam [4] memanfaatkan model preferensi aditif untuk integrasi informasi ulasan online. Walaupun demikian dalam analisisnya masih memasukkan peran dari ahli. Berbeda dengan [4], [5] melakukan pendekatan model preferensi pelanggan dari MADA menggunakan integral Choquet 2-aditif untuk mengantisipasi persyaratan kemandirian bersama, yang umumnya tidak terpenuhi dalam skenario preferensi pelanggan. Analisis pada penelitian ini seluruhnya menggunakan data yang didapatkan dari ulasan konsumen, namun nilai aspek yang didapatkan sudah diberikan oleh konsumen dan bukan melalui pengolahan analisis sentimen. Selain itu, dari sisi aplikasi praktis, metode ini tergolong lebih kompleks untuk diterapkan oleh perusahaan.

Terdapat penelitian sejenis yang dilakukan [6] telah memanfaatkan regresi linear bayesian untuk mengekstraksi bobot dari tiap aspek. Idealnya, bobot yang didapatkan totalnya berjumlah satu. Namun, pada penelitian ini kondisi tersebut belum dapat dicapai sehingga bobot yang didapatkan dari hasil regresi belum dapat langsung digunakan oleh pengambil keputusan. Maka dari itu, penelitian ini akan mencoba untuk mengatasi masalah tersebut sehingga bobot yang didapatkan dari hasil regresi bertotal sama dengan satu. Selain itu, penelitian terdahulu belum mempertimbangkan bahwa data yang dimiliki memiliki nilai nol, yang disebabkan karena konsumen tidak menyebutkan kata apapun yang terkait dengan aspek yang teridentifikasi, sehingga data ini dapat dikategorikan sebagai data yang hilang. Pada penelitian ini metode untuk mengatasi data yang hilang diterapkan.

Walaupun telah banyak penelitian dibidang disagregasi preferensi, belum banyak penelitian yang mengintegrasikan peringkat bintang dan ulasan teks untuk menganalisis preferensi pelanggan mengenai atribut atau kriteria yang lebih diperhatikan pelanggan. Maka, penelitian ini bertujuan untuk menganalisis preferensi konsumen berdasarkan ulasan yang diberikan oleh konsumen. Berbeda dengan sebagian besar penelitian yang dilakukan pada bidang ini, penelitian ini berfokus untuk mencerminkan kompleksitas dan ketidakpastian perilaku pasar di dunia nyata, memberikan dasar yang kuat untuk memprediksi tren konsumen dengan menggunakan pendekatan probabilistik melalui *Bayesian linear regression* (BLR). Selain itu penelitian ini mempertimbangkan kemungkinan data yang hilang atau tidak disebutkan dalam konteks analisis sentimen dari ulasan konsumen. Maka penelitian ini bertujuan untuk memahami pengaruh aspek produk terhadap evaluasi produk oleh konsumen untuk mendukung pengambilan keputusan dalam proses bisnis dan mengisi kesenjangan yang ada dalam literatur dengan menggabungkan teknik analisis sentimen dengan regresi bayesian untuk menganalisis statistik dan mengisi data yang hilang, memberikan kontribusi wawasan berharga baik untuk bidang akademis maupun aplikasi praktis dalam analisis pasar.

2. Metode Penelitian

Data yang digunakan dalam penelitian ini berupa data hasil pengolahan analisis sentimen berbasis aspek untuk sebagai variabel independen dan penilaian bintang oleh pelanggan sebagai variabel dependen. Data tersebut merupakan data ulasan telepon selular yang digunakan pada penelitian serupa dengan regresi linear bayes untuk mengekstraksi bobot aspek yang dilakukan oleh [6]. Data terdiri dari satu variabel dependen, yaitu peringkat keseluruhan atau *rating* (Y), dan enam variabel independen, X_1 (*worth*), X_2 (*application*), X_3 (*design*), X_4 (*network*), X_5 (*durability*), X_6 (*performance*). Data variabel independen telah dilakukan pengolahan dengan mengambil rata-rata dari nilai sentimen, maka dapat dianggap sebagai variabel kontinu. Pengolahan data dilakukan dengan menggunakan regresi bayesian.

Penelitian ini menerapkan dua skenario, dikarenakan hasil analisis sentimen memiliki nilai nol (0). Skenario pertama, data yang bernilai nol tetap dianggap valid, sehingga dapat langsung digunakan dalam analisis lebih lanjut. Skenario kedua, data yang bernilai nol (0) dianggap tidak valid, sehingga perlu memasukkan nilai pengganti yang pada penelitian ini menggunakan rerata

dari setiap variabel independen. Kemudian untuk mengekstraksi bobot dari tiap aspek yang kemudian dapat digunakan sebagai bobot dalam pengambilan keputusan, dilakukan analisis regresi linear bayes. Selanjutnya bobot yang telah didapatkan dari hasil bayesian linear regression dapat digunakan oleh pemilik bisnis dalam berbagai proses bisnisnya, terutama dalam pengambilan keputusan. Bobot ini dapat secara langsung digunakan untuk pembobotan dalam analisis pengambilan keputusan multi-atribut. Salah satu metode yang dapat digunakan dalam aplikasi praktisnya, yang akan dipraktikkan dalam penelitian ini, adalah dengan memetakan bobot dengan nilai rerata aspek terkait dalam matriks performa-kepentingan.

Bayesian Linear Regression merupakan salah satu metode model regresi yang menerapkan pendekatan bayesian untuk mengestimasi parameternya. Dengan mengimplementasikan pendekatan bayesian terhadap regresi memungkinkan untuk mendapatkan hasil yang lebih terpercaya walaupun dengan jumlah data yang sedikit. Konsep utama regresi bayesian adalah mencari estimasi parameter regresi melalui distribusi probabilitas untuk parameter tersebut. Persamaan regresi bayesian adalah sebagai berikut [6] :

$$Y_i = \sum_{j=1}^p X_{ij}\beta_j + \varepsilon_i \tag{1}$$

Kelebihan dari regresi linear bayesian adalah kemampuannya untuk mengatasi masalah kurangnya data [6], dapat mengestimasi ketidakpastian dalam prediksi, dan juga dapat meningkatkan performa model saat diketahui informasi mengenai data. Selain itu metode ini dapat memenuhi kebutuhan jumlah dari bobot yang didapatkan sama dengan 1 mengingat konteks dari analisis dilakukan untuk mendapatkan bobot penentuan keputusan, yang mana tidak dimungkinkan dalam regresi linear berganda standar. Regresi bayesian sangat erat kaitannya dengan penentuan prior dan likelihood dari model. Dalam penelitian ini untuk memenuhi kebutuhan ekstraksi bobot untuk pengambilan keputusan, digunakan distribusi Dirichlet yang dapat membatasi distribusi posterior bobot bertotal lebih dari 1. Batasan ini muncul dari sifat distribusi dirichlet yang memastikan bahwa vektor probabilitas memiliki total 1. Distribusi Dirichlet merupakan generalisasi multivariat dari distribusi beta. Sebagai *non-informative* prior, distribusi dirichlet yang digunakan dalam penelitian ini memiliki parameter alpha = [1,1,1,1,1,1]. Hal ini karena distribusi ini merupakan distribusi seragam pada simpleks, artinya semua titik dalam ruang parameter mempunyai kemungkinan yang sama [7].

Regresi linear bayesian menghasilkan data yang bersifat kontinyu, sedangkan pada dasarnya penilaian umum merupakan data ordinal, maka diperlukan transformasi data agar menjadi data diskrit. Diskritisasi dapat dilakukan dengan membagi data yang ada menjadi beberapa bin, namun juga dapat dilakukan dengan pembulatan. Dalam penelitian ini dilakukan pembulatan ke atas dengan asumsi pelanggan lebih cenderung memilih untuk memberikan angka yang lebih tinggi dibandingkan lebih rendah dan penilaian subyektifnya.

3. Hasil dan Pembahasan

3.1. Skenario 1

Hasil regresi yang dapat dilihat pada Tabel 1. menunjukkan bahwa variabel X_1 (*worth*) dengan bobot 0.689 memiliki bobot yang paling tinggi dibandingkan variabel yang lain. Kemudian kedua adalah X_4 (*design*) dengan bobot 0.156, diikuti oleh X_2 (*application*) dengan bobot 0.078, kemudian X_3 (*network*) dengan bobot 0.043, dan dua variabel yang memiliki bobot terkecil adalah X_6 (*performance*) dan terakhir X_5 (*durability*).

Tabel 1. Parameter Regresi Skenario 1 (Model 1)

Parameter	X_1	X_2	X_3	X_4	X_5	X_6
rerata	0.689	0.078	0.043	0.156	0.010	0.024
distribusi	normal	normal	weibull	normal	beta	weibull

Tabel 2. Interval Kepercayaan Model 1

interval kepercayaan	X_1	X_2	X_3	X_4	X_5	X_6
5.50%	0.6365317	0.01912066	0.00480746	0.09371095	0.0005978	0.0018684
94.50%	0.7411786	0.1426003	0.0937954	0.21812218	0.02913677	0.0601068

Berbeda dengan regresi linear OLS, dalam regresi linear bayes terdapat interval kepercayaan yang menunjukkan ketidakpastian terkait parameter yang diestimasi [8]. Pada Tabel 2, X_1 , dengan data observasi, terdapat probabilitas sebesar 89% bahwa nilai parameter bobot diantara 0.637 dan 0.7411. Jangkauan nilai bobot untuk X_1 cukup kecil, namun hal ini tidak terjadi pada bobot variabel independen lainnya. Untuk X_5 , jangkauan nilai bobot sangat lebar, yaitu dari 0.000597 sampai 0.02913. Jangkauan nilai bobot yang lebar juga terjadi kurang lebih pada 4 variabel independen lainnya. Hal ini mungkin terjadi karena variabilitas data

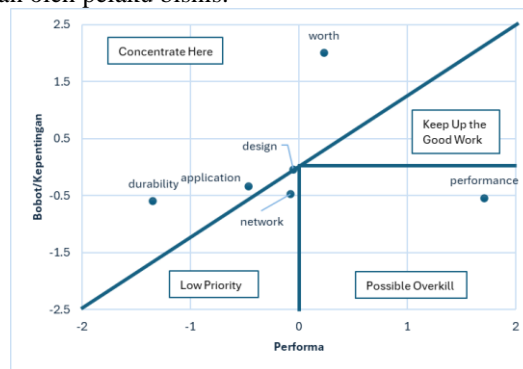
sangat tinggi. Alasan lain yang mungkin menyebabkan lebarnya jangkauan nilai bobot adalah kurangnya jumlah data.

		Referensi				
		1	2	3	4	5
Prediksi	1	55	31	31	24	18
	2	35	17	14	28	26
	3	16	18	15	20	37
	4	3	5	15	47	74
	5	0	0	4	17	34

Gambar 1. Confusion Matrix Model 1

Hasil dari regresi linear bayesian, kemudian dibulatkan ke atas. Gambar 1 menunjukkan *confusion matrix* dari hasil prediksi regresi. Dengan akurasi sebesar 28,77%, dapat dikatakan bahwa akurasi model ini tidak terlalu baik. Kesalahan prediksi nilai peringkat keseluruhan produk terlihat tersebar secara rata, contohnya model memprediksi nilai 1 baik untuk nilai aktual yang bernilai 2 sampai 5. Begitu pula untuk peringkat keseluruhan dengan nilai 5 banyak diprediksi berada pada nilai lain. Model ini kurang sensitif terhadap nilai 5. Mengingat data adalah data ordinal, maka hanya akurasi saja tidak bisa menjamin performa dari model. Kesalahan prediksi dari 1 menjadi 2 berbeda dengan nilai 1 diprediksi bernilai 5. Sehingga dapat disimpulkan performa model ini kurang baik.

Pada Gambar 2. telah dipetakan keenam aspek produk handphone X dengan nilai rerata aspek terkaitnya dalam matriks performa-kepentingan milik [9]. Terdapat tiga aspek yang berada pada kuadran I (*concentrate here*), yaitu *worth*, *application*, dan *durability*. Hal ini berarti kinerja produk dalam aspek terkait masih tergolong rendah, namun aspek aspek tersebut memiliki bobot yang tinggi dalam pertimbangan konsumen dalam memberikan peringkat tinggi pada produk. Oleh karena itu pelaku bisnis perlu meningkatkan kinerja produk dalam aspek terkait, sehingga aspek tersebut masuk kedalam kuadran II dan meningkatkan kepuasan pelanggan terhadap produk. Sedangkan pelaku bisnis tidak perlu terlalu memperhatikan aspek *network*, karena aspek ini terdapat pada kuadran III (*low priority*). Selain itu terdapat satu aspek yang terdapat pada kuadran IV (*possible overkill*), yaitu *performance*. Walaupun performa produk telpon seluler X dinilai sangat baik, namun aspek ini tidak terlalu memengaruhi penilaian peringkat dari konsumen, sehingga dapat diabaikan oleh pelaku bisnis.



Gambar 2. Matriks Performa-Kepentingan Model 1

3.2. Skenario 2 (Substitusi Rerata)

Tabel 3. Parameter Regresi Skenario 2 (Model 2)

Parameter	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
rerata	0.46	0.05	0.17	0.03	0.02	0.27
distribusi	beta	weibull	normal	weibull	weibull	normal

Hasil regresi yang dapat dilihat pada Tabel 3. menunjukkan bahwa variabel X₁ (*worth*) memiliki bobot yang paling tinggi dibandingkan variabel yang lain, dengan bobot 0.46. Kemudian kedua adalah X₆ (*performance*) dengan bobot 0.27, diikuti oleh X₃ (*network*) dengan bobot 0.17, kemudian X₂ (*application*) dengan bobot 0.05, dan dua variabel yang memiliki bobot terkecil adalah X₄ (*design*) dan terakhir X₅ (*durability*).

Tabel 4. Interval Kepercayaan Model 2

interval kepercayaan	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
5.50%	0.395964	0.00501617	0.09960298	0.0022872	0.0017708	0.2141065
94.50%	0.515271	0.10394305	0.24726518	0.07194914	0.05847879	0.3321659

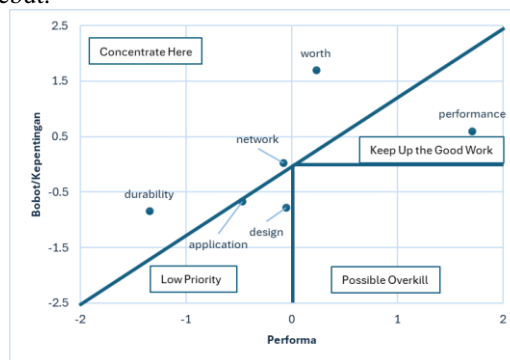
Berbeda dengan regresi linear OLS, dalam regresi linear bayes terdapat interval kepercayaan yang menunjukkan ketidakpastian terkait parameter yang diestimasi [8]. Hasil dapat dilihat pada Tabel 4., untuk X₁ dengan data observasi, terdapat probabilitas sebesar 89% bahwa nilai parameter bobot diantara 0.3959 dan 0.51527. Jangkauan nilai bobot untuk X₁ dan X₆ cukup kecil, namun hal ini tidak terjadi pada bobot variabel independen lainnya. Untuk X₂, jangkauan nilai bobot sangat lebar, yaitu dari 0.005016 sampai 0.1039. Jangkauan nilai bobot yang lebar juga terjadi kurang lebih pada 3 variabel independen lainnya. Hal ini mungkin terjadi karena variabilitas data sangat tinggi. Alasan lain yang mungkin menyebabkan lebarnya jangkauan nilai bobot adalah kurangnya jumlah data. Hal ini mungkin juga disebabkan oleh sifat data yang merupakan hasil analisis sentimen dari ulasan konsumen, sehingga data memiliki banyak variabilitas mengingat sulitnya untuk menyamaratakan perilaku konsumen.

Hasil dari regresi linear bayesian, kemudian dibulatkan ke atas. Gambar 3. menunjukkan *confusion matrix* dari hasil prediksi regresi. Akurasi sebesar 30,65% sekilas dapat dikatakan bahwa akurasi model ini tidak terlalu baik. Meskipun demikian, perlu diperhatikan bahwa hampir tidak ada peringkat keseluruhan yang diprediksi bernilai 1. Hal ini cukup baik dalam konteks tidak banyak kesalahan prediksi yang cukup signifikan (1 menjadi 5 atau 5 menjadi 1), namun disisi lain, hal ini juga berarti model yang ada kurang sensitif terhadap penilaian bernilai 1. Hal ini juga terjadi untuk penilaian yang bernilai 5, model bekerja cukup baik untuk mengidentifikasi penilaian yang tidak bernilai 5, hal ini ditunjukkan dari spesifitas bernilai 92%, yang berarti model sangat jarang salah memprediksi skenario penilaian yang tidak bernilai 5 sebagai nilai 5. Secara umum performa dari model cukup baik. Perbandingan performa model ini dengan model sebelumnya dibahas lebih lanjut pada subbab berikutnya.

		Referensi				
		1	2	3	4	5
Prediksi	1	2	0	0	0	0
	2	28	17	2	2	2
	3	59	40	41	41	31
	4	20	13	34	65	102
	5	0	1	2	28	54

Gambar 3. Confusion Matrix Model 2

Performa dan bobot dari model 2 dipetakan kedalam matriks performa-kepentingan pada Gambar 4. Berbeda dengan matriks model 1 pada Gambar 3. *performance* memiliki bobot yang lebih tinggi pada model 2, sehingga pada matriks terletak pada kuadran ke II (*keep up the good work*), dimana nilai performa sudah cukup tinggi dan merupakan aspek yang diperhatikan oleh konsumen dalam mempertimbangkan penilaian peringkat produk telepon seluler X. Hampir sama dengan model 1, secara umum performa *durability*, *network* dan *worth* perlu ditingkatkan sehingga aspek-aspek tersebut dapat bergeser dari kuadran I (*concentrate here*) menuju ke arah kuadran II, sehingga akan meningkatkan kepuasan konsumen terhadap produk telepon seluler X. Sedangkan hasil pembobotan pada model 2 menempatkan *application* dan *design* pada kuadran III (*low priority*) sehingga pelaku bisnis tidak perlu terlalu memperhatikan aspek aspek tersebut.



Gambar 4. Matriks Performa-Kepentingan Model 1

3.3. Perbandingan Model

Tabel 5. Perbandingan Evaluasi Performa Model

Parameter Evaluasi	Model 1	Model 2
MAE	1.265	0.913
RMSE	1.482	0.691
Kappa	0.417	0.567
P-value	<0.0001	<0.0001
Accuracy	0.2877	0.3065

Hasil dari model regresi linear bayesian dengan data yang telah melalui substitusi data nol dengan rata-rata dengan pembulatan ke atas, yang selanjutnya akan disebut model 1, menunjukkan akurasi sebesar 28.77 %, sedangkan untuk data yang melalui substitusi rerata, selanjutnya disebut model 2, memiliki akurasi sebesar 28.74%. Secara sekilas akurasi dari kedua model tersebut tidak memiliki perbedaan yang signifikan. Meskipun demikian, mengingat data bersifat ordinal penting juga untuk mengevaluasi data dari parameter lain untuk menentukan kesalahan prediksi yang dilakukan sangat jauh dari nilai peringkat sebenarnya atau tidak. Dapat dilihat bahwa dari *Mean absolute error* (MAE) dan *root mean squared error* (RMSE), Model 2 memiliki performa yang lebih baik. Dalam hal ini dapat disimpulkan bahwa kesalahan prediksi peringkat oleh model 2 tidak lebih dari 1 peringkat di atas atau di bawah peringkat sebenarnya. Evaluasi kedua model juga dapat dilihat dari nilai kappa. Nilai kappa untuk kedua model hampir mirip, walaupun model 2 memiliki nilai yang lebih tinggi namun secara umum kedua model memiliki kesepakatan yang cukup baik diantara nilai prediksi dan aktual.

Saat bekerja dengan data analisis sentimen, ada probabilitas memiliki data yang hilang, kemampuan dan performa regresi linear bayesian untuk mengekstraksi bobot terhadap skenario tersebut dan cara penanganannya telah diuji pada penelitian ini. Secara umum model 2 dapat memprediksi dan menggambarkan preferensi konsumen dengan lebih akurat. Seperti yang telah dibahas pada subbab sebelumnya, terlihat dari matriks performa-kepentingan bahwa model 2 dapat memetakan aspek-aspek mana yang penting diperhatikan oleh pelaku bisnis lebih jelas dibandingkan dengan hasil dari model 1. Hal ini juga membuktikan bahwa penanganan data yang hilang penting dalam skenario ini, dimana data yang bernilai 0 tidak dapat dianggap bahwa konsumen sangat tidak suka dengan aspek tersebut dan tetap diikuti pada analisis. Sehingga dengan substitusi data yang hilang dengan rata-rata nilai aspek, hasil analisis dapat lebih menggambarkan preferensi pasar secara umum.

4. Kesimpulan

Disagregasi preferensi dengan menggunakan regresi linear bayesian membantu mendapatkan bobot kepentingan aspek produk dari ulasan konsumen. Substitusi rerata pada data yang hilang memberikan penjelasan yang lebih baik tentang keputusan penilaian peringkat produk dari konsumen, sehingga bobot yang didapatkan dapat digunakan oleh pelaku bisnis dalam pengambilan keputusan seperti memfokuskan peningkatan kualitas produk dari aspek produk berdasarkan hasil matriks performa-kepentingan.

Terdapat keterbatasan dalam penelitian ini, penggunaan rerata dari tiap aspek memungkinkan nilai tidak merepresentasikan keadaan atau sentimen asli dari konsumen terhadap aspek terkait. Penelitian selanjutnya diharapkan dapat menerapkan metode lain, seperti pembobotan atau metode substitusi data lain seperti menggunakan *metaheuristic* atau *maximum likelihood estimator* (MLE). Selanjutnya, variabel dependen bersifat data ordinal, sehingga masih harus dilakukan pembulatan hasil prediksi. Maka dari itu, dapat menggunakan metode regresi lain seperti *logistic bayesian regression*, atau metode yang dapat langsung digunakan untuk analisis data ordinal.

Referensi

- [1] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA: ACM, Jul. 2010.
- [2] S. K. Lee, Y. H. Cho, and S. H. Kim, "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations," *Inf Sci (N Y)*, vol. 180, no. 11, pp. 2142–2155, Jun. 2010.
- [3] K. Hung, B. D. Guillet, and H. Q. Zhang, "Understanding Luxury Shopping Destination Preference Using Conjoint Analysis and Traditional Item-Based Measurement," *J Travel Res*, vol. 58, no. 3, pp. 411–426, Mar. 2019.
- [4] M. Guo, X. Liao, J. Liu, and Q. Zhang, "Consumer preference analysis: A data-driven multiple criteria approach integrating online information," *Omega (United Kingdom)*, vol. 96, Oct. 2020.
- [5] H. Liao, Q. Yang, and X. Wu, "Customer Preference Analysis From Online Reviews By A 2-Additive Choquet Integral-Based Preference Disaggregation Model," *Technological and Economic Development of Economy*, vol. 29, no. 2, pp. 411–437, Mar. 2023.
- [6] F. P. Andrianto, "Consumer Preference Analysis and Disaggregation Based on Aspect-Level Sentiment Classification and Bayesian Approach," Unpublished master thesis, National Taiwan University of Science and Technology, Taipei, Taiwan, Jan. 2024.
- [7] S. Kotz, N. Balakrishnan, and N. L. Johnson, "Continuous Multivariate Distributions, Volume 1: Models and Applications," New York, NY: Wiley, 2000,

ch. 49, "Dirichlet and Inverted Dirichlet Distributions,"

- [8] D. Makowski, M. Ben-Shachar, and D. Lüdecke, "bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework," *J Open Source Softw*, vol. 4, no. 40, p. 1541, Aug. 2019.
- [9] J. Abalo, J. Varela, and V. Manzano, "Importance values for Importance-Performance Analysis: A formula for spreading out values derived from preference rankings," *J Bus Res*, vol. 60, no. 2, pp. 115–121, Feb. 2007.