

Klasifikasi Kisaran Harga Tarif Endorsement Influencer Instagram dengan Metode Decision Tree

Jeshen Oktavian Nathanel, Alexander Setiawan

Program Studi Informatika Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031) – 2983455, Fax. (031) – 8417658

E-Mail: jeshentambuwun@gmail.com, alexander@petra.ac.id

ABSTRAK

Digital marketing pada era globalisasi ini merupakan hal yang wajib bagi small business owner ataupun bisnis yang sudah berkembang. Marketing menggunakan Instagram Influencer merupakan salah satu cara. Sering kali influencer mematok harga mereka sesuai hanya sesuai dengan followers yang mereka miliki sedangkan belum tentu post bisa meng-engage followers mereka yang aktif dan tidak palsu. Permasalahan yang ingin di selesaikan oleh peneliti adalah klasifikasi harga endorsement yang sesuai dengan data yang mereka miliki dengan melihat selain followers seperti engagement rate, average liker, dan average comment agar harga yang di tetapkan influencer lebih sesuai dengan data yang mereka miliki. Metode yang di gunakan untuk melakukan klasifikasi adalah Decision Tree dengan CART algorithm. Model yang di hasilkan ini akan di gunakan untuk mengklasifikasi harga endorsement influencer. Hasil dari pengujian model yang di gunakan hanya mampu mencapai akurasi 50% dengan nilai RMSE 1.12 untuk melakukan klasifikasi harga endorsement influencer.

Kata Kunci: *Decision Tree, Endorsement, Influencer, Instagram.*

ABSTRACT

Digital marketing in this era of globalization is mandatory for small business owners or businesses that have developed. Marketing using Instagram Influencers is one way. Oftentimes, influencers set their prices according to the followers they have, while not necessarily posting can engage their followers who are active and not fake. The problem that researchers want to solve is the classification of endorsement prices according to the data they have by looking at other than followers such as engagement rate, average liker, and average comment so that the prices set by influencers are more in line with the data they have. The method used for classification is Decision Tree with CART algorithm. The resulting model will be used to classify influencer endorsement prices. The results from testing the model used are only able to achieve 50% accuracy with a RMSE value of 1.12 to classify influencer endorsement prices.

Keywords: *Decision Tree, Endorsement, Influencer, Instagram*

1. PENDAHULUAN

Penggunaan internet dan sosial media telah mengubah tingkah laku konsumen dan perusahaan dalam menjalankan bisnisnya. Sosial marketing dan digital marketing menawarkan kesempatan yang sangat signifikan untuk berbagai organisasi untuk menurunkan pengeluaran, mengembangkan brand awareness dan meningkatkan penjualan[5]. Pada tahun 2018, terdapat sekitar 171.176.716

pengguna internet yang berada di Indonesia. Dari jumlah tersebut, sekitar 18,9 persen menggunakan internet untuk menjelajahi sosial media seperti Facebook, Instagram, Twitter, dan YouTube. Lebih dari setengah pengguna internet untuk tujuan komunikasi di Indonesia berumur 19-34 tahun (49,52%). Dengan jumlah orang dan tujuan penggunaan tersebut, dapat dikatakan bahwa sosial media adalah tempat yang ideal untuk melakukan pemasaran dan penjualan.[1]. Untuk mencapai tujuan tersebut, para pemegang bisnis menggunakan instrumen berupa endorsement influencer Instagram. Menurut website Instagram sendiri 75% pengguna Instagram mengambil aksi atau keputusan setelah terinspirasi dari sebuah post influencer dan 60% pengguna Instagram mengatakan kalau mereka mengetahui produk baru dari Instagram [7]. Menurut Forbes Micro (Influencer yang mempunyai followers 1000 - 10.000) dan nano-influencer (Influencer yang mempunyai followers 10.000 - 50.000) telah mendapatkan lebih banyak perhatian akhir-akhir ini di industri pemasaran influencer, dan untuk alasan yang baik: Pembuat konten kecil-kecilan ini cenderung memiliki lebih banyak koneksi pribadi dengan pengikut mereka dan lebih cenderung dianggap otentik dalam konten dan rekomendasi merek mereka [18]. Harga yang dipatok oleh influencer - influencer pun beragam (data beragam harga influencer) jadi sangat lah susah mendapatkan atau mengetahui harga yang "worth" dengan postingan influencer tersebut. Oleh karena itu dibutuhkan sebuah platform untuk menentukan kisaran harga Dengan menentukan harga yang sesuai dengan endorsement influencer tersebut maka mulai dari small business owner sampai dengan perusahaan yang sudah berkembang akan bisa mencocokkan harga yang sudah influencer patok dengan hasil dari klasifikasi dengan decision tree melihat apakah harga yang ditetapkan oleh Influencer itu sesuai dengan data pada page influencer

2. PENELITIAN SEBELUMNYA

Berdasarkan penelitian sebelumnya tentang mencari determinan harga rumah, Decision trees menyediakan pendekatan yang efektif untuk mengidentifikasi determinan dari harga rumah[6]. Decision tree menunjukkan banyak potensi dalam banyak masalah pengenalan pola seperti multi sumber penginderaan jauh data classification [6]. Jadi penelitian ini akan menggunakan decision tree sebagai metode untuk mengklasifikasi kisaran harga endorsement.

3. DATASET

Data yang digunakan di simpan pada google sheets berjumlah total 519 data influencer yang berisikan variable-variabel yang di butuhkan. Data yang akan di analisis merupakan data yang sudah di transformasi sesuai dengan ketentuan : Harga akan di kategorikan menjadi 9 kategori. Followers akan di kategorikan menjadi 8 kategori. Engagement rate akan di kategorikan menjadi

4 kategori.. Average liker akan di kategorikan dengan mencari rasio dari membagi average liker dengan follower dari influencer itu. Average Comment akan di kategorikan dengan mencari rasio dari membagi average comment dengan follower dari influencer itu. Kemudian data akan di bagi data training 80% dan data test 20%.

4. METODE

4.1 Endorsement

Endorsement adalah orang yang menggunakan jasa endorser yang artinya adalah pendukung iklan atau bintang iklan untuk mendukung iklan produknya [16]. dalam hal ini influencer instagram lah yang menjadi alat untuk melakukan endorsement. Menurut Shimp salah satu fungsi iklan yang paling penting adalah untuk mempublikasikan merek dan pengiklan memanfaatkan nilai kepercayaan dengan memilih endorser yang secara luas dianggap sebagai orang yang jujur, dapat dipercaya, dan dapat diandalkan [15].

4.2 Influencer sebagai Alat Marketing

Influencer marketing Adalah pendekatan baru yang terpenting dalam marketing dalam dekade ini untuk para profesional pada bidang purchasing decision-making . Kata “Influence” bisa didefinisikan sebagai kekuatan untuk mempengaruhi seseorang atau sebuah acara. [3]. Brown and Hayes mendefinisikan influencer adalah orang ke 3 yang membentuk secara signifikan keputusan pembelian customer[3]. influencer marketing harus di masukan ke perusahaan atau pemilik bisnis dalam strategi marketing mereka[2]. pengguna jasa influencer harus tau budget yang dimiliki agar menuntun dalam pemilihan influencer yang tepat[19]. Berdasarkan penjelasan dari hasil penelitian, dapat ditarik kesimpulan bahwa Celebrity Endorser, memiliki pengaruh terhadap keputusan pembelian produk-produk yang dipromosikan olehnya.[13].

4.3 Decision Tree

Decision tree adalah metode umum dalam data mining digunakan untuk membangun sistem klasifikasi berdasarkan beberapa kovariat atau untuk mengembangkan algoritma prediksi untuk variabel target. Metode ini mengklasifikasikan populasi ke dalam segmen seperti cabang yang membangun pohon terbalik dengan simpul akar, simpul internal, dan simpul daun[17]. Dari banyaknya metode machine learning lainnya Decision tree merupakan machine learning yang mudah di komprehensi [10]. Pada penelitian ini akan menggunakan Classification and Regression Trees (CART) CART dimulai dari sebuah root node T1, yang mengandung semua records pada dataset. Lalu Melakukan pencarian suatu nilai yang digunakan melakukan binary splits dari T1 sampai ditemukan optimal split sopt. Untuk attributes yang bersifat kontinu, bentuk split yang dilakukan adalah mengambil sebuah nilai yang menjadi titik tengah di antara sepasang nilai unique yang berdekatan yang dimiliki oleh attributes tersebut. Setelah ditemukan sopt root node memperoleh dua keturunan node yaitu TL dan TR, yang masing-masing memiliki subset dari semua records pada dataset. Proses Binary splits ini dilakukan setiap node secara recursive sampai mencapai suatu batas yang sudah ditentukan atau records yang dimiliki oleh sebuah node hanya terdiri dari 1 class saja.[4]. Pada penelitian yang di lakukan oleh Rathan, Sai & Manikanta menggunakan Decision Tree dan Regression techniques , kemudia membandingkan kedua metode itu . Decision tree menghasilkan nilai akurasi 95,9 % [14].

4.4 Confusion Matrix

Untuk confusion matrix pengukuran performa machine learning untuk masalah klasifikasi yang outputnya bisa 2 atau lebih kelas. Confusion matrix adalah tabel dengan 4 kombinasi yang berbeda dari nilai prediksi dan nilai asli [12]

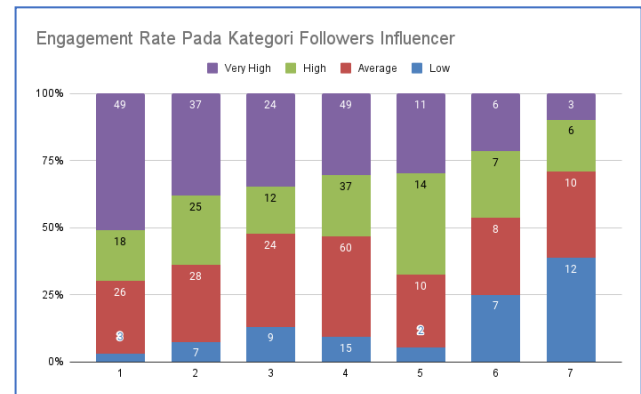
4.5 Root Mean Squared Error

RMSE adalah akar dari rata rata kuadrat dari semua kesalahan klasifikasi. Penggunaan RMSE sangat bagus untuk melihat standar deviasi dari distribusi kesalahan atau error pada model [8]. Mean Squared Error berguna tidak hanya untuk machine learning akan tetapi banyak masalah yang perlu di identifikasi presisinya [11].

5. PENGUJIAN

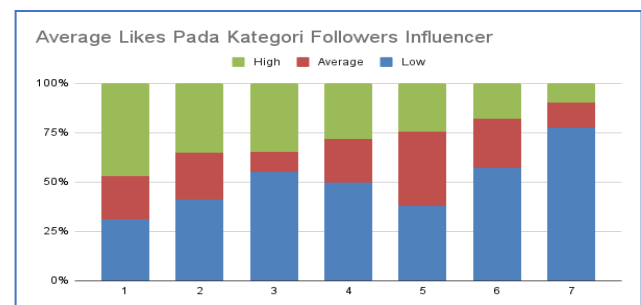
5.1 Analisis Data

Pada Gambar 1 data di transformasi di bagi dengan syarat <1% menjadi low, 1% - 3.49% menjadi average, 3.5% - 5.99% menjadi high, >=6% menjadi very high. Jika melihat pada persebaran data pada gambar 1 followers yang lebih tinggi cenderung lebih sulit untuk mendapatkan engagement rate yang lebih tinggi. Hal ini berarti influencer dengan followers yang tinggi yang jumlah followersnya sampai jutaan akan lebih sulit untuk mendapatkan engagement rate yang tinggi.



Gambar 1. Chart Persebaran Kategori Engagement Rate

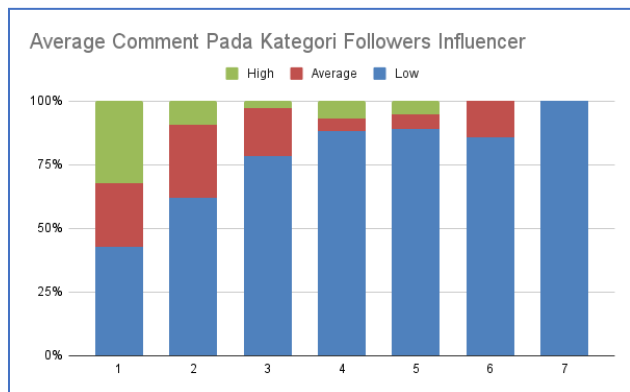
Pada Average like di Gambar 2 pun terjadi hal yang sama setelah di transformasi data followers yang lebih tinggi pun jauh lebih sulit untuk mendapatkan likes dari post yang mereka upload bahkan pada followers tinggi 75% average like mereka akan low.



Gambar 2. Chart Persebaran Kategori Average Like

Pada Gambar 3 sangat jelas bahwa pada kategori 6 dan 7 yaitu kategori dengan followers >1.000.000 akan memiliki average comment yang low. Dengan hal ini bisa di simpulkan bahwa influencer dengan followers yang jutaan akan lebih sulit untuk

mendapatkan engagement , average like , dan average comment yang tinggi.



Gambar 3. Chart Persebaran Kategori Average Comment

5.2 Pengujian Parameter pada Decision Tree dan Analisa Metode Lain

Pada sub bab ini akan di bagi menjadi 2 tuning parameter yaitu 'Gini' dan 'Entropy' kemudian 2 pembagian data training dan test yaitu 70% data training 30% data Testing dan 80% data training 20% data testing . Pengujian ini parameter dan pembagian data test, training bertujuan untuk menentukan model dengan parameter apa dan pembagian data apa yang memiliki nilai akurasi yang paling tinggi dan nilai RMSE yang paling rendah.

Tabel 1. Hasil dengan Parameter Gini

Pembagian Data	Akurasi	Recall	Precision	F-1 Score	RMSE	Waktu
70:30	45.5 %	Macro : 46.5% Weighted: 45.5%	Macro : 48.5% Weighted: 47.4%	Macro : 44.8% Weighted: 44.1%	1.16	0.036
80:20	50%	Macro : 51.9% Weighted: 50%	Macro : 53.2% Weighted: 51.1%	Macro : 51.2% Weighted: 49.2%	1.12	0.032

Pada Tabel 1 model akan di lakukan pengujian parameter yaitu gini entropy yang berasal dari library sk learn. Pengujian pada parameter gini ini di lakuakn 2 kali dengan menggunakan data set yang sudah di tentukan di atas, kemudian menggangi distribusi data lagi dan di lakukan pengujian 2 kali untuk mendapat hal yang di butuhkan yaitu Akurasi, Recall, Presicion, F-1 Score, Nilai RMSE , dan Waktu atau berapa lama model melakukan training pada dataset yang telah di berikanpengujian pada parameter gini mendapatkan hasil akurasi 45.5% pada dataset 70:30, Sedangkan pada dataset 80:20 hasilnya 4.5% lebih tinggi yaitu 50% dengan waktu 0.036 detik pada data 70:30 dan 0.032 detik pada data 80:20. Untuk hasil RMSE pada dataset 80:20 pun nilainya lebih rendah 0.04 di bandingan data yang di bagi 70:30.

Tabel 2. Hasil dengan Parameter Entropy

Pembagian Data	Akurasi	Recall	Precision	F-1 Score	RMSE	Waktu
70:30	45.5%	Macro : 46.5% Weighted: 45.5%	Macro : 48.5% Weighted: 47.4%	Macro : 44.8% Weighted: 44.1%	1.16	0.05 detik
80:20	50%	Macro : 51.9% Weighted: 50%	Macro : 54.7% Weighted: 52.3%	Macro : 51.2% Weighted: 49.2%	1.14	0.04 detik

Pada Tabel 2 pengujian pada parameter Entropy mendapatkan hasil akurasi 45.5% pada dataset 70:30, Sedangkan pada dataset 80:20 hasilnya 4.5% lebih tinggi yaitu 50% dengan waktu 0.05 detik pada data 70:30 dan 0.04 detik pada data 80:20. Untuk hasil RMSE pada dataset 80:20 pun nilainya lebih rendah 0.02 di bandingan data yang di bagi 70:30. Parameter Gini dengan data train 80% dan 20% data test merupakan mmodel terbaik dengan akurasi tertinggi , RMSE terendah , Dan waktu tercepat. Setelah menghasilkan akurasi yang masih tergolong rendah maka akan dilakukan Analisa dengan metode lainya yaitu Support Vector Machine dan MLP. Kedua metode ini masih menghasilkan akurasi yang relative rendah yaitu 44% untuk metode MLP dan 47% untuk metode Support Vector Machine.

5.3 Analisa Confusion Matrix

Yang dilakukan di sini adalah mencari TP, TN, FP dan FN untuk masing-masing kelas individu. Misalnya, jika mengambil suatu kelas, maka mari kita lihat berapa nilai matrik dari confusion matrix [9]. Pada model yang di gunakan model decision tree menghasilkan confusion matrix pada subab ini akan di Analisa hasil dari confusion matrix 10 kelas harga endorsement.

	0	1	2	3	4	5	6	7	8	9
0	8	1	0	0	0	0	0	0	0	0
1	1	5	2	1	0	0	0	0	0	0
2	1	1	4	4	0	0	0	0	0	0
3	0	0	7	7	2	0	0	0	0	0
4	0	0	0	3	6	2	1	1	0	0
5	0	0	1	1	5	1	1	0	0	0
6	0	0	0	0	2	0	7	1	0	0
7	0	0	0	1	2	0	5	0	2	0
8	0	0	0	0	1	0	0	3	1	0
9	0	0	0	0	0	0	0	0	0	5

Gambar 4. Confusion Matrix dari Model

Pada Gambar 4 yang merupakan hasil dari prediksi yang benar adalah garis diagonal dari (0,0), (1,1) sampai (9,9) . Jika di lihat sudah Sebagian besar dari prediksi berada pada titik yang benar seperti contoh harga yang di prediksi pada (0,0) mendapat 8 yang

benar prediksi dan 2 ter-prediksi pada (0,2) dan (0,1) . hal ini terjadi pada beberapa kelas akan tetapi kelas lain juga banya ayng prediksinya salah akan tetapi tidak jauh dari prediksi nya yang sebenarnya seperti pada contoh pada kelas no 4 yang terjadi prediksi benar yaitu pada titik (4,4) sebanyak 6 kali akan tetapi prediksi meleset di (4,5) sebnyak 5 kali , Hal ini lah yang membuat akurasi dari model menjadi rendah meskipun kesalahannya hanya menyimpang 1 sampai rata rata 3 dari kelas yang benar

5.4 Analisa Decision Rules

Klasifikasi menggunakan decision tree akan menghasilkan sebuah tree yang akan di sebut dengan decision rule pap sub bab ini hasil dari decision rule yang di buat oleh decision tree akan di bahas. Pada pengujian ini decision tree yang di gunakan akan adalah decision tree dari model yang terbaik yaitu model deangan pembagian data 80:20 dan menggunakan parameter gini impurity dengan akurasi 50% dan nilai RMSE 1.12. Berdasarkan text representation decision tree dapat di simpulkan bahwa atribut followers merupakan atribut yang paling berpengaruh dalam penentuan harga endorsement. Followers merupakan root dari decision tree ini dan juga followers pada decision tree ini terdapat hingga cabang ke 4. Disusul dengan atribut lainnya yaitu Engagement rate dan average like dan average comment pada cabang 4 dan cabang 5 akan tetapi lebih di dominasi oleh engagement rate dan average like pada bagian cabang ini, Yang berarti engagement rate dan average like membawa pengaruh cukup besar untuk klasifikasi. Akan tetapi jika melihat padah cabang yang lebih dalam engagement rate jauh lebih medominasi yang berarti engagement rate atribut yang lebih berpengaruh. Sedangkan untuk variable average comment merupakan variable dengan pengaruh paling sedikit pada decision tree ini.

6. KESIMPULAN

Decision Tree untuk melakukan klasifikasi pada tarif endorment influencer masih belum cocok karena akurasi yang didapatkan masi sangat rendah yaitu 50% akan tetapi untuk skor RMSE lebih bisa di andalkan yaitu 1.12 yang berarti hasil rata rata akan melenceng dari hasil asli sebesar 1.12 kategori . Untuk parameter yang di gunakan Gini Impurity merupakan parameter terbaik untuk melakukan split pada decision tree karena dengan menggunakan gini impurity proses pada data lebih cepat. Pada pembagian data 80% data training data 20% data training merupakan pembagian data dengan akurasi tertinggi yang berarti semakit banyak data akan semakin tinggi akurasi nya. Ada beberapa variable yang paling berpengaruh pada Tree yang sudah di hasilkan yang pertama merupakan followers, Jumlah followers merupakan factor utama dalam pembagian harga endorsement di susul dengan engagement rate dan kemudian average like, Sedangkan average comment merupakan atribut yang paling tidak berpengaruh. Hasil dari survei yang di lakukan menghasilkan kesalahan sebesar 30% dan 70% dari survei merupakan hasil klasifikasi yang benar, Yang berarti program ini berhasil 7 kali melakukan klasifikasi yang benar dari 10 percobaan akan tetapi untuk nilai akurasi masih tergolong rendah dan belum bisa di andalkan.

7. SARAN

Berdasarkan hasil dan berbagai pengujian yang telah dilakukan, maka saran untuk penelitian selanjutnya yaitu:

- Mencoba metode lainya seperti regression tree untuk mendapat nilai harga yang lebih tepat dan akurat .
- Mencoba beberapa fitur atau atribut lain dari API social blade seperti Grade influencer yang sudah di tentukan pada social

blade dan ada tidaknya centang biru pada akun tersebut. yang bisa membuat harga lebih akurat untuk Influencer yang merupakan Celebrity yang sudah ter verified.

- Menambahkan data influencer lebih banyak karena pada data training 80:20 akurasi menjadi lebh tinggi di bandingkan dengan data training yang terbagi 70:30. Kemudian lebih membuat pembagian kategori lebih banyak dan rata jumlahnya pada dataset dengan melakukan wawancara lagi terhadap social media expert dalam hal kategori influencer dan harga.
- Membuat rasio Low , Medium , dan High pada atribut seperti engagement rate untuk membuat pembagian data lebih adil pada influencer dengan follower tinggi dan juga embuat rasio average like dan average commet yang lebih adil untuk influencer dengan followers tinggi agar data transformasi lebih tersebar rata untuk low , medium ,dan high.

8. REFERENSI

- [1] APJII. 2018. Penetrasi dan Profil Perilaku Pengguna Internet Indonesia. Laporan Survei. Jakarta: Indonesian Internet Service Provider Association
- [2] Biaudet, S. (2017). Influencer Marketing as a Marketing Tool The process of creating an Influencer Marketing Campaign on Instagram
- [3] Brown, D., & Hayes, N. (2008). Influencer marketing. Amsterdam: Elsevier/Butterworth-Heinemann
- [4] Crawford, S. L. (1989). Extensions to the CART algorithm. International Journal of Man-MachineStudies,31(2),197-217
- [5] Dwivedi, Y., Ismagilova, E., Hughes, D., Carlson, J., Filieri, R., & Jacobson, J. et al. (2020). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal Of Information Management, 102168. doi: 10.1016/j.ijinfomgt.2020.102168.
- [6] Fan, G., Ong, S., & Koh, H. (2006). Determinants of House Price: A Decision Tree Approach. Urban Studies, 43(12), 2301-2315. doi: 10.1080/00420980600990928
- [7] Instagram.com. Why Instagram? Available from: <https://business.instagram.com/getting-started/> Accessed: 30.08.2021
- [8] Moody, j. (2022). What does RMSE really mean?. Retrieved 29 May 2022, from <https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e>
- [9] Mohajon, J. (2021). Confusion Matrix for Your Multi-Class Machine Learning Model. Retrieved 2 September 2021, from <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-m> Murphy, A., & Moore, C. (2019). Mean squared error. Radiopaedia.Org. doi: 10.53347/rid-68955achine-learning-model-ff9aa3bf7826
- [10] Murphy, A. (2017). Decision tree (machine learning). Radiopaedia.Org. doi: 10.53347/rid-57104
- [11] Murphy, A., & Moore, C. (2019). Mean squared error. Radiopaedia.Org. doi: 10.53347/rid-68955
- [12] Narkhede, S. (2021). Understanding Confusion Matrix. Retrieved 2 September 2021, from <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- [13] . Ramadhan, A., Nadia Naswandi, C., & Maharani Herman, C. (2020). FENOMENA ENDORSEMENT DI INSTAGRAM

STORY PADA KALANGAN SELEBGRAM. doi:
<https://doi.org/10.31947/kareba.vi.8207>

- [14] Rathan, K., Sai, S., & Manikanta, T. (2019). Crypto-Currency price prediction using Decision Tree and Regression techniques. 2019 3Rd International Conference On Trends In Electronics And Informatics (ICOEI). doi: 10.1109/icoei.2019.8862585
- [15] Shimp, T. (1993). *Transparencies to Accompany "Promotion Management and Marketing Communication"*. Canada: Nelson Education, Ltd
- [16] Shimp, T. (2010). *Advertising Promotion and Other Aspects of Integrated Marketing Communication*. Canada: Nelson Education, Ltd.
- [17] Song, Y. Y., & Lu, Y. (2015). *Decision tree methods: applications for classification and prediction*. Shanghai archives of psychiatry, 27(2), 130–135. <https://doi.org/10.11919/j.issn.1002-0829.215044>
- [18] Wiley, D. (2021). Council Post: Choosing The Right Influencer (And One Major Misstep To Avoid). Retrieved 21 September 2021, from <https://www.forbes.com/sites/forbesagencycouncil/2020/12/17/choosing-the-right-influencer-and-one-major-misstep-to-avoid/?sh=aae5e0b31b2d>
- [19] Vicuña, X. (2021). Council Post: Choosing The Right Influencers: The Metrics That Matter. Retrieved 22 September 2021, from <https://www.forbes.com/sites/forbesbusinesscouncil/2020/12/09/choosing-the-right-influencers-the-metrics-that-matter/?sh=78983ae6709a>