

BAB II

TINJAUAN PUSTAKA

2.1 Justifikasi Solusi

Berdasarkan hasil wawancara dengan Viicesaa Agency, proses seleksi influencer dilakukan melalui beberapa tahapan, mulai dari pengumpulan data kandidat, penyaringan awal berdasarkan informasi pada formulir pendaftaran, pemeriksaan profil media sosial, hingga pengambilan keputusan oleh stakeholder. Dalam dataset penelitian terdapat 2.385 kandidat influencer, sedangkan keseluruhan proses seleksi membutuhkan waktu sekitar 48 jam [2]. Waktu tersebut mencakup pengolahan data terstruktur, pemeriksaan profil Instagram dan TikTok, evaluasi kualitas konten, kesesuaian niche, karakteristik audiens, kebutuhan campaign, serta pengambilan keputusan stakeholder. Oleh karena itu, waktu sekitar 48 jam digunakan sebagai konteks permasalahan proses bisnis dan tidak dibandingkan secara langsung dengan waktu pemrosesan teknis sistem.

Justifikasi pengembangan solusi dalam penelitian ini didasarkan pada beberapa penelitian dan sumber ilmiah yang membahas efektivitas influencer marketing, pemanfaatan data dalam pemilihan influencer, supervised learning, dan Gradient Boosting.

2.1.1 Justifikasi Berdasarkan Penelitian Leung dkk.

Leung dkk. dalam penelitian berjudul *Influencer Marketing Effectiveness* menjelaskan bahwa efektivitas influencer marketing tidak hanya dipengaruhi oleh jumlah pengikut influencer. Faktor lain, seperti keterlibatan audiens, karakteristik influencer, dan kesesuaian antara influencer dengan brand, juga berperan dalam pelaksanaan influencer marketing [1].

Penelitian tersebut menunjukkan bahwa proses pemilihan influencer tidak seharusnya hanya dilakukan berdasarkan satu parameter, seperti jumlah followers tertinggi. Dalam konteks penelitian ini, temuan tersebut menjadi dasar bahwa pengurutan sederhana menggunakan spreadsheet belum sepenuhnya merepresentasikan pertimbangan stakeholder dalam memilih kandidat.

Penelitian ini menggunakan jumlah followers dan engagement rate Instagram serta TikTok sebagai fitur karena data tersebut tersedia secara konsisten pada formulir pendaftaran kandidat. Namun, hasil sistem tidak digunakan sebagai satu-satunya dasar keputusan. Faktor kualitatif seperti kualitas konten, niche, karakteristik audiens, dan kesesuaian dengan campaign tetap diperiksa secara manual oleh stakeholder.

2.1.2 Justifikasi Berdasarkan Penelitian Jha dan Ray

Jha dan Ray mengembangkan *Influencer Recommendation System* menggunakan pendekatan analisis jaringan untuk membantu proses pemilihan influencer yang sesuai. Penelitian tersebut menunjukkan bahwa data mengenai influencer dapat diolah secara komputasional untuk memberikan informasi pendukung dalam proses pemilihan influencer [4].

Relevansi penelitian Jha dan Ray terhadap penelitian ini terletak pada penggunaan pendekatan berbasis data untuk mendukung proses pengambilan keputusan. Pendekatan tersebut menunjukkan bahwa pemilihan influencer tidak harus sepenuhnya dilakukan melalui pemeriksaan manual, tetapi dapat dibantu oleh sistem yang mengolah karakteristik kandidat.

Meskipun demikian, penelitian ini tidak menghasilkan rekomendasi dalam bentuk ranking maupun menentukan influencer terbaik. Sistem yang dikembangkan hanya menghasilkan prediksi kelas dan probabilitas kandidat termasuk ke dalam kelas yang sebelumnya dipilih berdasarkan pola keputusan historis stakeholder Viicesaa Agency.

2.1.3 Justifikasi Berdasarkan Penelitian Gan dkk.

Gan dkk. dalam penelitian berjudul *Seeking Micro-Influencers for Brand Promotion* membahas proses pencarian micro-influencer untuk kebutuhan promosi brand dengan memanfaatkan karakteristik kandidat yang tersedia pada data [5]. Penelitian tersebut menunjukkan bahwa informasi kandidat dapat digunakan untuk mendukung proses identifikasi influencer yang sesuai dengan tujuan promosi.

Penelitian Gan dkk. menjadi dasar bahwa pendekatan berbasis data dapat digunakan untuk membantu proses evaluasi kandidat influencer. Penelitian ini mengadaptasi konsep tersebut dengan menggunakan data internal agency berupa jumlah followers dan engagement rate Instagram serta TikTok.

Perbedaannya adalah penelitian ini tidak secara khusus mencari micro-influencer dan tidak menghasilkan ranking kandidat untuk kebutuhan promosi. Sistem mempelajari keputusan historis berupa kandidat dipilih dan tidak dipilih, kemudian menghasilkan probabilitas kandidat termasuk ke dalam kelas dipilih berdasarkan fitur yang tersedia.

2.1.4 Justifikasi Berdasarkan Hastie dkk.

Hastie dkk. menjelaskan bahwa pada pendekatan supervised learning, model dilatih menggunakan data yang telah memiliki fitur masukan dan target keluaran. Model kemudian mempelajari hubungan antara fitur dan target tersebut untuk menghasilkan prediksi terhadap data baru [3].

Pendekatan supervised learning sesuai digunakan dalam penelitian ini karena dataset Viicesaa Agency telah memiliki label keputusan historis stakeholder, yaitu:

1. kelas 1 untuk kandidat yang dipilih; dan
2. kelas 0 untuk kandidat yang tidak dipilih.

Model mempelajari hubungan antara jumlah followers dan engagement rate Instagram serta TikTok dengan label keputusan historis tersebut. Dengan demikian, keluaran model menunjukkan probabilitas kandidat termasuk ke dalam kelas yang sebelumnya dipilih oleh stakeholder.

Probabilitas tersebut tidak menunjukkan kemungkinan keberhasilan campaign. Nilai yang dihasilkan hanya merepresentasikan pola yang dipelajari dari keputusan historis stakeholder berdasarkan fitur yang tersedia dalam dataset.

2.1.5 Justifikasi Berdasarkan Penelitian Friedman

Friedman memperkenalkan Gradient Boosting sebagai metode yang membangun model secara bertahap. Pada setiap tahap, model baru ditambahkan untuk memperbaiki kesalahan prediksi dari model sebelumnya sehingga terbentuk model gabungan [6].

Konsep Gradient Boosting relevan terhadap penelitian ini karena hubungan antara fitur kandidat dan keputusan stakeholder belum tentu dapat dijelaskan menggunakan aturan linear atau satu batas nilai tertentu. Metode berbasis boosting dapat mempelajari hubungan yang lebih kompleks antara beberapa fitur numerik dan target klasifikasi.

HistGradientBoosting digunakan sebagai model akhir setelah dilakukan perbandingan terhadap Logistic Regression, Random Forest, Extra Trees, HistGradientBoosting, Support Vector Machine dengan kernel RBF, dan XGBoost. Namun, keputusan pemilihan model akhir tidak hanya didasarkan pada teori Gradient Boosting. Pemilihan tersebut dilakukan berdasarkan hasil training, validation, hyperparameter tuning, dan final testing yang dijelaskan pada Bab IV.

2.1.6 Sintesis Justifikasi Solusi

Berdasarkan kondisi proses bisnis stakeholder dan penelitian terdahulu, solusi yang dikembangkan adalah sistem pendukung seleksi awal influencer berbasis Machine Learning yang diintegrasikan dengan website Streamlit.

Machine Learning digunakan karena dataset memiliki label keputusan historis yang dapat dipelajari menggunakan supervised learning. Model tidak hanya mengurutkan kandidat berdasarkan satu nilai, tetapi mempelajari hubungan antara kombinasi jumlah followers dan engagement rate Instagram serta TikTok dengan keputusan historis stakeholder.

Website Streamlit digunakan agar stakeholder dapat mengunggah data kandidat, melakukan validasi format data, menjalankan preprocessing, dan

memperoleh hasil prediksi tanpa perlu menjalankan kode pemrograman secara langsung. Sistem menghasilkan:

1. prediksi kelas dipilih atau tidak dipilih; dan
2. probabilitas kandidat termasuk ke dalam kelas dipilih berdasarkan pola historis stakeholder.

Sistem tidak menghasilkan ranking, tidak menentukan influencer terbaik, dan tidak memprediksi keberhasilan campaign. Hasil prediksi hanya digunakan sebagai informasi pendukung sebelum stakeholder melakukan pemeriksaan kualitatif dan menentukan keputusan akhir.

Pengujian waktu dalam penelitian ini digunakan untuk mengukur durasi pemrosesan teknis sistem yang mencakup validasi data, preprocessing, dan prediksi model. Pengujian tersebut tidak mencakup pemeriksaan profil media sosial, evaluasi kualitas konten, kesesuaian niche, karakteristik audiens, komunikasi dengan kandidat, maupun pengambilan keputusan akhir oleh stakeholder.

Dengan demikian, solusi yang dikembangkan tidak bertujuan menggantikan keseluruhan proses seleksi influencer. Sistem digunakan untuk mengotomatisasi pengolahan data terstruktur dan menyediakan informasi prediksi awal berdasarkan pola keputusan historis stakeholder.

2.2 Tinjauan Teori

2.2.1 Influencer Marketing

Influencer marketing merupakan strategi pemasaran yang memanfaatkan individu dengan pengaruh tertentu pada komunitas atau audiens digital untuk menyampaikan pesan pemasaran suatu brand. Efektivitas influencer marketing

tidak hanya dipengaruhi oleh jumlah pengikut, tetapi juga oleh kesesuaian influencer dengan brand, tingkat interaksi audiens, serta karakteristik komunitas yang dimiliki influencer [1].

Dalam proses influencer marketing, pemilihan influencer menjadi salah satu tahap penting karena influencer yang dipilih harus sesuai dengan tujuan campaign. Agency biasanya mempertimbangkan beberapa faktor seperti jumlah followers, engagement rate, kualitas konten, kesesuaian niche, serta karakteristik audiens. Namun, tidak seluruh faktor tersebut selalu tersedia dalam bentuk data terstruktur sehingga proses pemilihan masih banyak melibatkan evaluasi manual dari pihak agency [2].

2.2.2 Sistem Pendukung Keputusan

Sistem pendukung keputusan merupakan sistem berbasis komputer yang dirancang untuk membantu pengguna dalam mengambil keputusan berdasarkan data dan informasi yang tersedia. Sistem ini tidak menggantikan keputusan manusia, tetapi menyediakan informasi tambahan agar pengguna dapat membuat keputusan dengan lebih efektif.

Pada penelitian ini, sistem pendukung keputusan digunakan untuk membantu tahap penyaringan awal kandidat influencer. Sistem memanfaatkan data historis keputusan stakeholder untuk memberikan prediksi kandidat berdasarkan pola yang telah dipelajari oleh model Machine Learning.

2.2.3 Machine Learning

Machine Learning merupakan pendekatan komputasi yang memungkinkan sistem mempelajari pola dari data tanpa harus diberikan aturan secara eksplisit.

Pada pendekatan supervised learning, model dilatih menggunakan data yang telah memiliki pasangan antara fitur masukan dan target keluaran sehingga model dapat mempelajari hubungan antara keduanya [3].

Secara umum, proses supervised learning dapat digambarkan sebagai proses mencari fungsi:

$$f(X) = Y$$

dengan:

1. X merupakan fitur masukan,
2. Y merupakan target atau label,
3. f merupakan model yang dipelajari dari data.

Pada penelitian ini, fitur masukan berupa:

1. followers Instagram,
2. engagement rate Instagram,
3. followers TikTok,
4. engagement rate TikTok.

Sedangkan target yang diprediksi adalah label keputusan stakeholder:

1. 1 = kandidat dipilih,
2. 0 = kandidat tidak dipilih.

2.2.4 Gradient Boosting

Gradient Boosting merupakan metode ensemble learning yang membangun model secara bertahap dengan menggabungkan beberapa weak learner, umumnya berupa decision tree. Berbeda dengan metode ensemble yang membangun model secara independen, Gradient Boosting membuat model baru untuk memperbaiki kesalahan prediksi dari model sebelumnya [6].

Konsep dasar Gradient Boosting adalah membangun model akhir:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x)$$

dengan:

1. $F(x)$ = model prediksi akhir,
2. M = jumlah iterasi boosting,
3. $h_m(x)$ = model pembelajaran ke- m ,
4. γ_m = bobot model.

Pada setiap iterasi, model baru dibuat berdasarkan residual atau kesalahan prediksi model sebelumnya. Tujuannya adalah mengurangi nilai error sehingga model akhir memiliki kemampuan prediksi yang lebih baik.

2.2.5 HistGradientBoostingClassifier

HistGradientBoosting Classifier merupakan pengembangan dari metode Gradient Boosting yang menggunakan pendekatan histogram untuk meningkatkan efisiensi proses pelatihan. Pada metode ini, nilai fitur dikelompokkan ke dalam beberapa interval histogram sebelum proses pembentukan decision tree dilakukan.

Pendekatan histogram membuat proses pencarian split pada decision tree menjadi lebih cepat dibandingkan metode Gradient Boosting tradisional, terutama pada dataset dengan jumlah data yang besar. Penggunaan HistGradientBoosting sebagai model akhir dalam penelitian ini ditentukan berdasarkan hasil evaluasi empiris pada data validation dan final testing, bukan hanya berdasarkan karakteristik teoritis algoritma.

Pada penelitian ini, HistGradientBoosting digunakan karena dataset berbentuk tabel dengan fitur numerik dan memiliki target klasifikasi biner. Model ini digunakan untuk mempelajari pola keputusan historis stakeholder dan menghasilkan probabilitas kandidat masuk ke kelas yang dipilih.

Probabilitas keluaran model dihitung berdasarkan kemungkinan suatu data termasuk dalam kelas tertentu:

$$P(Y = 1 | X)$$

dengan:

1. $P(Y = 1 | X)$ = probabilitas kandidat termasuk kelas dipilih,
2. X = fitur kandidat.

Nilai probabilitas tersebut bukan merupakan probabilitas keberhasilan campaign. Nilai tersebut menunjukkan estimasi probabilitas kandidat termasuk ke dalam kelas dipilih berdasarkan pola keputusan historis stakeholder dan fitur yang tersedia dalam dataset.

2.2.6 Evaluasi Model Klasifikasi

Evaluasi model dilakukan untuk mengetahui kemampuan model dalam memprediksi label berdasarkan data pengujian yang belum digunakan saat pelatihan. Pada penelitian ini digunakan beberapa metrik evaluasi klasifikasi[7].

Penggunaan beberapa metrik diperlukan karena nilai accuracy dapat dipengaruhi oleh distribusi kelas. Oleh karena itu, evaluasi penelitian ini tidak hanya menggunakan accuracy, tetapi juga balanced accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, dan Matthews Correlation Coefficient untuk memberikan gambaran performa model yang lebih menyeluruh [7], [8].

Accuracy

Accuracy menunjukkan proporsi prediksi benar dibandingkan seluruh jumlah data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:

1. TP = True Positive
2. TN = True Negative
3. FP = False Positive
4. FN = False Negative

Precision

Precision menunjukkan tingkat ketepatan prediksi positif yang diberikan oleh model.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall menunjukkan kemampuan model dalam menemukan seluruh data positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN}$$

F1-score

F1-score merupakan rata-rata harmonik antara precision dan recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC-AUC

ROC-AUC digunakan untuk mengukur kemampuan model dalam membedakan dua kelas berdasarkan probabilitas prediksi. ROC menggambarkan hubungan antara True Positive Rate dan False Positive Rate pada berbagai threshold keputusan [8].

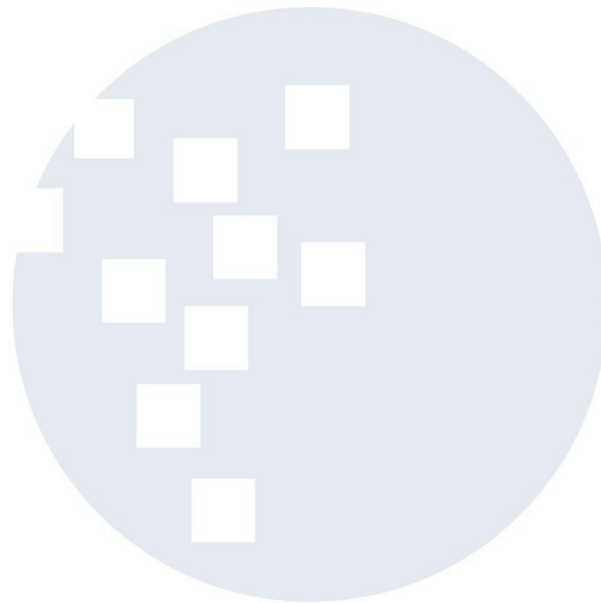
True Positive Rate dihitung sebagai:

$$TPR = \frac{TP}{TP + FN}$$

Sedangkan False Positive Rate:

$$FPR = \frac{FP}{FP + TN}$$

Nilai AUC menunjukkan luas area di bawah kurva ROC. Semakin besar nilai AUC, semakin baik kemampuan model dalam membedakan kelas positif dan negatif.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA