Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

ANALISIS BRAND LAYANAN AKADEMIK PERGURUAN TINGGI INDONESIA MENGGUNAKAN KLASIFIKASI TEKS DI MEDIA SOSIAL

Hashri Hayati [1], Muhammad Riza Alifi [2]

[1, 2] Politeknik Negeri Bandung
Jalan Gegerkalong Hilir, Desa Ciwaruga, Kec. Parongpong, Kab. Bandung Barat, 40559
[1]hashri.hayati@polban.ac.id, [2,]muhammad.riza@polban.ac.id*
**corresponding author*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis persepsi komunitas eksternal terhadap brand akademik perguruan tinggi di Indonesia melalui media sosial, khususnya Twitter/X. Seiring dengan tingginya jumlah perguruan tinggi dan angka partisipasi kasar (APK), kompetisi antar institusi pendidikan tinggi semakin kuat, mendorong perlunya diferensiasi *brand* yang disampaikan ke publik. Dalam studi ini, dikumpulkan post dari 30 akun resmi X perguruan tinggi di Indonesia yang kemudian diklasifikasikan ke dalam lima kategori brand akademik: *Innovative*, *Global Impact*, *Student Engagement*, *Career Focused*, dan *Research Excellent*. Proses klasifikasi dilakukan dengan membangun model pembelajaran menggunakan algoritma Naïve Bayes, yang diimplementasikan melalui pustaka pemrosesan bahasa alami di lingkungan Node.js. Untuk mengevaluasi kinerja model, dilakukan pengujian terhadap dataset uji terpisah, dan dihitung metrik evaluasi berupa *precision*, *recall*, dan *accuracy* berdasarkan nilai *True Positive*, *False Positive*, dan *False Negative* yang diperoleh melalui *confusion matrix* untuk setiap kelas. Hasil evaluasi menunjukkan bahwa model yang dikembangkan memiliki performa nilai rata-rata *precision* sebesar 80,8%, *recall* sebesar 78,8%, dan *accuracy* sebesar 80%, sehingga dapat diandalkan sebagai alat bantu untuk memahami kesesuaian antara brand yang dikomunikasikan dan persepsi publik secara daring.

***Kata Kunci*— brand akademik, brand perguruan tinggi, klasifikasi teks, naïve bayes, media sosial.**

ABSTRAKT

This study aims to analyze the perceptions of external communities regarding the academic branding of Indonesian universities through social media, particularly Twitter/X. With the growing number of higher education institutions and rising gross enrollment rates, competition among universities has intensified—prompting the need for more distinct and strategic public brand positioning. In this study, posts were collected from 30 official university X accounts in Indonesia and categorized into five academic brand themes: Innovative, Global Impact, Student Engagement, Career Focused, and Research Excellent. The classification process involved building a supervised machine learning model using the Naïve Bayes algorithm, implemented with a natural language processing library in the Node.js environment. To evaluate the model's performance, a separate test dataset was used, and evaluation metrics—namely precision, recall, and accuracy—were calculated for each class based on values of True Positive, False Positive, and False Negative derived from a confusion matrix. The results indicate that the developed model performs well, achieving average scores of 80,8% for precision, 78,8% for recall, and 80% for accuracy, making it a reliable tool for assessing the alignment between institutional brand communication and public perception in online discourse.

Keywords—academic brand, university brand, text classification, naïve bayes, social media.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

I. PENDAHULUAN

Sebagai negara berkembang, Indonesia mengalami pertumbuhan di berbagai sektor. Dalam sektor pendidikan, salah satu bentuk pertumbuhan tersebut terlihat dari jumlah perguruan tinggi yang beroperasi. Menurut Buku Statistik Perguruan Tinggi Indonesia tahun 2023, jumlah perguruan tinggi di Indonesia mencapai 3.107 di tahun 2022. Selain itu, dari tahun ke tahun angka partisipasi kasar (APK) perguruan tinggi mengalami peningkatan secara nasional, dari 34,58% pada tahun 2018 menjadi 40,88% pada 2023 [1].

Tingginya jumlah perguruan tinggi dan APK dapat memicu persaingan antar perguruan tinggi. Untuk menghadapi persaingan di tingkat nasional maupun internasional, perguruan tinggi di seluruh dunia mulai mencari keunikan yang dapat membedakan mereka dari institusi lainnya, guna menarik perhatian calon mahasiswa dan tenaga akademik [2]. Istilah seperti *branding*, komunikasi korporat, identitas, dan reputasi mulai muncul dalam ranah akademik. Hal ini membuat perguruan tinggi semakin menyadari pentingnya hubungan antara nilai dan karakteristik yang mereka bawa dengan bagaimana hal tersebut dipersepsikan [2].

Dalam industri jasa seperti perguruan tinggi, banyak pihak percaya bahwa organisasi adalah sebuah *brand*. Model *Service-Brand-Relationship-Value (SBRV)* oleh Brodie, Glynn, dan Little [3] menekankan pentingnya brand dalam konteks universitas, dengan fokus pada proses dan janji pemasaran yang menjadi dasar dari *brand* jasa. Pringle dan Naidoo [4] meneliti *brand* universitas dari persepsi para pegawai (dosen), yang secara langsung bertanggung jawab dalam menyampaikan janji brand kepada konsumen (mahasiswa). Studi tersebut menemukan bahwa meskipun staf dosen mengomentari adanya tekanan antara klaim *brand* dan kenyataan, mereka tidak meneliti bagaimana komunitas eksternal memandang *brand* mereka [4].

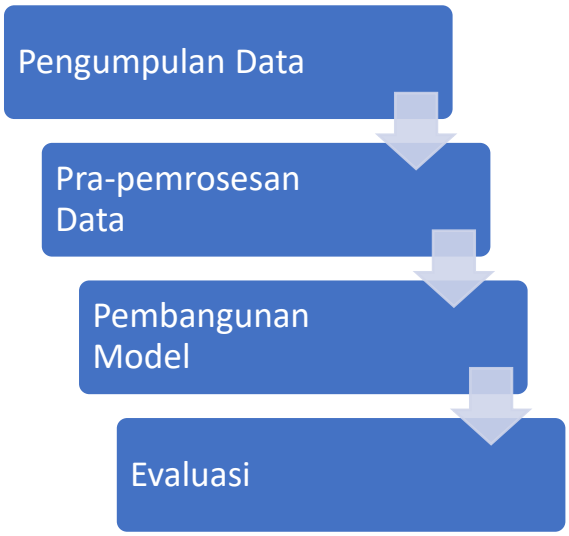
Salah satu media untuk menyampaikan *brand* kepada komunitas eksternal adalah melalui media sosial. Beberapa studi menyebutkan bahwa demi kepentingan *branding*, banyak perguruan tinggi beralih ke saluran media sosial seperti Facebook dan Twitter/X [4]. Bahkan, Rutter menyatakan bahwa Twitter/X dapat menjadi proksi kekuatan *brand* dan reputasi universitas [5]. Perguruan tinggi banyak memanfaatkan media sosial untuk mengukur *brand engagement* mereka [6]. Selain itu, menjaga hubungan dengan konsumen yang mendukung *brand* merupakan kunci keberhasilan penggunaan media sosial. Strategi media sosial yang konsisten dan relevan sangat penting dalam membangun *brand awareness* dan *engagement* di lingkungan kampus [7].

Berdasarkan penelitian [8], *brand* institusi berdampak positif pada minat, sehingga juga berdampak positif pada keputusan memilih. Dengan kata lain, semakin baik *brand* institusi, semakin besar kemungkinan seseorang memilih untuk bergabung dengan institusi tersebut [8]. Hal ini juga sejalan dengan hasil penelitian [9] yang menyatakan bahwa *Cyber Public Relation* memiliki pengaruh cukup kuat terhadap *brand* perguruan tinggi, yang didukung dengan nilai determinasi sebesar 42,8%.

Kebenaran dari janji *brand* sama pentingnya dengan komunikasi yang dilakukan secara rutin dan konsisten [4]. Selain itu, penting bagi janji *brand* ini untuk dapat dirasakan secara internal [4] maupun eksternal [10]. Penelitian ini akan berfokus pada bagaimana komunitas eksternal memandang *brand* dari perguruan tinggi melalui analisis media sosial. Analisis media sosial dilakukan dengan klasifikasi teks terhadap unggahan media sosial dari beberapa perguruan tinggi di Indonesia. Tujuan dari sistem klasifikasi teks ini adalah untuk menentukan kategori dokumen, yang mana kategorinya telah didefinisikan sebelumnya [11]. Melalui klasifikasi teks ini akan diperoleh kecenderungan *brand* layanan akademik berdasarkan konten media sosial perguruan tinggi. Informasi ini dapat digunakan sebagai bahan evaluasi bagi perguruan tinggi dalam menilai apakah persepsi eksternal sudah sesuai dengan *brand* yang telah dirancang oleh institusi.

II. METODE PENELITIAN

Penelitian ini dilakukan melalui empat tahapan utama mulai dari pengumpulan data sampai dengan evaluasi, sebagaimana digambarkan pada gambar 1:



Gambar 1. Metode Penelitian

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

A. Pengumpulan Data

Tahap ini bertujuan untuk mengumpulkan data yang akan digunakan dalam pembangunan dan pengujian model klasifikasi. Dalam penelitian ini, data diperoleh dari media sosial Twitter/X. X dipilih sebagai sumber data karena bersifat publik dan merupakan salah satu platform media sosial yang paling banyak digunakan. Data yang diklasifikasikan dalam studi ini berupa *post* dari akun resmi perguruan tinggi di Indonesia.

Untuk memperoleh data secara efisien dan terstruktur, proses pengambilan dilakukan dengan memanfaatkan X API dengan menggunakan Javascript pada *environment* Node.js. Data yang diperoleh melalui API kemudian disimpan ke dalam database MongoDB. Pemilihan MongoDB sebagai sistem penyimpanan dilakukan karena kemampuannya dalam mengelola data semi-struktur seperti dokumen JSON, serta skalabilitasnya dalam menangani volume data dalam jumlah besar.

B. Pra-pemrosesan Data

Tahap ini bertujuan untuk mengumpulkan data yang akan digunakan dalam pembangunan dan pengujian model klasifikasi. Dalam penelitian ini, data diperoleh dari media sosial X. X dipilih sebagai sumber data karena bersifat publik dan merupakan salah satu platform media sosial yang paling banyak digunakan. Data yang diklasifikasikan dalam studi ini berupa *post* dari akun resmi perguruan tinggi di Indonesia. Pada penelitian ini dilakukan 6 jenis para-pemrosesan, yaitu: mengubah semua huruf menjadi huruf kecil atau huruf besar; menghapus tanda baca, tanda aksen, dan diakritik lainnya; menghapus *white space* berlebih; memanjangkan singkatan; menghapus *stopword*, kata yang jarang muncul, dan kata tertentu; dan seleksi fitur.

C. Pembangunan Model Klasifikasi

Data yang telah melewati tahapan pra-pemrosesan siap digunakan untuk pengembangan dan pengujian model klasifikasi. Penelitian ini membangun model pembelajaran menggunakan pendekatan *machine learning*. Terdapat dua jenis utama dalam *machine learning*, yakni *supervised learning* dan *unsupervised learning*. Studi ini menggunakan pendekatan *supervised learning*, yaitu pembelajaran dengan menyediakan data berlabel oleh pengembang [12]. *Supervised learning* digunakan untuk melakukan klasifikasi, yaitu mengelompokkan dokumen ke dalam kelas-kelas yang telah ditentukan. Dalam penelitian ini, klasifikasi dilakukan menggunakan algoritma Naïve Bayes. Model Naïve Bayes digunakan untuk memprediksi kategori tweet (*post)* dari perguruan tinggi ke dalam enam kelas brand layanan akademik sebagaimana ditetapkan oleh [13], yakni: *Innovative*, *Global Impact*, *Student Engagement*, *Career Focused*, *Research Excellent*, dan *Diversity*.

Untuk membangun model klasifikasi tersebut, penelitian ini menggunakan pustaka natural, yaitu sebuah modul pemrosesan bahasa alami (*Natural Language Processing*/NLP) yang tersedia di lingkungan Node.js. Package natural menyediakan berbagai alat untuk klasifikasi teks, termasuk implementasi Naïve Bayes Classifier yang digunakan dalam penelitian ini. Modul ini diinstal dan dikelola melalui *Node Package Manager* (NPM), sehingga memudahkan integrasi dengan *pipeline* yang sebelumnya telah dibangun untuk pengambilan dan penyimpanan data dari X. Penggunaan pustaka natural memberikan efisiensi dalam proses pelatihan dan pengujian model karena kompatibel dengan format data JSON dari MongoDB.

D. Evaluasi

Evaluasi dilakukan dengan menghitung nilai *precision*, *recall*, dan *accuracy* sebagaimana dirumuskan oleh Junker [14] seperti yang ditunjukkan pada Tabel I. *Precision* mengukur proporsi prediksi positif yang benar-benar relevan. *Recall* mengukur seberapa banyak data positif yang berhasil dikenali oleh model. *Accuracy* adalah ukuran keseluruhan dari prediksi yang benar dibandingkan dengan semua data.

Metrik evaluasi diperoleh dari nilai pada *confusion matrix*. *Confusion matrix* adalah sebuah tabel yang digunakan untuk mengevaluasi performa dari algoritma klasifikasi, khususnya pada *supervised learning*. Tabel ini membandingkan antara hasil prediksi model dengan label sebenarnya *(ground truth)*, seperti diilustrasikan pada Gbr. 2. Ketiga metrik ini digunakan untuk menilai kinerja model klasifikasi teks yang dihasilkan. Pada penelitian ini terdapat 5 kelas klasifikasi sehingga akan dihasilkan nilai *precision* dan *recall* untuk setiap kelas dan nilai *accuracy* keseluruhan.

Tabel I
Formula Evaluasi

No.	Metrik Evaluasi	Formula
1	*Precision*	$\frac{\sum TP}{\sum TP + FP}$
2	*Recall*	$\frac{\sum TP}{\sum TP + FN}$
3	*Accuracy*	$\frac{\sum TP + TN}{\sum TP + FP + FN + TN}$

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

		Real Label: Positive	Real Label: Negative
Predicted Label	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Gambar 2. Confusion Matrix

III. HASIL DAN PEMBAHASAN

A. *Pengumpulan Data*

Pengumpulan data dilakukan dengan melakukan *crawling* terhadap akun X dari 30 perguruan tinggi di Indonesia yang terdiri dari 20 perguruan tinggi non-vokasional dan 10 perguruan tinggi vokasi yang memiliki akun X resmi. *Post* yang diambil berjumlah sekitar 400 *post* dari masing-masing akun resmi perguruan tinggi, sehingga secara total diperoleh 11.492 data.

B. *Pra-pemrosesan Data*

Berikut adalah tahapan pra-pemrosesan data yang dilakukan pada penelitian ini:

1) Mengubah semua huruf menjadi huruf kecil atau huruf besar: Kata yang ditulis dengan huruf kecil atau huruf besar memiliki makna yang sama. Agar sistem dapat mengenali bahwa "Pendidikan Tinggi" dan "pendidikan tinggi" mewakili entitas yang sama, maka seluruh teks diubah menjadi huruf kecil.

2) Menghapus tanda baca, tanda aksen, dan diakritik lainnya: Tanda baca, tanda aksen, dan diakritik lainnya perlu dihapus agar kata yang sama tetap dikenali meskipun berbeda dalam penulisan.

3) Menghapus white space berlebih: Pembersihan teks dari karakter spasi yang tidak diperlukan, seperti spasi ganda, tabulasi (\t), baris baru (\n), *carriage return* (\r), serta spasi di awal dan akhir kalimat.

4) Memanjangkan singkatan: Karena jumlah karakter dalam *post* terbatas, pengguna sering menggunakan singkatan. Agar sistem dapat mengenali entitas yang sama, semua singkatan yang relevan diekspansi, misalnya ITB menjadi Institut Teknologi Bandung.

5) Menghapus stopword, kata yang jarang muncul, dan kata tertentu: Penghapusan *stopword* dilakukan untuk menghapus kata-kata yang sering digunakan namun tidak secara langsung berhubungan dengan makna yang mempengaruhi klasifikasi, seperti "yang", "di", "ke", "ini", dan lain-lain.

6) Seleksi Fitur: Seleksi fitur adalah proses pemilihan fitur atau komponen data yang akan digunakan dalam pembentukan model. Dalam data teks dari media sosial seperti X, terdapat beberapa fitur yang dapat diambil, antara lain gambar profil pengguna, isi *post*, jumlah komentar, jumlah suka (*likes*), URL, nama pengguna, tanggal, dan *hashtag*. Penelitian ini tidak menggunakan fitur gambar profil, jumlah komentar, dan jumlah suka karena fokus penelitian adalah pada konten *post* yang dikeluarkan oleh perguruan tinggi.

C. *Model Pembelajaran (Naïve Bayes)*

Model pembelajaran dibangun menggunakan algoritma *Naïve Bayes*. Proses pengembangan model dilakukan dengan membagi data yang telah diperoleh ke dalam data pelatihan (*training data*) dan himpunan data (*dataset*). Proses pelatihan dilakukan dengan mengelompokkan data ke dalam beberapa kategori yang telah ditentukan. Terdapat enam kategori yang didefinisikan, yaitu: *Innovative, Global Impact, Student Engagement, Career Focused, Research Excellent, Diversity*. Namun, karena keterbatasan data yang diperoleh, kategori *Diversity* dikecualikan dalam penelitian ini.

Untuk membentuk model data, data dikelompokkan ke dalam lima set. Setiap data dalam himpunan tersebut dilabeli secara manual apakah termasuk dalam kategori tersebut atau tidak. Selanjutnya, setiap set dibagi menjadi 90% data latih dan 10% data uji. Himpunan data ini kemudian digunakan untuk membangun model klasifikasi. Model klasifikasi yang telah dibangun dapat digunakan untuk memprediksi data selanjutnya.

Berdasarkan data yang telah dikumpulkan, penelitian ini menggunakan 100 *post* untuk setiap kategori dalam proses pembangunan model. Sebanyak 500 *post* diberi label, kemudian dibagi menjadi 450 *post* sebagai data pelatihan dan 50 *post* sebagai data pengujian.

D. *Evaluasi*

Evaluasi dilakukan dengan menghitung nilai *precision*, *recall*, dan *accuracy*. Dari hasil evaluasi menggunakan model yang dikembangkan dan data pengujian, diperoleh rata-rata nilai *precision* sebesar 80,8%, recall sebesar 78,8%, dan akurasi sebesar 80%.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

Tabel II
Hasil Evaluasi

Kategori	*Precision*	*Recall*
Student Engagement	86%	67%
Career Focused	80%	80%
Global Impact	82%	90%
Innovative	80%	57%
Research Excellent	76%	100%
Average	**80,8%**	**78,8%**
Overall Accuracy	**80%**	

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa temuan penting. Dari lima kategori yang dibentuk, terdapat satu kategori dengan nilai *recall* terendah, yakni sebesar 57%. Hal ini disebabkan oleh tingginya tingkat kemiripan antara *post* yang diberi label *Innovative* dengan *post* yang dilabeli *Research*, sehingga keduanya tidak memiliki pembeda yang signifikan secara tekstual.

Jika dibandingkan dengan beberapa studi sebelumnya yang mengevaluasi *brand* perguruan tinggi di media sosial (Pringle, 2018), penelitian ini memiliki keunggulan dalam penerapan *machine learning* yang memungkinkan proses klasifikasi dilakukan berdasarkan konteks keseluruhan isi *post*. Sebaliknya, pada penelitian sebelumnya, analisis dilakukan dengan menghitung frekuensi kemunculan kata kunci tertentu. Penggunaan kata kunci secara eksklusif terkadang tidak mencerminkan konteks sebenarnya dari suatu *post*, karena beberapa institusi memiliki pola unggahan standar yang menggunakan *hashtag* tertentu yang mengandung kata kunci dalam suatu kategori, meskipun konteks *post* tidak relevan dengan kategori tersebut. Sebagai contoh, sebagian besar *post* akun @unpad menggunakan *hashtag* #UnpadMendunia, meskipun tidak selalu membahas mengenai *Global Impact*.

KESIMPULAN

Penelitian ini berhasil membangun sebuah model klasifikasi yang mampu mengelompokkan *post* perguruan tinggi ke dalam lima kategori, yaitu *Innovation*, *Research*, *Global Impact*, *Student Engagement*, dan *Career Focused*. Model tersebut dikembangkan berdasarkan data *post* dari 30 perguruan tinggi di Indonesia dengan menggunakan algoritma Naïve Bayes. Evaluasi terhadap kinerja model dilakukan melalui tiga metrik utama, yakni *precision*, *recall*, dan *accuracy*. Hasil evaluasi menunjukkan bahwa model yang dibangun memiliki nilai rata-rata *precision* sebesar 80,8%, *recall* sebesar 78,8%, dan *accuracy* sebesar 80%.

REFERENSI

[1] M. F. Rouf, A. N. R. Attamimi, D. A. V. Putri, I. Nirmala, and A. N. Fadhilah, Statistik Pendidikan Tinggi. Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi (Kemendikti), 2023.

[2] A. Wæraas and M. N. Solbakk, 'Defining the essence of a university: lessons from higher education branding', High Educ (Dordr), vol. 57, no. 4, pp. 449–462, 2009, doi: 10.1007/s10734-008-9155-z.

[3] R. J. Brodie, M. S. Glynn, and V. Little, 'The service brand and the service-dominant logic: missing fundamental premise or the need for stronger theory?', Marketing Theory, vol. 6, no. 3, pp. 363–379, 2006, doi: 10.1177/1470593106066797.

[4] J. Pringle and R. Naidoo, 'Branding and the commodification of academic labour. In R. Barnett, P. Temple, & P. Scott (Eds.), Valuing higher education: An appreciation for the work of Gareth Williams', UCL Institute of Education Press, pp. 158–177, 2016.

[5] R. Rutter, S. Roper, and F. Lettice, 'Social media interaction, the university brand and recruitment performance', J Bus Res, vol. 69, no. 8, pp. 3096–3104, Aug. 2016, doi: 10.1016/j.jbusres.2016.01.025.

[6] A. Peruta and A. B. Shields, 'Social media in higher education: understanding how colleges and universities use Facebook', Journal of Marketing for Higher Education, vol. 27, no. 1, pp. 131–143, Jan. 2017, doi: 10.1080/08841241.2016.1212451.

[7] M. Z. Rabbil, T. D. Gugat, R. J. Intiha, and D. H. Putri, 'Strategi Media Sosial yang Efektif untuk Meningkatkan Brand Awareness dan Engagement pada Kampus Politeknik Bina Madani', Masarin, vol. 1, no. 2, pp. 67–77, Dec. 2022.

[8] M. Yusuf, H. Saleh, and L. Setiawan, 'Pengaruh Media Sosial Dan Citra Institusi Terhadap Minat Pada Keputusan Memilih Kuliah Pada Jurusan Perjalanan Politeknik Pariwisata Makassar', Indonesian Journal of Business and Management, vol. 7, no. 1, pp. 152–159, 2024.

[9] T. P. Yazid, A. Rasyid, and M. Hatika, 'Pengaruh Cyber Public Relations terhadap Citra Perguruan Tinggi Swasta Terfavorit di Provinsi Riau', Edukatif: Jurnal Ilmu Pendidikan, vol. 4, no. 4, pp. 5734–5743, 2022.

[10] M. Beverland, Building brand authenticity, 1st ed. Basingstoke, England: Palgrave Macmillan, 2009.

[11] Z. Wang, X. Sun, D. Zhang, and X. Li, 'An Optimal SVM-Based Text Classification Algorithm', in 2006 International Conference on Machine Learning and Cybernetics, 2006, pp. 1378–1381. doi: 10.1109/ICMLC.2006.258708.

[12] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. in The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2011. [Online]. Available: https://books.google.co.id/books?id=pQws07tdpjoC

[13] J. Pringle and S. Fritz, 'The university brand and social media: using data analytics to assess brand authenticity', Journal of Marketing for Higher Education, vol. 29, no.

1, pp. 19–44, Jan. 2019, doi: 10.1080/08841241.2018.1486345.

[14] M. Junker, R. Hoch, and A. Dengel, 'On the evaluation of document analysis components by recall, precision, and accuracy', in Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318), 1999, pp. 713–716. doi: 10.1109/ICDAR.1999.791887.