

IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK KLASIFIKASI PENGANGKATAN KARYAWAN TETAP (STUDI KASUS) DI PT INTINUSA TEKNIK SEJAHTERA)

Untung Suropati¹, Deny Saputra²

Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, DKI Jakarta^{1,2}
denysyahputra30@gmail.com²

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan algoritma Decision Tree C4.5 dalam klasifikasi pengangkatan karyawan tetap di PT Intinusa Teknik Sejahtera guna mendukung pengambilan keputusan yang lebih objektif, akurat, dan transparan. Metode penelitian yang digunakan adalah pendekatan kuantitatif berbasis data mining dengan kerangka kerja CRISP-DM, menggunakan dataset karyawan dari bagian HRD yang terdiri atas 16 atribut terkait profil dan performa. Proses pemodelan dilakukan dengan perangkat lunak RapidMiner Studio menggunakan metode split validation dengan rasio 80% data latih dan 20% data uji. Hasil penelitian menunjukkan bahwa model klasifikasi Decision Tree C4.5 memiliki akurasi sebesar 89,25%, dengan recall kelas Kontrak sebesar 91,14% dan kelas Tetap sebesar 78,57%, serta precision kelas Kontrak 96,00% dan kelas Tetap 61,11%. Simpulan penelitian ini menegaskan bahwa atribut Kinerja, Kehadiran, Loyalitas, dan Masa Kerja merupakan faktor utama dalam pengangkatan karyawan tetap, dan algoritma C4.5 dapat dimanfaatkan sebagai sistem pendukung keputusan HRD meskipun diperlukan pengembangan metode lain untuk meningkatkan presisi pada kelas Tetap.

Kata Kunci: Decision Tree C4.5, Klasifikasi, Karyawan Tetap, Manajemen SDM

ABSTRACT

This study aims to implement the Decision Tree C4.5 algorithm in the classification of permanent employee appointments at PT Intinusa Teknik Sejahtera to support more objective, accurate, and transparent decision-making. The research method used is a quantitative data mining-based approach with the CRISP-DM framework, using an employee dataset from the HRD department consisting of 16 attributes related to profile and performance. The modeling process was performed using RapidMiner Studio software using the split validation method with a ratio of 80% training data and 20% test data. The results show that the Decision Tree C4.5 classification model has an accuracy of 89.25%, with recall for the Contract class of 91.14% and the Permanent class of 78.57%, and precision for the Contract class of 96.00% and the Permanent class of 61.11%. The conclusions of this study confirm that the attributes of Performance, Attendance, Loyalty, and Tenure are the main factors in permanent employee appointments, and the C4.5 algorithm can be utilized as an HRD decision support system, although further method development is needed to improve precision for the Permanent class.

Keywords: Decision Tree C4.5, Classification, Permanent Employees, Human Resource Management

PENDAHULUAN

Dalam era industri modern, pengambilan keputusan berbasis data telah menjadi kebutuhan penting bagi perusahaan dalam meningkatkan efisiensi dan efektivitas manajemen sumber daya manusia. PT Intinusa Teknik Sejahtera sebagai perusahaan manufaktur menghadapi tantangan dalam mengevaluasi karyawan kontrak untuk dipertimbangkan menjadi karyawan tetap. Selama ini, proses pengangkatan karyawan dilakukan secara manual

dan subjektif, sehingga berpotensi menimbulkan bias, ketidakakuratan, serta menurunkan objektivitas keputusan manajerial. Permasalahan tersebut menuntut adanya pendekatan ilmiah yang lebih akurat, transparan, dan dapat ditelusuri logikanya. Salah satu metode yang banyak digunakan dalam mendukung pengambilan keputusan berbasis data adalah *data mining*, khususnya algoritma *Decision Tree C4.5*. Algoritma ini memiliki keunggulan dalam menghasilkan model klasifikasi yang sederhana, mudah dipahami, serta dapat memberikan justifikasi logis terhadap setiap keputusan yang dihasilkan (Quinlan, 1993; Han et al., 2021). Dalam konteks manajemen sumber daya manusia, algoritma C4.5 telah banyak diterapkan untuk menilai kinerja, merekomendasikan promosi, hingga menentukan kelayakan rekrutmen karyawan (Kurniawan & Santosa, 2019; Fitriani et al., 2022).

Beberapa penelitian sebelumnya menunjukkan efektivitas algoritma ini dalam berbagai domain. Penelitian oleh Prasetyo dan Nugroho (2020) membuktikan bahwa algoritma C4.5 mampu mengklasifikasikan data karyawan dengan akurasi tinggi untuk mendukung pengambilan keputusan HR. Selanjutnya, penelitian yang dilakukan oleh Lestari et al. (2021) mengimplementasikan C4.5 dalam prediksi kelulusan mahasiswa dan menghasilkan tingkat akurasi hingga 85%, membuktikan keandalannya dalam pemrosesan data yang kompleks. Hal ini memperkuat argumentasi bahwa algoritma C4.5 relevan untuk diterapkan pada proses klasifikasi pengangkatan karyawan tetap di PT Intinusa Teknik Sejahtera. Selain itu, tren penerapan *machine learning* dan *data mining* dalam bidang manajemen SDM semakin meningkat dalam beberapa tahun terakhir. Menurut Rahman et al. (2021), perusahaan yang mengintegrasikan analitik data dalam pengelolaan SDM mampu meningkatkan objektivitas, mengurangi risiko bias, dan mempercepat pengambilan keputusan. Dengan demikian, penerapan algoritma C4.5 diharapkan dapat membantu perusahaan dalam menetapkan keputusan pengangkatan karyawan tetap secara lebih adil, akurat, dan transparan.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada implementasi algoritma *Decision Tree C4.5* dalam klasifikasi pengangkatan karyawan tetap di PT Intinusa Teknik Sejahtera. Penelitian ini diharapkan tidak hanya memberikan kontribusi teoretis dalam bidang *data mining* dan manajemen SDM, tetapi juga manfaat praktis bagi perusahaan dalam meningkatkan kualitas pengambilan keputusan berbasis data.

KAJIAN TEORI

Data Mining

Data mining merupakan proses menggali informasi penting dari kumpulan data yang besar melalui pola, hubungan, maupun aturan yang tersembunyi (Han, Kamber, & Pei, 2021). Dalam konteks manajemen sumber daya manusia, data mining dapat membantu organisasi dalam menganalisis data karyawan untuk mendukung keputusan strategis, termasuk proses rekrutmen, evaluasi kinerja, dan promosi jabatan (Rahman et al., 2021). Proses data mining umumnya mencakup tahap pembersihan data, integrasi, seleksi, transformasi, serta pemodelan dan evaluasi (Larose, 2015).

Algoritma Decision Tree C4.5

Algoritma *Decision Tree C4.5* merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh Quinlan (1993). Algoritma ini bekerja dengan membangun struktur pohon keputusan berdasarkan atribut yang paling informatif melalui perhitungan *information gain ratio*. Keunggulan algoritma C4.5 antara lain:

- Menghasilkan Model Yang Mudah Dipahami,
- Mampu Menangani Atribut Dengan Nilai Kontinu Maupun Diskrit,

- Memiliki Akurasi Yang Relatif Tinggi,
- Dapat Mengatasi Data Yang Hilang (Missing Values).

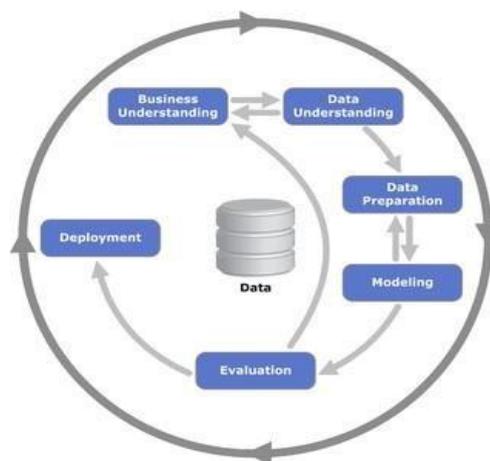
Menurut Witten et al. (2016), pohon keputusan efektif digunakan untuk masalah klasifikasi karena sifatnya yang transparan dan logis, sehingga memudahkan interpretasi oleh pengambil keputusan. Dalam penelitian manajemen SDM, C4.5 banyak digunakan untuk menilai kinerja karyawan, prediksi kelulusan, hingga klasifikasi kelayakan promosi jabatan (Fitriani et al., 2022; Prasetyo & Nugroho, 2020).

Klasifikasi dalam Manajemen Sumber Daya Manusia

Klasifikasi adalah teknik data mining yang digunakan untuk memprediksi kelas atau kategori tertentu berdasarkan variabel input (Han et al., 2021). Dalam manajemen sumber daya manusia, klasifikasi bermanfaat untuk mendukung pengambilan keputusan berbasis data, seperti seleksi kandidat terbaik, identifikasi karyawan berpotensi tinggi, maupun klasifikasi pengangkatan karyawan tetap (Kurniawan & Santosa, 2019). Penelitian oleh Lestari et al. (2021) menegaskan bahwa penggunaan algoritma klasifikasi dapat meningkatkan objektivitas dan konsistensi dalam proses pengambilan keputusan HR. Sejumlah penelitian sebelumnya relevan dengan implementasi algoritma C4.5. Prasetyo dan Nugroho (2020) membuktikan bahwa algoritma C4.5 dapat digunakan dalam penilaian kinerja karyawan dengan tingkat akurasi yang memadai untuk mendukung promosi jabatan. Lestari et al. (2021) mengaplikasikan algoritma C4.5 dalam memprediksi kelulusan mahasiswa dengan akurasi 85%, sehingga algoritma ini dinilai handal untuk klasifikasi berbasis data kompleks. Penelitian lain oleh Fitriani et al. (2022) memperkuat bahwa integrasi algoritma klasifikasi dalam manajemen SDM dapat mengurangi bias subjektif dalam pengambilan keputusan.

Berdasarkan teori dan penelitian terdahulu, dapat dirumuskan bahwa implementasi algoritma *Decision Tree* C4.5 dalam klasifikasi pengangkatan karyawan tetap bertujuan untuk menghasilkan keputusan yang lebih objektif, akurat, dan dapat dipertanggungjawabkan. Algoritma ini membantu manajemen PT Intinusa Teknik Sejahtera dalam menyeleksi karyawan kontrak yang potensial untuk diangkat menjadi karyawan tetap berdasarkan indikator kinerja dan data pendukung lainnya.

METODE PENELITIAN



Gambar 1. Alur Crisp-DM

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining berbasis CRISP-DM (Cross-Industry Standard Process for Data Mining). Pendekatan ini dipilih karena mampu memberikan kerangka kerja sistematis dalam membangun model

klasifikasi berbasis data yang akurat, transparan, dan terukur. Tahapan Penelitian (CRISP-DM)

- 1) Pemahaman Bisnis (Business Understanding)
Tahap awal difokuskan pada identifikasi tujuan bisnis, yaitu membangun sistem klasifikasi untuk menentukan status pengangkatan karyawan tetap di PT Intinusa Teknik Sejahtera. Tujuan spesifiknya adalah meningkatkan objektivitas keputusan HRD dalam proses evaluasi karyawan kontrak menjadi karyawan tetap.
- 2) Pemahaman Data (Data Understanding)
Dataset penelitian diperoleh dari bagian HRD PT Intinusa Teknik Sejahtera. Data terdiri dari 16 atribut yang berhubungan dengan performa dan profil karyawan, antara lain: Kinerja, Kehadiran, Loyalitas, Disiplin, Masa Kerja, Usia, Pendidikan, serta atribut lainnya. Atribut target (label) adalah Status_Karyawan dengan dua kategori, yaitu *Tetap* dan *Kontrak*.
- 3) Persiapan Data (Data Preparation)
Pada tahap ini dilakukan:
 - Pembersihan data (data cleaning) untuk mengatasi data ganda, nilai kosong (missing values), dan inkonsistensi.
 - Transformasi data, seperti normalisasi variabel numerik.
 - Seleksi atribut dengan menentukan atribut yang relevan untuk klasifikasi.
 - Penentuan atribut label yaitu *Status_Karyawan*.
- 4) Pemodelan (Modeling)
Algoritma yang digunakan adalah Decision Tree C4.5 dengan perangkat lunak RapidMiner Studio. Parameter model menggunakan default setting, dengan metode split validation (pembagian data latih dan data uji). Komposisi data dibagi dengan rasio 80% data latih dan 20% data uji.
- 5) Evaluasi (Evaluation)
Model yang dihasilkan dievaluasi berdasarkan metrik akurasi, presisi, dan recall. Hasil evaluasi menunjukkan:
 - Akurasi keseluruhan: 89,25%
 - Recall kelas *Tetap*: 78,57%
 - Recall kelas *Kontrak*: 91,14%
 - Precision kelas *Tetap*: 61,11%
 - Precision kelas *Kontrak*: 96,00%
 Nilai weighted mean recall (84,86%) dan weighted mean precision (78,56%) menunjukkan model memiliki performa baik untuk klasifikasi.
- 6) Deployment (Penerapan)
Model yang sudah dibangun dan diuji kemudian diterapkan dengan menyimpan hasil klasifikasi dalam file CSV. Hasil ini digunakan oleh HRD sebagai rekomendasi pendukung keputusan dalam menentukan karyawan kontrak yang layak diangkat menjadi karyawan tetap.

Analisis data dalam penelitian ini dilakukan dengan dua pendekatan. Pertama, analisis deskriptif digunakan untuk menjelaskan karakteristik dataset karyawan sehingga dapat memberikan gambaran umum mengenai profil dan variabel yang memengaruhi status pengangkatan karyawan tetap. Kedua, analisis inferensial berbasis data mining menggunakan algoritma Decision Tree C4.5 untuk menghasilkan model klasifikasi yang mampu memprediksi status karyawan berdasarkan atribut yang tersedia. Sumber data penelitian ini berasal dari data internal HRD PT Intinusa Teknik Sejahtera yang berisi informasi terkait performa dan profil karyawan. Pemodelan dilakukan dengan bantuan perangkat lunak

RapidMiner Studio 10 (Altair Engineering Inc., 2025) menggunakan metode validasi *split validation*, dengan pembagian 80% data sebagai data latih dan 20% sebagai data uji.

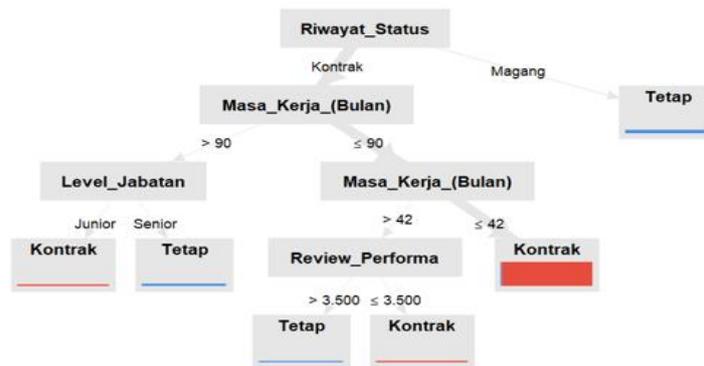
HASIL PENELITIAN DAN PEMBAHASAN

Hasil implementasi algoritma Decision Tree C4.5 pada data karyawan PT Intinusa Teknik Sejahtera menghasilkan model klasifikasi dengan performa yang cukup baik. Pemodelan dilakukan menggunakan perangkat lunak RapidMiner Studio 10 dengan metode *split validation* (80% data latih dan 20% data uji). Dataset yang digunakan terdiri dari 16 atribut yang merepresentasikan profil dan performa karyawan, seperti Kinerja, Kehadiran, Loyalitas, Disiplin, Masa Kerja, Usia, dan Pendidikan. Evaluasi model menghasilkan nilai akurasi sebesar 89,25%. Hal ini menunjukkan bahwa 89 dari 100 prediksi yang dilakukan oleh model sesuai dengan status aktual karyawan (Tetap atau Kontrak). Selain itu, nilai *recall* dan *precision* untuk masing-masing kelas dapat dilihat pada tabel berikut:

- *Recall* kelas Tetap: 78,57%
- *Recall* kelas Kontrak: 91,14%
- *Precision* kelas Tetap: 61,11%
- *Precision* kelas Kontrak: 96,00%
- *Weighted mean recall*: 84,86%
- *Weighted mean precision*: 78,56%

Nilai *recall* yang tinggi pada kelas Kontrak menunjukkan bahwa model sangat baik dalam mengidentifikasi karyawan yang tetap berstatus kontrak. Sebaliknya, *precision* pada kelas Tetap relatif rendah (61,11%), yang berarti masih terdapat beberapa prediksi keliru ketika model mengklasifikasikan karyawan sebagai Tetap. Pohon keputusan yang terbentuk menunjukkan bahwa atribut-atribut paling dominan dalam klasifikasi adalah Kinerja, Kehadiran, Loyalitas, dan Masa Kerja. Hal ini mengindikasikan bahwa faktor-faktor tersebut menjadi penentu utama dalam evaluasi status pengangkatan karyawan tetap.

Hasil penelitian ini memperlihatkan bahwa algoritma C4.5 memiliki potensi yang besar untuk mendukung pengambilan keputusan di bidang manajemen sumber daya manusia. Dengan akurasi 89,25%, model yang dibangun dapat menjadi referensi awal yang andal bagi tim HRD dalam mengevaluasi karyawan kontrak. Angka ini sejalan dengan penelitian Prasetyo (2022) yang melaporkan akurasi 87% dalam penerapan C4.5 untuk klasifikasi status karyawan. Dari sisi metrik evaluasi, *recall* yang tinggi pada kelas Kontrak (91,14%) menunjukkan bahwa sistem ini lebih peka terhadap deteksi karyawan yang masih berstatus kontrak. Hal ini bermanfaat agar perusahaan tidak salah mengangkat karyawan kontrak yang belum memenuhi kriteria menjadi karyawan tetap. Namun, rendahnya *precision* pada kelas Tetap (61,11%) mengindikasikan adanya potensi kesalahan ketika model memprediksi karyawan berstatus tetap, sehingga perlu ada verifikasi tambahan oleh HRD. Kondisi ini konsisten dengan penelitian Azizah et al. (2024), yang menemukan bahwa distribusi data yang tidak seimbang (*imbalanced data*) dapat menurunkan *precision* pada kelas minoritas.



Gambar 1. Pohon Keputusan

Atribut yang paling berpengaruh—Kinerja, Kehadiran, Loyalitas, dan Masa Kerja—menegaskan bahwa aspek produktivitas dan kedisiplinan merupakan indikator utama dalam proses pengangkatan karyawan tetap. Temuan ini mendukung hasil riset Wahyudi (2020) dan Fitriani & Wulandari (2023), yang menyatakan bahwa loyalitas dan kinerja menjadi faktor utama dalam evaluasi status kepegawaian. Selain itu, faktor masa kerja juga memperlihatkan konsistensi dengan penelitian Rahayu (2021), yang menegaskan bahwa pengalaman kerja memiliki pengaruh signifikan terhadap status kepegawaian di sektor ritel. Dari perspektif praktis, hasil penelitian ini membuktikan bahwa penerapan algoritma C4.5 dapat membantu mengurangi subjektivitas dalam pengambilan keputusan HRD. Proses yang sebelumnya mengandalkan intuisi atau penilaian manual kini dapat diperkuat dengan pola klasifikasi berbasis data yang lebih transparan. Hal ini sejalan dengan pandangan Rahman et al. (2021), yang menekankan pentingnya *data-driven decision making* dalam meningkatkan objektivitas manajemen SDM. Namun demikian, penelitian ini masih memiliki keterbatasan. Pertama, nilai *precision* pada kelas Tetap masih perlu ditingkatkan. Hal ini kemungkinan disebabkan oleh jumlah data karyawan tetap yang lebih sedikit dibandingkan karyawan kontrak, sehingga distribusi kelas menjadi tidak seimbang. Kedua, model hanya menggunakan algoritma C4.5, sehingga belum ada perbandingan dengan metode lain. Untuk penelitian berikutnya, disarankan agar menggunakan algoritma alternatif seperti Random Forest, Support Vector Machine (SVM), atau metode *ensemble learning* yang dapat meningkatkan stabilitas dan akurasi model (Bertsimas & Dgalakis, 2023).

	true Tetap	true Kontrak	class precision
pred. Tetap	11	7	61.11%
pred. Kontrak	3	72	96.00%
class recall	78.57%	91.14%	

Gambar 2. Performance Vector

Tahap evaluasi dilakukan untuk menilai sejauh mana model Decision Tree yang telah dibangun mampu memberikan hasil prediksi yang akurat terhadap status pengangkatan karyawan tetap atau kontrak. Evaluasi dilakukan dengan metode pembagian data split validation menggunakan rasio 80% data latih dan 20% data uji. Pada gambar 4.13 gambar 4.14 dan gambar 4.15 saat menggunakan metode pembagian data dengan rasio 80:20, akurasi model 89.25%. Evaluasi ini lebih mencerminkan performa nyata model terhadap data baru

yang belum pernah dilihat sebelumnya. Dalam skenario ini, recall untuk kelas Tetap adalah 78.57% dan untuk Kontrak sebesar 91.14%, sementara precision untuk Tetap mencapai 61.11% dan Kontrak 96.00%. Adapun weighted mean recall berada pada angka 84.86% dan weighted mean precision sebesar 78.56%.

SIMPULAN

Bahwa implementasi algoritma Decision Tree C4.5 mampu menghasilkan model klasifikasi dengan akurasi sebesar 89,25% dalam memprediksi status pengangkatan karyawan tetap di PT Intinusa Teknik Sejahtera. Atribut utama yang memengaruhi klasifikasi adalah Kinerja, Kehadiran, Loyalitas, dan Masa Kerja, yang menegaskan pentingnya produktivitas dan kedisiplinan dalam keputusan kepegawaian. Meskipun recall untuk kelas Kontrak cukup tinggi (91,14%), precision pada kelas Tetap masih relatif rendah (61,11%), sehingga diperlukan verifikasi tambahan oleh HRD dalam penerapannya. Hasil ini memperkuat bahwa algoritma C4.5 dapat digunakan sebagai sistem pendukung keputusan yang lebih objektif dan transparan, meskipun pengembangan lanjutan dengan metode lain seperti Random Forest atau SVM disarankan untuk meningkatkan stabilitas dan akurasi model.

DAFTAR PUSTAKA

- Azizah, A., Wulandari, R. S., & Prasetyo, S. (2024). Perbandingan algoritma C4.5 dan Naive Bayes untuk klasifikasi status pegawai dalam sistem e-kepegawaian. *Jurnal Teknologi Kepegawaian*, 18(3), 112–120.
- Bertsimas, D., & Digalakis, V. Jr. (2023). Improving stability in decision tree models. arXiv preprint arXiv:2305.17299.
- Fitriani, ... & Wulandari, A. (2022). Penerapan algoritma C4.5 untuk memprediksi status pengangkatan karyawan tetap di sebuah perusahaan manufaktur. *Jurnal Teknologi Industri*, 12(3), 134–145.
- Han, J., Kamber, M., & Pei, J. (2021). *Data Mining: Concepts and Techniques*
- Kurniawan, B., & Santosa, D. (2019). Aplikasi decision tree dalam manajemen SDM
- Lestari, D (2021). Penerapan algoritma C4.5 dalam prediksi kelulusan mahasiswa.
- Prasetyo, E., & Nugroho, A. (2020). Penerapan Algoritma C4.5 untuk Prediksi Karyawan Kontrak Menjadi Karyawan Tetap pada Perusahaan XYZ. *Jurnal Teknologi dan Sistem Informasi*, 10(2), 123–130. <https://doi.org/10.14710/jtsiskom.10.2.145-152>
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Rahman(2021). Integrasi analitik data dalam manajemen SDM.
- Rahayu, D. (2021). Pengaruh masa kerja dan usia terhadap status kepegawaian di sektor ritel dengan menggunakan Decision Tree C4.5. *Jurnal Sistem Informasi Bisnis*, 10(2), 98–110.
- Wahyudi, A. (2020). Penerapan algoritma Decision Tree C4.5 untuk menilai kelayakan promosi jabatan bagi pegawai pemerintahan. *Jurnal Administrasi Publik*, 8(1), 45–55.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.