**Machine Learning Model untuk Merekomendasikan Buku**

**pada Yolo Book Café**

**Dayton\_11\*, Christopher Khomali\_22, Ellwan Edy Wei \_33, Kelvyn \_4⁴**

Alamat Institusi: Sei Ladi, Jl. Gajah Mada, Baloi Permai, Kec. Sekupang, Kota Batam, Kepulauan Riau 29442

[dayton159@gmail.com](mailto:dayton159@gmail.com)

***Abstract***

*Small-Medium Entreprises (SMEs) are the type of business that is most affected during this pandemic. The pandemic that occurred in Indonesia has forced SMEs to evolve in order to survive. The SME which is the partner of the author is the Food and Beverages SME which provides book lending services and currently requires assistance in the book lending system in selecting the appropriate genre for visitors. This genre classification process can be done using the Decision Tree algorithm. This research will compare the performance of the model developed with 2 different tools, namely KNIME Workflows and Sci-Kit Learn by comparing Accuracy, Precision, F1-Score, and Recall. The results showed that the performance tests using KNIME had the best performance compared to the model developed using Sci-Kit Learn and it is the best model to be used for classifying the book genre for our SME partner.*

***Abstrak***

*Usaha Mikro Kecil Menengah (UMKM) merupakan jenis usaha yang paling terdampak selama masa pandemi ini. Pandemi yang terjadi di Indonesia telah memaksa UMKM untuk berevolusi demi terus bertahan. UMKM yang menjadi mitra dari penulis merupakan UMKM Food and Beverages yang menyediakan layanan peminjaman buku dan saat ini membutuhkan bantuan dalam sistem peminjaman buku tersebut dalam memilihkan genre yang sesuai bagi pengunjung. Proses klasifikasi genre ini dapat dilakukan dengan menggunakan algoritma Decision Tree. Penelitian kali ini akan melakukan perbandingan performa dari model yang dikembangkan dengan 2 tools berbeda, yaitu menggunakan KNIME Workflows dan juga Sci-Kit Learn dengan membandingkan Accuracy, Precision, F1-Score, dan Recall.  Hasil penelitian menunjukkan bahwa hasil pengujian dengan menggunakan KNIME memiliki performa yang paling baik dibandingkan model yang dikembangkan dengan Sci-Kit Learn untuk digunakan sebagai model pengklasifikasian genre buku bagi UMKM.*

**PENDAHULUAN**

Usaha Mikro Kecil Menengah atau yang biasa dikenal dengan UMKM merupakan pilar fondasi dari pertumbuhan perekonomian Indonesia. UMKM memiliki peran penting dalam menyediakan lapangan kerja yang memadai bagi masyarakat Indonesia serta berkontribusi dalam menyumbangkan sebagian besar pada pendapatan nasional Indonesia. Sebagaimana diketahui, pandemi Covid-19 telah berhasil meluluhlantakkan perekonomian Indonesia, khususnya pada UMKM. Sebagian besar UMKM yang terpengaruh oleh pandemi merupakan UMKM yang bergerak dalam bidang *Food and Beverages* (F&B).

Di tengah tantangan ekonomi selama pandemi, UMKM memerlukan sebuah terobosan baru untuk terus bertahan dalam keadaan ini. Terlepas dari beberapa solusi yang dilakukan seperti pemotongan biaya dan mengubah bisnis model, pemerintah sendiri menekankan pentingnya transformasi digital yang dilakukan terhadap UMKM. Hal tersebut juga merupakan target di Universitas Internasional Batam, yang mana para mahasiswa maupun mahasiswi ditugaskan untuk membantu mitra UMKM di kota Batam dengan cara melakukan penerapan teknologi dalam proses bisnis UMKM.

UMKM mitra penulis sendiri adalah Yolo Book Café yang merupakan UMKM bidang F&B yang juga menyediakan layanan peminjaman buku dari perpustakaan yang mereka miliki. Berdasarkan observasi yang dilakukan oleh penulis, sejauh ini UMKM terkait masih belum memiliki sistem atau platform digital beserta teknik pemasaran yang memadai dalam menjalankan proses bisnis mereka. Kondisi inilah yang mendorong penulis guna membangun dan mengoptimalisasi sistem UMKM demi meningkatkan performa proses bisnis UMKM, yang salah satunya akan dilakukan dengan mengimplementasikan teknologi *Machine Learn*ing sebagaimana yang menjadi fokus dari penelitian ini.

Saat ini salah satu jenis penerapan teknologi yang paling populer terhadap UMKM adalah *Machine Learning*. Istilah *Machine Learning* merupakan suatu metode guna mengajarkan sesuatu hal terhadap komputer dengan memberikan data-data yang kemudian akan dilakukan pengolahan. *Machine Learning* sangatlah praktis dan efektif digunakan, khususnya dalam pengambilan keputusan penting yang biasa dilakukan oleh manusia kini sudah dapat dipercayakan kepada sebuah sistem.

Kemampuan inilah yang akan kami terapkan pada UMKM terkait. *Machine Learning* ini akan dibangun berdasarkan pada layanan peminjaman buku yang disediakan oleh UMKM. *Machine Learning* yang akan dikembangkan memiliki fungsi untuk membantu pengguna dalam merekomendasikan genre-genre buku apa saja yang cocok untuk dibaca pengunjung berdasarkan data diri pengunjung seperti umur dan gender.

Penelitian ini difokuskan untuk menjelaskan metodologi beserta proses-proses yang ditempuh penulis dalam membangun model *Machine Learning* dengan menggunakan algoritma berbasis *Decision Tree* yang akan diterapkan pada layanan peminjaman buku UMKM terkait. Hal ini dilakukan guna membantu mempermudah proses bisnis UMKM. Penelitian ini juga akan membahas mengenai hasil akhir dan performa model dalam mengklasifikasi genre buku, serta akan dilakukan perbandingan berdasarkan 2 tools yang digunakan guna mencari tools mana yang menghasilkan performa yang lebih baik.

**METODE**

**Tahapan Penelitian**

Penelitian diawali dengan pengumpulan data-data yang berasal dari mitra UMKM sendiri yaitu Yolo Book Café. Setelah data-data berkualitas yang akan digunakan untuk melatih algoritma diperoleh, tahap selanjutnya adalah melakukan pengolahan data-data lebih lanjut sebelum diproses dengan menggunakan *tool-tool* pada penelitian ini. Data yang akan diterapkan pada *tool-tool* yang digunakan tentunya merupakan data serupa dengan algoritma yang memiliki basis serupa pula guna memudahkan proses perbandingan performa model.

Berdasarkan hasil yang diperoleh melalui proses pengolahan data dan pengujian dengan algoritma, akan dilakukan analisa terhadap performa model guna mendapatkan kesimpulan mengenai algoritma dan tools yang menghasilkan performa paling baik untuk diimplementasikan pada mitra UMKM.

**Data yang Digunakan**

Data yang digunakan merupakan sekumpulan dataset yang diperoleh penulis secara manual melalui UMKM. Dikarenakan istilah *Machine Learning* masih merupakan suatu hal yang baru bagi UMKM, maka dari itu data masih diperoleh secara manual langsung dari UMKM guna dilakukan pelatihan pertama kali pada algoritma.

Data yang berhasil dikumpulkan memiliki jumlah sebanyak 2641 contoh data, yang mana sebagian dari dataset yang akan digunakan dalam pelatihan *Machine Learning* ditunjukkan sebagaimana yang terlihat pada Gambar 1 berikut ini:

Table

Description automatically generated

**Gambar 1**Data yang Digunakan

Sumber: Data Diolah (2021)

**Tools dan Algoritma yang Diterapkan**

Pengujian terhadap data dilakukan dengan menggunakan 2 jenis aplikasi, yaitu dengan menggunakan aplikasi perangkat lunak KNIME 4.3.2 dan Sci-Kit Learn 0.24.1 (Python 3.8.5). Penulis akan menguji perbedaan performa yang ada pada penggunaan S*oftware Open Source*yang merupakan suatu platform analisis dengan modul *software* yang dikembangkan secara langsung oleh penulis dengan menggunakan bahasa pemrograman Python. Algoritma yang akan dibandingkan melalui kedua *tools* tersebut tidak lain adalah algoritma *Decision Tree*.

**HASIL DAN PEMBAHASAN**

Pada penelitian ini, proses dan hasil dari pembuatan model merupakan komponen utama dalam penelitian ini. Selain dari mengolah dataset dengan menggunakan *tools* dan algoritma tertentu, akan dibahas lebih lanjut pula mengenai proses-proses yang dilakukan dalam melakukan pengolahan dan partisi pada dataset, pelatihan dan pengujian model, hingga memunculkan laporan performa dari model itu sendiri.

**KNIME Workflows**

Pengujian yang dilakukan dengan KNIME tentunya memiliki tahap-tahap pengolahan data hingga akhirnya menjadi suatu model yang dapat diukur performanya. Tahap-tahap pemrosesan data yang dilakukan pada KNIME dapat dilihat pada Gambar 2:

A picture containing chart

Description automatically generated

**Gambar 2**KNIME Workflows

Sumber: Data Diolah (2021)

Dikarenakan dataset yang akan digunakan pada penelitian ini telah dibahas sebelumnya pada Gambar 1, maka pembahasan mengenai proses akan dimulai dari *pre-processing data.*

**KNIME Data Pre-processing**

Pada tahap data *pre-processing* sebagaimana yang terlihat pada Gambar 3*,* dilakukan pengecekan untuk mengantisipasi adanya *missing values* yang dapat berpotensi mengganggu perancangan model. Selanjutnya, akan dilakukan partisi data untuk membagi data-data yang akan digunakan baik untuk melatih model algoritma hingga untuk pengujian terhadap model itu sendiri. Penulis menggunakan persentase data yang sama dalam proses partisi data baik pada KNIME maupun Sci-Kit Learn. Data akan dibagi sebanyak 80% untuk pelatihan model dan 20% sisanya untuk melakukan pengujian.

Chart

Description automatically generated with low confidence

**Gambar 3:**Partitioning Node

Sumber: Data Diolah (2021)

Setelah partisi data dilakukan, data yang sudah terpartisi akan ditandai dengan warna merah dan hijau guna memudahkan proses analisis hasil secara visual sebagaimana tampak pada Gambar 4:

Diagram

Description automatically generated

**Gambar 4:**Color Interpretation Node

Sumber: Data Diolah (2021)

**KNIME Training and Testing Model**

Tahap selanjutnya dalam membangun model adalah melakukan pelatihan terhadap algoritma *Decision Tree*, yang kemudian akan diikuti dengan pengujian terhadap algoritma yang telah dilatih untuk memprediksi data-data testing dengan menggunakan data yang sebelumnya telah dilakukan partisi. Pada tahap ini model Machine Learning dengan algoritma Decision Tree telah berhasil dibuat seperti pada Gambar 5 dan sudah siap untuk dilakukan evaluasi terhadap performa model.

Diagram

Description automatically generated

**Gambar 5:**Decision Tree Node

Sumber: Data Diolah (2021)

**KNIME Model Evaluation**

Tahap ini merupakan tahap terakhir dari pembuatan model *Machine Learning*. Berdasarkan model yang telah diimplentasikan sebelumnya, akan dilakukan evaluasi performa model dalam mengklasifikasi data seperti pada Gambar 6.

A screenshot of a computer

Description automatically generated with medium confidence

**Gambar 6:**Scorer and Report Node

Sumber: Data Diolah (2021)

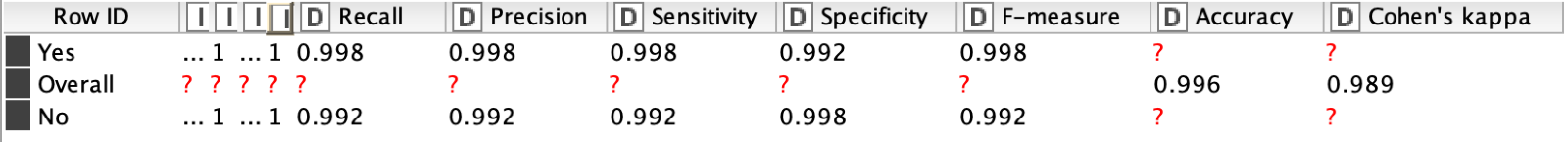
Hasil performa model kemudian akan ditampilkan beberapa atribut seperti *Confusion Matrix*, *Accuracy*, *Precision*, *F1-Score*, dan *Recall*. Atribut-atribut ini akan ditampilkan dalam bentuk report pada KNIME sebagaimana yang dapat dilihat pada Gambar 7 dan Gambar 8. Berdasarkan hasil yang ditampilkan, dapat diambil kesimpulan bahwa model yang dibangun pada KNIME ini memiliki performa yang tinggi dengan 526 dari 528 data telah berhasil terklasifikasi dengan benar. Sesuai dengan statistik performa pada laporan yang dihasilkan juga dapat terlihat bahwa atribut-atribut performa pada model seperti *Accuracy*, *Precision*, *F1-Score*, dan *Recall* seluruhnya memiliki nilai diatas 99%.

Application

Description automatically generated with medium confidence

**Gambar 7:**Confusion Matrix Report

Sumber: Data Diolah (2021)

**Gambar 8:**Model Performance Report

Sumber: Data Diolah (2021)

**Sci-Kit Learn**

Pengujian yang dilakukan dengan Sci-Kit Learn dilaksanakan dengan bahasa pemrograman Python yang dikembangkan langsung oleh penulis. Penerapan algoritma *Decision Tree* yang diterapkan menggunakan tools ini dapat dilihat pada Gambar 9:

Diagram

Description automatically generated

**Gambar 9:**Decision Tree Algorithm

Sumber: Data Diolah (2021)

Sebagaimana yang telah diketahui bahwa dataset yang akan digunakan pada penelitian ini ada pada Gambar 1, maka pembahasan mengenai proses akan dimulai dari *pre-processing data*.

**Sci-kit Learn Data Pre-Processing**

Pada tahap ini sebagaimana yang terlihat pada Gambar 10, perlu adanya transformasi terhadap dataset penelitian dikarenakan Sci-Kit Learn mengharuskan dataset yang akan diproses dalam bentuk angka desimal. Transformasi data kemudian dilakukan pada gender dan jenis genre pada dataset penelitian.

Text

Description automatically generated

**Gambar 10:**Sci-kit Learn Data Preprocessing

Sumber: Data Diolah (2021)

**Sci-Kit Learn Training and Testing Model**

Tahap selanjutnya adalah membangun model dengan melakukan pelatihan dan pengujian terhadap algoritma *Decision Tree* itu sendiri. Untuk melakukan hal ini, dilakukan partisi terhadap dataset yang telah berhasil ditransformasi sebelumnya sebagaimana yang terlihat pada Gambar 11. Partisi data dilakukan dengan komposisi 80% data training dan 20% data testing sama seperti perancangan model dengan tools sebelumnya. Hal ini kemudian diikuti dengan pelatihan dan pengujian klasifikasi data langsung terhadap model dengan algoritma *Decision Tree* ini. Setelah melalui proses ini, model telah siap untuk dilakukan evaluasi terhadap performanya.

Text

Description automatically generated

**Gambar 11:**Decision Tree Training and Testing

Sumber: Data Diolah (2021)

**Sci-Kit Learn Model Evaluation**

Tahap ini merupakan tahap terakhir dari pembuatan model *Machine Learning* dengan menggunakan Sci-Kit Learn. Berdasarkan model yang telah diimplentasikan sebelumnya, akan dilakukan evaluasi performa model dalam mengklasifikasi data seperti pada Gambar 12.

Text

Description automatically generated

**Gambar 12:**Sci-Kit Learn Model Evaluation

Sumber: Data Diolah (2021)

Berdasarkan hasil dari performa model sebagaimana yang dapat dilihat pada Gambar 13, atribut-atribut performa model seperti *Accuracy*, *Precision*, *F1-Score*, dan *Recall* memiliki nilai diantara 87% hingga 91%.

Calendar

Description automatically generated

**Gambar 13:**Sci-Kit Model Performance

Sumber: Data Diolah (2021)

**Hasil Perbandingan Performa**

Berdasarkan hasil performa dari model yang diuji baik menggunakan KNIME Workflows maupun Sci-Kit Learn, hasil klasifikasi data menggunakan algoritma Decision Tree menggunakan Tool KNIME mengungguli Sci-Kit Learn dalam hal performa.

Model yang dikembangkan menggunakan KNIME Workflows memiliki hasil Accuracy, Precision, F1-Score, dan Recall yang seluruhnya memiliki nilai diatas 99%. Sementara itu, model yang dikembangkan menggunakan Sci-Kit Learn memiliki hasil Accuracy, Precision, F1-Score, dan Recall yang memiliki nilai diantara 87% hingga 91%. Oleh sebab itu, dapat diambil kesimpulan bahwa pembuatan model dengan menggunakan KNIME Workflows merupakan pilihan yang lebih baik dibandingkan Sci-Kit Learn dimana hasil dari model tersebut melampaui model Sci-Kit Learn dari seluruh aspek performa dan memiliki nilai performa yang paling mendekati angka 1 (nilai terbaik).

**KESIMPULAN**

Tujuan daripada penelitian ini adalah untuk mencari tools yang terbaik serta untuk mendokumentasikan proses dalam pembuatan model Machine Learning yang dapat membantu mempermudah UMKM dalam melakukan proses bisnisnya terutama pada layanan peminjaman buku. Penelitian ini difokuskan pada perbandingan performa dari model dengan menggunakan algoritma Decision Tree yang dikembangkan pada 2 tools penelitian yang meliputi KNIME Workflows dan Sci-Kit Learn.

Melalui hasil penelitian ini dapat disumpulkan bahwa KNIME merupakan tools yang paling baik dalam penerapan algoritma *Decision Tree* pada model Machine Learning untuk membantu salah satu proses bisnis UMKM, yaitu membantu sistem perpusatakaan UMKM dalam merekomendasikan genre buku yang paling layak untuk pengunjung. Namun tak dapat dipungkiri bahwa meskipun performa dari model Sci-Kit Learn ini berada di bawah model KNIME yang memiliki performa 99% atau nyaris sempurna, performa yang dimiliki oleh model berbasis Sci-Kit Learn ini juga memiliki performa memuaskan yang mencapai nilai 87% hingga 91% sehingga sebenarnya sudah dapat tergolong model yang baik dan layak untuk digunakan.

**UCAPAN TERIMAKASIH**

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Yefta Christian, S. Kom., M. Kom. selaku dosen pengampu mata kuliah Machine Learning yang telah memberikan banyak bimbingan baik secara teknis maupun pemberian masukan terhadap penulisan penelitian ini sehingga penelitian ini dapat terlaksana dengan baik.