

## Perbandingan Algoritma K-Nearest Neighbor (K-NN) Dengan Decision Tree (C4.5) Untuk Menentukan Judul Buku Berdasarkan Peminatan pada Perpustakaan Kantor Pusat BPK RI

Ari Ardiansyah<sup>1\*</sup>, Tezar Maula<sup>1</sup>, Agus Budiyantera<sup>1</sup>

<sup>1</sup>Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Widuri Jakarta, Indonesia.

Email : ari.ardiansyah@bpk.go.id, tezarmaula27@gmail.com, agusbudiyantera@kampuswiduri.ac.id

Article Info

Received: 29 August 2023

Accepted: 18 September 2023

Published: 30 September 2023

### ABSTRAK

Buku adalah kumpulan kertas atau bahan lain yang berisi teks, gambar, atau lampiran yang digabungkan menjadi satu pada satu ujungnya. Penanda buku yang pertama adalah judul, yaitu judul atau judul artikel, dan biasanya muncul di sampul depan. Judulnya yang singkat, padat, jelas dan biasanya menarik membuat orang tertarik untuk membaca buku tersebut. Terjadi permasalahan pada perpustakaan BPK RI penjaga perpustakaan kesulitan saat membuka kembali catatan data peminjaman buku dan kesulitan mengumpulkan judul buku yang paling banyak dipinjam mengingat banyaknya data peminjam yang terdapat di dalam buku catatan. Tujuan dari penelitian ini adalah untuk menganalisa akurasi dan mengklasifikasikan judul buku favorit berdasarkan catatan peminjaman buku pada perpustakaan kantor pusat BPK RI dan untuk membandingkan algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5) dalam menentukan judul buku berdasarkan peminatan. Dataset yang digunakan dalam penelitian ini berjumlah 644 data. Dari 644 data tersebut di proses atau dibagi menjadi data *training* sebanyak 451 dan data *testing* sebanyak 193 data menggunakan *tools* rapidminer. Dari hasil perbandingan algoritma *K-Nearest Neighbor* (K-NN) memperoleh hasil akurasi sebesar 94,82% sedangkan *Decision Tree* (C4.5) memperoleh hasil akurasi 95,34%. Dapat disimpulkan bahwa hasil akurasi *Decision Tree* lebih baik dari pada nilai akurasi *K-Nearest Neighbor*.

**Kata kunci :** buku, *decision tree* (c4.5), *k-nearest neighbor* (k-nn).

### ABSTRACT

A book is a bound collection of papers or other materials containing text, images, or appendices bound together at one end. The first bookmark is the title, which is the article's title or title and is normally found on the front cover. The title is brief, concise, straightforward, and usually engaging, piqueing the reader's interest in the book. The BPK RI library had an issue; the library keeper had trouble re-opening the book borrowing data records and gathering the most borrowed book titles due to the vast volume of borrower data stored in the notebook. The purpose of this study is to examine the accuracy and categorize preferred book titles based on book borrowing records at the BPK RI headquarters library and to compare the *K-Nearest Neighbor* (K-NN) algorithm with *Decision Tree* (C4.5) in determining book titles based on specialization. The dataset used in this research consisted of 644 data. Of the 644 data points, they were processed or divided into 451 training data and 193 testing data using the Rapidminer tool. From the comparison results, the *K-Nearest Neighbor* (K-NN) algorithm obtained an accuracy result of 94.82%, while the *Decision Tree* (C4.5) obtained an accuracy result of 95.34%. It can be concluded that the *Decision Tree* accuracy results are better than the *K-Nearest Neighbor* accuracy values.

**Keywords:** books, *decision tree* (c4.5), *k-nearest neighbor* (k-nn).

## 1. PENDAHULUAN

Buku adalah kumpulan kertas atau dokumen lain yang dihubungkan pada salah satu ujungnya, termasuk teks, gambar, atau lampiran. Setiap sisi buku disebut halaman.. Maksud dari slogan buku “*The Window of Knowledge*” ini adalah agar dengan membaca berbagai buku kita memperoleh wawasan dan pengetahuan yang luas, karena membaca buku memperluas wawasan kita. Kemudian merangsang otak, meningkatkan rasa ingin tahu dan keinginan untuk belajar (*Deepublish Store, 2022*).

Perpustakaan adalah organisasi yang mengumpulkan informasi tercetak dan terekam serta mengelola informasi tersebut dengan cara yang memenuhi kebutuhan intelektual penggunaannya melalui berbagai jenis interaksi informasi. Perpustakaan juga dapat dipahami sebagai tempat mengumpulkan informasi untuk memenuhi kebutuhan dasar manusia seperti ilmu pengetahuan, hiburan dan ibadah. Akibatnya, perpustakaan modern telah didefinisikan ulang sebagai tempat tersedianya informasi dalam bentuk apa pun, baik disimpan di gedung perpustakaan atau tidak. Di perpustakaan modern ini, selain koleksi tercetak, perpustakaan juga mempunyai sejumlah judul, koleksinya dalam bentuk digital atau data yang dapat diakses melalui jaringan komputer (Pengembangan dan Pengelolaan Sumber Daya) (Badan Pengawasan Keuangan dan Pembangunan Perpustakaan (BPKP, 2022).

Pustaka Riset BPK didirikan untuk mendukung penelitian di bidang pengelolaan keuangan dan akuntabilitas pemerintah dengan menyediakan data, informasi, dan pengetahuan tentang keuangan sektor publik sebanyak 21.852 jilid yang tersebar dalam 17.410 judul, serta e-book dan e-magazine. Perpustakaan BPK juga bekerjasama dengan Pusat Informasi dan Komunikasi BPK untuk menyediakan Laporan Hasil Pemeriksaan BPK kepada publik sebagai bahan penelitian. BPK menyadari bahwa pengelolaan dan akuntabilitas keuangan pemerintah yang kredibel tidak dapat dicapai hanya melalui audit, tetapi juga penting bagi publik untuk memahami pengelolaan dan akuntabilitas keuangan pemerintah. Pemahaman masyarakat tentang pengelolaan dan administrasi keuangan pemerintah akan meningkat bila didukung dengan literasi yang baik. Dalam pengelolaan keuangan dan pertanggungjawaban pemerintah, terdapat isu-isu penting seperti korupsi, pengelolaan utang pemerintah, pengelolaan aset negara, pengelolaan BUMN, dan kontrak yang berkaitan dengan aset pemerintah, yang seringkali menjadi perhatian publik. Masalah tersebut selalu menjadi perhatian BPK dan karenanya menjadi bahan pemeriksaan BPK.

Melihat hal tersebut, BPK memandang perlu melakukan terobosan dalam mendorong peran aktif masyarakat dalam melakukan penelitian tentang tata kelola dan akuntabilitas keuangan publik. Hasil penelitian ini akan bermanfaat dalam meningkatkan pemahaman masyarakat tentang tata kelola keuangan dan akuntabilitas pemerintah. Oleh karena itu, BPK harus membangun dan menyediakan perpustakaan dengan koleksi buku dan jurnal pengelolaan keuangan publik (BPK RI, 2022).

Perpustakaan BPK RI untuk memprediksi judul favorit berdasarkan catatan peminjaman. Memprediksi

judul buku favorit yang akan dipinjam untuk memudahkan inventarisasi buku dan penataan buku di ruang perpustakaan, sehingga juga memudahkan pembaca untuk menginginkan buku favoritnya dapat ditemukan. Permasalahan pustakawan membuat prediksi yaitu ketika pustakawan mengumpulkan informasi kembali tentang data peminjaman dan nama buku yang sering dipinjam, metode ini sangat sulit diterapkan karena banyaknya data di perpustakaan tentang orang yang dipinjam dalam buku perpustakaan BPK RI.

Penambahan data adalah suatu penarikan informasi dari pemrosesan dan eksplorasi data yang besar menggunakan persamaan matematis, kecerdasan buatan, dan pembelajaran mesin yang berguna bagi sains (Hasanah et al., 2022).

Penelitian dilakukan oleh Edi Kurniawan dengan penerapan data mining judul buku dalam menganalisis model peminjaman buku di perpustakaan dengan metode aturan asosiasi. Biasanya, data transaksi peminjaman buku perpustakaan hanya diperbolehkan oleh perpustakaan untuk terakumulasi pada sistem basis data tanpa adanya penggunaan atau pengolahan data lebih lanjut terhadap data yang sudah lama tersimpan. Penggunaan teknik penambahan data, metode legal yang dipadukan dengan *FP-Growth*, data ini akan sangat berguna. Karena data pinjaman perpustakaan dapat mengetahui informasi baru mengenai buku-buku yang sering dipinjam dan menemukan pola hubungan antar buku yang dipinjam, kemudian dapat digunakan untuk mengorganisasikan buku-buku sesuai dengan buku-buku yang ada. meminjam *template* untuk memudahkan pengunjung perpustakaan dalam mencari buku (Kurniawan, 2019).

Penelitian yang dilakukan oleh Angelina Hadriani tentang analisis data *bibliografi* pola peminjaman buku perpustakaan menggunakan metode algoritma asosiasi apriori. Penelitian ini menggunakan teknik untuk mengklasifikasikan data peminjaman buku perpustakaan.

Berdasarkan tren yang terjadi bersamaan dalam kegiatan kunjungan perpustakaan. Untuk mempermudah proses data mining, peneliti menggunakan Anaconda Navigator dari tabel data transaksi kredit tahun 2016-2019 yang menjadi pokok bahasan penelitian ini. Menguji algoritma apriori menggunakan *Python* dan aplikasi Jupiter disertakan dengan Anaconda. Selama pengujian, peneliti mengonversi data ke dalam format file CSV. Berdasarkan hasil pengujian dan analisis penerapan algoritma apriori, dapat disimpulkan bahwa algoritma apriori dapat digunakan untuk menentukan buku mana yang banyak terdapat di perpustakaan. Proses peminjaman buku dan mencari rekanan untuk meminjam buku siswa Sandikta adalah (1-2 3600 detik - Pocong atau biasa disebut Pocong) dan (12-20 4R - *Always With Me*) (Hadriani et al., 2022).

Penelitian yang dilakukan oleh Ari Ardiansyah dengan judul penerapan data mining dalam menentukan judul buku berdasarkan peminatan menggunakan algoritma *K-Nearest Neighbor* (K-NN) pada perpustakaan kantor pusat BPK RI. Pada proses pengolahan data dan penerapan data mining dalam menentukan judul buku berdasarkan peminatan ada langkah-langkah penelitian yang harus dikerjakan mulai dari mencari *dataset* buku perpustakaan, proses transformasi data, pemilihan metode yang digunakan, menentukan nilai *cross validation*, pengujian *confusion*

matrix, diketahui hasil akurasi dari algoritma tersebut adalah sebesar 95,72% (Ari Ardiansyah, 2022).

Tujuan dari penelitian yang dilakukan peneliti berdasarkan permasalahan dan penelitian sebelumnya adalah untuk menganalisa akurasi dan memprediksi judul buku favorit berdasarkan catatan peminjaman buku dan untuk membandingkan algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5) dalam menentukan judul buku berdasarkan peminatan pada perpustakaan kantor pusat BPK RI.

Ada beberapa alasan peneliti memilih algoritma *K-Nearest Neighbor* (K-NN) di antaranya adalah sederhana dan mudah dipahami: Algoritma K-NN relatif sederhana dan mudah dipahami. Konsepnya sangat intuitif, yaitu mencari kelas mayoritas dari tetangga terdekat dalam ruang fitur. Ini membuatnya menjadi pilihan yang baik untuk pemula dalam pemodelan data atau untuk kasus sederhana yang tidak memerlukan kompleksitas model yang tinggi. Non-parametrik dan pembelajaran *lazy*: K-NN adalah algoritma non-parametrik, yang berarti tidak ada asumsi tertentu tentang distribusi data. Ini membuatnya lebih fleksibel dalam menangani data yang tidak terstruktur atau kompleks. Selain itu, K-NN juga merupakan algoritma pembelajaran *lazy*, yang berarti tidak memerlukan proses pelatihan yang signifikan. Model hanya perlu menyimpan data pelatihan sebagai referensi untuk mengklasifikasikan data baru. Toleran terhadap data *noise*: K-NN memiliki toleransi yang cukup tinggi terhadap data yang tidak sempurna atau berisik. Karena algoritma ini mempertimbangkan tetangga terdekat, beberapa data yang salah klasifikasi atau *outlier* tidak akan secara signifikan mempengaruhi hasil akhir. Ini membuatnya menjadi pilihan yang baik ketika data yang digunakan memiliki tingkat kebisingan yang tinggi atau ketidakpastian yang cukup besar.

Sedangkan alasan peneliti memilih algoritma *Decision Tree* (C4.5) adalah kemampuan untuk mengekstraksi aturan yang dapat dipahami: *Decision Tree*, seperti algoritma C4.5, memiliki kemampuan yang baik dalam mengekstraksi aturan yang dapat dipahami. Struktur pohon keputusan yang dihasilkan memberikan interpretabilitas yang tinggi, yang memudahkan interpretasi dan pengambilan keputusan berdasarkan aturan yang dihasilkan oleh algoritma. Mendukung fitur diskrit dan kontinu: C4.5 dapat menangani baik fitur diskrit maupun kontinu (Muryono et al., 2021).

Ini memungkinkan Anda untuk bekerja dengan berbagai jenis data, termasuk data kategorikal dan numerik. Algoritma ini juga dapat menangani data yang hilang dengan baik dengan menggunakan metode seperti penggantian nilai rata-rata atau menggunakan variabel terbanyak. Mampu menangani multiklasifikasi dan klasifikasi berimbang: *Decision Tree* adalah salah satu algoritma yang efektif dalam menangani masalah multiklasifikasi, yaitu ketika terdapat lebih dari dua kelas yang harus diprediksi. Selain itu, C4.5 juga dapat mengatasi masalah klasifikasi berimbang di mana distribusi kelas dalam data tidak seimbang. Dengan menggunakan teknik seperti penyesuaian bobot atau penggunaan pengukuran pemisahan yang tepat, algoritma ini dapat menghasilkan model yang dapat menangani kasus klasifikasi berimbang dengan baik.

Pada penelitian ini mempunyai keterbaruan dibandingkan penelitian terdahulu yaitu dari segi dataset yang digunakan, studi kasus penelitian, dan algoritma yang penulis gunakan ada dua buah metode yaitu *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5).

## 2. METODOLOGI

### Teknik Pengumpulan Data

#### 1. Observasi

Observasi dilakukan dengan mengunjungi langsung tempat penelitian dalam hal ini di Perpustakaan Kantor Pusat BPK RI. Peneliti melakukan observasi untuk mengetahui tentang lingkungan perpustakaan, untuk mengajukan izin penelitian, dan untuk mengumpulkan data.

#### 2. Wawancara

Wawancara dilakukan kepada kepala perpustakaan, guna mendapatkan informasi mengenai permasalahan, sistem yang berjalan, dan data peminjam buku.

#### 3. Studi Pustaka

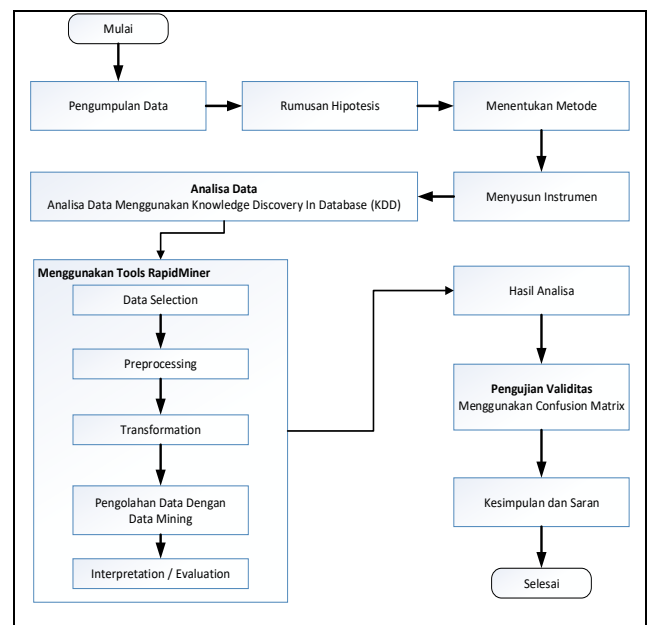
Penelitian kepustakaan dilakukan dengan mencari referensi pada jurnal, buku, dan lain-lain yang terdahulu. Untuk mengumpulkan sumber data yang diperlukan untuk penulisan penelitian.

### Metode Penelitian

Metode yang peneliti gunakan untuk menentukan judul buku berdasarkan peminatan pada perpustakaan kantor pusat BPK RI adalah algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5) (Budiyantara et al., 2020).

### Langkah Langkah Penelitian

Langkah langkah penelitian dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Langkah-Langkah Penelitian

Penjelasan tahapan-tahapan penelitian pada gambar 1 sebagai berikut:

#### 1. Pengumpulan Data

Pengumpulan data adalah untuk mendapatkan informasi yang akurat untuk memperoleh tujuan dari penelitian. Langkah yang dilakukan dalam mengumpulkan data pada penelitian ini adalah dengan melakukan observasi, wawancara dan studi pustaka (Muryono et al., 2020).

2. Rumusan Hipotesis

Rumusan hipotesis ini merupakan jawaban sementara terhadap suatu masalah yang masih berupa praduga, karena kebenarannya harus diuji terlebih dahulu kebenarannya.

3. Menentukan Metode

Pada langkah ini, menentukan metode yang bertujuan untuk menganalisis dan mengklasifikasikan data peminjam buku pada perpustakaan kantor BPK RI.

4. Menyusun Instrumen

Menyusun instrumentasi adalah tahap yang dimana akan dijelaskan alat-alat yang akan digunakan saat melakukan penelitian.

5. Analisa Data

Analisa data yang peneliti gunakan menggunakan *Knowledge Discovery In Database (KDD)* yang terdiri dari *data selection, pre-processing/cleaning, transformation, data mining, interpretation / evaluation.*

6. Menggunakan Tools RapidMiner

*Tools RapidMiner* digunakan untuk memproses analisa data mulai dari tahap *data selection, pre-processing/cleaning, transformation, data mining, interpretation/evaluation.*

7. Hasil analisa

Pada langkah ini dimana suatu analisa telah dilakukan sehingga mendapatkan hasil akurasi atau informasi berdasarkan data yang telah dianalisa dengan menggunakan *tools RapidMiner.*

8. Pengujian Validitas

Pengujian validitas menggunakan rumus dari *confusion matrix* dilakukan untuk menguji hasil akurasi dari perbandingan yang didapatkan pada *tools RapidMiner* antara metode K-NN dengan C4.5.

9. Kesimpulan dan Saran

Pada langkah terakhir peneliti memberikan kesimpulan berdasarkan tujuan dari pada penelitian, dan memperoleh hasil yang didapat dari penelitian yang dilakukan dari awal hingga akhir dari penelitian dimana untuk mendapatkan suatu hasil yang dilakukan pada perbandingan algoritma *K-Nearest Neighbor (K-NN)* dengan *Decision Tree (C4.5)* untuk menentukan judul buku berdasarkan peminatan pada perpustakaan kantor pusat BPK RI.

3. HASIL DAN PEMBAHASAN

Proses Pengolahan Data Mining

Dalam penelitian ini data yang diperoleh berasal dari Perpustakaan Kantor Pusat BPK RI. Data mentah yang didapat berjumlah 764 data bulan Januari 2021 sampai dengan Desember 2022 dengan format *microsoft excel*, kemudian data mentah tersebut diolah menjadi data set berjumlah 644 data. Dari 644 data tersebut di proses menjadi data *training* sebanyak 451 dan data *testing* sebanyak 193 data menggunakan *tools rapidminer*. Dapat dilihat pada gambar 2 dan 3 dibawah ini:

1. Data Mentah Daftar Peminjaman Buku

Data mentah adalah data yang diambil oleh peneliti dari kantor perpustakaan BPK RI dan belum diproses menjadi dataset atau data yang siap untuk digunakan dalam penelitian.

Gambar 2. Contoh Data Mentah Daftar Peminjaman Buku Perpustakaan

2. Data Set Daftar Peminjaman Buku Perpustakaan

Data set adalah data yang siap untuk digunakan dalam penelitian.

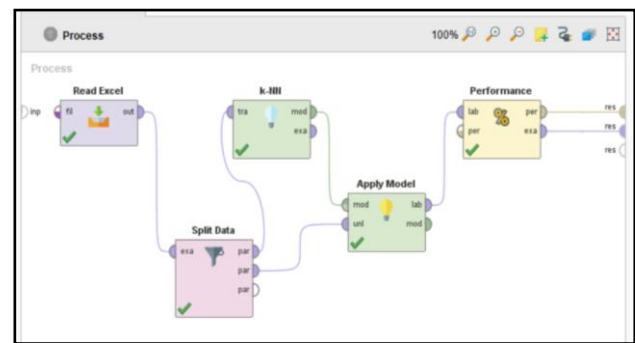
Gambar 3. Contoh DataSet Daftar Peminjaman Buku Perpustakaan

3. Evaluasi Model K-NN dan Decision Tree (C4.5) Menggunakan Tools RapidMiner

Tujuan dari penelitian adalah untuk menganalisa akurasi dan memprediksi judul buku favorit dengan menerapkan teknik klasifikasi data mining dengan Algoritma *K-Nearest Neighbors* dan *Decision Tree*.

Proses Import Data K-NN Dengan Tools RapidMiner

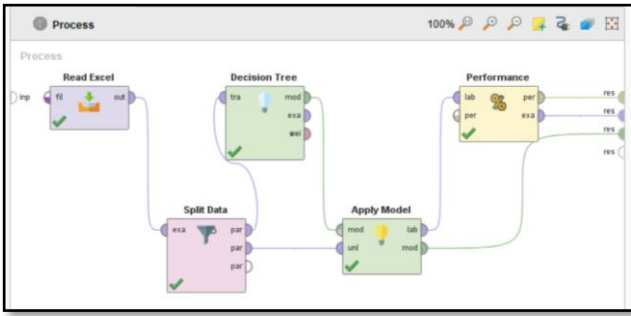
Untuk menghasilkan akurasi menggunakan metode K-NN maka perlu dilakukan proses desain model seperti gambar 4 dibawah ini:



Gambar 4. Proses Import Data K-NN Dengan Tools RapidMiner

Proses Import Data C4.5 Dengan Tools RapidMiner

Untuk menghasilkan akurasi menggunakan metode C4.5 maka perlu dilakukan proses desain model seperti gambar 4 dibawah ini:



Gambar 5. Proses Import Data C4.5 Dengan Tools RapidMiner

1. *Read Excel* : Operator ini digunakan untuk membuat informasi dari *spreadsheet Microsoft Excel*.
2. *Split Data*: Operator yang digunakan untuk membagi data antara data *training* dengan data *testing*, Dalam penelitian ini membagi datanya menjadi 0,7 untuk data *training* dan 0.3 untuk data *testing*.
3. K-NN dan C4.5: Algoritma yang dipakai untuk melakukan penelitian pada Perpustakaan Kantor Pusat BPK RI.
4. *Apply Model* : Operator yang digunakan untuk penghubung algoritma K-NN dan C4.5 ke *performance*.
5. *Performance* : Operator yang dipakai buat mengukur *performance* akurasi dari model dari algoritma K-NN dan C4.5.

**4. Evaluasi Model K-NN**

Proses pada penelitian ini dengan melakukan evaluasi model K-NN. Dapat kita lihat pada gambar 6 yang peneliti *capture* dari *tools* RapidMiner sebagai berikut:

accuracy: 94,82%			
	true Tidak Minat	true Minat	class precision
pred. Tidak Minat	179	10	94,71%
pred. Minat	0	4	100,00%
class recall	100,00%	28,57%	

Gambar 6. Evaluasi Model K-NN

**5. Evaluasi Model C4.5**

Proses pada penelitian ini dengan melakukan evaluasi model C4.5 dengan menggunakan *cross validation*. Dapat kita lihat pada gambar 7 yang peneliti *capture* dari *tools* RapidMiner sebagai berikut:

accuracy: 95,34%			
	true Tidak Minat	true Minat	class precision
pred. Tidak Minat	179	9	95,21%
pred. Minat	0	5	100,00%
class recall	100,00%	35,71%	

Gambar 7. Evaluasi Model C4.5

**6. PerformaVector Dari Hasil Akurasi Algoritma K-NN**

Setelah peneliti melakukan proses evaluasi model menggunakan *cross validation* diperoleh keterangan tentang *performancevector*. Dapat dilihat pada gambar 8 berikut ini:

```

PerformanceVector
PerformanceVector:
accuracy: 94.82%
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   10
Minat:         0    4
precision: 100.00% (positive class: Minat)
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   10
Minat:         0    4
recall: 28.57% (positive class: Minat)
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   10
Minat:         0    4
AUC (optimistic): 0.997 (positive class: Minat)
AUC: 0.962 (positive class: Minat)
AUC (pessimistic): 0.927 (positive class: Minat)
    
```

Gambar 8. PerformanceVector Algoritma K-NN

**7. PerformaVector Dari Hasil Akurasi Algoritma C4.5**

Setelah peneliti melakukan proses evaluasi model menggunakan *cross validation* diperoleh keterangan tentang *performancevector*. Dapat dilihat pada gambar 9 berikut ini:

```

PerformanceVector
PerformanceVector:
accuracy: 95.34%
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   9
Minat:         0    5
precision: 100.00% (positive class: Minat)
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   9
Minat:         0    5
recall: 35.71% (positive class: Minat)
ConfusionMatrix:
True:  Tidak Minat  Minat
Tidak Minat:  179   9
Minat:         0    5
AUC (optimistic): 1.000 (positive class: Minat)
AUC: 0.714 (positive class: Minat)
AUC (pessimistic): 0.429 (positive class: Minat)
    
```

Gambar 9. PerformanceVector Algoritma C4.5

**8. Evaluasi Model Pengujian Menggunakan Confusion Matrix**

Dibawah ini merupakan perhitungan akurasi dari gambar 6 menggunakan algoritma K-NN dan gambar 7 menggunakan algoritma C4.5.

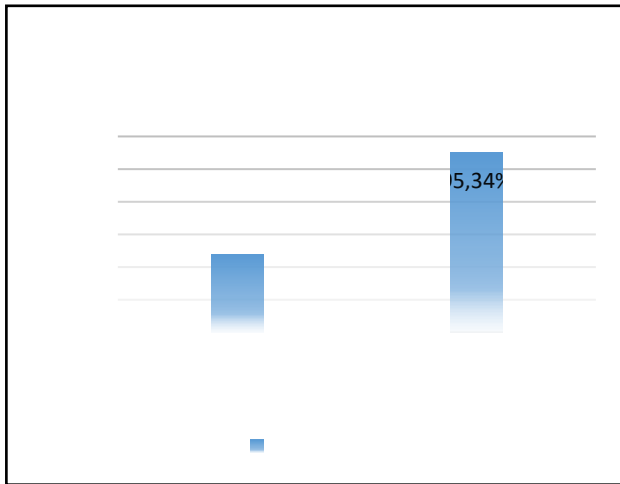
$$Accuracy\ K - NN = \frac{TP+TN}{TP+TN+FP+FN} + \frac{4+179}{4+179+0+10} = \frac{183}{193} = 0,948186 = 94,82\% \dots\dots\dots(1)$$

$$Accuracy\ C4.5 = \frac{TP+TN}{TP+TN+FP+FN} + \frac{5+179}{5+179+0+9} = \frac{184}{193} = 0,953367 = 95,34\% \dots\dots\dots(2)$$

**Grafik Perbandingan Algoritma K-Nearest Neighbor (K-NN) Dengan Decision Tree (C4.5)**

Berikut ini merupakan hasil perbandingan Algoritma *K-Nearest Neighbor* (K-NN) akurasinya 94,82% dengan *Decision Tree* (C4.5) akurasinya 95,34% dapat dilihat pada gambar 10 berikut ini:





Gambar 10. Grafik Hasil Akurasi dari Perbandingan Algoritma K-NN dengan C4.5

#### 4. KESIMPULAN

Kesimpulan pada penelitian perbandingan algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5) untuk menentukan judul buku berdasarkan peminatan pada perpustakaan kantor pusat BPK RI adalah sebagai berikut: Pada proses pengolahan data dan penerapan data mining dalam menentukan judul buku berdasarkan peminatan ada langkah-langkah penelitian yang harus dikerjakan mulai dari mencari *dataset* buku perpustakaan, proses transformasi data, pemilihan metode yang digunakan, menentukan nilai *cross validation*, pengujian *confusion matrix*, diketahui hasil akurasi dari algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5). *Data mining* dapat diterapkan dalam perbandingan antara algoritma *K-Nearest Neighbor* (K-NN) dengan *Decision Tree* (C4.5) sehingga dapat memprediksi hasil akurasi judul buku berdasarkan peminatan pada perpustakaan kantor pusat BPK RI. Hasil akurasi algoritma *Decision Tree* (C4.5) memiliki tingkat akurasi (95,50%) yang lebih baik dibandingkan dengan algoritma *K-Nearest Neighbor* (K-NN) (94,88%). Saran untuk penelitian selanjutnya gunakan *tools* selain *rapidminer* yaitu seperti bahasa *python* dengan *library machine learning*, bahasa *R*, *matlab*, dll dan juga penelitian ini dapat dievaluasi kembali dengan metode atau algoritma yang lain sebagai bahan perbandingan.

#### DAFTAR PUSTAKA

Ari Ardiansyah. (2022). *penerapan data mining dalam menentukan judul buku berdasarkan peminatan*

*menggunakan algoritma K-Nearest Neighbor (K-NN) pada perpustakaan kantor pusat BPK RI*. STMIK Widuri.

- Badan Pengawasan Keuangan dan Pembangunan Perpustakaan (BPKP). (2022). *Pengertian, Tujuan dan Peran Perpustakaan*. <https://www.bpkp.go.id/pustakabpkp/index.php?p=pengertian,tujuan>
- BPK RI. (2022). *Peresmian Perpustakaan Riset Bpk Untuk Mendukung Pemahaman Publik Tentang Tata Kelola Dan Akuntabilitas Keuangan Negara*. <https://www.bpk.go.id/news/peresmian-perpustakaan-ri-set-bpk-untuk-mendukung-pemahaman-publik-tentang-tata-kelola-dan-akuntabilitas-keuangan-negara>
- Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020). *Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu*. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 265–270. <https://doi.org/10.33480/jitk.v5i2.1214>
- Deepublish Store. (2022). *Pengertian Buku, Jenis dan Manfaatnya*. <https://deepublishstore.com/pengertian-buku/>
- Hadriani, A., Informatika, T., Pamulang, U., & Selatan, T. (2022). *Analisa Data Pola Peminjaman Buku Perpustakaan Dengan Metode Algoritma Asosiasi Apriori*. I(1).
- Hasanah, F., Suprapti, T., Rahaningsih, N., & Ali, I. (2022). *Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Buku Berdasarkan Peminatan*. *Jurnal Accounting Information System (AIMS)*, 5(1), 102–111. <https://doi.org/10.32627/aims.v5i1.467>
- Kurniawan, E. (2019). *Implementasi Data Mining Dalam Analisa Pola Peminjaman Buku Di Perpustakaan Menggunakan Metode Association Rule*. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(1), 89–96. <https://doi.org/10.33330/jurteksi.v5i1.324>
- Muryono, T. T., Irwansyah, I., & Budiyantara, A. (2020). *Penentuan Penerimaan Pegawai Menggunakan Metode Perbandingan Eksponensial (Mpe)*. *Infotech: Journal of Technology Information*, 6(2), 57–62. <https://doi.org/10.37365/jti.v6i2.98>
- Muryono, T. T., Taufik, A., & Irwansyah, I. (2021). *Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit*. *Infotech: Journal of Technology Information*, 7(1), 35–40. <https://doi.org/10.37365/jti.v7i1.104>