

# **METODE *DECISION TREE* UNTUK KLASIFIKASI HASIL SELEKSI KOMPETENSI DASAR PADA CPNS 2019 DI ARSIP NASIONAL REPUBLIK INDONESIA**

*Rama Dian Syah*  
*Fakultas Teknologi Informasi, Universitas Gunadarma,*  
*Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat*  
*rama\_ds@staff.gunadarma.ac.id*

## **Abstrak**

*Pelaksanaan seleksi Calon Pegawai Negeri Sipil (CPNS) 2019 di institusi Arsip Nasional Republik Indonesia (ANRI) dilakukan berdasarkan peraturan Badan Kepegawaian Negara di lingkungan instansi pemerintah. Seleksi tes cpns dilakukan dengan beberapa tahapan. Seleksi Kompetensi Dasar (SKD) merupakan tahapan yang diuji berdasarkan parameter penilaian Tes Wawasan Kewarganegaraan (TWK), Tes Intelegensi Umum (TIU), dan Tes Karakter Kepribadian (TKP). Algoritma *Decision Tree* diimplementasikan untuk mengklasifikasikan hasil tes Seleksi Kemampuan Dasar (SKD). Hasil dari SKD dianalisis menggunakan *Data Mining*. Metode *Data Mining* yang digunakan yaitu *Decision Tree*. Berdasarkan penelitian yang dilakukan pada 344 peserta SKD di Arsip Nasional Republik Indonesia adalah nilai *Accuracy* sebesar 92.23%, *Classification Error* sebesar 7.77%, *Kappa* sebesar 0.879, *Recall* sebesar 94.84%, dan *Precision* sebesar 95.79%.*

**Kata Kunci:** *Decision Tree, Seleksi Kompetensi Dasar, CPNS*

## **Abstract**

*2019 Civil Servant Enrollment Test (CPNS) held by the National Archives of the Republic of Indonesia (ANRI) is based on the National Civil Service Agency (BKN) regulation within government agencies. Civil Servant Enrollment Test (CPNS) is done in several stages. The Basic Competency Selection (SKD) is a stage that tested based on assessment parameters of the National Insight Test (TWK), General Intelligence Test (TIU), and Personal Characteristics Test (TKP). *Decision Tree* algorithm is implemented to classify the results of Basic Competency Selection (SKD). The results of Basic Competency Selection (SKD) are analyzed using *Data Mining*. The *Data Mining* method used is *Decision Tree*. Based on research conducted on 344 Basic Competency Selection (SKD) participants in the National Archives of the Republic of Indonesia, the value of *Accuracy* is 92.23%, *Classification Error* is 7.77%, *Kappa* is 0.879, *Recall* is 94.84%, and *Precision* is 95.79%.*

**Keywords:** *Decision Tree, Basic Competency Selection, Civil Servant Enrollment Test*

## **PENDAHULUAN**

Seleksi Calon Pegawai Negeri Sipil (CPNS) merupakan hal yang sangat ditunggu oleh banyak masyarakat untuk mengikuti seleksi menjadi Aparatur Sipil Negara (ASN). Pengadaan seleksi CPNS dilakukan oleh

Badan Kepegawaian Nasional berdasarkan peraturan Badan Kepegawaian Negara (BKN) Nomor 14 tahun 2018 [1]. Jumlah pendaftar yang sangat besar membutuhkan seleksi yang sangat selektif untuk menghasilkan calon ASN yang berkompeten. Formasi dan

persyaratan ditentukan oleh instansi yang melakukan seleksi CPNS.

Seleksi CPNS dilakukan dengan beberapa tahap yaitu seleksi administrasi, seleksi kompetensi dasar (SKD), dan seleksi kompetensi bidang (SKB). Peserta yang telah lulus dari seleksi administrasi bisa mengikuti seleksi kompetensi dasar. Penilaian yang dilakukan pada tahap seleksi kompetensi dasar dilakukan berdasarkan tiga parameter penilaian yaitu Tes Wawasan Kewarganegaraan (TWK), Tes Intelegensi Umum (TIU), dan Tes Karakter Kepribadian (TKP). Peserta yang telah dinyatakan lulus pada SKD dapat mengikuti seleksi kompetensi bidang (SKB).

Data hasil tes diumumkan melalui situs SSCN- Badan Kepegawaian Negara atau melalui situs instansi masing-masing. Data dapat dilihat oleh peserta seleksi dan masyarakat umum sehingga dari data tersebut dapat diambil informasi mengenai kriteria penilaian dari seleksi di berbagai instansi pemerintahan. Data yang banyak diolah sehingga menghasilkan pengetahuan menggunakan suatu metode.

*Data mining* merupakan ekstraksi pola dan informasi untuk menghasilkan pengetahuan dari data yang besar [2]. Teknik *data mining* yang paling sederhana dan banyak digunakan yaitu klasifikasi. Teknik klasifikasi merupakan teknik untuk penyusunan data ke dalam kategori agar mudah digunakan dan efisien [3]. Teknik klasifikasi dapat diimplementasikan menggunakan algoritma *Decision Tree*. Struktur *Decision tree*

merupakan struktur yang menyerupai pohon, dimana setiap simpul menunjukkan atribut, setiap cabang menunjukkan hasil tes, dan daun mewakili kelas atau distribusi kelas [4].

Beberapa penelitian telah dilakukan terhadap *data mining* menggunakan algoritma *Decision Tree*. Siregar dkk [5] melakukan penelitian dengan mengklasifikasi mahasiswa berdasarkan kemampuan ekonomi untuk biaya kuliah. Hasil penelitiannya adalah klasifikasi mahasiswa dikategorikan menjadi tujuh kategori. Nilai akurasi yang didapatkan menggunakan algoritma *Decision Tree* yaitu 80.52%.

Penelitian lainnya, Topirceanu dan Grosseck [6], melakukan penelitian dengan mengklasifikasi tingkat partisipasi kursus *online* berdasarkan pada ciri-ciri individu siswa. Data yang diambil dari hasil kuisioner para peserta kursus *online*. Hasil penelitiannya adalah algoritma *Decision Tree* mampu mengklasifikasikan 8 kelas sesuai dengan pola dasar siswa pada penyelesaian kursus online.

Pada penelitian ini, algoritma *data mining* yang digunakan yaitu *decision tree*. Algoritma *decision tree* mampu mengklasifikasikan variable data berdasarkan kelas target sehingga dapat menghasilkan pengetahuan. Algoritma *Decision Tree* diimplementasikan untuk mengklasifikasikan hasil tes Seleksi Kemampuan Dasar (SKD) pada CPNS 2019 di institusi Arsip Nasional Republik Indonesia (ANRI). Jangkauan nilai hasil tes Seleksi Kemampuan Dasar (SKD) akan diklasifikasikan

berdasarkan kategori P/L, P, TL dan TH. Data yang digunakan berjumlah 344 peserta seleksi CPNS.

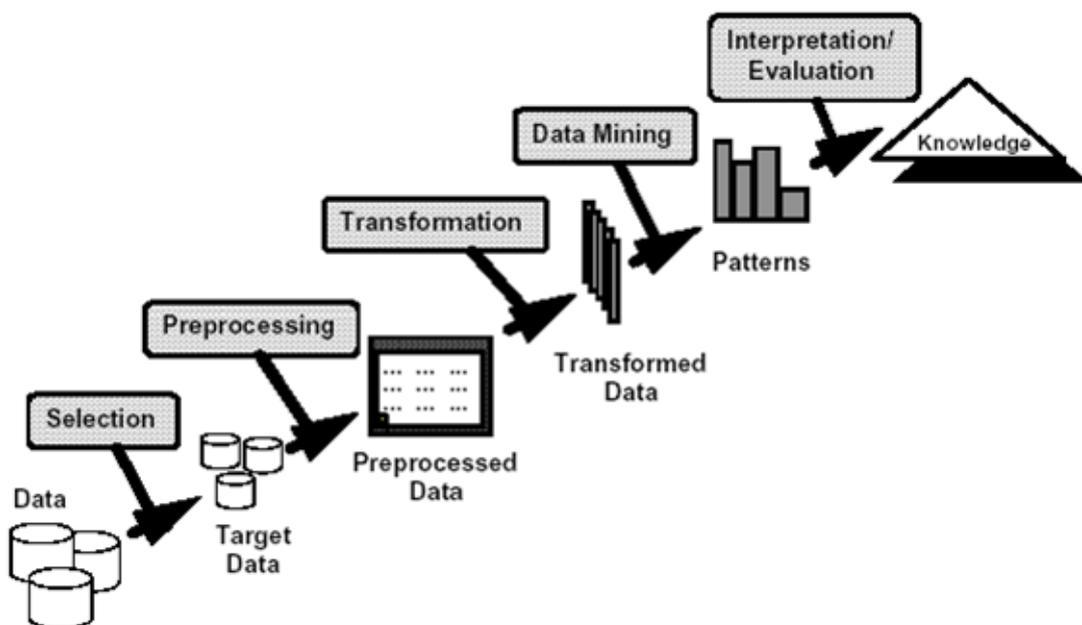
## METODE PENELITIAN

Metode penelitian yang dilakukan dengan menggunakan *Knowledge Discovery in Databases* (KDD). KDD merupakan metode untuk mencari dan mengidentifikasi pola dalam data yang dapat digunakan sebagai pengetahuan. Terdapat 5 tahapan pada KDD yaitu *Selection*, *Preprocessing*, *Transformasi Transformation*, *Data Mining*, dan *Evaluation*. Struktur KDD dapat dilihat pada Gambar 1. *Selection* merupakan pengambilan data yang berkaitan dengan analisis yang dilakukan dari database. *Preprocessing* merupakan proses pembersihan data dengan menghilangkan duplikasi data.

*Transformation* merupakan proses perubahan data menjadi bentuk data yang valid dan siap dilakukan untuk proses *data mining*. *Data Mining* adalah proses implementasi metode atau algoritma yang sesuai dengan informasi atau pengetahuan yang akan digali. *Evaluasi* merupakan proses yang dilakukan dengan mengevaluasi hasil dari proses *data mining*.

### Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data peserta Seleksi Kemampuan Dasar (SKD) pada CPNS di Arsip Nasional Republik Indonesia (ANRI) tahun 2019. Penelitian ini berfokus pada 344 peserta seleksi. Dataset terdiri dari 344 data dengan 8 atribut yang menggambarkan hasil dari nilai tes SKD. Berikut atribut data terdapat pada Tabel 1.



Gambar 1. Knowledge Discovery in Database

**Tabel 1. Atribut Data**

No	Atribut	Tipe Data
1	No Peserta	17 Digit Number
2	Nama	String
3	Pendidikan	Alphanumeric
4	TWK	Number
5	TIU	Number
6	TKP	Number
7	Total	Number
8	Keterangan	P/L, P, TL, TH

### **Persiapan Data**

Persiapan atau *Preprocessing Data* berguna untuk manipulasi dataset sebelum digunakan pada model. Beberapa persiapan data yang dilakukan adalah pembersihan data (*Data Cleaning*) dengan menghilangkan redundansi data dan memperbaiki data yang tidak konsisten. Pemilihan data dari banyaknya jenis formasi yang ada pada data seleksi CPNS ANRI 2019 yang akan diujikan menggunakan Data Mining.

Transformasi dataset menjadi bentuk yang lebih cocok untuk Data Mining agar lebih mudah untuk mendapatkan pengetahuan dari dataset.

### **Pengembangan Model**

Model yang dikembangkan menggunakan algoritma *C4.5 Decision Tree*. Tools pemodelan menggunakan framework RapidMiner.

Struktur *Decision Tree* dan nilai akurasi akan didapatkan setelah dataset

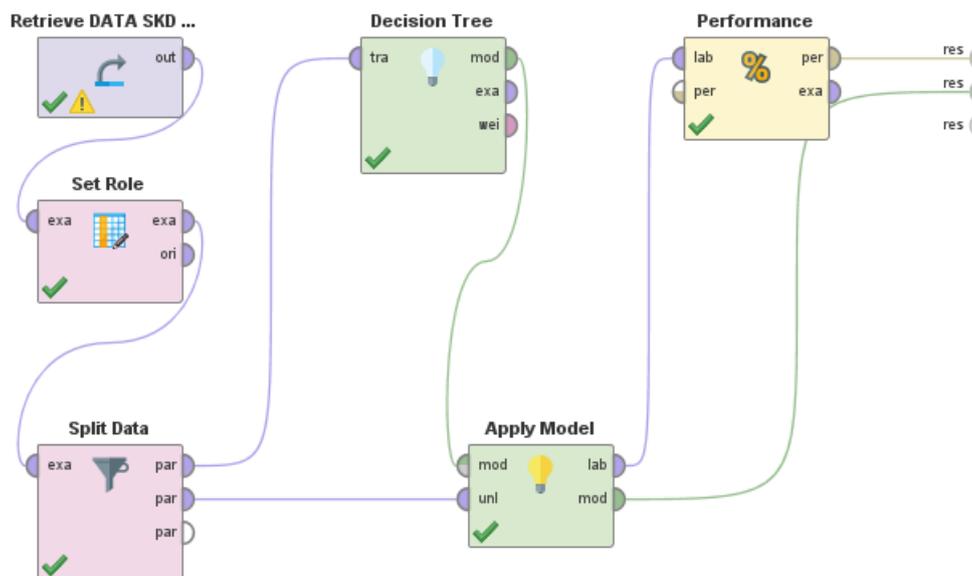
diproses pada RapidMiner. Nilai peforma diukur dengan *confusion matrix*.

### **HASIL DAN PEMBAHASAN**

Data yang digunakan untuk pengklasifikasian hasil tes Seleksi Kompetensi Dasar (SKD) CPNS di Arsip Nasional Republik Indonesia tahun 2019 sebanyak 344 data dengan 8 atribut. Proses klasifikasi dilakukan dengan menggunakan RapidMiner versi 9 untuk mengolah data tersebut.

### **Model Algoritma Decision Tree**

Model yang dirancang pada RapidMiner menggunakan beberapa operator yaitu *Set Role*, *Split Data*, *Decision Tree*, *Apply Model*, dan *Performance*. Berikut model algoritma *Decision Tree* pada RapidMiner terdapat pada Gambar 2. Pada Gambar 2 terdapat beberapa operator seperti operator *Retrieve Data* yang merupakan data yang dipakai untuk proses Data Mining.



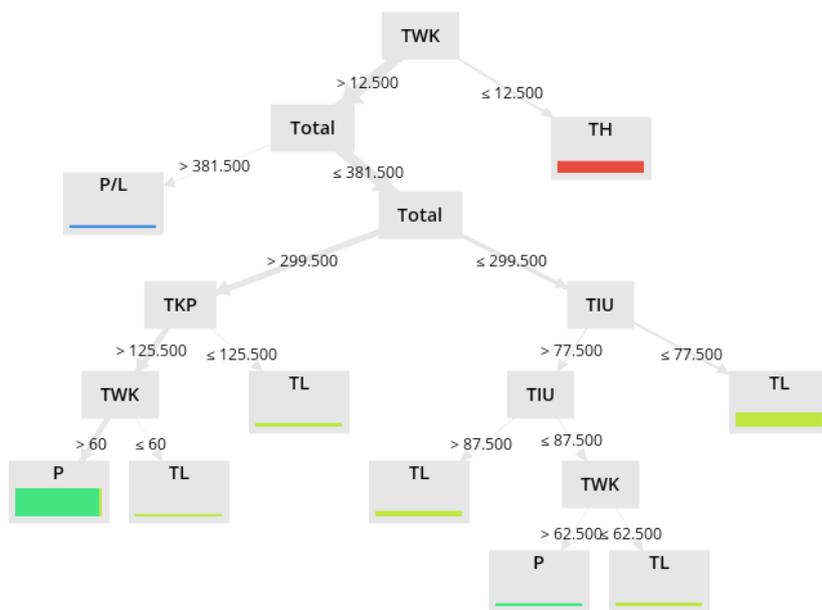
**Gambar 2. Model Algoritma *Decision Tree***

Operator *Set Role* digunakan untuk mengubah peran satu atribut atau lebih dimana atribut yang diubah adalah atribut ‘Keterangan’ dengan *Target Role* menjadi label.

Operator *Split Data* digunakan untuk membagi data menjadi partisi dengan *ratio* 0.7 dan 0.3. Operator *Decision Tree* digunakan untuk membuat model *Decision Tree* yang merupakan model yang pada penelitian ini. Operator *Apply Model* digunakan untuk menerapkan model *Decision Tree* pada dataset. Operator *Performance* digunakan untuk evaluasi kinerja dari model yang diterapkan. Evaluasi kinerja yang digunakan yaitu *Accuracy*, *Classification Error*, *Kappa*, *Recall*, dan *Precision*.

### **Model Tree**

*Model Tree* merupakan struktur pohon yang dihasilkan dari proses data SKD menggunakan Algoritma *Decision Tree* di RapidMiner. Pada struktur pohon dapat dilihat pembagian kelas menjadi P/L, P, TL, dan TH berdasarkan nilai TIU, TWK, TKP, dan Total. Dari model *Tree* tersebut didapatkan penjabaran hasil klasifikasi. Berikut model *Tree* dari algoritma *Decision Tree* di RapidMiner dapat dilihat pada Gambar 3 dan hasil klasifikasi pada Tabel 2. Pada Gambar 3 merupakan model *tree* yang didapatkan dari proses yang dilakukan di RapidMiner. *Root* atau akar dari model *tree* tersebut yaitu TWK sedangkan *Leaf* atau daunnya yaitu P/L, TL, P dan TH.



Gambar 3. Model Tree

Tabel 2. Hasil Klasifikasi

No	Kelas	Nilai
1	P/L	Total > 381
2	TH	TWK < 12
3	P	Total > 299 && TKP > 125 && TWK > 60 Total ≤ 299 && TIU ≤ 87 && TWK > 62
4	TL	Total > 299 && TWK ≤ 60 && TKP > 125 Total > 299 && TKP ≤ 125 Total ≤ 299 && TIU > 87 Total ≤ 299 && TIU ≤ 87 && TWK ≤ 62

Tabel 2 menjelaskan hasil klasifikasi yang diperoleh pada model *tree*. Hasil klasifikasi terdiri dari beberapa kondisi nilai untuk menentukan kelas.

Jika salah satu kondisi nilai terpenuhi, maka nilai tersebut masuk ke kelas yang sesuai. Kelas P/L dan TH terdapat 1 kondisi. Kelas P memiliki 2 kondisi. Kelas TL memiliki 4 kondisi.

### Performa Model *Decision Tree*

Performa didapat dari model *Decision Tree* yang didesain untuk mengklasifikasikan hasil tes SKD. Performa diukur berdasarkan *Confusion Matrix*. Beberapa performa yang didapat yaitu *accuracy*, *classification error*, *kappa*, *recall*, dan *precision*. Berikut *Confusion Matrix* terdapat pada Gambar 4 dan Performa Model pada Tabel 3.

True:	P/L	P	TL	TH
P/L:	3	0	0	0
P:	0	44	7	0
TL:	0	1	31	0
TH:	0	0	0	17

Gambar 4. Confusion Matrix

Tabel 3. Performa Model *Decision Tree*

No	Jenis Performa	Nilai
1	<i>Accuracy</i>	92.23%
2	<i>Classification error</i>	7.77%
3	<i>Kappa</i>	0.879
4	<i>Recall</i>	94.84%
5	<i>Precision</i>	95.79%

Beberapa performa pada Tabel 3 didapatkan dari *Confusion Matrix* pada Tabel 3. Performa *accuracy* merupakan nilai ketepatan dari model yang digunakan. Nilai akurasi yang didapatkan adalah 92.23% dimana model cukup baik dalam hasil prediksi. Performa *Classification error* merupakan nilai kesalahan yang didapatkan dari model yang digunakan. Nilai kesalahan yang didapatkan 7.77% di mana kesalahan hanya sedikit pada model yang digunakan. Performa *kappa* merupakan normalisasi nilai *accuracy*. Nilai *Kappa* yang didapatkan adalah 0.879 dimana angka mendekati 1 berarti model mendekati sempurna. Nilai *Recall* dan *Precision* yang didapatkan adalah 94.84% dan 95.79%.

## KESIMPULAN DAN SARAN

Berdasarkan penelitian pada model *Decision Tree* untuk klasifikasi hasil SKD pada CPNS di ANRI tahun 2019 dengan data

sebanyak 344 peserta SKD dan 8 atribut didapatkan bahwa model *Decision Tree* berhasil mengklasifikasikan hasil SKD menjadi 4 kelas berdasarkan nilai yang didapatkan peserta seleksi. Nilai performa model *Decision Tree* yang didapatkan yaitu *Accuracy* = 92.23%, *Classification Error* = 7.77%, *Kappa* = 0.879, *Recall* = 94.84%, dan *Precision* = 95.79%. Hal ini membuktikan bahwa Algoritma *Decision Tree* cocok sebagai model untuk klasifikasi pada penelitian ini. Penelitian ini dapat dikembangkan dengan mengimplementasikan algoritma *data mining* yang lain dan melakukan perbandingan performa antara algoritma-algoritma yang digunakan. Dengan perbandingan performa tersebut dapat dilihat kinerja berbagai algoritma *data mining*.

## DAFTAR PUSTAKA

- [1] Peraturan Badan Kepegawaian Negara Nomor: 14 tahun 2018. Tentang

- Petunjuk Teknis Pengadaan Pegawai Negeri Sipil.
- [2] A. Unnisabegum, M. A. Hussain, dan M. Shaik, "Data mining techniques for big data," *International Journal of Advanced Research in Science Engineering and Technology*, vol. 6, Agustus, hal. 396-399, 2019.
- [3] C. Jalota dan R. Agrawal, "Analysis of educational data mining using classification," Dalam Prosiding 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, hal. 243-247.
- [4] Harwati dan S. Amby, "Application of decision tree approach to student selection model-a case study," IOP Conference Series: Material Science and Engineering 105 012104, 2016.
- [5] B. Siregar, dkk, "Tuition single classification using decision tree method and C4.5," IOP Conf. Series: Journal of Physics: Conf. Series 1175 012105, 2019.
- [6] A. Topirceanu dan G. Grosseck, "Decision tree learning used for the classification of student archtypes in online course," *Procedia Computer Science*, vol. 112, hal. 51 – 60, 2017.
- [7] M. V. Amazona dan A. A. Hernandez, "Modelling student performance using data mining techniques: input for academic program development," Dalam Prosiding 2019 5<sup>th</sup> International Conference on Computing and Data Engineering (ICCDE'19), 2019, hal. 36-4.