



Reduction of the Number of Demonstrative Word Determination Rules in Arabic Using the ID3 Algorithm

Reduksi Jumlah Aturan Penentuan Kata Tunjuk dalam Bahasa Arab Menggunakan Algoritma ID3

Inggih Permana^{1*}, Febi Nur Salisah²

^{1,2}Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Corresponden E-Mail: ¹inggihpermana@uin-suska.ac.id

Makalah: Diterima 03 September 2021; Diperbaiki 04 September 2021; Disetujui 30 September 2021

Abstract

Arabic has no less than ten forms of demonstrative words that can be used. To determine the designation must pay attention to four variables on the object that is pointed. From the four variables, 24 conditions (rules) may appear to determine the point of view. The number of rules that may appear is one of the reasons why Arabic students make mistakes using pointing words. Based on these problems, this research has reduced the rules by using the ID3 Algorithm. The decision tree resulting from the ID3 Algorithm has succeeded in reducing the rules by 50%. In addition, based on the decision tree, it can be concluded that to determine the demonstrative word, it is enough to pay attention to only three variables.

Keywords: Arabic, ID3, demonstrative word, decision tree, reduction

Abstrak

Bahasa Arab memiliki tidak kurang dari sepuluh bentuk kata tunjuk yang bisa digunakan. Untuk menentukan kata tunjuk tersebut harus memperhatikan empat variabel pada objek yang ditunjuk. Dari keempat variabel tersebut menghasilkan 24 kondisi (aturan) yang mungkin muncul untuk menentukan kata tunjuk. Banyaknya aturan yang mungkin muncul tersebut menjadi salah satu sebab pelajar Bahasa Arab melakukan kesalahan menggunakan kata tunjuk. Berdasarkan permasalahan tersebut, maka penelitian ini telah mereduksi aturan tersebut dengan menggunakan Algoritma ID3. Pohon keputusan hasil Algoritma ID3 berhasil mereduksi aturan sebanyak 50%. Selain itu, berdasarkan pohon keputusan tersebut dapat disimpulkan bahwa untuk menentukan kata tunjuk cukup memperhatikan tiga variabel saja.

Kata Kunci: Bahasa Arab, ID3, kata tunjuk, pohon keputusan, reduksi

1. Pendahuluan

Kata tunjuk (*isim isyaroh*) adalah kata-kata yang digunakan untuk menunjuk objek yang posisinya bisa dekat, sedang, maupun jauh dari pembicara [1]. Seperti yang terlihat pada buku-buku pelajaran bahasa arab [1, 2, 3], kata tunjuk merupakan salah satu materi yang selalu diajarkan. Dalam Bahasa Arab terdapat tidak kurang dari sepuluh bentuk kata tunjuk yang bisa digunakan [3, 4]. Untuk menentukan kata tunjuk yang dipakai, pembicara harus memperhatikan empat variabel dari objek yang ditunjuk, yaitu: jenis kelamin (laki-laki atau perempuan); jarak (jauh atau dekat); jumlah (tunggal, ganda, atau jamak); dan status akal (berakal atau tidak berakal) dari objek yang ditunjuk. Sehingga kondisi objek yang mungkin muncul adalah 24 kondisi. Contoh kondisi yang dimaksud pada konteks ini sebagai berikut: anggap ada sebuah objek yang ditunjuk dengan nilai jenis kelamin adalah laki-laki; nilai jarak adalah dekat; nilai jumlah adalah jamak; dan nilai status akal adalah berakal.

Pada penjelasan paragraf sebelumnya, dapat dilihat bahwa ada banyak kondisi yang mungkin terjadi untuk menentukan kata tunjuk. Sehingga hal ini menjadi salah satu penyebab pelajar Bahasa Arab banyak melakukan kesalahan dalam menggunakan kata tunjuk. Menurut penelitian terdahulu [5, 6], kesalahan penggunaan kata tunjuk merupakan salah satu kesalahan dalam struktur Bahasa Arab yang mungkin dilakukan oleh peserta didik. Berdasarkan hasil penelitian Rusdi dkk. tahun 2020 [7] tentang kesalahan penggunaan kata tunjuk yang

dilakukan oleh siswa kelas X di Sekolah Menengah Kejuruan (SMK), kesalahan penggunaan kata haadzih mencapai 17,03%, kesalahan penggunaan kata haadzaa mencapai 31,38%, kesalahan penggunaan kata dzaalika mencapai 20,46%, dan kesalahan penggunaan kata tilka mencapai 30,42%.

Berdasarkan permasalahan tersebut, maka penelitian ini bertujuan untuk mereduksi jumlah kondisi yang mungkin terjadi pada penentuan kata tunjuk dalam Bahasa Arab dengan menggunakan teknik *machine learning*. Teknik ini akan mempelajari data yang ada sehingga menghasilkan pengetahuan [8] baru yang diinginkan. Pada konteks penelitian ini pengetahuan baru yang diinginkan adalah jumlah kemungkinan kondisi untuk menentukan kata tunjuk yang lebih sedikit dari biasanya. Sehingga hal ini dapat mempermudah pelajar menguasai kata tunjuk yang digunakan serta dapat digunakan sebagai bahan ajar oleh pengajar Bahasa Arab.

Penelitian ini menggunakan pohon keputusan (*decision tree*) untuk menentukan kata tunjuk yang digunakan. Pohon keputusan merupakan salah satu metode *machine learning* [9] yang sering digunakan. Jenis *machine learning* yang ini akan menghasilkan sebuah pohon keputusan untuk menyelesaikan kasus yang diberikan. Pohon keputusan tersebut dapat diterjemahkan dalam aturan (*rule*) *IF-THEN* (JIKA-MAKA). Contohnya: *IF* jenis kelamin: laki-laki *AND* jarak: dekat *AND* jumlah: jamak *AND* status akal: tidak berakal *THEN* kata tunjuk: haadzih. Pada kasus ini, kondisi yang mungkin muncul untuk menentukan kata tunjuk dapat dianalogikan sebagai sebuah aturan. Karena hasil pengetahuan berupa pohon keputusan, maka pengetahuan tersebut bisa digunakan dengan mudah untuk orang non komputer sains seperti pelajar dan pengajar Bahasa Arab.

Algoritma pembuat pohon keputusan yang digunakan adalah Algoritma *Iterative Dichotomiser 3* (ID3). Algoritma ini tawarkan awalnya oleh Quinlan tahun 1986 [10]. ID3 dipilih karena variabel yang digunakan untuk menentukan kata tunjuk (jenis kelamin, jarak, jumlah, status akal) memiliki nilai nominal. ID3 memang didesain untuk data yang memiliki variabel bertipe nominal [11]. Algoritma ini juga telah berhasil diterapkan di berbagai kasus pendidikan, seperti prediksi pencapaian siswa pada suatu materi pelajaran di *playgroup* [12], lama masa studi mahasiswa [13], dan untuk penerimaan beasiswa [14].

Paper ini disusun dalam lima bab. Bab 2 membahas teori-teori terkait kata tunjuk dalam Bahasa Arab dan Algoritma ID3. Bab 3 membahas tentang metode penelitian yang digunakan. Bab 4 berisi hasil dan pembahasan dari penelitian yang dilakukan. Bab terakhir membahas tentang kesimpulan dari penelitian ini.

2. Teori-Teori Terkait

Pada bab ini akan dijelaskan tentang Algoritma ID3 dan kata tunjuk dalam Bahasa Arab.

2.1. Algoritma ID3

ID3 menggunakan *information gain* sebagai kriteria pembuatan cabang pohon keputusan, algoritma ini berhenti membuat cabang pohon keputusan ketika semua *information gain* sama dengan nol. Untuk mendapatkan nilai *information gain* dapat digunakan Persamaan 1, Persamaan 2, dan Persamaan 3 [15]. Persamaan 1 dan Persamaan 2 disebut juga dengan istilah *entropy*.

$$Info(D) = \sum_{i=1}^m p_i \log_2(p_i) \quad (\text{Persamaan 1})$$

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (\text{Persamaan 2})$$

$$Gain(A) = Info(D) - Info_A(D) \quad (\text{Persamaan 3})$$

Keterangan:

1. Info(D) = Nilai *entropy* seluruh data
2. Info_A(D) = Nilai *entropy* variabel A
3. Gain(A) = *Information gain* variabel A
4. |D_j| = jumlah kemunculan jenis nilai tertentu dari variabel A pada seluruh data
5. |D| = jumlah seluruh data
6. p_i = peluang *class* i
7. m = jumlah *class*
8. v = jumlah jenis nilai pada variabel A

2.2. Kata Tunjuk

Sebagaimana dijelaskan pada awal Bab 1, terdapat setidaknya sepuluh kata tunjuk yang bisa digunakan untuk menunjuk objek dalam Bahasa Arab. Kesepuluh kata tunjuk tersebut dapat dilihat pada Tabel 1.

Tabel 1. Kata tunjuk dalam Bahasa Arab (disarikan dari [1, 2, 3])

| No. | Kata Tunjuk |
|-----|-------------|
| 1 | Haadzaa |
| 2 | Haadzihi |
| 3 | Haadzani |
| 4 | Haatani |
| 5 | Haaulai |
| 6 | Dzaalika |
| 7 | Tilka |
| 8 | Dzaanika |
| 9 | Taanika |
| 10 | Ulaaika |

Untuk menentukan kata tunjuk dalam Bahasa Arab harus memperhatikan empat variabel pada objek yang ditunjuk. Keempat variabel tersebut beserta nilainya dapat dilihat pada Tabel 2.

Tabel 2. Variabel yang harus diperhatikan untuk menentukan kata tunjuk (disarikan dari [1, 2, 3])

| No. | Nama Variabel | Nilai |
|-----|---------------|---------------------------------|
| 1 | Jenis kelamin | - Laki-laki - Perempuan |
| 2 | Jarak | - Jauh - Dekat |
| 3 | Jumlah | - Tunggal - Ganda - Jamak |
| 4 | Status akal | - Berakal - Tidak berakal |

Dari variabel dan kata tunjuk yang ada didapat 24 kondisi (aturan) yang mungkin untuk menentukan kata tunjuk, lebih detailnya dapat dilihat pada Tabel 3.

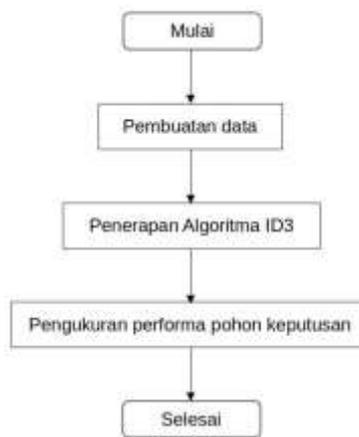
Tabel 3. Aturan penggunaan kata tunjuk dalam Bahasa Arab (disarikan dari [1, 2, 3])

| | | Jarak | Jauh | | Dekat | |
|---------------|--------|-------------|---------|---------------|---------|---------------|
| | | Status Akal | Berakal | Tidak Berakal | Berakal | Tidak Berakal |
| Jenis Kelamin | Jumlah | | | | | |

| | | | | | | |
|-----------|---------|--|----------|----------|----------|----------|
| Laki-laki | Tunggal | | Dzaalika | Dzaalika | Haadza | Haadza |
| | Ganda | | Dzaanika | Dzaanika | Haadzani | Haadzani |
| | Jamak | | Ulaaika | Tilka | Haaulai | Haadzihi |
| Perempuan | Tunggal | | Tilka | Tilka | Haadzihi | Haadzihi |
| | Ganda | | Taanika | Taanika | Haatani | Haatani |
| | Jamak | | Ulaaika | Tilka | Haaulai | Haadzihi |

3. Metode Penelitian

Secara garis besar langkah-langkah penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Langkah-langkah penelitian

Data dibuat dengan menuliskan semua kombinasi yang mungkin terjadi berdasarkan nilai-nilai dari variabel-variabel yang ada. Jumlah data yang dihasilkan adalah 24 baris. *Class* pada data ini adalah jenis kata tunjuk yang digunakan. Data tersebut dapat dilihat pada Tabel 4.

Tabel 4. Data untuk proses Algoritma ID3

| Jarak | Jenis Kelamin | Jumlah | Status Akal | Kata Tunjuk |
|-------|---------------|---------|---------------|-------------|
| Jauh | Laki-laki | Tunggal | Berakal | Dzaalika |
| Jauh | Laki-laki | Tunggal | Tidak Berakal | Dzaalika |
| Jauh | Laki-laki | Ganda | Berakal | Dzaanika |
| Jauh | Laki-laki | Ganda | Tidak Berakal | Dzaanika |
| Jauh | Laki-laki | Jamak | Berakal | Ulaaika |
| Jauh | Laki-laki | Jamak | Tidak Berakal | Tilka |
| Jauh | Perempuan | Tunggal | Berakal | Tilka |
| Jauh | Perempuan | Tunggal | Tidak Berakal | Tilka |
| Jauh | Perempuan | Ganda | Berakal | Taanika |
| Jauh | Perempuan | Ganda | Tidak Berakal | Taanika |
| Jauh | Perempuan | Jamak | Berakal | Ulaaika |
| Jauh | Perempuan | Jamak | Tidak Berakal | Tilka |
| Dekat | Laki-laki | Tunggal | Berakal | Haadzaa |

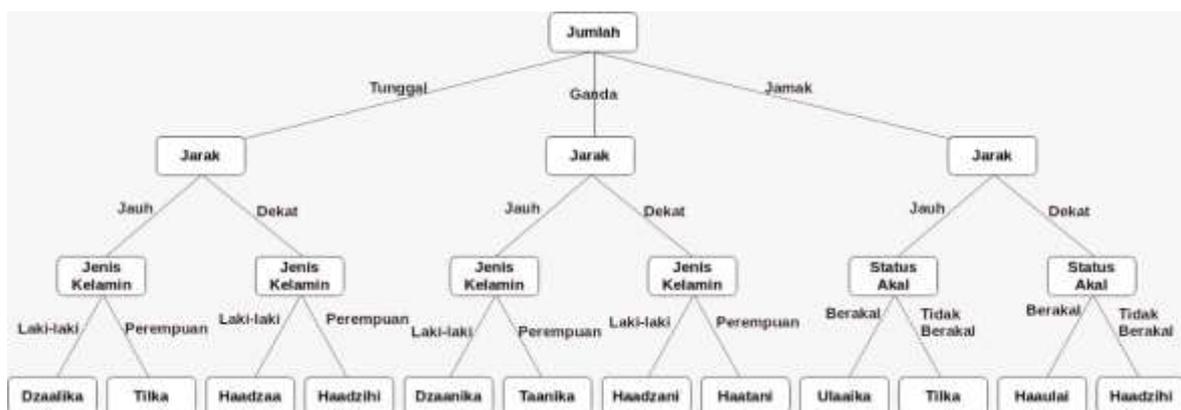
| | | | | |
|-------|-----------|---------|---------------|----------|
| Dekat | Laki-laki | Tunggal | Tidak Berakal | Haadzaa |
| Dekat | Laki-laki | Ganda | Berakal | Haadzani |
| Dekat | Laki-laki | Ganda | Tidak Berakal | Haadzani |
| Dekat | Laki-laki | Jamak | Berakal | Haaula'i |
| Dekat | Laki-laki | Jamak | Tidak Berakal | Haadzihi |
| Dekat | Perempuan | Tunggal | Berakal | Haadzihi |
| Dekat | Perempuan | Tunggal | Tidak Berakal | Haadzihi |
| Dekat | Perempuan | Ganda | Berakal | Haatani |
| Dekat | Perempuan | Ganda | Tidak Berakal | Haatani |
| Dekat | Perempuan | Jamak | Berakal | Haaula'i |
| Dekat | Perempuan | Jamak | Tidak Berakal | Haadzihi |

Penerapan Algoritma ID3 menggunakan *Software Waikato Environment for Knowledge Analysis (WEKA)* versi 3.8.5. Algoritma ID3 tidak ada pada *package* bawaan Weka, maka pada penelitian telah di-*install package* tambahan agar bisa menggunakan Algoritma ID3, yaitu *simpleEducationalLearningSchemes*. Data yang digunakan untuk membuat pohon keputusan (data latih) adalah semua data.

Peforma pohon keputusan yang dihasilkan diukur menggunakan akurasi. Data yang digunakan untuk mengukur akurasi (data uji) adalah semua data. Akurasi dari pohon keputusan yang dihasilkan sudah tersedia pada *Software Weka*.

4. Hasil dan Pembahasan

Pohon keputusan hasil Algoritma ID3 dapat dilihat pada Gambar 2. Dari pohon keputusan diatas terlihat, untuk menentukan kata tunjuk yang digunakan hanya perlu memperhatikan tiga variabel dari empat variabel yang ada. Variabel pertama yang perlu dicek adalah jumlah. Variabel kedua yang dicek adalah jarak. Sedangkan variabel ketiga yang dicek tergantung nilai dari variabel jumlah. Jika nilai variabel jumlah tunggal atau ganda, maka variabel ketiga yang dicek adalah jenis kelamin, sedangkan jika nilai variabel jumlah adalah jamak, maka variabel ketiga yang dicek adalah status akal.



Gambar 2. Pohon keputusan hasil Algoritma ID3

Tabel 5. Bentuk aturan *IF-THEN* dari pohon keputusan Algoritma ID3

| No. | Aturan <i>IF-THEN</i> |
|-----|--|
| 1 | <i>IF</i> jumlah = tunggal <i>AND</i> jarak = jauh <i>AND</i> jenis kelamin = laki-laki <i>THEN</i> kata_tunjuk = dzaalika |
| 2 | <i>IF</i> jumlah = tunggal <i>AND</i> jarak = jauh <i>AND</i> jenis kelamin = perempuan <i>THEN</i> kata_tunjuk = tilka |
| 3 | <i>IF</i> jumlah = tunggal <i>AND</i> jarak = dekat <i>AND</i> jenis kelamin = laki-laki <i>THEN</i> kata_tunjuk = haadzaa |
| 4 | <i>IF</i> jumlah = tunggal <i>AND</i> jarak = dekat <i>AND</i> jenis kelamin = perempuan <i>THEN</i> |

| | |
|----|--|
| | kata_tunjuk = haadzihi |
| 5 | <i>IF</i> jumlah = ganda <i>AND</i> jarak = jauh <i>AND</i> jenis kelamin = laki-laki <i>THEN</i> kata_tunjuk = dzaanika |
| 6 | <i>IF</i> jumlah = ganda <i>AND</i> jarak = jauh <i>AND</i> jenis kelamin = perempuan <i>THEN</i> kata_tunjuk = taanika |
| 7 | <i>IF</i> jumlah = ganda <i>AND</i> jarak = dekat <i>AND</i> jenis kelamin = laki-laki <i>THEN</i> kata_tunjuk = haadzani |
| 8 | <i>IF</i> jumlah = ganda <i>AND</i> jarak = dekat <i>AND</i> jenis kelamin = perempuan <i>THEN</i> kata_tunjuk = haatani |
| 9 | <i>IF</i> jumlah = jamak <i>AND</i> jarak = jauh <i>AND</i> status akal = berakal <i>THEN</i> kata_tunjuk = ulaaika |
| 10 | <i>IF</i> jumlah = jamak <i>AND</i> jarak = jauh <i>AND</i> status akal = tidak berakal <i>THEN</i> kata_tunjuk = tilka |
| 11 | <i>IF</i> jumlah = jamak <i>AND</i> jarak = dekat <i>AND</i> status akal = berakal = ya <i>THEN</i> kata_tunjuk = hauulai |
| 12 | <i>IF</i> jumlah = jamak <i>AND</i> jarak = dekat <i>AND</i> status akal = tidak berakal <i>THEN</i> kata_tunjuk = hadzihi |

Akurasi pohon keputusan yang dihasilkan dari Algoritma ID3 adalah 100%. Hal ini artinya pohon keputusan yang dihasilkan berhasil menentukan seluruh kata tunjuk dengan benar untuk semua kondisi. Dari pohon keputusan yang dihasilkan didapat 12 aturan sementara sebelumnya ada 24 aturan untuk menentukan kata tunjuk. Ini berarti jumlah aturan untuk menentukan kata tunjuk berhasil direduksi sebanyak 50%. Aturan-aturan yang tereduksi tersebut dalam bentuk aturan *IF-THEN* dapat dilihat pada Tabel 5. Selain itu, aturan tersebut juga bisa dibuat dalam bentuk *NESTED IF* seperti Algoritma di bawah ini.

Algoritma Penentuan Kata Tunjuk

Input: jumlah, jarak, jenis kelamin dan status akal dari objek yang ditunjuk

Output: jenis kata tunjuk yang digunakan

```

IF jumlah = tunggal THEN
  IF jarak = jauh THEN
    IF jenis kelamin = laki-laki THEN
      kata tunjuk = dzaalika
    ELSE
      kata tunjuk = tilka
    ENDIF
  ELSE
    IF jenis kelamin = laki-laki THEN
      kata tunjuk = haadzaa
    ELSE
      kata tunjuk = haadzihi
    ENDIF
  ENDIF
ELSEIF jumlah = ganda THEN
  IF jarak = jauh THEN
    IF jenis kelamin = laki-laki THEN
      kata tunjuk = dzaanika
    ELSE
      kata tunjuk = taanika
    ENDIF
  ELSE
    IF jenis kelamin = laki-laki THEN
      kata tunjuk = haadzani
    ELSE
      kata tunjuk = haatani
    ENDIF
  ENDIF
ELSE
  IF jarak = jauh THEN
    IF status akal = berakal THEN
      kata tunjuk = ulaaika
    ELSE
      kata tunjuk = tilka
    ENDIF
  ELSE
    IF status akal = tidak berakal THEN

```

```
        kata tunjuk = haaulai
    ELSE
        kata tunjuk = haadzihi
    ENDIF
ENDIF
ENDIF
```

5. Kesimpulan

Berdasarkan penerapan ID3 yang telah dilakukan, dapat disimpulkan bahwa ID3 berhasil mereduksi jumlah aturan untuk penentuan jenis kata tunjuk yang digunakan yang awalnya 24 aturan menjadi 12 aturan atau sebanyak 50%. Untuk menentukan kata tunjuk yang digunakan hanya perlu memperhatikan tiga variabel dari empat variabel yang ada. Variabel pertama dan kedua yang perlu dicek berturut-turut adalah jumlah dan jarak. Variabel ketiga yang dicek tergantung nilai dari variabel jumlah. Jika nilai variabel jumlah tunggal atau ganda, maka variabel ketiga yang dicek adalah jenis kelamin, sedangkan jika nilai variabel jumlah adalah jamak, maka variabel ketiga yang dicek adalah status akal.

Ucapan Terima Kasih

Terima kasih penulis ucapkan kepada Ustadz Muhammad Karim, Lc. yang telah bersedia menjadi pakar dalam penelitian ini. Semoga selalu dalam rahmat Allah Subhanahu Wa Ta'ala.

Referensi

- [1] Huda, N. "Mudah Belajar Bahasa Arab". Jakarta: Amzah. 2012.
- [2] Abdur-Raheem, V. "*Durusul Lughah Al-'Arabiyyah Li Ghairin Nathiqina Biha*". Madinah. 1998
- [3] Rahmawati, S. "*Al-'Arabiyyatul Muyassarotu Lifahmits Tsaqofatil Muta'addidati*". Diktat Bahasa Arab, UIN KH. Achmad Siddiq, Jember. nd.
- [4] Ilyas, H. "*Al-Nakirah wa Al-Ma'rifah*". Shaut Al Arabiyyah, 3(2), 7-15. 2015.
- [5] Karomah, S. "*Arabic Language Education Program at Islamic Boarding School Ibnul Qoyyim Putri Yogyakarta: Study of Code Mixing, Code Switching, and Interference*". Jurnal Pendidikan Islam, 10(2), 181-218. 2021.
- [6] Khairunnas, K. "Kesalahan Nahwu dan Saraf dalam Insha'Tahriri Mahasiswa Jurusan Pendidikan Bahasa Arab Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar (Tesis, Universitas Islam Negeri Alauddin Makassar)". 2017.
- [7] Rusdi, R., Muthmainnah, S., dan Asri, W.K. "Analisis Kesalahan Penggunaan Isim Isyarah dalam Menulis Kalimat Sederhana Bahasa Arab Siswa Kelas X SMK Muhammadiyah Bungoro Kabupaten Pangkep (Universitas Negeri Makassar)". 2020.
- [8] Shalev-Shwartz, S., dan Ben-David, S. "*Understanding Machine Learning: From Theory to Algorithms*". Cambridge University Press. 2014.
- [9] Jordan, M.I., dan Mitchell, T.M. "*Machine Learning: Trends, Perspectives, and Prospects*". Science, 349(6245), 255-260. 2015.
- [10] Quinlan, J. R. "*Induction of Decision Trees. Machine learning*", 1(1), 81-106. 1986.
- [11] Mak, B., dan Munakata, T. "*Rule Extraction from Expert Heuristics: A Comparative Study of Rough Sets with Neural Networks and ID3*". European journal of operational research, 136(1), 212-229. 2002.
- [12] Agus, N., Lusi, M., dan Deasy, P. "*Implementation ID3 Algorithm to Predict Children Achievement in Response (Case Study Children Playgroup School)*". Journal of Engineering and Applied Sciences, 12(2), 204-207. 2017.
- [13] Putra, N.P., dan Maulany, G.J. "*Classification System for Student Study Duration on Department of Information Systems at Musamus University, Using ID3*". International Journal of Mechanical Engineering and Technology (IJMET), 9(12), 878-885. 2018.
- [14] Rohman, A., Arif, I., Harianti, I., Hidayat, A. S., Teniwut, W. A., Baharun, H., ... dan Suradi, A. "*ID3 algorithm approach for giving scholarships*". In Journal of Physics: Conference Series (Vol. 1175, No. 1, p. 012116). IOP Publishing. 2019, March.
- [15] Han, J., Pei, J., & Kamber, M. "*Data Mining: Concepts and Techniques*". Elsevier. 2011.