Pengujian Akurasi Akor Pada Suara Musik Piano Dengan Metode Convolutional Neural Network (CNN)

Muhammad Kamal Yasya1, Onny Marleen2

1, 2Business Information System, Gunadarma University, Depok, West Java, Indonesia-16424

Email Address: 1kamalyasya[@gmail.com](mailto:nadiailhaq@gmail.com), 2onny\_marleen@staff.gunadarma.ac.id

***Abstract*—***Implementasi kecerdasan buatan dalam dunia permusikan sudah tergolong banyak, namun banyak dari implementasi tersebut yang masih kurang akurat dalam hal akurasi seperti Chord Recognition. Chord Recognition adalah sebuah sistem atau aplikasi untuk mengetahui atau mengukur akor secara tepat. Untuk membangun sistem tersebut, dibutuhkan metode-metode agar akurasi yang dihasilkan tinggi. Beberapa metode yang sering digunakan adalah metode Convolutional Neural Network (CNN)*.

*Penelitian ini bertujuan untuk mengukur akurasi pengenalan akor dari audio piano dan bass dengan menggunakan metode Convolutional Neural Network (CNN) dan memanfaatkan chromagram dalam transformasi audio ke bentuk gambar. Data yang digunakan dalam penelitian ini terdiri dari 12 kelas major dan 12 kelas minor. Berdasarkan hasil pengujian yang dilakukan, akurasi pada masing-masing kelas memberikan nilai F1 yang cukup baik. Rata-rata nilai F1 dari keseluruhan pengujian pengenalan akor berdasarkan nada root adalah sebesar 87%.*

***Keywords—*** *Convolutional Neural Network (CNN), Chromagram, Chord Recognition.*

# Introduction

Implementasi kecerdasan buatan dalam dunia permusikan sudah tergolong banyak, namun masih terdapat banyak kekurangan dalam hal akurasi seperti dalam hal *Chord Recognition*. *Chord Recognition* adalah sebuah sistem atau aplikasi untuk mengetahui atau mengukur akor secara tepat. Untuk membangun sistem tersebut, dibutuhkan metode-metode agar akurasi yang dihasilkan tinggi. Beberapa metode yang sering digunakan adalah metode *Convolutional Neural Network (CNN)*.

*Convolutional Neural Network (CNN)* adalah salah satu metode *Machine Learning* dari pengembangan *Multi Layer Perceptron (MLP)* yang didesain untuk mengolah data dua dimensi. *CNN* dikatakan pengembangan lebih lanjut dari *MLP* karena *CNN* menggunakan metode yang mirip dengan *MLP*, namun menggunakan dimensi yang lebih banyak. *CNN* juga sering digunakan sebagai ekstraktor fitur yang kuat dari data yang telah disatukan, seperti gambar. Metode ini dapat diperluas ke berbagai tugas klasifikasi sinyal audio dengan merepresentasikan sinyal *input* dalam domain frekuensi waktu.

Penelitian ini akan mengukur akurasi pengenalan akor audio dari alat musik *piano* dan *bass*. Metode yang digunakan adalah metode *Convolutional Neural Network (CNN)* dan untuk mengubah bentuk audio menjadi sebuah gambar akan digunakan *chromagram*. Akor yang akan digunakan dalam penelitian ini terdiri dari 12 kelas *major* dan 12 kelas *minor*.

# THEORY REVIEW AND REFERENCES

1. Audio

Audio adalah sebuah deret waktu, dimana sumbu y adalah amplitudo arus yang sesuai dengan sebuah membran *loudspeaker* dan sumbu x adalah sumbu yang sesuai dengan satuan waktu data tersebut (Ingo Miersawa & Katharina Morik, 2005). Telinga manusia hanya dapat mendengar bunyi dengan rentang frekuensi antara 20 Hz hingga 20 KHz (20.000Hz). Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 KHz merupakan frekuensi tertinggi yang dapat didengar.[[1]](#footnote-0) Gelombang suara mengandung sejumlah komponen penting, seperti amplitudo, panjang gelombang, dan frekuensi. Komponen-komponen tersebut mampu membuat suara yang satu berbeda dengan suara yang lain.

1. Akor

Akor merupakan rangkaian nada-nada dasar yang tersusun secara teratur dari sebuah tangga nada dan bisa merepresentasi tangga nada tersebut (Tito Galit Permana, Dr. Ir. Bambang Hidayat, dan Eko Susatio S.T,M.T., 2014). Akor tiga not yang berjarak tiga *scale* pada *root* sampai not ketiga disebut *triad*. Untuk mempermudah pemusik dalam menentukan apakah suatu kombinasi dari tiga nada tertentu merupakan sebuah *triad* atau bukan, maka pemusik dapat membentuk diagram *circle of thirds*. *Scale* atau skala nadamerupakan jarak yang menandakan lokasi suatu nada dengan titik mulai tertentu. Terdapat empat jenis *triad* yang dibedakan berdasarkan skalanadanya, diantaranya :

### Major Triad : Major triad merupakan tiga nada yang dibunyikan dengan skala nada 1-2-1,5. Akor major menunjukan identitas suatu bagian lagu lebih dekat dengan root karena ada satu buah akor major yang menempati Scale I.

### Minor Triad : Minor triad merupakan tiga nada yang dibunyikan dengan skala 1-1,5-2.

### Diminished Triad : Diminished triad merupakan tiga nada yang dibunyikan dengan skala 1-1,5-1,5.

### Augmented Triad : Augmented triad merupakan tiga nada yang dibunyikan dengan skala 1-1,5-2,5.

1. *Beat Tracking*

*Beat tracking* digunakan untuk menentukan contoh waktu dalam rekaman *audio*, di mana pendengar manusia cenderung mengetuk kakinya ke musik (Tavish Srivastava, 2019). *Beat tracking* pada rancangan menggunakan metode *dynamic programming. Dynamic programming* atau *dynamic optimization* bukan merupakan rumus matematika yang mampu memberikan jawaban dengan hanya sekedar memberikan input. Sebaliknya, *dynamic programming* adalah kombinasi pemikiran terstruktur dan pola pikir analitis untuk menyelesaikan sebuah permasalahan. Untuk melakukan *beat tracking*, hal yang harus dilakukan terlebih dahulu adalah mencari sinyal yang berupa *onset*. *Onset* merupakan lokasi dimana sinyal berada pada nilai yang tinggi secara tiba-tiba.

1. *Chromagram*

*Chromagram* merupakan pencapaian nada dalam durasi tertentu (Eka Angga Laksana & Feri Sulianta, 2017). Chromagram adalah transformasi properti frekuensi-waktu sinyal menjadi prekursor pitch yang berubah-ubah untuk sementara waktu. Transformasi ini didasarkan pada pengamatan persepsi tentang sistem pendengaran dan telah terbukti memiliki beberapa sifat matematika yang menarik. Chromagram memperluas konsep chroma untuk memasukkan dimensi waktu. Seperti halnya kita menggunakan spektrogram untuk menyimpulkan properti tentang distribusi energi sinyal dari frekuensi dan waktu, chromagram dapat digunakan untuk menyimpulkan properti tentang distribusi energi sinyal terhadap kroma dan waktu.

1. *Windowing*

*Windowing* berfungsi untuk meminimalisir sinyal yang tak kontinu pada awal dan akhir masing-masing frame (Marwa A.Nasr, 2018). Nilai panjang filter didapatkan dengan menggunakan rumus di bawah. Nilai panjang tersebut akan dipakai untuk membuat respon impulse di simulasi dan implementasi. Nilai panjang filter juga akan berpengaruh kepada faktor roll-off atau faktor kelandaian dari suatu filter.

1. *Convolutional Neural Network (CNN)*

*Convolutional Neural Network (CNN)* adalah salah satu jenis neural network yang biasa digunakan pada data image (Samuel Sena, 2019). *CNN* bisa digunakan untuk mendeteksi dan mengenali objek pada sebuah image. *CNN* terdiri dari neuron yang memiliki weight, bias, dan fungsi aktivasi. Cara kerja *CNN* memiliki kesamaan pada *MLP*, namun dalam *CNN* setiap neuron dipresentasikan dalam bentuk dua dimensi, tidak seperti *MLP* yang setiap neuron hanya berukuran satu dimensi. Pada CNN, data yang dipropagasikan oleh jaringan adalah data dua dimensi, sehingga operasi linear dan parameter bobot pada CNN berbeda. Pada CNN operasi linear menggunakan operasi konvolusi, sedangkan bobot tidak lagi satu dimensi saja, namun berbentuk empat dimensi yang merupakan kumpulan kernel konvolusi. Karena sifat proses konvolusi, maka CNN hanya dapat digunakan pada data yang memiliki struktur dua dimensi seperti citra dan suara.

# RESEARCH METHOD

Kerangka pemikiran dalam penelitian ini memiliki hasil akhir yakni mengetahui akurasi dari metode *CNN* dalam mengenali sebuah akor dengan memanfaatkan *Chromagram*. Langkah awal yang akan dilakukan adalah menerima input file audio yang memiliki format .wav. Selanjutnya akan dilakukan beat tracking. Pada proses beat tracking, file audio yang telah diunggah akan dipotong tiap detik dan akan dianalisa bentuk gelombangnya sehingga terbentuklah sebuah onset. Setelah mendapatkan onset dari proses beat tracking, selanjutnya onset tersebut akan diubah ke bentuk chromagram. Pada proses chromagram, onset yang telah didapatkan akan dibentuk ke bentuk gambar. Gambar tersebut nantinya akan dilatih dengan menggunakan metode *CNN* untuk menemukan akor yang sesuai. Setelah menemukan akor yang sesuai, selanjutnya melakukan proses pengaturan direktori. Proses ini dibutuhkan agar penempatan *parenthesis* dan *child* package bagian backend dan frontend dapat berkomunikasi satu sama lain. Dan terakhir adalah ada proses pendekatan backend. Pendekatan backend dibutuhkan agar memudahkan dalam mengakses fungsi yang ada pada pada *script Python*.

# DISCUSSION

Hasil dari kerangka penelitian adalah mengukur akurasi data akor yang dihasilkan pada Chord Recognition dengan menggunakan metode CNN, kemudian menarik kesimpulan dari hasil yang telah didapatkan. Pada bab ini disajikan beberapa hal yang berkaitan dengan proses, hasil, dan pembahasan hasil penelitian, diantaranya, hasil pengujian prediksi model menggunakan Confusion Matrix, dan hasil pengujian akurasi akor beserta pembahasannya.

1. Pengujian Prediksi Model Menggunakan *Confusion Matrix*

*Confusion matrix* berbentuk tabel yang sering digunakan untuk memberikan keterangan performa dari model klasifikasi terhadap sekumpulan data tes.[[2]](#footnote-1) Pada aplikasi sistem terdapat 1192 data latih dan 247 data validasi dengan kelas sebanyak 24. Terdapat tiga variabel untuk melakukan perhitungan tingkat akurasi terhadap klasifikasi yang dilakukan oleh model. Ketiga variabel itu adalah *precision*, *recall*, dan *F1-score*. Untuk mendapatkan nilai dari ketiga variabel tersebut dibutuhkan *confusion matrix* yang memiliki nilai-nilai sebagai berikut:[[3]](#footnote-2)

1. *True Negative* (TN) : TN merupakan jumlah prediksi benar untuk data salah.
2. *True Positive* (TP) : TP merupakan jumlah prediksi benar untuk data benar.
3. *False Negative* (FN) : FN merupakan jumlah prediksi salah untuk data salah.
4. *False Positive* (FP) : FP merupakan jumlah prediksi salah untuk data benar.
5. Hasil Pengujian Model

Hasil *confusion matrix* pada pengujian dapat dilihat pada Tabel 1. Berdasarkan nilai dari *confusion matrix* tersebut, maka nilai *precision*, *recall*, dan *F1-score* dapat dilihat pada Tabel 2. Hasil dari pengujian, nilai akurasi yang didapatkan sebesar 89% terhadap 72 data yang diuji. Pada Tabel 2 dapat dilihat bahwa prediksi perbedaan akor dapat mudah dilakukan, sementara itu prediksi jenis akor lebih sukar dibedakan. Sebagai contoh untuk kelas akor A#m terdapat kesalahan prediksi menjadi kelas akor A#. Terdapat juga kesalahan prediksi pada kelas akor G#. Model memprediksi kelas akor G# menjadi kelas akor G dan kelas akor G#m. Hal ini dikarenakan akor *major* dan *minor* hanya memiliki perbedaan nada kedua sebesar setengah langkah. Akor A# (A# major) terdiri dari nada A#, D, dan F, sedangkan akor A#m (A# minor) terdiri dari nada A#, C#, dan F. Perbedaan nada yang tidak terlalu signifikan mengakibatkan sistem sukar membedakan akor *minor* dan *major*.

Sebagai salah satu hasil pengujian akor, untuk akor C *Major* terdapat 30 visualisasi yang didapat dari data lagu masukan akor C dengan nilai F1 yang didapat sebesar 95 persen. Terdapat beberapa kelas akor lain yang diprediksi oleh model yaitu, kelas akor C minor dan kelas akor F major. Kesalahan ini dikarenakan akor C major dan akor C minor hanya memiliki satu buah nada yang berbeda. Akor C major terdiri dari nada C, E, dan G. Sementara akor C minor terdiri dari nada C, D#, dan G. Frekuensi nada E adalah 329.63 Hz sedangkan frekuensi nada D# adalah 311.13 Hz. Kedua frekuensi tersebut saling berdekatan sehingga model salah mengenali akor C major sebagai akor C minor sebanyak dua kali. Detil hasil pengujian pada kelas akor C major dapat dilihat pada Tabel 3.

Table 1. Hasil *Confution Matrix*

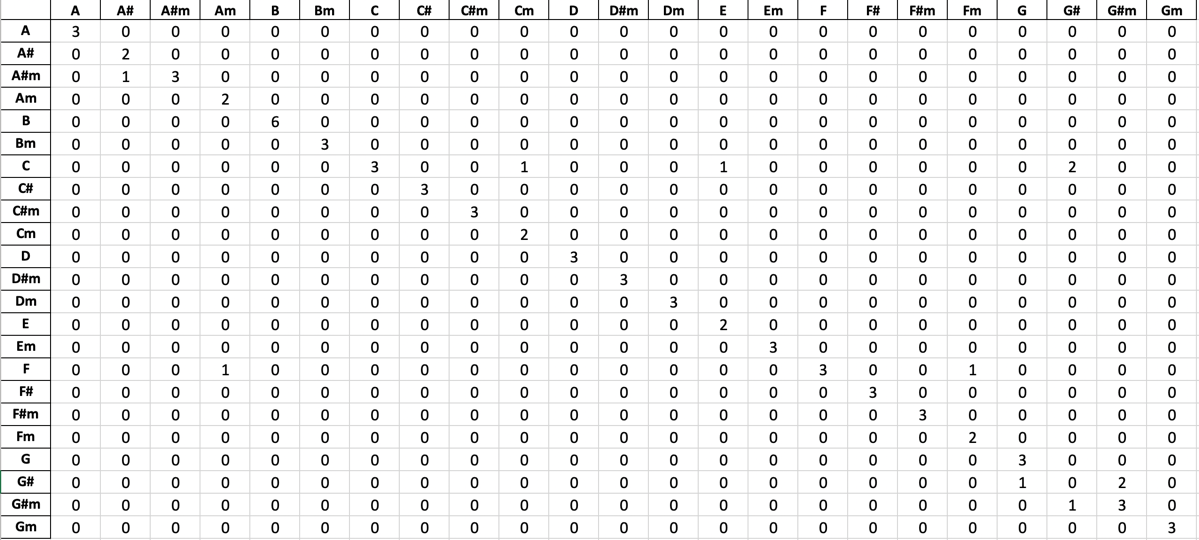


Table 2. Perhitungan Precision, Recall, F1-score, dan Support

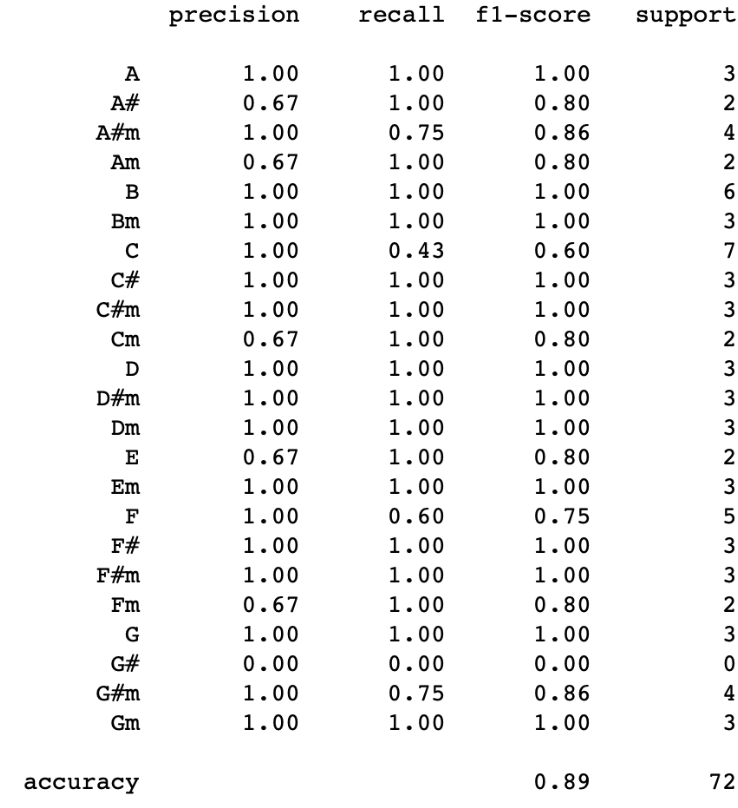


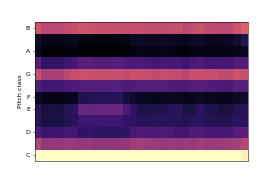
Table 3. Hasil Pengujian pada Akor C *Major*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-Score** | **Support** |
| C | 0,90 | 1,00 | 0,95 | 27 |
| Cm | 0,00 | 0,00 | 0,00 | 2 |
| F | 0,00 | 0,00 | 0,00 | 1 |
| Macro Avg | 0,30 | 0,33 | 0,32 | - |
| Weighted Avg | 0,82 | 0,90 | 0,85 | - |

Untuk data latih akor C major dapat dilihat pada Gambar 1 dan untuk data validasi C major dapat dilihat pada Gambar 2. Dapat dilihat pada Gambar 1 terdapat baris yang terang, yakni baris akor G dan akor C. Setelah melalui proses *CNN* dimana berperan sebagai validator, ditemukan bahwa akor dari data latih tersebut adalah akor C (Gambar 2).



Gambar 1. Data Latih Akor C *Major*



Gambar 2. Data Validasi Akor C *Major*

1. Hasil Rinci Pengujian

Dari pengujian akurasi yang telah dilakukan, dapat disimpulkan bahwa pengujian akurasi pada masing-masing kelas memberikan nilai F1 yang cukup baik. Besar nilai F1 pada masing-masing kelas dapat dilihat pada Tabel 4.

Tabel 4. Rincian Nilai F1 Keseluruhan Pengujian Pengenalan Akor Berdasarkan Root

|  |  |  |
| --- | --- | --- |
| Nomor | Kelas Akor | Nilai F1 |
| 1 | C | 0,95 |
| 2 | Cm | 0,77 |
| 3 | C# | 0,91 |
| 4 | C#m | 0,80 |
| 5 | D | 0,83 |
| 6 | Dm | 0,75 |
| 7 | D# | 0,81 |
| 8 | D#m | 0,82 |
| 9 | E | 0,89 |
| 10 | Em | 0,84 |
| 11 | F | 0,99 |
| 12 | Fm | 0,85 |
| 13 | F# | 0,84 |
| 14 | F#m | 1,00 |
| 15 | G | 0,83 |
| 16 | Gm | 0,98 |
| 17 | G# | 0,92 |
| 18 | G#m | 0,86 |
| 19 | A | 0,98 |
| 20 | Am | 0,80 |
| 21 | A# | 0,82 |
| 22 | A#m | 1,00 |
| 23 | B | 0,86 |
| 24 | Bm | 0,86 |
| Rata-rata | | 0,87 |

# CONCLUSIONS AND SUGGESTIONS

1. Conclusions

Dari pengujian akurasi yang telah dilakukan, dapat disimpulkan bahwa pengujian akurasi pada masing-masing kelas memberikan nilai F1 yang cukup baik. Nilai F1 yang paling besar diperoleh pada pengujian akor F# minor dan A# minor yaitu sebesar 100 persen. Nilai F1 yang paling kecil diperoleh pada pengujian akor D minor yaitu sebesar 75 persen. Rata-rata nilai F1 dari keseluruhan pengujian pengenalan akor berdasarkan nada root adalah sebesar 87 persen.

1. Suggestions

Saran yang dapat diberikan untuk memperluas penelitian ini adalah sebagai berikut:

1. Meneliti pengenalan akor dari suara alat musik selain piano dan bass.
2. Melakukan penelitian menggunakan metode KNN dan membandingkan hasilnya dengan metode CNN.

References

1. AlAmeri, Bakheet, 2017, The Implications Of Implementing Electronic Human Resource Management In Abu Dhabi, Department, Liverpool John Moores University.
2. Adli, M., Gharib, S., Hakami, M., and Pourmahdi, K., 2014, A Survey on Electronic Human Resource Management Consequences, Its Outcomes and Performances, Journal of Basic and Applied Scientific Research, 4(10) 71-82.
3. Alfa, A.AG., Rachmatin, D., and Agustina, F., 2017, Analysis of the Effect of Consumer Decision Factors With Structural Equation Modeling Partial Least Square, EurekaMatika, Vol.5, No.2, pp. 59 – 71.
4. Atallah, A.A., 2016, The Impact of Electronic Human Resource Management (E-HRM) on Organizational Development of UNRWA in Gaza Strip, The Islamic University Gaza.
5. Businesswire, 2018, As Global Leader in Digital Transformation Market Research, IDC Reveals Worldwide Digital Transformation Predictions. Available at

https://www.businesswire.com/news/home/20181107005026/

1. Davoudi, S.M.M., and Fartash, K., 2012, Electronic Human Resource Management: New Avenues Which Leads To Organizational Success, Spectrum: A Journal of Multidisciplinary Research, Vol.1, Issue 2.
2. Dewi, D.A.N.N., 2018, Validity and Reliability Test Module, Available at

https://www.researchgate.net/publication/328600462

1. Fındıklı, M.A., and Bayarçelik, E.B., 2015, Exploring the outcomes of Electronic Human Resource Management (E-HRM)?, Procedia - Social and Behavioral Sciences, 207, 424 – 431.
2. Haryono, Siswoyo, 2016, Practical Module Analysis Sem Method for Management Research with PLS, Yogyakarta : Pascasarjana Universitas Muhammadiyah Yogyakarta.
3. Iqbal, N., Ahmad, M., Allen, M., Raziq, M.M., 2018, Does E-HRM Improve Labour Productivity? A Study of Commercial Bank Workplaces in Pakistan, Employee Relations, 40(2).
4. Juliandi, A.*,* 2018*,* Structural Equation Model Based Partial Least Square (SEM-PLS): Use SmartPLS. Practice SEM-PLS Program Pascasarjana. Universitas Batam on December,16-17 2018, Batam: Universitas Batam.
5. Lengnick-Hall, Mark L., and Moritz, S., 2003, The Impact of E-HR on the Human Resource Management Function, Journal Of Labor Research, Vol. XXIV, No. 3, pp. 365 – 379.
6. Ma, Li, and Ye, Maolin, 2015, The Role Of Electronic Human Resource Management In Contemporary Human Resource Management, Open Journal of Social Sciences, 3, 71-78.
7. Maatman, Marco, 2006, Measuring the effectiveness of E-HRM The development of an analytical framework for the measurement of e-HRM and its application within a Dutch Ministry, University of Twente.
8. Nivlouei, F.B., 2014, Electronic Human Resource Management System: The Main Element in Capacitating Globalization Paradigm, International Journal of Business and Social Science, Vol. 5 No. 2, pp. 147 – 159.
9. Parry, E., and Tyson, S., 2011, Desired goals and actual outcomes of E-HRM, Human Resource Management Journal, Vol. 21, No. 3, Pp. 335 –354.
10. Priyono, 2016, Quantitative Research Methods, Surabaya : Zifatama Publishing.
11. Setyawan, Nurlina, 2015, Analysis of Factors Affecting the Intention to Reuse Information-Based Mobile Applications: Case Study of Abc Applications on PT. XYZ, Depok : Universitas Indonesia.
12. Wright, P. M., and Dyer, Lee., 2000, People in the E-Business: New Challenges, New Solutions, Center for Advanced Human Resource Studies : DigitalCommons@ILR.

1. Sri Waluyanti, Buku Direktorat PSMK Untuk Tehnik Audio Video, (Jakarta: Direktorat Pembinaan SMK, 2008), h.1. [↑](#footnote-ref-0)
2. Data School, Simple Guide to Confusion Matrix, https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/, 14 Desember 2019. [↑](#footnote-ref-1)
3. Muthu, Understanding the Classification Report through Sklearn, https://muthu.co/understanding-the-classification-report-in-sklearn/, 15 Desember 2019. [↑](#footnote-ref-2)