

ANALISIS PREDIKSI PIALA DUNIA 2022 DENGAN MENGGUNAKAN METODE DECISION TREE DAN ALGORITMA CART

Prihananti Liazizah Putri; Agung Riyardi

Ilmu Ekonomi Studi Pembangunan, Fakultas Ekonomi dan Bisnis, Universitas Muhammadiyah Surakarta

Abstrak

Sepakbola merupakan salah satu cabang olahraga dengan jumlah penggemar terbanyak di dunia. Salah satu ajang pertandingan sepakbola terbesar yaitu Piala Dunia. Piala Dunia diadakan setiap empat tahun sekali dan diikuti oleh berbagai negara di belahan dunia. Salah satu yang dinantikan ketika menonton pertandingan sepakbola yaitu hasil prediksi hasil pertandingan kedua tim. Meskipun hasil prediksi tak selamanya benar, akan tetapi hasil prediksi ini nantinya akan digunakan untuk menyusun strategi dan pola ketika akan bertanding melawan tim dengan mempertimbangkan hasil prediksi tersebut. Penelitian ini bertujuan untuk memprediksi hasil pertandingan Piala Dunia 2022 dengan menggunakan metode *decision tree* dan algoritma *CART* melalui *google collab* dan menggunakan bahasa pemrograman *python*. *CART* digambarkan sebagai struktur yang menyerupai sebuah pohon. Hasil dari penelitian ini membuktikan bahwa algoritma *CART* memberikan kinerja yang sangat baik dengan hasil akurasi sebesar 0.6659, sehingga metode *decision tree* dan algoritma *CART* sangat cocok diterapkan untuk memprediksi hasil pertandingan Piala Dunia 2022.

Kata kunci: Sepakbola, *Decision Tree*, *CART*

Abstract

Football is one of the sports with the most fans in the world. One of the biggest football match events is the World Cup. The World Cup is held every four years and is participated in by various countries around the world. One of the things to look forward to when watching a football match is the prediction of the results of the match between the two teams. Even though the prediction results are not always correct, the results of this prediction will later be used to develop strategies and patterns when playing against a team by considering the results of these predictions. This study aims to predict the results of the 2022 World Cup matches using the decision tree method and the *CART* algorithm via *Google collab* and using the *Python* programming language. *CART* is described as a structure that resembles a tree. The results of this study prove that the *CART* algorithm provides very good performance with an accuracy of 0.6659, so the decision tree method and the *CART* algorithm are very suitable to be applied to predict the results of the 2022 World Cup matches.

Keyword: Football, *Decision Tree*, *CART*

1. PENDAHULUAN

Dalam mempersiapkan mahasiswa untuk menghadapi perubahan sosial, budaya, teknologi dan dunia kerja. Mahasiswa dituntut harus mempersiapkan kebutuhan zaman yang begitu cepat. Perguruan Tinggi harus dapat merancang proses pembelajaran yang inovatif. Digagaskan kebijakan Merdeka Belajar- Kampus Merdeka oleh Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi (Kemendikbud-Dikti) untuk menjawab tuntutan tersebut. Diharapkan melalui program Kampus Merdeka dapat mendorong mahasiswa untuk menguasai berbagai keilmuan secara otonom dan fleksibel. Sehingga tercipta kultur belajar yang inovasi, tidak mengekang dan sesuai dengan kebutuhan mahasiswa untuk menghadapi masa depan. Dalam mendukung program tersebut, Kemendikbud - Dikti mengajak untuk berkolaborasi bersama mitra yang terdiri dari perusahaan multinasional, perusahaan teknologi global, startup teknologi, organisasi multilateral, BUMN, BUMD dan organisasi nirlaba.

Program Synthesis Academy merupakan kegiatan Studi Independen berbasis *Artificial Intelligence For GenZ Jobseekers. AI 4 GenZ Jobseekers* berpusat pada kompetensi untuk membangun kesiapan AI (*AI Readiness*) dengan meningkatkan keterampilan individu dalam bidang AI yang dibutuhkan sebagai persiapan untuk memasuki dunia pekerjaan masa depan. Program ini juga memiliki cakupan dalam pengembangan mindset dan skillset untuk mencapai keberhasilan dalam perkembangan teknologi dunia yang sangat dinamis menuju pemutakhiran dan kebergantungan di bidang AI.

Dalam laporan ini, secara khusus penulis menguraikan dan menggambarkan kegiatan *Syntesis Academy* melalui program *Studi Independen Artificial Intelligence For GenZ Jobseekers* Mempersiapkan generasi muda dalam memasuki dunia kerja maupun merintis usaha yang berbasis inovasi berkelanjutan serta terciptanya pengembangan skill dan mindset terhadap perkembangan AI untuk tercapainya keberhasilan dalam dunia teknologi dunia.

2. METODE

2.1 Tahap Persiapan

Tahap persiapan mengenai magang studi independen dalam *Artificial Intelligence*. Metode penelitian ini menggunakan pendekatan kualitatif,

dengan metode kualitatif deskriptif. Penulis menggunakan teknik pengumpulan data yang bersifat kualitatif. Metode penelitian kualitatif yang dapat digunakan dalam penelitian dengan observasi, dan dokumentasi dalam pengumpulan data. Penelitian ini menjelaskan bagaimana memprediksi pertandingan sepak bola FIFA 2022.

2.2 Tahap Pelaksanaan

Dalam tahap pelaksanaan dalam metode penelitian ini dilakukan dengan mengumpulkan data, menganalisis informasi, menjalankan observasi. Tahap ini merupakan kegiatan mahasiswa melaksanakan dan mengaplikasikan di sektor teknologi dalam bidang *Artificial Intelligence*. Pada kegiatan ini mahasiswa belajar dan mendapatkan pengetahuan mengenai kemajuan dan kecanggihan teknologi di masa depan, serta keterampilan yang dapat disesuaikan oleh mahasiswa itu sendiri dan juga pada kegiatan magang ini bertujuan untuk membangun Teknologi kecerdasan buatan (AI) memiliki beberapa keterampilan penting yang dapat membantu mahasiswa dalam berbagai aspek kehidupan dan karier. Magang Studi Independen (MSIB) teknologi AI dapat mengembangkan keterampilan ini, mahasiswa dapat membantu mahasiswa mengembangkan keterampilan *soft skill*. Pengembangan diri dan mahasiswa dapat menjadikannya pekerja yang lebih berkembang dan mampu beradaptasi dalam kecerdasan buatan.

2.3 Tahap Penilaian

Tahap ini merupakan evaluasi penilaian dari kegiatan Magang Studi Independen (MSIB) selama proses kegiatan belajar berlangsung. Tahap ini digunakan untuk mengukur dan mengevaluasi seberapa besar peserta Magang Studi Independen (MSIB) Mengevaluasi hasil yang telah diperoleh selama magang. Tahap penilaian dalam metode penelitian ini mengevaluasi kualitas data, analisis, dan kesimpulan yang dihasilkan.

Dengan verifikasi data dengan sumbernya dan pastikan konsistensi antara data yang tercatat dan data yang terkumpul. Dengan memelalui tahapan tersebut mahasiswa dapat mengevaluasi hasil magang yang telah dicapai yang berkualitas. Serta dapat membantu mahasiswa dalam memperoleh umpan

balik dan saran dari pembimbing magang dan pihak yang terkait untuk pengembangan diri di masa depan.

2.4 Tahap Tindak Lanjut

Tahap tindak lanjut ini penulis menyelesaikan menyelesaikan tugas dari data penelitian yang telah di observasi. Penulis dapat mengevaluasi pencapaian mereka selama penelitian, merumuskan rencana pengembangan diri, dan menyusun langkah-langkah perbaikan serta pengembangan untuk masa depan berdasarkan hasil penelitian yang telah dicapai. Dalam tahap tindak lanjut, penting untuk memastikan bahwa implementasi penelitian dilakukan dengan benar dan efektif, serta mengevaluasi hasilnya untuk memastikan keberhasilan dan keberlanjutan penelitian ini.

3. HASIL DAN PEMBAHASAN

Data Acquisition, merupakan serangkaian proses pengumpulan dan validasi data yang dibutuhkan dalam proyek. Pada proyek ini, saya membantu tim untuk melakukan pengumpulan dataset, pengambilan dataset diambil dari salah satu website yaitu *Kaggle*.

Data *Eksplorasi*, bertujuan untuk memahami karakteristik data serta membantu dalam pemilihan *tools* yang tepat dalam melakukan preprocessing atau analisis data melalui pattern data. Membantu dalam membuat model algoritma yang dibutuhkan seperti *Decision Tree*, *Algoritma Classification and Regression Tree (CART)*. Penyusunan laporan akhir, menyusun laporan akhir proyek dimulai dari penentuan latar belakang, masalah dan solusi yang akan diselesaikan, lalu dilanjutkan dengan penyusunan bab selanjutnya yaitu latar belakang proyek akhir. Bab dan sub bab selanjutnya pada laporan akhir dan artikel disusun bersama. Penyusunan jurnal, membuat dan menentukan pendahuluan dan metode yang digunakan pada penulisan akhir, selain itu mengumpulkan beberapa artikel penelitian terdahulu yang dapat dijadikan referensi dalam penyusunan projekakhir. Penyusunan PPT yang berisikan hasil pembuatan jurnal, yang digunakan untuk dipresentasikan.

Sepak bola merupakan salah satu cabang olahraga yang sangat populer di dunia. Sepakbola bertujuan untuk mencetak gol ke gawang lawan sebanyak mungkin guna memperoleh kemenangan. Sepakbola saat ini bukan hanya milik dari mereka yang

terlibat langsung, tapi juga merupakan olahraga milik seluruh masyarakat dunia. Setiap harinya selalu ada berita, ulasan pertandingan, bahkan prediksi atau perkiraan hasil pertandingan sepakbola, baik media cetak, media elektronik ataupun dari para penggemar olahraga ini. Salah satu pertandingan terbesar dari sepakbola yaitu Piala Dunia. Piala dunia merupakan kompetisi olahraga sepakbola bertaraf internasional yang diikuti oleh tim nasional dari berbagai negara yang diselenggarakan setiap empat tahun sekali. Pada tahun ini, tuan rumah piala dunia 2022 berada di Qatar dan diikuti oleh 32 tim dari berbagai negara yang bersaing untuk memperebutkan gelar juara. Selain itu, sepakbola tak hanya menjadi olahraga tetapi juga menjadi industri yang terdiri dari beberapa aspek, seperti bisnis, hiburan, pengetahuan, dan kemajuan teknologi.

Semakin populernya olahraga sepakbola, kebutuhan prediksi terhadap hasil pertandingan kian semakin meningkat. Hasil prediksi sangat diperlukan untuk berbagai bidang. Salah satu tujuan dari hasil prediksi sepak bola yaitu sebagai bahan dalam penyusunan strategi dan pola yang akan digunakan untuk melawan tim dengan mempertimbangkan hasil prediksi tersebut. Prediksi merupakan proses memperkirakan sesuatu yang akan terjadi di masa depan berdasarkan data yang diperoleh di masa lalu hingga sekarang. Selain itu, prediksi bola merupakan bahan yang sedang hangat-hangatnya diperbincangkan bagi pecinta bola yang menginginkan tim kesayangannya untuk menang. Hasil prediksi tidak sepenuhnya benar, bahkan tak seorang pun dapat memprediksi hasil pertandingan sepakbola mana yang akan menang, kalah ataupun seri. Presentase hasil prediksi akan sama sesuai dengan fenomena yang terjadi sangatlah kecil. Hal inilah yang menjadi masalah. Untuk mengatasi permasalahan ini, banyak orang telah membuat sistem untuk membantu dalam menyelesaikan permasalahan tersebut. Salah satunya dengan menggunakan *decision tree* (pohon keputusan).

Pada penelitian sejenis yang dilakukan untuk memprediksi luas lahan panen tanaman padi di Kabupaten Karawang menunjukkan hasil bahwa algoritma *CART* memiliki performa lebih baik dari algoritma KNN serta dinyatakan sebagai korelasi sempurna. Kemudian penelitian sejenis untuk menentukan jurusan siswa menunjukkan hasil bahwa algoritma *CART* sangat direkomendasikan untuk mengklasifikasikan penjurusan siswa, karena berdasarkan perhitungannya,

algoritma *CART* memang lebih baik.

Dalam penelitian ini *decision tree* dengan algoritma *CART* digunakan untuk memprediksi hasil pertandingan Piala Dunia 2022. Karena piala dunia merupakan ajang kompetisi sepakbola terbesar di dunia yang dilaksanakan setiap 4 tahun sekali dan memiliki banyak penggemar di seluruh dunia. Selain itu, hasil yang diperoleh dari algoritma *CART* menunjukkan hasil yang lebih akurat jika dibandingkan dengan algoritma lain.

Proses pelaksanaan proyek akhir Studi Independen Bersertifikat dilaksanakan dengan beberapa kegiatan, diantaranya : Penentuan tim yang digunakan untuk proses penyelesaian proyek akhir; Menentukan ide untuk proyek akhir yang akan dilakukan; Menganalisis ide yang telah ditentukan bersama tim dan menuliskan hasil diskusi kedalam laporan proyek akhir; Melakukan *Data Acquisition* bersama tim untuk mengumpulkan data yang dibutuhkan dalam proyek akhir. Dataset yang diperlukan oleh tim didapatkan dari website *Kaggle*; Melakukan *Data Exploration*, selanjutnya tim melakukan proses eksplorasi data yang bertujuan untuk memahami karakteristik data yang dapat membantu dalam pemilihan *tools*, langkah pertama tim mengumpulkan *code* yang digunakan seperti *code* untuk menampilkan info dan *describe* data. Setelah itu tim akan melakukan data visualization dengan tujuan menampilkan data menjadi grafis atau visual dengan tujuan yaitu timakan lebih mudah memahami pola atau *outlier* pada data. Tim menggunakan *library matplotlib*; *Modelling*, tim melakukan *modelling* dengan tujuan merepresentasikan masalah menjadi bentuk yang lebih sederhana sehingga jelas dan mudah dikerjakan; Penggunaan Algoritma dalam implentasi pendukung jurnal yang kami susun disini yaitu *Machine Learning* dengan *Algoritma Decision Tree Algoritma Classification and Regression Tree (CART)*, lalu mengevaluasi beberapa model tersebut.

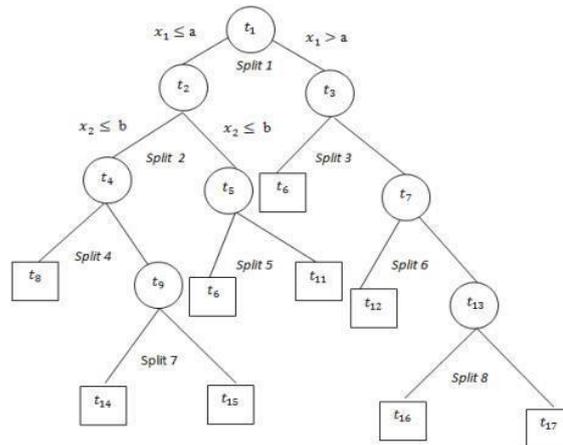
Decision Tree (pohon keputusan) merupakan teknik pengambilan keputusan dengan menggunakan diagram alir (*flowchart*) yang mirip pohon. *Decision Tree* terdiri dari beberapa node (simpul) dan *edge* (rusuk). Simpul pada *decision tree* terbagi menjadi 3 bagian yaitu *root node*, *branch node*, dan *leaf node*. Metode ini sangat umum dan populer untuk digunakan. Ada beberapa algoritma dalam decision tree, seperti *C4.5*, *CART*, dan *ID3*.

Decision Tree memberikan gambaran sederhana mengenai metode analisis dari atribut yang telah diuji untuk kelompok kelas tertentu dimana *node* akar sebagai *node* teratas memiliki dampak paling besar. Untuk nama atributnya sendiri menggunakan simpul akar dan simpul internal, nilai atributnya yaitu rusuk-rusuknya, sedangkan kelompok kelas yang berbeda menggunakan simpul daun. Proses yang terjadi dalam *decision tree* yaitu mengubah data tabel menjadi model pohon lalu mengubah model pohon menjadi aturan. Pada *decision tree* terdapat beberapa tahapan dalam algoritma yaitu menyiapkan data *traininig*, memilih atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan mengunlangi proses untuk setiap cabang pada kasus tersebut. Tujuan menggunakan *decision tree* ini adalah untuk membuat training model yang dapat digunakan untuk memproduksi sesuai keputusan yang telah diambil. Cara kerja pohon keputusan ini yaitu dengan memperlihatkan peluang/faktor-faktor kemungkinan yang dapat mempengaruhi pemilihan terhadap keputusan tersebut disertai dengan perkiraan hasil akhir yang diperoleh jika memilih alternatif tersebut.

Adapun manfaat dan kelebihan dari pohon keputusan yaitu dapat memperjelas/merinci proses pengambilan keputusan sehingga menjadi lebih sederhana dan menjelaskan solusi dari permasalahan tersebut serta meneliti data dan menemukan hubungan antara calon variabel input dengan variabel target. Sedangkan kelemahan dari pohon keputusan yaitu hasil keputusan sangat bergantung dari desain pohon, sering terjadi *overlapping* jika jumlahnya banyak, sulit dalam membuat pohon keputusan, dan derajat *error* yang dihasilkan sangat besar. Dalam hal ini, manusia dan teknologi memiliki peran masing-masing. Manusia adalah sumber pengetahuan karena mereka yang mengembangkan dan menyebarkannya. Teknologi adalah komponen kunci dalam memungkinkan orang untuk menjalankan proses dan membuat pengetahuan tersedia di mana saja dan kapan saja.

Classification and Regression Tree (CART) merupakan salah satu algoritma dari pohon keputusan (*decision tree*) yang paling umum dan banyak digunakan dengan menggabungkan berbagai faktor yang berasal dari berbagai sumber yang berbeda untuk permasalahan klasifikasi dan regresi berdasarkan *biner rekursif*. *CART* tergolong ke dalam data non parametrik yang tidak memerlukan kerangka

fungsional dan merupakan alat yang mampu memprediksi suatu permasalahan dan penggolongan data. *CART* digambarkan struktur yang menyerupai sebuah pohon.



Gambar 1. Model Klasifikasi *CART*

Keterangan: Dalam membuat pohon klasifikasi, *CART* bekerja dalam 5 tahapan, yaitu : *Tree building* yaitu proses pembuatan pohon klasifikasi. Proses ini terdiri dari proses pemecahan simpul utama menjadi 2 buah simpul anak melalui *splitting rules* dan dilakukan secara berulang. Proses *class assignment* yaitu proses identifikasi simpul yang terbentuk dalam suatu. *Stopping the trees building process*. Dalam tahapan ini pohon akhir telah terbentuk. *Pruning the tree* yaitu proses memotong pohon terakhir menjadi pohon yang lebih kecil. *Optimal tree selection process* yaitu proses memilih pohon klasifikasi yang baik

Penelitian ini bertujuan untuk menganalisis hasil prediksi Piala Dunia 2022 dengan menggunakan metode *decision tree* dan algoritma *CART*. Data yang digunakan yaitu hasil skor pertandingan Piala Dunia 2022 yang diambil dari website <https://www.kaggle.com>. Disini kami menggunakan tiga dataset dalam menentukan tingkat *accuracy* klasifikasi prediksi yaitu: *FIFA rankings from 1993 to 2022*. Kami menggunakan ini untuk mendapatkan peringkat dan poin FIFA untuk tim, yang merupakan peringkat perubahan bulanan yang sebelumnya ditampilkan sebagai prediktor kinerja tim yang layak.

International Soccer matches from 1872 to 2022. Untuk mengetahui seberapa besar selisih *point*, rank dan rank tim saat ini mempengaruhi *outcome* suatu pertandingan.

```
rankings = pd.read_csv('fifa_ranking.csv')
rankings = rankings.loc[:, ['rank', 'country_full', 'country_abrv', 'cur_year_avg_weighted', 'rank_date',
                            'two_year_ago_weighted', 'three_year_ago_weighted']]
rankings = rankings.replace({'IR Iran': 'Iran'})
rankings['weighted_points'] = rankings['cur_year_avg_weighted'] + rankings['two_year_ago_weighted'] + rankings['three_year_ago_weighted']
rankings['rank_date'] = pd.to_datetime(rankings['rank_date'])
rankings.head()
```

	rank	country_full	country_abrv	cur_year_avg_weighted	rank_date	two_year_ago_weighted	three_year_ago_weighted	weighted_points
0	1	Germany	GER	0.0	1993-08-08	0.0	0.0	0.0
1	2	Italy	ITA	0.0	1993-08-08	0.0	0.0	0.0
2	3	Switzerland	SUI	0.0	1993-08-08	0.0	0.0	0.0
3	4	Sweden	SWE	0.0	1993-08-08	0.0	0.0	0.0
4	5	Argentina	ARG	0.0	1993-08-08	0.0	0.0	0.0

Gambar 2. Dataset FIFA_ranking

```
matches = pd.read_csv('results2018.csv')
matches = matches.replace({'Germany DR': 'Germany', 'China': 'China PR'})
matches['date'] = pd.to_datetime(matches['date'])
matches.head()
```

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False

Gambar 3. Dataset results

FIFA *World Cup* 2022 dataset digunakan untuk mendapatkan pertandingan yang akan datang. *Feature Extraction* merupakan set awal yang berupa data mentah direduksi menjadi grup yang lebih mudah dikelola untuk diproses. Karakteristik kumpulan data besar ini adalah sejumlah besar variabel yang membutuhkan banyak sumber daya komputasi untuk diproses.

```

world_cup = pd.read_csv('World Cup 2018 Dataset.csv')
world_cup = world_cup.loc[:, ['Team', 'Group', 'First match \n against', 'Second match\n against', 'Third match\n against']]
world_cup = world_cup.dropna(how='all')
world_cup = world_cup.replace({"IRAN": "Iran",
                              "Costarica": "Costa Rica",
                              "Porugal": "Portugal",
                              "Columbia": "Colombia",
                              "Korea": "Korea Republic"})
world_cup = world_cup.set_index('Team')
world_cup.head()

```

Group	First match \n against	Second match\n against	Third match\n against
Russia	A	Saudi Arabia	Egypt
Saudi Arabia	A	Russia	Uruguay
Egypt	A	Uruguay	Russia
Uruguay	A	Egypt	Saudi Arabia
Portugal	B	Spain	Morocco

Gambar 4. Dataset World Cup 2022

	average_rank	rank_difference	point_difference	is_stake
10901	49.5	95.0	0.0	False
10940	69.0	126.0	0.0	True
7049	119.0	-60.0	0.0	True
10031	99.5	-75.0	0.0	False
15530	42.5	-23.0	262.1	False

X_test.head()

	average_rank	rank_difference	point_difference	is_stake
290	4.0	2.0	0.00	False
5903	130.0	20.0	0.00	False
15920	91.5	-27.0	42.01	True
17338	109.0	-40.0	121.16	True
4288	65.5	-39.0	0.00	True

Gambar 5. Source Code Feature Extraction

Modelling dengan menggunakan kriteria gini index untuk menemukan skor akurasi.training set.

```

from sklearn.metrics import accuracy_score

print('Model accuracy score with criterion gini index: {0.04f}'.format(accuracy_score(y_test, y_pred_gini)))

Model accuracy score with criterion gini index: 0.6639

y_pred_train_gini = clf_gini.predict(X_train)
y_pred_train_gini

array([False, False, True, ..., False, True, False])

print('Training-set accuracy score: {0.04f}'.format(accuracy_score(y_train, y_pred_train_gini)))

Training-set accuracy score: 0.6829

print('Training set score: {:.4f}'.format(clf_gini.score(X_train, y_train)))

print('Test set score: {:.4f}'.format(clf_gini.score(X_test, y_test)))

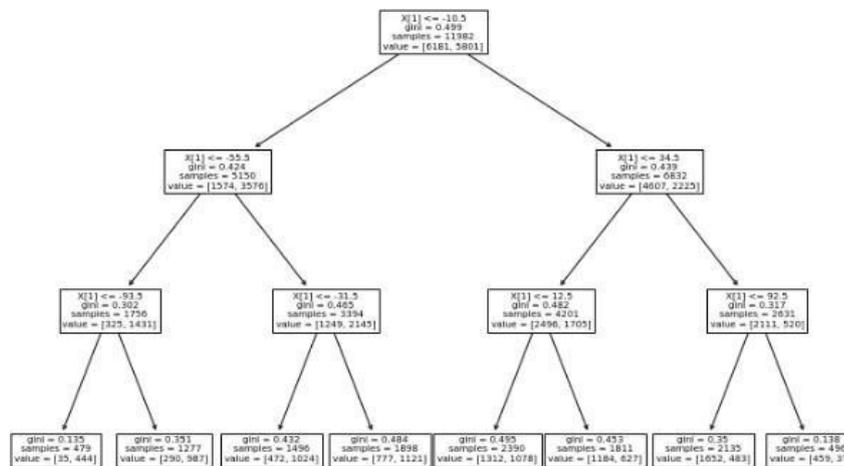
Training set score: 0.6829
Test set score: 0.6639

```

Gambar 6 Decision Tree Classifier dengan kriteria gini inde

Di sini, skor *akurasi training-set* adalah 0.6829 sedangkan akurasi test-set menjadi 0.6659. Kedua nilai ini cukup sebanding. Jadi, tidak ada tanda-tanda *overfitting*. Selanjutnya kami melakukan *modelling Decision Tree Classifier* dengan kriteria *entropy*.

Kita dapat melihat bahwa skor set pelatihan dan skor set tes sama seperti di atas. Skor akurasi *training-set* adalah 0,7865 sedangkan akurasi *test-set* menjadi 0,8021. Kedua nilai ini cukup sebanding. Jadi, tidak ada tanda-tanda *overfitting*.



Gambar 8 Visualisasi Pohon Keputusan

Confusion Matrix adalah alat untuk meringkas kinerja algoritma klasifikasi. *Confusion matrix* akan memberi kita gambaran yang jelas tentang kinerja model klasifikasi dan jenis kesalahan yang dihasilkan oleh model tersebut. Ini memberi kita ringkasan prediksi yang benar dan salah yang dikelompokkan berdasarkan masing-masing kategori. Ringkasan disajikan dalam bentuk tabel.

```
# Print the Confusion Matrix and slice it into four pieces

from sklearn.metrics import confusion_matrix

cm = confusion_matrix(y_test, y_pred_en)

print('Confusion matrix\n\n', cm)

Confusion matrix

[[2227 823]
 [1149 1704]]
```

Gambar 9 *Confusion Matrix*

Classification Report adalah cara lain untuk mengevaluasi kinerja model klasifikasi. Ini menampilkan skor presisi, daya ingat, f1, dan dukungan untuk model tersebut. Saya telah menjelaskan istilah-istilah ini nanti.

```
from sklearn.metrics import classification_report

print(classification_report(y_test, y_pred_en))

precision recall f1-score support

False 0.66 0.73 0.69 3050
True 0.67 0.60 0.63 2853

accuracy 0.67 5903
macro avg 0.67 0.66 0.66 5903
weighted avg 0.67 0.67 0.66 5903
```

Gambar 10 *Classification Report*

4. PENUTUP

4.1 Kesimpulan

Kampus merdeka merupakan wujud pembelajaran di perguruan tinggi yang otonom dan fleksibel sehingga tercipta kultur belajar yang inovatif, tidak mengekang, dan sesuai dengan kebutuhan mahasiswa. Pelaksanaan Program Magang Studi Independen Bersertifikat di Synthesis Academy telah

memberikan pengalaman dan pengetahuan yang mendalam mengenai teknologi di AI dalam dunia industri. *Synthesis Academy* menyelenggarakan program Studi Independen Bersertifikat salah satunya *Artificial Intelligence For GenZ Job Seekers*. Program ini berlangsung selama 5 bulan yang dimulai pada 16 Agustus - 31 Desember 2022. Dari mengikuti Program Studi Independen ini, penulis mendapatkan pengalaman pengembangan *karier* dan *skill*. Setelah mengikuti Program Kegiatan Magang Studi Independen Bersertifikat Batch 3 tahun 2022 *Academy (Artificial Intelligence)*, maka kesimpulan yang dapat saya ambil adalah sebagai berikut:

Dengan adanya program ini salah satu inovasi dan wawasan luas di mana mahasiswa diberikan kesempatan mengikuti kegiatan yang diselenggarakan pemerintah melalui Kampus Merdeka. Mahasiswa diajak untuk mengembangkan kemandirian dalam mencari dan menemukan pengetahuan luar di luar kampus melalui kenyataan dan dinamika lapangan seperti interaksi sosial, manajemen diri, tuntutan kinerja, target dan pencapaiannya.

Pihak perusahaan *Pt Synthesis Academy* yang menjadi penyelenggara kegiatan program Magang Studi Independen Bersertifikat membantu mahasiswa turut mendapatkan pengetahuan dan pengalaman diluar kemampuan yang belum diketahui mahasiswa. sehingga dengan adanya kegiatan program tersebut sangat berguna bagi mahasiswa, perusahaan, dan pihak kampus yang dapat diimplementasikan setelah mengikuti kegiatan tersebut.

Melalui program Kampus Merdeka yang dirancang dan diimplementasikan dengan baik dan *soft skills* mahasiswa akan terbentuk dengan lebih baik. dengan mengikuti Program Kampus Merdeka diharapkan dapat menjawab tantangan Perguruan Tinggi untuk menghasilkan lulusan yang sesuai perkembangan zaman, kemajuan IPTEK, tuntutan dunia usaha, dunia industri dan dinamika masyarakat menjadi lebih luas dan berkembang.

Berdasarkan Proyek Akhir, kami membuat model Pengklasifikasi Pohon Keputusan untuk memprediksi Piala Dunia 2022. Saya membangun dua model, satu dengan indeks kriteria gini dan satu lagi dengan entropi kriteria. Model menghasilkan kinerja yang sangat baik yang ditunjukkan oleh akurasi

model pada kedua kasus yang ditemukan sebesar 0.6659. Pada model dengan indeks gini kriteria, skor akurasi *training-set* adalah 0.6829 sedangkan akurasi *test-set* menjadi 0.6659. Kedua nilai ini cukup sebanding. Jadi, tidak ada tanda-tanda *overfitting*. Demikian pula, pada model dengan *entropi* kriteria, skor akurasi set pelatihan adalah 0.6829 sedangkan akurasi set tes menjadi 0.6659. Kami mendapatkan nilai yang sama seperti dalam kasus dengan kriteria gini. Jadi, tidak ada tanda-tanda *overfitting*. Dalam kedua kasus tersebut, skor akurasi set pelatihan dan set tes adalah sama. *Matriks* konfusi dan laporan klasifikasi menghasilkan kinerja model yang cukup baik.

4.2 Saran

Setelah kurang lebih selama 5 bulan mengikuti Program Magang Studi Independen Bersertifikat (MSIB), berikut beberapa saran tersebut : Dalam pelaksanaan MSIB, diharapkan kepada mitra selaku penyelenggara dan pelaksana kegiatan memberikan informasi terkait kegiatan salam mengikuti program tersebut kepada pihak kampus, kepala program studi dan dosen pembimbing, serta jadwal dalam melakukan pertemuan lebih fleksibel dan tidak terlalu dekat.

Untuk melakukan prediksi hasil pertandingan sepakbola dapat menambahkan sampel data yang baru dan lebih lengkap lagi serta data dalam jumlah besar, artinya semakin besar jumlah data yang digunakan maka semakin akurat hasil prediksi yang dihasilkan. Serta memilih atribut yang berbobot yang dapat meningkatkan nilai prediksi. Menambahkan model Deployment agar hasil prediksi dapat diakses secara online

DAFTAR PUSTAKA

- A. R. Arrahimi, M. K. Ihsan, D. Kartini, M. R. Faisal, and F. Indriani, “Teknik Bagging Dan Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa,” *J. Sains dan Inform.*, vol. 5, no. 1, pp. 21–30, 2019, doi: 10.34128/jsi.v5i1.171.
- D. . Prabowo, “Prediksi hasil pertandingan sepakbola english premier league dengan

- menggunakan algoritma k-nearest neighbors dan naive bayes classifier,” Skripsi, 2020.
- Hermawati and Fajar Astuti, *Data Mining*. Yogyakarta: AndiOffset, 2013.
- H. S. Kurniawan, “Prediksi Hasil Pertandingan Liga Sepak Bola Menggunakan Metode Distribusi Poisson,” no. June, 2021.
- J. A. Sidette, E. Eko, and O. D. Nurhayati, “Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS,” *J. Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 75–86, 2014, doi: 10.21456/vol4iss2pp75-86.
- J. Han, K. Mao, T. Xu, J. Guo, Z. Zuo, and C. Gao, “A soil moisture estimation framework based on the CART algorithm and its application in China,” *Sci. Direct*, vol. 563, pp. 65–75, 2018, doi: <https://doi.org/10.1016/j.jhydrol.2018.05.051>.
- Juna Eska, “Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9–13, 2019.
- Mardi Yuli, “Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2014, [Online]. Available: <https://core.ac.uk/download/pdf/229192374.pdf>.
- Marwana, “Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola,” *J. Inform. Multimed.*, p. 55, 2014.
- M. F. Aziz, S. Defiyanti, and B. N. Sari, “Perbandingan Algoritma CART Dan K-Nearest Neighbor Untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang,” *J. TAM (Technology Accept. Model.)*, vol. 9, no. 2, pp. 74–78, 2018.
- M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, “Decision tree- 20 based diagnosis of coronary artery disease: CART model,” *Comput Methods Programs Biomed*, vol. 192, p. 105400, 2020, doi: 10.1016/j.cmpb.2020.105400.
- R. D. Oktayona, “Tinjauan Kemampuan Teknik Dasar Sepak Bola,” *J. Patriot*, vol. 2, no. 1, pp. 1259–1273, 2020.
- S. M. H. Ziabari and T. Bakhshpoori, “Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5’ and CART algorithms,” *Appl. Soft Comput.*, vol. 68, pp. 147–161, 2018, doi: <https://doi.org/10.1016/j.asoc.2018.03.052>.
- S. Monalisa and F. Hadi, “Penerapan Algoritma CART Dalam Menentukan Jurusan Siswa di MAN 1 Inhil,” *J. SISFOKOM (Sistem Inf. dan Komputer)*, vol. 09, no. 03, pp.

387–394, 2020.

Z. W. Mardika, M. A. Mukid, and H. Yasin, “Pembentukan Pohon Klasifikasi Biner Dengan Algoritma CART (Classification And Regression Trees) (Studi Kasus: Kredit Macet di PD. BPR-BKK Purwokerto Utara),” *J. Gaussian*, vol. 5, no. 3, pp. 583–592, 2016.