International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 12; Issue: 5; May 2025
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Research Paper

Digitalization of Disability Athlete Management System Using Decision Tree Cart Algorithm to Identify Potential Athletes in NPCI Kediri

Ar Rasyid Sarifullah Gilbijatno[1], Resty Wulanningrum[2], Siti Rochana[3]

Department of Informatics Engineering,
Universitas Nusantara PGRI Kediri, Kampus 2, Mojoroto Gang I, No. 6, Mojoroto, Kediri, East Java, Indonesia

Corresponding Author: Resty Wulanningrum

DOI: https://doi.org/10.52403/ijrr.20250545

ABSTRACT

This research addressed inefficient data management and subjective assessment in disability athlete selection by developing an artificial intelligence-based system for the National Paralympic Committee of Indonesia (NPCI) Kediri. The study employed Research and Development methodology to create a digitalized platform that integrated a centralized data management system with a Decision Tree CART algorithm enhanced through standard deviation-based data augmentation. The research utilized a dataset of 72 disability athletes categorized into four types: Daksa Lower, Daksa Upper, Cerebral Palsy, and Tuna Daksa. Data collection encompassed anthropometric measurements (height, weight), sociodemographic information (age, gender), disability classification, and performance test results (40-meter sprint, 600-meter run, arm hang, sit up, vertical jump). The system implementation included data collection, augmentation to address limited sample size, algorithm implementation with weighted balance parameters for class imbalance, model validation, and web-based interface development. Rigorous testing employed 10-fold cross-validation across multiple data splits (90:10, 80:20, 70:30), with the 80:20 split achieving perfect consistency (100% accuracy, 0% standard deviation). Disability type emerged as the primary discriminative feature, creating homogeneous groupings with pure Gini values at terminal nodes. The implemented system transformed subjective talent identification into an evidence-based process while streamlining administrative workflows through responsive interfaces tailored to various stakeholder needs.

Keywords: Decision Tree CART, Data Augmentation, Athlete Disability Management System, Artificial Intelligence

INTRODUCTION

NPCI (National Paralympic Committee of Indonesia) is the organization that houses athletes with special needs in Indonesia [1]. NPCI is derived from the International Paralympic Committee (IPC), with each country having its own National Paralympic Committee (NPC) which serves as a sports organization focused on achievement for athletes with disabilities [2]. As the national body responsible for disability sports development, NPCI operates through regional branches across Indonesia. In its implementation at the district level, NPCI Kediri faces several challenges in data management and the selection process of new athlete candidates. The manual data collection system can lead to unnecessary burden for administrators, resulting in annoyance, frustration, and potential abandonment of proper documentation processes [3]. Sports organizations can contribute more optimally in transforming sports data information management by utilizing management system digitalization [4].

Digitalization of management systems refers to the intensified application of information and communication technologies across all domains of economic and social activities [5]. When applied to sports organizations, this digital transformation addresses not only data management inefficiencies but also helps mitigate another significant challenge, the subjectivity factor in the selection process of new athlete candidates. The issue of subjectivity in athlete selection has been identified with evidence suggesting that the accuracy of subjective talent decisions by coaches and scouts is relatively low, which may lead to less accurate assessments and inconsistencies between individual decision-makers [6]. Subjective assessment can lead to errors in recognizing athlete potential, especially for athletes with disabilities who have special characteristics and capabilities.

The unique capabilities of athletes with disabilities are often overlooked due to systemic barriers in sport participation. Typical obstacles include inadequate awareness among non-disabled individuals about proper inclusive practices, insufficient training and competition opportunities, limited accessible facilities resulting from physical barriers, and restricted access to essential resources and information [7]. These challenges emphasize the need for greater visibility of the disability community in mainstream media to foster inclusion, establish role models and mentors, and support broader disability awareness initiatives [8].

Addressing these challenges requires robust systems that can effectively support athletes with disabilities in developing their unique capabilities. Athlete management systems play a crucial role in this regard, serving to refine training processes, enhance performance readiness, and minimize injury and illness risks while monitoring the development of physical capabilities and overall athlete progress [9]. Furthermore, the integration of advanced technology into these management systems significantly enhances measurement accuracy and streamlines the data collection process, liberating practitioners from the tedious manual recording methods that were standard practice in the past [10].

Along with technological developments, the integration of artificial intelligence into athlete data management systems has opened new opportunities in optimizing analysis and decision-making processes. Rather than being a single technology, Artificial Intelligence (AI) represents an evolving frontier of emerging computational capabilities that continually expands and transforms [11]. Contemporary AI is built upon machine learning technologies that demonstrate unprecedented levels of autonomy, learning capacity, and complexity compared to previous intelligent systems [12]. Recent advancements in AI have created computing systems capable of seeing, hearing, and learning, establishing innovative platforms that enhance analytical capabilities and skill development [13]. Among various artificial intelligence solutions available, this research focuses specifically on the implementation of the Decision Tree.

Decision Tree (DT) have emerged as a powerful ai model that can effectively support athlete classification and selection processes [14,15]. DT represent one of the data classification methods capable of transforming the complexity of decision-making into a simple and easily interpretable process [16]. DT consist of nodes and branches, where each node represents a feature or category, while each branch depicts rules or decisions for dividing data based on certain threshold values until the data can be classified [17]. This hierarchical structure allows DT to be tailored to different classification needs, ensuring that the assessment process remains both objective and transparent, which is essential for fair athlete selection in disability sports. One of the most well-known and widely used DT algorithms is CART.

CART (Classification and Regression Tree) algorithm is a sophisticated decision tree method that partitions data into two distinct subsets by determining the optimal split based on variables through separation rules at each node [18]. CART employs a binary division approach that not only simplifies the decision-making process but also produces an easily interpretable tree structure [14]. In constructing DT, CART utilizes advanced heuristic approaches that employ various criteria to specifically select features and determine branching conditions [19].

The algorithm strategically applies the Gini index to classify decision points within datasets, implementing separation rules that systematically enhance the algorithm's performance [18]. Compared to other segmentation techniques such as CHAID, AID, and QUEST, CART's main advantage lies in its ability to process quantitative variables and determine separation criteria based on the concept of impurity [20]. This impurity represents the level of data heterogeneity in CART analysis, where the Gini index plays a crucial role in determining class purity after division based on certain attributes [21].

The Gini index specifically measures class purity, operating on the principle that the best division will increase the purity of the resulting subsets [22]. In the context of this research, the Gini index calculation serves as a fundamental metric for the CART algorithm to identify the most discriminative features among the collected athlete data, enabling the system to effectively distinguish between potential and non-potential athletes based on their physical performance metrics and disability characteristics.

While these algorithmic approaches offer powerful classification capabilities, their application within sensitive contexts demands careful ethical consideration. The integration of artificial intelligence systems into decision-making frameworks raises significant concerns, as research indicates that poorly designed AI technologies can exhibit biases against individuals with disabilities, potentially leading to increased marginalization and negative impacts [23]. This risk is particularly evident in classification contexts, where conventional big data classification methods often fail to adequately address the distinctive individual characteristics exhibited by people with disabilities [24]. To address these limitations, this research employs data augmentation techniques that expand the range of available data points by artificially increasing the dataset's size and diversity, helping the DT model avoid overfitting to specific characteristics in the limited training data [25]. The augmented dataset enables the CART algorithm to better recognize natural variations among athletes with disabilities and generate more reliable classification outcomes.

Building on this approach, this research proposes the development of an artificial intelligence-based disability athlete management system at NPCI Kediri that integrates a centralized data management platform with an enhanced CART classification model. The proposed system architecture consists of three fundamental components: data acquisition, algorithmic processing, and results presentation. The data acquisition module facilitates comprehensive collection of athlete information including demographic data, disability classification, and anthropometric measurements. This process culminates in the determination of athlete potential with binary outcomes: potential athletes and non-potential athletes. The processing module implements the CART algorithm enhanced through standard deviation-based data augmentation to analyze athlete data and identify potential talent specifically for athletics (track and field) disciplines. The output interface provides customized dashboards with varying levels of detail for different stakeholders, including administrators, coaches, and the assessor.

LITERATURE REVIEW

Firstly, DT algorithms have been increasingly applied in sports contexts for talent identification and performance prediction. Wahyu Romadhonia et al. (2023) conducted a comprehensive study analyzing 11 sociodemographic and anthropometric variables within a dataset of 113 prospective athletes. Their research demonstrated the high efficacy of the DT model, achieving accuracy and precision rates exceeding 80% in identifying potential athletes.

In the domain of disability classification, Hassan and Mokhtar (2019) employed DT algorithms to analyze datasets from the National Database for Autism Research comprising nearly 3,000 individuals. Their analysis successfully identified 15 medical conditions highly associated with autism spectrum disorder (ASD) diagnoses and revealed six potentially hereditary medical conditions linked to ASD. The DT approach demonstrated remarkable effectiveness, achieving 90% accuracy in these associations.

Similarly, Sarabia et al. (2021) demonstrated the effectiveness of DT in classifying para-footballers with cerebral palsy by impairment severity. Their study evaluated classification models for different types of cerebral palsy, with accuracy rates ranging from 86.5% to 90.9% across different impairment profiles. For athletes with athetosis and ataxia, the model utilized five nodes and 11 leaves to achieve 86.5% precision, while the classification model for those with spastic hemiplegia achieved 90.9% accuracy with only two nodes and five leaves.

Regarding the implementation of digital systems in disability sports contexts, Haynes et al. (2024) highlighted the perceived value of Decision Support Systems (DSS) among fitness facility exercisers and adapted fitness center trainers. The study found unanimous agreement that such systems would be beneficial for exercise prescription and progression for individuals with disabilities. However, participants expressed concerns about individualization and safety, suggesting that effective systems should incorporate evidence-based, disease-specific exercise prescriptions, prescreening for contraindications, and fall risk assessment. Additionally, participants indicated that behavioral change strategies were desirable features for promoting sustained exercise participation among individuals with disabilities.

In exploring the readiness of sports professionals to adopt digital technologies, Shutova and Andryushchenko (2020) revealed significant gaps in digital literacy among physical education and sports experts. The study found that 32% of experts acknowledged the need to improve their professional IT skills, while 68% demonstrated limited understanding of the digitalization process in physical education. Moreover, 60% of coaches exhibited critical deficiencies in utilizing modern data collection techniques and information processing technologies. Despite 79% of respondents reporting regular use of internet search engines, only 3.8% maintained online courses or held online education certifications.

Finally, Qi et al. (2024) conducted research on data-driven systems in sports contexts provided important insights into the balance between technological implementation and human expertise. The study concluded that while data-driven systems can significantly improve operations and athlete outcomes, over-reliance on such systems risks diminishing the value of human expertise. The research emphasized that responsible integration of digital technologies must align with organizational goals and values. Through an integrated methodology examining technology adoption trends, perceived impacts, and stakeholder perspectives, the study offered evidence-based guidance for maximizing the benefits of digital transformation in sports organizations while addressing potential ethical considerations and implementation challenges.

MATERIALS & METHODS

This research employed a Research and Development (R&D) methodology to systematically develop, implement, and evaluate the disability athlete management system using the CART algorithm. R&D methodology was selected for its particular suitability in creating and validating practical products in organizational settings, making it ideal for the development of technological solutions for sports management. The design followed a structured sequential approach consisting of four primary phases: (1) preliminary study and needs analysis within NPCI Kediri to identify specific requirements; (2) system design and initial prototyping based on stakeholder input; (3) implementation and development of the CART algorithm enhanced through data augmentation techniques; (4) evaluation and refinement through iterative testing with actual user groups.

The subjects of this research comprise disability athletes registered with the NPCI Kediri. The athletes in this study represent four primary disability classifications relevant to paralympic athletics; daksa lower (lower limb impairment), daksa upper (upper limb impairment), cerebral palsy, and tuna daksa (physical/mobility impairment).

Tuna daksa (physical/mobility impairment) refers to limitations that restrict a person's ability to perform routine daily activities due to physical constraints, individuals with this condition may experience difficulties with movement, walking, limb functionality, or may lack the endurance necessary to complete everyday tasks due to physical limitations or bodily restrictions [26]. Cerebral Palsy (CP) is characterized by irregular muscle tone, abnormal posture, and movement disorders, with clinical classification based on the predominant motor symptoms, with an incidence rate of 2-3 per 1,000 live births [27]. CP is also historically referred to as Little's disease, named after William John Little who first described the condition in 1843, noting that spasticity results from brain damage during infancy, premature birth, or birth asphyxia [28].

Daksa lower (lower limb impairment) refers to physical disabilities specifically affecting the lower extremities of the body, including the legs, feet, and hip function [29]. This condition may result from congenital factors, injury, or disease, impacting an individual's mobility, balance, and ability to perform activities that require lower body strength and coordination. Daksa upper (upper limb impairment) encompasses disabilities affecting the upper body, including the arms, hands, shoulders, and related functions [30]. Individuals with daksa upper may experience limitations in fine motor skills, reaching, grasping, and other activities requiring upper body strength and dexterity, necessitating adaptive techniques in both daily life and athletic pursuits.

To effectively analyze these diverse disability profiles and their relationship to athletic potential, a comprehensive dataset was compiled from two primary sources. Primary data was collected through direct measurements and assessments of disability athletes at NPCI Kediri during the 2024 period, including anthropometric measurements, sociodemographic information, and results from standardized physical tests specific to athletics disciplines. Secondary data was gathered through documentation to obtain athletes historical records, direct observation of athlete performance processes, and discussions with coaches to validate assessment criteria. The research includes a two-part validation process that evaluates both the accuracy of the CART algorithm in correctly identifying potential athletes for athletics disciplines and the system's ease of use for NPCI Kediri staff, coaches, and administrators who will operate the platform in their daily management activities. The complete dataset consists of 72 records of disability athletes from NPCI Kediri collected throughout 2024. This dataset underwent data augmentation processing to address the limited sample size challenge in disability sports research, enhancing the robustness of the CART algorithm implementation within the management system.

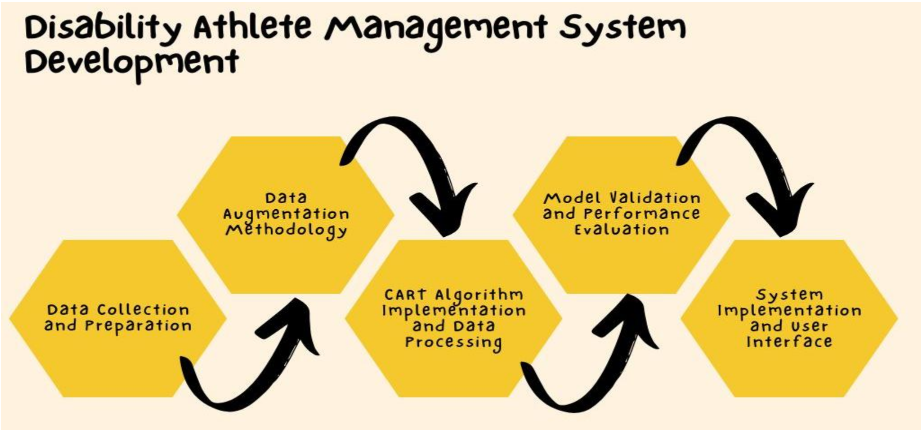

Figure 1 Disability athlete management system development

Figure 1 illustrates the five interconnected components of the research implementation process: data collection and preparation, data augmentation methodology, cart algorithm implementation, model validation and performance evaluation, and system implementation and user interface. The framework demonstrates the sequential yet iterative nature of the methodology, highlighting how each component builds upon and informs the others to create a comprehensive approach to disability athlete classification.

The research implementation process consisted of five major components that will be examined in detail in the result section. Figure 1 presents the overall methodology framework illustrating the relationship between these components:

1. Data Collection and Preparation

This component involved gathering anthropometric measurements (height, weight), demographic information (age, gender), disability classifications, and performance test results from NPCI Kediri athletes. The data was structured and pre processed to ensure compatibility with the classification algorithm, including the conversion of categorical variables through label encoding processes.

2. Data Augmentation Methodology

To address the limited sample size challenge inherent in disability sports research, standard deviation-based data augmentation techniques were applied to expand the dataset while preserving the essential characteristics of the original data. This process created synthetic data points that maintained the statistical distribution of the original dataset, providing more robust training data for the algorithm.

3. CART Algorithm Implementation and Data Processing

This phase involved the application of the CART algorithm to the augmented dataset, with particular focus on the Gini index calculations for determining optimal splitting points. The implementation specifically addressed class imbalance through weighted balance parameters to ensure fair representation of both potential and non-potential athlete classifications.

4. Model Validation and Performance Evaluation

Multiple validation techniques were employed to assess the algorithm's effectiveness, including K-fold cross-validation with varying data split ratios (90:10, 80:20, and 70:30) and confusion matrix analysis to evaluate precision, recall, and overall accuracy metrics of the classification model.

5. System Implementation and User Interface

The final component involved the development of a web-based management system integrating the validated CART algorithm, with specialized interfaces for data input, athlete assessment, results visualization, and administrative functions tailored to the specific needs of NPCI Kediri stakeholders.

Each of these methodological components was designed to address specific challenges in disability athlete management and classification, with particular attention to the unique characteristics and requirements of paralympic sports contexts. The interconnected nature of these components ensured that the final management system would effectively integrate data collection, algorithmic processing, and user-friendly interfaces to support evidence-based decision-making in identifying potential athletes with disabilities.

RESULT

Data Collection and Preparation

The research implementation began with the collection and preparation of data from disability athletes registered with NPCI Kediri. The data used in this study encompassed test results conducted on athletes with disabilities. These data included various parameters relevant to the athletes physical condition and abilities, such as 40 meter sprint test, 600 meter run test, arm hang, sit up, vertical jump, as well as additional anthropometric measurements including weight and height. Additionally, the data also involved information related to the athletes personal identities, including name, age, gender, and type of disability. As shown in Table 1, Table 2, and 3, these collected measurements and information formed the essential basis for developing an effective classification model to identify promising athletic talent.

Table 1 Performance test results of athletes with disabilities

No	40m sprint	600m run	Arm hang	Sit up	Vertical jump
1	13.88	832.8	24	59	42
2	10.83	145.75	0	34	42
3	8.38	114.52	0	55	54

Table 2 Identity and anthropometric measurements of athletes with disabilities

No	Age	Gender	Weight	Height
1	17	Laki-Laki	67.95	165.2
2	20	Perempuan	37.6	148
3	13	Laki-Laki	47.05	158

Table 3 Disability types data

No	ID	Disability Types
1	0	Daksa Lower
2	1	Daksa Upper
3	2	Cerebral Palsy (CP)
4	3	Tuna Daksa

To facilitate data analysis, the ID column in the table 3 was generated through a label encoding process, which systematically converts disability categories into numerical representations.

The three tables above present data that will be processed using the CART classification system, obtained from NPCI Kediri district for the 2024 period. In this classification process, two classes will be identified; potential athletes and non-potential athletes. Overall, the dataset to be analyzed includes 72 athletes, including data that has been enriched through the data augmentation process.

Data Augmentation Methodology

Data augmentation is a data processing technique used to increase the volume, quality, and diversity of training data through a series of transformation techniques capable of generating new information variations from limited data collections [31]. Data augmentation represents a set of algorithms that generate synthetic data, where the synthetic data contains minor changes that should not affect model predictions [32]. This method essentially addresses two main problems; first, it can generate additional data from limited data quantities, and second, it reduces the risk of overfitting in analytical models [33].

In this research, standard deviation-based data augmentation was employed to address the limited sample size of disability athletes.

The standard deviation calculation was performed using the following formula:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad \ldots(1)$$

Information:

a. S = Standard deviation
b. x = Data value
c. $\bar{x}$ = Mean value
d. n = Number of data points

This process was applied to various features, including age, gender, type of disability, height, weight, 400-meter sprint test, 600 meter run test, arm hang, sit up, and vertical jump. As an example of age feature augmentation, the following data was available:

Table 4 Age data of athlete with disabilities

No	Age
1	17
2	20
3	14
4	36
5	18
6	24
7	13
8	22
9	20

Based on this data, the standard deviation calculation using equation (1) yielded a value of 6.8211. The calculation is as follows:

Table 5 Standard deviation calculation

No.	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	17	-3,444	11.861136
2	20	-0,444	0.197136
3	14	-6,444	41.525136
4	36	15,556	241.989136
5	18	-2,444	5.973136
6	24	3,556	12.645136
7	13	-7,444	55.413136
8	22	1,556	2.421136
9	20	-0,444	0.197136
	$\sum x_i = 184$		$\sum (x_i - \bar{x})^2 = 372.222224$

$$S = \sqrt{\frac{372.222224}{8}}$$

$$S = \sqrt{46{,}527778}$$

$$S = 6{,}8211$$

This value was then added to each athlete's age data, resulting in new augmented age values, as shown in the following table:

Table 6 Results of athlete age data augmentation

No	Age	Age(rounded)
1	23,82113	24
2	26,82113	27
3	20,82113	21
4	42,82113	43
5	24,82113	25
6	30,82113	31
7	19,82113	20
8	28,82113	29
9	26,82113	27

The similar augmentation process was performed on other features to increase data diversity. The athlete data that has been enriched through augmentation would become the input for the subsequent stage of the research.

CART Algorithm Implementation and Data Processing

After completing the data augmentation process, the next step was preparing the athlete data to be used as input for the CART algorithm. This data included all the original features previously available, plus new features generated through the augmentation process.

Table 7 Athlete input data

Age	TD	H	W	40m	600m	AH	SU	VJ
17	0	165,2	67.95	13,88	832,8	24	59	42
20	1	148	37,6	10,83	145,75	0	34	42
13	3	158	47,05	8,38	114,53	0	55	54
14	2	144,8	41,15	11,5	198,95	11	35	18

The athlete data encompassed information such as age, type of disability (TD), height (H), weight (W), 40-meter sprint (40m), 600-meter run (600m), arm hang (AH), sit up (SU), and vertical jump (VJ). These comprehensive athlete profiles, structured in tabular format, served as input for the CART algorithm to determine whether an athlete falls into the "potential" or "non-potential" category.

The first stage in CART classification involved analyzing data distribution and determining the best feature for the initial split. In this case, the dataset consisted of 72 samples, with a distribution of 57 training data points and 15 testing data points. To understand the degree of homogeneity in this dataset, the level of heterogeneity needed to be calculated using the Gini index.

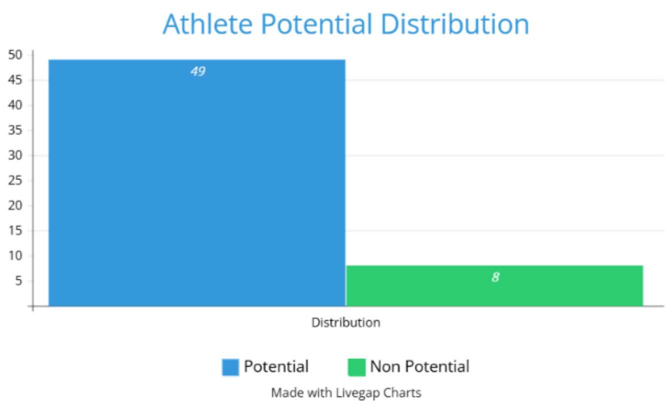

Figure 2 Athlete potential distribution

In this process, the weighted balance parameter was used to address the imbalance in class distribution, where the number of samples for the "yes" label (49) was significantly more dominant compared to the "no" label (8). This approach ensured that impurity level calculations accurately reflected the proportional contribution from each class. The value resulting from the weighted balance calculation was 28.5, obtained by calculating the average of the two classes based on the number of samples in each class. This balanced representation ensured that the algorithm would give appropriate consideration to both potential and non-potential athlete categories despite their uneven distribution in the training data. The weighted balance formula is shown below:

$$Weighted\ Balance\ =\ \sum_{i=1}^{n}\left(\frac{N_i}{N_{Total}} *\ W_i\right)\ \ldots(2)$$

Where n_1 refers to the number of samples in class i, which is the amount of data in a particular class. n represents the total number of samples from all existing classes. While w_1 is the value or weight assigned to class i. To calculate the weighted balance, the number of samples and weights can be substituted into the following equation:

$$Weighted\ Balance = \frac{49}{49+8} *\ 49 + \frac{8}{49+8} *\ 8$$
$$Weighted\ Balance = 28.5$$

The analysis continued by calculating the probability of each class. These probabilities form the basis for determining the extent to which data is evenly distributed between classes. Based on the previous calculations, the probabilities for the "yes" and "no" classes are as follows:

$$P_{yes} = \frac{28.5}{57} = 0{,}5\ and\ P_{no} = \frac{28.5}{57} = 0{,}5$$

Subsequently, these probabilities were substituted into the Gini index formula to measure the impurity level of the dataset. The Gini index is calculated using the following equation:

$$GINI\ =\ 1\ -\ \sum_{i\,=\,1}^{n} Pi^2\ \ldots(3)$$

For binary classification problem involving potential athletes "yes" and non-potential athletes "no" with balanced class weights, the Gini index calculation is applied as follows:

$$GINI = 1\ -\ \left(P_{yes^2} + P_{no^2}\right)$$
$$GINI = 1\ -\ \left((0.5)^2 + (0.5)^2\right)$$
$$GINI = 1 - (0.25 + 0.25)$$
$$GINI =\ 1\ -\ 0.5$$
$$GINI\ =\ 0.5$$

The Gini index ranges from 0 to 1, where 0 signifies perfect equality or homogeneity within a dataset, and 1 indicates complete inequality or maximum heterogeneity [34]. This value describes the extent to which the class distribution in the dataset is diverse. With a Gini value of 0.5, it can be concluded that the dataset has a moderate level of impurity. This value represents the theoretical maximum impurity in binary classification scenarios, indicating equal probability distribution between the "potential" and "non-potential" classes prior to any node splitting operations in the CART algorithm.

CART algorithm's next crucial step involved dividing the dataset based on the most optimal feature to reduce impurity. In this case, the **disability type** feature was selected as the primary splitting feature, denoted as x [1]. This feature was divided based on values converted to labels 0 through 3, with the following class distribution:

a. Label 0: Daksa Lower (16 samples with "yes" class, 0 samples with "no" class)
b. Label 1: Daksa Upper (20 samples with "yes" class, 0 samples with "no" class)
c. Label 2: Cerebral Palsy (13 samples with "yes" class, 0 samples with "no" class)
d. Label 3: Tuna Daksa (0 samples with "yes" class, 8 samples with "no" class)

To determine the optimal dataset division, experiments were conducted using a threshold of x [1] ≤ 2.5. This threshold was chosen because it systematically separated disability types with "yes" class distribution (labels 0, 1, and 2) from disability types with "no" class distribution (label 3). The division at this threshold resulted in two branches: the left branch with 49 samples (labels 0, 1, and 2) all having the "yes" class, and the right branch with 8 samples (label 3) all having the "no" class. Both branches had pure Gini values (0.0), indicating that the division produced completely homogeneous groups, as shown in figure 3 below.

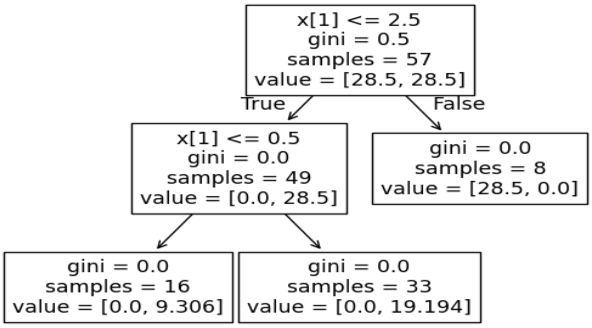

Figure 3 Decision tree diagram result

Further division was performed on the left branch using the threshold x[1] ≤ 0.5. This threshold was used to separate label 0 (16 samples) from labels 1 and 2 (33 samples). This division resulted in two homogeneous branches: the first consisting of 16 samples with the "yes" class (label 0) and the second consisting of 33 samples with the "yes" class (labels 1 and 2). These branches also had pure Gini values (0.0), demonstrating perfect homogeneity at each terminal node. This indicates that further division based on x[1] ≤ 0.5 successfully grouped the data optimally according to their labels.

Model Validation and Performance Evaluation

Data testing represents the validation phase for assessing the accuracy and effectiveness of the CART algorithm in classifying potential disability athletes at NPCI Kediri. This process was conducted by analyzing the algorithm's performance on an athlete dataset that included anthropometric parameters (height, weight), sociodemographic information, and physical test results (40-meter sprint, 600-meter run, arm hang, sit up, vertical jump) associated with each athlete's type of disability. To ensure the validity of test results and avoid bias in model evaluation, this research implemented the K-fold cross validation method.

The principle of K-fold cross validation involves dividing the dataset into K equal subsets, where each subset alternately functions as validation data to evaluate the model trained on the other subsets [35]. In this research, a 10-fold cross validation approach was implemented, where the dataset was divided into ten equal parts, providing a robust evaluation framework that balances computational efficiency with comprehensive performance assessment. Additionally, various data split ratios were tested to evaluate model stability and generalization capabilities. A data split ratio refers to the proportion of data allocated to training versus testing; for example, a 90:10 split indicates that 90% of the dataset is used for model training while the remaining 10% is reserved for performance evaluation. The implementation of K-fold cross validation on the NPCI Kediri disability athlete dataset yielded the following results:

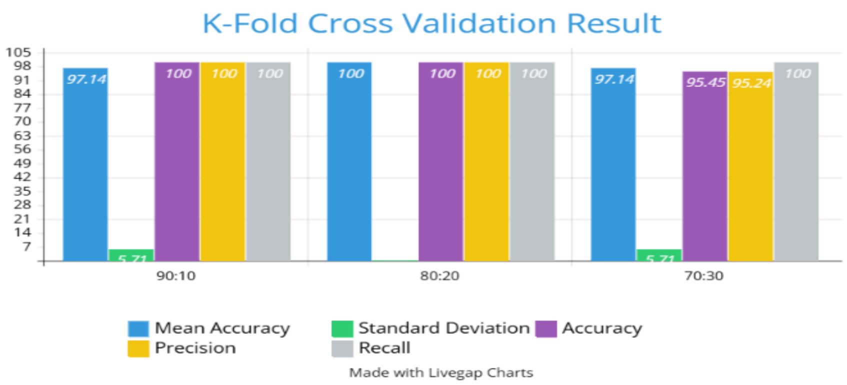

Figure 4 K-fold cross validation result

In the implementation of K-fold cross validation with a 90:10 data split, the CART model demonstrated excellent performance with slight variations between folds. Of the 10 folds tested, eight folds achieved perfect 100% accuracy, while two folds (the 3rd and 8th folds) showed slightly lower accuracy at 85.71%. The average cross validation accuracy reached 97.14% with a standard deviation of 5.71%, indicating relatively high consistency in performance despite minor fluctuations. After the model was trained using 90% of the complete dataset, evaluation on 10% testing data yielded perfect accuracy, precision, and recall (100%).

The K-fold cross validation testing with an 80:20 data split produced remarkably consistent performance. All 10 folds achieved perfect 100% accuracy, resulting in an average cross validation accuracy of 100% with a standard deviation of 0%. The absence of variation between folds demonstrates that the CART algorithm was able to extract classification patterns excellently and consistently from various training data subsets. The evaluation results on the final testing data were also perfect, with accuracy, precision, and recall reaching 100%.

In the testing scenario with a 70:30 data split, a similar pattern to the 90:10 split was observed, where eight folds achieved 100% accuracy and two folds (the 3rd and 8th folds) showed 85.71% accuracy. The average cross validation accuracy was 97.14% with a standard deviation of 5.71%. However, unlike the 90:10 data split, evaluation on the final testing data showed a slight decrease in performance with 95.45% accuracy and 95.24% precision, although recall remained perfect at 100%. This decline was attributed to the more limited training data (70% of the dataset), which caused the model to not fully capture some complex patterns in the data.

To further validate this optimal model configuration, detailed performance analysis was conducted using more rigorous metrics. The classification performance evaluation of the NPCI Kediri disability athlete management system was conducted using a confusion matrix. The effectiveness of a classification algorithm can be analyzed, graphically visualized, and summarized through a table structure known as a confusion matrix [36]. This evaluation technique maps the classification distribution into four categories: True Positive (TP) – potential athletes correctly predicted as potential; True Negative (TN) – non-potential athletes correctly predicted as non-potential; False Positive (FP) – non-potential athletes incorrectly predicted as potential; and False Negative (FN) – potential athletes incorrectly predicted as non-potential. From the model testing with an 80:20 data split, a confusion matrix was obtained with 13 TP data points, 2 TN data points, and 0 data points for both FP and FN, demonstrating perfect classification without prediction errors. The confusion matrix visualization is shown in the figure 5 below:

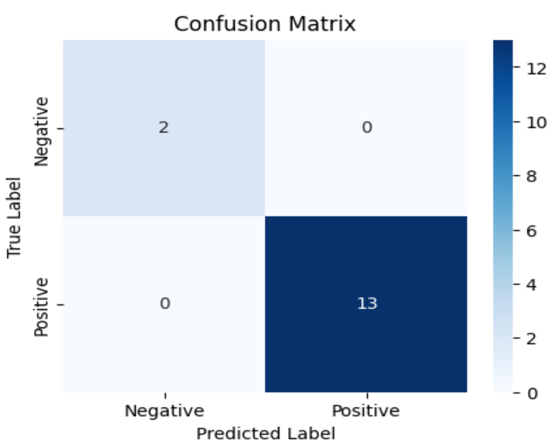

Figure 5 Confusion matrix CART algorithm

Based on these confusion matrix results; various model performance evaluation metrics were calculated. The 100% accuracy indicates the proportion of total correct predictions from all cases. The 100% precision indicates that all athletes predicted as potential were indeed potential (no false positives). The 100% recall demonstrates the model's ability to identify all potential athletes without missing any (no false negatives). Despite the class imbalance in the testing dataset, with positive cases (potential athletes) being more dominant than negative cases, the model was still able to perfectly classify both classes. These optimal evaluation results confirm the reliability of the CART model in the NPCI Kediri system for classifying the potential of disability athletes based on anthropometric parameters, physical test results, and disability types.

System Implementation and User Interface

Following the successful validation of the CART algorithm's performance, the research proceeded with the implementation of a comprehensive web-based disability athlete management system for NPCI Kediri. This section presents the actual implementation of the system and its user interface components. The system was developed as a web application with responsive design principles to ensure accessibility across various devices. The user interface was specifically designed to accommodate the needs of NPCI Kediri administrators, coaches, and assessment personnel. The following figures illustrate the key interfaces of the implemented system, showcasing how the theoretical model has been transformed into a practical and user-friendly application for disability athlete management.

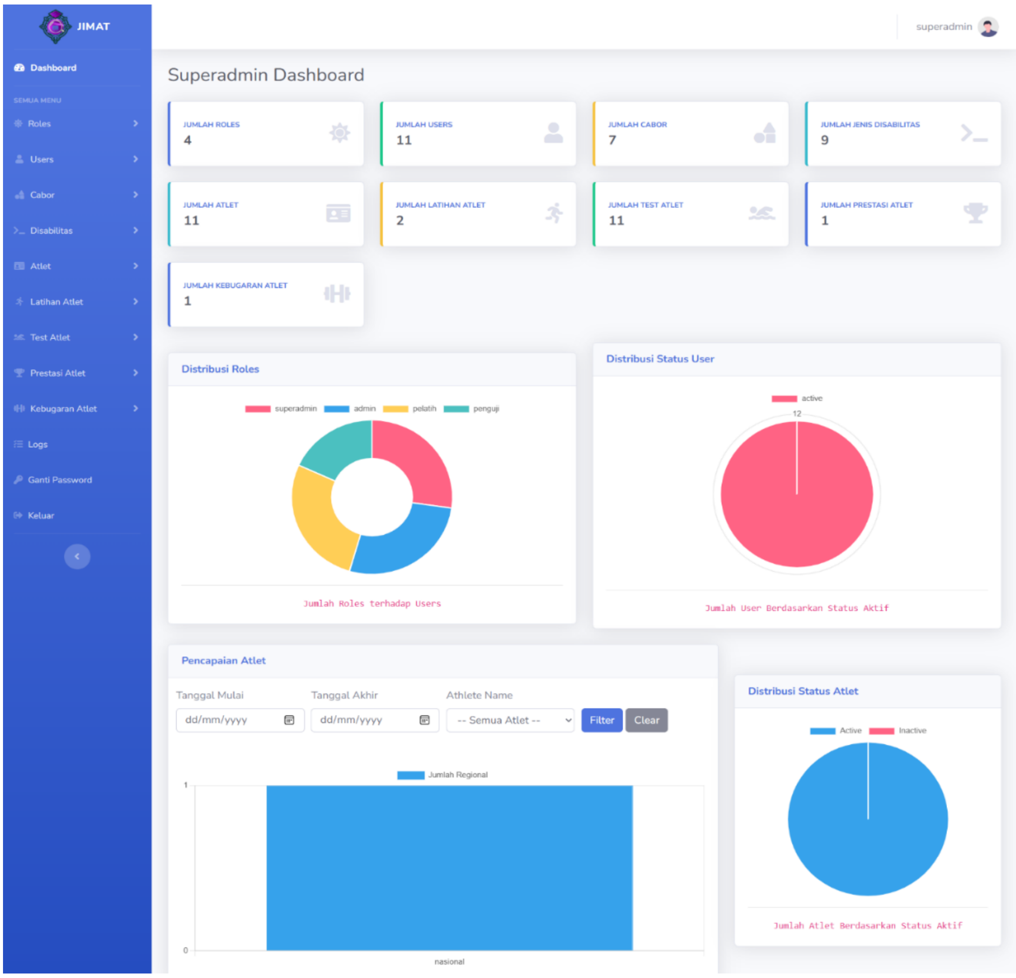

Figure 6 Main dashboard athlete management system

Figure 6 presents the main dashboard of the system, providing a visual overview of critical information and real-time analytics. This dashboard interface displays comprehensive metrics including total registered users, athlete counts categorized by disability type, distribution across various athletics disciplines, and recently entered assessment data. The dashboard implements intuitive data visualization components such as charts and summary cards that enable administrators to quickly identify trends, monitor classification outcomes, and track system usage patterns, facilitating evidence-based decision-making for NPCI Kediri management.

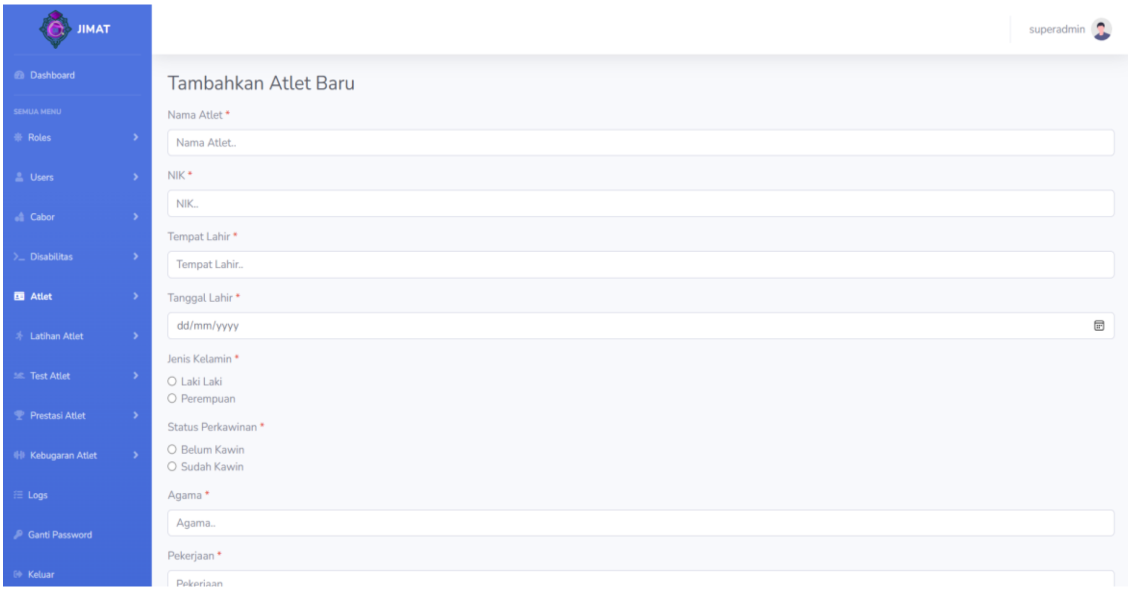

Figure 7 Athlete demographic data entry form for personal information collection

Figure 7 displays the athlete demographic data entry form. This page contains structured input fields specifically designed to capture athletes' personal information and demographic data, including fields for personal identification details. The form also includes functionality for document uploads to validate the athlete's identity information.

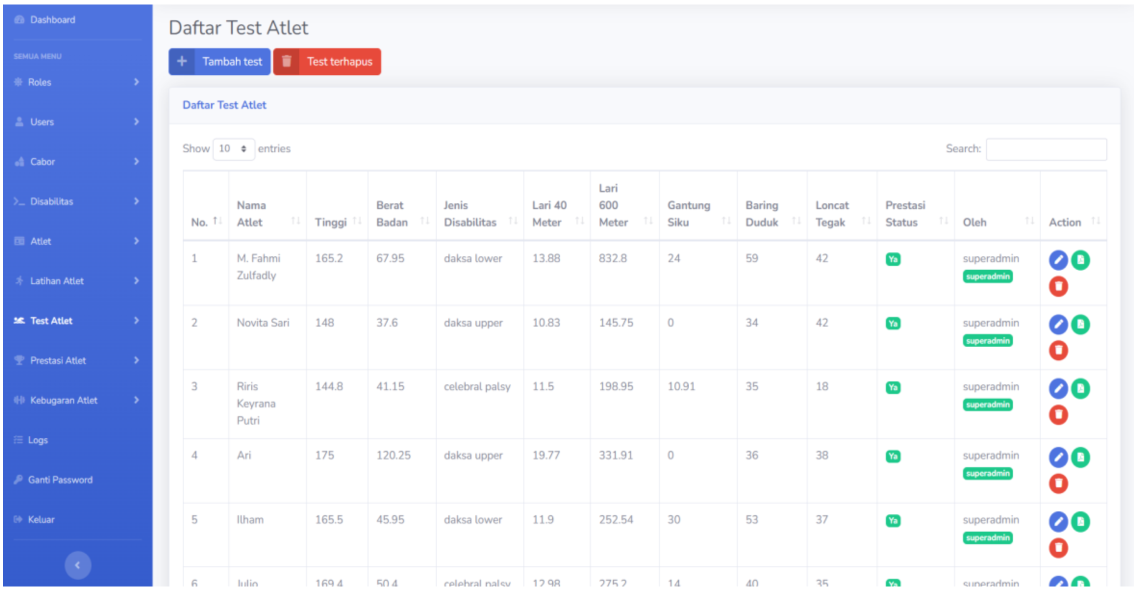

Figure 8 Athlete assessment and classification with potential prediction results

Figure 8 illustrates the athlete assessment and classification page. This interface contains input fields for anthropometric measurements and disability type classification, which serve as the primary parameters for the CART algorithm. After data entry is complete, the page displays the prediction results, clearly indicating whether an athlete is classified as "potential" or "non-potential" based on the algorithm's analysis of the input parameters. As shown in Figure 8, the system presents classification results using "prestasi status: ya" to indicate potential athletes and "prestasi status: tidak" to designate non-potential athletes.

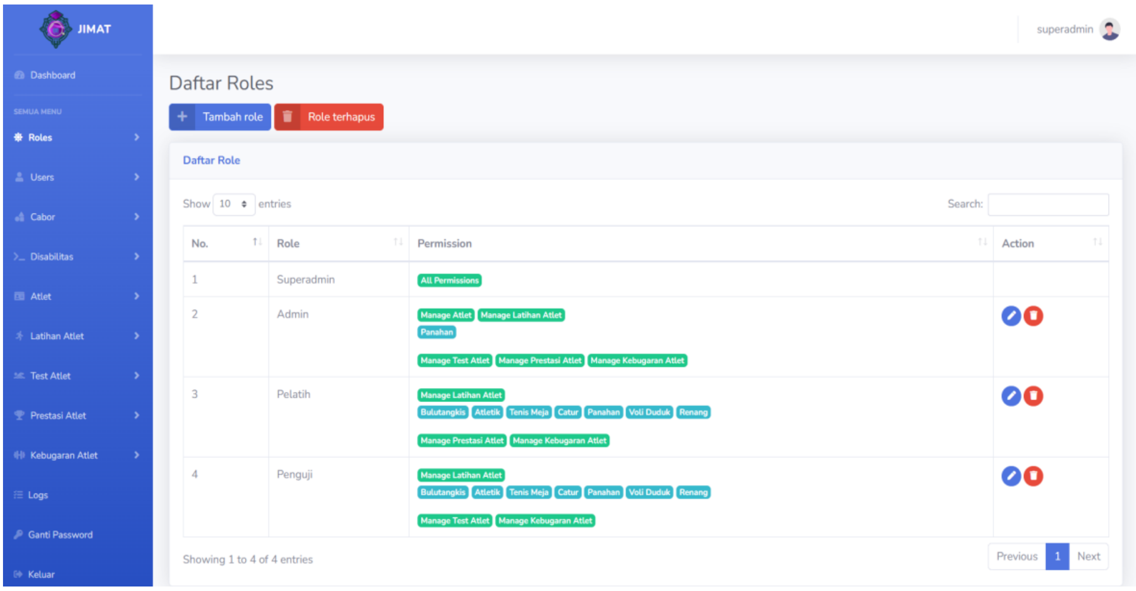

Figure 9 Administrative interface for user account and system settings management

Figure 9 shows the administrative interface for managing user accounts and system settings. This component allows NPCI Kediri administrators to control access permissions, customize assessment parameters, and configure system behavior according to organizational requirements. Figure 9 shows the administrative interface for managing user accounts and system settings. This component allows NPCI Kediri administrators to control access permissions, customize assessment parameters, and configure system behavior according to organizational requirements.

The implementation of this web-based system represents the practical application of the theoretical model described in previous sections. By transforming the CART algorithm into an accessible, user-friendly interface, the system enables NPCI Kediri staff to efficiently manage athlete data and objectively identify potential talent among disability athletes without requiring extensive technical knowledge of the underlying classification algorithms.

DISCUSSION

The results of this study demonstrate the effectiveness of implementing the CART algorithm within a digitalized management system for identifying potential athletes with disabilities at NPCI Kediri. The exceptional classification performance—achieving 100% accuracy, precision, and recall with the 80:20 data split—significantly exceeds the performance metrics reported in comparable studies. For instance, Wahyu Romadhonia et al. (2023) reported accuracy rates exceeding 80% in their application of CART models for athlete talent identification [14], while the present implementation achieved perfect classification in the optimal configuration.

The exceptional performance of the developed system can be attributed to several key methodological innovations. First, the application of standard deviation-based data augmentation techniques effectively addressed the limited sample size challenge that typically constrains disability sports research. Data augmentation provided a data-space solution to the limited dataset challenge by encompassing various techniques that enhanced both the size and quality of the training data, enabling the development of more robust classification models while maintaining the integrity of the original patterns [37]. This approach aligns with emerging practices in machine learning, where appropriate data augmentation not only generates additional data from limited quantities but also reduces overfitting risks [33].

Second, the weighted balance parameter implementation successfully mitigated the class imbalance issue, where the "potential" class (49 samples) significantly outnumbered the "non-potential" class (8 samples). This methodological approach employed Gini impurity as the splitting criterion for decision trees and strategically adjusted the weights for the two different classes of the imbalanced data to make the learning algorithm to properly account for the uneven class distribution [38].

From a practical implementation perspective, the development of a web-based management system incorporated specialized interfaces for different user roles. This design approach addresses the digital transformation needs that emphasize responsible integration of digital technologies aligned with organizational goals and values [39]. The system's user interface design specifically considered the documented gaps in digital literacy among sports professionals by creating intuitive interfaces that accommodate varying levels of technological proficiency.

However, several limitations should be acknowledged. The relatively small sample size (72 records), even after augmentation, may limit the generalizability of the findings to larger and more diverse paralympic populations. Furthermore, the dataset is restricted to athletes registered with NPCI Kediri, which represents only one regional branch of the National Paralympic Committee of Indonesia. Additionally, while the model achieved perfect classification results with the 80:20 data split, these results are specific to the current dataset and context. Future validation with expanded and diverse datasets would further strengthen the model's applicability across varied paralympic settings. Research suggests that more comprehensive datasets enable advanced modeling techniques that can identify complex and nuanced patterns not detectable in limited samples which could eventually lead to more robust classification systems [40].

The system's implementation at NPCI Kediri addresses the data management inefficiencies and subjective assessment biases identified in the initial problem statement. By transforming manual data collection processes into a digitalized system and replacing subjective talent identification with an evidence-based algorithm, the project directly responds to the challenges outlined by Visuri et al. (2017) regarding administrative burden [3] and by Höner et al. (2021) concerning the low accuracy of subjective talent decisions [6]. The integration of a centralized data management platform with an enhanced DT classification model creates a comprehensive solution that could serve as a model for other paralympic organizations facing similar challenges.

CONCLUSION

This research has successfully developed and implemented a digital management system for disability athletes at NPCI Kediri that integrates the CART algorithm to identify potential athletes. The system addresses two critical challenges faced by disability sports organizations: inefficient manual data management and subjective assessment in athlete selection processes.

The implementation of the CART algorithm, enhanced through standard deviation-based data augmentation techniques, demonstrated exceptional performance in classifying potential athletes. With optimal data splitting (80:20 ratio), the model achieved perfect accuracy, precision, and recall (100%), as confirmed through rigorous K-fold cross validation and confusion matrix analysis. The algorithm effectively identified disability type as the primary discriminating feature, creating a classification model that produced homogeneous groupings with pure Gini values (0.0) at terminal nodes.

From a practical standpoint, this research has significant implications for NPCI Kediri and potentially other paralympic organizations. The digitalized system transforms data management processes, reducing administrative burden while increasing accuracy and consistency in talent identification. By establishing an objective, evidence-based framework for assessing athletic potential among individuals with disabilities, the system helps overcome traditional barriers in disability sports talent recognition, potentially expanding opportunities for athletes who might otherwise be overlooked through subjective assessment methods.

Future research should focus on extending this model to other paralympic sports beyond athletics, exploring additional anthropometric and performance parameters specific to different disability categories, and investigating the long-term predictive validity of the classification system through longitudinal studies. Additionally, future development initiatives should be directed toward expanding the dataset by collecting more comprehensive data from a larger and more diverse population of paralympic athletes across various regions in Indonesia. A more extensive dataset would improve the model's robustness, reduce the dependency on data augmentation techniques, and potentially reveal more nuanced patterns specific to different disability classifications, ultimately enhancing the precision and generalizability of the talent identification system.

Declaration by Authors

Acknowledgement: None

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Haris M Al, Doewes M, Liskustyawati H. Development of Boccia Cerebral Palsy's National Athlete Achievement in the Indonesian National Paralympic Committee. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal 2020; 3:784–94. https://doi.org/10.33258/birle.v3i2.940.
2. Kardiyanto DW, Setijono H, Mintarto E. The evaluation of indonesia paralympic coaching. European Journal of Special Education Research 2017. https://doi.org/10.46827/ejse.v0i0.1032.
3. Visuri A, van Berkel N, Luo C, Goncalves J, Ferreira D, Kostakos V. Predicting interruptibility for manual data collection. Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, New York, NY, USA: ACM; 2017, p. 1–14. https://doi.org/10.1145/3098279.3098532.
4. Fatkhul Imron, Joko Sulistyono, Moh Erkamim, Zandra Dwanita Widodo. Sportif.App as an Effort to Increase Sports Human Resources Readiness and Digitalization of Sports Data Information Management. International Journal of Economics and Management Research 2023; 2:287–95. https://doi.org/10.55606/ijemr.v3i1.223.
5. Urbach N, Ahlemann F, Böhmann T, Drews P, Brenner W, Schaudel F, et al. The Impact of Digitalization on the IT Department. Business & Information Systems Engineering 2019; 61:123–31. https://doi.org/10.1007/s12599-018-0570-0.
6. Höner O, Murr D, Larkin P, Schreiner R, Leyhr D. Nationwide Subjective and Objective Assessments of Potential Talent Predictors in Elite Youth Soccer: An Investigation of Prognostic Validity in a Prospective Study. Front Sports Act Living 2021; 3:638227. https://doi.org/10.3389/fspor.2021.638227.
7. Kiuppis F. Inclusion in sport: disability and participation. Sport Soc 2018; 21:4–21. https://doi.org/10.1080/17430437.2016.1225882.
8. Kamberidou I, Bonias A, Patsantaras N. Sport as a means of inclusion and integration

for" those of us with disabilities". European Journal of Physical Education and Sport Science 2019. https://doi.org/10.5281/zenodo.3464696.
9. Coutts AJ, Crowcroft S, Kempton T. Developing athlete monitoring systems. Recovery and Well-being in Sport and Exercise, London: Routledge; 2021, p. 17–31. https://doi.org/10.4324/9781003258117-3.
10. Windt J, MacDonald K, Taylor D, Zumbo BD, Sporer BC, Martin DT. "To Tech or Not to Tech?" A Critical Decision-Making Framework for Implementing Technology in Sport. J Athl Train 2020; 55:902–10. https://doi.org/10.4085/1062-6050-0540.19.
11. Berente N, Gu B, Recker J, Santhanam R. Managing artificial intelligence. MIS Quarterly 2021;45. https://doi.org/10.25300/MISQ/2021/16274.
12. Baird A, Maruping LM. The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. MIS Quarterly 2021; 45:315–41. https://doi.org/10.25300/MISQ/2021/15882.
13. Javaid M, Haleem A, Singh RP, Suman R. Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study. Journal of Industrial Integration and Management 2022; 07:83–111. https://doi.org/10.1142/S2424862221300040.
14. Romadhonia RW, Sofro A, Ariyanto D, Maulana DA, Prihanto JB. Application of Decision Trees in Athlete Selection: A Cart Algorithm Approach. INTI Journal 2023; JODS:2805–5160. https://doi.org/10.61453/jods.v2023no14.
15. Sarabia JM, Roldan A, Henríquez M, Reina R. Using Decision Trees to Support Classifiers' Decision-Making about Activity Limitation of Cerebral Palsy Footballers. Int J Environ Res Public Health 2021; 18:4320. https://doi.org/10.3390/ijerph18084320.
16. Priyanka NA, Kumar D. Decision tree classifier: a detailed survey. International Journal of Information and Decision Sciences 2020; 12:246. https://doi.org/10.1504/IJIDS.2020.108141.
17. Charbuty B, Abdulazeez A. Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends 2021; 2:20–8. https://doi.org/10.38094/jastt20165.
18. Daniya T, Geetha M, Suresh Kumar K. CLASSIFICATION AND REGRESSION TREES WITH GINI INDEX. Advances in Mathematics: Scientific Journal 2020; 9:8237–47. https://doi.org/10.37418/amsj.9.10.53.
19. Günlük O, Kalagnanam J, Li M, Menickelly M, Scheinberg K. Optimal decision trees for categorical data via integer programming. Journal of Global Optimization 2021; 81:233–60. https://doi.org/10.1007/s10898-021-01009-y.
20. Pappalardo G, Cafiso S, Di Graziano A, Severino A. Decision Tree Method to Analyze the Performance of Lane Support Systems. Sustainability 2021; 13:846. https://doi.org/10.3390/su13020846.
21. Bishnoi S, Hooda BK. Decision Tree Algorithms and their Applicability in Agriculture for Classification. Journal of Experimental Agriculture International 2022; 44:20–7. https://doi.org/10.9734/jeai/2022/v44i730833.
22. Tangirala S. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm. International Journal of Advanced Computer Science and Applications 2020; 11:612–9. https://doi.org/10.14569/IJACSA.2020.0110277.
23. Newman-Griffis D, Rauchberg JS, Alharbi R, Hickman L, Hochheiser H. Definition drives design: Disability models and mechanisms of bias in AI technologies 2022. https://doi.org/https://doi.org/10.48550/arXiv.2206.08287.
24. Wald M. AI Data-Driven Personalisation and Disability Inclusion. Front Artif Intell 2021; 3:571955. https://doi.org/10.3389/frai.2020.571955.
25. Wangiyana S, Samczyński P, Gromek A. Data Augmentation for Building Footprint Segmentation in SAR Images: An Empirical Study. Remote Sens (Basel) 2022; 14:2012. https://doi.org/10.3390/rs14092012.
26. Murtaza H, Kausar N, Mazhar M. SELF-ESTEEM AND QUALITY OF LIFE AMONG STUDENTS WITH AND WITHOUT PHYSICAL DISABILITY. Pakistan Journal of Social Research 2022; 04:267–73. https://doi.org/10.52567/pjsr.v4i03.711.
27. Patel DR, Neelakantan M, Pandher K, Merrick J. Cerebral palsy in children: a

clinical overview. Transl Pediatr 2020;9: S125–35. https://doi.org/10.21037/tp.2020.01.01.
28. Paul S, Nahar A, Bhagawati M, Kunwar AJ. A Review on Recent Advances of Cerebral Palsy. Oxid Med Cell Longev 2022; 2022:2622310. https://doi.org/https://doi.org/10.1155/2022/2622310.
29. Stolt M, Laitinen A, Ruutiainen J, Leino-Kilpi H. Research on lower extremity health in patients with multiple sclerosis: a systematic scoping review. J Foot Ankle Res 2020; 13:54. https://doi.org/10.1186/s13047-020-00423-x.
30. Valè N, Gandolfi M, Mazzoleni S, Battini E, Dimitrova EK, Gajofatto A, et al. Characterization of Upper Limb Impairments at Body Function, Activity, and Participation in Persons with Multiple Sclerosis by Behavioral and EMG Assessment: A Cross-Sectional Study. Front Neurol 2020; 10:1395. https://doi.org/10.3389/fneur.2019.01395.
31. Mumuni A, Mumuni F. Data augmentation: A comprehensive survey of modern approaches. Array 2022; 16:100258. https://doi.org/10.1016/j.array.2022.100258.
32. Shorten C, Khoshgoftaar TM, Furht B. Text Data Augmentation for Deep Learning. J Big Data 2021; 8:101. https://doi.org/10.1186/s40537-021-00492-0.
33. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings 2022; 3:91–9. https://doi.org/10.1016/j.gltp.2022.04.020.
34. Lian M, Shi P, Zhang L, Yao W, Gielis J, Niklas KJ. A generalized performance equation and its application in measuring the Gini index of leaf size inequality. Trees 2023; 37:1555–65. https://doi.org/10.1007/s00468-023-02448-8.
35. Zhang X, Liu C-A. Model averaging prediction by K-fold cross-validation. J Econom 2023; 235:280–301. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.04.007.
36. Nurul Ehsan Ramli, Zainor Ridzuan Yahya, Nor Azinee Said. Confusion Matrix as Performance Measure for Corner Detectors. Journal of Advanced Research in Applied Sciences and Engineering Technology 2022; 29:256–65. https://doi.org/10.37934/araset.29.1.256265.
37. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data 2019; 6:60. https://doi.org/10.1186/s40537-019-0197-0.
38. Disha RA, Waheed S. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity 2022; 5:1. https://doi.org/10.1186/s42400-021-00103-8.
39. Qi Y, Sajadi SM, Baghaei S, Rezaei R, Li W. Digital technologies in sports: Opportunities, challenges, and strategies for safeguarding athlete wellbeing and competitive integrity in the digital era. Technol Soc 2024; 77:102496. https://doi.org/https://doi.org/10.1016/j.techsoc.2024.102496.
40. Mujahid M, Rehman A, Alam T, Alamri FS, Fati SM, Saba T. An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. Diagnostics 2023;13. https://doi.org/10.3390/diagnostics13152489.

How to cite this article: Ar Rasyid Sarifullah Gilbijatno, Resty Wulanningrum, Siti Rochana. Digitalization of disability athlete management system using decision tree cart algorithm to identify potential athletes in NPCI Kediri. *International Journal of Research and Review*. 2025; 12(5): 442-460.
DOI: *https://doi.org/10.52403/ijrr.20250545*

******