March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

Classification Algorithm Based on Machine Learning to Optimize Athletes Talent Detection

Bryam Soto[1], Hugo Vega[2], Yudi Guzmán[3], Ciro Rodriguez[4], Daniel Quinto [5]

[1,2,3,4, 5] NationalUniversity Mayor de San Marcos

Article Info
Volume 83
Page Number: 13464 - 13471
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 20 April 2020*

Abstract

The objective of this research is to automate the process of detecting youth athletic talent using a classification algorithm based on Machine Learning in a mobile application. The research addresses the problem of detecting talented athletes of young people aged between 6 and 18, which is deficient due to the complexity and amount of information that is processed to classify an athlete in the Academy in an accurate and timely manner in the corresponding discipline. “La Academia del IPD” is the massive support program in sport of the Peruvian Sports Institute, which seeks to promote sports activity and the training of sports talents.

Keywords; *Classification algorithm, Machine Learning, mobile application, automation, talent.*

I. INTRODUCTION

According to [1], the Peruvian Sports Institute (IPD) is the governing body of the National Sports System and, in conjunction with various organizations, responsible for formulating and executing sports, recreational and physical education policies. Its mission is to promote the development of competitive and recreational sports as a tool for social change, which contributes to improving the quality of life of Peruvian society.

That is why the National Direction of Recreation and Sport Promotion (DNRPD), internal body in charge of promoting the practice of physical, recreational and sports activity at the national wide through its programs aimed atmassive support and sports initiation, has been developing since 2018 the La Academia sports program, which consists of providing sports initiation workshops to the Peruvian population from 6 to 18 years old through a sports model that allows them to develop necessary skills and abilities, to detect sports talents and their training to discharge competition.

The program is developed at the national wide in all regions of Peru and is directly managed by the National Directorate for Recreation and Sports Promotion (DNRPD). As shown in table 1, there are three main processes to allow the development of the program, starting with themassive support in sport through which sports workshops are held in all disciplines that will enable the sports initiation of all the beneficiaries of the program, continuing with the detection of sports talents, a process in which the highest performing athletes are evaluated to pass them to the next process, which consists of sports training and consolidation of the athlete for high competition.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

Table 1: Sports program processes of La Academia

Get in of children from 8 to 14 years		Sports Training Process		Output: DP/DC
Starting	**Recruitment**	**Training**	**Consolidation**	**Projection Athlete**
Three months of induction according to profile and sports preferences	Physical tests, a battery of physical tests, abilities, skills and anthropometric tests	Comprehensive basic training: TECHNICAL PHYSICAL PSYCHOLOGICAL NUTRITIONAL SOCIAL 70%	General preparation with projection towards high performance 30%	DINADAF FDN
Sports initiation	Basic tests of physical capabilities, abilities, skills and anthropometric tests	Basic technical principles	Complete technical elements	

Currently, our web system allows the comprehensive management and digitization of its processes, including the virtual registration of all beneficiaries of the program, the management of sports workshops, the traceability of information on athletes, detection of talents, sports and its subsequent publication in a mobile application for Android & iOS devices. [2]

As detailed in figure 1, the evaluation of an athlete is divided into 5 levels.

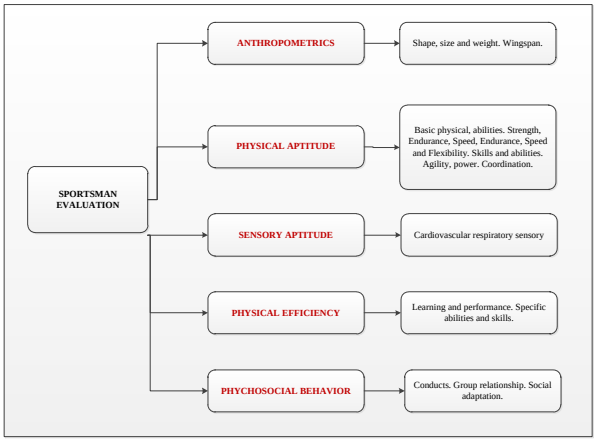


Figure 1: Scopes for the evaluation of the athlete

However, the talent detection process is deficient, the process currently begins when the teacher of a sports workshop identifies a potential athlete, registers it in the system and waits for a specialist to evaluate it by taking measurements in the field, This information is then uploaded to the system so that the program coordinators can evaluate the information and decide through their experience if the beneficiary can be qualified as sports talent. [3]

Theevaluator problem when interpreting data is that the evaluator may err when it comes to qualifying a talent since he or she must evaluate all the athletes identified as potential talents at the national wide and in all sports disciplines, which it involves processing a large amount of information and having extensive experience in all sports disciplines.

II. PROBLEM STATEMENT

2.1. Problem Description

Currently, the analysis for sports talents detection is performed based on the experience of the sports analyst, it is an estimate calculated by comparing the beneficiary's physical capacity indicators with qualified athlete standards in the discipline being evaluated. This type of analysis allows to identify potential talents, but they have an error bias, since the evaluator may err in a data, make a bad calculation or confuse the interpretation of the athlete's indicators, due to the fact that they have to do this work for all athletes at the national wide and from all sports disciplines, which involves processing a large amount of information. The main problem identified is the imperfect talent detection process reflected in the annual publication of sports

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

talents through the mobile application of the La Academia.

2.2. Problem Basis

When evaluating athletes for talent, a sports analyst conducts an empirical assessment of the athlete's physical fitness indicators. This work must be carried out for all the beneficiaries who were registered as potential talents at the national wide and in all disciplines, making this work difficult due to the amount of information and the complexity of interpretation, causing an imperfect talent detection process reflected in the Annual publication of sports talents through the mobile application La Academia IPD. As a result, the time it takes for the sports talent screening process is excessive, increasing the workload for sports analysts and giving inaccurate results on sports talent qualification.

Variable 1: Total time it takes for the sports talent classification process = 60 days

Variable 2: Error rate in the qualification of sports talents = 12.5%

2.2. Research objective

Develop a classification algorithm based on Machine Learning that automates the analysis of sports information and talent classification to optimize the talent detection process reflected in the annual publication of sports talents through the La Academia IPD mobile application. Reducing the total time it takes for the sports analyst to qualify talents, in addition to reducing his workload and the number of erroneous results, generally improving the performance of the process.

Variable 1: Total time it takes for the sports talent classification process = 1 day

Variable 2: Error rate in the qualification of sports talents = 1%

Specific objectives

• SO1: Determine the Machine Learning algorithm that allows analyzing and classifying sports talents

• SO2: Automate the analysis of information on the physical abilities of athletes classified as potential talents

• SO3: Automate the classification of talents based on a comparative analysis with qualified athletes and the system data.

III. THEORETICAL AND CONCEPTUAL FRAMEWORK

3.1. Machine learning

The author [4] gives us an overview of the concept of Machine Learning as an extensive set of tools to understand the data. These tools can be classified as supervised or unsupervised. Generally speaking, supervised statistical learning involves building a statistical model to predict or estimate output based on one or more inputs. Problems of this nature occur in fields as diverse as business, medicine, astrophysics, and public policy. With unsupervised statistical learning, there are inputs, but no outputs of supervision; however, we can learn the relationships and structure from these data.

According to [5], as we can see in figure 2, there are several applications to use Machine Learning.

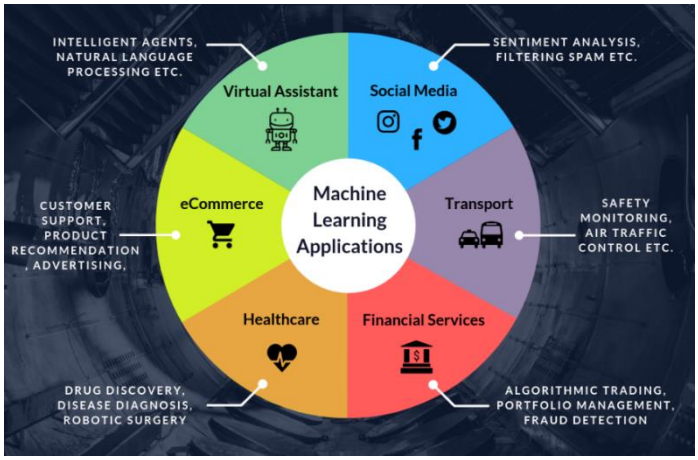


Figure 2: Fields Machine learning applications

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

Types of Machine Learning

According to [6], as we can see in table 2, we can divide Machine Learning concepts into the following areas:

- Supervised learning
- Unsupervised learning
- Reinforced learning

Table 2: Types of machine learning

Types of machine learning			
	Supervised	**Unsupervised**	**Reinforcement**
Data types	Data with objective	Data without objective	Data without objective
Functions	Classification Prediction	Segmentation Feature extraction	Robotic Games Navigation
Algorithms	K nearest neighbors Artificial neural networks Support vector machine Bayesian networks Decision Tree Logistic Regression Linear Regression	K-medias Gaussian Mix Hierarchical Group Self-organizing maps	Dynamic programing Q-learning SARSA

In figure 3, we can see more details about three types of machine learning, which we will describe later in more detail [7]

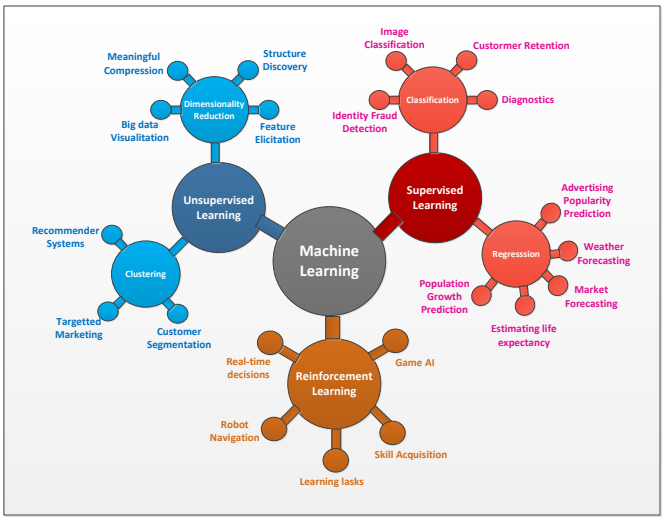


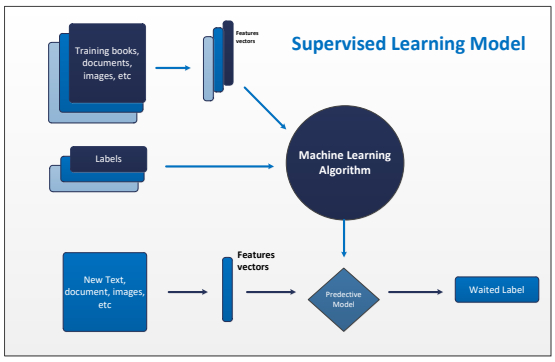
Figure 3. Types of Machine Learning

Supervised Learning

According to [8], Supervised Learning allows predictions to be made in the future based on characteristics or behaviors that have already been recorded in the historical data (historical label). Supervised learning searches for patterns in historical data linking all fields to a target field. An example is when emails are tagged as "spam" or "legitimate" by users. The prediction process begins with an analysis of what characteristics have the emails already marked with both labels. See figure 4.

Unsupervised Learning

According to [8], Unsupervised Learning lacks historical labels, so you must uncover patterns based on what is shown to you. To carry out this process, you must be fed by a comprehensive data source on which you can explore and find a structure or main attributes that allow you to segment and identify data. A classic example is the segmentation of clients based on their behavior, which is found within a full data history.See figure 4



Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

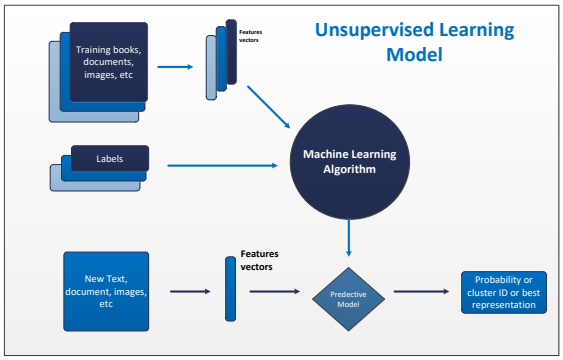


Figure 4. Supervised and Unsupervised Learning

Reinforced Learning

According to [8], reinforced learning defines a model focused on maximizing the actions that produce the most significant rewards and the environment in which the intelligent agent will perform. This algorithm is based on the behavioral psychology of the human being, since its action-reward model, seeks that the algorithm fits the best reward given by the same environment, and its actions to take are subject to these rewards. A widely used example is used by robots learning to perform different tasks based on trial and error. See figure 5.

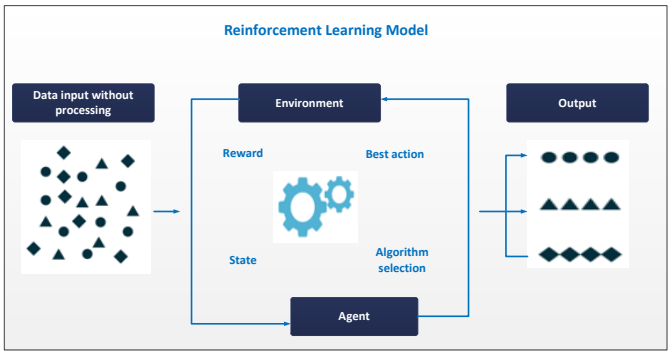


Figure 5. Reinforced Learning

Supervised learning is quite common in classification problems because the goal is often to make the computer learn a classification system that we have created. [7]

3.2. Classification algorithms

According to [9], Machine Learning algorithms are organized in a taxonomy based on the desired result of the algorithm. Supervised learning generates a function that maps inputs to desired outputs. Unprecedented data generation has caused machine learning techniques to become sophisticated from time to time.

This has been used for various algorithms for both supervised and unsupervised machine learning. Supervised learning is quite common in classification problems because the goal is often to make the computer learn a classification system that we have created.

This is why it describes various classification techniques based on Machine Learning, comparing multiple supervised learning algorithms, and determining the most efficient one according to the data set, the number of instances and variables. In table 3, we can see seven different machine learning algorithms that were considered: Decision Table, Random Forest, Naïve Bayes, Support Vector Machine, Neural Networks, JRip, and Decision Tree.

Table 3: Supervised Learn

Classification Techniques
Supervised Learning Algorithms
Decision Table
Random Forest
Naive Bayes
Support Vector Machine
Neural Networks
JRip
Decisions Tree

3.3. Detection of sports talents

According to [10], the detection of sports talents requires very well-structured processes to meet the planned objectives, in which the training and long-term development of the athlete is significant. However, to carry out these processes, it is necessary to define the meaning of the word talent.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

Like [11], detection is signal important configure algorithms.

Figure 6: Ranking of qualified athletes in the Men's 10K Half Marathon discipline

It is very common to hear from talented athletes in certain sports, but without having valid grounds to refer in this way about a player or athlete. When the word talent is heard, various questions arise, such as:

• What does this word imply?

• What characteristics does a talented athlete have?

• What types of talent are there?

• How does talent develop?

The meaning of the word talent must be understood in all its dimensions to understand this concept and justify development processes.

3.4. Classification of human talent

According to [10], the concept of talent can be applied to various social and behavioral fields of the human being, which can be appreciated in the different spheres of the human being. As we can see in figure 7, from the study of this concept, the following categories have emerged:

• Academic talent

• Creative talent

• Psychosocial talent

• Talent in the performing arts

• Kinesthetic talent

• The motor talent

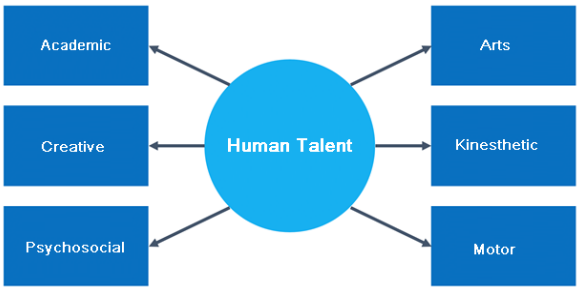


Figure 7: Categories of human talent

There is no doubt that one can be a talent in music, in mathematics, in language, in drawing, in sports, in interpersonal relationships and thus in each of the intelligence, and not in all the same. How many examples of this does life give how many brilliant musicians are just that, talented musicians; how many athletes are just that, outstanding people in their sport; how many writers are nothing but successful in that area.[12]

IV. WEB APPLICATION DEVELOPMENT

The project involves the development of a digital platform for the comprehensive management of the La Academia program, which has a classification algorithm based on Machine Learning to optimize the process of detecting sports talents. Part of the platform is a web application for browsers; it has various profiles that have permissions and functionalities so that they can carry out their corresponding processes.

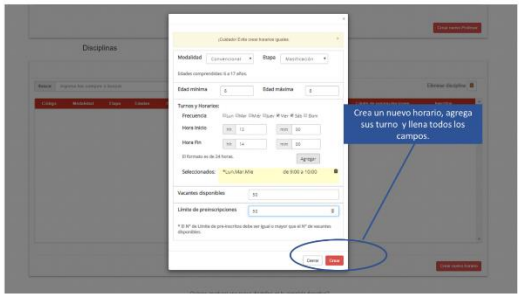


Figure 8: Schedule Creation

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

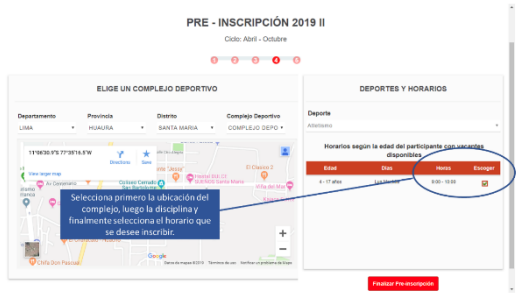


Figure 9: Virtual Inscription

Figure 10: Direct Inscription

Figure 11: Registered Management

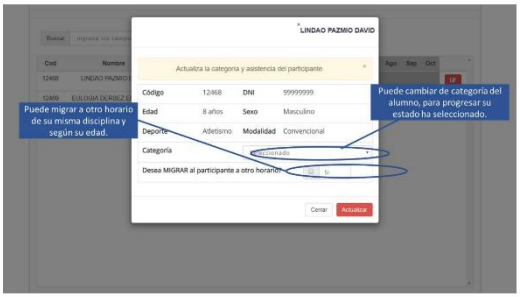


Figure 12: Schedule Management

Figure 13: Data Exportation

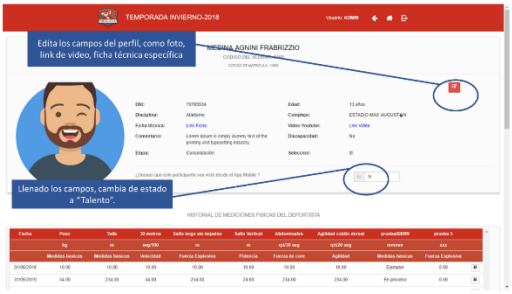


Figure 14: Talent Detection

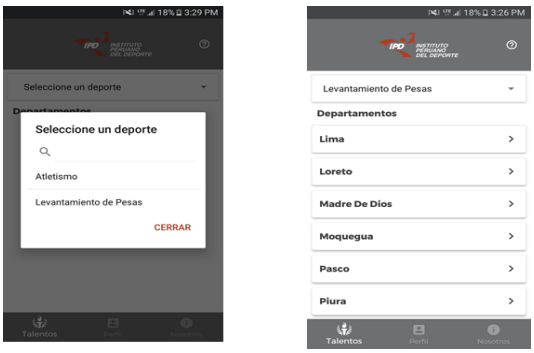


Figure 15: Data Search

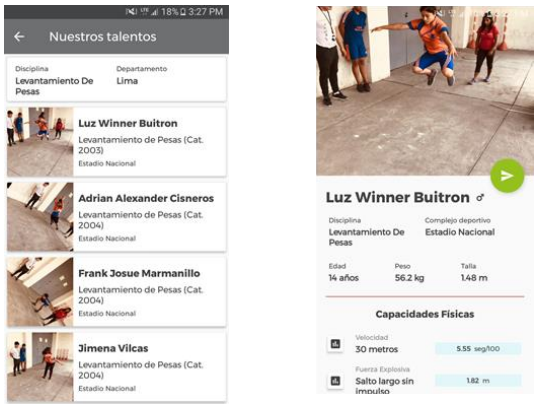


Figure 16: Results of Data Search

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 13464 - 13471

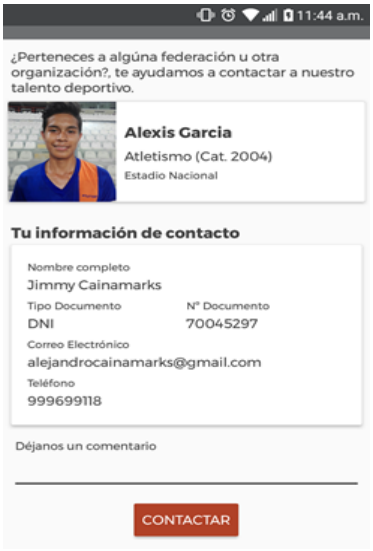


Figure 17: Talent Search -Contacts

V. CONCLUSIONS

The implementation of the classification algorithm based on Machine Learning achieved in the first instance supporting the processes that the program requires to cover the number of beneficiaries nationwide, from registering their virtual registrations, managing schedules, sports complexes, registering their attendance and digitizing all the information necessary to make decisions to implement improvements in program resource management. With the sports information of the beneficiaries, it was possible to improve the talent detection process by increasing the number of athletes detected and reducing the evaluation time and the margin of error in determining whether or not a beneficiary is a sports talent.

6. Recommendations

Implement other strategies and techniques to increase the precision of this process, such as Big Data and predictive analytics, which also allow for more personalized exploration.

VII. ACKNOWLEDGMENTS

A special recognition to my mentors from the National University Mayor de San Marcos, who have made it possible to put into practice the knowledge acquired during the period of stay in the study classrooms, as well as to collaborators who contribute to understanding spread internationally.

REFERENCES.

1. Peruvian Institute of Sport (2018), Technical Methodological Manual, Guidelines for the Organization of the Activities of La Academia.
2. Peruvian Institute of Sport (2018), Annual Report.
3. Peruvian Institute of Sport (2017), Annual Report.
4. Gareth James (2013), An Introduction to Statistical Learning: With Applications in R.
5. IswantoIswanto, E. Laxmi Lydia, K. Shanka&PhongThanh Nguyen (2019) Identifying Diseases and Diagnosis using Machine Learning
6. Paul Mueller, Luca Massaron (2016) Machine Learning for Dummies.
7. Machine Learning TIMES. The maching learning professional's premier resource. Retrieved November, 2019. Available on the World Wide Web at:https://www.predictiveanalyticsworld.com /machinelearningtimes/machine-learning-can-please-just-agree-means/9190/machinelearning-image/
8. Silvia Vicente Martinez (2019), The use of Smart Devices and Machine Learning for the prediction of diseases.
9. Alex Smola y S.V.N. Vishwanathan (2008) Introduction to Machine Learning.
10. GershonTenenbaum, Robert C. Eklund (2007) Handbook of Sport Psychology.
11. J. Bustamante, C. Rodriguez, D. Esenarro. "Real Time Facial Expression Recognition System Based on Deep Learning.". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019
12. Fernando Lapalma (2005),Multiple intelligences and talent development.