

Fairness in Football Player Transfer Value Prediction

Franciszek Wysocki and Gabriel Meira

Course code: 91257 - Ethics in Artificial Intelligence

May 30th, 2024

Abstract

This project involves the development of a machine learning model designed to predict football players' transfer values using a dataset that includes age at the moment of the transfer, nationality, position, market value, remaining contract days, and changes in market values (delta value 1 and delta value 2). The dataset will be complemented by actual transfer amounts for players who have been acquired, allowing the model to be trained on real-world data. The primary aim is to examine potential biases and fairness issues within the model, especially focusing on variables such as nationality and age. We analyze how these biases manifest using metrics to check fairness and bias in each of these features and tools that will help us, in particular Aequitas.

Contents

1	Introduction	4
2	Methodology	5
3	Problem Understanding	6
4	Data Understanding	7
4.1	Data Collection	7
4.2	Data Characterization	7
4.3	Data Preparation	7
4.4	Model Training	8
5	Fairness and Bias Analysis	9
5.1	Pre-Aequitas exploration	9
5.2	Using Aequitas	10
5.2.1	Equal Parity	10
5.2.2	False Positive Rate Parity	11
5.2.3	False Discovery Rate Parity	12
5.2.4	False Negative Rate Parity	14
5.2.5	False Omission Rate Parity	15
6	Discussion	18
7	Conclusion	19

1 Introduction

Football has seen a huge increase in player transfer actions and money being spent by clubs all over the world since the last two decades or so. A scenario in which the football transfer value is accurately predicted [5] would, therefore, be particularly valuable to clubs, agents and other stakeholders involved in player transactions. Given its position at the interface between number crunching and decision making, management finds itself wrestling with forecasts that can now be based on trained machine learning models of — for example — what price to expect a sale value to be, depending on player qualities or market dynamics.

This project focuses on developing a simple transfer value prediction model based on a dataset containing various information related to these transactions, such as age, market value, and remaining contract days at the time of sale, as well as nationality, position, and market value variations. By associating these features with the actual transfer value, the model seeks to make accurate predictions of transfer values, which would greatly assist clubs in player buying and selling decisions.

However, the development of such a model raises concerns related to fairness and biases inherent to the dataset and learned by the model. Therefore, this project also aims to identify and analyze possible discriminatory practices or inequalities present in the data, aiming to ensure fair outcomes for all players.

Thus, one of the project’s objectives is also to identify potential biases through a comprehensive and visual analysis using the Aequitas [6] tool. This analysis will involve examining how different groups are treated by the model, identifying areas where fairness may be compromised. Finally, we will discuss the impacts of these issues and the possible causes for the model to have learned these injustices.

2 Methodology

The methodology employed in this project aims to ensure robust development of a fair model for predicting transfer values. Initially, data collection was carried out using **Transfermarkt** [9], a platform that groups player and transfer data and calculates their market values based on several external factors, such as performance and media. The initial datasets included player information, transfer information and historical market values for each player.

Given the datasets, the data preparation and cleaning process began. At this stage, missing values were treated and an exploratory data analysis was carried out. Furthermore, new features were built and used in the model as well as some categorical variables were encoded.

Next, we move on to defining the regression model to be used to predict the values. We tested several possibilities such as Linear Regression, Gradient Boosting and ended up deciding on Random Forrest [8], which reached satisfactory accuracy. In this process we also tested different hyper parameters in order to obtain the most accurate results possible.

Fairness and bias analyses using Aequitas are a fundamental part of the methodology adopted, making it possible to really analyze the proposed model regarding these aspects and fuel discussions with respect to the fairness of our model. At this stage, we used the methods and tools provided by Aequitas to evaluate the model’s performance among different demographic groups of age and nationality. Fairness metrics were calculated to assess disparities between groups, in addition to the use of visualizations to facilitate understanding of the injustices highlighted by the information calculated by Aequitas.

Finally, we discussed the possible causes for these biases in football and how these inequalities may be mitigated in the model, proposing strategies that would make predictions fairer and respecting ethical principles in the use of Artificial Intelligence.

3 Problem Understanding

To understand the problem it is necessary to analyze the entire context of the football transfer market. In this market, it is unlikely that a player's transfer fee will be equal to the estimated market value for that player, as this fee is influenced by several other factors, such as age, nationality, remaining contract days, position, value variations. market (indicating decline or rise) among others. However, there is growing concern in the current context about the bias and fairness presented by a model, especially with demographic characteristics such as nationality and age.

These concerns are fundamental due to the problems that can be caused by the presence of bias related to these characteristics in the process of predicting transfer values. Therefore, an unfair model may result in transfers with unfair values, where players are overvalued or undervalued due to their nationalities and age groups compared to other players with similar characteristics. Furthermore, these practices can perpetuate existing inequalities in football, negatively affecting players from certain backgrounds or age groups.

4 Data Understanding

4.1 Data Collection

Data collection was performed using the Transfermarket website, which is a widely recognized online platform used for the collection, analysis and distribution of data related to football. The website offers information such as Player Data (including performance statistics, career history, field position, estimated market value, etc.), Transfer Information and estimated market values for players and teams, calculated based on a variety of factors, including performance recentness, future potential, age, and market demand.

We used three datasets:

- **Players_Info** has all the personal information about each of the more than 44 thousand players in the database, such as age, dominant foot, position, secondary, height, primary position, full name, etc. This dataset is used to obtain the id, nationality and position of each of the players present in a transfer.
- **Transfers** has information related to each of the base's more than 120,000 market movements, such as: type of transfer (e.g., sale, loan), player ID, competition and team of origin, competition and team of destination, coach of origin and coach of destination, transfer value, market value at the time of transfer, date of transfer and end date of the contract at the time of transfer. This dataset provides the main data on which we will work.
- **Market_Values** is the simplest and contains only 5 columns: player id, age, market value date, market value and team. This way, it serves as a historical record of all the market values that a player has ever had, we will use it to calculate the player's price variations.

4.2 Data Characterization

The Data Characterization process of this project involved examining the **players.info**, **transfers**, and **market_values** datasets in order to understand their structure and the values found in each of their columns, in addition to the statistical analysis of each of them. For **players.info** key elements such as nationality, primary position and loan status and their unique values were analyzed, in addition to the distribution and variability of these columns. Similarly, for **transfers** we analyzed attributes such as transfer type, ages at the time of transfer and loan status, which made it possible to better understand the distribution of the data. Finally, **market_values** was the simplest, and we only did the statistical analysis of its data.

4.3 Data Preparation

After loading the datasets and performing exploratory analysis, we moved on to data preparation, which began by removing columns considered irrelevant for the model predicting transfer values. Additionally, columns containing dates had their format corrected to homogenize the data format, and a cutoff date was established to discard any transfers that occurred after the chosen date, as we did not have market values beyond the cutoff date, making it impossible to calculate market value variations.

Next, we performed some data filtering to retain only attributes that are relevant to the problem. The first filter aimed to keep only the transfers categorized as 'sale' and with the loan boolean set to False, since the model should consider only transfers with a transfer fee, justifying the removal of all instances where the 'VL_FEE' is ≤ 0 , ensuring that we only kept paid and definitive transfers.

The next step was handling missing features for the model. The feature 'CONTRACT_DAYS_LEFT' extracts the number of days remaining on the contract at the time of sale, a relevant factor for determining how much needs to be paid to hire a player. Subsequently, the process of building market value variation features was carried out as follows:

- **Delta_value_1:** represents the difference between the market value of the player at the time of the transfer and the immediately preceding market value.

$$\delta_{\text{value1}} = V_{\text{market, at transfer}} - V_{\text{market, previous}}$$

- **Delta_value_2:** indicates the variation between two market value points prior to the transfer, that is, the difference between the value immediately before the last and the even older value.

$$\delta_{\text{value2}} = V_{\text{market, previous}} - V_{\text{market, before previous}}$$

One of the challenges in using old values is the need to adjust them financially to today's terms, so we created a function to adjust the values according to the year they occurred and correct them according to the inflation. We applied this function to all columns representing a financial value. Furthermore, it was also necessary to extract the first values from the attributes '**citizenship**' and '**Position**', since players may have more than one value in these columns, but the first was the main one.

Finally, it was necessary to create groups for the categorical column of nationality. Our strategy was to create 4 groups, following the FIFA [4] ranking, with group 1 composed of the top 5 countries, group 2 the 6-10 best, and group 3 the 11-15 best, with the others in group 4. We also eliminated all instances that had any of these features with NaN.

At the end of all these steps, the data were all ready to be used in the model, except for the encoding of the categorical variables **Position** and **Nationality**, so that they could be used in the regression model. We decided to use the one-hot encoding method for them, given that they were not ordinal variables.

4.4 Model Training

The methodology adopted for training the model follows a standard approach. Firstly, we define our feature set, composed of age, market value, remaining contract days at the time of transfer, delta_value_1, delta_value_2, nationality, and position, and our target: transfer fee. Next, we split the data into training and testing sets and apply the encoding decided and explained in the **Data Preparation** stage.

With the data prepared and separated into training and testing sets, we need to choose which algorithm to use. To do this, we test the algorithms **Linear Regression**, **Random Forest Regressor**, and **Gradient Boosting Regressor**. We vary the hyperparameters for each of these algorithms and analyze the cross-validated Mean Squared Error (MSE) and R^2 (R-squared) of each to determine the best option [7]:

```

Random Forest - Cross-validated MSE: 32848620965309.883, R2: 0.951692604545831
Linear Regression - Cross-validated MSE: 33342987553055.32, R2: 0.8112976625246552
Gradient Boosting - Cross-validated MSE: 32967646704716.188, R2: 0.9067323733083208

Best model selected from cross-validation is Random Forest with MSE: 32848620965309.883

Final test results - MSE: 27553833838774.582, RMSE: 5249174.58642543, R2: 0.8139959671563838

```

Figure 1: Models performance

Thus, we decide on the Random Forest Regressor, which achieved quite satisfactory performance, namely an $R^2 = 0.81$.

Finally, we prepare the dataset with the results in the format required by Aequitas. The Aequitas tool only works with binarized scores and *label_values* and does not accept continuous data, so we need to define our evaluation procedure considering these constraints. The sub-optimal solution found was to create two measures for comparison between the predicted and the real transfer fee values. The real measure, assigned to the *label_value* column, is obtained by the difference between the transfer fee and the market value at the time of sale: if the player selling value was above the market value, otherwise it receives a label 0. The measure representing our prediction is stored in the *score* column and is calculated following the same logic but replacing the actual transfer fee with the transfer fee predicted by the model. In this way, it becomes possible to make the comparison between what our model predicted and the real value in a binarized setup.

The dataset constructed to be input into the Aequitas tool uses only instances from the test set to avoid the risk of contamination and over-fitting, and consists of the *score* and *label_values* columns and the two demographic variables that we will use to evaluate the fairness and bias of the model: age and nationality. The age attribute was grouped for better visualization of the plots generated by the Aequitas tool and the groups are 'U20', '20 a 23', '23 a 27', '27 a 31', and '31+'.

5 Fairness and Bias Analysis

5.1 Pre-Aequitas exploration

In this first step we will do a simple analysis of the differences in *label_values*(actual difference) and *scores*(predicted difference) for each of the demographic groups.

- **AGE_GROUPS:** Here we can see the difference between the labels and scores for each age group, with a clear overvaluation of players from the 'U20' and '20 to 23' groups.

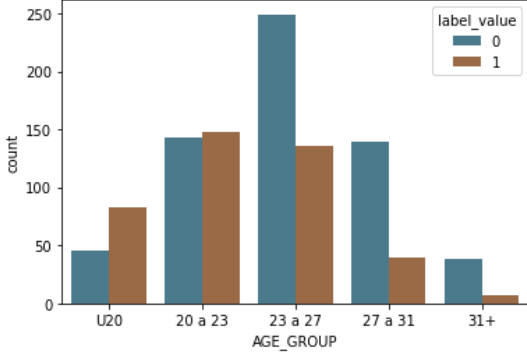


Figure 2: Distribution of labels between age groups

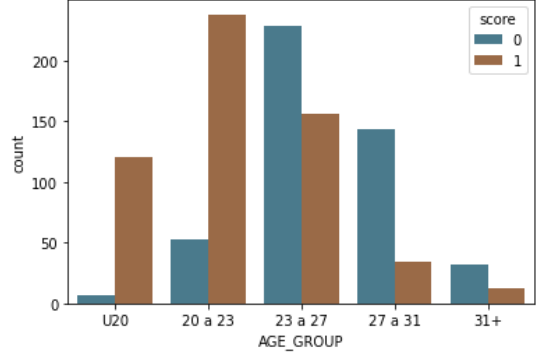


Figure 3: Distribution of scores between age groups

- **CITIZENSHIP:** In this plot we can see that the model predicts more sales values higher than the market value for all nationality groups, so we can not identify any bias easily.

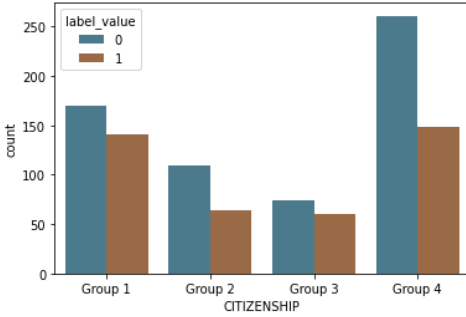


Figure 4: Distribution of labels between nationality groups

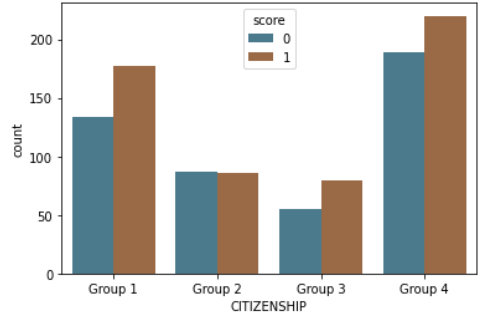


Figure 5: Distribution of scores between nationality groups

5.2 Using Aeiquitas

For the application of Aeiquitas and analysis of the results, we employed two different strategies. Initially, we utilized the Aeiquitas web interface [3], which generates an automated report. To further elucidate and justify the findings of the report, we leveraged the Aeiquitas [1] library along with Python. This enabled us to delve deeper into the analysis and visualization of the metrics employed. The reference group used was to automatically select, for each bias metric, the group on each attribute with the lowest value to serve as a baseline for calculating relative disparities in this audit.

5.2.1 Equal Parity

This criteria considers an attribute to have equal parity if every group is equally represented in the selected set. For example, if citizenship (with 4 possible values) has equal parity, it implies that all four citizenship are equally represented (25% each) in the selected set. The result of this criterion was a **FAILURE**: this occurred unbalanced representation of some groups in the selected sample, that is, the size of some groups were much smaller than the others. In Figure 6 we can see, for example, that the 20 to 23 year old group has many more representatives than the 31+ group.

Which groups failed the audit:

For CITIZENSHIP (with reference group as Group 3)
Group 1 with 2.24X Disparity
Group 4 with 2.78X Disparity

For AGE_GROUP (with reference group as 31+)
27 a 31 with 2.62X Disparity
U20 with 9.31X Disparity
23 a 27 with 12.00X Disparity
20 a 23 with 18.31X Disparity

Figure 6: Equal parity outcome

5.2.2 False Positive Rate Parity

This criteria considers an attribute to have False Positive parity if every group has the same False Positive Error Rate. For example, if age group has false positive parity, it implies that all groups of age have about the same False Positive Error Rate. Using this criteria allowed us to make sure that our model is not making false positive mistakes about any single group disproportionately. The result of this criterion was a **FAILURE** (Figure 7):

For AGE_GROUP (with reference group as 27 a 31)
20 a 23 with 4.64X Disparity
23 a 27 with 2.06X Disparity
31+ with 1.50X Disparity
U20 with 5.62X Disparity

Figure 7: FPR parity outcome

- CITIZENSHIP: We can see that the FPR [2] disparity for each of the nationality groups does not exceed the defined fairness limit (80%), so taking into account only nationalities the model would be fair (Figures 8 and 9).

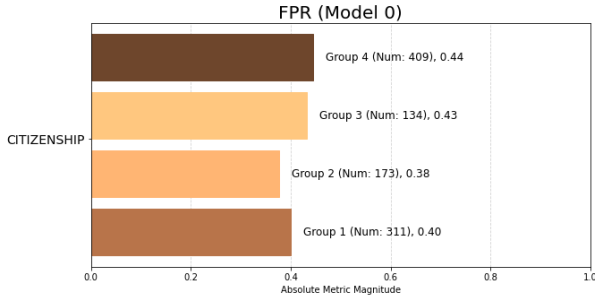


Figure 8: FPR of CITIZENSHIP

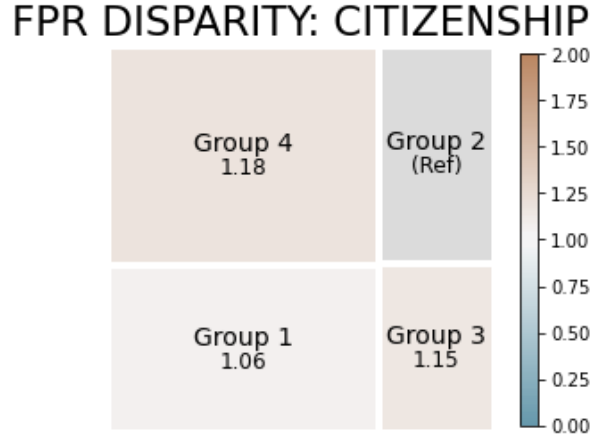


Figure 9: FPR disparity of CITIZENSHIP

- **AGE_GROUP:** On the other hand (Figures 10 and 11), in the different age groups, we can observe a more significant disparity related to the false positive rates for each group, which justifies the issue reported by Aequitas. This indicates that the proportion of false positives, predicted as positives (Predicted Transfer Fee > Market Value), but actually labeled as negatives (Transfer Fee $\leq$ Market Value), highlights an overvaluation of younger players compared to others. In other words, U20 players are falsely identified as having transfer fee greater than their market value 5.6 times the rate for 27 to 31 years old players

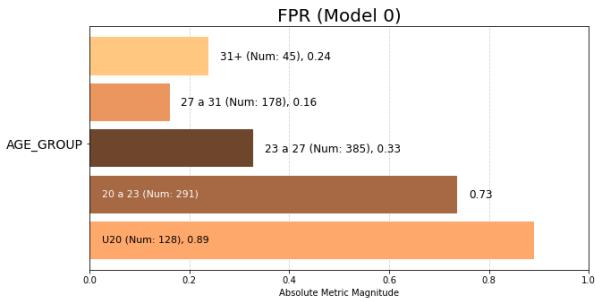


Figure 10: FPR of Age Group

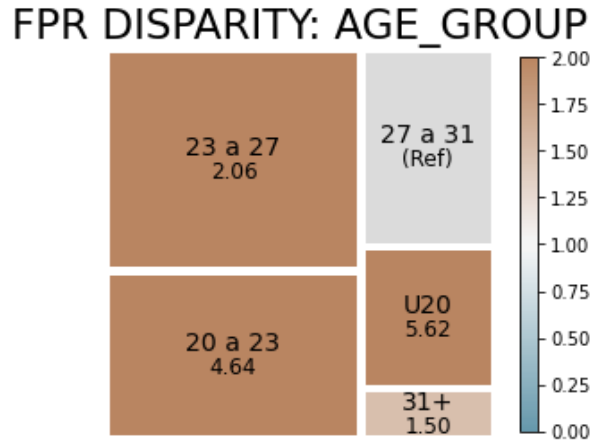


Figure 11: FPR disparity of Age Group

5.2.3 False Discovery Rate Parity

This criteria considers an attribute to have False Discovery Rate parity if every group has the same False Discovery Error Rate. For example, if age group has false discovery parity, it implies that all groups of age have the same False Discovery Error Rate. This metric seeks to show the fraction of false positives in a group in relation to predicted positives, that is, how many of the predicted positives were wrongly predicted. The result of this criterion was a **FAILURE** (Figure 12):

Which groups failed the audit:

For CITIZENSHIP (with reference group as Group 1)
 Group 4 with **1.37X** Disparity

For AGE_GROUP (with reference group as U20)
 31+ with **2.09X** Disparity
 20 a 23 with **1.33X** Disparity
 27 a 31 with **1.96X** Disparity
 23 a 27 with **1.57X** Disparity

Figure 12: FDR parity outcome

- CITIZENSHIP: We can see that the FDR disparity of group 4 in comparison to the chosen reference group exceeds the defined threshold. Even though they do not have a very large disparity, group 4 points out that more than 50% of the positive predicted values are false positives (Figures 13 and 14).

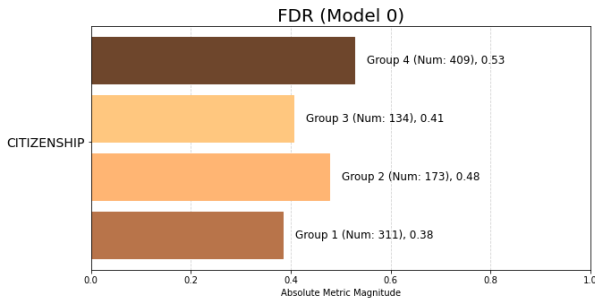


Figure 13: FDR of CITIZENSHIP

FDR DISPARITY: CITIZENSHIP

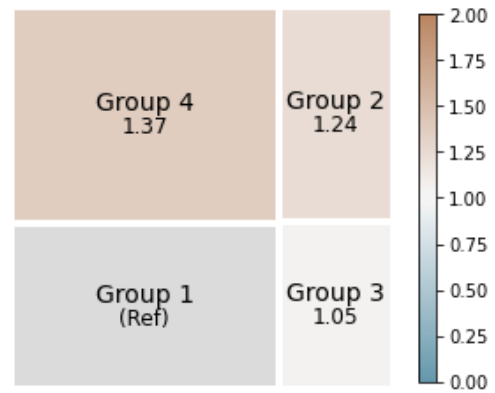


Figure 14: FDR disparity of CITIZENSHIP

- AGE_GROUP: When comparing the age groups, we see significant disparities, indicating unfairness and bias found in the model (Figures 15 and 16). An interesting point is that the older player groups, despite having low FPRs, which indicates that labeled negatives are indeed predicted as negatives, have high FDRs. This means that even though there are fewer values erroneously predicted as positives compared to younger players, these values still represent a larger proportion of the total predicted positives.

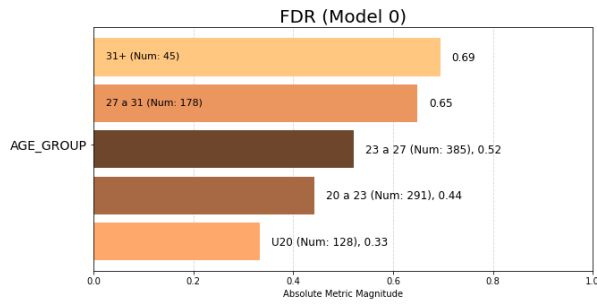


Figure 15: FDR of Age Group

FDR DISPARITY: AGE_GROUP

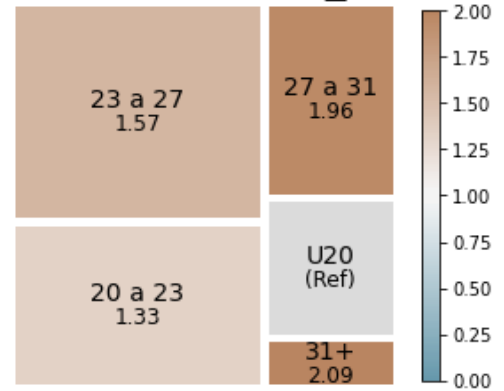


Figure 16: FDR disparity of Age Group

5.2.4 False Negative Rate Parity

This criteria considers an attribute to have False Negative parity if every group has the same False Negative Error Rate. For example, if age group has false negative parity, it implies that all age groups have the same False Negative Error Rate. Using this criteria allows you to make sure that you are not missing people from certain groups disproportionately. This metric calculates the proportion of false negatives relative to the labeled positives, that is, what proportion of the labeled positive values were incorrectly predicted as negatives. In other words, what proportion of players had their Transfer Fee > Market Value, but the model predicted otherwise. The result of this criterion was a **FAILURE** (Figure 17):

Which groups failed the audit:

For CITIZENSHIP (with reference group as Group 3)
 Group 4 with **1.37X** Disparity
 Group 2 with **1.37X** Disparity

For AGE_GROUP (with reference group as U20)
 23 a 27 with **18.61X** Disparity
 27 a 31 with **28.73X** Disparity
 20 a 23 with **4.21X** Disparity
 31+ with **17.79X** Disparity

Figure 17: FNR parity outcome

- CITIZENSHIP: Again, we can observe greater fairness and less disparity among the nationality groups compared to the age groups (Figures 18 and 19). However, in this case both Group 4 and Group 2 exceed the disparity threshold in relation to the reference group. This indicates that the model, proportionally to the players

with a positive label, undervalues players from Groups 2 and 4 more than those from Groups 1 and 3.

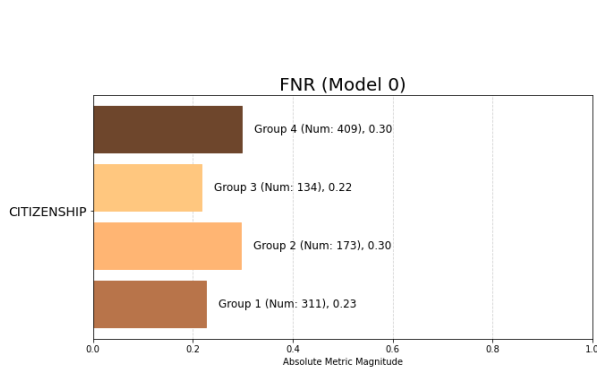


Figure 18: FNR of CITIZENSHIP

FNR DISPARITY: CITIZENSHIP

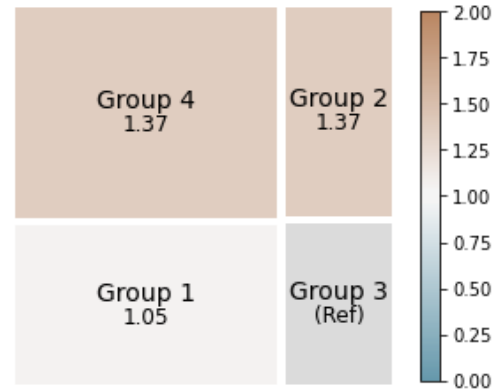


Figure 19: FNR disparity of CITIZENSHIP

- **AGE_GROUP:** The graphs (Figures 20 and 21) show us that, once again, there is a significant disparity among age groups, with the younger groups (especially U20) having the lowest false negative rate metric. In other words, the likelihood of a player's transfer fee being erroneously undervalued is very small for this group, while players in the 27 to 31 age group have a much higher likelihood of experiencing such undervaluation. Furthermore, the disparity graph shows us the same information, making it possible to state that players aged 27 to 31 are 28.7 times more likely to be falsely identified as having a lower transfer fee compared to their market value than U20 players are.

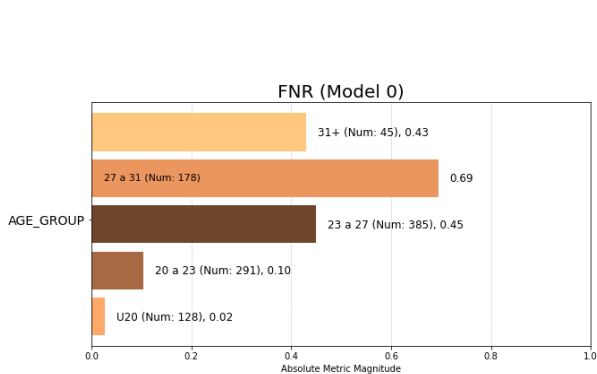


Figure 20: FNR of Age Group

FNR DISPARITY: AGE_GROUP

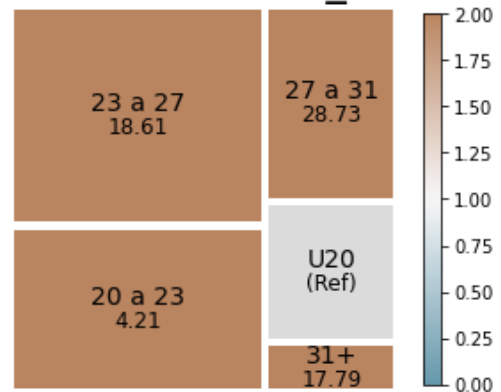


Figure 21: FNR disparity of Age Group

5.2.5 False Omission Rate Parity

This criteria considers an attribute to have False Omission Rate parity if every group has the same False Omission Error Rate. For example, if an age grouping has false omission parity, it implies that all age groups have the same False Omission Error Rate. Using this criteria allows you to make sure that you're not missing people from certain groups

disproportionately. This metric aims to indicate the proportion of false negatives relative to the predicted negatives. In other words, it shows what proportion of players predicted to have a transfer fee $\leq$ market value should have actually been predicted to have a transfer fee $>$ market value. The result of this criterion was a **FAILURE** (Figure 22):

Which groups failed the audit:

For AGE_GROUP (with reference group as 31+)
 U20 with **3.05X** Disparity
 23 a 27 with **2.84X** Disparity
 20 a 23 with **3.02X** Disparity
 27 a 31 with **2.00X** Disparity

Figure 22: FOR parity outcome

- **CITIZENSHIP**: We can see that the FOR disparity for each of the nationality groups does not exceed the defined fairness limit (80%), so taking into account only nationalities the model would be fair (Figures 23 and 24).

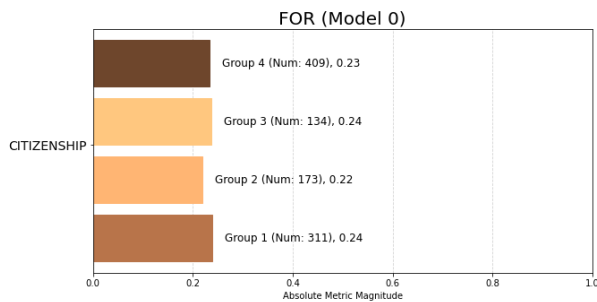


Figure 23: FOR of CITIZENSHIP

FOR DISPARITY: CITIZENSHIP

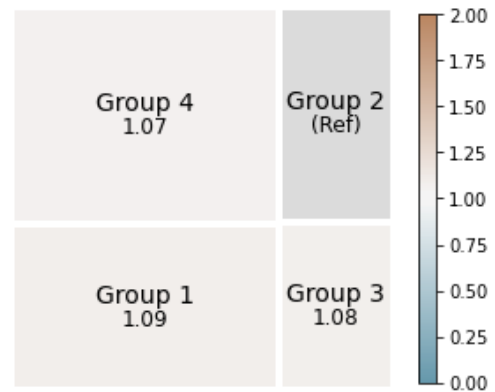


Figure 24: FOR disparity of CITIZENSHIP

- **AGE_GROUP**: Finally, among the age groups, the metrics indicate a greater unfairness once again. We can observe that the FOR indicates that older player groups are unlikely to incorrectly predict a negatively labeled player as positive, while this proportion is higher for younger players despite having a low FNR. This indicates that, even with few false negatives relative to the positively labeled players, these few represent a considerable proportion of the predicted negatives (Figures 25 and 26).

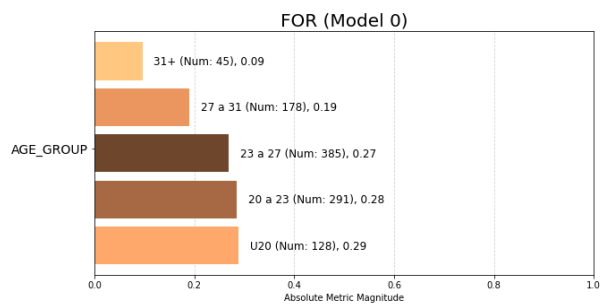


Figure 25: FOR of Age Group

FOR DISPARITY: AGE_GROUP

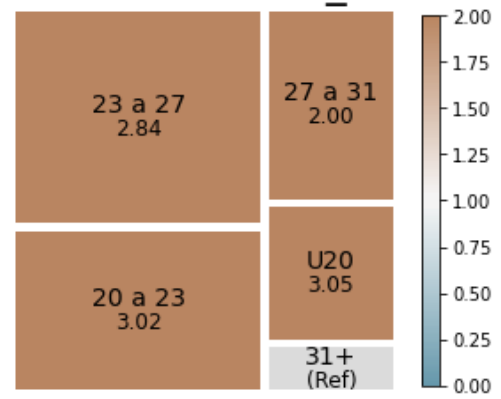


Figure 26: FOR disparity of Age Group

6 Discussion

After this analysis of bias and disparity metrics, we were able to verify the existence of significant bias in the model, especially in the age demographic group, indicating an overvaluation of the transfer values of younger players. When analyzing the nationality groups, we did not notice a significant impact of bias, possibly because the countries are divided into groups with at least five countries each. Additionally, the groups were defined based on FIFA’s ranking of the best football countries, which may justify the model’s consistency in these groups.

Based on the results analyzed in the previous section, we identified that the model learned overvalues the transfer value of players in general, but it also became clear that it tends to value more players in the U20 and 20 to 23 age groups, as we observed from the numbers presented by the False Positive Rate. Furthermore, we also noticed an undervaluation of the transfer fee for older players, as observed when analyzing the False Negative Rate, which indicates a much higher proportion of false negatives relative to the labeled positives for the older groups compared to the younger ones. In other words, the proportion of players whose transfer fee is predicted to be less than or equal to the market value when it should actually be higher is much more significant in the older player groups than in the younger ones.

7 Conclusion

In this project, we were able to successfully develop a machine learning model using Random Forest Regressor to predict transfer values for football players, using a dataset with characteristics such as age, nationality, position, market value and remaining days on contract. The model demonstrated satisfactory accuracy and would meet the demands of the current football market. However, a significant point of the project was the analysis of potential biases and fairness issues within the model, focusing on the demographic variables nationality and age.

Using the Aequitas tool, we performed an in-depth analysis of fairness and bias, revealing relevant disparities, especially in age-related predictions. The model tended to overvalue younger players (under-20 and age groups from 20 to 23 years old) and undervalue older players, indicating a significant bias. Visualizations and fairness metrics, such as false positive rate parity and false negative rate parity, provided clear evidence of these biases, demonstrating the importance of this analysis to ensure fairer results.

Future research should aim to mitigate these biases through techniques such as re-sampling or introducing fairness constraints during model training. Furthermore, the use of new features and validation of the model with professionals in the field can further improve the model. We then conclude stating the importance of this type of analysis aiming to build fair and bias-free predictive models, guaranteeing similar treatment to different demographic groups and without harming individuals due to personal characteristics.

References

- [1] Aequitas. Aequitas demonstration in Python, Accessed in May 28, 2024.
- [2] Aequitas. Bias Metrics by Aequitas, Accessed in May 28, 2024.
- [3] Center for Data Science and Public Policy at the University of Chicago. Aequitas info, Accessed in May 28, 2024.
- [4] FIFA. FIFA, Accessed in May 29, 2024.
- [5] Dibyanshu Patnaik, Harsh Praharaj, Kartikeya Prakash, and Krishna Samdani. A study of prediction models for football player valuations by quantifying statistical and economic attributes for the global transfer market. In *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, pages 1–7. IEEE, 2019.
- [6] Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, and Rayid Ghani. Aequitas: A bias and fairness audit toolkit, 2019.
- [7] SKLearn. SKLearn supervise learning models, Accessed in May 29, 2024.
- [8] Toward data science. Regression models, Accessed in May 28, 2024.
- [9] Transfermarket. Transfermarket, Accessed in May 28, 2024.