PREDICTIVE PRICING

Sara Sklenka, Aliza Kajani, Tommy Hessel, CJ Keim, Scott McConnell

INTRODUCTION

The FIFA console game by EA Sports is one of the most played games year in and year out, and within the game there are a handful of different modes that the player can enjoy, but none is more prominent than FIFA Ultimate Team (FUT) It is a game mode that allows individuals to build their dream squad from scratch and compete with it in a variety of single player and online games.

You begin FIFA Ultimate Team with a starter pack of players, most of which are lower tier players. However, by winning matches and completing challenges throughout FUT, you earn more coins and have the opportunity to buy new players on the FUT marketplace. Within this marketplace there are well over 15,000 different players available to bid on at any given time. It is similar to an eBay format, where each player has an expiring date and then can be bid on or bought at the *buy it now* price until time expires. With all this being said, we arrive at the two main goals of our project:

(1) Based on a number of their different individual statistics in the game, predict whether the player is "elite," which we define as having a price in the top 1% (over 20,000) in the FUT marketplace.
(2) After classifying the players into "elite" and "non-elite," use the player's individual statistics to predict their specific trading price in the FUT marketplace.

DATA

We utilized two main datasets, which can be understood as (1) an independent variables dataset and (2) a dependent variable dataset. The independent variables dataset comes from Kaggle, while the dependent dataset comes from a public git repository. More details on our datasets can be seen below:

The independent dataset, which we reference as the player characteristics dataset, is a dataset containing a kaleidoscope of player statistics, both objective ones (age, name, nationality, height, weight, jersey number, position, wage, preferred foot, etc.) and subjective ones (accuracy, agility, shot power, standing tackle, strength, vision, etc.), all of which are provided by the makers of FIFA (EA Sports).

The dependent dataset, which we denote by the player price dataset, represents all the player's prices in the FUT marketplace as of February 2, 2019.

In order to accomplish our goals for this project, one of the first things we needed to do was merge these two datasets. The merging process consisted of creating a dictionary of players mapped to their statistics, FUT value, and FUT card type. To avoid associating players with the same name with wrong prices, we matched players on country of origin as well as name.

IMPLEMENTATION

The implementation of our project can be broken down into three main steps. First, merge our two datasets. Second, create an accurate classifier to split the 18,000+ players into two groups "elite" and "non-elite," where elite is defined as player's who have a FUT market price of above 20,000 coins (top 1%). Third, create separate regression models for both subsets of players in order to accurately try to predict their FUT market price.

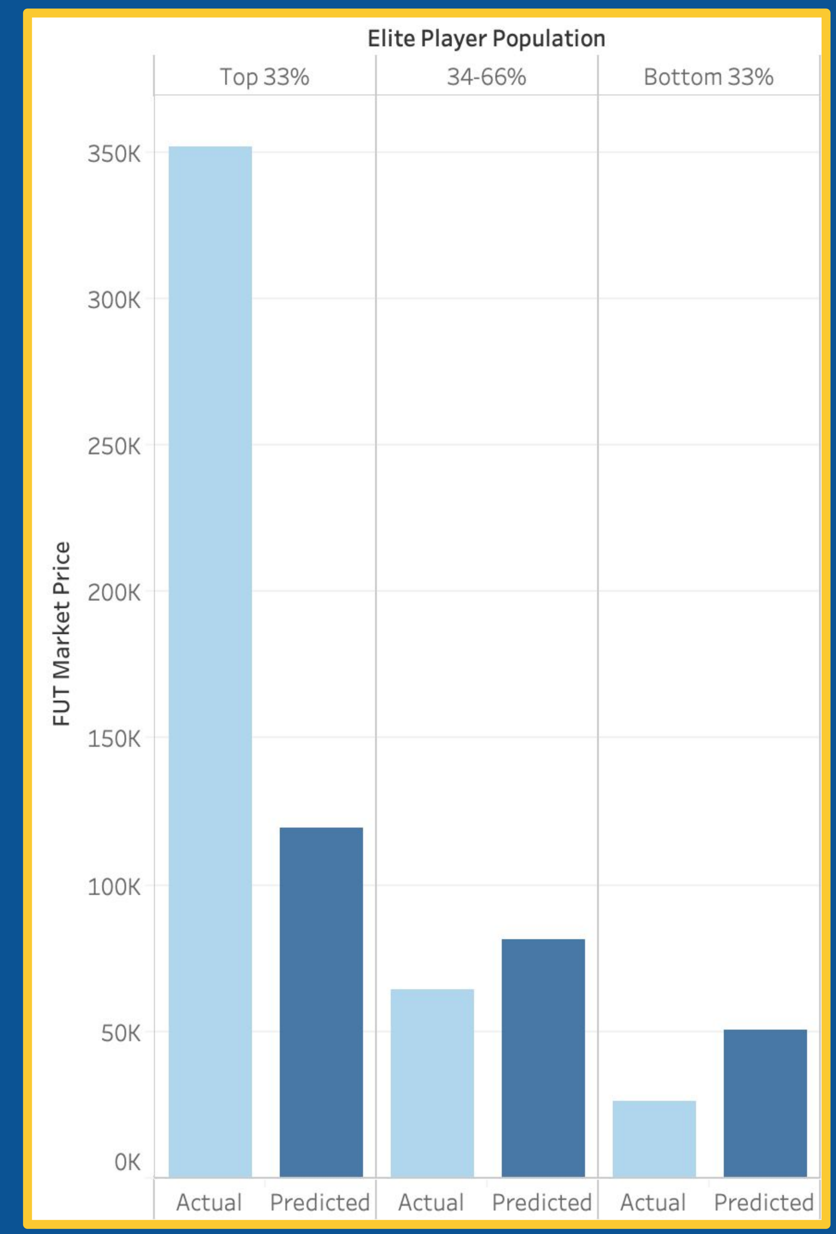
CLASSIFIERS

KNN Classifier	Naive-Bayes Classifier
90%	92%

From the initial regression testing and visualizations, it was apparent that those with high prices, the "elite," needed a different model to predict their prices than those within the lower prices, the "non-elite." In order to account for this and maintain the ability to input a random player and see that player's predicted price, we built two binary classification models using sklearn. We chose to test both the Naive Bayes and K-Nearest Neighbor models and achieved accuracy ~90% for both. Since the data did not have many elite players, we made sure to conduct separate trainings & testings for equal numbers of elite and non-elite before testing the entire set. We achieved slightly higher accuracy with the Naive Bayes model due to the dimensionality of our set (8 features), pushing the KNN neighbors further away. Additionally, our accuracy improved after label encoding the positions to account for the "goalie problem," where goalies ranked rather poorly on field stats but very highly on goalie specific stats.

ELITE REGRESSION

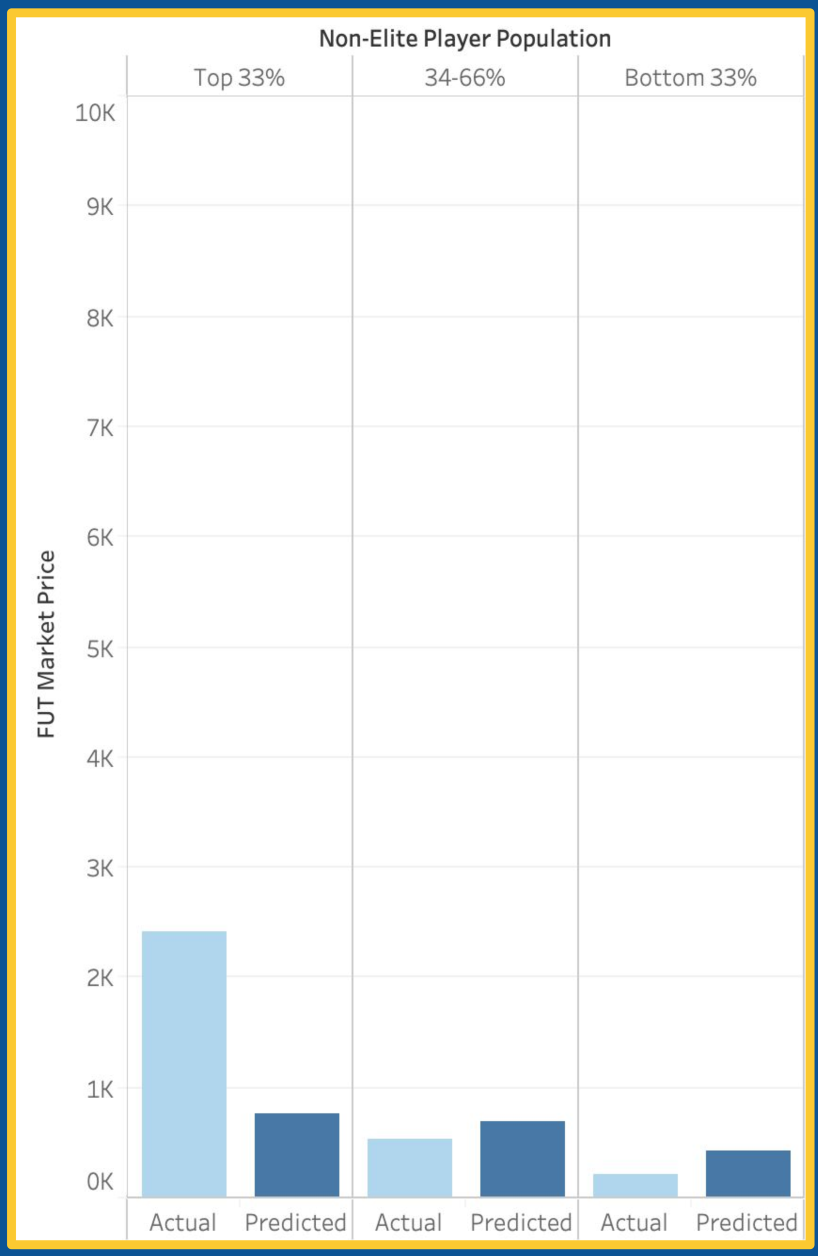
Variable	Coefficient	P-Value
potential	.1077	.000
forward	.3875	.012
is10	.2939	.168
maxPos	-.1417	.063
maxSkills	.0272	.122
eliteClub	.2431	.055
overall	.1431	.051
MSE		.8114
R-Squared		.3862



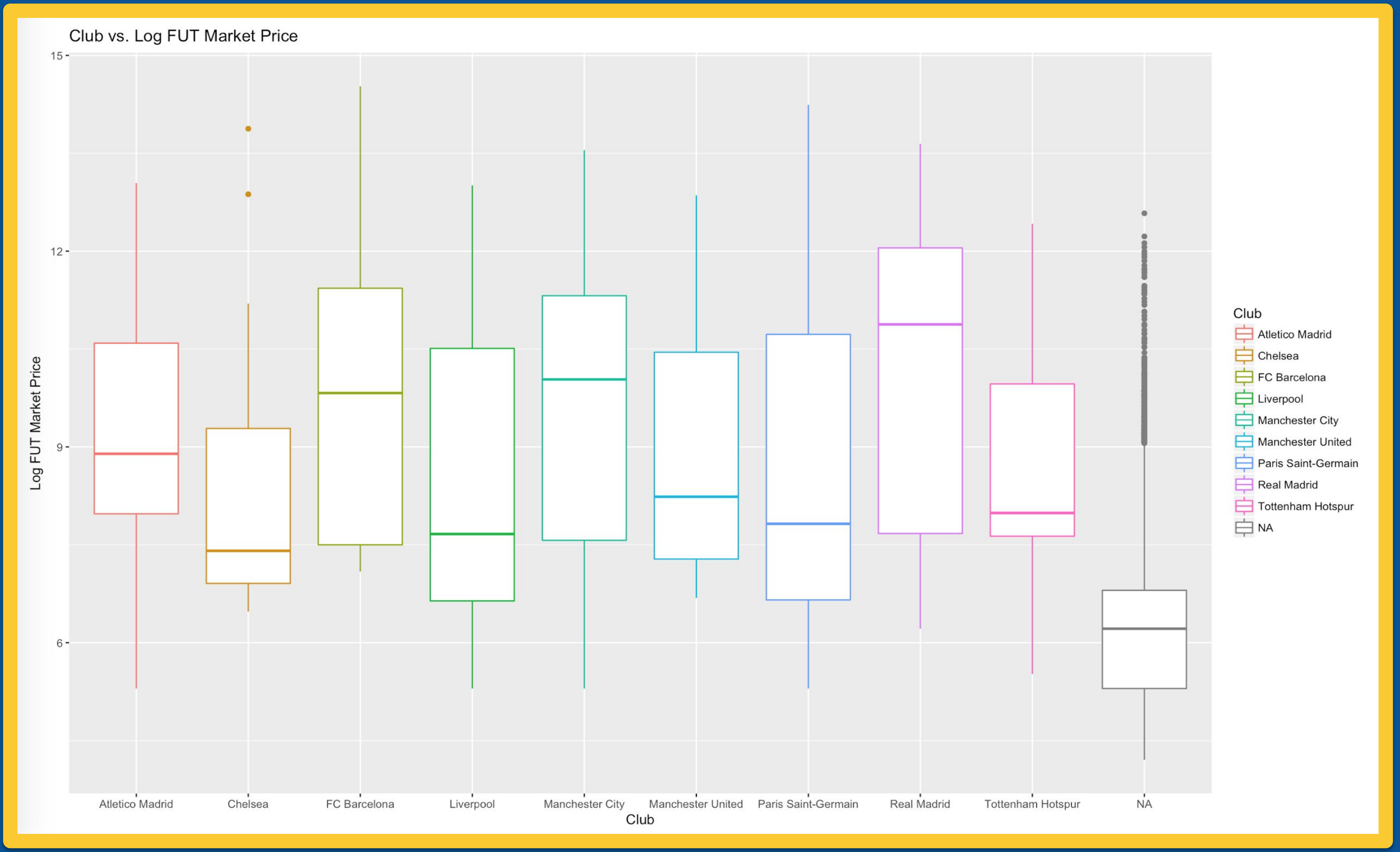
When regressing on elite players, we found the highlighted variables in the chart above to be the ones that are statistically significant at the 10% level. The bar chart on the right shows that our model consistently underpredicts the top 33% of the elite players' average price and overpredicts the bottom 67% of the elite players' average price.

NON-ELITE REGRESSION

Variable	Coefficient	P-Value
potential	.0034	.090
eliteClub	.6158	.000
is10	-.1277	.004
maxPos	.0655	.000
goldCard	.1436	.000
silverCard	.0904	.000
MSE		.8515
R-Squared		.2384



When regressing on non-elite players, we found the highlighted variables in the chart above to be the ones that are statistically significant at the 10% level. The bar chart on the right again shows that for non-elite players as well, our model consistently underpredicts the top 33% of the elite players' average price and overpredicts the bottom 67% of the elite players' average price.

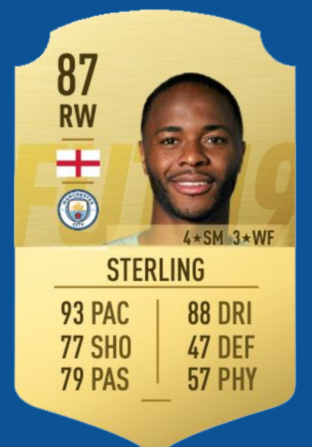


ELITE CLUBS VS. PRICE

The graph on the left plots the logarithmic FUT prices on the y-axis and the clubs we considered "elite" on the x-axis. In both the elite and the non-elite regression models above, one of the statistically significant variables was the "eliteClub" variable. This variable demonstrates that if players are apart of "elite" clubs, then they are much more likely to have a higher price in the FUT marketplace. The box-plot to the left further emphasizes this, as one can see that all of the elite clubs have a much higher median price than those that are not part of the elite clubs, denoted by the NA box column.

PLAYER PREDICTION EXAMPLE

From our elite regression model shown above, we omitted a handful of random players in the elite category, and then utilized these players to test our predictive model. This omittance allowed for a more unbiased test. Raheem Sterling was one of these omitted players.



Prediction	142,343
Actual	110,968

CONCLUSION

Overall, predicting a player's price on the FUT marketplace proved to be very challenging. One reason for this was due to the distribution of the player's prices in the FUT marketplace, as it is largely not representative of the real world price distribution amongst players. In particular, there are thousands of mediocre to good players that flood the FUT marketplace with low prices under the 2,000 level, while the elite players rack in large prices in a vast range on the market. Hence, this exponential price increase in the market makes it very difficult to create accurate predictive models based on their statistics.

We believe user psychology plays a role in these large price discrepancies. In particular, the statistically significant variables in both regressions indicate that users care more about player prestige (if a player belongs to an elite club, if they have a face card, if they are a gold card) more so than the statistics themselves. So is the price really a reflection of the player's skills or simply a reflection of their name and fame?