# Team Members: Shraman Basyal, John Slack, Shane Deng, Jonny Howard

# 1. Project Title.

Predicting the value of Football/Soccer players from their FIFA 2019 statistics.

# 2. Description of data source and web link(s).

Data includes players attributes from the 2019 edition FIFA video game.

Web link to data:[**https://www.kaggle.com/karangadiya/fifa19**](https://www.kaggle.com/karangadiya/fifa19)

# 3. # of records & # of attributes with description of each attribute.

18,207 records, 47 attributes chosen from dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Feature Name** | **Example Data** | **Feature Type** | **Data Type 1** | **Data Type 2** | **Note** | **Unit** |
| Age | 31 | Objective Feature | int | Numerical |  | Years |
| Nationality | Argentina | Objective Feature | str | Categorical | Option to encode or drop | N/A |
| Club | FC Barcelona | Objective Feature | str | Categorical | Option to encode or drop | N/A |
| Value | €110.5M | Target Variable | str | Numerical | Need to reformat to int from str | Euros |
| Preferred Foot | Left | Objective Feature | str | Categorical |  | Categorical |
| International Reputation | 5 | Subjective Variable | int | Numerical | Could be treated as ordinal. Higher is better | Rating 1-5 |
| Weak Foot | 4 | Subjective Variable | int | Numerical | Could be treated as ordinal. Higher is better | Rating 1-5 |
| Skill Moves | 4 | Subjective Variable | int | Numerical | Could be treated as ordinal. Higher is better | Rating 1-5 |
| Work Rate | Medium/ Medium | Subjective Variable | str | Ordinal | Convert to attacking/defensive workrate |  |
| Position | RF | Objective Feature | str | Categorical |  |  |
| Contract Valid Until | 2021 | Objective Feature | int | Numerical | Covert to Years left on contract | Year |
| Height | 5'7 | Objective Feature | str | Numerical | Convert to inches/int | ft/inches |
| Weight | 159lbs | Objective Feature | str | Numerical | Convert to int | lbs |
| Crossing | 84 | Subjective Variable | int | Numerical |  | Rating /100 |
| Finishing | 95 | Subjective Variable | int | Numerical |  | Rating /100 |
| HeadingAccuracy | 70 | Subjective Variable | int | Numerical |  | Rating /100 |
| ShortPassing | 90 | Subjective Variable | int | Numerical |  | Rating /100 |
| Volleys | 86 | Subjective Variable | int | Numerical |  | Rating /100 |
| Dribbling | 97 | Subjective Variable | int | Numerical |  | Rating /100 |
| Curve | 93 | Subjective Variable | int | Numerical |  | Rating /100 |
| FKAccuracy | 94 | Subjective Variable | int | Numerical | FK = Free Kick | Rating /100 |
| LongPassing | 87 | Subjective Variable | int | Numerical |  | Rating /100 |
| BallControl | 96 | Subjective Variable | int | Numerical |  | Rating /100 |
| Acceleration | 91 | Subjective Variable | int | Numerical |  | Rating /100 |
| SprintSpeed | 86 | Subjective Variable | int | Numerical |  | Rating /100 |
| Agility | 91 | Subjective Variable | int | Numerical |  | Rating /100 |
| Reactions | 95 | Subjective Variable | int | Numerical |  | Rating /100 |
| Balance | 95 | Subjective Variable | int | Numerical |  | Rating /100 |
| ShotPower | 85 | Subjective Variable | int | Numerical |  | Rating /100 |
| Jumping | 68 | Subjective Variable | int | Numerical |  | Rating /100 |
| Stamina | 72 | Subjective Variable | int | Numerical |  | Rating /100 |
| Strength | 59 | Subjective Variable | int | Numerical |  | Rating /100 |
| LongShots | 94 | Subjective Variable | int | Numerical |  | Rating /100 |
| Aggression | 48 | Subjective Variable | int | Numerical |  | Rating /100 |
| Interceptions | 22 | Subjective Variable | int | Numerical |  | Rating /100 |
| Positioning | 94 | Subjective Variable | int | Numerical |  | Rating /100 |
| Vision | 94 | Subjective Variable | int | Numerical |  | Rating /100 |
| Penalties | 75 | Subjective Variable | int | Numerical |  | Rating /100 |
| Composure | 96 | Subjective Variable | int | Numerical |  | Rating /100 |
| Marking | 33 | Subjective Variable | int | Numerical |  | Rating /100 |
| StandingTackle | 28 | Subjective Variable | int | Numerical |  | Rating /100 |
| SlidingTackle | 26 | Subjective Variable | int | Numerical |  | Rating /100 |
| GKDiving | 6 | Subjective Variable | int | Numerical | GK = Goalkeeper | Rating /100 |
| GKHandling | 11 | Subjective Variable | int | Numerical | GK = Goalkeeper | Rating /100 |
| GKKicking | 15 | Subjective Variable | int | Numerical | GK = Goalkeeper | Rating /100 |
| GKPositioning | 14 | Subjective Variable | int | Numerical | GK = Goalkeeper | Rating /100 |
| GKReflexes | 8 | Subjective Variable | int | Numerical | GK = Goalkeeper | Rating /100 |

# 4. Some general statistics of the dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **unique** | **top** | **freq** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **Age** | 18207 |  |  |  | 25.12 | 4.67 | 16 | 21 | 25 | 28 | 45 |
| **Nationality** | 18207 | 164 | England | 1662 |  |  |  |  |  |  |  |
| **Club** | 17966 | 651 | FC Barcelona | 33 |  |  |  |  |  |  |  |
| **Preferred Foot** | 18159 | 2 | Right | 13948 |  |  |  |  |  |  |  |
| **International Reputation** | 18159 |  |  |  | 1.11 | 0.39 | 1 | 1 | 1 | 1 | 5 |
| **Weak Foot** | 18159 |  |  |  | 2.95 | 0.66 | 1 | 3 | 3 | 3 | 5 |
| **Skill Moves** | 18159 |  |  |  | 2.36 | 0.76 | 1 | 2 | 2 | 3 | 5 |
| **Work Rate** | 18159 | 9 | Medium/ Medium | 9810 |  |  |  |  |  |  |  |
| **Position** | 18147 | 27 | ST | 2152 |  |  |  |  |  |  |  |
| **Contract Valid Until** | 17918 | 36 | 2019 | 4819 |  |  |  |  |  |  |  |
| **Height** | 18159 | 21 | 6'0 | 2881 |  |  |  |  |  |  |  |
| **Weight** | 18159 | 57 | 165lbs | 1483 |  |  |  |  |  |  |  |
| **Crossing** | 18159 |  |  |  | 49.73 | 18.36 | 5 | 38 | 54 | 64 | 93 |
| **Finishing** | 18159 |  |  |  | 45.55 | 19.53 | 2 | 30 | 49 | 62 | 95 |
| **HeadingAccuracy** | 18159 |  |  |  | 52.30 | 17.38 | 4 | 44 | 56 | 64 | 94 |
| **ShortPassing** | 18159 |  |  |  | 58.69 | 14.70 | 7 | 54 | 62 | 68 | 93 |
| **Volleys** | 18159 |  |  |  | 42.91 | 17.69 | 4 | 30 | 44 | 57 | 90 |
| **Dribbling** | 18159 |  |  |  | 55.37 | 18.91 | 4 | 49 | 61 | 68 | 97 |
| **Curve** | 18159 |  |  |  | 47.17 | 18.40 | 6 | 34 | 48 | 62 | 94 |
| **FKAccuracy** | 18159 |  |  |  | 42.86 | 17.48 | 3 | 31 | 41 | 57 | 94 |
| **LongPassing** | 18159 |  |  |  | 52.71 | 15.33 | 9 | 43 | 56 | 64 | 93 |
| **BallControl** | 18159 |  |  |  | 58.37 | 16.69 | 5 | 54 | 63 | 69 | 96 |
| **Acceleration** | 18159 |  |  |  | 64.61 | 14.93 | 12 | 57 | 67 | 75 | 97 |
| **SprintSpeed** | 18159 |  |  |  | 64.73 | 14.65 | 12 | 57 | 67 | 75 | 96 |
| **Agility** | 18159 |  |  |  | 63.50 | 14.77 | 14 | 55 | 66 | 74 | 96 |
| **Reactions** | 18159 |  |  |  | 61.84 | 9.01 | 21 | 56 | 62 | 68 | 96 |
| **Balance** | 18159 |  |  |  | 63.97 | 14.14 | 16 | 56 | 66 | 74 | 96 |
| **ShotPower** | 18159 |  |  |  | 55.46 | 17.24 | 2 | 45 | 59 | 68 | 95 |
| **Jumping** | 18159 |  |  |  | 65.09 | 11.82 | 15 | 58 | 66 | 73 | 95 |
| **Stamina** | 18159 |  |  |  | 63.22 | 15.89 | 12 | 56 | 66 | 74 | 96 |
| **Strength** | 18159 |  |  |  | 65.31 | 12.56 | 17 | 58 | 67 | 74 | 97 |
| **LongShots** | 18159 |  |  |  | 47.11 | 19.26 | 3 | 33 | 51 | 62 | 94 |
| **Aggression** | 18159 |  |  |  | 55.87 | 17.37 | 11 | 44 | 59 | 69 | 95 |
| **Interceptions** | 18159 |  |  |  | 46.70 | 20.70 | 3 | 26 | 52 | 64 | 92 |
| **Positioning** | 18159 |  |  |  | 49.96 | 19.53 | 2 | 38 | 55 | 64 | 95 |
| **Vision** | 18159 |  |  |  | 53.40 | 14.15 | 10 | 44 | 55 | 64 | 94 |
| **Penalties** | 18159 |  |  |  | 48.55 | 15.70 | 5 | 39 | 49 | 60 | 92 |
| **Composure** | 18159 |  |  |  | 58.65 | 11.44 | 3 | 51 | 60 | 67 | 96 |
| **Marking** | 18159 |  |  |  | 47.28 | 19.90 | 3 | 30 | 53 | 64 | 94 |
| **StandingTackle** | 18159 |  |  |  | 47.70 | 21.66 | 2 | 27 | 55 | 66 | 93 |
| **SlidingTackle** | 18159 |  |  |  | 45.66 | 21.29 | 3 | 24 | 52 | 64 | 91 |
| **GKDiving** | 18159 |  |  |  | 16.62 | 17.70 | 1 | 8 | 11 | 14 | 90 |
| **GKHandling** | 18159 |  |  |  | 16.39 | 16.91 | 1 | 8 | 11 | 14 | 92 |
| **GKKicking** | 18159 |  |  |  | 16.23 | 16.50 | 1 | 8 | 11 | 14 | 91 |
| **GKPositioning** | 18159 |  |  |  | 16.39 | 17.03 | 1 | 8 | 11 | 14 | 90 |
| **GKReflexes** | 18159 |  |  |  | 16.71 | 17.96 | 1 | 8 | 11 | 14 | 94 |
| **Value** | 18207 | 217 | €1.1M | 431 |  |  |  |  |  |  |  |

# 5. Tools / methods your team (plan to) use in your study.

We will utilize Python and MATLAB to analyze data and build the model. We will use an ensemble method of combination of linear regression and SVM models for 4 categories of player that we will define: Goalkeeper, Defender, Midfielder, Attacker.

# 6. Detail description of what problems/questions your team plans to predict / study.

Question 1. How can we classify player position (4 categories of player that we will define: Goalkeeper, Defender, Midfielder, Attacker) based on their statistics?

We are planning to utilize SVM model to predict the class/position of the player.

Question 2. How can we predict players’ value based on statistics?

We will be doing a linear regression model for each position group from question 1 to predict the player’s value.