Let’s play FIFA! 

(AI-1 Project)

Himanshu Kandpal
Samyak R Jain
Amit Gupta
Akshul Mittal

Introduction (Part A)

Problem | Approach | Reasoning

- Overall score prediction
- How is the best player in the world?

Data:

- Player Overall Score
- Player attributes like attacking, goalkeeping, defense, skills, etc.
- Player personal data like Nationality, Club, Salary, etc.

- Data Cleaning
- Basic Analysis
- Baseline Model
- Approaching Different models
- Remove correlated columns.
- Retraining on models
- Plotting top 10 player barchart using best model

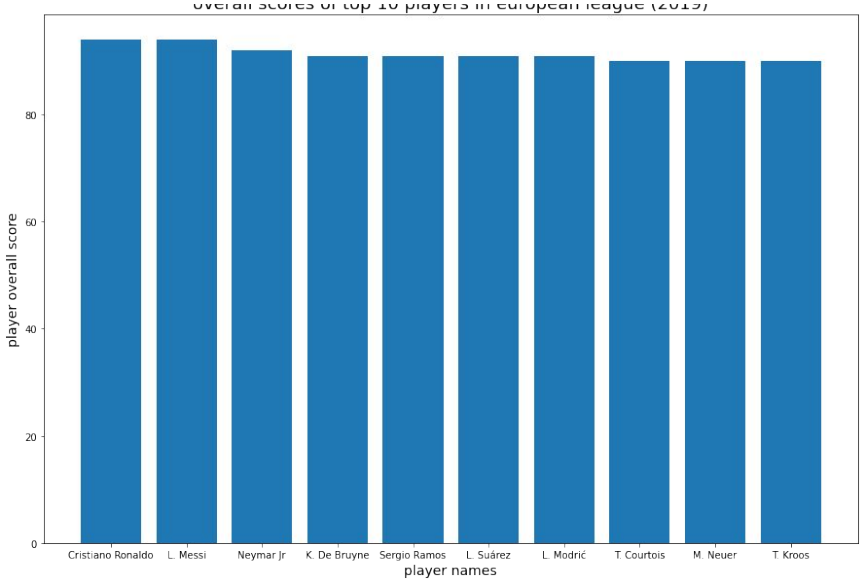
- Linear regression (Baseline)
- Lasso (regulates overfitting)
- MLP (4 hidden layers)

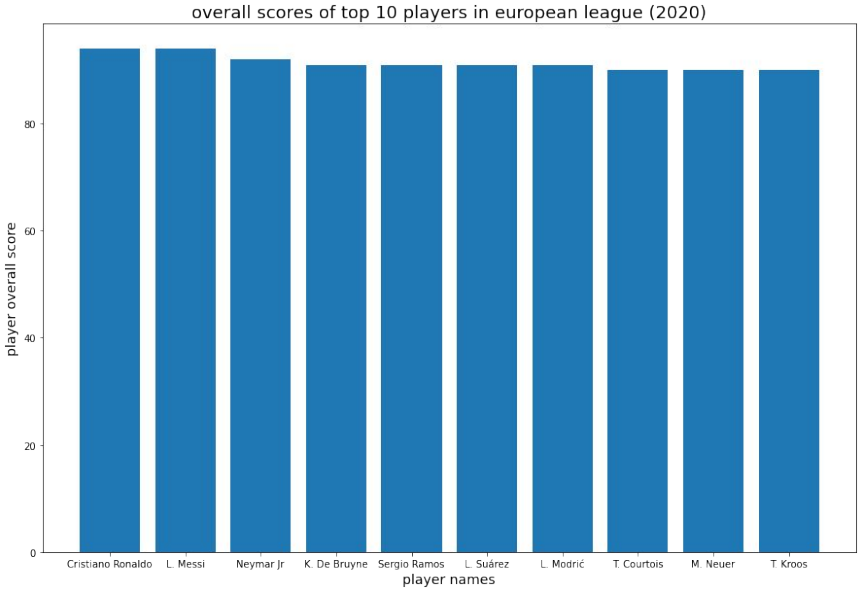
Methodology (Part A)

- Data Cleaning, Preprocessing
- Simple Linear Regression (Baseline)
- Overfitting
- Lasso Regularization
- MLP
- Removing correlated columns
- Final model

Results & Conclusion (Part A)

- Test_mse = 160, test_rmse = 13





Introduction (Part B)

Problem

- **Player position classification**
- **Can a professional defender become a forward?**

Data:

- Player positions
- Player attributes like attacking, goalkeeping, defense, skills, etc.
- Player personal data like Nationality, Club, Salary, Age, etc.

Approach

- Data Preparation - handling missing values, processing categorical variables, train-val-test splits
- Feature selection done to remove unnecessary features and similar features
- Classification Models - **neural networks with and without regularization**
- Comparison of the various neural network models

Reasoning

- We choose neural network over logistic regression for the classification problem
- Neural networks are more complex than logistic regression
- We have **over 90 features** in the dataset so neural networks can best represent this complexity

Methodology (Part B)

Data Preparation

- 100+ Features cleaned, missing values handled
- Categorical features one-hot encoded or target encoded
- Some players have multiple positions - we **choose the position with the highest rating**
- Removing irrelevant columns
- Added single feature for goalkeeper skill as it was missing
- Training-Validation split of 75-25

Player Position Classification

- Unregularized NN
- NN with Early Stopping
- NN with Dropout
- NN with Early Stopping and Dropout

NN Arch:
5 hidden layers
1 output layer - softmax
Adam Optimizer
Dropout layers added after all hidden layers
Early stopping patience = 10

Can a defender become a forward?

- Retrieve all defenders from FIFA 19 dataset
- Map them to their stats from FIFA 20 dataset
- Check player's position in FIFA 20 dataset
- Note down the number of defenders in 2019 who played at forward positions in 2020

Results & Conclusion (Part B)

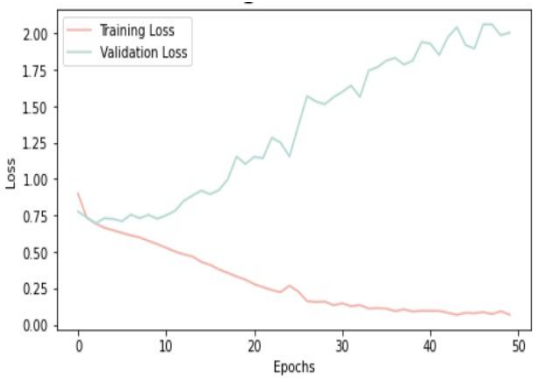
Player Position Classifier	Accuracy	Loss
Unregularized Neural Network	72.82%	1.78
Regularized NN - Early Stopping	74.22%	0.67
Regularized NN - Dropout	74.22%	0.77
Regularized NN - Dropout & Early Stopping	75.96%	0.63

Observations:

- Accuracy(unregularized) < Accuracy (regularized)
- Regularized NN with Dropout AND Early Stopping outperforms other models

Possible Future Improvements:

- Better feature selection can be done
- Deeper NN may increase performance



Can a defender become a forward?	
Total defenders in FIFA 2019	4387
2019 defenders that became forwards in FIFA 2020	4
Defender to Forward Conversion Rate	0.09%

Observations:

- Only about 0.1% of the defenders go on to become forwards
- Highly unlikely that defenders will become forward based on FIFA 19 and 20 data

Possible Future Improvements:

- Can see difference in skills of those defenders who went on to become forwards and those who did not.

Introduction (Part C)

Problem | Approach | Reasoning

Problem:

- Make a custom score based on historical data of players.
- Rank the clubs with best staffs.

Data:

- Player Overall Score
- Player attributes like weight, weak foot, work type, skills, and international reputation considered.

- Data cleaning
- Vectorization
- Data splitting into 1:4
- Feature importance
- Dropping insignificant features
- Making custom score on the historical data.
- Getting the results.

- Overall score, Potential, international reputation are directly related to the personal growth of the player
- However, weight of the player is not.

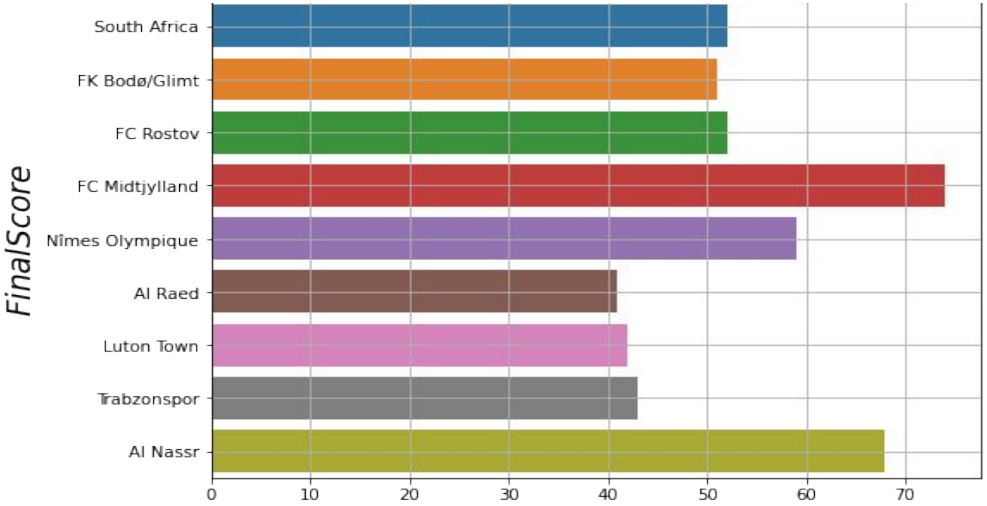
Methodology (Part C)

- **avg_wt** = (p_15_wt + p_16_wt + p_17_wt + p_18_wt + p_19_wt + p_20_wt) / 6
- **potential_trend** = (potential_2016 - potential_2015) + (potential_2017 - potential_2016) + (potential_2018 - potential_2017) + (potential_2019 - potential_2018) + (potential_2020 - potential_2019)
- **int_rel** = int_rel_15 + int_rel_16 + int_rel_17 + int_rel_18 * int_rel_19 + int_rel_20
- **wf** = wf_15 + wf_16 + wf_17 + wf_18 + wf_19 + wf_20
- **skill_moves** = sm_15 + sm_16 + sm_17 + sm_18 + sm_19 + sm_20

> **Overall_trend** = (overall_2016 - overall_2015) + (overall_2017 - overall_2016) + (overall_2018 - overall_2017) + (overall_2019 - overall_2018) + (overall_2020 - overall_2019)

```
Final Score = ((Overall_trend + Potential_trend)
* int_rel  * wf * skill_moves) / avg_wt
```

Results & Conclusion (Part C)



FC Midtylland and **Al Nasar** have turned out to be the best club with best staff with mean overall score with 74 and 68 score.

Notebook link : [P1.ipynb - Colaboratory (google.com)]