



Soccer Player Market Value Prediction – FIFA18 Dataset

Developer: Vincent Wang

Motivation

- › The most popular soccer simulation video game
- › Players Rating and Simulation System
- › Player Transfer Value Prediction

Problems

- › Implicit to rate player overall performance
- › Teams suffer losses when underestimate player value
- › Negotiation is unnecessarily long to determine player value

Impact

- › Create latent attributes to evaluate players
- › Model to predict player transfer value
- › Assist club manager to gain negotiation power
- › Offer wise transfer fee

Web App Demo

Player	Age	Wage	Actual wage	Overall Rating	Potential	Composure	Marking	Reactions	Vision	Volleys	Num Positions	Real Value	True Value	Result
Cristiano Ronaldo	32	€565K	565000	94	94	95	22	96	85	88	ST LW	€95.5M	95500000	€ 106,500,000.00
L. Messi	30	€565K	565000	93	93	96	13	95	90	85	RW	€105M	10500000	€ 99,290,000.00
Neymar	25	€280K	280000	92	94	92	21	88	80	83	LW	€123M	12300000	€ 111,425,000.00
P. Dybala	23	€215K	215000	88	93	84	14	84	84	88	ST CAM	€79M	7900000	€ 76,370,000.00
G. Bale	27	€370K	370000	89	89	85	51	87	79	76	RW	€69.5M	6950000	€ 52,570,000.00
L. Bonucci	30	€210K	210000	88	88	84	84	86	74	58	CB	€44M	4400000	€ 41,645,000.00
E. Hazard	26	€295K	295000	90	91	87	25	85	86	79	LW	€90.5M	9050000	€ 65,205,000.00



- › Player Attribute Data
 - › Ball control
 - › Crossings
 - › etc
- › Player Personal Data
 - › Age
 - › Potential
 - › etc
- › Player Playing Position Data
 - › Preferred Position

Datasets

Photo	Name	Age	Nationality	Club	Value
	Neymar	25	Brazil	Paris Saint-Germain	€123M
	L. Messi	30	Argentina	FC Barcelona	€105M
	L. Suárez	30	Uruguay	FC Barcelona	€97M
	Cristiano Ronaldo	32	Portugal	Real Madrid CF	€95.5M
	R. Lewandowski	28	Poland	FC Bayern Munich	€92M

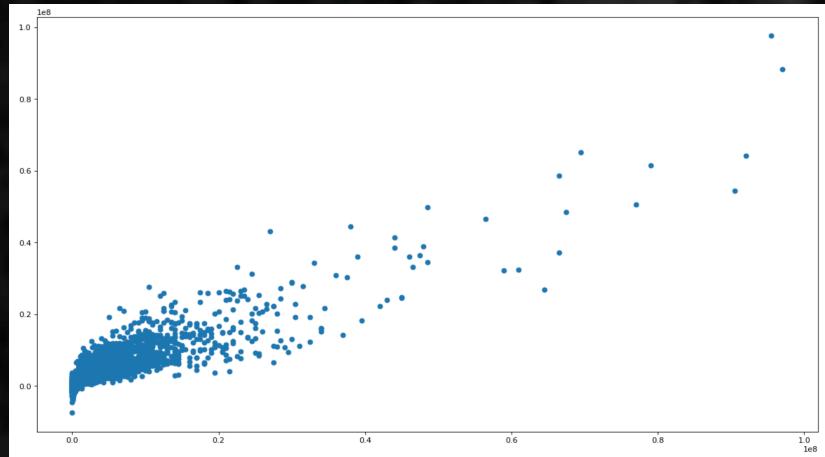
Top value players



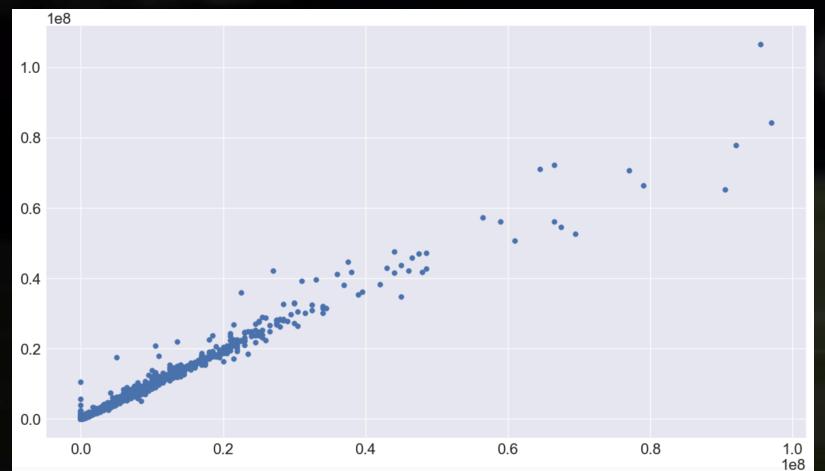
The relations of player's rating, value, and potential score



EDA



Linear
Regression



Random
Forest

Evaluation

MODELS

Linear Regression

Testing R²

0.784

Ridge Regression

0.444

Lasso Regression

0.778

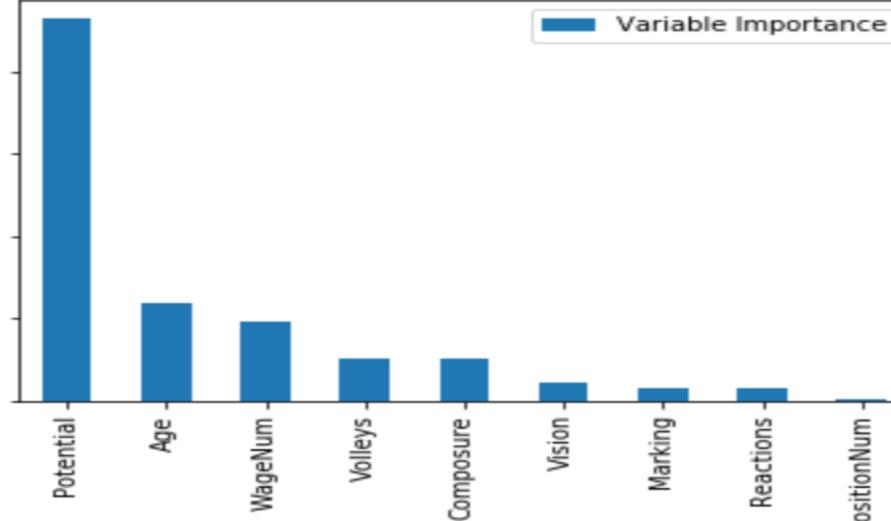
Random Forest

0.979

Neural Network

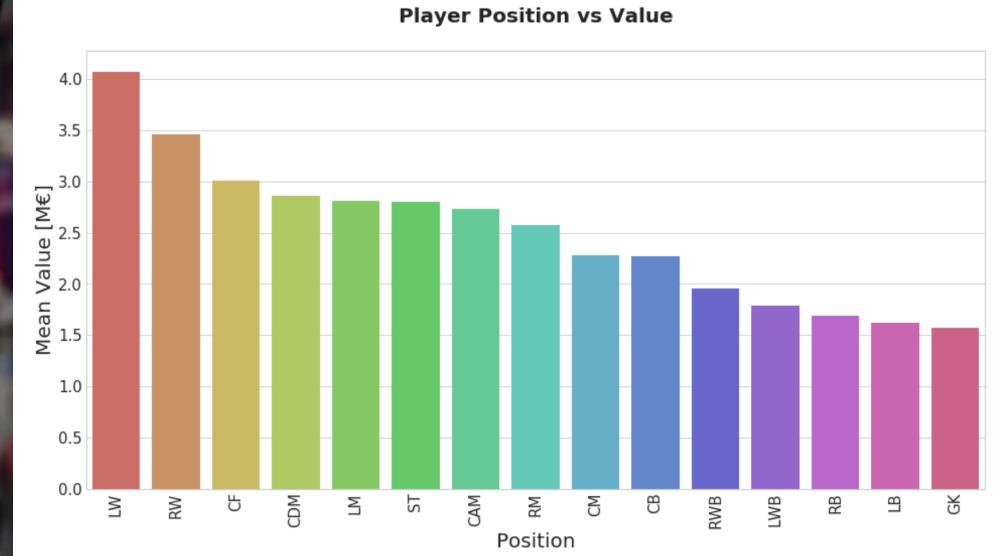
0.863

Algorithms

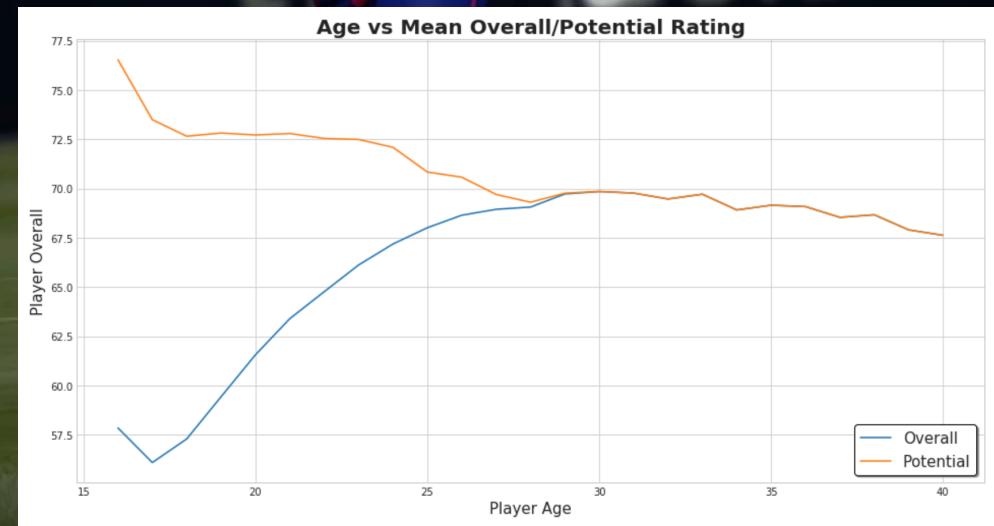


- › Potential and Age are the most influential attributes
- › Attacking attributes are more important than defense attributes
- › Position flexibility is not important

Variable Importance



Position “Discrimination”



Age influence on rating and potential

Insights

Vincent Wang – Master of Science in Analytics' 18

Responsibilities :

- Application Developer
- Data Scientist



-
- As a passionate data detective using cutting edge data science techniques to find the impactful, game-changing business insights, Vincent is currently pursuing a Master of Science in Analytics (MSiA) at Northwestern University.
 - Through collaboration with colleagues at MSiA, Vincent is studying holistic data science toolkits with machine learning methodologies through interdisciplinary curriculum and hands on industry projects.
 - Vincent is thrilled to combine careful analysis and wild creativity to optimize the solutions and contribute to the real-world business with the power of data science.



Thank you!