

**Московский авиационный институт
(Национальный исследовательский университет)**

Факультет: «Информационные технологии и прикладная математика»
Кафедра: 806 «Вычислительная математика и программирование»
Дисциплина: «Искусственный интеллект»

Лабораторная работа № 1
Тема: Azure machine learning

Студент: Шевчук П.В.

Группа: М80-304Б

Преподаватель: Ахмед Самир Халид

Москва, 2019

1. Постановка задачи

Ваша задача познакомиться с платформой Azure Machine Learning, реализовав полный цикл разработки решения задачи машинного обучения, используя три различных алгоритма, реализованные на этой платформе.

2. Требования

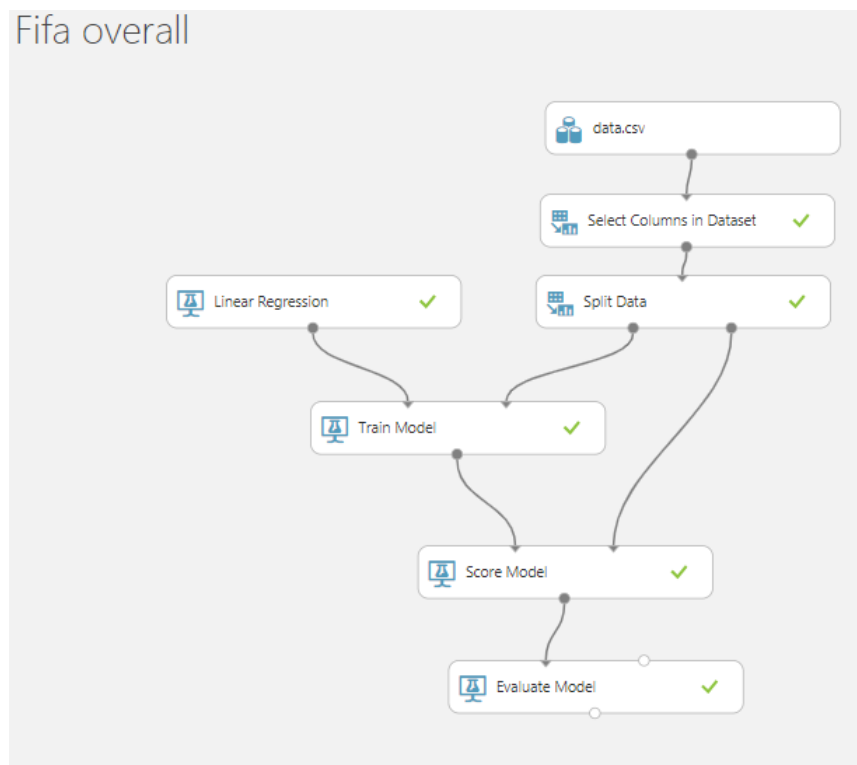
- 1) Уникальность решения
- 2) Обоснованность выбора той или иной операции
- 3) В отчете должны быть указаны алгоритмы, которые применялись, результаты применения этих алгоритмов, а также скрины некоторых этапов обработки данных

3. Решение задачи

Датасет: <https://www.kaggle.com/karangadiya/fifa19>, данные об игроках FIFA 19

1 алгоритм: Определить рейтинг игрока с помощью его данных о возрасте и цене.





Выбираем интересующие поля: рейтинг, цена, возраст и имя игрока (для удобства). Для обучения алгоритма воспользуемся классом линейной регрессии, т.к. рейтинг почти линейно зависит от цены. Подадим 80% данных на обучение, а остальные 20% на проверку. В результате видим, что предсказанные значения не сильно отличаются от действительного результата.



Fifa overall > Score Model > Scored dataset

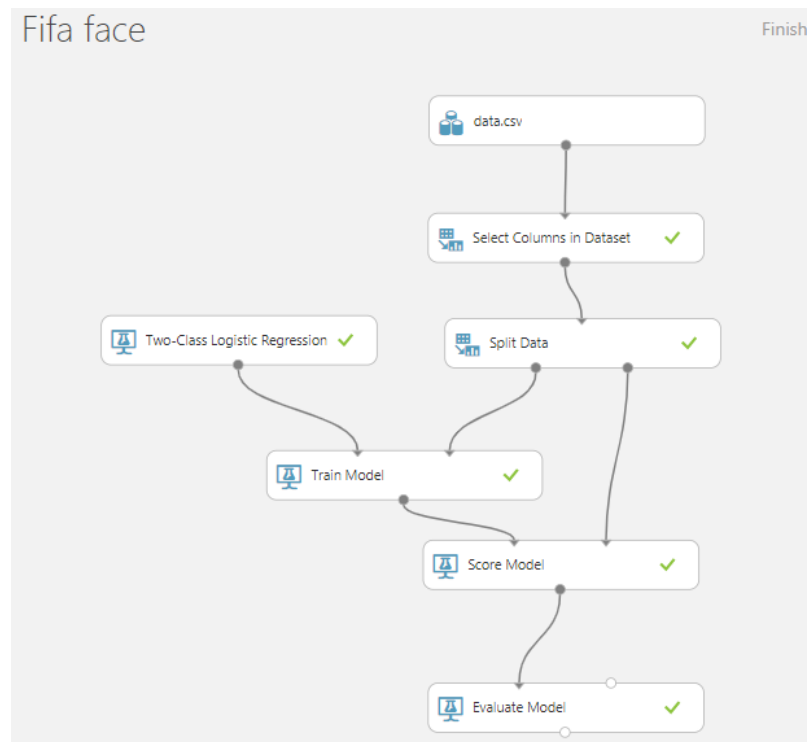
rows
9103

columns
5

	Name	Age	Overall	Value	Scored Labels
view as  					
	C. Buonavito	25	70	€2.1M	69.671763
	L. McGee	22	64	€550K	63.050988
	Luo Senwen	25	63	€450K	63.246687
	B. Morris	22	64	€750K	65.549849
	B. Vazquez	19	63	€700K	63.271361
	J. Sparkes	17	52	€100K	53.618792
	J. Donachie	25	67	€775K	66.057609
	S. Doumbia	21	60	€290K	58.499969
	E. Cerna	20	61	€475K	61.963256
	L. Štetina	26	72	€3.1M	72.224717
	A. Bordin	19	61	€425K	60.083018
	Danilo	28	68	€900K	70.287063

2 алгоритм: Определить, есть ли gameface у игрока с помощью его данных о возрасте, цене.

Надо предсказать категориальный признак, состоящий из двух вариантов: “Да” и “Нет”. Для этого воспользуемся методом two-class logistic regression, т.к. опять предполагается линейная зависимость от рейтинга. В результате мы видим, что большинство игроков не имеют gameface, но есть игроки с не самым большим рейтингом, которые имеют его. Это может быть связано с историей игрока или с другими причинами, поэтому алгоритм не может в точности дать верный результат при исходных данных.



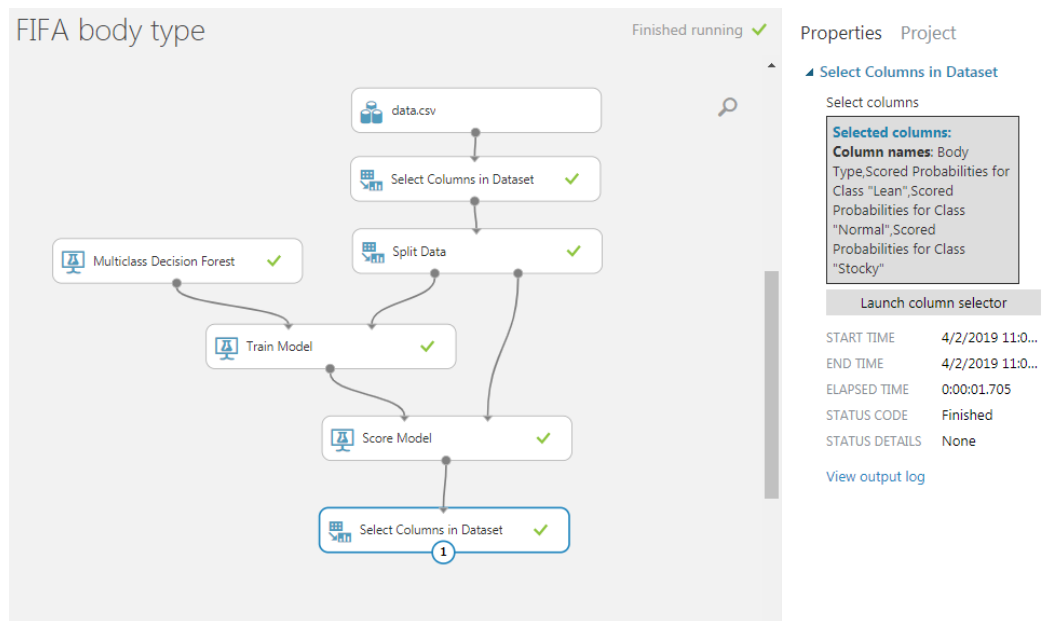
Fifa face > Score Model > Scored dataset

rows 3641
columns 7

	Name	Age	Overall	Value	Real Face	Scored Labels	Scored Probabilities
view as							
	S. Żurkowski	20	72	€5.5M	No	No	0.121197
	J. Zacařía	28	69	€1.1M	No	No	0.062472
	F. Langsdorf	22	53	€80K	No	No	0.002354
	C. Tătarușanu	32	79	€7M	No	No	0.436939
	R. Swan	22	55	€120K	No	No	0.002712
	Romarinho	27	72	€3.3M	No	No	0.141592
	Aguilera	32	69	€600K	No	No	0.095097
	M. Brink	20	58	€230K	No	No	0.006497
	T. Oar	26	72	€3.3M	No	No	0.138027
	Angeliño	21	77	€10.5M	Yes	No	0.260269
	G. Gönül	33	76	€3M	No	No	0.43446
	T. Monconduit	27	76	€7.5M	No	No	0.281065






3 алгоритм: Определить, какая вероятность того или иного телосложения у игрока.

Выбираем поля: возраст, рейтинг, рост, вес и вид телосложения. Так как категориальный признак имеет 3 варианта, то воспользуемся методом multiclass decision forest. Этот метод быстро изучает машину и работает с высокой точностью. Выведем признаки, относящиеся только к типу телосложения. Наблюдаем вероятность одного из трёх видов, которые может получить игрок.



FIFA body type > Select Columns in Dataset > Results dataset

rows 3641
 columns 4

	Body Type	Scored Probabilities for Class "Lean"	Scored Probabilities for Class "Normal"	Scored Probabilities for Class "Stocky"
view as 				
	Normal	0.352329	0.632043	0.015627
	Normal	0.139317	0.816806	0.043877
	Normal	0.267474	0.718509	0.014017
	Lean	0.121446	0.733089	0.145466
	Normal	0.466705	0.524633	0.008662
	Lean	0.588389	0.402196	0.009415
	Lean	0.468887	0.501031	0.030081
	Normal	0.276113	0.705276	0.018611
	Lean	0.655724	0.210297	0.133979
	Stocky	0.413799	0.560785	0.025417
	Lean	0.063867	0.90106	0.035074
	Normal	0.035714	0.710714	0.253571

4. Вывод

В ходе данной лабораторной работы освоен алгоритм работы с Azure machine learning. С помощью этого облачного сервиса я научился прогнозировать разные признаки. Надеюсь, эта система поможет мне для дальнейшей работы в машинном обучении.