

Отчет по лабораторной работе №1 по курсу «Искусственный интеллект»

Выполнил студент группы М8О-3046-16 Величутин Андрей

Тема: Azure ML

Задача:

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

Оборудование студента:

Intel(R) Core(TM) i5-5200 CPU @ 2.20GHz 2.19GHz
ОЗУ 6,00 ГБ

Программное обеспечение:

Windows 10, Python 3.7.4(С библиотеками Pandas, Numpy, Seaborn и Scikit-Learn), Jupyter notebook 4.4.0

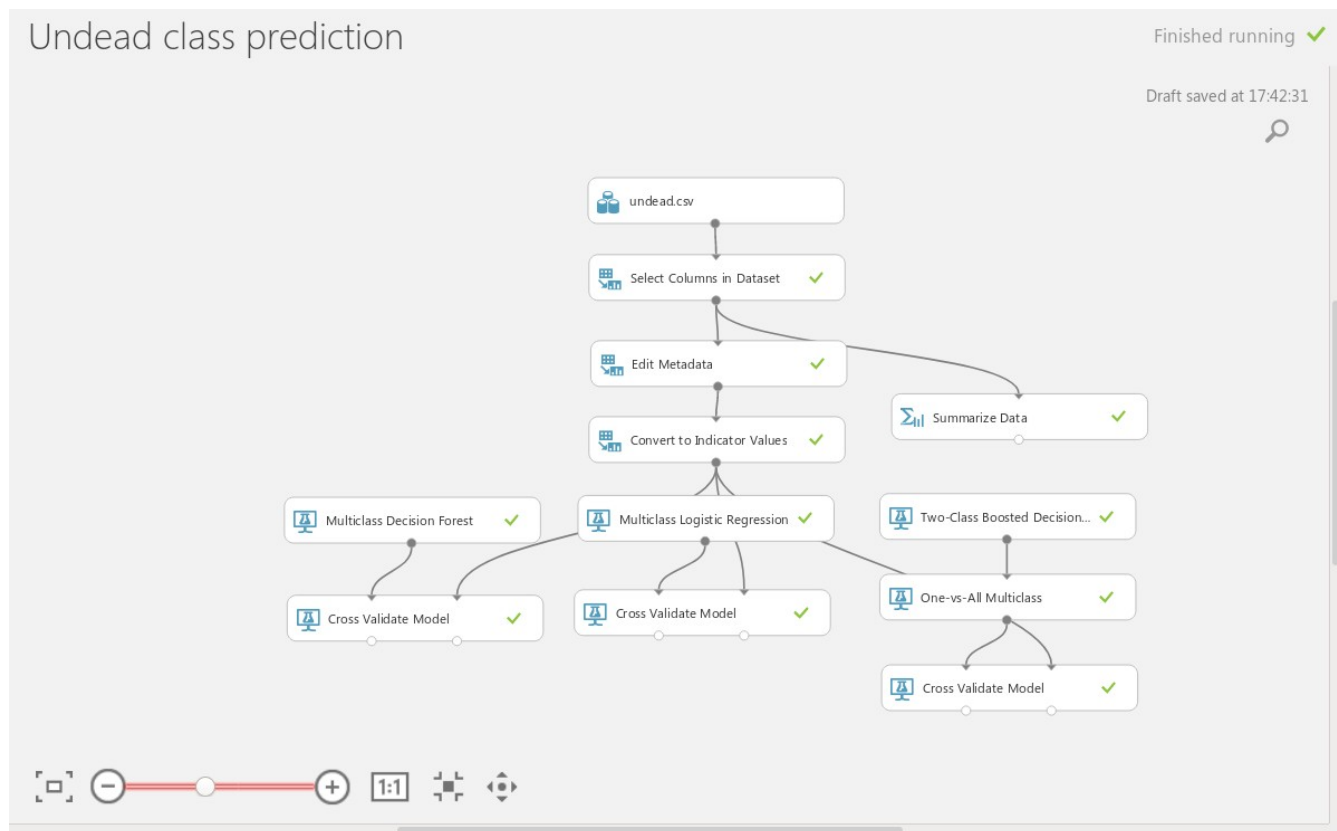
Ход работы:

Azure Machine Learning – облачный сервис для выполнения задач прогнозной аналитики, с помощью которого можно легко создавать модели и интегрировать их в промышленные решения.

Среди особенностей данной платформы можно отметить:

- Интуитивно понятный графический интерфейс
- Готовые для работы реализации самых популярных алгоритмов машинного обучения
- Широкие возможности предобработки данных
- Отличная визуализация всех результатов
- Облачные вычисления
- Возможность конвертации в формат ноутбука на Python

На одном поле можно отобразить почти весь цикл решения задачи машинного обучения. Вот так, например, выглядит весь процесс работы с датасетом из моей предыдущей лабораторной работы:



Весь процесс работы свелся к тому, что я расставлял готовые блоки алгоритмов. И смотрел на результаты работы каждого из них.

Исходный датасет (Блок undead.csv):

Undead class prediction > undead.csv > dataset

rows
371

columns
7

	id	bone_length	rotting_flesh	hair_length	has_soul	color	type
view as							
	0	0.354512	0.350839	0.465761	0.781142	clear	Ghoul
	1	0.57556	0.425868	0.531401	0.439899	green	Goblin
	2	0.467875	0.35433	0.811616	0.791225	black	Ghoul
	4	0.776652	0.508723	0.636766	0.884464	black	Ghoul
	5	0.566117	0.875862	0.418594	0.636438	green	Ghost
	7	0.40568	0.253277	0.44142	0.280324	green	Goblin
	8	0.399331	0.568952	0.618391	0.467901	white	Goblin
	11	0.516224	0.536429	0.612776	0.468048	clear	Ghoul
	12	0.314295	0.67128	0.417267	0.227548	blue	Ghost
	19	0.280942	0.701457	0.179633	0.141183	white	Ghost
	22	0.431685	0.438959	0.239212	0.47182	clear	Goblin
	23	0.584543	0.593082	0.681166	0.935721	clear	Ghoul

>
Statistics

Visualizations



Подсчет различных статистик для датасета (Блок Summarize Data):

Undead class prediction > Select Columns in Dataset > Results dataset

rows
6

columns
22

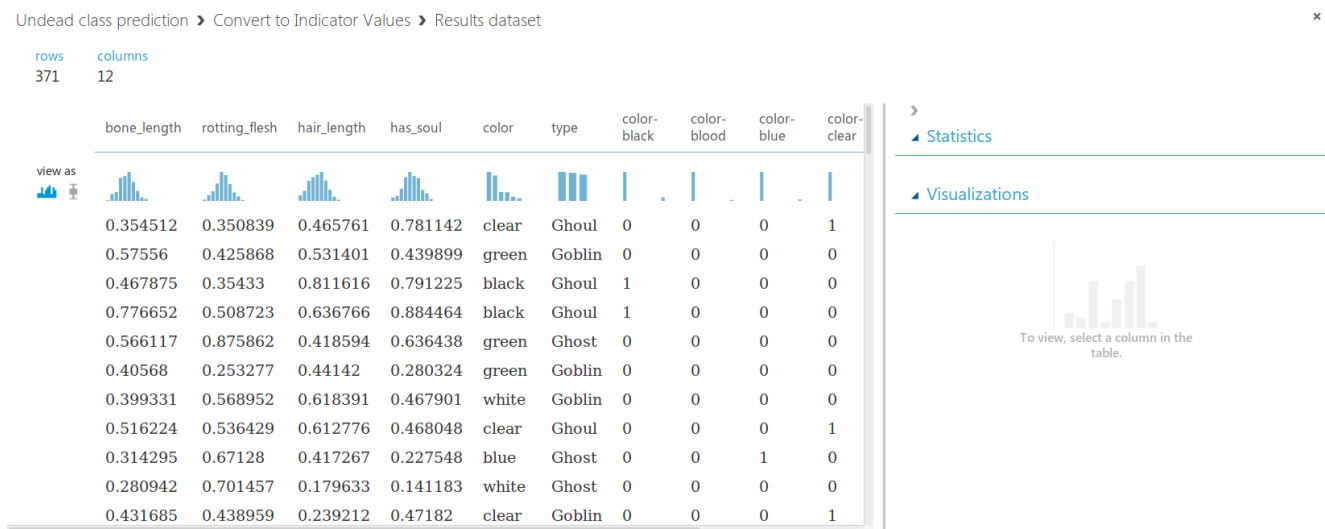
	Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile
view as									
	bone_length	371	371	0	0.061032	0.817001	0.43416	0.10731	0.340006
	rotting_flesh	371	371	0	0.095687	0.932466	0.506848	0.116092	0.414812
	hair_length	371	371	0	0.1346	1	0.529114	0.138535	0.407428
	has_soul	371	371	0	0.009402	0.935721	0.471392	0.142226	0.348002
	color	371	6	0					
	type	371	3	0					

>
Statistics

Visualizations



Конвертация категориальных признаков в индикаторные значения (Блок Convert to Indicator Values):



Следующие блоки отвечают непосредственно за обучение моделей и оценку качества на кросс-валидации по 10-ти блокам. В данной лабораторной работе я использовал следующие алгоритмы:

- Мультиклассовая логистическая регрессия
- Мультиклассовый случайный лес
- Один-Против-Всех бустинг над решающими деревьями

Чтобы не нагромождать отчет множеством скриншотов, я выписал среднюю оценку результатов работы для каждого класса:

Модель	Precision для класс Ghost	Recall для класса Ghost	Precision для класса Ghoul	Recall для класса Ghoul	Precision для класс Goblin	Recall для класса Goblin
Логрег	0.827963	0.886969	0.728761	0.799573	0.669113	0.560759
Случайный лес	0.819128	0.8388	0.738211	0.762051	0.632464	0.578708
Бустинг	0.808901	0.870467	0.761365	0.730214	0.60163	0.573378

Желтым выделены наилучшие значения в каждом столбце. Как можно заметить, логистическая регрессия показала в основном результаты лучше, чем случайный лес или бустинг. Возможно это произошло, так как данных в датасете не так много, а тот же бустинг показывает хорошие результаты как раз на больших данных. С другой стороны, скорее всего подкрутив параметры случайного леса или бустинга, можно добиться большей точности и полноты

Выводы:

В данной ЛР я познакомился с облачной платформой Azure ML. Лично мне она показалась достаточно удобной. Мне кажется, что пользователь без каких либо навыков программирования, с её помощью способен решить задачу машинного обучения. Сами же решения выглядят очень компактно и наглядно. Данный инструмент может очень пригодиться как начинающим аналитикам так и профессионалам