

## ПРАКТИЧНИЙ ПРИКЛАД ВИКОРИСТАННЯ MICROSOFT AZURE MACHINE LEARNING STUDIO

Наразі стан інформатизації суспільства досягнув того, що інновації буквально заповнили всі сфери життєдіяльності. Вражаючі темпи оновлення технологій змушують наукову спільноту миттєво реагувати на виклики сьогодення. Приєднання України до Болонської системи освіти зумовило її докорінну модернізацію у сфері інформатизації освітнього простору.

Новітні технології пронизують усі рівні й аспекти педагогічної діяльності – від застосування інформаційних технологій навчання під час викладання певної дисципліни до впровадження систем управління вищим навчальним закладом.

Серед новітніх технологій чинне місце посідають «хмарні» технології, які все частіше проникають у вітчизняну систему освіти. Дійсно, освітня галузь не залишається осторонь процесів оновлення, а одним із шляхів вирішення проблеми взаємодії кількох віддалених систем підтримки навчального процесу, їх мобільності й економічності є використання «хмарних» обчислень, коли ресурси для опрацювання даних надаються кінцевим користувачам в якості інтернет-сервісу.

Розглянемо докладніше використання Microsoft Azure Machine Learning Studio.

У програмний продукт інтегровано велику кількість статистичних моделей (Рис. 1), перевірених мільйонами користувачів. На першому етапі вивчення Data Science немає необхідності досконально розбиратися в цих моделях, досить розуміти область застосування кожної і вибрати кращу.

Для початку роботи з програмою і перевірки її можливостей був згенерований найпростіший набір даних, що містить 3 змінні color, mileage, crash.

Від пробігу аварія не залежить, а залежить тільки від кольору автомобіля. Використовувалися наступні кольори (перераховані в порядку ймовірності аварії): white; black; green; blue; violet; red; orange; yellow.

Експорт набору даних в ML Studio.

Дані в Azure ML Studio легко експортуються з будь-якого джерела за адресою в інтернеті (URL).

Візуалізація даних. На початковому етапі можна «подивитися» на вхідні дані і, якщо пощастить, «побачити» деякі закономірності, які можуть бути корисні в процесі подальшого аналізу.

---

<sup>1</sup> Студент групи М КН 2020-1, ХНУМГ ім. О. М. Бекетова

<sup>2</sup> Доцент кафедри КНтаІТ ХНУМГ ім. О. М. Бекетова

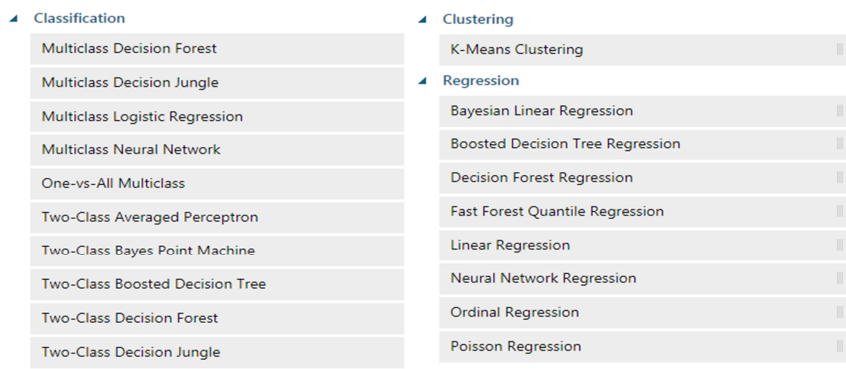


Рисунок 1 – Статистичні моделі Microsoft Azure Machine Learning Studio

Загальний вигляд вхідних даних представлено на Рис. 2.

The screenshot shows the data view for a dataset named 'tr\_2\_50000.csv'. The dataset has 50,000 rows and 3 columns: 'color', 'mileage', and 'crash'. Below the summary, there are small bar charts for each column and a table of sample data.

color	mileage	crash
red	507966	0
orange	156470	0
orange	619959	0
violet	602456	0
violet	820961	0
red	584710	0
blue	560267	0

Рисунок 2 – Вхідні дані.

Вибір і тренування моделі на першому наборі даних. Обрана модель: Two-Class Bayes Point Machine (Рис. 3).

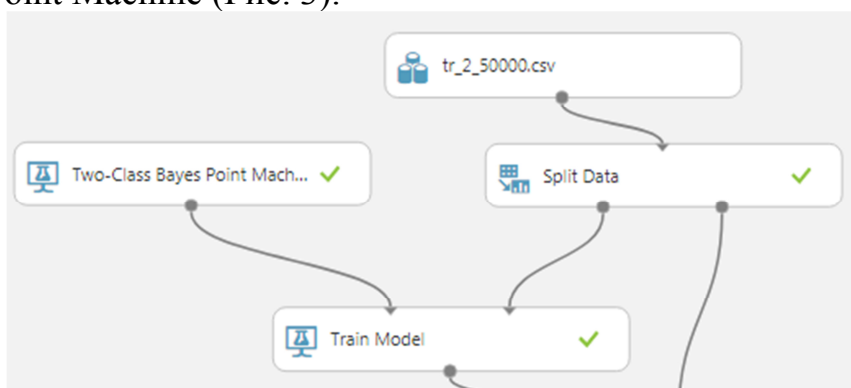


Рисунок 3 – Вибір і тренування моделі.

На цьому етапі можна оцінити значимість (ваги) змінних. В даному випадку вже на першому етапі можна визначити, що змінна 'mileage' практично не має значення (вага близька до нуля). Найрідше в аварію потрапляють машини червоного, жовтого і помаранчевого кольору, найчастіше – білого і чорного (Рис. 4).

### Feature Weights

Feature	Weight Mean	Weight Standard Deviation
color_red_4	-1.09328	0.0279437
color_yellow_7	-1.08931	0.0275149
color_orange_3	-1.06011	0.0272349
color_black_0	0.819723	0.0256965
color_white_6	0.806502	0.0258081
color_blue_1	-0.507002	0.0238587
color_green_2	-0.459744	0.0226556
color_violet_5	-0.0723266	0.022419
Bias	-0.00253621	0.00888862
mileage	3.65604e-8	1.45724e-8
color#unknown_9	0	39537100000000

Рисунок 4 – Значимість (ваги) змінних

Azure ML Studio будує і виводить на екран (Рис. 6) ROC-криву (receiver operating characteristic), що дозволяє оцінити якість бінарної класифікації і відображає співвідношення між часткою об'єктів від загальної кількості носіїв ознаки, вірно класифікованих що несуть ознаку.

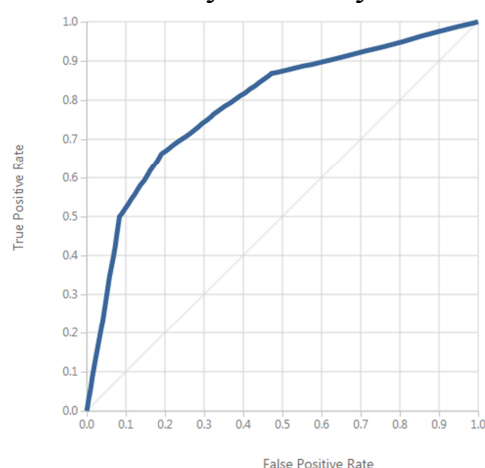


Рисунок 5 – ROC-крива

Якість прогнозу (класифікації) оцінюється і за допомогою таблиці спряженості (Рис. 6).

True Positive	False Negative
<b>4988</b>	<b>4997</b>
False Positive	True Negative
<b>1259</b>	<b>13756</b>
Positive Label	Negative Label
<b>1</b>	<b>0</b>

Рисунок 6 – Оцінка якості прогнозу.

### ПЕРЕЛІК ПОСИЛАНЬ:

1. Tan P-N., Steinbach M., Karpatne A. and Kumar V. “Introduction to data mining”. Pearson; 2nd edition. 2018. 864p.
2. Abu-Mostafa S., Magdon-Ismail M., Lin H-T. “Learning from data”. AMLBook. 2012. 213p.