**Таблиця результатів лабораторної 6**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Метод відбору | Пул  (%) | Точність до відбору (%) | Точність після відбору (%) | Час виконання (секунди) | Переваги | Недоліки |
| Least Confidence | 90 | 33,5 | 86,05 | 1109.23 | Легко реалізувати та використовувати. | Ризик вибору неякісних зразків замість інформативних  Обмежена адаптивність до специфіки конкретної задачі |
| 75 | 71 | 88,8 | 1425.52 |
| 50 | 77,6 | 88,7 | 2059.77 |
| 25 | 76,8 | 91,4 | 3254.40 |
| 10 | 84,1 | 91,4 | 4172.54 |
| Margin Sampling | 90 | 63,6 | 85,65 | 1060.48 | Ефективна для моделей, коли різні класи близькі за характеристиками. | Менш ефективна на малих датасетах або при розбалансованих класах. |
| 75 | 64,8 | 89,1 | 1403.56 |
| 50 | 77 | 90,4 | 2456.24 |
| 25 | 85 | 89,6 | 2842.17 |
| 10 | 86,85 | 90 | 3949.39 |
| Ratio Confidence | 90 | 58,8 | 83,5 | 968.28 | Якщо модель може помилятися між кількома схожими класами, ця стратегія може вибрати більш інформативні зразки | Є складнішою в реалізації і менш ефективною для класифікації на простих класах. |
| 75 | 72,5 | 84,6 | 1517.56 |
| 50 | 80,6 | 88,3 | 2136.65 |
| 25 | 80,3 | 89,8 | 2702.82 |
| 10 | 79,8 | 92 | 7742.25 |
| Entropy | 90 | 41,8 | 82,1 | 973.18 | Враховує всю невпевненість моделі щодо всіх класів, що робить її дуже інформативною | Найбільш обчислювально складна з усіх стратегій. |
| 75 | 69,7 | 88,3 | 1346.44 |
| 50 | 76,4 | 91,5 | 2423.97 |
| 25 | 77,4 | 89,6 | 3006.57 |
| 10 | 84,7 | 86,3 | 4104.26 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Пояснення + Опис власної проблеми**

З цих результатів можна зробити висновок що при збільшенні відсодковості пулу (відносно навчальних даних) точність падає, але швидкість набагато краща.Аналізуючи експеримент я б обрала б пул 50% або 90%, тому що при 50 відсотках активно піднялась точність(надалі вона не сильно змінюється, а десь навіть впала) і середній час очікування відносно інших. 90 ж хороший своєю швидкістю та достатній по точності(вона підійматися дуже швидко відносно ітерацій).

Стосується 10 % пула : Результати не ідеальні.По графіках можна зрозуміти що точність стрибає.Пояснити на 100 відсотків не можу.

Можливо модель перенавчається тому точність регресує, але не зрозуміло чому least\_confidence на графіку при першій ітерації так спустилась зразу(далі такі інциденти теж були саме з least\_confidence).

Надалі буду виводити точність на кожній ітерації, щоб можна було простежити чому це відбувається (не планую знов перенавчати модель на 10% пулу, бо це зайняло майже 6 годин, не знаю чому, я пробувала ще всілякі методи, але я не хочу прям зовсім зменшувати обсяг даних, бо тоді точність буде ще менше)

PS.Я намагаюся виконати цю лабу вже цілий тиждень, постійно знаходила десь помилки в моїй реалізації алгоритму і правила тому була необхідність знов перенавчати.Я питалася одногрупників скільки часу в них займає навчання моделі на вісіх відсодковостях пулу то вони казали від 40 хв до 80 хв.В мене ж тільки одне проходження зайняло 6 годин(дата сети в нас однакові).Я не розумію в чому проблема, якщо б ви мали ще якісь поради напишіть в коментарях під лабою.

Далі що я зобила(додала): Було додано зупинку на рівні епох.Логіка така : якщо модель протягом 2 епох регресує в точності, то навчання зупиняється і береться та де найкращі результати до цього.Це допомогло передбачити зниження точності при навчанні.