

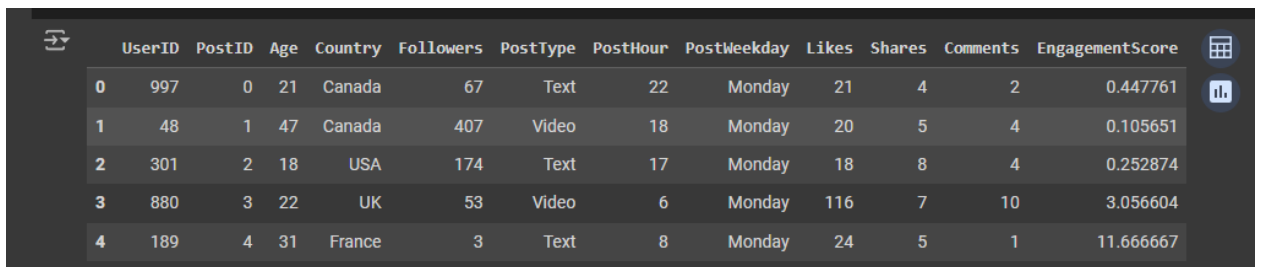
## Завдання

1. **Зрозуміти алгоритм:** Визначити, як обчислюється EngagementScore на основі різних характеристик користувача та посту. Дослідити, які фактори впливають на цей скор.
2. **Прогнозування:** Оцінити, чи можна точно передбачити EngagementScore для новостворених постів.

## Методологія

1. **Збір та підготовка даних:**

Завантаження даних та їх попередній огляд.



	UserID	PostID	Age	Country	Followers	PostType	PostHour	PostWeekday	Likes	Shares	Comments	EngagementScore
0	997	0	21	Canada	67	Text	22	Monday	21	4	2	0.447761
1	48	1	47	Canada	407	Video	18	Monday	20	5	4	0.105651
2	301	2	18	USA	174	Text	17	Monday	18	8	4	0.252874
3	880	3	22	UK	53	Video	6	Monday	116	7	10	3.056604
4	189	4	31	France	3	Text	8	Monday	24	5	1	11.666667

2. **Аналіз даних:**

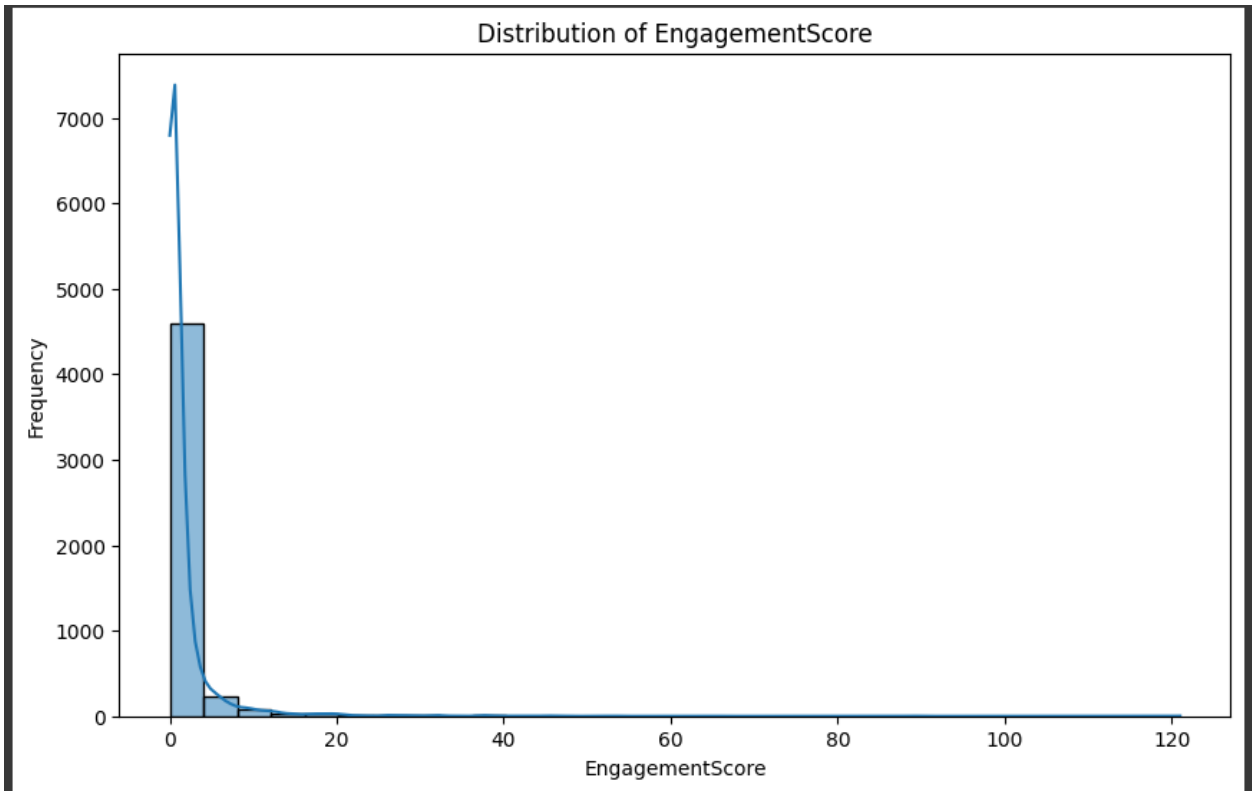
- Визначення кореляції між EngagementScore та іншими характеристиками.
- Дослідження розподілу EngagementScore.

3. **Моделювання:**

- Використання різних алгоритмів машинного навчання для прогнозування EngagementScore.
- Оцінка точності моделей за допомогою метрик (RMSE, MAE,  $R^2$ ).

# Аналіз даних

Дослідження розподілу EngagementScore.



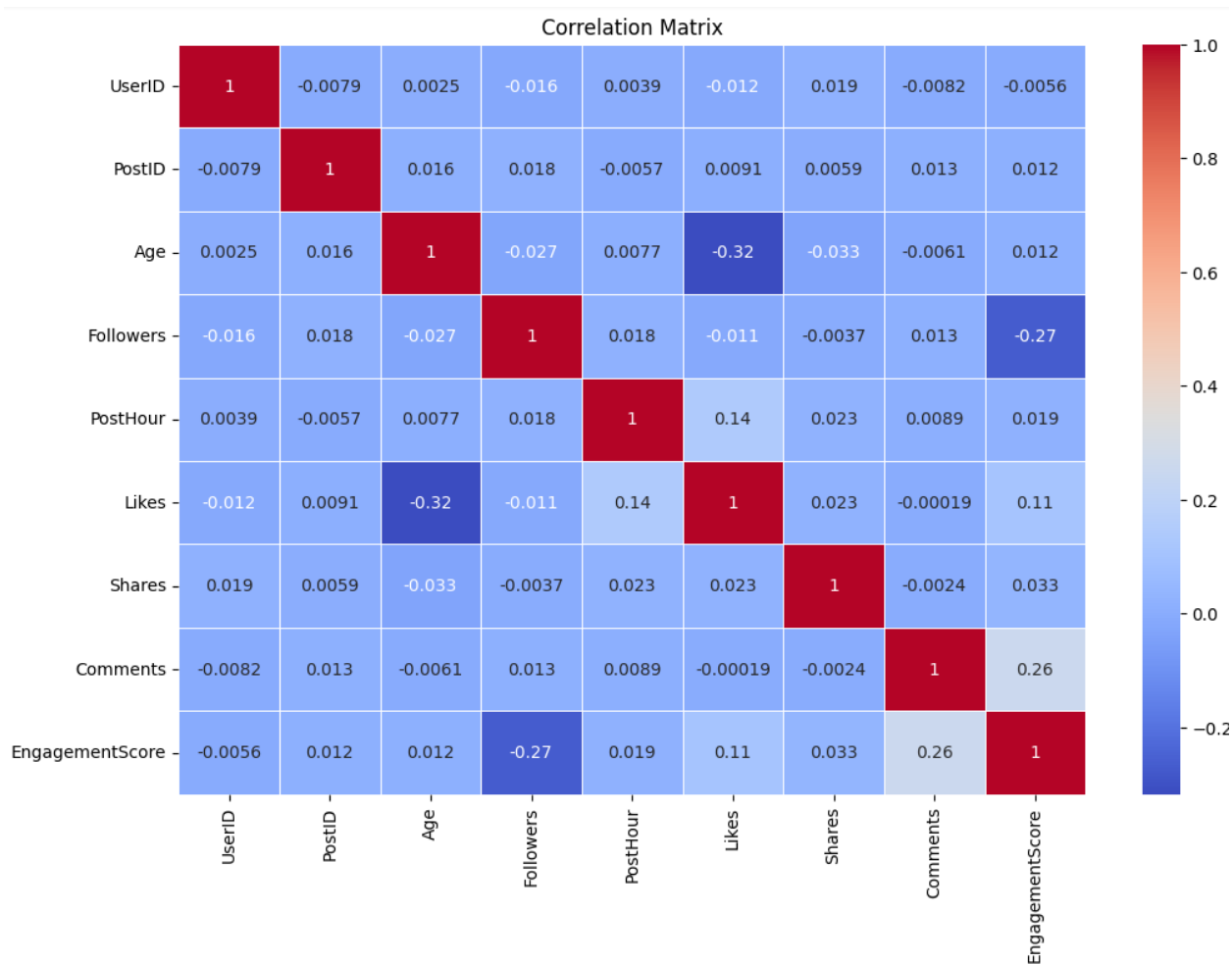
	Likes	Shares	Comments	EngagementScore
count	5000.000000	5000.000000	5000.000000	5000.000000
mean	28.977600	4.965400	15.295200	1.547303
std	27.176946	2.245353	21.652303	4.254233
min	7.000000	0.000000	0.000000	0.017450
25%	17.000000	3.000000	2.000000	0.181470
50%	20.500000	5.000000	4.000000	0.440183
75%	24.000000	6.000000	8.000000	1.233675
max	177.000000	17.000000	61.000000	121.000000

Значення розподілу мають дуже широкий діапазон від 0 до 121, при цьому середє значення дорівнює 1.54

**Визначення кореляції між EngagementScore та іншими характеристиками.**

Матриця кореляцій: Є помітна кореляція між деякими змінними, наприклад, Likes, Shares, Comments мають позитивну кореляцію з EngagementScore, при цьому від’ємну кореляцію створює

змінна Followers.



Моделювання

Для прогнозування EngagementScore було використано кілька моделей:

- 1. Лінійна регресія
- 2. Градієнтний бустинг

Висновки

- **Основні фактори:** Є помітна кореляція між деякими змінними, наприклад, Likes, Shares, Comments мають позитивну кореляцію з EngagementScore, при цьому від’ємну кореляцію створює змінна Followers.
- **Прогнозування:** Модель градієнтного бустингу показала найкращу точність з  $R^2 = 0.994$ , У моделі були використані такі параметри:
  - ('regressor\_\_learning\_rate': 0.2,
  - 'regressor\_\_max\_depth': 3,
  - 'regressor\_\_min\_samples\_leaf': 1,
  - 'regressor\_\_min\_samples\_split': 10,

- 'regressor\_\_n\_estimators': 300,
- 'regressor\_\_subsample': 1.0},
- MSE = 0.05777943852321589,
- $R^2 = 0.9942550462431922$ )

що свідчить про високу ймовірність точного прогнозування EngagementScore для нових постів.