

Revenue: Q1 2020

\$232M  
Q4 2019: \$1.87M

Expense: Q1 2020

\$47.6K  
Q4 2019: \$148.90K

Profit: Q1 2020

\$186.4K  
Q4 2019: \$471.63K

Revenue per Employee

\$110,917

Active Customers

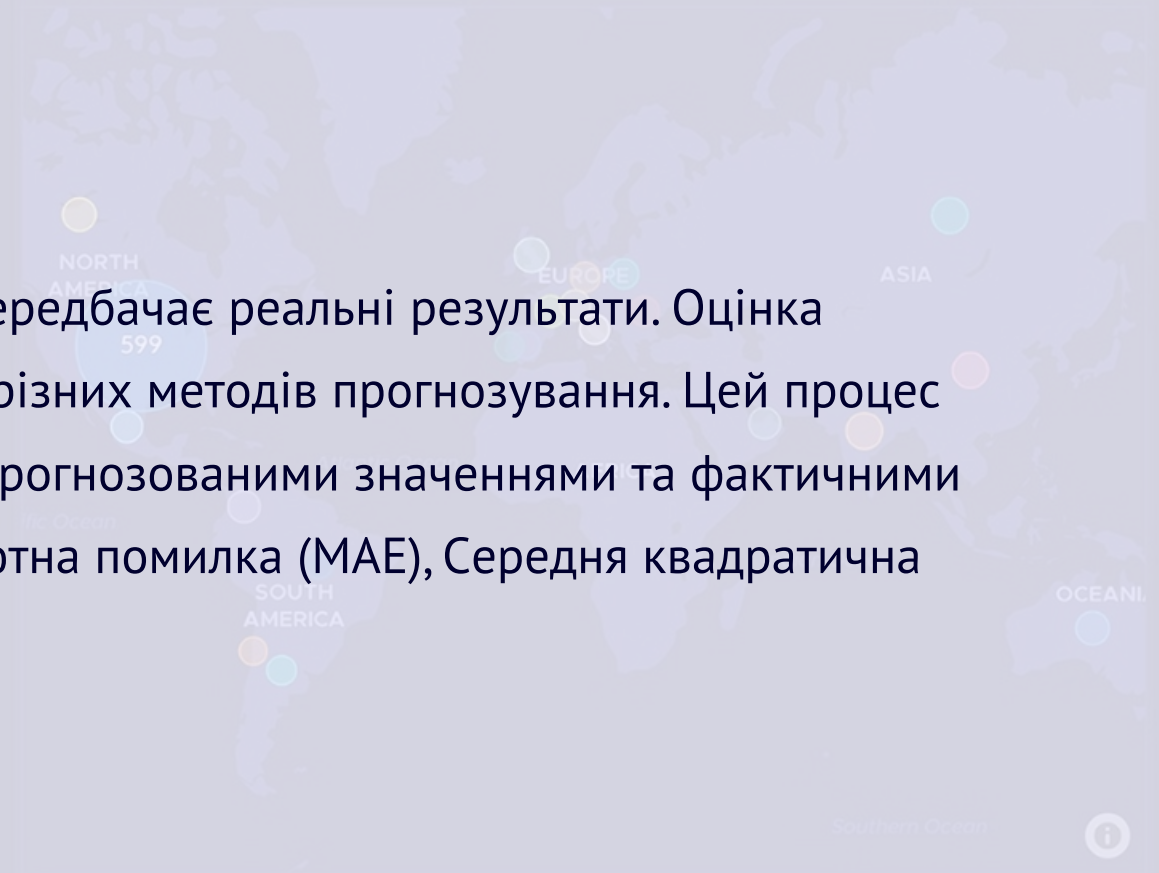
5.63K

# Оцінка точності прогнозування:

## показники та порівняння

### моделей

У прогнозуванні надзвичайно важливо оцінити, наскільки добре модель передбачає реальні результати. Оцінка точності гарантує надійність моделі та допомагає порівняти ефективність різних методів прогнозування. Цей процес передбачає розрахунок показників помилки, які вимірюють різницю між прогнозованими значеннями та фактичними значеннями. До поширених показників помилки належать Середня абсолютна помилка (MAE), Середня квадратична помилка (MSE) та Середньоквадратична помилка (RMSE).



# Показники помилки

Показники помилки використовуються для кількісної оцінки точності прогнозної моделі, вимірюючи різницю між прогнозованими значеннями та фактичними значеннями. Ось три поширені показники помилки:

- Середня абсолютна помилка (MAE)
- Середня квадратична помилка (MSE)
- Середньоквадратична помилка (RMSE)



# Середня абсолютна помилка (MAE)

MAE вимірює середнє значення величини помилок прогнозування, незалежно від їх напрямку. Це дає просте тлумачення точності прогнозування.

Формула:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Де:

- $y_i$ : Фактичне значення в момент часу  $i$
- $\hat{y}_i$ : Прогнозоване значення в момент часу  $i$
- $n$ : Кількість спостережень

Тлумачення:

- MAE надає просте вимірювання помилки прогнозування в тих самих одиницях, що й дані.
- Нижче MAE вказує на кращу ефективність моделі.

# Середня квадратична помилка (MSE)

MSE вимірює середнє значення квадратів різниць між фактичними та прогнозованими значеннями. Піднесення помилок у квадрат підкреслює більші відхилення, караючи моделі із значними помилками.

Формула:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Тлумачення:

- MSE чутливий до великих помилок, що робить його придатним для сценаріїв, де уникнення великих помилок прогнозування має вирішальне значення.
- Нижче MSE вказує на кращу відповідність моделі.

# Середньоквадратична помилка (RMSE)

RMSE - це квадратний корінь із MSE. Він зберігає ту саму одиницю, що й дані, що робить його більш зрозумілим порівняно з MSE.

Формула:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Тлумачення:

- RMSE карає великі помилки сильніше, ніж MAE, але його легше інтерпретувати, ніж MSE, завдяки його еквівалентності одиниці до даних.
- Нижче RMSE вказує на кращу точність моделі.

# Кроки для оцінки точності прогнозування

Оцінка точності прогнозування передбачає порівняння прогнозованих значень із фактичними значеннями. Ось кроки, які слід виконати:

1. Зібрати прогнози та фактичні значення
2. Розрахувати показники помилки.
3. Порівняти показники помилки.

# Приклад порівняння моделей

Припустимо, набір даних містить щомісячні дані про продажі. Зазвичай застосовують три методи прогнозування: зважене середнє, експоненціальне згладжування та лінійна регресія. Потім вони розраховують показники помилки для кожного методу.

Метод	MAE	MSE	RMSE
Зважене середнє	50.2	4200.5	64.8
Експоненціальне згладжування	45.6	3800.2	61.6
Лінійна регресія	40.1	3500.8	59.2

Тлумачення:

- Лінійна регресія має найнижчі MAE, MSE та RMSE, що вказує на те, що це найточніша модель у цьому випадку.
- Ковзне середнє має найвищі помилки, що свідчить про те, що воно менш підходить для цього набору даних.

## 1. Зважене середнє (Moving Average)

Метод **зваженого середнього** обчислює середнє значення фіксованої кількості попередніх точок даних для прогнозування наступного значення. Цей метод згладжує коливання у даних, що робить його корисним для короткострокового прогнозування, де важливо врахувати загальні тенденції, а не сезонні або патерни.

### Як це працює:

- Вибирається розмір вікна (наприклад, 3 місяці або 6 місяців), що визначає кількість точок даних для усереднення.
- Обчислюється середнє значення останніх точок даних, яке використовується як прогноз для наступного періоду.
- Коли додається нове значення, найстаріше значення у вікні замінюється новим, створюючи "рухоме" середнє.



## Переваги:

- Простий у розумінні та реалізації.
- Згладжує шуми у даних, дозволяючи краще бачити тенденції.

## Недоліки:

- Не враховує тенденцій чи сезонних патернів у даних.
- Результати сильно залежать від вибору розміру вікна.



## 2. Експоненціальне згладжування (Exponential Smoothing)

Метод **експоненціального згладжування** присвоює зменшувани експоненційно ваги старішим спостереженням, надаючи більше значення останнім даним. Це робить його більш чутливим до змін тенденцій порівняно з методом зваженого середнього.

### Як це працює:

• Прогноз оновлюється шляхом зваженого середнього попереднього прогнозу та фактичного значення поточного періоду:

$$F_t = \alpha Y_t + (1 - \alpha) F_{t-1}$$

- $F_t$ : Прогноз на поточний період.
- $Y_t$ : Фактичне значення поточного періоду.
- $\alpha$ : Параметр згладжування ( $0 < \alpha < 1$ ).

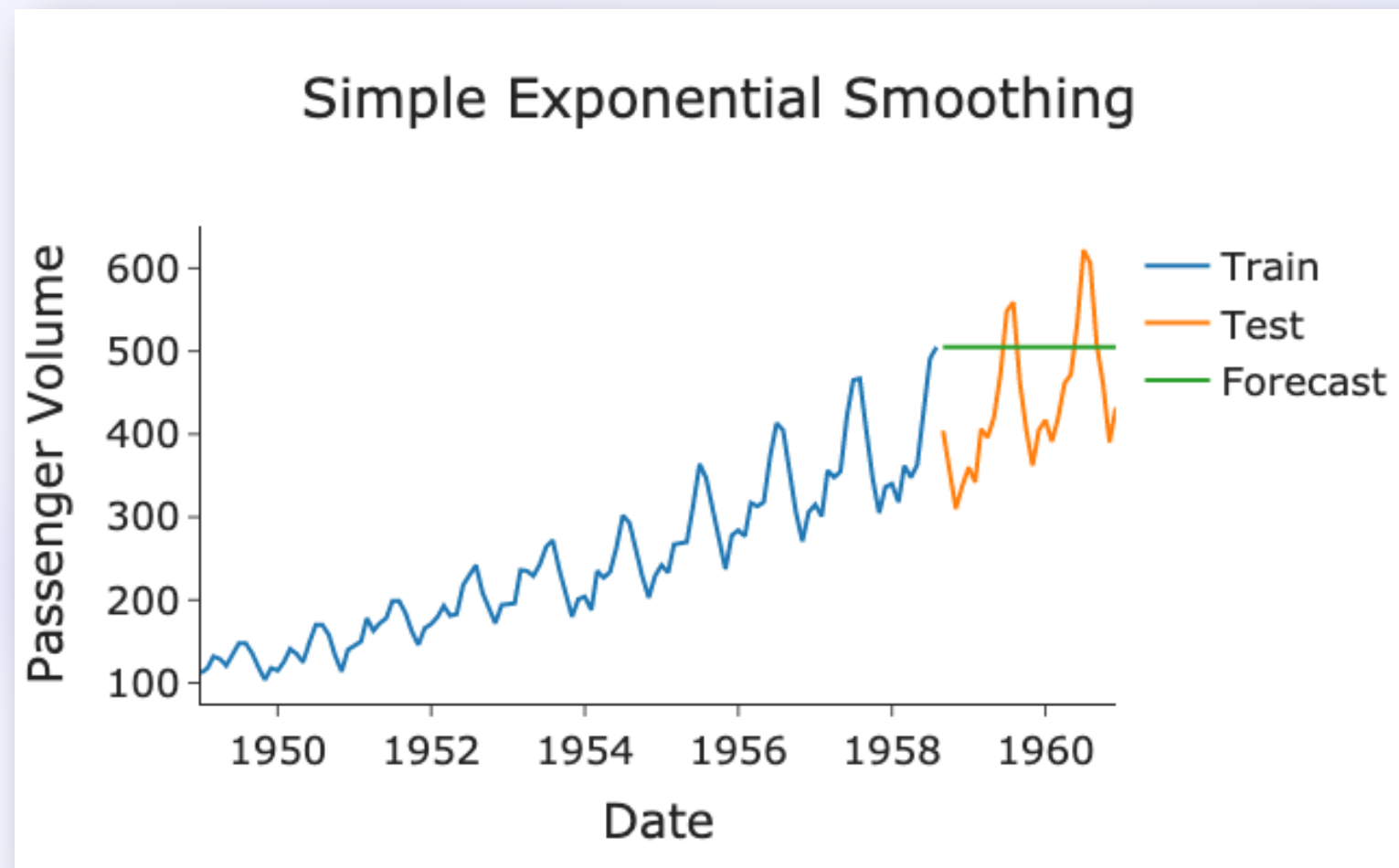
Параметр згладжування ( $\alpha$ ) визначає, наскільки сильно враховуються останні дані. Вищі значення  $\alpha$  надають більше ваги останнім спостереженням.

## Переваги:

- Враховує тенденції та коригується відповідно до змін у даних.
- Гнучкий завдяки налаштуванню параметра згладжування.

## Недоліки:

- Слабко працює із сильноволатильними даними або складними сезонними патернами.
- Вимагає налаштування параметра згладжування.



### 3. Лінійна регресія (Linear Regression)

**Лінійна регресія** – це статистичний метод, який моделює залежність між залежною змінною (тим, що прогнозується) та однією або декількома незалежними змінними (предикторами, наприклад, часом). У найпростішому випадку вона визначає прямолінійну залежність між змінними.

#### Як це працює:

• Метод будує лінію, яка мінімізує суму квадратів відхилень (помилки) між фактичними точками даних та передбаченими значеннями:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- $Y$ : Залежна змінна (прогнозоване значення).
- $X$ : Незалежна змінна (наприклад, час).
- $\beta_0$ : Перетин (значення  $Y$ , коли  $X=0$ ).
- $\beta_1$ : Нахил лінії (швидкість зміни  $Y$  відносно  $X$ ).
- $\epsilon$ : Помилка моделі.

Отримане рівняння можна використовувати для прогнозування майбутніх значень, підставляючи нові значення  $X$  (наприклад, майбутні періоди).

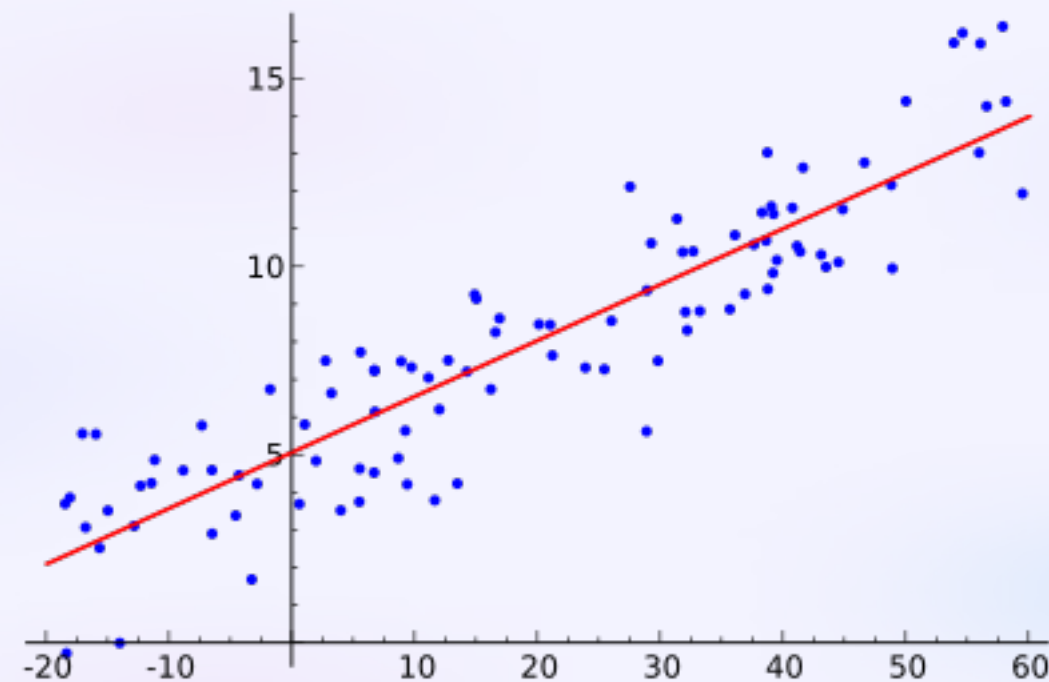
## Лінійна регресія (Linear Regression)

### Переваги:

- Ефективно моделює довгострокові тенденції, якщо зв'язок між змінними лінійний.
- Корисний для виявлення залежностей між змінними.

### Недоліки:

- Припускає лінійний зв'язок, який може не підходити для всіх наборів даних.
- Чутливий до викидів, які можуть спотворити модель.



## Оцінка методу прогнозування

На основі показників помилок (MAE, MSE, RMSE), розрахованих для трьох методів прогнозування, застосованих до місячного набору даних про продажі, Лінійна регресія стає найточнішим методом прогнозування для цього сценарію.

### 1. Порівняння показників помилок:

- Лінійна регресія має найнижчий MAE (40,1), що вказує на те, що вона створює найменшу середню помилку порівняно з фактичними значеннями продажів.
- Він також має найнижчий MSE (3500,8), що свідчить про те, що він менше штрафує за великі помилки, ніж інші методи, що є критичним для мінімізації значних відхилень у прогнозах.
- Найнижчий RMSE (59,2) підтверджує, що лінійна регресія досягає найкращої загальної ефективності з точки зору точності, залишаючись інтерпретованою в тих самих одиницях, що й дані про продажі.

### 2. Відносна ефективність інших методів:

- Експоненціальне згладжування працює краще, ніж метод ковзного середнього, з меншими значеннями MAE, MSE і RMSE. Це означає, що він краще підходить для фіксації тенденцій або закономірностей у даних про продажі.
- Ковзне середнє мало найбільші помилки за всіма показниками, що свідчить про те, що воно найменше підходить для цього набору даних. Цей результат очікується, якщо дані про продажі містять тенденції або сезонні зміни, які метод ковзного середнього не може ефективно моделювати.

### 3. Наслідки:

- Результати свідчать про те, що дані про продажі демонструють сильну лінійну тенденцію, яку лінійна регресія добре впорається.
- Чудова продуктивність експоненціального згладжування над ковзним середнім підкреслює його здатність більш ефективно включати останні дані, що робить його кращим варіантом, коли існують тенденції або сезонність.

### 4. Рекомендація:

- Виходячи з отриманих даних, для прогнозування цих конкретних даних про продажі рекомендується Лінійна регресія, особливо якщо основний зв'язок між часом і продажами продовжує демонструвати лінійні характеристики.
- Однак, якщо шаблон даних змінюється з часом, можуть знадобитися більш вдосконалені моделі або коригування існуючих методів.

# Ключові моменти для обговорення: Оцінка точності прогнозування

При оцінці точності моделей прогнозування важливо розуміти компроміси між різними показниками помилки, придатність методів для різних типів даних та наслідки вибору моделі. Ці обговорення допомагають студентам критично аналізувати методи прогнозування та вибирати найкращий підхід для конкретних застосувань.

**1** **Компромід між показниками**  
Різні показники помилки виділяють різні аспекти ефективності моделі. Розуміння цих компромісів забезпечує обґрунтоване прийняття рішень.

**2** **Придатність методів**  
Кожен метод прогнозування має свої сильні та слабкі сторони. Розуміння контексту та характеристик даних допомагає визначити, який метод використовувати.

**3** **Наслідки вибору моделі**  
Вибір моделі прогнозування повинен узгоджуватися з характеристиками даних та метою прогнозування. Роздуми над цими факторами є ключовими для точних та змістовних прогнозів.