

ДОДАТОК А. СЛАЙДИ ПРЕЗЕНТАЦІЇ

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки
Кафедра інформаційно-мережної інженерії



КВАЛІФІКАЦІЙНА РОБОТА

Модель прогнозу для даних про продажі на
основі керованих веб-сервісів Amazon

Виконав ст. гр. ІМІм-20-2 Юр'єв Я. В.
Керівник: доц. Кривенко С. А.
Харків 2022

МЕТА РОБОТИ

Аналіз керованих веб-сервісів Amazon, що пропонують послуги машинного навчання для прогнозування, розробка моделі прогнозу продажів для роздрібно́ї торгівлі і подальше її тестування на заготовлених експериментальних даних про попит з використанням сервісу Amazon Forecast.



ЗМІСТ

- 1 Машинне навчання і прогнозування
- 2 Аналіз веб-сервісів машинного навчання Amazon для прогнозування
- 3 Модель прогнозування продажів
- 4 Тестування створення прогнозів в Amazon Forecast

3

1 МАШИННЕ НАВЧАННЯ І ПРОГНОЗУВАННЯ

Прогнозування даних пройшло довгий шлях після того, як були представлені неймовірні технології, що прискорюють обробку даних, як-от **машинне навчання**. Прогнозні моделі, засновані на ML, сьогодні можуть враховувати залежні від часу компоненти – сезонність, тенденції, цикли, нерегулярні компоненти тощо – щоб максимізувати точність прогнозів і прогнозів на основі даних.

4

1 МАШИННЕ НАВЧАННЯ І ПРОГНОЗУВАННЯ

Це називається **прогнозуванням** за допомогою машинного навчання, і воно може бути найбільш вигідним для всіх видів бізнесу, включаючи прогнозування продажів і попиту, прогнозування набору кадрів, прогноз погоди, прогнозування споживання контенту, прогнозне планування та обслуговування тощо.

5

1 МАШИННЕ НАВЧАННЯ І ПРОГНОЗУВАННЯ



Рисунок 1

Галузі застосування машинного навчання для прогнозування

6

2 АНАЛІЗ ВЕБ-СЕРВІСІВ МАШИННОГО НАВЧАННЯ AMAZON ДЛЯ ПРОГНОЗУВАННЯ

Amazon SageMaker – це повністю керована служба машинного навчання. За допомогою SageMaker науковці та розробники даних можуть швидко й легко створювати й навчати моделі машинного навчання, а потім безпосередньо розгорнути їх у готовому для виробництва середовищі розміщення.

Сервіс надає інтегрований екземпляр блокнота для розробки Jupyter для легкого доступу до джерел даних для дослідження та аналізу, тому користувачу не потрібно керувати серверами.

7

2 АНАЛІЗ ВЕБ-СЕРВІСІВ МАШИННОГО НАВЧАННЯ AMAZON ДЛЯ ПРОГНОЗУВАННЯ

Він також надає загальні алгоритми машинного навчання, які оптимізовані для ефективної роботи з надзвичайно великими даними в розподіленому середовищі. Завдяки вбудованій підтримці алгоритмів і фреймворків користувача,

SageMaker пропонує гнучкі розподілені варіанти навчання, які адаптуються до конкретних робочих процесів. Розгортання моделі у безпечному та масштабованому середовищі відбувається кількома клацаннями з SageMaker Studio або консолі SageMaker.

8

2 АНАЛІЗ ВЕБ-СЕРВІСІВ МАШИННОГО НАВЧАННЯ AMAZON ДЛЯ ПРОГНОЗУВАННЯ

Amazon Forecast – це повністю керований сервіс, який використовує статистичні алгоритми та алгоритми машинного навчання для надання високоточних прогнозів часових рядів. На основі тієї ж технології, що використовується для прогнозування часових рядів на Amazon.com,

Forecast надає найсучасніші алгоритми для прогнозування майбутніх даних часових рядів на основі історичних даних і не вимагає досвіду машинного навчання.

9

2 АНАЛІЗ ВЕБ-СЕРВІСІВ МАШИННОГО НАВЧАННЯ AMAZON ДЛЯ ПРОГНОЗУВАННЯ

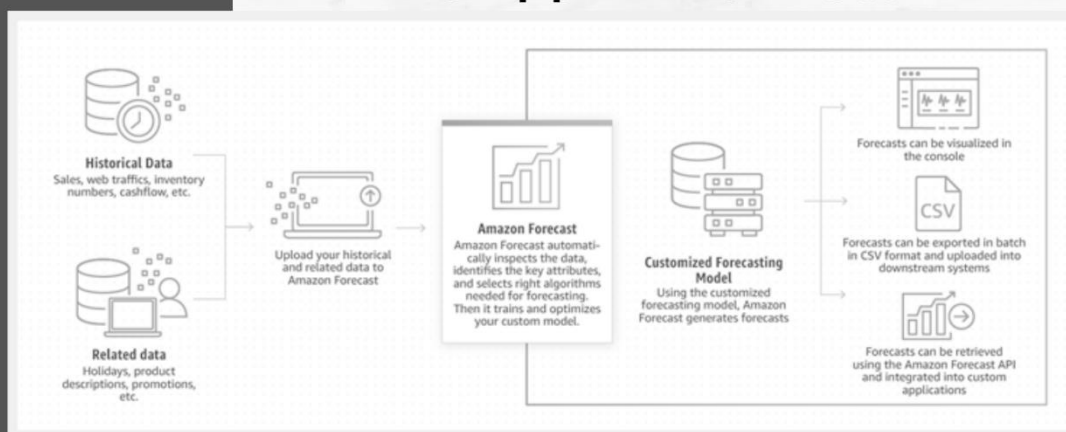


Рисунок 2

Схема роботи Amazon Forecast

10

2 АНАЛІЗ ВЕБ-СЕРВІСІВ МАШИННОГО НАВЧАННЯ AMAZON ДЛЯ ПРОГНОЗУВАННЯ

Користувач може використовувати **Python** і **R** вбудовані в ядрах блокнотів Amazon SageMaker. Існують також ядра, які підтримують певні фреймворки. Дуже популярним способом розпочати роботу з SageMaker є використання Amazon SageMaker Python SDK.

Він надає API та контейнери Python з відкритим вихідним кодом, які полегшують навчання та розгортання моделей у SageMaker, а також приклади для використання з кількома різними фреймворками машинного навчання та глибокого навчання.

11

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Прогнозування є важливою областю машинного навчання. Це важливо, оскільки багато можливостей для прогнозування майбутніх результатів базуються на **історичних даних**. Багато з цих можливостей включають часовий компонент. Незважаючи на те, що компонент часу додає більше інформації, він також ускладнює вирішення проблем часових рядів, ніж інші типи передбачень.

12

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Дані часових рядів фіксуються в хронологічній послідовності протягом певного періоду часу. Введення часу в модель машинного навчання має позитивний вплив, оскільки модель може отримувати значення через зміну точок даних з часом. Дані часових рядів, як правило, корелюються, що означає, що існує залежність між точками даних.

13

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

На додаток до даних часового ряду, можна додати **пов'язані дані** для доповнення моделі прогнозування. Наприклад, для прогнозування роздрібних продажів ми можемо включити інформацію про продукт, що продається (наприклад, ідентифікацію товару або ціну продажу). Ця інформація є доповненням до кількості одиниць, які продаються за період часу.

14

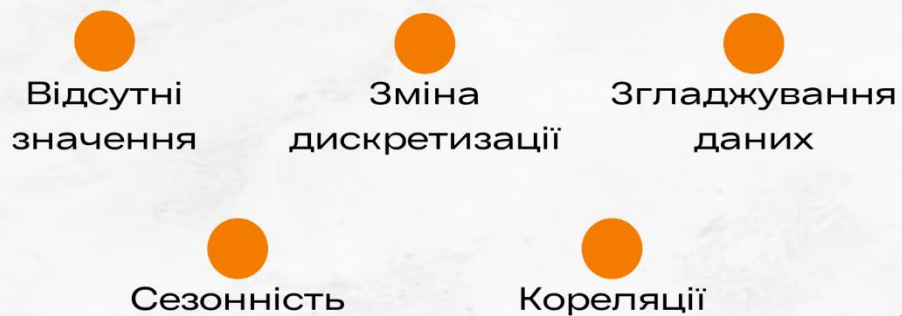
3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Третій тип даних – це **метадані** про набір даних. Наприклад, для набору даних роздрібної торгівлі ми можемо включити метадані до групових результатів, наприклад назву бренду. Іншим прикладом метаданих може бути включення жанру для музики чи відео.

15

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Обробка даних часових рядів



16

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

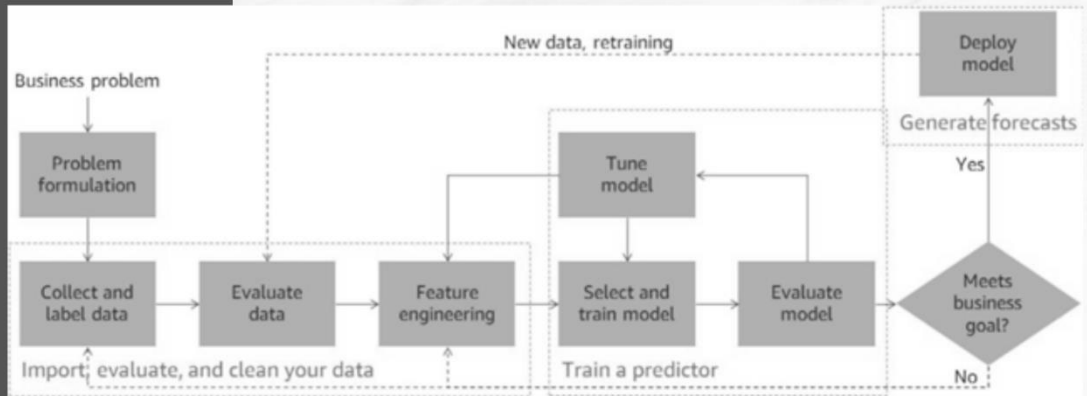


Рисунок 3

Робочий процес Amazon Forecast

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Алгоритми для навчання моделей в Amazon Forecast

●
ARIMA

●
DeepAR+

●
ETS

●
NPTS

●
Prophet

3 МОДЕЛЬ ПРОГНОЗУВАННЯ ПРОДАЖІВ



Рисунок 4

Оцінка прогнозу зворотним тестуванням

19

4 ТЕСТУВАННЯ СТВОРЕННЯ ПРОГНОЗІВ В AMAZON FORECAST

Набір даних Online Retail II, який буде використовуватися як експериментальні дані для тестування, містить усі транзакції, які відбулися між 12 січня 2009 року та 12 вересня 2011 року для онлайн-магазину, який зареєстрований та розташований у Великобританії. В основному компанія продає унікальні сувеніри на всі випадки життя. Багато клієнтів компанії – оптові торговці.

20

4 ТЕСТУВАННЯ СТВОРЕННЯ ПРОГНОЗІВ В AMAZON FORECAST

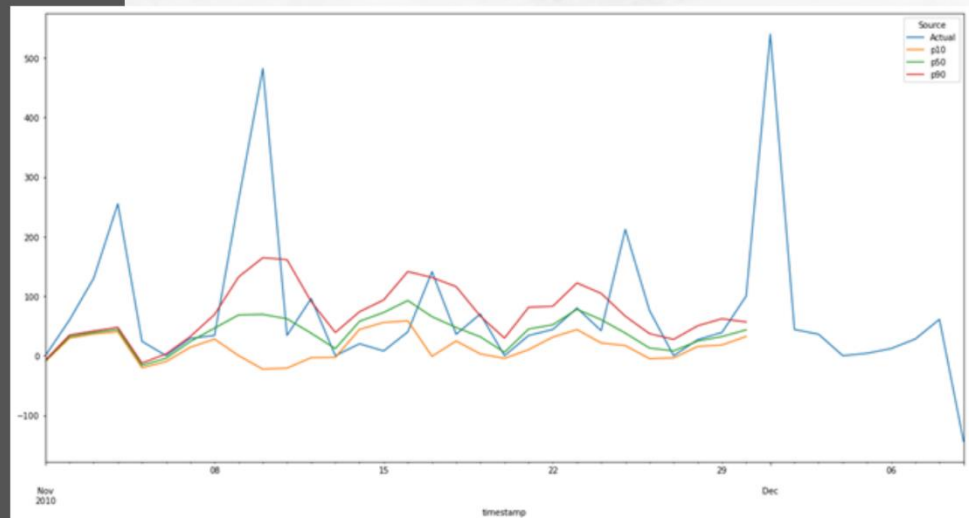


Рисунок 5

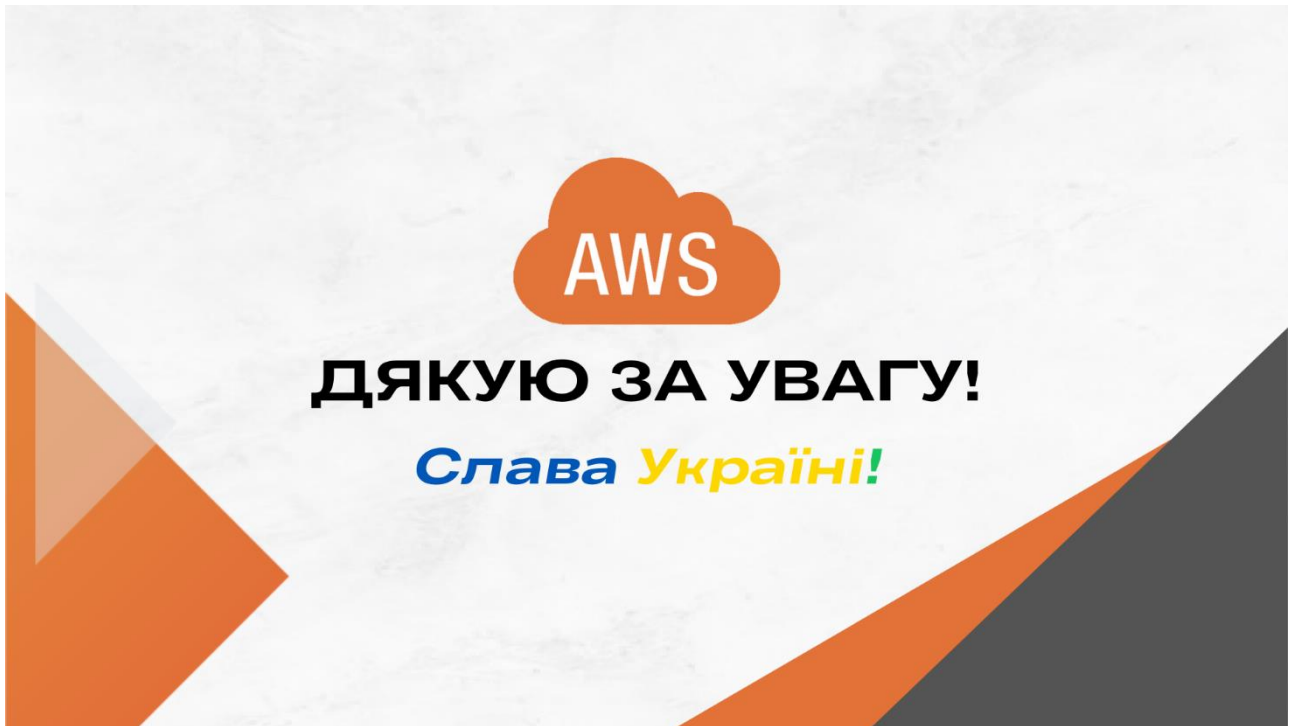
Графік порівняння реальних і прогнозованих значень

21

ВИСНОВКИ

В роботі зроблено огляд використання машинного навчання для створення прогнозів у різних галузях підприємництва, зокрема у роздрібній торгівлі, наведено різні алгоритми прогнозування, можливі виклики в галузі і їх рішення. Проведено аналіз керованих веб-сервісів Amazon, які використовуються для навчання і розгортання моделей прогнозування. Також розроблено експериментальну модель для прогнозу продажів, і потім виконано її тестування на заготовлених даних онлайн-магазину про попит на товари. Виконано порівняльний аналіз отриманих даних прогнозу і реальних даних за потрібний період і виявлено кореляцію між ними. Додатково підкреслено важливість використання прогнозування машинного навчання в умовах конкурентної боротьби бізнесу.

22



ДОДАТОК Б. ПУБЛІКАЦІЇ ЗА ТЕМАТИКОЮ РОБОТИ

УДК 004.032.2:621.391

**МОДЕЛЬ ПРОГНОЗУ ДЛЯ ДАНИХ ПРО ПРОДАЖІ НА ОСНОВІ
КЕРОВАНИХ ВЕБ-СЕРВІСІВ AMAZON**

Юр'єв Я.В.

Науковий керівник – к.т.н., доц. Кривенко С.А.
Харківський національний університет радіоелектроніки
61166, Харків, пр. Науки, 14, каф. ІМІ,
тел. +38(095) 553-95-71, e-mail: yaroslav.yuriev@nure.ua.

This paper describes the use of machine learning to predict data and the use of artificial intelligence algorithms to predict sales. To achieve this goal, it is proposed to use Amazon services: SageMaker (machine learning modeling and placement service) and Forecast (high-precision forecasting service based on machine learning technologies).

Дані часових рядів фіксуються в хронологічній послідовності протягом визначеного періоду часу. Введення часу в модель машинного навчання має позитивний вплив, оскільки модель може отримати сенс із зміни точок даних з часом. Дані часових рядів, як правило, корелюють, що означає, що існує залежність між точками даних.

Оскільки у нас є проблема з регресією, а також оскільки регресія передбачає незалежність точок даних, нам необхідно розробити модель обробки залежності даних.

Метою розробки цієї моделі і відповідного методу є підвищення достовірності прогнозів для даних про продажі.

Машинне навчання (Machine Learning) — це тип штучного інтелекту (Artificial Intelligence), який дозволяє програмним додаткам стати більш точними в прогнозуванні результатів, не будучи явно запрограмованим на це. Алгоритми машинного навчання використовують історичні дані як вхідні дані для прогнозування нових вихідних значень.

Прогнозування є важливою областю машинного навчання. Це важливо, оскільки багато можливостей для прогнозування майбутніх результатів базуються на історичних даних. Багато з цих можливостей включають часовий компонент. Хоча компонент часу додає більше інформації, він також ускладнює вирішення проблем часових рядів, ніж інші типи передбачень.

Amazon SageMaker — це повністю керована служба машинного навчання. За допомогою SageMaker науковці та розробники даних можуть швидко й легко створювати й навчати моделі машинного навчання, а потім безпосередньо розгортати їх у готовому для виробництва середовищі розміщення. Вона надає інтегрований екземпляр віртуальної машини для розробки Jupyter для легкого

доступу до джерел даних для дослідження та аналізу без потреби керувати серверами. Служба також надає загальні алгоритми машинного навчання, які оптимізовані для ефективної роботи з надзвичайно великими даними в розподіленому середовищі. Завдяки вбудованій підтримці алгоритмів і фреймворків від користувача, SageMaker пропонує гнучкі розподілені варіанти навчання, які адаптуються до конкретних робочих процесів користувачів. Розгортання моделі у безпечному та масштабованому середовищі відбувається за допомогою SageMaker Studio або консолі SageMaker.

Amazon Forecast — це повністю керований сервіс, який використовує машинне навчання для надання високоточних прогнозів. На основі тієї ж технології, що використовується на Amazon.com, Forecast використовує машинне навчання для поєднання даних часових рядів з додатковими змінними для побудови прогнозів. Щоб розпочати створення прогнозу, не потрібен досвід машинного навчання. Користувачу потрібно надати лише історичні дані, а також будь-які додаткові дані, які, на його думку, можуть вплинути на прогнози. Наприклад, попит на певний товар може змінюватися в залежності від сезону та розташування магазину. Цей складний зв'язок важко визначити самостійно, але машинне навчання ідеально підходить для його розпізнавання. Щойно користувач надає свої дані, Forecast автоматично перевірить їх, визначить, що є значущим, і створить модель прогнозування, здатну робити прогнози, які є на 50% точнішими, ніж перегляд лише даних часового ряду.

Amazon Forecast автоматизує більшу частину процесу прогнозування часових рядів, дозволяючи зосередитися на підготовці наборів даних та інтерпретації ваших прогнозів. Forecast надає вказані нижче функції.

Автоматичне машинне навчання – Forecast автоматизує складні завдання машинного навчання, знаходячи оптимальну комбінацію алгоритмів машинного навчання для наборів даних.

Найсучасніші алгоритми – застосування комбінацій алгоритмів машинного навчання, які базуються на тій же технології, що й на Amazon.com. Forecast пропонує широкий спектр алгоритмів навчання, від широко використовуваних статистичних методів до складних нейронних мереж.

Підтримка відсутніх значень – Forecast надає кілька методів заповнення для автоматичної обробки відсутніх значень у наборах даних.

Додаткові вбудовані набори даних – Forecast може автоматично включати вбудовані набори даних, щоб покращити модель. Ці набори даних уже розроблені і не потребують додаткової конфігурації.

Запропонована модель і відповідні методи підвищують достовірність прогнозів для даних про продажі [1].

Список використаних джерел:

1. AWS Documentation. Amazon Web Services. URL: <https://docs.aws.amazon.com/index.html> (дата звернення: 20.02.2022).

ДОДАТОК В. ДАНІ ТЕСТОВОГО КОДУ ЗАПАСУ

```

    {'Forecast': {'Predictions': {'p10': [{'Timestamp': '2010-11-01T00:00:00',
'Value': -9.7571210861}, {'Timestamp': '2010-11-02T00:00:00', 'Value':
29.4540252686}, {'Timestamp': '2010-11-03T00:00:00', 'Value': 36.668056488},
{'Timestamp': '2010-11-04T00:00:00', 'Value': 40.2012252808}, {'Timestamp':
'2010-11-05T00:00:00', 'Value': -20.3314857483}, {'Timestamp': '2010-11-
06T00:00:00', 'Value': -9.8204021454}, {'Timestamp': '2010-11-07T00:00:00',
'Value': 14.1882171631}, {'Timestamp': '2010-11-08T00:00:00', 'Value':
27.8944339752}, {'Timestamp': '2010-11-09T00:00:00', 'Value': 0.0503234863},
{'Timestamp': '2010-11-10T00:00:00', 'Value': -22.5631103516}, {'Timestamp':
'2010-11-11T00:00:00', 'Value': -20.8479499817}, {'Timestamp': '2010-11-
12T00:00:00', 'Value': -3.3248710632}, {'Timestamp': '2010-11-13T00:00:00',
'Value': -2.8114709854}, {'Timestamp': '2010-11-14T00:00:00', 'Value':
44.1766777039}, {'Timestamp': '2010-11-15T00:00:00', 'Value': 55.8097457886},
{'Timestamp': '2010-11-16T00:00:00', 'Value': 58.5392341614}, {'Timestamp':
'2010-11-17T00:00:00', 'Value': -1.0365142822}, {'Timestamp': '2010-11-
18T00:00:00', 'Value': 24.7277812958}, {'Timestamp': '2010-11-19T00:00:00',
'Value': 2.967792511}, {'Timestamp': '2010-11-20T00:00:00', 'Value': -
4.4655151367}, {'Timestamp': '2010-11-21T00:00:00', 'Value': 10.1678295135},
{'Timestamp': '2010-11-22T00:00:00', 'Value': 31.2767086029}, {'Timestamp':
'2010-11-23T00:00:00', 'Value': 43.8765182495}, {'Timestamp': '2010-11-
24T00:00:00', 'Value': 21.3958320618}, {'Timestamp': '2010-11-25T00:00:00',
'Value': 17.0110282898}, {'Timestamp': '2010-11-26T00:00:00', 'Value': -
4.9480218887}, {'Timestamp': '2010-11-27T00:00:00', 'Value': -3.7253046036},
{'Timestamp': '2010-11-28T00:00:00', 'Value': 15.5467720032}, {'Timestamp':
'2010-11-29T00:00:00', 'Value': 17.9712162018}, {'Timestamp': '2010-11-
30T00:00:00', 'Value': 32.2370758057}], 'p50': [{'Timestamp': '2010-11-
01T00:00:00', 'Value': -8.6056346893}, {'Timestamp': '2010-11-02T00:00:00',
'Value': 32.2053871155}, {'Timestamp': '2010-11-03T00:00:00', 'Value':
39.0319213867}, {'Timestamp': '2010-11-04T00:00:00', 'Value': 43.9337882996},
{'Timestamp': '2010-11-05T00:00:00', 'Value': -16.2842712402}, {'Timestamp':
'2010-11-06T00:00:00', 'Value': -4.2147541046}, {'Timestamp': '2010-11-
07T00:00:00', 'Value': 23.8522014618}, {'Timestamp': '2010-11-08T00:00:00',
'Value': 46.3265037537}, {'Timestamp': '2010-11-09T00:00:00', 'Value':
68.4266052246}, {'Timestamp': '2010-11-10T00:00:00', 'Value': 69.2805404663},
{'Timestamp': '2010-11-11T00:00:00', 'Value': 62.0143089294}, {'Timestamp':
'2010-11-12T00:00:00', 'Value': 37.7757949829}, {'Timestamp': '2010-11-
13T00:00:00', 'Value': 11.8608989716}, {'Timestamp': '2010-11-14T00:00:00',
'Value': 57.7246780396}, {'Timestamp': '2010-11-15T00:00:00', 'Value':
72.5650405884}, {'Timestamp': '2010-11-16T00:00:00', 'Value': 92.6500930786},
{'Timestamp': '2010-11-17T00:00:00', 'Value': 65.4984970093}, {'Timestamp':
'2010-11-18T00:00:00', 'Value': 47.34324646}, {'Timestamp': '2010-11-
19T00:00:00', 'Value': 31.6962127686}, {'Timestamp': '2010-11-20T00:00:00',
'Value': 5.7113828659}, {'Timestamp': '2010-11-21T00:00:00', 'Value':
44.7913475037}, {'Timestamp': '2010-11-22T00:00:00', 'Value': 51.9129867554},
{'Timestamp': '2010-11-23T00:00:00', 'Value': 77.8986358643}, {'Timestamp':
'2010-11-24T00:00:00', 'Value': 60.5322151184}, {'Timestamp': '2010-11-
25T00:00:00', 'Value': 37.955696106}, {'Timestamp': '2010-11-26T00:00:00',
'Value': 12.8957977295}, {'Timestamp': '2010-11-27T00:00:00', 'Value':
8.0049095154}, {'Timestamp': '2010-11-28T00:00:00', 'Value': 24.597951889},
{'Timestamp': '2010-11-29T00:00:00', 'Value': 32.1774024963}, {'Timestamp':
'2010-11-30T00:00:00', 'Value': 43.3186378479}], 'p90': [{'Timestamp': '2010-11-
01T00:00:00', 'Value': -7.5065326691}, {'Timestamp': '2010-11-02T00:00:00',
'Value': 34.7498283386}, {'Timestamp': '2010-11-03T00:00:00', 'Value':
41.5090408325}, {'Timestamp': '2010-11-04T00:00:00', 'Value': 47.7141838074},
{'Timestamp': '2010-11-05T00:00:00', 'Value': -12.509645462}, {'Timestamp':
'2010-11-06T00:00:00', 'Value': 3.0610613823}, {'Timestamp': '2010-11-

```

```

07T00:00:00', 'Value': 32.2038230896}, {'Timestamp': '2010-11-08T00:00:00',
'Value': 69.0688095093}, {'Timestamp': '2010-11-09T00:00:00', 'Value':
132.3823547363}, {'Timestamp': '2010-11-10T00:00:00', 'Value': 164.4650115967},
{'Timestamp': '2010-11-11T00:00:00', 'Value': 161.2159423828}, {'Timestamp':
'2010-11-12T00:00:00', 'Value': 90.8257598877}, {'Timestamp': '2010-11-
13T00:00:00', 'Value': 38.9117431641}, {'Timestamp': '2010-11-14T00:00:00',
'Value': 73.7878952026}, {'Timestamp': '2010-11-15T00:00:00', 'Value':
93.5307006836}, {'Timestamp': '2010-11-16T00:00:00', 'Value': 141.247253418},
{'Timestamp': '2010-11-17T00:00:00', 'Value': 131.273727417}, {'Timestamp':
'2010-11-18T00:00:00', 'Value': 115.9075241089}, {'Timestamp': '2010-11-
19T00:00:00', 'Value': 66.4336776733}, {'Timestamp': '2010-11-20T00:00:00',
'Value': 29.4721832275}, {'Timestamp': '2010-11-21T00:00:00', 'Value':
81.2667388916}, {'Timestamp': '2010-11-22T00:00:00', 'Value': 83.025680542},
{'Timestamp': '2010-11-23T00:00:00', 'Value': 122.101890564}, {'Timestamp':
'2010-11-24T00:00:00', 'Value': 104.6154022217}, {'Timestamp': '2010-11-
25T00:00:00', 'Value': 66.2911911011}, {'Timestamp': '2010-11-26T00:00:00',
'Value': 37.0283470154}, {'Timestamp': '2010-11-27T00:00:00', 'Value':
27.4108543396}, {'Timestamp': '2010-11-28T00:00:00', 'Value': 50.3027801514},
{'Timestamp': '2010-11-29T00:00:00', 'Value': 62.0303039551}, {'Timestamp':
'2010-11-30T00:00:00', 'Value': 56.574016571}}], 'ResponseMetadata':
{'RequestId': 'ef860a17-251a-4765-9764-e35bf2708c54', 'HTTPStatusCode': 200,
'HTTPHeaders': {'content-type': 'application/x-amz-json-1.1', 'date': 'Sun, 01
May 2022 01:10:58 GMT', 'x-amzn-requestid': 'ef860a17-251a-4765-9764-
e35bf2708c54', 'content-length': '5270', 'connection': 'keep-alive'},
'RetryAttempts': 0}}

```

