



# 分析時事的原油價格預測

主辦單位

林本然、陳乙慈、黃聖民



# 01 Motivations

# Motivations

- 原油價格對市場的影響很大，對股票漲幅、進出口策略都有一定程度的影響
- 原油也會影響台灣消費者加油時的油價、民生用品價格
- 若能準確預測原油走勢，可以提早制定策略避免商品價格波動過大，消費者也能自行規劃汽油的使用。



# 02 Crawling



產量

其他作物

需求

# Expected Dataset to be Crawled

歷史價格

OilPrice.com

OPEC Basket Price

產量指標

EnerData

EIA

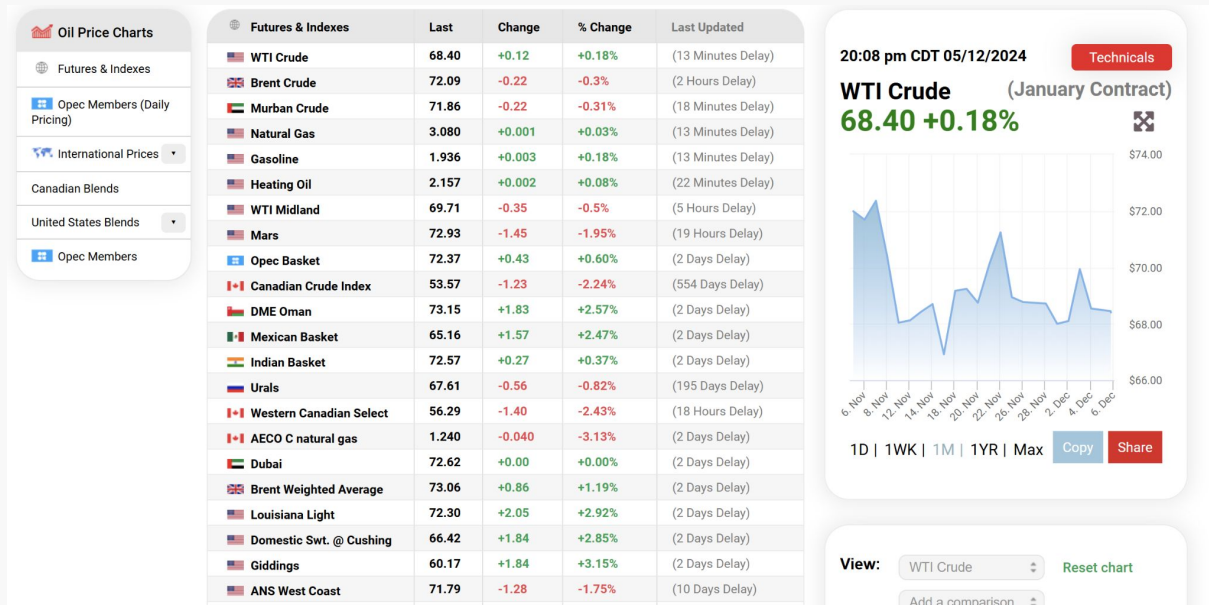
需求指標

OCED

國際新聞

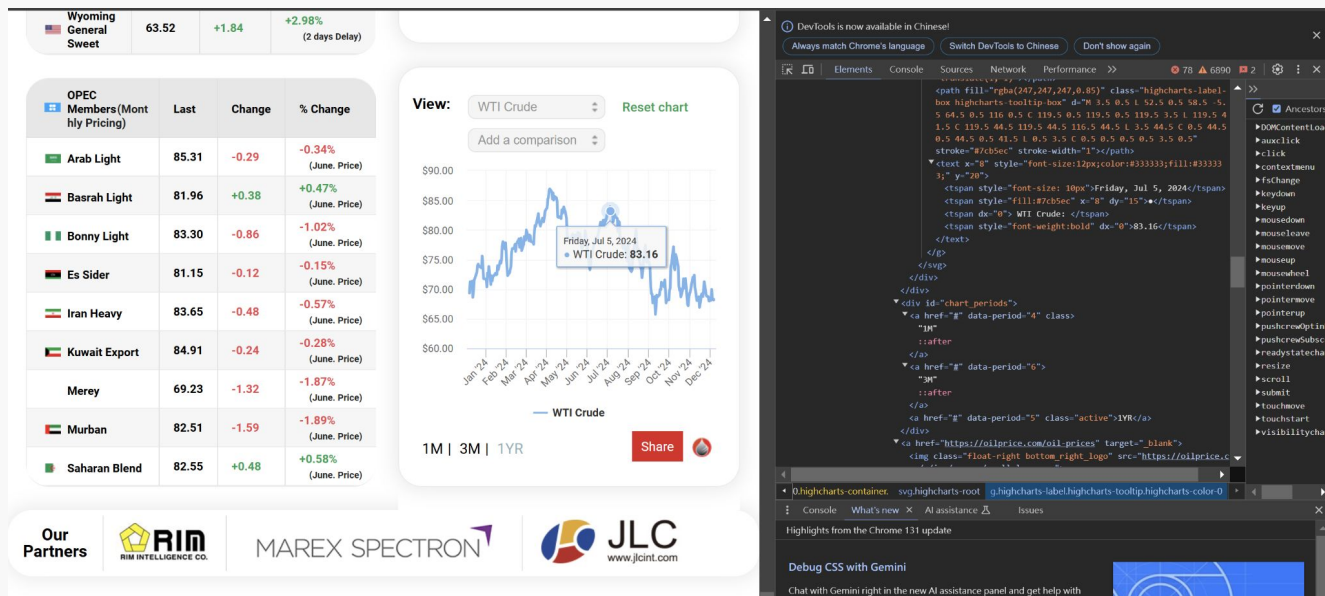
WorldOil.com

# Expected Dataset to be Crawled



爬取當前價格與變動比率

# Expected Dataset to be Crawled



獲取歷史原油價格 (天)



# Expected Dataset to be Crawled



NEWS

EAGLE FORD SHALE PRODUCTION

## Crescent Energy to acquire Ridgemar Eagle Ford assets in \$900M deal

December 05, 2024

Crescent Energy Company has announced the signing of a definitive agreement to acquire Eagle Ford assets from Ridgemar Energy for upfront consideration of \$905 million plus future oil price contingent consideration



NEWS

SAUDI ARABIA CARBON CAPTURE

## Saudi Aramco advances plan to build large-capacity CCS facility

December 04, 2024

(Bloomberg) – Saudi Aramco is moving closer to building a giant carbon capture and storage project at its main refining and chemicals hub at Jubail, on the kingdom's Gulf coast.



NEWS

OFFSHORE UK/UKCS ONSHORE

## Kent awarded 3-year global EPCm framework with bp

December 04, 2024

Kent has secured a three-year global Engineering, Procurement, and Construction Management (EPCm) framework contract by bp (P&O Projects). The framework covers the full scope of EPCm services, supporting bp's asset lifecycle, from small and large projects to minor modifications.

使用**LDA**來分析新聞情緒

# Expected Dataset to be Crawled

## MORE EDITORIALS

**Kazakhstan's Ambitious Plan to Achieve Carbon Neutrality by 2060**  
PUBLISHED 12 HOURS AGO


**Data Centers Are Sending Global Electricity Demand Soaring**  
PUBLISHED 13 HOURS AGO

**Bitcoin Shatters \$100,000 Ceiling, Fueled by Trump's Pro-Crypto Stance**  
PUBLISHED 14 HOURS AGO

**Small Nuclear Reactors Are Gaining Traction Around the Globe**  
PUBLISHED 16 HOURS AGO

## LATEST ENERGY NEWS

### BREAKING NEWS




**Oil Prices Set for Weekly Dip Despite OPEC+ Decision**

**Putin Relaxes Gas Payment Rules, But European Energy Security Still Shaky**  
PUBLISHED 10 HOURS AGO

**Oil Prices Steady After OPEC+ Pushes Back Output Hike**  
PUBLISHED 13 HOURS AGO

## MOST POPULAR



**OPEC+ May Be Facing Long-Term Production Cuts**



**U.S.-Based Semiconductor Company Tied to Russian Military Supply Chain**

**The Critical Metal That Skyrocketed From \$6,000 to \$38,000 Per Ton**



**Gas Prices Set for a Breakout in 2025**

爬取各新聞的標題，用於後續短文本分析

# Expected Dataset to be Crawled

 Sources & Uses | Topics | Geography | Tools | Education | News 

+ SORT / ORDER \*optional

 SUBMIT  RESET


API URL:

```
https://api.eia.gov/v2/crude-oil-imports/data/?
frequency=monthly&data[0]=quantity&sort[0]
[column]=period&sort[0]
[direction]=desc&offset=0&length=5000
```

METHOD: GET

SERIES DESCRIPTION:

Crude oil imports by country to destination, includes type, grade, quantity. Source: EIA-814 Interactive data product:  
[www.eia.gov/petroleum/imports/companylevel/](http://www.eia.gov/petroleum/imports/companylevel/)

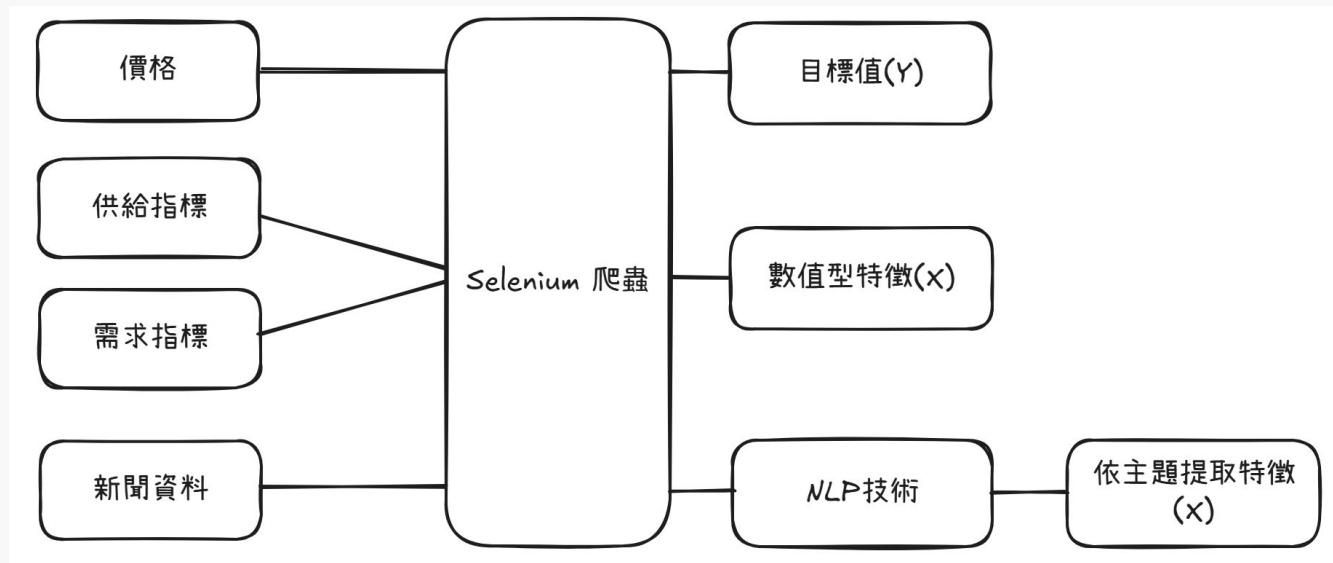
 Need help using our API? Read our [API Documentation](#) to learn more.

HEADER:

```
X-Params: {
  "frequency": "monthly",
  "data": [
    "quantity"
  ],
  "facets": {},
  "start": null,
  "end": null,
  "sort": [
    {
      "column": "period",
      "direction": "desc"
    }
  ],
  "offset": 0,
  "length": 5000
}
```

爬取 **EIA** 的 **API** 獲得需求、供給指標

# Expected Dataset to be Crawled



The background features abstract, organic shapes in dark blue and gold. In the top left, a dark blue shape with white speckles is partially visible, along with a gold line. In the top right, a dark blue shape with two small dark blue spots is present. In the bottom left, there are gold shapes and a dark blue speckled area. In the bottom right, a gold shape is accompanied by several dark blue spots.

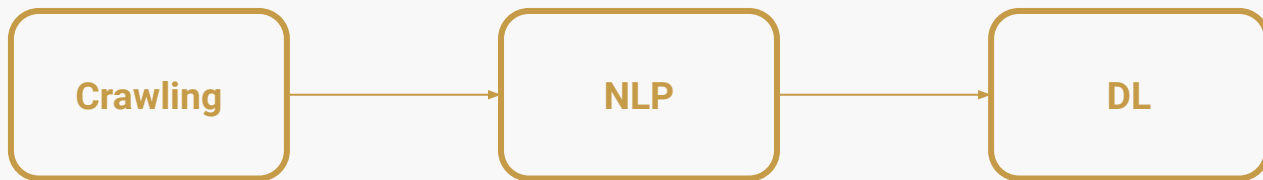
03

# Problem Statement

# Problem Statement

**目標:**根據新聞資料輔助原油價格預測

**架構:**



**結構化資料:**整理成數 值型

**非結構化資料:**根據 NLP技術整理成有用特徵

# Problem Statement

Input(numeric):

特徵(X)	時間單位	從哪裡獲得
歷史漲幅	每日	<a href="https://oilprice.com">oilprice.com</a>
原油產量(區域性)	每月	<a href="https://eia.gov">eia.gov</a>
原油庫存	每週	<a href="https://investing.com">investing.com</a>

# Problem Statement

Input(numeric):

特徵(X)	時間單位	從哪裡獲得
原油消耗量 (區域性)	每月	EIA API
美金指數	每日	<a href="https://www.investing.com">investing.com</a>
日期		



# Problem Statement

Input(non-structural):

特徵(X)	NLP 技術	從哪裡獲得
情緒分數	VADER→標題	<a href="http://worldoil.com">worldoil.com</a> <a href="http://oilprice.com">oilprice.com</a> <a href="http://investing.com">investing.com</a> <a href="http://reuters.com">reuters.com</a>
主題權重	LDA	
關鍵詞出現頻率	TF-IDF	

# Problem Statement

Output(numeric):

目標變數 (Y)	時間單位	從哪裡獲得
Brant Crude	每日	<a href="https://www.oilprice.com">oilprice.com</a>
WTI Crude	每日	
Dubai Crude	每日	

中油油價: 70% Brant + 30% Dubai

# Problem Statement



**Brant Crude**

歐洲、非洲



**WTI Crude**

美洲



**Dubai Crude**

亞洲

The background features abstract, organic shapes in dark blue and gold. In the top left, a dark blue shape with white speckles is partially visible, along with thin gold lines. The top right shows a dark blue shape with a few white dots. The bottom left has a gold shape with a dark blue speckled area. The bottom right features a gold shape with a dark blue speckled area and a small dark blue shape. The overall style is modern and artistic.

04

# Technical Challenge

# Technical Challenges - Crawling

內容是動態生成的

反爬蟲機制

需要參考很多網站

網頁需要進行某些操作才能繼續取得內容

# Technical Challenges - Features

原油受到國際政策、地緣政治、大宗農作物等影響

原油易受經濟成長影響

時事造成短期的價格波動

現行較少以 AI 考慮各方因素預測原油價格的專案

The background features abstract, organic shapes in dark blue and gold. In the top left, a dark blue shape with white speckles is partially visible, along with a gold line. In the top right, a dark blue shape with two small dark blue spots is present. In the bottom left, there are gold shapes and a cluster of dark blue speckles. In the bottom right, a gold shape is accompanied by several dark blue spots.

05

# Preliminary Methods

# Preliminary Methods - Crawling

內容是動態生成的

網頁需要進行某些操作才能繼續取得內容

反爬蟲機制



## Selenium

會跑出一個模擬器，  
可以在上面進行和真實網頁一樣的操作



# Preliminary Methods - Features

## 數值型資料

線性平滑：

將每月/週為單位的平滑處理，產生每日資料

具時間性的資料：

預計爬取疫情後 (2020~2023) 當作 train  
今年 (2024) 作為 test

特徵工程：

特徵相關性分析、資料清洗、尺度縮放、特徵選擇

# Preliminary Methods - Features

## 時事分析

VADER: 情緒特徵

短文本分析 → 得到三種情緒特徵的分數

LDA: 主題分布特徵

每個新聞對應指定標題數的矩陣 →  
自動找出隱藏的相關主題

TF-IDF: 關鍵詞頻率

計算頻率最高的詞彙 → 當作特徵

# Preliminary Methods - Model

## 模型設計

LSTM: Baseline

最常用於有時序性的資料

Transformer: 我們的模型設計

更適合去捕捉全局的趨勢

較適合新聞資料與數據資料混合的訓練集

適合捕捉突發事件造成的巨大影響

一些模型設置:

Dropout、正則化、Early Stopping → 防過擬合

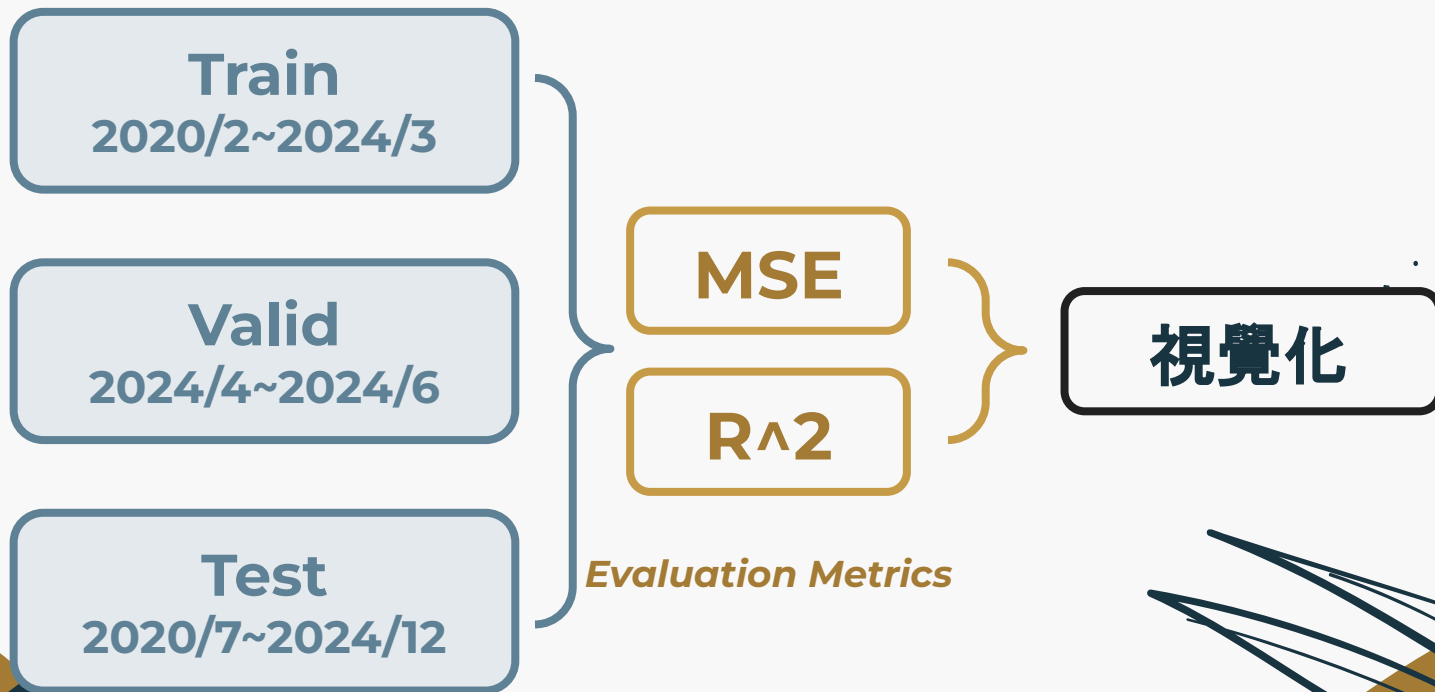
Stacking → 集成, 綜合模型結果, 較泛化



06

# Evaluation Plan

# Evaluation Plan



*Training Setting*



# 07 Schedule

# Evaluation Plans

	11/01 - 11/15	11/16 - 11/30	12月
方法調査			
爬虫			
特徴工程			
模型設計			
效能評估			