International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

EIA Data as a Forecasting Tool: Enhancing Accuracy in Oil Production Predictions

Pankaj Dureja

Email: *pankaj.dureja[at]gmail.com*

Abstract: ***This paper examines how much U. S. Energy Information Administration (EIA) data oil companies use to forecast supply and demand, as well inventory balances which result from them. Oil companies can enhance their production strategies and improve efficiency using sophisticated forecasting models powered by EIA data. This was done using the EIA data to predict renewable expansion and linear regression time - series analysis with a solid example of application, to demonstrate how these analyses are employed in interpreting this kind on data which can be used for future market trends & production needs.***

Keywords: U. S. Energy Information Administration (EIA), Oil production forecasting, Linear regression, Time - series analysis, Inventory management

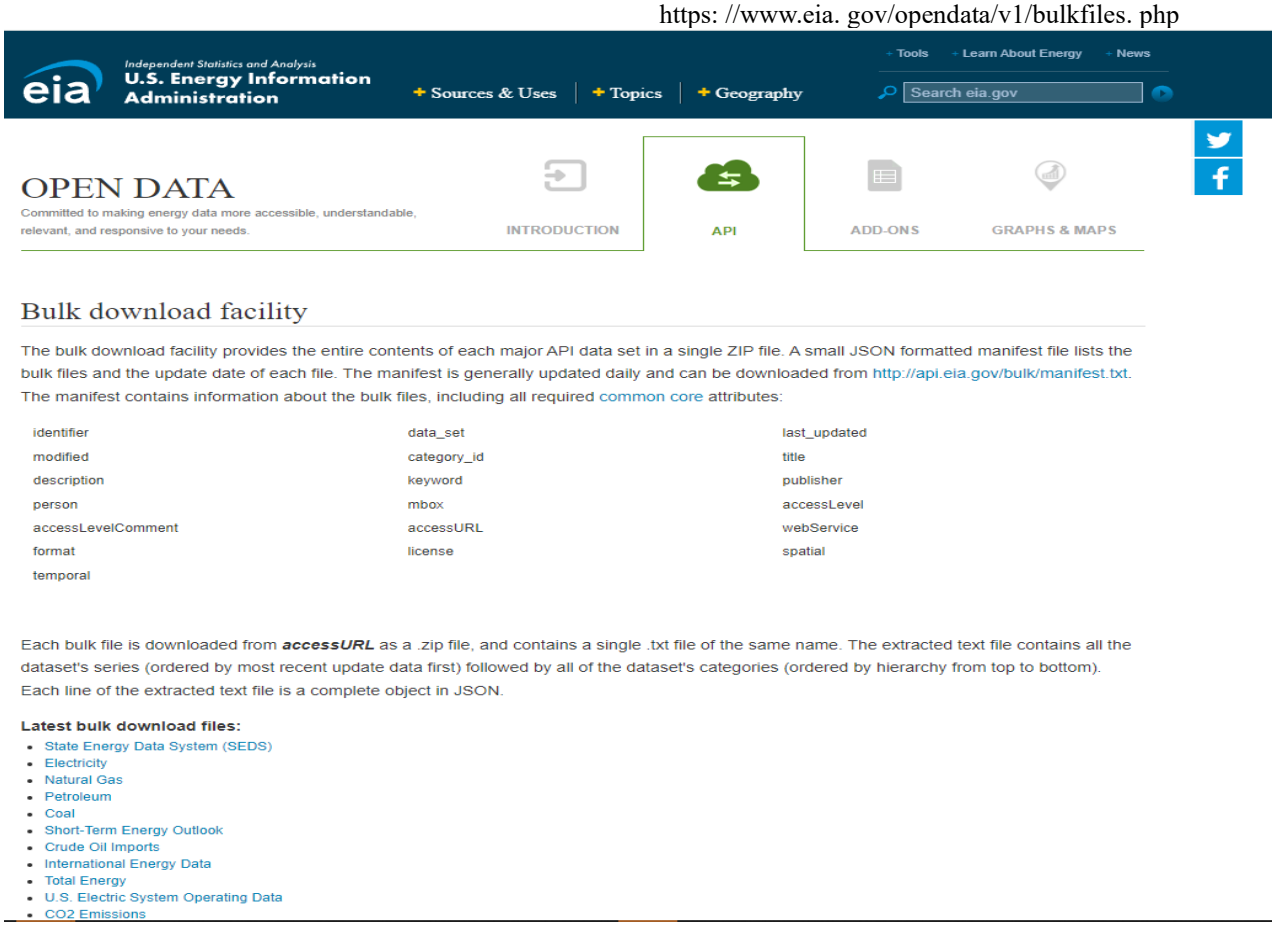
1. Introduction

In the oil market - where prices fluctuate very erratically as companies try to find an equilibrium between supply and demand - predicting what will be produced is critical. For predictive analytics much of the valuable data comes from deeply entrenched statistics covering production, consumption, reserves and price that is housed at EIA. This study examines how oil companies gather this data and use it to anticipate future events and prepare by applying statistical models combined with machine learning that improves the quality of their predictions.

2. Problem Statement

Oil companies face significant challenges in forecasting due to the volatile nature of the market influenced by geopolitical, environmental, and economic factors. The complexity of accurately processing and analyzing vast amounts of data from the EIA often results in forecasting errors, leading to financial losses and operational inefficiencies. I am employed at an oil and gas company, where we have been assigned the task of processing various data sets provided by the EIA. Each data record is formatted as JSON. The EIA data sets are open data and can be freely downloaded from the following link:

https: //www.eia. gov/opendata/v1/bulkfiles. php



Solution Implemented:

In this paper we will explore the loading of the Petroleum file as mentioned in the above screen shot. The link of the Petroleum file is mentioned below and once clicked the zip file will be downloaded.

https: //www.eia. gov/opendata/bulk/PET. zip

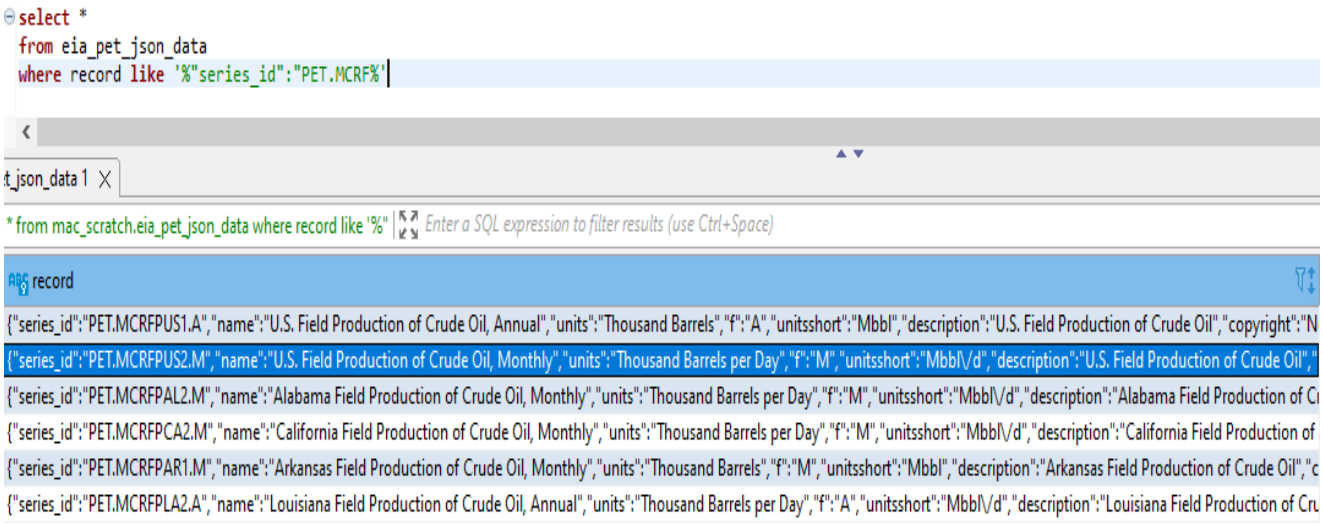
```
-rw-r--r--. 1 data_load domain users  49057013 Apr 27  2022 PET_04272022151234.zip
-rw-r--r--. 1 data_load domain users  48950099 May  2  2022 PET_04292022161423.zip
-rw-r--r--. 1 data_load domain users  48537833 May  4  2022 PET _05042022150901.zip
-rw-r--r--. 1 data_load domain users  48640003 May 12  2022 PET_05112022150903.zip
-rw-r--r--. 1 data_load domain users  48586184 May 19  2022 PET_05182022150804.zip
-rw-rw-rw-. 1 data_load domain users 334770876 May 25  2022 PET.txt
```

The petroleum data is formatted in JSON, and the EIA uses a unique system to represent this data, specifically through a 'series_id'. This 'series_id' is comprised of three segments separated by dots (.). Each segment conveys distinct and meaningful information. For instance, in the example series_id 'PET. MCRFPUS2. M', the first three letters 'PET' denote the file name, which stands for petroleum. The subsequent eight characters 'MCRFPUS2' following the first dot indicate the key, which represents data available on the EIA website concerning U. S. Field Production of Crude Oil. The final character, 'M', indicates the data's reporting frequency, which is monthly in this case.

Sample Data:


```
{"series_id":"PET.MCRFPUS2.M"
,"name":"U.S. Field Production of Crude Oil
, Monthly"
,"units":"Thousand Barrels per Day"
,"f":"M"
,"unitsshort":"Mbbl\/d"
,"description":"U.S. Field Production of Crude Oil"
,"copyright":"None"
,"source":"EIA, U.S. Energy Information Administration"
,"iso3166":"USA"
,"geography":"USA"
,"start":"192001"
,"end":"202202"
,"last_updated":"2022-04-29T11:00:30-04:00"
,"data":[["202202",11258],["202201",11480],["202112",11678]
,["202111",11782],["202110",11564],["202109",10921]
,["202108",11276],["202107",11392]]}
```


The data from the Petroleum file is loaded as - is which is basically raw data.



Each record is then loaded by a MySQL Procedure which calls an INSERT SQL to break down the json key value pairs into row and columns by monthly data into the target table.

```
INSERT INTO eia_pet_raw_data
(series_id,     name,   units,  frequency,  unitsshort,     description,
copyright,  source,     iso3166,    geography,  geography2,     start,
end,    last_updated,   date_value,     volume
)
SELECT
series_id,  name,   units,  frequency,  unitsshort,     description,
copyright,  source,     iso3166,    geography,  geography2,     start,
end,    last_updated,   r.date_value,   r.volume
FROM
(
with json_data
as
(
SELECT json_extract_string(record, 'series_id') as series_id
, json_extract_string(record, 'name') as name
, json_extract_string(record, 'units') as units
, json_extract_string(record, 'f') as frequency
, json_extract_string(record, 'unitsshort') as unitsshort
, json_extract_string(record, 'description') as description
, json_extract_string(record, 'copyright') as copyright
, json_extract_string(record, 'source') as source
, json_extract_string(record, 'iso3166') as iso3166
, json_extract_string(record, 'geography') as geography
, json_extract_string(record, 'geography2') as geography2
, json_extract_string(record, 'start') as start
, json_extract_string(record, 'end') as end
, json_extract_string(record, 'last_updated') as last_updated
, REPLACE(REPLACE(REPLACE(REPLACE(json_extract_string(record, 'data'), '[', ''), ']', ''), '",', '"~'), '"', '') as json_data
FROM eia_pet_json_data
WHERE record LIKE '%series_id%'
)
SELECT d.*
, SUBSTRING_INDEX(TRIM(SUBSTRING_INDEX(SUBSTRING_INDEX(csv_data, ',', l.level), ',', -1)), '~', 1) date_value
, SUBSTRING_INDEX(TRIM(SUBSTRING_INDEX(SUBSTRING_INDEX(csv_data, ',', l.level), ',', -1)), '~', -1) volume
FROM json_data d, mac_level l
WHERE l.level <= CHAR_LENGTH(csv_data) - CHAR_LENGTH(REPLACE(csv_data, ',', '')) + 1
) r;
```

After the data is loaded, the row and columns looks like as shown below

```
select series_id
  , name
  , frequency
  , date_value
  , volume
  from eia_pet_raw_data
  where series_id = 'PET.MCRFPUS2.M'
  and date_value like '2021%'
```

t_raw_data 1

series_id , name , frequency , date_value , volume from mac | Enter a SQL expression to filter results (use Ctrl+Space)

series_id	name	frequency	date_value	volume
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202101	11,137
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202102	9,916
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202103	11,351
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202104	11,318
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202105	11,390
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202106	11,366
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202107	11,392
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202108	11,276
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202109	10,921
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202110	11,564
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202111	11,782
PET.MCRFPUS2.M	U.S. Field Production of Crude Oil, Monthly	M	202112	11,678

The data is then loaded into the respective app to be used by end user for forecasting purpose.

Potential Extended Use Cases:
Higher resolution data, such as that provided by the EIA powers not only oil production forecasts but also more accurate energy price predictions and compliance with environmental regulations allowed strategic planning for long term regional renewables power investments. Beyond that, companies use this data in extended applications to reduce risks and be compliant with regulations or make decisions for their investment.

Impact:
When EIA data is incorporated into forecasting models, operational costs are reduced through lower levels of over production and stock inventory, while directly aligning supply with market demand. So, for example, a 10% increase in forecast accuracy could mean millions of dollars less sitting as inventory or being lost to stockouts.

Scope:
This research is dedicated to shaping quantitative forecasting models on the basis of historical data from (EIA) over the last ten years. It looks at how they can be applied in the current market, where there is room for improvement with existing methodologies and where things are not necessarily ideal to suggest some lines of further research - namely combining real - time data analytics and machine learning techniques that could yield significant improvements over predictions / forecasting.

3. Conclusion

The paper claims that EIA data is significant information for oil production forecasting, but its maximal utility requires both sophisticated statistical techniques and ongoing fine tuning of models. The industry needs to improve data analytics capabilities in keeping pace with changes occurred in marketplace and new technological improvements.

References

[1] U. S. Energy Information Administration, Available at https: //www.eia. gov/
[2] Bulk files download facility by different categories, Available at https: //www.eia. gov/opendata/v1/bulkfiles. php
[3] Petroleum file download link, Available at https: //www.eia. gov/opendata/bulk/PET. zip
[4] Crude Oil Production, Monthly Thousands Barrel Per day, Available at https: //www.eia. gov/dnav/pet/pet_crd_crpdn_adc_mbblpd_m. htm