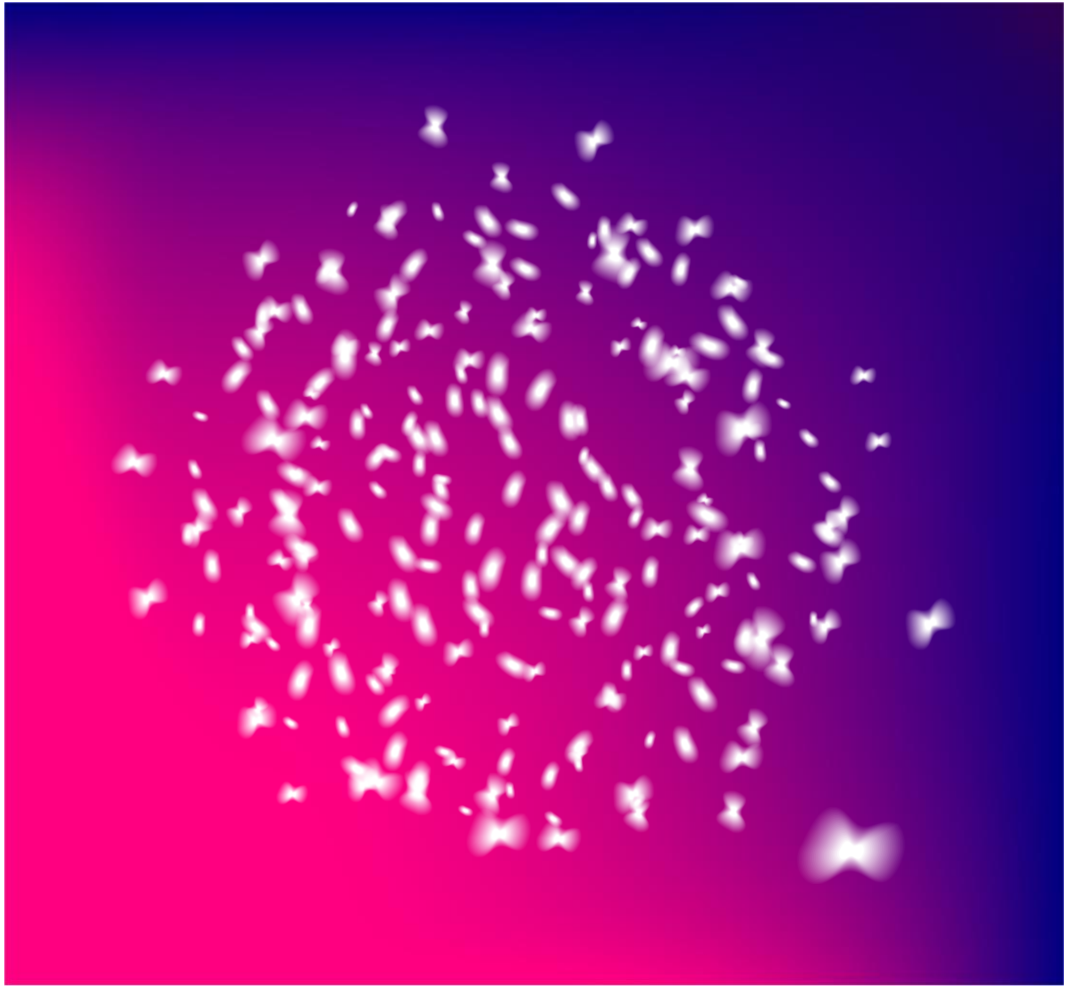


PyData Bristol

POLECAT™

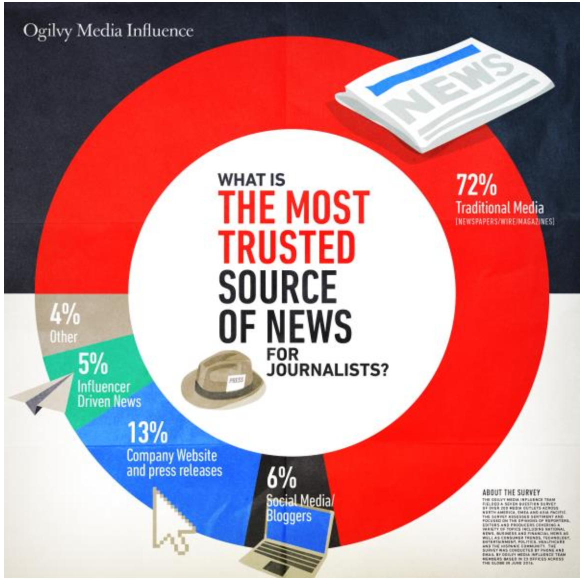
Coding with ChatGPT

Alon Samuel

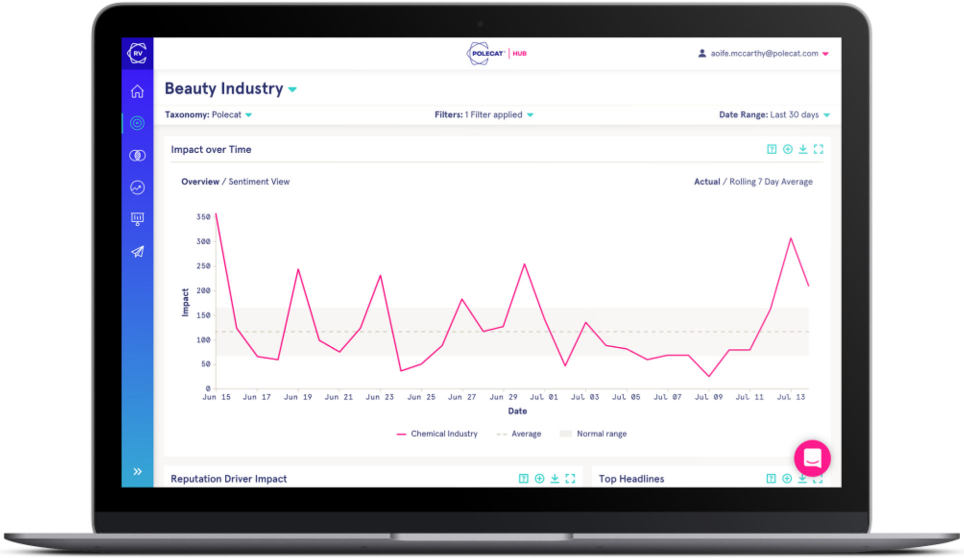
July 20th, 2023

Later: Live demo - open problem

Why is it needed – Polecat.com



This Photo by Unknown author is licensed under CC BY-SA-NC.



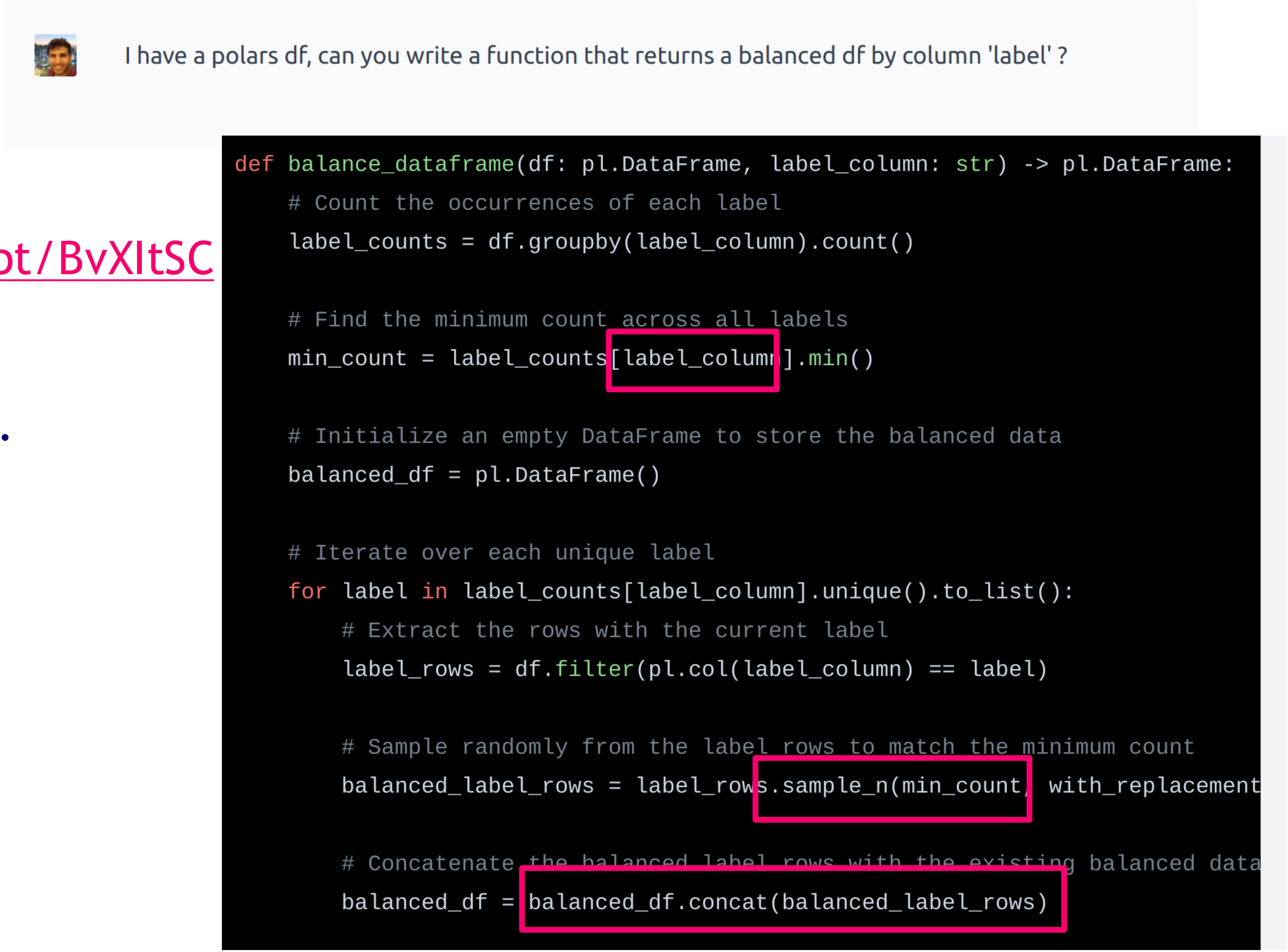
Taxonomy: Polecat
Filters: 1 Filter applied
Date Range: Last 30 days

Example 1

https://shareg.pt/BvXItSC

Using Polars package.

Mistakes are in Red



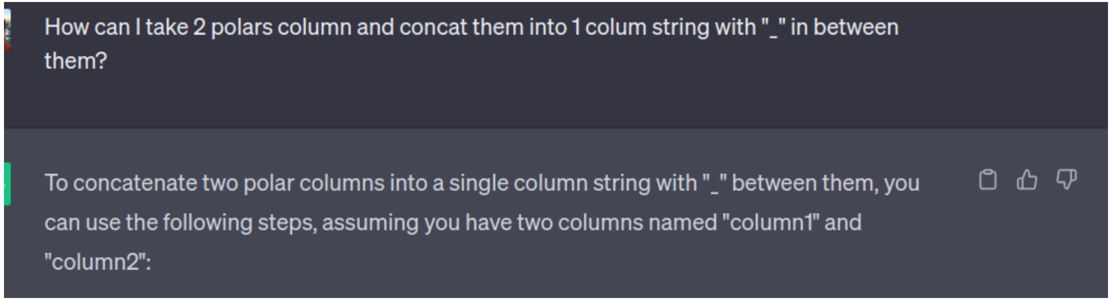
I have a polars df, can you write a function that returns a balanced df by column 'label' ?

```
def balance_dataframe(df: pl.DataFrame, label_column: str) -> pl.DataFrame:
    # Count the occurrences of each label
    label_counts = df.groupby(label_column).count()

    # Find the minimum count across all labels
    min_count = label_counts[label_column].min()

    # Initialize an empty DataFrame to store the balanced data
    balanced_df = pl.DataFrame()

    # Iterate over each unique label
    for label in label_counts[label_column].unique().to_list():
        # Extract the rows with the current label
        label_rows = df.filter(pl.col(label_column) == label)

        # Sample randomly from the label rows to match the minimum count
        balanced_label_rows = label_rows.sample_n(min_count, with_replacement

        # Concatenate the balanced label rows with the existing balanced data
        balanced_df = balanced_df.concat(balanced_label_rows)
```

Example 2

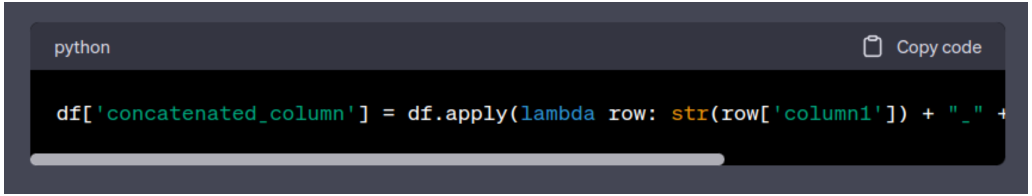
https://shareg.pt/IwtOe0W

Another Polars mistake



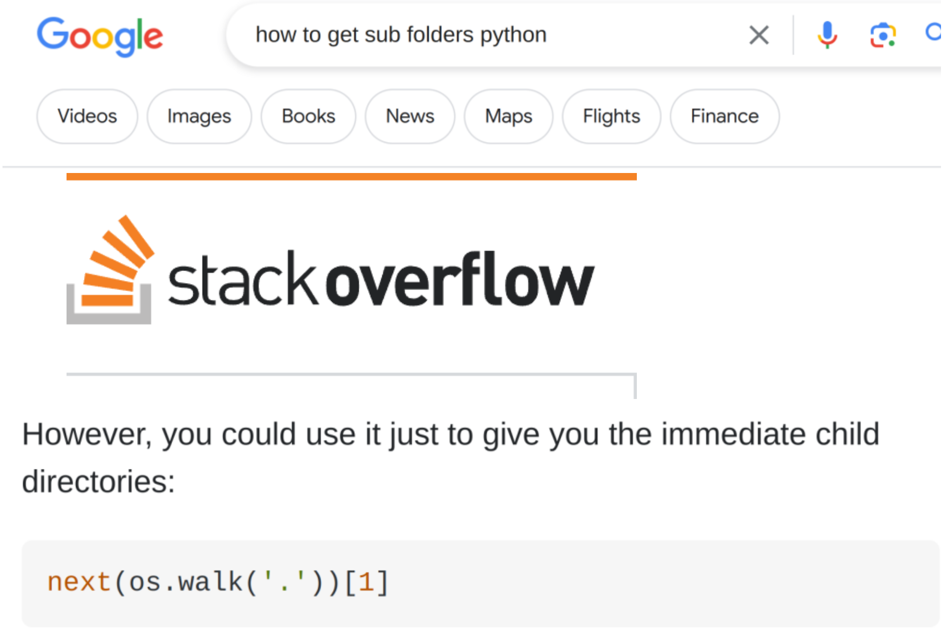
How can I take 2 polars column and concat them into 1 colum string with "_" in between them?

To concatenate two polar columns into a single column string with "_" between them, you can use the following steps, assuming you have two columns named "column1" and "column2":



```python
df['concatenated_column'] = df.apply(lambda row: str(row['column1']) + "_" +
```

Example 3

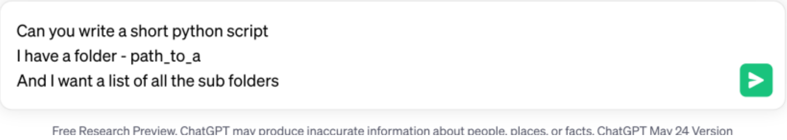


how to get sub folders python

However, you could use it just to give you the immediate child directories:

```
next(os.walk('.'))[1]
```

ChatGPT

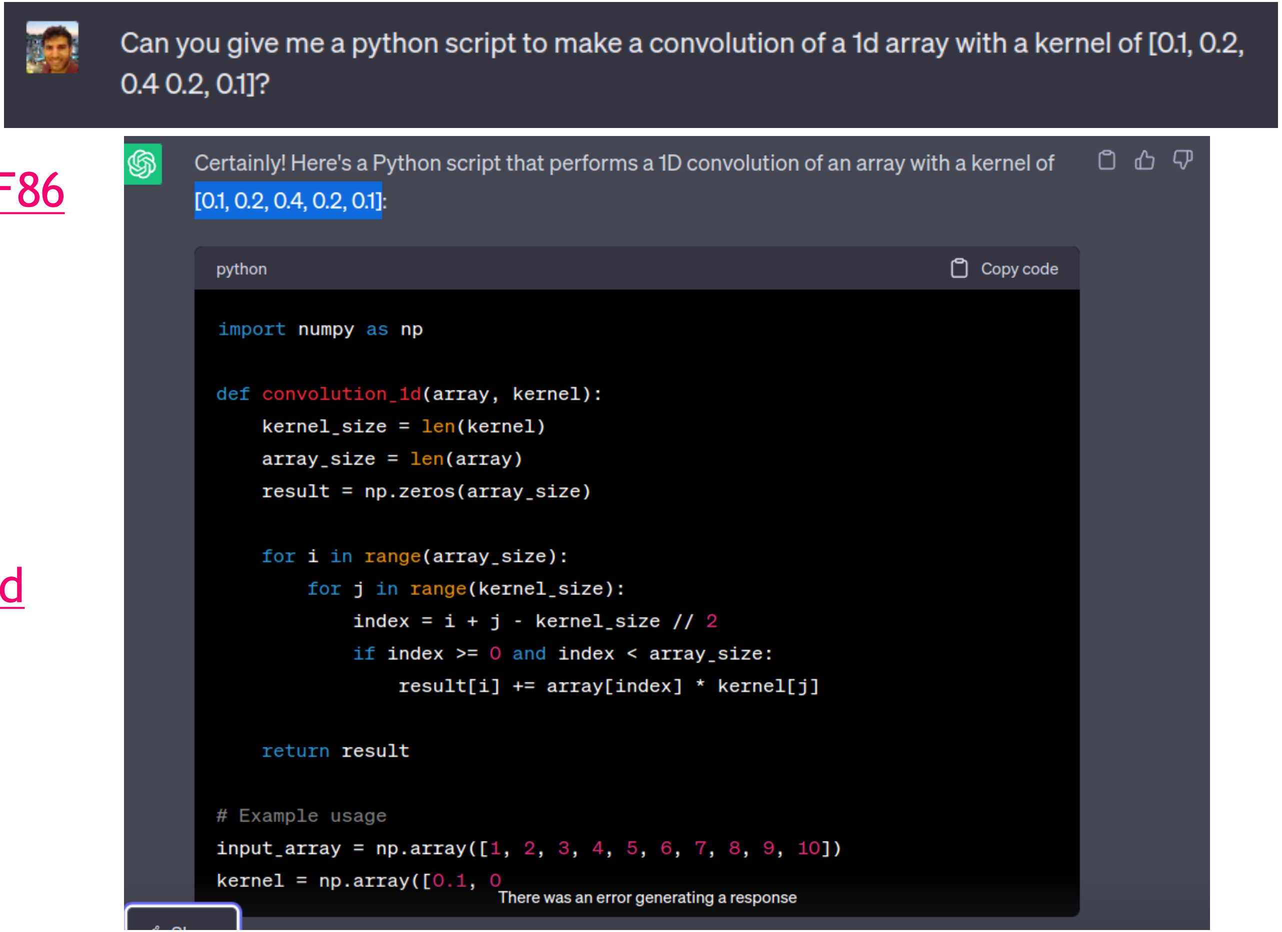


Can you write a short python script
I have a folder - path_to_a
And I want a list of all the sub folders

Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 24 Version

Example 4

https://shareg.pt/cjYvF86



Can you give me a python script to make a convolution of a 1d array with a kernel of [0.1, 0.2, 0.4 0.2, 0.1]?

Certainly! Here's a Python script that performs a 1D convolution of an array with a kernel of [0.1, 0.2, 0.4, 0.2, 0.1]:

```python
import numpy as np

def convolution_1d(array, kernel):
    kernel_size = len(kernel)
    array_size = len(array)
    result = np.zeros(array_size)

    for i in range(array_size):
        for j in range(kernel_size):
            index = i + j - kernel_size // 2
            if index >= 0 and index < array_size:
                result[i] += array[index] * kernel[j]

    return result

# Example usage
input_array = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
kernel = np.array([0.1, 0
```

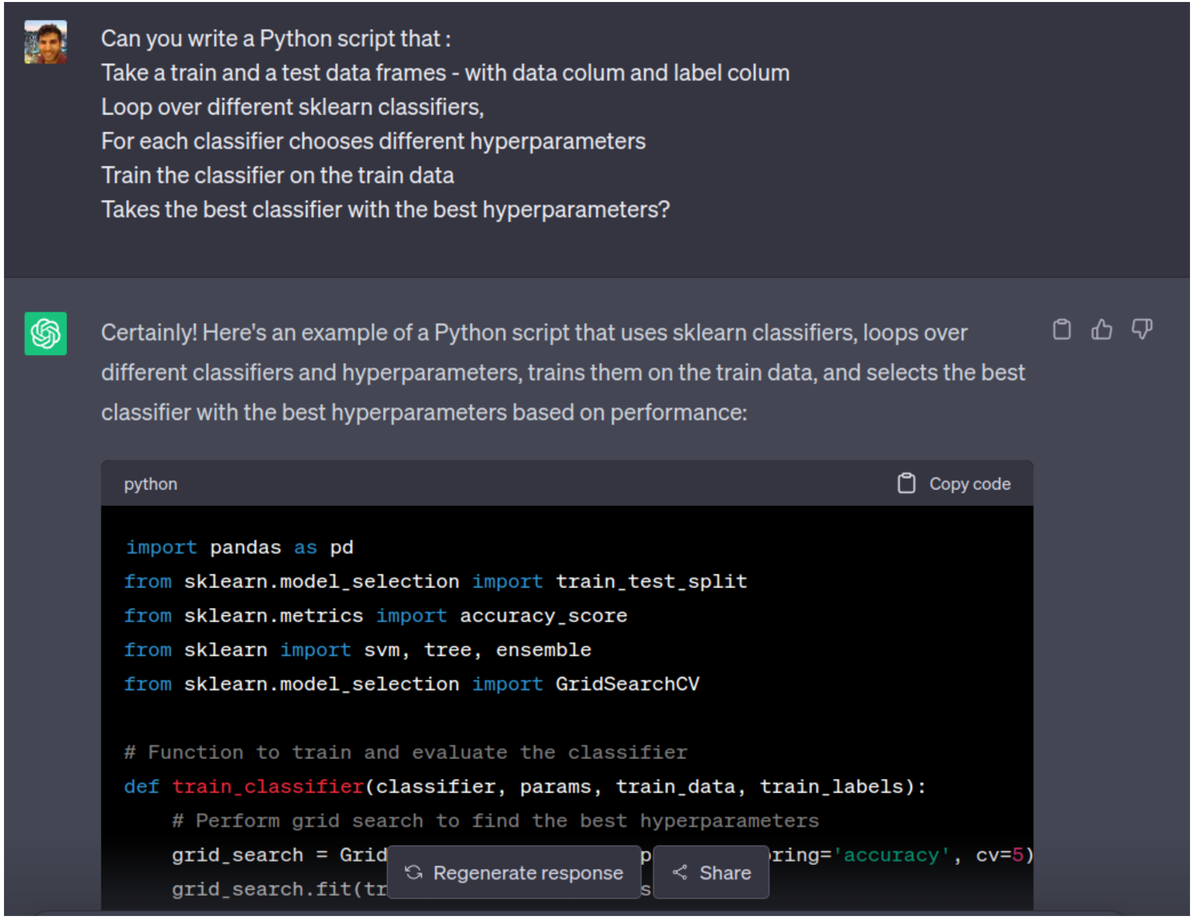
There was an error generating a response

Mistake with 1-d filter
There is a ready-made function to do this - scipy.ndimage.convolve1d

Example 5

https://sharegpt.com/c/isKVu5v

Creating a whole training script



Can you write a Python script that :
Take a train and a test data frames - with data colum and label colum
Loop over different sklearn classifiers,
For each classifier chooses different hyperparameters
Train the classifier on the train data
Takes the best classifier with the best hyperparameters?

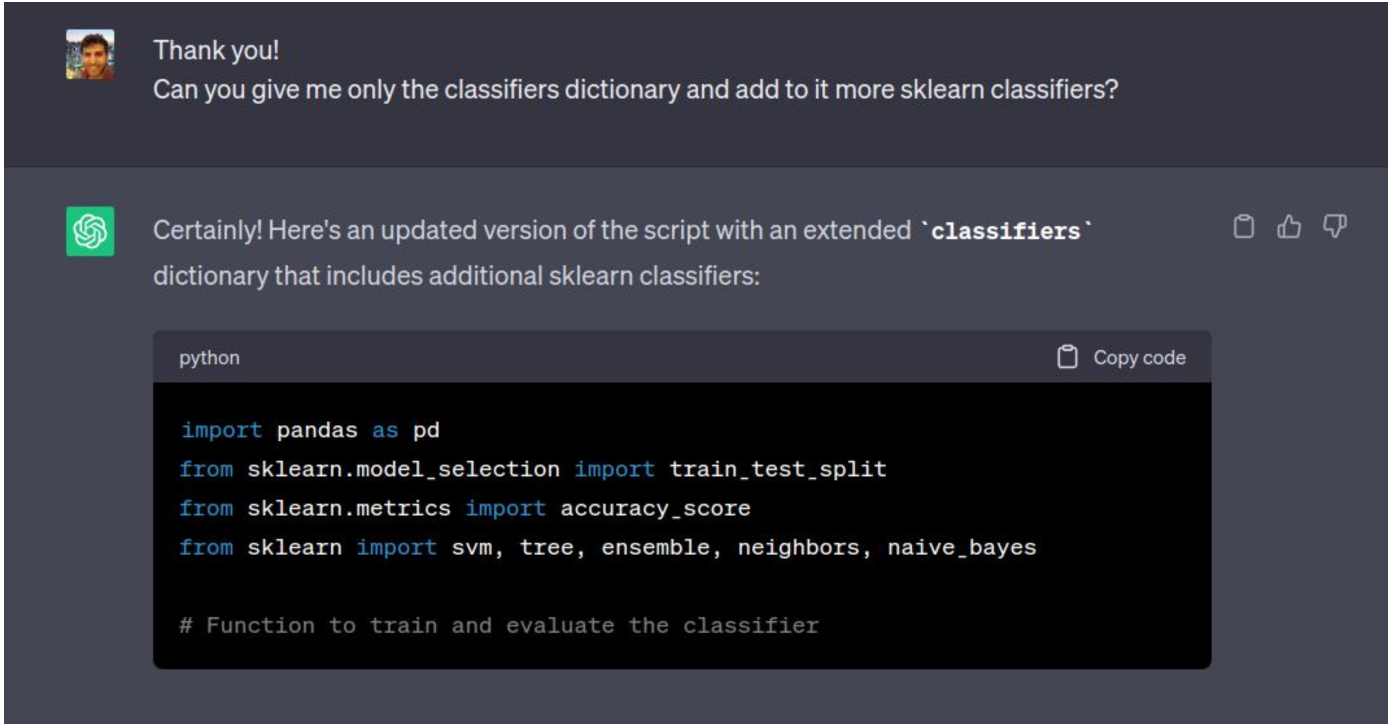
Certainly! Here's an example of a Python script that uses sklearn classifiers, loops over different classifiers and hyperparameters, trains them on the train data, and selects the best classifier with the best hyperparameters based on performance:

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn import svm, tree, ensemble
from sklearn.model_selection import GridSearchCV

# Function to train and evaluate the classifier
def train_classifier(classifier, params, train_data, train_labels):
    # Perform grid search to find the best hyperparameters
    grid_search = Grid                    p          ring='accuracy', cv=5)
    grid_search.fit(tr                    s
```

Example 5 Continue

sharegpt.com/c/isKVu5v

Oops!



Thank you!
Can you give me only the classifiers dictionary and add to it more sklearn classifiers?

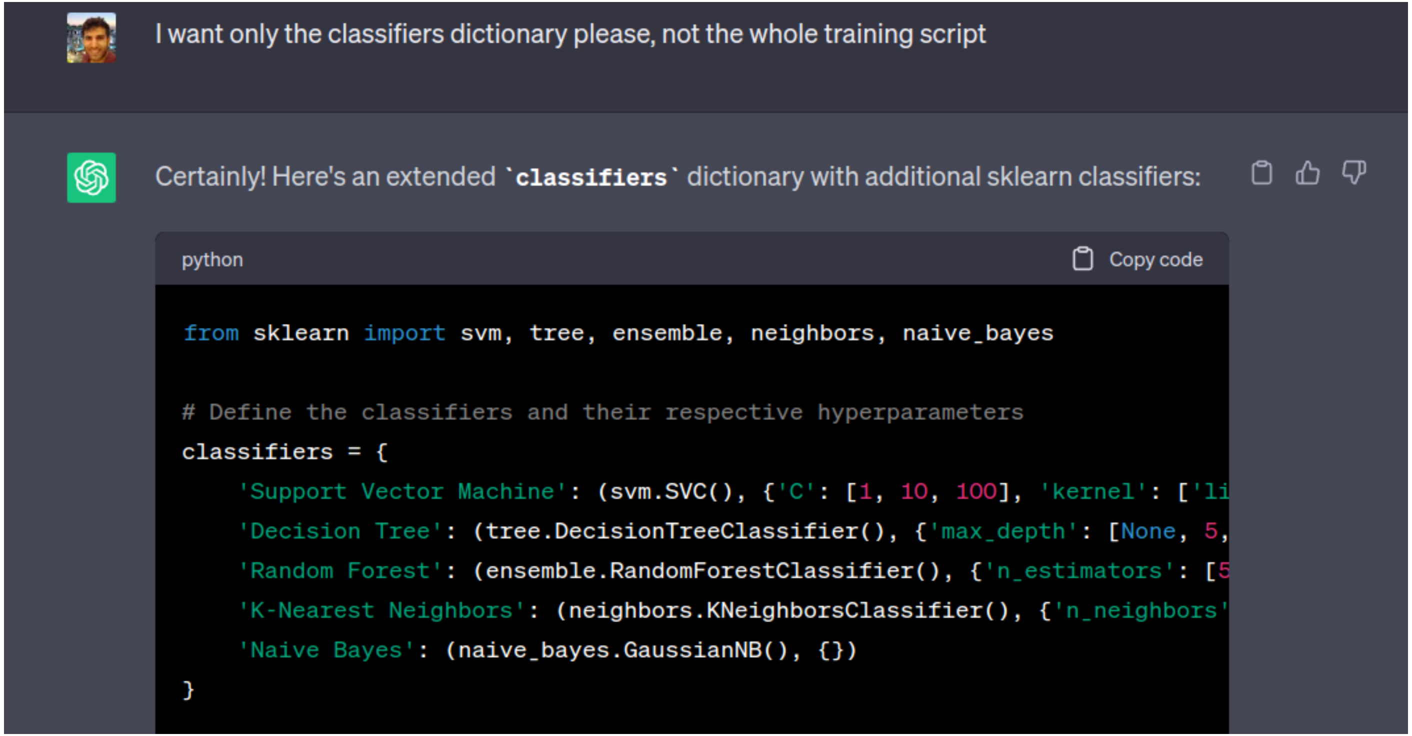
Certainly! Here's an updated version of the script with an extended `classifiers` dictionary that includes additional sklearn classifiers:

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn import svm, tree, ensemble, neighbors, naive_bayes

# Function to train and evaluate the classifier
```

Example 5 Continue

sharegpt.com/c/isKVu5v

That's better...



I want only the classifiers dictionary please, not the whole training script

Certainly! Here's an extended `classifiers` dictionary with additional sklearn classifiers:

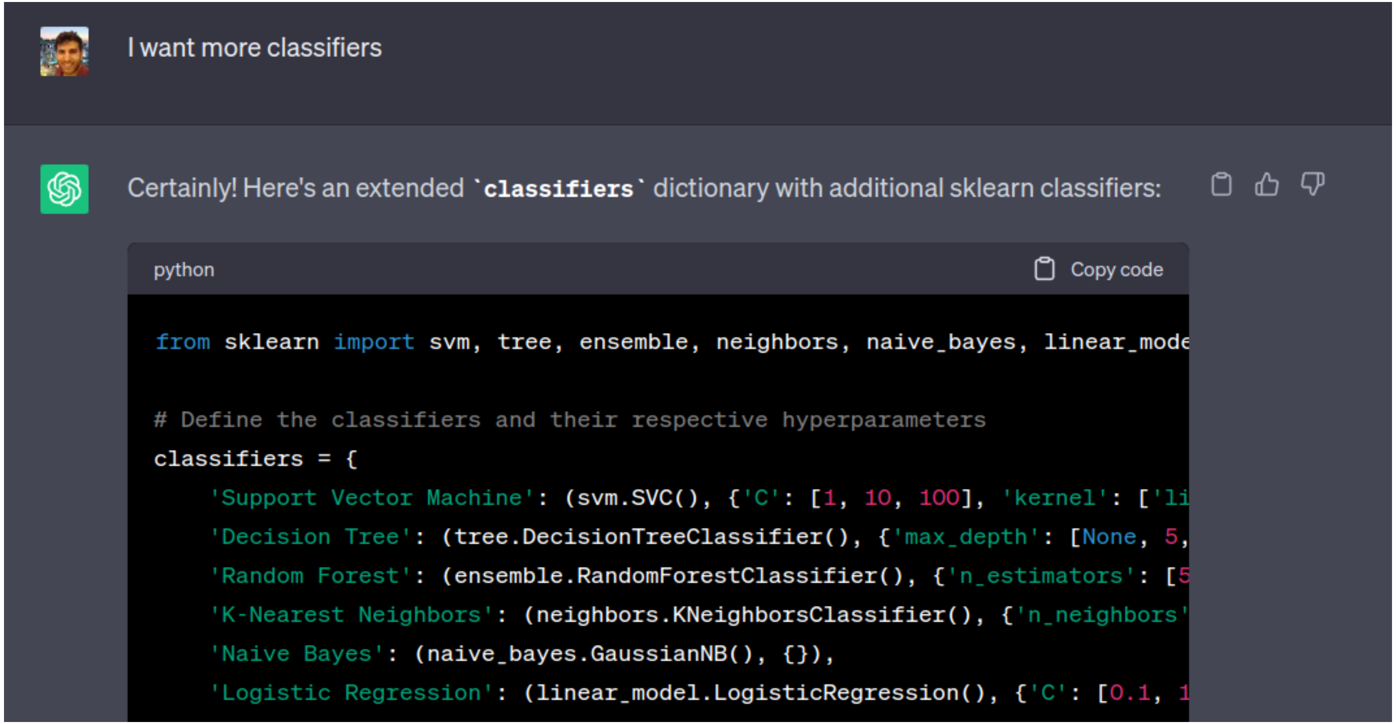
```python
from sklearn import svm, tree, ensemble, neighbors, naive_bayes

# Define the classifiers and their respective hyperparameters
classifiers = {
    'Support Vector Machine': (svm.SVC(), {'C': [1, 10, 100], 'kernel': ['li
    'Decision Tree': (tree.DecisionTreeClassifier(), {'max_depth': [None, 5,
    'Random Forest': (ensemble.RandomForestClassifier(), {'n_estimators': [5
    'K-Nearest Neighbors': (neighbors.KNeighborsClassifier(), {'n_neighbors'
    'Naive Bayes': (naive_bayes.GaussianNB(), {})
}
```

Example 5 Continue

sharegpt.com/c/isKVu5v

More

Also works
with HuggingFace



I want more classifiers

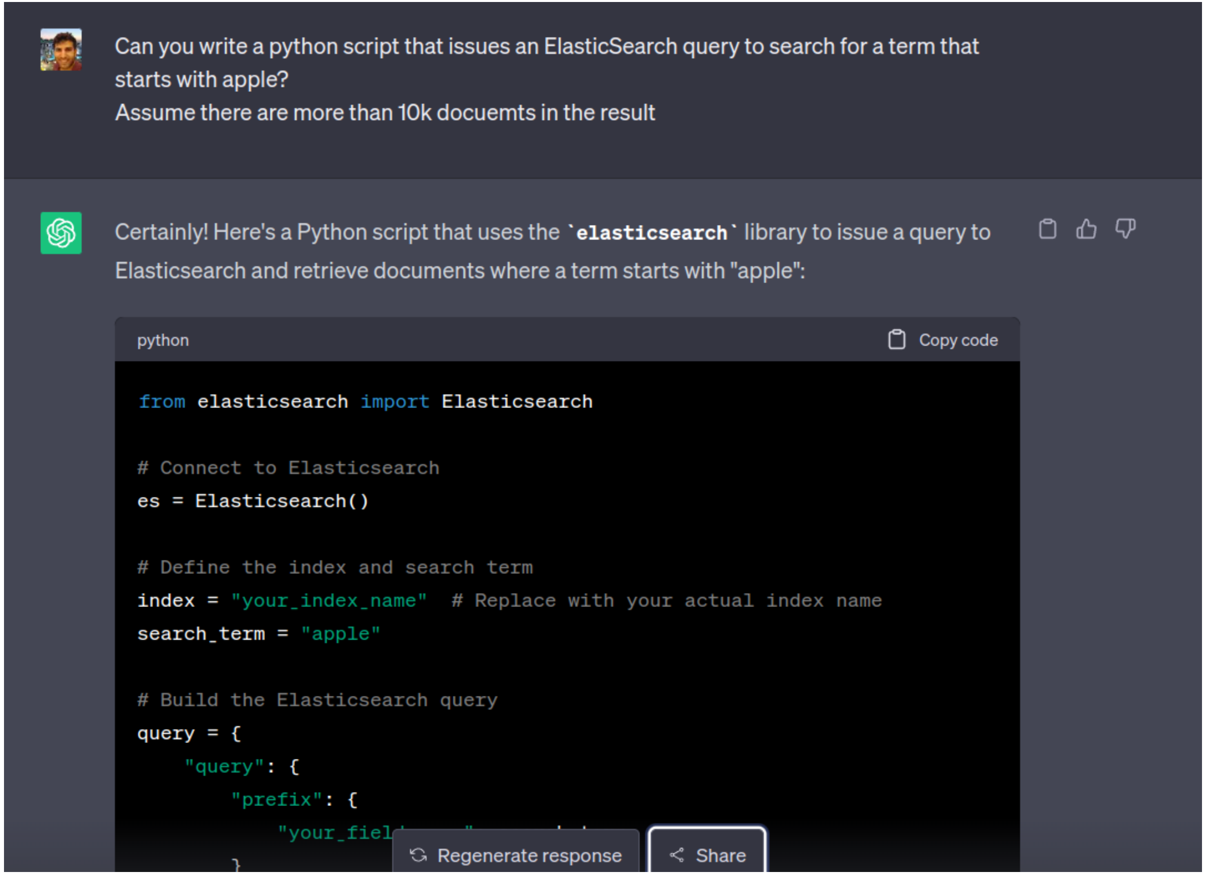
Certainly! Here's an extended `classifiers` dictionary with additional sklearn classifiers:

```python
from sklearn import svm, tree, ensemble, neighbors, naive_bayes, linear_mode

# Define the classifiers and their respective hyperparameters
classifiers = {
    'Support Vector Machine': (svm.SVC(), {'C': [1, 10, 100], 'kernel': ['li
    'Decision Tree': (tree.DecisionTreeClassifier(), {'max_depth': [None, 5,
    'Random Forest': (ensemble.RandomForestClassifier(), {'n_estimators': [5
    'K-Nearest Neighbors': (neighbors.KNeighborsClassifier(), {'n_neighbors'
    'Naive Bayes': (naive_bayes.GaussianNB(), {}),
    'Logistic Regression': (linear_model.LogisticRegression(), {'C': [0.1, 1
```

Example 6

https://sharegpt.com/c/x2zO7jj

Using ElasticSearch



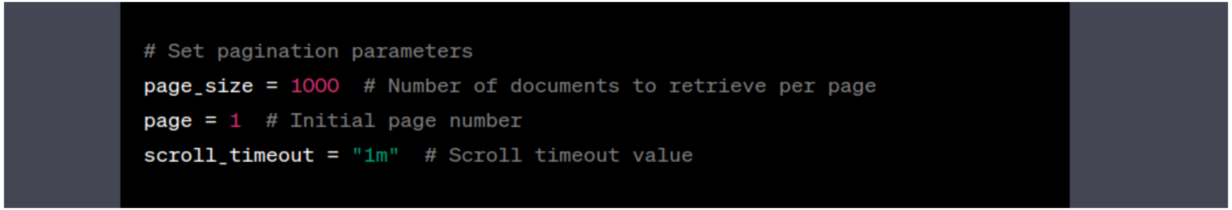
Can you write a python script that issues an ElasticSearch query to search for a term that starts with apple?
Assume there are more than 10k docuemts in the result

Certainly! Here's a Python script that uses the `elasticsearch` library to issue a query to Elasticsearch and retrieve documents where a term starts with "apple":

```python
from elasticsearch import Elasticsearch

# Connect to Elasticsearch
es = Elasticsearch()

# Define the index and search term
index = "your_index_name"  # Replace with your actual index name
search_term = "apple"

# Build the Elasticsearch query
query = {
    "query": {
        "prefix": {
            "your_fiel
        }
```

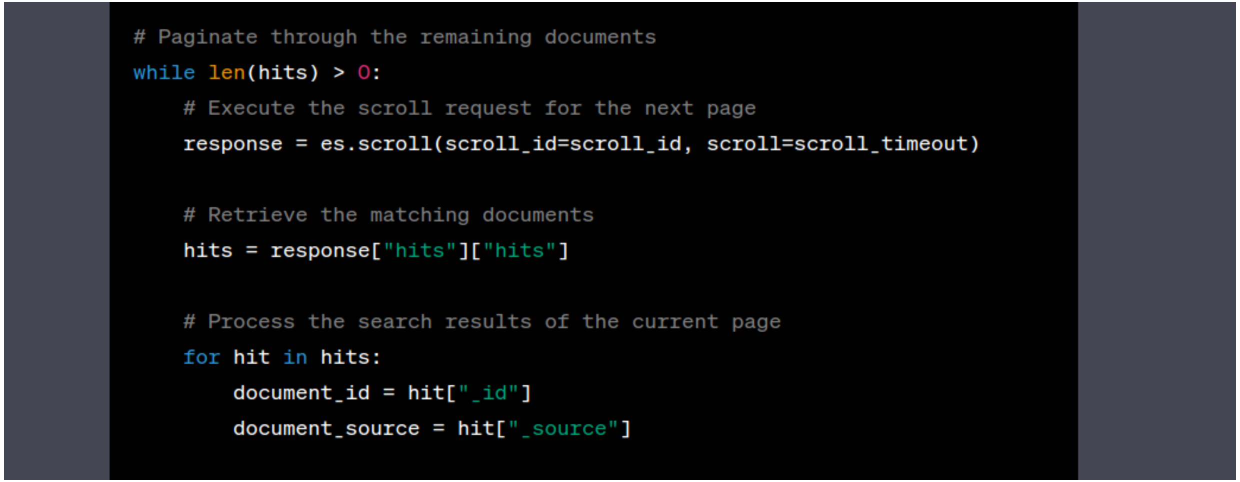
Example 6 Continue

https://sharegpt.com/c/x2zO7jj



```
# Set pagination parameters
page_size = 1000  # Number of documents to retrieve per page
page = 1  # Initial page number
scroll_timeout = "1m"  # Scroll timeout value
```

Using ElasticSearch
With results pagination



```
# Paginate through the remaining documents
while len(hits) > 0:
    # Execute the scroll request for the next page
    response = es.scroll(scroll_id=scroll_id, scroll=scroll_timeout)

    # Retrieve the matching documents
    hits = response["hits"]["hits"]

    # Process the search results of the current page
    for hit in hits:
        document_id = hit["_id"]
        document_source = hit["_source"]
```

Live demo - open problem?

Links

- Chat: chat.openai.com
- Prompt engineering - link
  - Notebooks
- Code with Python - link

This Photo by Unknown author is licensed under CC BY-SA-NC.

POLECAT™

For further information, please contact:

✉ Alon.samuel@polecat.com