

# Worksheet: Automated Feature Extraction via LLMs

---

This worksheet guides you through using LangChain to enrich a Pandas DataFrame with new features extracted from raw text.

## Step 1: Setup & Mock Data

```
import pandas as pd
df = pd.DataFrame({'feedback': ["The battery is terrible.", "Amazing service!", "Too expensive."]}))
```

## Step 2: Define the Extraction Chain

```
from langchain_openai import ChatOpenAI
from langchain_core.prompts import ChatPromptTemplate
from langchain_core.output_parsers import StrOutputParser

prompt = ChatPromptTemplate.from_template("Classify the sentiment of this text as Positive, Negative, or Neutral: {input}")
llm = ChatOpenAI(model="gpt-4o", temperature=0)
chain = prompt | llm | StrOutputParser()
```

## Step 3: Enrich the DataFrame

```
df['sentiment'] = [chain.invoke({"input": text}) for text in df['feedback']]
print(df)
```

## Exercise

Try extending this pipeline to extract a second feature (e.g., "Category") from the feedback and add it to the DataFrame.

### Sample Answer

You can create a second chain or modify your prompt to return both values. Here is a simple approach using a separate chain:

```
# Second chain for Category
cat_prompt = ChatPromptTemplate.from_template("Classify the
feedback category (Hardware, Service, Pricing): {input}")
cat_chain = cat_prompt | llm | StrOutputParser()

# Apply to DataFrame
df['category'] = [cat_chain.invoke({"input": text}) for text in
df['feedback']]
print(df)
```