

Jupyter Notebook Execution Report

Name: Jatin Narayan Bhoir
Project Title: Fake Job Posting Detection
Date: December 29, 2025

Cell 1: ■ Code

```
import pandas as pd

df = pd.read_csv("fake_job_postings.csv")
df.head()
```

Output:

```
job_id    ...   fraudulent
0         1   ...        0
1         2   ...        0
2         3   ...        0
3         4   ...        0
4         5   ...        0

[5 rows x 18 columns]
```

Cell 2: ■ Code

```
df.info()

RangeIndex: 17880 entries, 0 to 17879
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
  0   job_id          17880 non-null   int64  
  1   title            17880 non-null   object  
  2   location         17534 non-null   object  

```

```
3    department            6333 non-null  object
4    salary_range          2868 non-null  object
5    company_profile        14572 non-null  object
6    description           17879 non-null  object
7    requirements          15184 non-null  object
8    benefits              10668 non-null  object
9    telecommuting         17880 non-null  int64
10   has_company_logo     17880 non-null  int64
11   has_questions         17880 non-null  int64
12   employment_type       14409 non-null  object
13   required_experience  10830 non-null  object
14   required_education    9775 non-null  object
15   industry              12977 non-null  object
16   function              11425 non-null  object
17   fraudulent             17880 non-null  int64
dtypes: int64(5), object(13)
memory usage: 2.5+ MB
```

Cell 3: ■ Code

```
df['fraudulent'].value_counts()
```

Output:

```
fraudulent
0      17014
1       866
Name: count, dtype: int64
```

Cell 4: ■ Code

```
df.isnull().sum()
```

Output:

```
job_id                  0
title                   0
location                346
```

```
department           11547
salary_range        15012
company_profile     3308
description          1
requirements         2696
benefits            7212
telecommuting        0
has_company_logo      0
has_questions         0
employment_type       3471
required_experience    7050
required_education      8105
industry              4903
function                6455
fraudulent              0
dtype: int64
```

Cell 5: ■ Code

```
df['text'] = (
    df['title'].fillna('') + ' ' +
    df['company_profile'].fillna('') + ' ' +
    df['description'].fillna('') + ' ' +
    df['requirements'].fillna('')
)
```

Cell 6: ■ Code

```
pip install nltk
```

Error:

```
Traceback (most recent call last):
  File "c:\Users\JATIN\.vscode\extensions\ganeshkumbhar.nb2pdf-1.1.9\scripts\nb2pdf.py", line 403
    result = eval(lines[-1], glb)
  File "<string>", line 1
    pip install nltk
    ^^^^^^
SyntaxError: invalid syntax
```

```
During handling of the above exception, another exception occurred:
```

```
Traceback (most recent call last):
```

```
  File "c:\Users\JATIN\vscode\extensions\ganeshkumbhar.nb2pdf-1.1.9\scripts\nb2pdf.py", line 408
    exec(source, glb)
  File "<string>", line 1
    pip install nltk
    ^^^^^^
SyntaxError: invalid syntax
```

Cell 7: ■ Code

```
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
nltk.download('stopwords')
```

Output:

```
True
[ STDERR ]
[nltk_data]  Downloading package stopwords to
[nltk_data]      C:\Users\JATIN\AppData\Roaming\nltk_data...
[nltk_data]  Package stopwords is already up-to-date!
```

Cell 8: ■ Code

```
stop_words = set(stopwords.words('english'))
ps = PorterStemmer()

def clean_text(text):
    text = text.lower()
    text = re.sub(r'http\S+|www\S+|https\S+', '', text)
    text = re.sub(r'<.*?>', '', text)
    text = re.sub(r'^[a-zA-Z]', ' ', text)
    words = text.split()
    words = [ps.stem(w) for w in words if w not in stop_words]
    return " ".join(words)
```

Cell 9: ■ Code

```
df['clean_text'] = df['text'].apply(clean_text)  
df[['text', 'clean_text']].head()
```

Output:

	text	clean_text
0	Marketing Intern We're Food52, and we've creat...	market intern food creat groundbreak award
1	Customer Service - Cloud Video Production 90 S...	custom servic cloud video product second wo
2	Commissioning Machinery Assistant (CMA) Valor ...	commis machineri assist cma valor servic p
3	Account Executive - Washington DC Our passion ...	account execut washington dc passion improv
4	Bill Review Manager SpotSource Solutions LLC i...	bill review manag spotsourc solut llc globa

Cell 10: ■ Code

```
df.to_csv("cleaned_fake_job_postings_cleaned.csv", index=False)
```