

```

1 import sys
2 import pandas as pd
3 import numpy as np
4 import random
5 import re
6 import joblib
7 import json
8 from sklearn.feature_extraction.text import CountVectorizer
9 from sklearn.naive_bayes import MultinomialNB
10 from sklearn.metrics import accuracy_score
11
12 def load_data(csvfile):
13     """
14     Read in CSV file data
15     *This function assumes held-out data has the same format as training data
16
17     """
18
19     df = pd.read_csv(csvfile)
20
21     ids = np.array(df["complaint_id"])
22     product_groups = np.array(df["product_group"])
23     text_list = np.array(df["text"])
24
25     return ids, product_groups, text_list
26
27 def vectorize_documents(document_list, vectorizer):
28     """
29     Function to vectorize input text documents
30
31     """
32     X = vectorizer.transform(document_list)
33
34     return X
35
36 if __name__ == "__main__":
37     input_data_file = sys.argv[1]
38
39     print("\nLoading Data...")
40     ids, product_groups, text_list = load_data(input_data_file)
41
42     print("\nLoading Group Labels...")
43     with open("group_mapping.json", "r") as file:
44         group_to_label = json.load(file) # Get mapping for model numbers to group labels
45
46     print("\nLoading Vectorizer...")
47     vectorizer = joblib.load("NWCaseStudyvectorizer.joblib")
48
49     print("\nLoading Model...")
50     clf = joblib.load("NWCaseStudyNBModel.joblib")
51
52     print("\nVectorizing Input Data...")
53     X = vectorize_documents(text_list, vectorizer)
54     y = np.array([group_to_label[group] for group in product_groups])
55
56     print("\nMaking Predictions...")
57     y_pred = clf.predict(X)
58
59     acc = accuracy_score(y_pred, y)
60     print("Model Accuracy: %f" %acc)

```