In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.preprocessing import LabelEncoder
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.optimizers import Adam
```

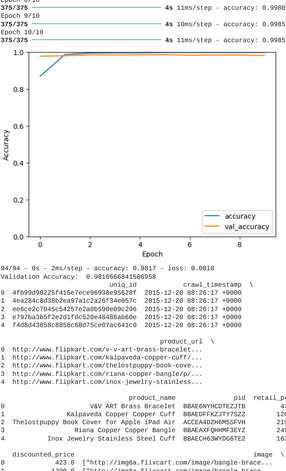
In [5]:
```
# Display the column names
print(train_data.columns)
```

```
Index(['uniq_id', 'crawl_timestamp', 'product_url', 'product_name',
       'product_category_tree', 'pid', 'retail_price', 'discounted_price',
       'image', 'is_FK_Advantage_product', 'description', 'product_rating',
       'overall_rating', 'brand', 'product_specifications'],
      dtype='object')
```

In [6]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.optimizers import Adam
import matplotlib.pyplot as plt
import joblib

# Load the dataset
train_data = pd.read_csv("/Users/varunkumar/Downloads/hackathon/train_product_data.csv")
test_data = pd.read_csv("/Users/varunkumar/Downloads/hackathon/test_data.csv")

# Display basic information about the dataset
print(train_data.head())
print(train_data.columns)

# Handle missing values
train_data.dropna(subset=['description', 'product_category_tree'], inplace=True)

# Extract category from product_category_tree
train_data['category'] = train_data['product_category_tree'].apply(lambda x: x.split('>>')[0].strip())

# Encode the labels
label_encoder = LabelEncoder()
train_data['category'] = label_encoder.fit_transform(train_data['category'])

# TF-IDF Vectorization
tfidf = TfidfVectorizer(max_features=5000, stop_words='english')
X = tfidf.fit_transform(train_data['description']).toarray()
y = train_data['category']

# Split data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Logistic Regression Model
log_reg = LogisticRegression(max_iter=1000)
log_reg.fit(X_train, y_train)

# Predictions and Evaluation
y_pred = log_reg.predict(X_val)
print("Accuracy:", accuracy_score(y_val, y_pred))
print("Classification Report:\n", classification_report(y_val, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_val, y_pred))

# Neural Network Model
model = Sequential([
    Dense(512, activation='relu', input_shape=(X_train.shape[1],)),
    Dropout(0.3),
    Dense(256, activation='relu'),
    Dropout(0.3),
    Dense(len(np.unique(y)), activation='softmax')
])

model.compile(optimizer=Adam(learning_rate=0.001), loss='sparse_categorical_crossentropy', metrics=['accuracy'])
history = model.fit(X_train, y_train, epochs=10, batch_size=32, validation_data=(X_val, y_val))

# Plot training history
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label='val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0, 1])
plt.legend(loc='lower right')
plt.show()

# Evaluate on validation set
val_loss, val_acc = model.evaluate(X_val, y_val, verbose=2)
print("Validation Accuracy: ", val_acc)

# Preprocess test data
test_data['description'].fillna('', inplace=True)
test_data_tfidf = tfidf.transform(test_data['description']).toarray()

# Predict using the trained logistic regression model
test_pred = log_reg.predict(test_data_tfidf)
test_data['predicted_category'] = label_encoder.inverse_transform(test_pred)
print(test_data.head())

# Save the trained model
joblib.dump(log_reg, 'logistic_regression_model.pkl')

# Save the label encoder
joblib.dump(label_encoder, 'label_encoder.pkl')
```

```
                   uniq_id          crawl_timestamp  \
0  c2d766ca982eca8304150849735ffef9  2016-03-25 22:59:23 +0000
1  f449ec65dcbc041b6ae5e6a32717d01b  2016-03-25 22:59:23 +0000
2  0973b37acd0c664e3de26e97e5571454  2016-03-25 22:59:23 +0000
3  ce5a6818f7707e2cb61fdcdbba615f3ad  2016-03-25 22:59:23 +0000
4  29c8d290ceaa451f97b1c32df64477a2c  2016-03-25 22:59:23 +0000

                                         product_url  \
0  http://www.flipkart.com/alisha-solid-women-s-c...
1  http://www.flipkart.com/aw-bellies/p/itmeh4grg...
2  http://www.flipkart.com/alisha-solid-women-s-c...
3  http://www.flipkart.com/alisha-solid-women-s-c...
4  http://www.flipkart.com/dilli-bazaar-bellies-...

                                        product_name product_category_tree  \
0              Alisha Solid Women's Cycling Shorts            Clothing
1                                    AW Bellies           Footwear
2              Alisha Solid Women's Cycling Shorts            Clothing
3              Alisha Solid Women's Cycling Shorts            Clothing
4  dilli bazaaar Bellies, Corporate Casuals, Casuals           Footwear

          pid  retail_price  discounted_price  \
0  SRTEH2FF9KEDEFGF          999.0             379.0
1  SHOEH4GRSUBJGZXE          999.0             499.0
2  SRTEH2FGHUZMQ6SJ          699.0             267.0
3  SRTEH2FVVKRBAXHB         1199.0             479.0
4  SHOEH3DZBFR8E5CK          699.0             349.0

                                               image  is_FK_Advantage_product  \
0  ["http://img5a.flixcart.com/image/short/u/4/a/...                    False
1  ["http://img5a.flixcart.com/image/shoe/7/z/z/r...                    False
2  ["http://img5a.flixcart.com/image/short/6/2/h/...                    False
3  ["http://img6a.flixcart.com/image/short/p/j/z/...                    False
4  ["http://img6a.flixcart.com/image/shoe/b/p/n/p...                    False

                                         description  product_rating  \
0  Key Features of Alisha Solid Women's Cycling S...  No rating available
1  Key Features of AW Bellies Sandals Wedges Heel...  No rating available
2  Key Features of Alisha Solid Women's Cycling S...  No rating available
3  Key Features of Alisha Solid Women's Cycling S...  No rating available
4  Key Features of dilli bazaaar Bellies, Corpora...  No rating available

       overall_rating          brand  \
0  No rating available        Alisha
1  No rating available            AW
2  No rating available        Alisha
3  No rating available        Alisha
4  No rating available  dilli bazaaar

                                product_specifications
0  {"product_specification"=>[{"key"=>"Number of ...
1  {"product_specification"=>[{"key"=>"Ideal For"...
2  {"product_specification"=>[{"key"=>"Number of ...
3  {"product_specification"=>[{"key"=>"Number of ...
4  {"product_specification"=>[{"key"=>"Occasion",...
Index(['uniq_id', 'crawl_timestamp', 'product_url', 'product_name',
       'product_category_tree', 'pid', 'retail_price', 'discounted_price',
       'image', 'is_FK_Advantage_product', 'description', 'product_rating',
       'overall_rating', 'brand', 'product_specifications'],
      dtype='object')
Accuracy: 0.9703333333333334
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       198
           1       0.93      0.75      0.83        51
           2       0.93      0.84      0.88        31
           3       0.97      1.00      0.98      1083
           4       0.92      0.92      0.92       107
           5       1.00      0.97      0.98       220
           6       0.95      0.97      0.96       154
           7       0.99      1.00      1.00       608
           8       0.98      0.94      0.96       114
           9       0.93      0.96      0.95       171
          10       0.96      0.66      0.78        41
          11       1.00      0.95      0.97        79
          12       0.78      0.79      0.78        48
          13       1.00      1.00      1.00        95

    accuracy                           0.97      3000
   macro avg       0.95      0.91      0.93      3000
weighted avg       0.97      0.97      0.97      3000

Confusion Matrix:
 [[ 191    0    0    0    2    0    1    0    0    4    0    0    0    0]
 [   1   38    0    0    0    0    2    2    0    0    0    0    0    0]
 [   0    0   26    5    0    0    0    0    0    0    0    0    0    0]
 [   0    1    0 1081    0    0    1    0    0    0    0    0    0    0]
 [   0    0    0    1    2   98    0    0    0    0    6    0    0    0]
 [   0    0    0    0    5    1  214    0    0    0    0    0    0    0]
 [   0    0    0    0    2    0    0  150    1    1    0    0    0    0]
 [   0    0    0    0    1    0    0    0  607    0    0    0    0    0]
 [   0    2    0    2    0    0    2    0  107    1    0    0    0    0]
 [   0    0    0    2    5    0    0    0    0  164    0    0    0    0]
 [   0    0    0    1    0    0    0    1    0    1   27    0   11    0]
 [   0    0    0    2    0    0    1    0    1    0    0   75    0    0]
 [   3    0    1    2    0    1    2    0    0    0    1    0   38    0]
 [   0    0    0    0    0    0    0    0    0    0    0    0    0   95]]
```

```
/Users/varunkumar/anaconda3/lib/python3.11/site-packages/keras/src/layers/core/dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When using Sequential models, prefer using an `Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
Epoch 1/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 5s 12ms/step - accuracy: 0.7005 - loss: 1.0314 - val_accuracy: 0.9780 - val_loss: 0.0773
Epoch 2/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 10ms/step - accuracy: 0.9884 - loss: 0.0427 - val_accuracy: 0.9807 - val_loss: 0.0585
Epoch 3/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9958 - loss: 0.0170 - val_accuracy: 0.9853 - val_loss: 0.0588
Epoch 4/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9966 - loss: 0.0117 - val_accuracy: 0.9850 - val_loss: 0.0612
Epoch 5/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9972 - loss: 0.0089 - val_accuracy: 0.9857 - val_loss: 0.0629
Epoch 6/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9979 - loss: 0.0061 - val_accuracy: 0.9863 - val_loss: 0.0630
Epoch 7/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 10ms/step - accuracy: 0.9983 - loss: 0.0051 - val_accuracy: 0.9837 - val_loss: 0.0694
Epoch 8/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9980 - loss: 0.0057 - val_accuracy: 0.9847 - val_loss: 0.0712
Epoch 9/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 10ms/step - accuracy: 0.9985 - loss: 0.0038 - val_accuracy: 0.9847 - val_loss: 0.0725
Epoch 10/10
375/375 ━━━━━━━━━━━━━━━━━━━━ 4s 11ms/step - accuracy: 0.9985 - loss: 0.0060 - val_accuracy: 0.9817 - val_loss: 0.0810
```




```
94/94 - 0s - 2ms/step - accuracy: 0.9817 - loss: 0.0810
Validation Accuracy:  0.9816666841506958
                   uniq_id          crawl_timestamp  \
0  4fb99d98225f415e7ece96938e95628f  2015-12-20 08:26:17 +0000
1  4ea284c8d38b2ea97a1c2a26f34e057c  2015-12-20 08:26:17 +0000
2  ee6c2c7045c54257e2ab059d0e09c796  2015-12-20 08:26:17 +0000
3  e797ba3b5f2e2d1fdc520e484864ab60e  2015-12-20 08:26:17 +0000
4  f4d8d43858c8858c68d75ce07ac641c0  2015-12-20 08:26:17 +0000

                                         product_url  \
0  http://www.flipkart.com/v-v-art-brass-bracelet...
1  http://www.flipkart.com/kalpaveda-copper-cuff/...
2  http://www.flipkart.com/thelostpuppy-book-cove...
3  http://www.flipkart.com/riana-copper-bangle/p/...
4  http://www.flipkart.com/inox-jewelry-stainless...

                                        product_name             pid  retail_price  \
0                  V&V ART Brass Bracelet  BBAE6NYHCDTEZJTB         470.0
1              Kalpaveda Copper Copper Cuff  BBAE6FFKZJTY7SZZ        1200.0
2  Thelostpuppy Book Cover for Apple iPad Air  ACCEA4DZHGMSFXVH        2199.0
3                Riana Copper Copper Bangle  BBAEAXFQHHMF3EYZ        2499.0
4          Inox Jewelry Stainless Steel Cuff  BBAECH63WYDGGTE2        1629.0

   discounted_price                                              image  \
0             423.0  ["http://img6a.flixcart.com/image/bangle-brace...
1            1200.0  ["http://img8a.flixcart.com/image/bangle-brace...
2             599.0  ["http://img5a.flixcart.com/image/cases-covers...
3             649.0  ["http://img5a.flixcart.com/image/bangle-brace...
4            1222.0  ["http://img8a.flixcart.com/image/bangle-brace...

  is_FK_Advantage_product                                        description  \
0                   False  V&V ART Brass Bracelet - Buy V&V ART Brass Bra...
1                   False  Kalpaveda Copper Copper Cuff\n                ...
2                   False  Thelostpuppy Book Cover for Apple iPad Air (Mu...
3                   False  Riana Copper Copper Bangle - Buy Riana Copper ...
4                   False  Inox Jewelry Stainless Steel Cuff\n            ...

  product_rating      overall_rating          brand  \
0  No rating available  No rating available        V&V ART
1  No rating available  No rating available     Kalpaveda
2  No rating available  No rating available  Thelostpuppy
3                  5.0                  5         Riana
4  No rating available  No rating available  Inox Jewelry

                                product_specifications      predicted_category
0  {"product_specification"=>[{"key"=>"Brand", "v...                Jewellery
1  {"product_specification"=>[{"key"=>"Stretchabl...                Jewellery
2  {"product_specification"=>[{"key"=>"Brand", "v...  Mobiles & Accessories
3  {"product_specification"=>[{"key"=>"Collection...                Jewellery
4  {"product_specification"=>[{"key"=>"Stretchabl...                Jewellery
['label_encoder.pkl']
```

Out[6]:

In [ ]: