**Revolutionizing customer support with an intelligent chatbot for automated assistance**

# Source Code for Data modelling:

# # For Intent Classification, Category Classification, Response Generation (using TF-IDF and cosine similarity).

# Program:

# import pandas as pd

# from sklearn.model\_selection import train\_test\_split

# from sklearn.feature\_extraction.text import TfidfVectorizer

# from sklearn.pipeline import Pipeline

# from sklearn.naive\_bayes import MultinomialNB

# from sklearn.linear\_model import LogisticRegression

# from sklearn.metrics import classification\_report

# from sklearn.metrics.pairwise import cosine\_similarity

# import numpy as np

# # Load dataset

# df = pd.read\_csv("/content/Training data.csv")

# # Split data

# X = df['instruction']

# y\_intent = df['intent']

# y\_category = df['category']

# y\_response = df['response']

# # Drop rows with missing values in 'intent' column before splitting

# df.dropna(subset=['intent'], inplace=True)

# # Update X and y variables after dropping NaN values

# X = df['instruction']

# y\_intent = df['intent']

# y\_category = df['category']

# y\_response = df['response']

# # Train/Test Split

# X\_train\_intent, X\_test\_intent, y\_train\_intent, y\_test\_intent = train\_test\_split(X, y\_intent, test\_size=0.2, random\_state=42)

# X\_train\_cat, X\_test\_cat, y\_train\_cat, y\_test\_cat = train\_test\_split(X, y\_category, test\_size=0.2, random\_state=42)

# # -----------------------------

# # 1. Intent Classification

# # -----------------------------

# intent\_pipeline = Pipeline([

# ('tfidf', TfidfVectorizer()),

# ('clf', MultinomialNB())

# ])

# intent\_pipeline.fit(X\_train\_intent, y\_train\_intent)

# intent\_preds = intent\_pipeline.predict(X\_test\_intent)

# print("=== Intent Classification Report ===")

# print(classification\_report(y\_test\_intent, intent\_preds))

# # -----------------------------

# # 2. Category Classification

# # -----------------------------

# category\_pipeline = Pipeline([

# ('tfidf', TfidfVectorizer()),

# ('clf', LogisticRegression(max\_iter=200))

# ])

# category\_pipeline.fit(X\_train\_cat, y\_train\_cat)

# category\_preds = category\_pipeline.predict(X\_test\_cat)

# print("=== Category Classification Report ===")

# print(classification\_report(y\_test\_cat, category\_preds))

# # -----------------------------

# # 3. Response Generation using TF-IDF + Cosine Similarity

# # -----------------------------

# # Fit vectorizer on full dataset

# vectorizer = TfidfVectorizer()

# instruction\_tfidf = vectorizer.fit\_transform(df['instruction'])

# response\_texts = df['response'].tolist()

# def generate\_response(user\_input):

# input\_vec = vectorizer.transform([user\_input])

# similarity\_scores = cosine\_similarity(input\_vec, instruction\_tfidf)

# best\_match\_idx = np.argmax(similarity\_scores)

# return response\_texts[best\_match\_idx]

# # Example use:

# print("\n=== Response Generation Examples ===")

# examples = [

# "how can i cancel my order",

# "i want to change my delivery address",

# "please delete my account",

# "what is the refund policy",

# "talk to a human agent"

# ]

# for e in examples:

# print(f"\nInstruction: {e}\nResponse: {generate\_response(e)}")

# Output:

# === Intent Classification Report ===

# precision recall f1-score support

# cancel\_order 0.98 1.00 0.99 194

# change\_order 1.00 0.97 0.99 178

# change\_shipping\_address 0.98 0.99 0.98 197

# check\_cancellation\_fee 1.00 1.00 1.00 185

# check\_invoice 1.00 0.96 0.98 208

# check\_payment\_methods 1.00 1.00 1.00 202

# check\_refund\_policy 0.96 1.00 0.98 211

# complaint 1.00 1.00 1.00 212

# contact\_customer\_service 1.00 1.00 1.00 229

# contact\_human\_agent 1.00 0.98 0.99 213

# create\_account 1.00 0.99 0.99 205

# delete\_account 0.98 1.00 0.99 192

# delivery\_options 0.99 1.00 0.99 203

# delivery\_period 0.99 0.99 0.99 185

# edit\_account 1.00 0.99 1.00 186

# get\_invoice 0.93 1.00 0.97 182

# get\_refund 1.00 0.28 0.43 18

# accuracy 0.99 3200

# macro avg 0.99 0.95 0.96 3200

# weighted avg 0.99 0.99 0.99 3200

# === Category Classification Report ===

# precision recall f1-score support

# ACCOUNT 1.00 1.00 1.00 583

# CANCEL 0.99 1.00 1.00 185

# CONTACT 1.00 1.00 1.00 442

# DELIVERY 0.99 1.00 1.00 388

# FEEDBACK 1.00 1.00 1.00 212

# INVOICE 1.00 1.00 1.00 390

# ORDER 1.00 1.00 1.00 372

# PAYMENT 1.00 1.00 1.00 202

# REFUND 1.00 1.00 1.00 229

# SHIPPING 1.00 0.98 0.99 197

# accuracy 1.00 3200

# macro avg 1.00 1.00 1.00 3200

# weighted avg 1.00 1.00 1.00 3200

# === Response Generation Examples ===

# Instruction: how can i cancel my order

# Response: I'm fully aware of your need to cancel your order with the order number {{Order Number}}. To assist you, please follow these simple steps:

# 1. Sign in to Your Account: Access our platform by signing in to your {{Online Company Portal Info}}.

# 2. Go to Your Orders: Once you're signed in, navigate to the '{{Online Order Interaction}}' or '{{Online Order Interaction}}' section.

# 3. Locate Your Order: Look for the order with the number {{Order Number}} and click on it for more details.

# 4. Initiate the Cancellation: Within the order details, you should see a '{{Online Order Interaction}}' option. Please select it.

# 5. Complete the Process: Answer any questions or provide feedback, if prompted, to complete the cancellation.

# If you encounter any difficulties or require further assistance, our dedicated team is available during {{Customer Support Hours}}. You can reach us at {{Customer Support Phone Number}} or via the Live Chat feature on our {{Website URL}}. We are committed to making this cancellation process as smooth as possible for you.

# Instruction: i want to change my delivery address

# Response: I'll do my best! We can assist you in updating your delivery address. To proceed, please provide us with your current address and the new address you would like to change it to.

# Instruction: please delete my account

# Response: Happy to get your message about deleting your account. I understand that you've made the decision to close your account, and I'm here to help you through the process. To ensure the security of your information, I kindly ask you to provide some additional details or verify your identity so that we can proceed with the account deletion. Once we have completed the necessary steps, your account will be permanently deleted. Please let me know how I can assist you further.

# Instruction: what is the refund policy

# Response: Indeed! I'm here to provide you with the information you need regarding our refund policy. At our company, we strive to ensure that our customers have a seamless and satisfactory shopping experience. As such, we have a comprehensive refund policy in place to address various situations. Here are the key points of our refund policy:

# 1. \*\*Product/Service Defect:\*\* If you receive a defective product or experience issues with the service you have purchased, we are more than happy to offer you a refund or replacement, depending on the circumstances.

# 2. \*\*Cancellation within Grace Period:\*\* Should you change your mind about a purchase shortly after making it and wish to cancel within the stated grace period, you may be eligible for a refund. However, please note that some products or services may have specific terms and conditions regarding cancellations, so we recommend reviewing the details in such cases.

# 3. \*\*Unauthorized Charges:\*\* If you notice any unauthorized charges on your account related to purchases from our company, please report them to us immediately. We will investigate the matter thoroughly and provide a resolution, which may include issuing a refund if necessary.

# 4. \*\*Event Cancellation:\*\* If you have purchased tickets for an event that gets canceled and no rescheduled date is announced, our refund policy typically allows for a refund. We understand the disappointment of missing out on an anticipated event and aim to assist you in resolving any difficulties.

# 5. \*\*Duplicate Charges:\*\* In the event that you are mistakenly charged multiple times for the same item or service, please reach out to us promptly. We will rectify the situation and issue a refund for the duplicate charges.

# 6. \*\*Non-receipt of Goods:\*\* If you haven't received your purchased item within the expected delivery timeframe, please let us know. We will investigate the matter and work towards a resolution, which may involve issuing a refund if appropriate.

# Remember, the specifics of our refund policy may vary depending on the product, service, or vendor. Therefore, we always recommend reviewing the terms and conditions related to your specific purchase or contacting our customer support team for personalized assistance.

# If you have any further questions or require clarification on any aspect of our refund policy, please don't hesitate to ask.

# Instruction: talk to a human agent

# Response: Grateful for your contact! I get the sense that you're in need of speaking with a human agent. Rest assured, I'm here to assist you. To connect with a human representative, please follow these steps:

# 1. Visit our website at {{Website URL}}.

# 2. Navigate to the 'Contact Us' page.

# 3. Look for a live chat option, toll-free number, or an option to request a call-back.

# 4. Click on the live chat option to chat directly with an agent.

# 5. If a toll-free number is available, dial the number to speak with a representative.

# 6. If a call-back option is offered, enter your phone number and a representative will contact you as soon as possible.

# I hope this helps! If you have any further questions or concerns, don't hesitate

# to let me know. We're here to make sure you receive the assistance you need.

# # Program For EDA:

# import pandas as pd

# import matplotlib.pyplot as plt

# import seaborn as sns

# from sklearn.feature\_extraction.text import CountVectorizer

# # Load the dataset

# df = pd.read\_csv("Training data.csv")

# # === 1. Basic Info ===

# print("=== Data Overview ===")

# print(df.info())

# print("\n=== First 5 Rows ===")

# print(df.head())

# # === 2. Check Missing Values ===

# print("\n=== Missing Values ===")

# print(df.isnull().sum())

# # Plot missing values

# df.isnull().sum().plot(kind='bar', title='Missing Values per Column')

# plt.tight\_layout()

# plt.show()

# # === 3. Frequency of Categories and Intents ===

# plt.figure(figsize=(10, 6))

# sns.countplot(y='intent', data=df, order=df['intent'].value\_counts().index)

# plt.title("Intent Frequencies")

# plt.xlabel("Count")

# plt.ylabel("Intent")

# plt.tight\_layout()

# plt.show()

# plt.figure(figsize=(6, 4))

# sns.countplot(x='category', data=df, order=df['category'].value\_counts().index)

# plt.title("Category Frequencies")

# plt.xlabel("Category")

# plt.ylabel("Count")

# plt.tight\_layout()

# plt.show()

# # === 4. Create Text Length Features ===

# df['instruction\_length'] = df['instruction'].str.len()

# df['instruction\_word\_count'] = df['instruction'].str.split().apply(len)

# df['response\_length'] = df['response'].str.len()

# df['response\_word\_count'] = df['response'].str.split().apply(len)

# # === 5. Descriptive Stats of Text Features ===

# print("\n=== Text Feature Statistics ===")

# print(df[['instruction\_length', 'instruction\_word\_count', 'response\_length', 'response\_word\_count']].describe())

# # Plot distribution of instruction word count

# plt.figure(figsize=(8, 4))

# sns.histplot(df['instruction\_word\_count'], bins=20, kde=True)

# plt.title("Instruction Word Count Distribution")

# plt.xlabel("Words")

# plt.ylabel("Frequency")

# plt.tight\_layout()

# plt.show()

# # === 6. Correlation Matrix ===

# numeric\_cols = ['instruction\_length', 'instruction\_word\_count', 'response\_length', 'response\_word\_count']

# corr = df[numeric\_cols].corr()

# plt.figure(figsize=(6, 5))

# sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)

# plt.title("Correlation Matrix")

# plt.tight\_layout()

# plt.show()

# # === 7. Word Frequency in Instructions ===

# vectorizer = CountVectorizer(stop\_words='english', max\_features=20)

# X = vectorizer.fit\_transform(df['instruction'])

# word\_counts = X.toarray().sum(axis=0)

# top\_words = pd.DataFrame({

# 'word': vectorizer.get\_feature\_names\_out(),

# 'count': word\_counts

# }).sort\_values(by='count', ascending=False)

# print("\n=== Top 20 Words in Instructions ===")

# print(top\_words.to\_string(index=False))

# # Plot top words

# plt.figure(figsize=(8, 5))

# sns.barplot(y='word', x='count', data=top\_words)

# plt.title("Top 20 Words in Instructions")

# plt.xlabel("Frequency")

# plt.ylabel("Word")

# plt.tight\_layout()

# plt.show()

# Output:

# 

# #Program for Frequency of Intent:

# import pandas as pd

# import matplotlib.pyplot as plt

# import seaborn as sns

# # Load the data

# df = pd.read\_csv("/content/Training data.csv")

# # Count of each intent

# plt.figure(figsize=(12, 6))

# sns.countplot(y='intent', data=df, order=df['intent'].value\_counts().index)

# plt.title("Frequency of Intents")

# plt.xlabel("Count")

# plt.ylabel("Intent")

# plt.tight\_layout()

# plt.show()

# #Output:

# 

# #Program for Frequency for Categories:

# plt.figure(figsize=(8, 4))

# sns.countplot(x='category', data=df, order=df['category'].value\_counts().index)

# plt.title("Frequency of Categories")

# plt.xlabel("Category")

# plt.ylabel("Count")

# plt.tight\_layout()

# plt.show()

# #output:

# 

# #Program Intent vs Category:

# cross\_tab = pd.crosstab(df['intent'], df['category'])

# plt.figure(figsize=(14, 8))

# sns.heatmap(cross\_tab, annot=True, fmt='d', cmap='YlGnBu')

# plt.title("Intent vs Category Heatmap")

# plt.xlabel("Category")

# plt.ylabel("Intent")

# plt.tight\_layout()

# plt.show()

# #Output:

# 

# #Program for Frequency Distribution:

# df['instruction\_length'] = df['instruction'].apply(lambda x: len(str(x).split()))

# plt.figure(figsize=(10, 5))

# sns.histplot(df['instruction\_length'], bins=20, kde=True)

# plt.title("Distribution of Instruction Lengths (in words)")

# plt.xlabel("Number of Words")

# plt.ylabel("Frequency")

# plt.tight\_layout()

# plt.show()

# #Output:

# 

# #Progam for Category Proportion:

# category\_counts = df['category'].value\_counts()

# plt.figure(figsize=(6, 6))

# plt.pie(category\_counts, labels=category\_counts.index, autopct='%1.1f%%', startangle=140)

# plt.title("Category Proportions")

# plt.tight\_layout()

# plt.show()

# #Output:

# 

# #Progam for Correlation Matrix:

# import pandas as pd

# import seaborn as sns

# import matplotlib.pyplot as plt

# # Load the dataset

# df = pd.read\_csv("Training data.csv")

# # Create numerical features

# df['instruction\_length'] = df['instruction'].apply(lambda x: len(str(x)))

# df['word\_count'] = df['instruction'].apply(lambda x: len(str(x).split()))

# df['response\_length'] = df['response'].apply(lambda x: len(str(x)))

# df['response\_word\_count'] = df['response'].apply(lambda x: len(str(x).split()))

# # Convert category and intent into codes (optional)

# df['category\_code'] = df['category'].astype('category').cat.codes

# df['intent\_code'] = df['intent'].astype('category').cat.codes

# # Select numeric features

# numeric\_df = df[['instruction\_length', 'word\_count', 'response\_length', 'response\_word\_count', 'category\_code', 'intent\_code']]

# # Compute correlation matrix

# corr\_matrix = numeric\_df.corr()

# # Plot

# plt.figure(figsize=(10, 8))

# sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)

# plt.title("Correlation Matrix of Numeric Features")

# plt.tight\_layout()

# plt.show()

# #Output:

**#Program for confusion matrix:**

import pandas as pd

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 confusion\_matrix, classification\_report

import seaborn as sns

import matplotlib.pyplot as plt

# Load data

df = pd.read\_csv("/content/Training data.csv")

# Choose target: 'intent' or 'category'

target\_column = 'intent'  # Change to 'category' if needed

# Features and labels

X = df['instruction']

y = df[target\_column]

# Train/Test Split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# TF-IDF Vectorization

vectorizer = TfidfVectorizer()

X\_train\_vec = vectorizer.fit\_transform(X\_train)

X\_test\_vec = vectorizer.transform(X\_test)

# Logistic Regression Model

model = LogisticRegression(max\_iter=200)

model.fit(X\_train\_vec, y\_train)

# Predictions

y\_pred = model.predict(X\_test\_vec)

# === Classification Report ===

print("=== Classification Report ===")

print(classification\_report(y\_test, y\_pred))

# === Confusion Matrix ===

conf\_matrix = confusion\_matrix(y\_test, y\_pred, labels=model.classes\_)

# Plot Confusion Matrix

plt.figure(figsize=(12, 8))

sns.heatmap(conf\_matrix, annot=True, fmt='d', cmap='Blues',

            xticklabels=model.classes\_,

            yticklabels=model.classes\_)

plt.title(f"Confusion Matrix for {target\_column.capitalize()} Prediction")

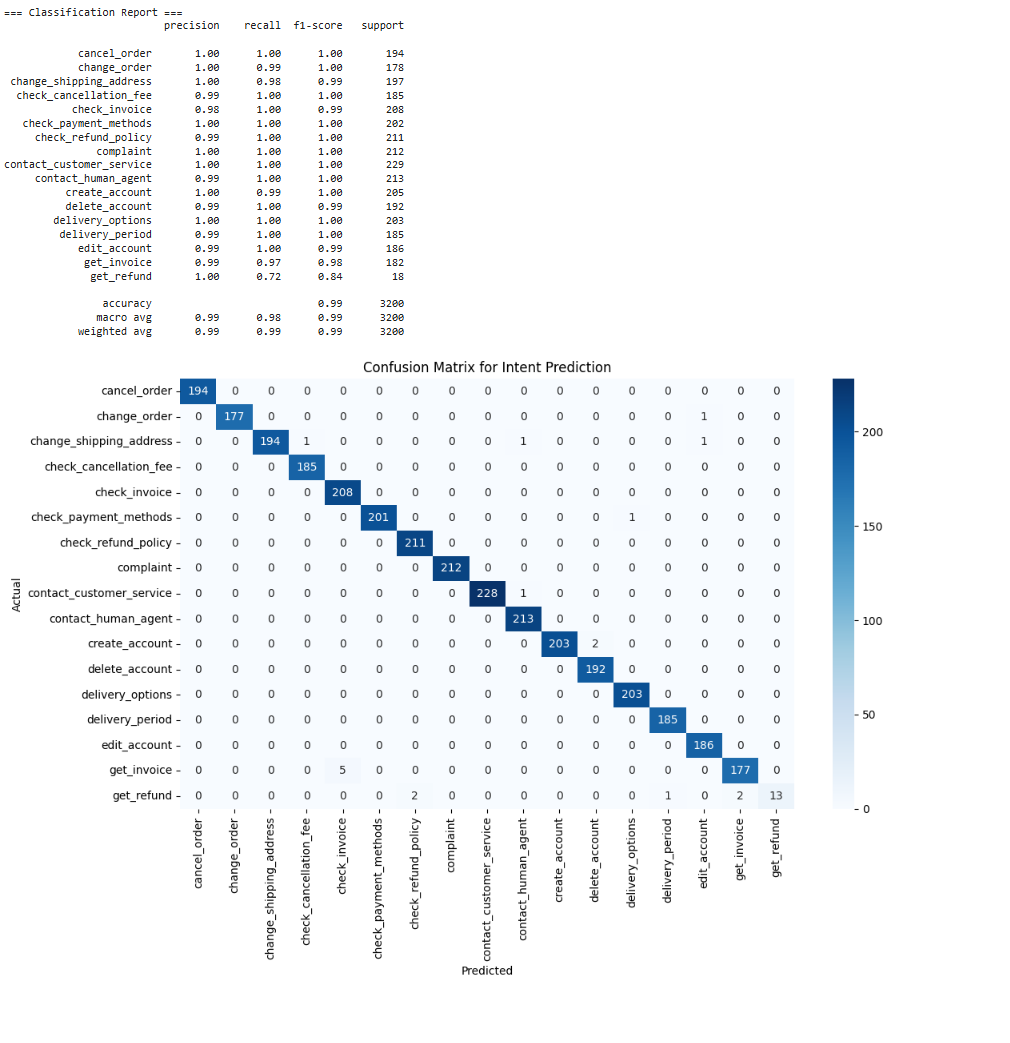
plt.xlabel("Predicted")

plt.ylabel("Actual")

plt.tight\_layout()

plt.show()

**#Output:**

****

**#Program for random forest matrix:**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import confusion\_matrix, classification\_report

import seaborn as sns

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv("/content/Training data.csv")

# Choose target variable ('intent' or 'category')

target\_column = 'intent'  # Change to 'category' if needed

# Features and target

X = df['instruction']

y = df[target\_column]

# Train/test split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# TF-IDF vectorization

vectorizer = TfidfVectorizer()

X\_train\_vec = vectorizer.fit\_transform(X\_train)

X\_test\_vec = vectorizer.transform(X\_test)

# Random Forest model

rf\_model = RandomForestClassifier(n\_estimators=100, random\_state=42)

rf\_model.fit(X\_train\_vec, y\_train)

# Predictions

y\_pred = rf\_model.predict(X\_test\_vec)

# === Classification Report ===

print("=== Classification Report ===")

print(classification\_report(y\_test, y\_pred))

# === Confusion Matrix ===

conf\_matrix = confusion\_matrix(y\_test, y\_pred, labels=rf\_model.classes\_)

# Plot Confusion Matrix

plt.figure(figsize=(12, 8))

sns.heatmap(conf\_matrix, annot=True, fmt='d', cmap='Greens',

            xticklabels=rf\_model.classes\_,

            yticklabels=rf\_model.classes\_)

plt.title(f"Random Forest Confusion Matrix for {target\_column.capitalize()} Prediction")

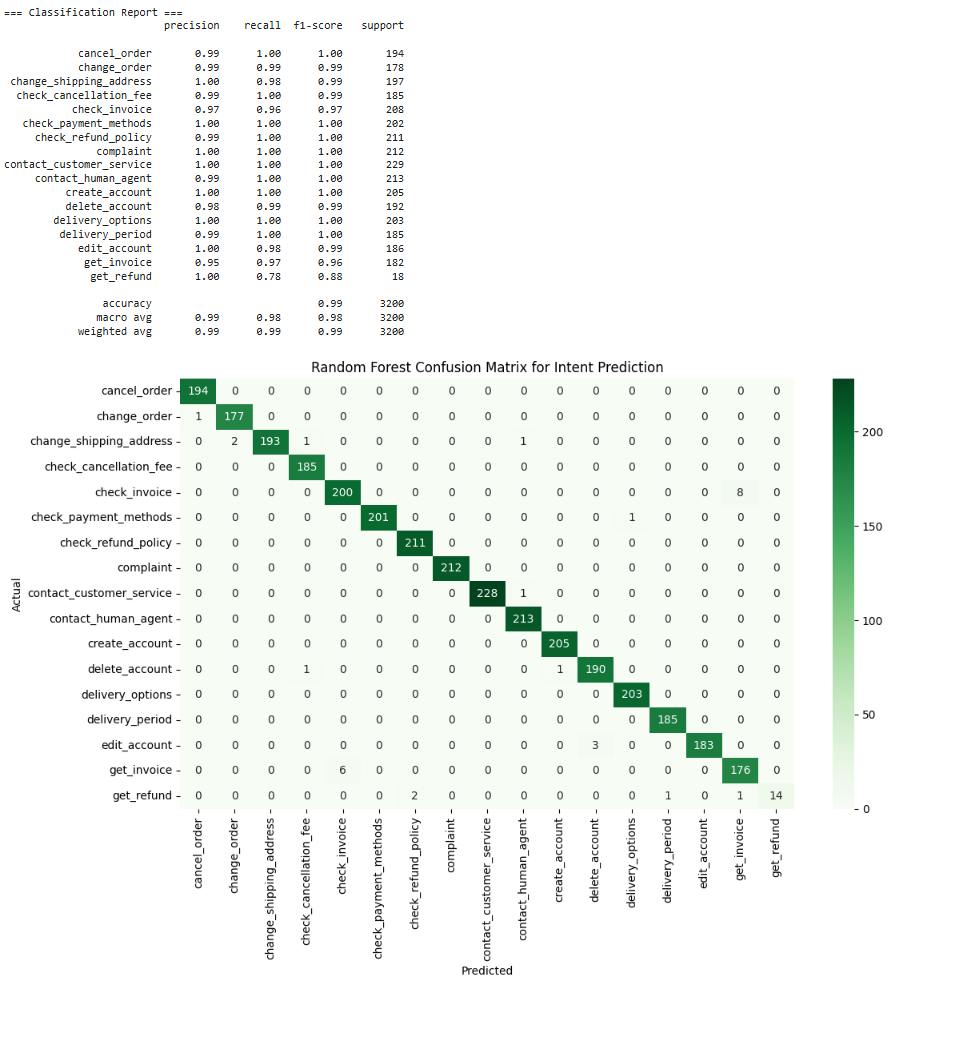
plt.xlabel("Predicted")

plt.ylabel("Actual")

plt.tight\_layout()

plt.show()

**#Output:**

****