

▼ Automatic Text Summarization

- Given that time is scarce, one of the main goals when dealing with large volumes of information is to increase the efficiency with which it can be processed.
- This paper will look at five of the major Automatic Text Summarization Techniques (ATS) methods used for processing linguistic textual data to determine which method yields the results for the DoQA dataset.
- An assumption will be made that given the nature of this dataset; the best result will be yielded through the use of a Query-Based approach.
- Methods will be compared numerically based on the ROGUE metric and F1 score amongst others.
- Furthermore, the authors present the Cooking dataset in three separate forms consisting of training, development and testing formats an attempt will be made to see if there is any statistical significance distinguishing these three possible sets of the same subject domain.

▼ Import Necessary Libraries

```
import warnings
warnings.filterwarnings("ignore")

pip install transformers

Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.35.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.13.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.10/dist-packages (from tra
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.23.
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (2
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.1
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transform
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.66.1
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggi
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from request
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transform
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->tra
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->tra
```

```
import transformers
```

```
import pandas as pd
import numpy as np
from numpy import hstack
import matplotlib.pyplot as plt
#from keras.utils.np_utils import to_categorical
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
#from keras.preprocessing.sequence import pad_sequences # from keras_preprocessing.sequence import pad_sequences
#from keras_preprocessing.sequence import pad_sequences
from keras.models import Sequential
from keras.layers import Dense, LSTM, GRU, Embedding
from keras import optimizers
from sklearn.metrics import r2_score, mean_absolute_error
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingRegressor, RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from fractions import Fraction
import seaborn as sns
import matplotlib.pyplot as plt
import json
import pandas as pd
from pathlib import Path
import re
import math
import random
import sys
import os# os.environ["CUDA_VISIBLE_DEVICES"] = "-1" # Force TF to use only the CPU
from keras.utils import to_categorical
from keras import layers
from keras import models
from matplotlib.ticker import StrMethodFormatter

import time
import datetime
print(sys.version)
import numpy as np
print(np.__version__)

3.10.12 (main, Nov 20 2023, 15:14:05) [GCC 11.4.0]
1.23.5
```

```
pip install rouge-score
```

```
Requirement already satisfied: rouge-score in /usr/local/lib/python3.10/dist-packages (0.1.2)
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.4.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (from rouge-score) (3.8.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.23.5)
Requirement already satisfied: six>=1.14.0 in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.16.0)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (2021.8.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (4.66.1)
```

```
!pip install --upgrade tensorflow
```

```
Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (2.15.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.1)
Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (23.5.26)
Requirement already satisfied: gast!=0.5.0,!0.5.1,!0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.5.3)
Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.1.2)
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.9.0)
Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (16.0.0)
```

```
Requirement already satisfied: ml-dtypes~==0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.
Requirement already satisfied: numpy<2.0.0,>=1.23.5 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow) (23.2)
Requirement already satisfied: protobuf!=4.21.0,!!=4.21.1,!!=4.21.2,!!=4.21.3,!!=4.21.4,!!=4.21.5,<5.0.0dev,>=3.20.3
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow) (67.7.2)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.
Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: tensorboard<2.16,>=2.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (fr
Requirement already satisfied: keras<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow)
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse)>=1
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorbo
Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in /usr/local/lib/python3.10/dist-packages (from ten
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-au
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-aut
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from request
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=
Requirement already satisfied: MarkupSafe>=2.1.1 in /usr/local/lib/python3.10/dist-packages (from werkzeug>=1.0.
Requirement already satisfied: pyasn1<0.6.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modu
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthli
```

```
import tensorflow
```

```
! pip install transformers
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.35.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.13.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.10/dist-packages (from tra
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.23.
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (2
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.1
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transform
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.66.1
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggi
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from request
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transform
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->tra
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->tra
```

Extract Data From Json Files

```
data = pd.read_json("/content/doqa-travel-test-v2.1.json")
data.head()
```

```

data 
0 {"title": "Japanese train etiquette: Is it acc... 
1 {"title": "UK online visa application: travell...
2 {"title": "Do layovers count in the 90-in-180 ...
3 {"title": "Do shops in Germany open on 2nd Jan...

# Function to extract data
def extract_data(data):
    cols = ["Context", "Question", "Answer", "OrigAnswer"]
    comp_list = []
    for entry in data["data"]:
        for paragraph in entry["paragraphs"]:
            for qa in paragraph["qas"]:
                # Extracting all answers for a question
                for answer in qa["answers"]:
                    temp_list = [
                        paragraph["context"],
                        qa["question"],
                        answer["text"],
                        qa["orig_answer"]["text"]
                    ]
                    comp_list.append(temp_list)
    # Creating a pandas DataFrame
    data_df = pd.DataFrame(comp_list, columns=cols)
    return data_df

```

▼ Calculate F1 Score

The F1 score ranges from 0 to 1, with a score of 1 indicating perfect precision and recall and 0 indicating poor performance. A high F1 score means that the model has high precision and high recall, which indicates that it is a good model for the binary classification task.

```
df1 = extract_data(data)
```

```
df1.head()
```

	Context	Question	Answer	OrigAnswer
0	Yes, it is acceptable. However, it is etiquett...	Japanese train etiquette: Is it acceptable to ...	Yes, it is acceptable. However, it is etiquett...	it is acceptable. However, it is etiquette tha...
1	Yes, it is acceptable. However, it is etiquett...	Will our family feel welcome with a baby in th...	CANNOTANSWER	CANNOTANSWER
2	Yes, it is acceptable. However, it is etiquett...	How accommodating will the bathrooms be for th...	CANNOTANSWER	CANNOTANSWER
3	Yes, it is acceptable. However, it is etiquett...	What sort of special perks come with sitting i...	CANNOTANSWER	CANNOTANSWER
4	In general, they are trying to find out if, fo...	What do I do if my companion does not have a v...	you can simply reword your text above (i.e., y...	you can then go to Part 9 of the form and ente...

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1713 entries, 0 to 1712
```

```
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          --    
 0   Context     1713 non-null    object 
 1   Question    1713 non-null    object 
 2   Answer      1713 non-null    object 
 3   OrigAnswer  1713 non-null    object 
dtypes: object(4)
memory usage: 53.7+ KB
```

▼ Calculation Of F1 Metric Through Precision And Recall

```
def calculate_f1_score(true_answer, predicted_answer):
    true_set = set(true_answer.lower().split())
    pred_set = set(predicted_answer.lower().split())
    # Calculate precision and recall
    precision = len(true_set.intersection(pred_set)) / len(pred_set) if len(pred_set) > 0 else 0
    recall = len(true_set.intersection(pred_set)) / len(true_set) if len(true_set) > 0 else 0
    # Calculate F1 score
    f1 = 2 * (precision * recall) / (precision + recall) if (precision + recall) > 0 else 0
    return f1

display(df1.Context.values.tolist()[0])

'Yes, it is acceptable. However, it is etiquette that if the baby does start to make noise that you take it to the "deck" area beyond the doors (where the bathrooms, telephone, vending machines, etc.) are. This is the same protocol if you have to make or receive a telephone call, want to have a loud conversation with your seatmate, or do anything that might disturb other passengers. I should note that this has nothing to do with the green cars -- the main cars nominally operate under these protocols, it's just that the other passengers (and conductors) are less likely to enforce them. CANNOTANSWER'

trueAnswer = df1.OrigAnswer.values.tolist()
PredictAnswer = df1.Answer.values.tolist()

f1_scores = []

for ta in range(len(trueAnswer)):
    true_answer = trueAnswer[ta].lower()
    predicted_answer = PredictAnswer[ta].lower()
    f1_score = calculate_f1_score(true_answer, predicted_answer)
    f1_scores.append(f1_score)

# Create a DataFrame from the results
F1Score_df = pd.DataFrame({"F1 Score": f1_scores})

F1Score_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1713 entries, 0 to 1712
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          --    
 0   F1 Score  1713 non-null    float64
dtypes: float64(1)
memory usage: 13.5 KB
```

▼ Calculation Of Exact Match

- Each Question Has One Correct Or "Gold" Answer
- Each Question Is Asked 3 Separate Times From The Same Context

- "Exact Match" Occurs When The Question Answer Is Identical To The Gold Answer
- Each Question Is Asked 3 Times And Is Either Correctly(Exact Matched) atleast once Thus It Will Have A Matching Average Of 1/3, 2/3, or 1 but not 0
- The Datasets Was Created Using two crowdworkers in Amazon Mechanical Turk(AMT) in accordance with the original source paper (Campos et al., 2020).

```
df2 = extract_data(data)
```

```
df2['Match'] = (df2['Answer'] == df2['OrigAnswer']).astype(float)
```

```
df2['EM3QAVG'] = df2.groupby('Question')['Match'].transform('mean')
```

```
df2.head(12)
```

	Context	Question	Answer	OrigAnswer	Match	EM3QAVG	grid icon	bar chart icon
0	Yes, it is acceptable. However, it is etiquette...	Japanese train etiquette: Is it acceptable to ...	Yes, it is acceptable. However, it is etiquette...	it is acceptable. However, it is etiquette tha...	0.0	0.0		
1	Yes, it is acceptable. However, it is etiquette...	Will our family feel welcome with a baby in th...	CANNOTANSWER	CANNOTANSWER	1.0	1.0		
2	Yes, it is acceptable. However, it is etiquette...	How accommodating will the bathrooms be for th...	CANNOTANSWER	CANNOTANSWER	1.0	1.0		
3	Yes, it is acceptable. However, it is etiquette...	What sort of special perks come with sitting i...	CANNOTANSWER	CANNOTANSWER	1.0	1.0		
4	In general, they are trying to find out if, fo...	What do I do if my companion does not have a v...	you can simply reword your text above (i.e., y...	you can then go to Part 9 of the form and ente...	0.0	0.0		
5	In general, they are trying to find out if, fo...	The application only gives me a yes or no opti...	You would answer the question honestly and che...	You would answer the question honestly and che...	0.0	0.0		
6	In general, they are trying to find out if, fo...	Will that delay my application?	CANNOTANSWER	CANNOTANSWER	1.0	1.0		
7	It's actually very simple, there are no comple...	Do layovers count in the 90-in-180 Schengen rule?	it counts as one day, whether it's a day trip,...	if you enter and leave immediately, or during ...	0.0	0.0		
8	It's actually very simple, there are no comple...	How do you calculate it?	if you enter and leave immediately, or during ...	if you enter 1 hour before midnight and leave ...	0.0	0.0		
9	It's actually very simple, there are no comple...	Okay. Where are you required to go through first?	you have to go through the border check and ge...	Each time you cross an external border, you sh...	0.0	0.0		
10	The 2nd of January is not a public	Do shops in Germany	The 2nd of January is not	most shops should be	0.0	0.5		

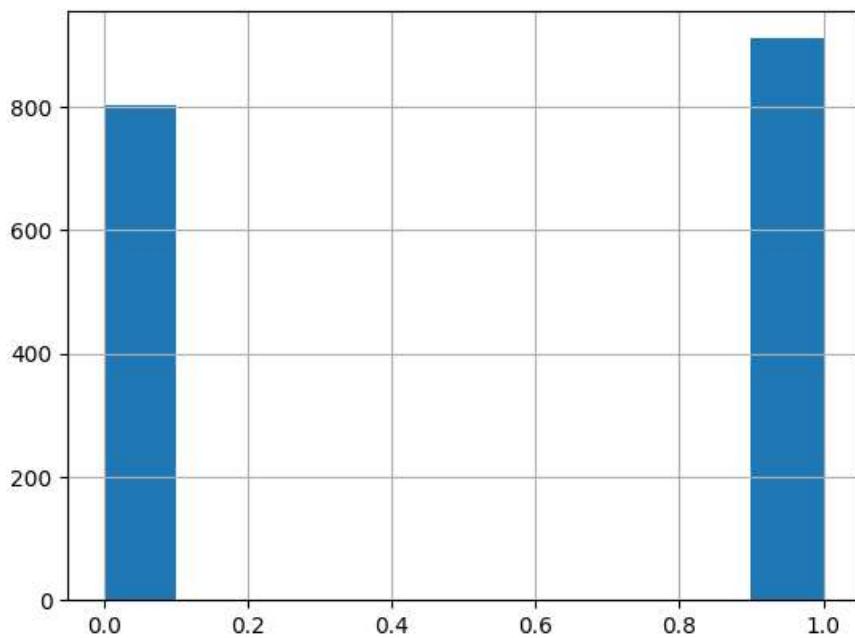
```
df2.describe()
```

	Match	EM3QAVG
count	1713.000000	1713.000000
mean	0.531232	0.531232
std	0.499169	0.496726
min	0.000000	0.000000
25%	0.000000	0.000000
50%	1.000000	1.000000
75%	1.000000	1.000000
max	1.000000	1.000000

▼ Match Visualization

```
df2["Match"].hist()
```

<Axes: >



```
df2.head()
```



	Context	Question	Answer	OrigAnswer	Match	EM3QAVG
0	Yes, it is acceptable. However, it is	Japanese train etiquette:	Yes, it is acceptable.	it is acceptable. However,	0.0	0.0

Exact Match Visualizatioin For Each Question

- Each Question Will Have An Exact Match Average Of 1/3, 2/3 or 1

res, it is acceptable. How accomodating will

```
#Take Every Third Row Of The
df3 = df2[df2.index % 3 == 0]
```

0	Yes, it is acceptable. How accomodating will	CANNOTANSWER	CANNOTANSWER	0.0	0.0
---	--	--------------	--------------	-----	-----

```
#Create A New DataFrame With Context, Question, and ExactMatch
```

```
#df4rows=df3/3
```

```
df4 = df3[['Context', 'Question', 'EM3QAVG']].copy()
```

0	fo...	have a v...	your local above (i.e., y...	or the form and circ...
---	-------	-------------	------------------------------	-------------------------

```
df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 571 entries, 0 to 1710
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Context     571 non-null    object  
 1   Question    571 non-null    object  
 2   EM3QAVG    571 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 17.8+ KB
```

```
df4=df4.reset_index(drop=True)
```

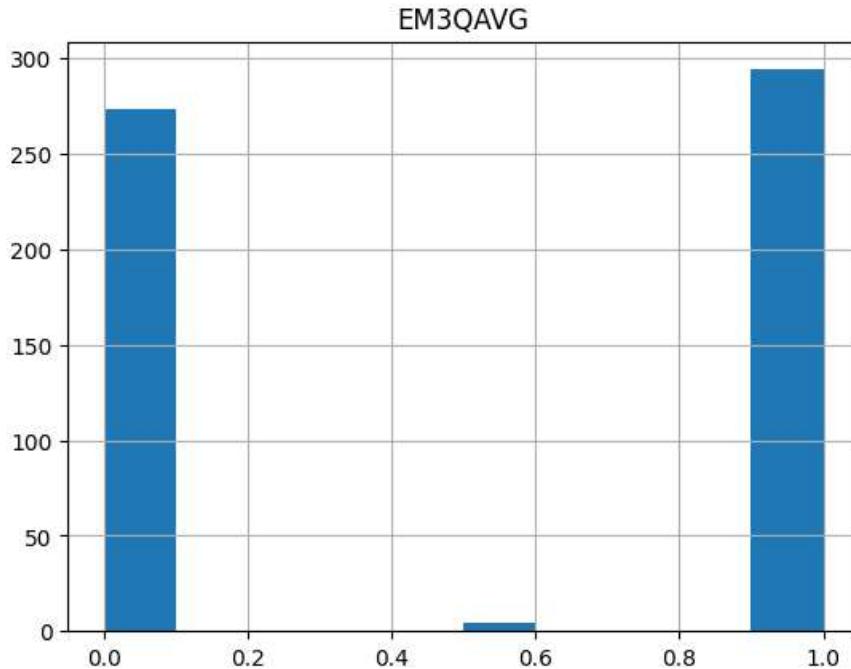
```
df4.head(10)
```

	Context	Question	EM3QAVG
0	Yes, it is acceptable. However, it is etiquette...	Japanese train etiquette: Is it acceptable to ...	0.0
1	Yes, it is acceptable. However, it is etiquette...	What sort of special perks come with sitting i...	1.0
2	In general, they are trying to find out if, fo...	Will that delay my application?	1.0
3	It's actually very simple, there are no comple...	Okay. Where are you required to go through first?	0.0
4	The 2nd of January is not a public holiday in ...	So you think the malls and big brand shops wil...	1.0
5	The actual process would depend on your locati...	After passing through immigration at the airpo...	0.0
6	The actual process would depend on your locati...	What is the process to cancel my travel and le...	1.0
7	He should call the courthouse for the relevant...	Should he stick around to file a police report?	1.0
8	He should call the courthouse for the relevant...	Does he have to stay in the area to see if the...	0.0
9	No, you can't use PackStation as a transient t...	Can you register for a "packstation" in German...	0.0

```
df4.describe()
```

```
EM3QAVG
count    571.000000
mean     0.518389
std      0.498343
min      0.000000
25%     0.000000
df4.hist()

array([[<Axes: title={'center': 'EM3QAVG'}>]], dtype=object)
```



❖ F1 Score Analysis

```
df5=df2
```

```
df5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1713 entries, 0 to 1712
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype  
 ---  --          -----          ----- 
 0   Context      1713 non-null   object  
 1   Question     1713 non-null   object  
 2   Answer       1713 non-null   object  
 3   OrigAnswer   1713 non-null   object  
 4   Match        1713 non-null   float64 
 5   EM3QAVG     1713 non-null   float64 
dtypes: float64(2), object(4)
memory usage: 80.4+ KB
```

```
df6 = F1Score_df
```

```
df5.reset_index(drop=True, inplace=True)
df6.reset_index(drop=True, inplace=True)
```

```
df7 = pd.concat([df5, df6], axis=1)
```

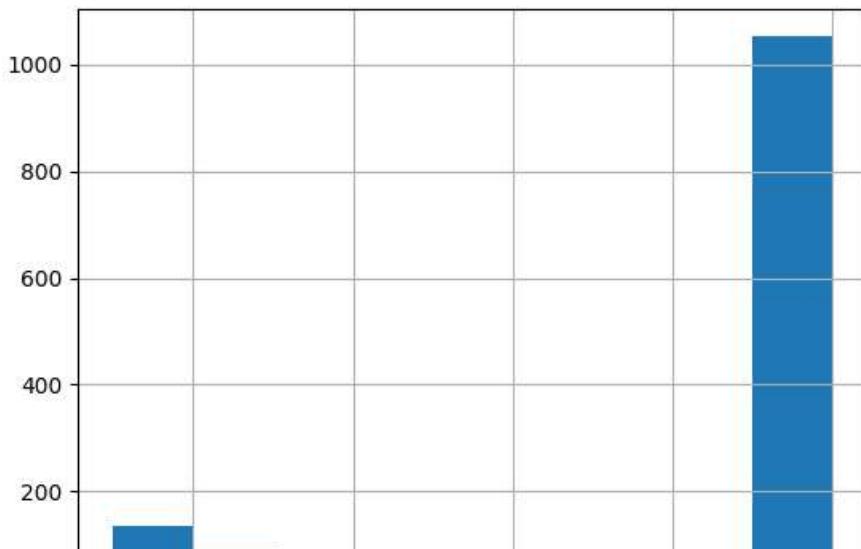
```
df7.head(12)
```

	Context	Question	Answer	OrigAnswer	Match	EM3QAVG	F1 Score
0	Yes, it is acceptable. However, it is etiquette...	Japanese train etiquette: Is it acceptable to ...	Yes, it is acceptable. However, it is etiquette...	it is acceptable. However, it is etiquette tha...	0.0	0.0	0.975610
1	Yes, it is acceptable. However, it is etiquette...	Will our family feel welcome with a baby in th...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
2	Yes, it is acceptable. However, it is etiquette...	How accommodating will the bathrooms be for th...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
3	Yes, it is acceptable. However, it is etiquette...	What sort of special perks come with sitting i...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
4	In general, they are trying to find out if, fo...	What do I do if my companion does not have a v...	you can simply reword your text above (i.e., y...	you can then go to Part 9 of the form and ente...	0.0	0.0	0.307692
5	In general, they are trying to find out if, fo...	The application only gives me a yes or no opti...	You would answer the question honestly and che...	You would answer the question honestly and che...	0.0	0.0	0.947368
6	In general, they are trying to find out if, fo...	Will that delay my application?	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
7	It's actually very simple, there are no comple...	Do layovers count in the 90-in-180 Schengen rule?	it counts as one day, whether it's a day trip,...	if you enter and leave immediately, or during ...	0.0	0.0	0.739130
8	It's actually very simple, there are no comple...	How do you calculate it?	if you enter and leave immediately, or during ...	if you enter 1 hour before midnight and leave ...	0.0	0.0	0.292683
9	It's actually very simple, there are no comple...	Okay. Where are you required to go through first?	you have to go through the border check and ge...	Each time you cross an external border, you sh...	0.0	0.0	0.222222
10	The 2nd of January is not a public holiday in ...	Do shops in Germany open on 2nd Jan?	The 2nd of January is not a public holiday in ...	most shops should be open with their Saturday ...	0.0	0.5	0.529412
11	The 2nd of January is not a public holiday in	We will be in Stuttgart from the 1st to the 3r...	The 2nd of January is not a public holiday in	The 2nd of January is not a public holiday in	1.0	1.0	1.000000

```
#F1Score_df.hist(bins=10)
```

```
df7["F1 Score"].hist(bins=[0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00])
```

<Axes: >



df7["F1 Score"].describe()

```

count    1713.000000
mean     0.739018
std      0.364878
min     0.000000
25%     0.457143
50%     1.000000
75%     1.000000
max     1.000000
Name: F1 Score, dtype: float64

```

✓ Use The Huggingface Transforms QA Model For Generation

From The Wikipedia Pace Of The AI Firm HuggingFace With Respect To Their Transformers Libraries The Transformers library is a Python package that contains open-source implementations of transformer models for text, image, and audio tasks. It is compatible with the PyTorch, TensorFlow and JAX deep learning libraries and includes implementations of notable models like BERT and GPT-2.[12] The library was originally called "pytorch-pretrained-bert"[13] which was then renamed to "pytorch-transformers" and finally "transformers."

Part 1 - Import The DataFrame

DF7 = df4

DF7.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 571 entries, 0 to 570
Data columns (total 3 columns):
 #   Column   Non-Null Count Dtype  
--- 
 0   Context   571 non-null   object  
 1   Question  571 non-null   object  
 2   EM3QAVG   571 non-null   float64 
dtypes: float64(1), object(2)
memory usage: 13.5+ KB

```

Step 2 - Take A Random Sample Of 50 Of The Origan Question-Answer Pairs

```
DF8 = DF7.sample(50)

! pip install pipeline

Requirement already satisfied: pipeline in /usr/local/lib/python3.10/dist-packages (0.1.0)

from transformers import AutoModelForQuestionAnswering, AutoTokenizer, pipeline
model_name = "deepset/roberta-base-squad2"
# a) Get predictions
nlp = pipeline('question-answering', model=model_name, tokenizer=model_name)
qa_output = []
# Iterate through the DataFrame and print the output
for index, row in DF8.iterrows():
    context = row['Context']
    question = row['Question']
    output = nlp(context=context, question=question)
    qa_output.append(output)
# Create a DataFrame from the results
Output_df = pd.DataFrame({"QA_Output": qa_output})
Output_df.head(10)

Output_df
```

QA_Output

```
0  {'score': 0.006162862293422222, 'start': 247, ...  
1  {'score': 0.32942333817481995, 'start': 633, '...  
2  {'score': 0.00223145610652864, 'start': 143, '...  
3  {'score': 0.00022367003839462996, 'start': 383...  
4  {'score': 0.007422285154461861, 'start': 0, 'e...  
5  {'score': 0.0003609021659940481, 'start': 578,...  
6  {'score': 0.01796315610408783, 'start': 351, '...  
7  {'score': 0.017307279631495476, 'start': 376, ...  
8  {'score': 0.014497920870780945, 'start': 547, ...  
9  {'score': 0.0007491487194783986, 'start': 414,...  
10  {'score': 0.2221788614988327, 'start': 743, 'e...  
11  {'score': 0.0033249398693442345, 'start': 80, ...  
12  {'score': 0.07213296741247177, 'start': 652, '...  
13  {'score': 0.011496561579406261, 'start': 124, ...  
14  {'score': 0.00034591712756082416, 'start': 0, ...  
15  {'score': 0.00779048353433609, 'start': 582, '...  
16  {'score': 0.161871999502182, 'start': 228, 'en...  
17  {'score': 0.009122165851294994, 'start': 866, ...  
18  {'score': 0.21475781500339508, 'start': 223, '...  
19  {'score': 0.11749587953090668, 'start': 457, '...  
20  {'score': 0.0003846264153253287, 'start': 402,...  
21  {'score': 0.008594973012804985, 'start': 76, '...  
22  {'score': 0.21398583054542542, 'start': 117, '...  
23  {'score': 0.017828304320573807, 'start': 314, ...  
24  {'score': 0.0005116212996654212, 'start': 119,...  
25  {'score': 0.04666643217206001, 'start': 515, '...  
26  {'score': 0.033046476542949677, 'start': 310, ...  
27  {'score': 4.6136970922816545e-05, 'start': 44,...
```

❖ Compute The ROUGE Metric

...

ROUGE (Recall-Oriented Understudy for Gisting Evaluation), is a set of metrics and a software package specifically designed for evaluating automatic summarization, but that can be also used for machine translation. The metrics compare an automatically produced summary or translation against reference (high-quality and human-produced) summaries or translations.

ROUGE-N ROUGE-N measures the number of matching n-grams between the model-generated text and a human-produced reference. Then, ROUGE-1 F1-score can be directly obtained from the ROUGE-1 precision and recall using the standard F1-score formula. ROUGE-1 F1-score = $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$

```
41      "score": 0.1755162225513458, "start": 350 }
dr = df8 = extract_data(data)
```

```
dr.head()
```

	Context	Question	Answer	OrigAnswer
0	Yes, it is acceptable. However, it is etiquette...	Japanese train etiquette: Is it acceptable to ...	Yes, it is acceptable. However, it is etiquette...	it is acceptable. However, it is etiquette tha...
1	Yes, it is acceptable. However, it is etiquette...	Will our family feel welcome with a baby in th...	CANNOTANSWER	CANNOTANSWER
2	Yes, it is acceptable. However, it is etiquette...	How accommodating will the bathrooms be for th...	CANNOTANSWER	CANNOTANSWER
3	Yes, it is acceptable. However, it is etiquette...	What sort of special perks come with sitting i...	CANNOTANSWER	CANNOTANSWER
4	In general, they are trying to find out if, fo...	What do I do if my companion does not have a v...	you can simply reword your text above (i.e., y...	you can then go to Part 9 of the form and ente...

```
! pip install rouge
```

```
Requirement already satisfied: rouge in /usr/local/lib/python3.10/dist-packages (1.0.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from rouge) (1.16.0)
```

```
import rouge
```

```
! pip install rouge-score
```

```
Requirement already satisfied: rouge-score in /usr/local/lib/python3.10/dist-packages (0.1.2)
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.4.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (from rouge-score) (3.8.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.23.5)
Requirement already satisfied: six>=1.14.0 in /usr/local/lib/python3.10/dist-packages (from rouge-score) (1.16.0)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (4.66.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk->rouge-score) (4.66.1)
```

```
from rouge_score import rouge_scorer
```

```
scorer = rouge_scorer.RougeScorer(['rouge1', 'rouge2', 'rougeL'], use_stemmer=True)
```

```
sc_output = []
```

```
for index, row in dr.iterrows():
    hyps = row['Answer']
    refs = row['OrigAnswer']

    # Calculate Rouge scores
    scores = scorer.score(hyps, refs)

    # Append scores to the list
    sc_output.append(scores)
```

```
# Create a DataFrame from the results
SCOutput_df = pd.DataFrame({"RougeScore_Output": sc_output})

SCOutput_df.head(10)
```

index	RougeScore_Output
0	{'rouge1': Score(precision=1.0, recall=0.9642857142857143, fmeasure=0.9818181818181818), 'rouge2': Score(precision=1.0, recall=0.9629629629629629, fmeasure=0.9811320754716981), 'rougeL': Score(precision=1.0, recall=0.9642857142857143, fmeasure=0.9818181818181818)}
1	{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
2	{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
3	{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
4	{'rouge1': Score(precision=0.25925925925925924, recall=0.4117647058823529, fmeasure=0.3181818181818182), 'rouge2': Score(precision=0.11538461538461539, recall=0.1875, fmeasure=0.14285714285714285), 'rougeL': Score(precision=0.18518518518518517, recall=0.29411764705882354, fmeasure=0.22727272727272727)}
5	{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
6	{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
7	{'rouge1': Score(precision=0.625, recall=1.0, fmeasure=0.7692307692307693), 'rouge2': Score(precision=0.6129032258064516, recall=1.0, fmeasure=0.76), 'rougeL': Score(precision=0.625, recall=1.0, fmeasure=0.7692307692307693)}
8	{'rouge1': Score(precision=0.28, recall=0.4117647058823529, fmeasure=0.3333333333333333), 'rouge2': Score(precision=0.125, recall=0.1875, fmeasure=0.15), 'rougeL': Score(precision=0.24, recall=0.35294117647058826, fmeasure=0.28571428571428564)}
9	{'rouge1': Score(precision=0.29411764705882354, recall=0.4166666666666667, fmeasure=0.3448275862068966), 'rouge2': Score(precision=0.125, recall=0.181818181818182, fmeasure=0.14814814814814), 'rougeL': Score(precision=0.29411764705882354, recall=0.4166666666666667, fmeasure=0.3448275862068966)}

Show 25 ▾ per page



Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

WARNING:root:Quickchart encountered unexpected dtypes in columns: "(['RougeScore_Output'],)"
No charts were generated by quickchart