

## ▼ 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-cooking-test-v2.1.json")
data.head()
```

```

data  
0 {"background": "We're going to try canning som...
1 {"background": "Other forms of poultry are reg...
2 {"background": "I've seen this question and ob...
3 {"background": "How do you prepare fresh fenne...

# 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()
```

	<b>Context</b>	<b>Question</b>	<b>Answer</b>	<b>OrigAnswer</b>
0	If you are going to be canning this for long t...	Canning chili with beans	CANNOTANSWER	CANNOTANSWER
1	If you are going to be canning this for long t...	What material would you advice using for canning?	If you are going to be canning this for long t...	If you are going to be canning this for long t...
2	If you are going to be canning this for long t...	Do I need a particular piece of machiner for t...	If you are going to be canning this for long t...	If you are going to be canning this for long t...
3	If you are going to be canning this for long t...	How much would I expect to pay for a pressure ...	CANNOTANSWER	CANNOTANSWER
4	If you are going to be canning this for long t...	What are the best vegetables to can?	CANNOTANSWER	CANNOTANSWER

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1797 entries, 0 to 1796
```

```
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   Context     1797 non-null    object  
 1   Question    1797 non-null    object  
 2   Answer      1797 non-null    object  
 3   OrigAnswer  1797 non-null    object  
 dtypes: object(4)
 memory usage: 56.3+ 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])

'If you are going to be canning this for long term storage you will need to be pressure canning it, and as that
will mean some pretty high heat you'd be better off not cooking you beans too long before canning. I'd add them
in a few minutes before the end of cooking just to get them warmed up in preparation for canning it as a half h
our of cooking plus pressure canning would probably mean mush. 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: 1797 entries, 0 to 1796
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   F1 Score   1797 non-null    float64
 dtypes: float64(1)
memory usage: 14.2 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	Actions
0	If you are going to be canning this for long t...	Canning chili with beans	CANNOTANSWER	CANNOTANSWER	1.0	1.0	
1	If you are going to be canning this for long t...	What material would you advice using for canning?	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	
2	If you are going to be canning this for long t...	Do I need a particular piece of machiner for t...	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	
3	If you are going to be canning this for long t...	How much would I expect to pay for a pressure ...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	
4	If you are going to be canning this for long t...	What are the best vegetables to can?	CANNOTANSWER	CANNOTANSWER	1.0	1.0	
5	Rare duck meat is safe to eat because it does ...	Is rare duck breast safe?	Rare duck meat is safe to eat because it does ...	Rare duck meat is safe to eat because it does ...	1.0	1.0	
6	Rare duck meat is safe to eat because it does ...	Does it taste better when it's rare?	CANNOTANSWER	CANNOTANSWER	1.0	1.0	
7	Carbon steel is more malleable and less brittl...	Why would I prefer carbon steel (rust prone) k...	Carbon steel is more malleable	Carbon steel is more malleable and less brittl...	0.0	0.0	
8	Carbon steel is more malleable and less brittl...	Is there a way to prevent it from rusting so q...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	
9	Carbon steel is more malleable and less brittl...	Why do I prefer carbon steel over stainless st...	Carbon steel is more malleable and less brittl...	Carbon steel is more malleable and less brittl...	0.0	0.0	
10	Carbon steel is more malleable and less brittl...	What extra care does carbon steel require?	maintain an extremely sharp edge	This means that it is easier to hone on a knif...	0.0	0.0	
	Sometimes the outer	How do you prepare		Sometimes the outer			

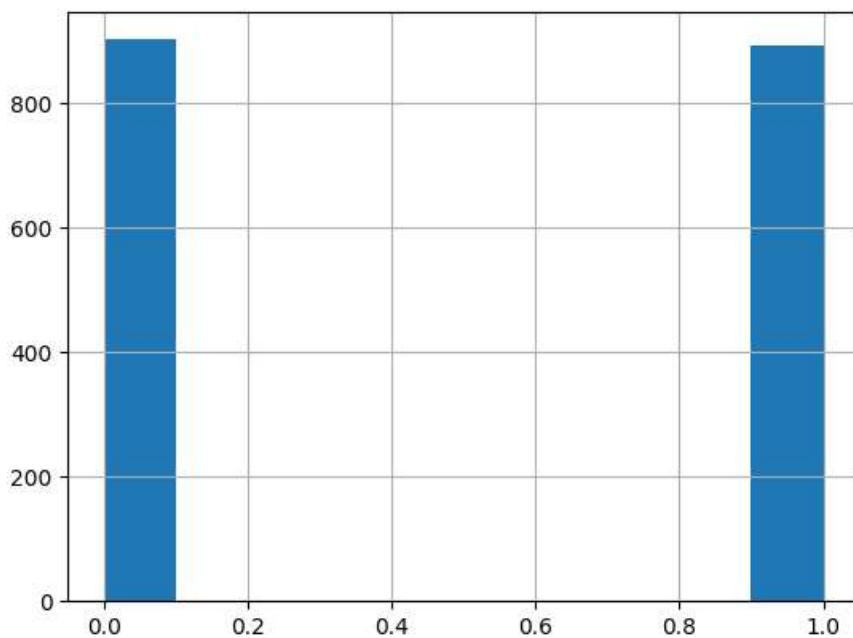
```
df2.describe()
```

	Match	EM3QAVG	grid icon
count	1797.000000	1797.000000	bar icon
mean	0.497496	0.497496	
std	0.500133	0.499297	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	

## ▼ Match Visualization

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

<Axes: >



```
df2.head()
```

	Context	Question	Answer	OrigAnswer	Match	EM3QAVG	grid icon
0	If you are going to be canning this for long t...	Canning chili with beans	CANNOTANSWER	CANNOTANSWER	1.0	1.0	bar icon
1	If you are going to be canning this for long t...	What material would you advice using for canning?	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	
2	If you are going to be canning this for long t...	Do I need a particular piece of machiner for t...	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	
3	If you are going to be canning this for long t...	How much would I expect to pay for a pressure ...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	

## ▼ Exact Match Visualizatioin For Each Question

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

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

```
#Create A New DataFrame With Context, Question, and ExactMatch
#df4rows=df3/3
df4 = df3[['Context', 'Question', 'EM3QAVG']].copy()
```

```
df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 599 entries, 0 to 1794
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Context     599 non-null    object  
 1   Question    599 non-null    object  
 2   EM3QAVG    599 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 18.7+ KB
```

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

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

	Context	Question	EM3QAVG	
0	If you are going to be canning this for long t...	Canning chili with beans	1.0	
1	If you are going to be canning this for long t... How much would I expect to pay for a pressure ...		1.0	
2	Rare duck meat is safe to eat because it does ...	Does it taste better when it's rare?	1.0	
3	Carbon steel is more malleable and less brittl...	Why do I prefer carbon steel over stainless st...	0.0	
4	Sometimes the outer leaves are a bit leathery ...	What does roughly chop mean?	1.0	
5	i usually use half of a previously-opened pist...	I have severe arthritis and joint pains, so I ...	1.0	
6	i usually use half of a previously-opened pist... How can I open them without the nut inside get...		1.0	
7	Your scale is measuring weight, not volume.Som...	The bag of coffee I am using instructs to "Add...	1.0	
8	If it has a shiny/reflective surface and doesn...	Does it affect the chocolate at all?	1.0	
9	If it has a shiny/reflective surface and doesn...	Can I untemper the chocolate?	1.0	

```
df4.describe()
```

EM3QAVG



df4.hist()

array([{'

EM3QAVG

EM3QAVG Bin	Count
[0.0, 1.0]	310
[1.0, 2.0]	290

## ❖ F1 Score Analysis

df5=df2

df5.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1797 entries, 0 to 1796
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          ----- 
 0   Context     1797 non-null   object  
 1   Question    1797 non-null   object  
 2   Answer      1797 non-null   object  
 3   OrigAnswer  1797 non-null   object  
 4   Match       1797 non-null   float64 
 5   EM3QAVG    1797 non-null   float64 
dtypes: float64(2), object(4)
memory usage: 84.4+ 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)

[https://colab.research.google.com/drive/1\\_g8C6OAPAiLmLGmlSbgRqe\\_SNDzElPQQ#scrollTo=lil4NQPr-TvU&printMode=true](https://colab.research.google.com/drive/1_g8C6OAPAiLmLGmlSbgRqe_SNDzElPQQ#scrollTo=lil4NQPr-TvU&printMode=true)

9/15



	Context	Question	Answer	OrigAnswer	Match	EM3QAVG	F1 Score
0	If you are going to be canning this for long t...	Canning chili with beans	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
1	If you are going to be canning this for long t...	What material would you advice using for canning?	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	1.000000
2	If you are going to be canning this for long t...	Do I need a particular piece of machiner for t...	If you are going to be canning this for long t...	If you are going to be canning this for long t...	1.0	1.0	1.000000
3	If you are going to be canning this for long t...	How much would I expect to pay for a pressure ...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
4	If you are going to be canning this for long t...	What are the best vegetables to can?	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
5	Rare duck meat is safe to eat because it does ...	Is rare duck breast safe?	Rare duck meat is safe to eat because it does ...	Rare duck meat is safe to eat because it does ...	1.0	1.0	1.000000
6	Rare duck meat is safe to eat because it does ...	Does it taste better when it's rare?	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
7	Carbon steel is more malleable and less brittl...	Why would I prefer carbon steel (rust prone) k...	Carbon steel is more malleable	Carbon steel is more malleable and less brittl...	0.0	0.0	0.312500
8	Carbon steel is more malleable and less brittl...	Is there a way to prevent it from rusting so q...	CANNOTANSWER	CANNOTANSWER	1.0	1.0	1.000000
9	Carbon steel is more malleable and less brittl...	Why do I prefer carbon steel over stainless st...	Carbon steel is more malleable and less brittl...	Carbon steel is more malleable and less brittl...	0.0	0.0	0.540541
10	Carbon steel is more malleable and less brittl...	What extra care does carbon steel require?	maintain an extremely sharp edge	This means that it is easier to hone on a knif...	0.0	0.0	0.454545
	Sometimes the	How do you		Sometimes the outer			

```
#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])
```



```
df7["F1 Score"].describe()
```

count	1797.000000
mean	0.723808
std	0.363851
min	0.000000
25%	0.363636
50%	0.971429
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()
```

RangeIndex:	599 entries, 0 to 598		
Data columns (total 3 columns):			
#	Column	Non-Null Count	Dtype
---	---	---	---
0	Context	599 non-null	object
1	Question	599 non-null	object
2	EM3QAVG	599 non-null	float64
	dtypes:	float64(1), object(2)	
	memory usage:	14.2+ 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
```

```
# 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': 0.01916523650288582, 'start': 155, 'end': 170}		
1	{'score': 8.062283995968755e-07, 'start': 87, 'end': 102}		
2	{'score': 0.029724912717938423, 'start': 198, 'end': 213}		
3	{'score': 0.33788934350013733, 'start': 108, 'end': 123}		
4	{'score': 0.013945959508419037, 'start': 400, 'end': 415}		
5	{'score': 0.00039197091246023774, 'start': 71, 'end': 86}		
6	{'score': 0.162758931517601, 'start': 80, 'end': 95}		
7	{'score': 0.23387962579727173, 'start': 367, 'end': 382}		
8	{'score': 0.04699481278657913, 'start': 164, 'end': 179}		
9	{'score': 0.28508952260017395, 'start': 125, 'end': 140}		
10	{'score': 0.044869545847177505, 'start': 17, 'end': 32}		
11	{'score': 0.02592082880437374, 'start': 358, 'end': 373}		
12	{'score': 0.06752008944749832, 'start': 116, 'end': 131}		
13	{'score': 0.004725998267531395, 'start': 42, 'end': 57}		
14	{'score': 0.02621706947684288, 'start': 384, 'end': 399}		
15	{'score': 0.04623829573392868, 'start': 140, 'end': 155}		
16	{'score': 0.05574658513069153, 'start': 135, 'end': 150}		
17	{'score': 4.862391733695404e-07, 'start': 366, 'end': 381}		
18	{'score': 0.00039197091246023774, 'start': 71, 'end': 86}		

## ✓ 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.  

$$\text{ROUGE-1 F1-score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

```
28     {'score': 0.11442896723747253, 'start': 437, 'end': 452}
dr = df8 = extract_data(data)

dr.head()
```



	Context	Question	Answer	OrigAnswer
0	If you are going to be canning this for long t...	Canning chili with beans	CANNOTANSWER	CANNOTANSWER
	! 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)			
44	{'sdpred4252019100530008t06mstarw121.expect to		CANNOTANSWER	CANNOTANSWER
	import rouge			
	If you are going to be    What are the best vegetables		CANNOTANSWER	CANNOTANSWER
	! pip install rouge-score			
	Requirement already satisfied: rouge-score in /usr/local/lib/python3.10/dist-packages (0.1.2)			
	Requirement already satisfied: abs1-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=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)}
1	{'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)}
2	{'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)}
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=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)}
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=0.1724137931034483, recall=1.0, fmeasure=0.29411764705882354), 'rouge2': Score(precision=0.14285714285714285, recall=1.0, fmeasure=0.25), 'rougeL': Score(precision=0.1724137931034483, recall=1.0, fmeasure=0.29411764705882354)}
8	{'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)}
9	{'rouge1': Score(precision=1.0, recall=0.3793103448275862, fmeasure=0.5499999999999999), 'rouge2': Score(precision=1.0, recall=0.3793103448275862, fmeasure=0.5499999999999999), 'rougeL': Score(precision=1.0, recall=0.3793103448275862, fmeasure=0.5499999999999999)}