



Training and Development of a Table-to-Text Transformer-Based Model for Contextual Summarization of Tabular Data

Department of Informatics

Despoina Angelonidi

MSc in Data Science Thesis

Supervisor : Dr. Panagiotis Louridas

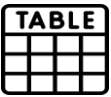
Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Training
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

Data, Data Everywhere



Data, like people, comes in all shapes and sizes. The most common data structure is tables.



A table with many rows and columns requires in some cases lots of time to go through.



Scope: Train and Develop a **Table-to-Text Transformer-based** model that generates targeted **Natural Language Summaries** from Tabular Data.



2 Open-source Datasets : ToTTo & QTSumm

Outline

1. Introduction
2. **Datasets**
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Training
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

ToTTo

Description & Task

Task: Given a Wikipedia Table and a set of Highlighted Cells, produce a One-Sentence Description.

The Dataset includes : Wikipedia Tables, Metadata, Highlighted Cells, Annotations.

Rank ↕	Home team ↕	Score ↕	Away team ↕	Attendance ↕	Date ↕	Stadium ↕
1	Atlético de Kolkata	3–0	Mumbai City	65,000	12 October 2014	Salt Lake Stadium
2	Kerala Blasters	0–1	Chennaiyin FC	61,323	30 November 2014	Jawaharlal Nehru Stadium
3	Kerala Blasters	3–0	Chennaiyin FC	60,900	13 December 2014	Jawaharlal Nehru Stadium
4	Kerala Blasters	2–1	Atlético de Kolkata	57,296	21 November 2014	Jawaharlal Nehru Stadium
5	Atlético de Kolkata	1–1	Delhi Dynamos	55,793	19 October 2014	Salt Lake Stadium
6	Atlético de Kolkata	0–0	Goa	53,173	14 December 2014	Salt Lake Stadium
7	Kerala Blasters	1–0	Goa	49,517	6 November 2014	Jawaharlal Nehru Stadium
8	Atlético de Kolkata	0–0	Chennaiyin FC	46,288	14 November 2014	Salt Lake Stadium
9	Kerala Blasters	1–0	Pune City	44,532	9 December 2014	Jawaharlal Nehru Stadium
10	Kerala Blasters	0–0	NorthEast United	43,299	4 December 2014	Jawaharlal Nehru Stadium

[Source](#)

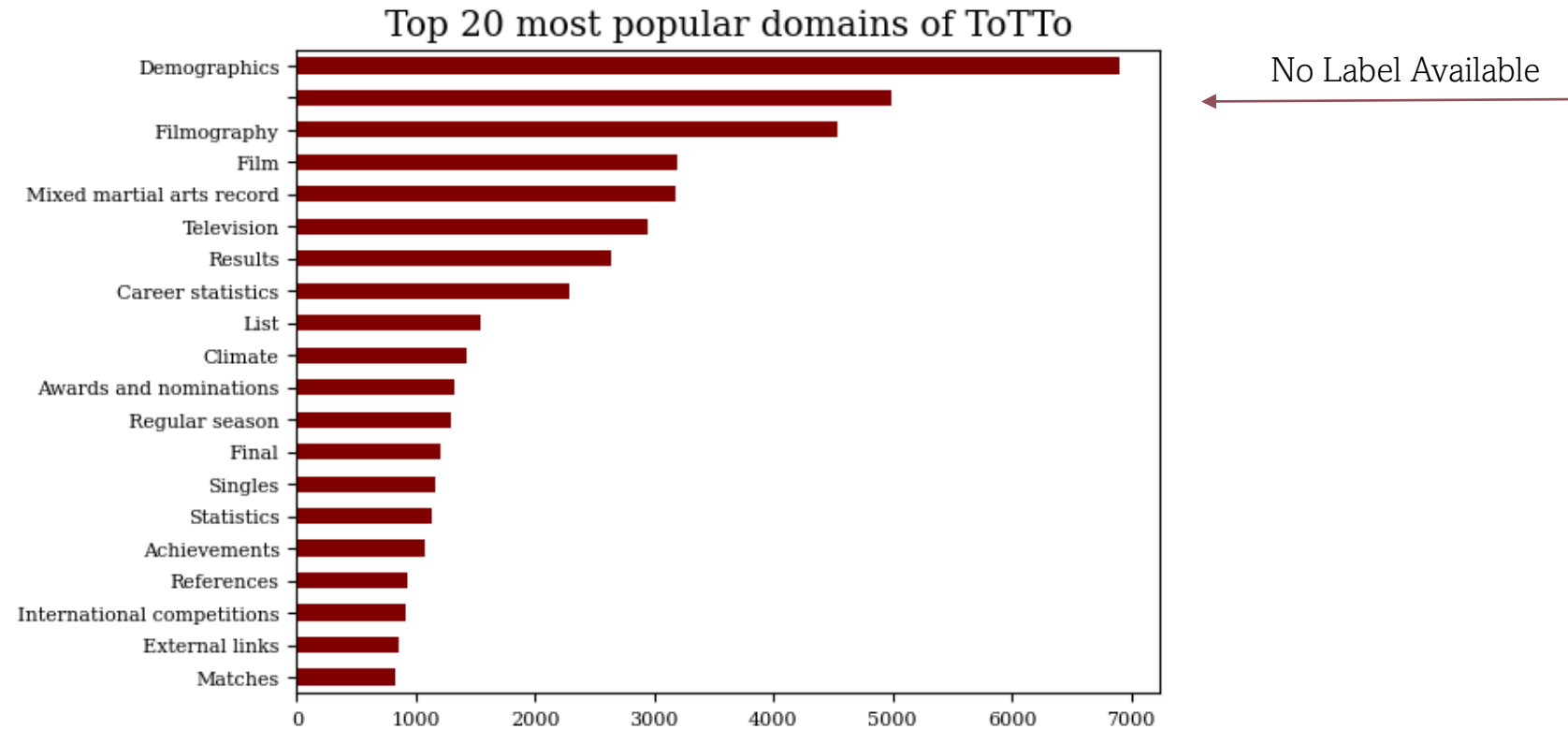


$[1,1], [1,2], [1,3], [1,5], [1,6]$



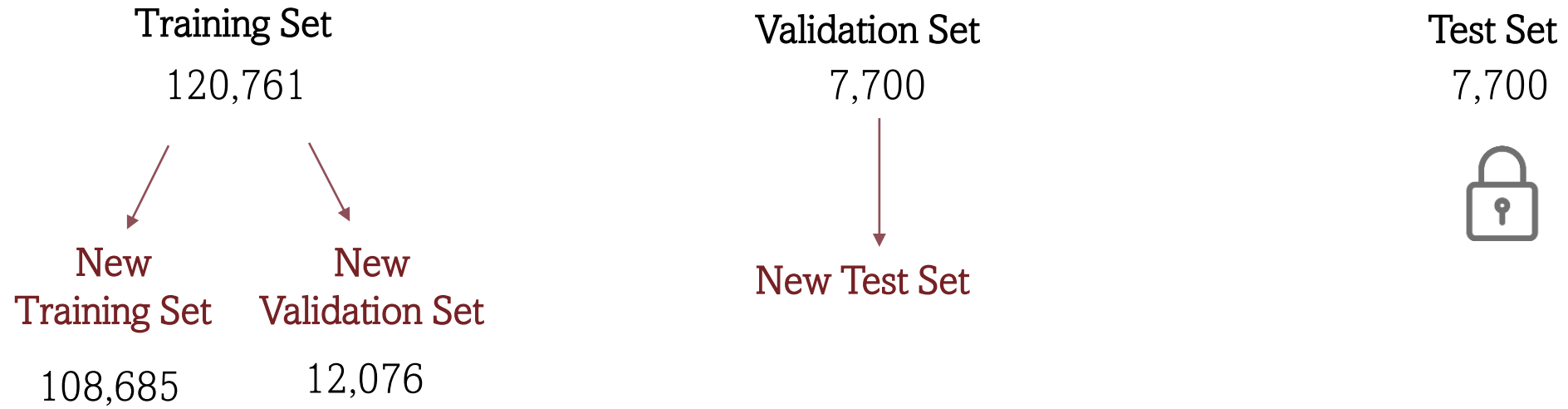
The season started on **12 October** when **Atlético de Kolkata** defeated **Mumbai City 3–0**, at the **Salt Lake Stadium**.

Most Popular Domains



Datasets

Dataset Split



QTSumm

QTSumm

Query	Annotation
Do Australian rules football matches receive higher attendance than cricket matches when analyzing the top 10 attendance records at the Bellerive Oval?	On the basis of top 10 records of attendance at Bellerive Oval, it seems cricket matches have more people attending than Australian rules football game. In top 10 match, 7 of them cricket matches, only 3 of them Australian rules football game. Plus, highest record of attendance for cricket match have 18,149 person, but highest for Australian rules football is 17,844 person.
What is the relationship between the percentage of seats won and the party's position as an opposition or the majority government in the given years?	The percent of seats win by the National Democratic Party (British Virgin Islands) direct influence party's position as opposition or majority government in the give years. When party win higher percent of seats (52.4% in 2003, 52.5% in 2011, and 60.2% in 2015), they have majority government. But, when their percent of seats lower (37.3% in 1999, 39.6% in 2007, and not know amount in 2019), they be opposition.
Compare the electricity production from wind power in Germany and Japan in 2011.	In 2011, Germany produced 45.3 terawatt-hours of electricity from wind power, and Japan produced 4.35 terawatt-hours of electricity from wind power.
Summarize the basic information of the wives of Amasa Lyman that had no children.	Amasa Lyman married one women who had no children. Diontha Walker, age 27 when she married Lyman in 1843, was the women with no children.
What was the range of attendances seen at events at The Pearl at The Palms venue in 2009?	the range of attendances seen at events at The Pearl at The Palms venue in 2009 was from 1,741 to 13,027.

- **Task:** Query-Focused Table Summarization.
- 7,111 Query-Summary pairs over 2,934 Wikipedia Tables.
- The Dataset consists of : the Tables, the Queries, the Annotations and the Highlighted Rows.

Training Set
4,981

Validation Set
1,052

Test Set
1,078

Datasets

Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Training
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

From Tables to Natural Language

Seq2Seq Transformers
accept **text** as input.



How?

“Header is Value”

Table Templates

Rank is 1, Home team is Atlético de Kolkata, Score is 30, Away team is Mumbai City, Attendance is 65,000, Date is 12 October 2014, Stadium is Salt Lake Stadium. Rank is 2, Home team is Kerala Blasters, Score is 01, Away team is Chennaiyin, Attendance is 61,323, Date is 30 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 3, Home team is Kerala Blasters, Score is 30, Away team is Chennaiyin, Attendance is 60,900, Date is 13 December 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 4, Home team is Kerala Blasters, Score is 21, Away team is Atlético de Kolkata, Attendance is 57,296, Date is 21 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 5, Home team is Atlético de Kolkata, Score is 11, Away team is Delhi Dynamos, Attendance is 55,793, Date is 19 October 2014, Stadium is Salt Lake Stadium. Rank is 6, Home team is Atlético de Kolkata, Score is 00, Away team is Goa, Attendance is 53,173, Date is 14 December 2014, Stadium is Salt Lake Stadium. Rank is 7, Home team is Kerala Blasters, Score is 10, Away team is Goa, Attendance is 49,517, Date is 6 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 8, Home team is Atlético de Kolkata, Score is 00, Away team is Chennaiyin, Attendance is 46,288, Date is 14 November 2014, Stadium is Salt Lake Stadium. Rank is 9, Home team is Kerala Blasters, Score is 10, Away team is Pune City, Attendance is 44,532, Date is 9 December 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 10, Home team is Kerala Blasters, Score is 00, Away team is NorthEast United, Attendance is 43,299, Date is 4 December 2014, Stadium is Jawaharlal Nehru Stadium.

Rank ♦	Home team ♦	Score ♦	Away team ♦	Attendance ♦	Date ♦	Stadium ♦
1	Atlético de Kolkata	3-0	Mumbai City	65,000	12 October 2014	Salt Lake Stadium
2	Kerala Blasters	0-1	Chennaiyin FC	61,323	30 November 2014	Jawaharlal Nehru Stadium

Preprocessing

ToTTo

ToTTo Templates

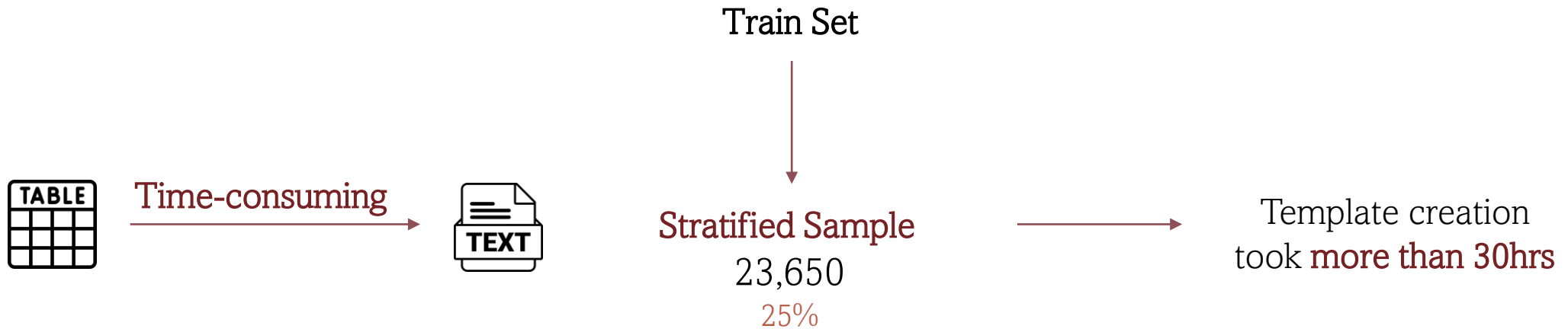
Metadata + Table converted to Text + Highlighted Cell Text + Table Structure

“Summarize:” +

Table page title:	2014 Indian Super League season,	Table section title	Highest attendances,
Table section text:	Source:,,	Table converted to text:	Rank is 1, Home team is Atlético de Kolkata, Score is 30, Away team is Mumbai City, Attendance is 65,000, Date is 12 October 2014, Stadium is Salt Lake Stadium. Rank is 2, Home team is Kerala Blasters, Score is 01, Away team is Chennaiyin, Attendance is 61,323, Date is 30 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 3, Home team is Kerala Blasters, Score is 30, Away team is Chennaiyin, Attendance is 60,900, Date is 13 December 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 4, Home team is Kerala Blasters, Score is 21, Away team is Atlético de Kolkata, Attendance is 57,296, Date is 21 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 5, Home team is Atlético de Kolkata, Score is 11, Away team is Delhi Dynamos, Attendance is 55,793, Date is 19 October 2014, Stadium is Salt Lake Stadium. Rank is 6, Home team is Atlético de Kolkata, Score is 00, Away team is Goa, Attendance is 53,173, Date is 14 December 2014, Stadium is Salt Lake Stadium. Rank is 7, Home team is Kerala Blasters, Score is 10, Away team is Goa, Attendance is 49,517, Date is 6 November 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 8, Home team is Atlético de Kolkata, Score is 00, Away team is Chennaiyin, Attendance is 46,288, Date is 14 November 2014, Stadium is Salt Lake Stadium. Rank is 9, Home team is Kerala Blasters, Score is 10, Away team is Pune City, Attendance is 44,532, Date is 9 December 2014, Stadium is Jawaharlal Nehru Stadium. Rank is 10, Home team is Kerala Blasters, Score is 00, Away team is NorthEast United, Attendance is 43,299, Date is 4 December 2014, Stadium is Jawaharlal Nehru Stadium,
	Focus on:	Atlético de Kolkata 30 Mumbai City 12 October 2014 Salt Lake Stadium,	Table Structure: 11 rows, 7 columns.

Preprocessing

ToTTo Templates



 Test Set : 1hr and 15 minutes

 Validation Set : 5hrs



- HTML tags
- Redundant punctuation marks
- Repeated consecutive words → 1 word
- # → "number"
- No Text Normalization was applied

Preprocessing

QTSumm

QTSumm Templates

Query + Table converted to Text + Highlighted Row Text + Table Structure

Without Text
Cleaning

Which competitors had their season's best (SB) performance in this championship and how did their results compare to each other? Table converted to text: Rank is 1, Name is Brianne Theisen-Eaton, Nationality is Canada, Time is 2:09.99, Points is 965, Notes is SB, Rank is 2, Name is Barbara Nwaba, Nationality is United States, Time is 2:10.07, Points is 964, Notes is SB, Rank is 3, Name is Györgyi Zsivoczky-Farkas, Nationality is Hungary, Time is 2:18.48, Points is 844, Notes is SB, Rank is 4, Name is Makeba Alcide, Nationality is Saint Lucia, Time is 2:18.65, Points is 842, Notes is SB, Rank is 5, Name is Salcia Slack, Nationality is Jamaica, Time is 2:19.00, Points is 837, Notes is , Rank is 6, Name is Kateřina Cachová, Nationality is Czech Republic, Time is 2:19.97, Points is 824, Notes is , Rank is 7, Name is Georgia Ellenwood, Nationality is Canada, Time is 2:20.18, Points is 821, Notes is , Rank is 8, Name is Morgan Lake, Nationality is Great Britain, Time is 2:20.40, Points is 818, Notes is , Rank is 9, Name is Alina Fyodorova, Nationality is Ukraine, Time is 2:20.42, Points is 818, Notes is , Rank is 10, Name is Kendell Williams, Nationality is United States, Time is 2:22.82, Points is 786, Notes is , Rank is 11, Name is Anastasiya Mokhnyuk, Nationality is Ukraine, Time is 2:23.19, Points is 781, Notes is , Rank is 12, Name is Celina Leffler, Nationality is Germany, Time is 2:24.01, Points is 770, Notes is Focus on: 1, Brianne Theisen-Eaton, Canada, 2:09.99, 965, SB, 2, Barbara Nwaba, United States, 2:10.07, 964, SB, 3, Györgyi Zsivoczky-Farkas, Hungary, 2:18.48, 844, SB, 4, Makeba Alcide, Saint Lucia, 2:18.65, 842, SB, Table Structure: Number of rows: 12, Number of columns: 6

Without Text
Normalization

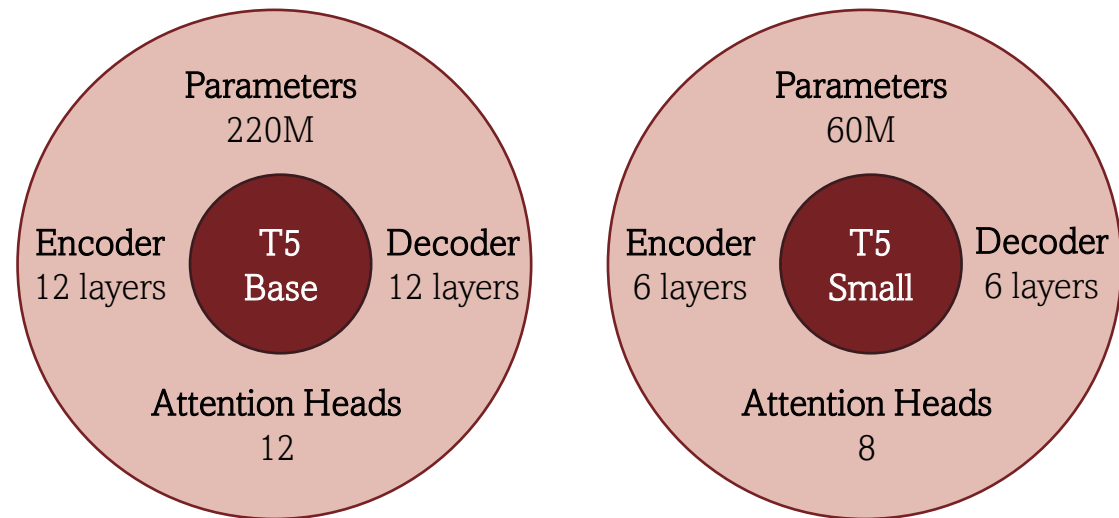
Preprocessing

Outline

1. Introduction
2. Datasets
3. Preprocessing
4. **Models**
5. Evaluation Metrics
6. Training
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

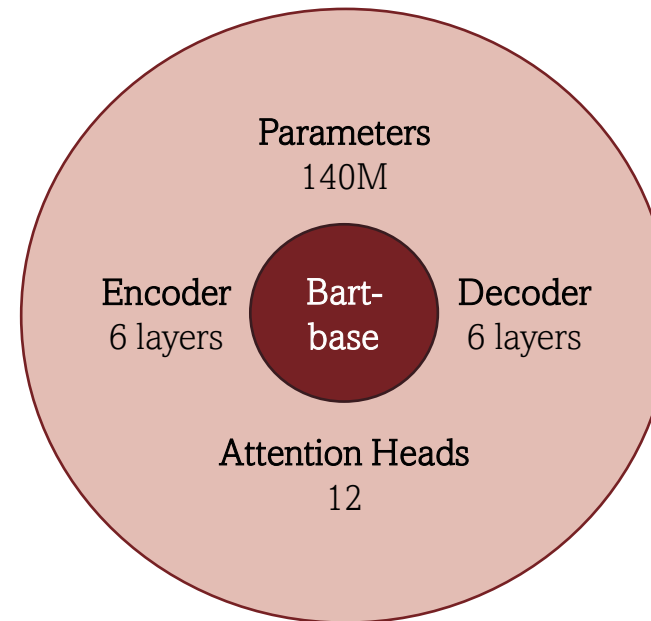
T5

- Google Research
- Seq2Seq Model
- Encoder – Decoder Architecture
- Treats various NLP tasks such as :
 - Translation
 - **Summarization**
 - Question-Answering
 - Other



BART

- Facebook
- Seq2Seq Model
- Encoder – Decoder Architecture
- Treats various NLP tasks such as :
 - Translation
 - **Summarization**
 - Text-Generation
 - Other



Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
- 5. Evaluation Metrics**
6. Training
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

ROUGE : Recall-Oriented Understudy for Gisting Evaluation

- It measures the **overlap** between a Candidate Summary and a Reference Summary.
- Ranges from 0 to 1. The higher, the better.

ROUGE 1

Unigrams

ROUGE 2

Bigrams


ROUGE L

Longest Common
Subsequence
(LCS)

ROUGE Lsum

LCS per sentence

SACREBLEU

- Based on **BLEU** (Bilingual Evaluation Understudy)
- **Precision-based** measure.
- Measures the similarity between machine-translated text and Reference  N-gram Overlap
- Brevity Penalty: penalizes the predictions that are too short when compared to the references.
- Ranges from 0 to 100. The higher, the better.

METEOR : Metric for Evaluation of Translation with Explicit Ordering

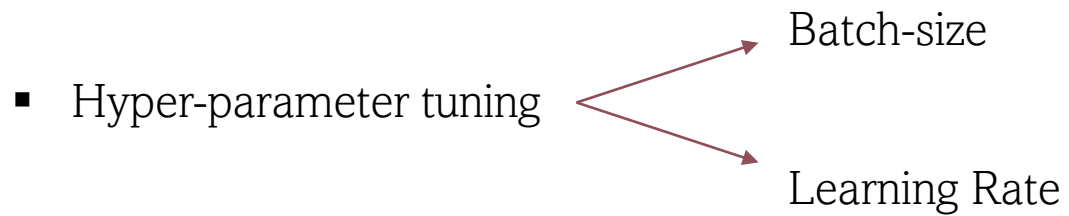
- METEOR creates alignments between the words in the Candidate Translation and the Reference Translation, focusing on exact, stem, synonym, and paraphrase matches.
- Considers both **Precision** and **Recall** while evaluating a match.
- Ranges from 0 to 1. The higher, the better.

Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
- 6. Training**
7. Results
8. Comparative Data Analysis
9. Conclusion and Future Work

ToTTo - QTSumm

- Fine-tuning **without** freezing any layers.



ToTTo

Trained for:

- 4 epochs : Bart-base, T5-small
- 1 epoch: T5-base

QTSumm

Trained for:

- 9 epochs : T5-small
- 6 epochs: Bart-base
- 2 epochs: T5-base

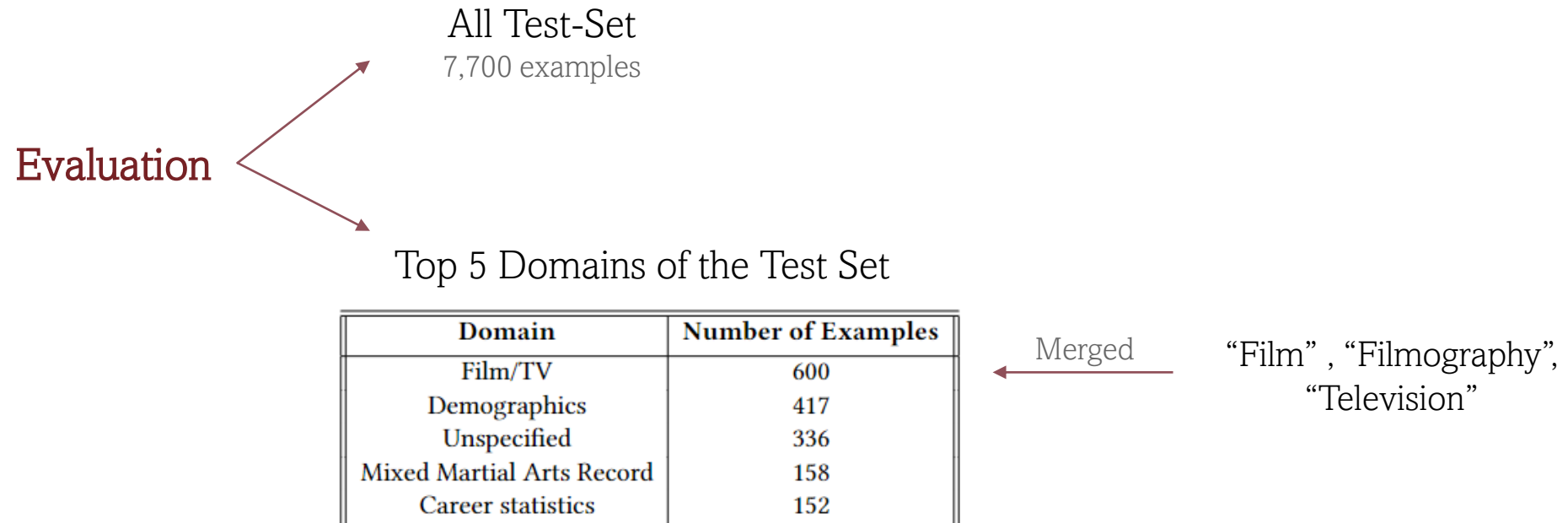
With **Early Stopping**

Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Training
- 7. Results**
8. Comparative Data Analysis
9. Conclusion and Future Work

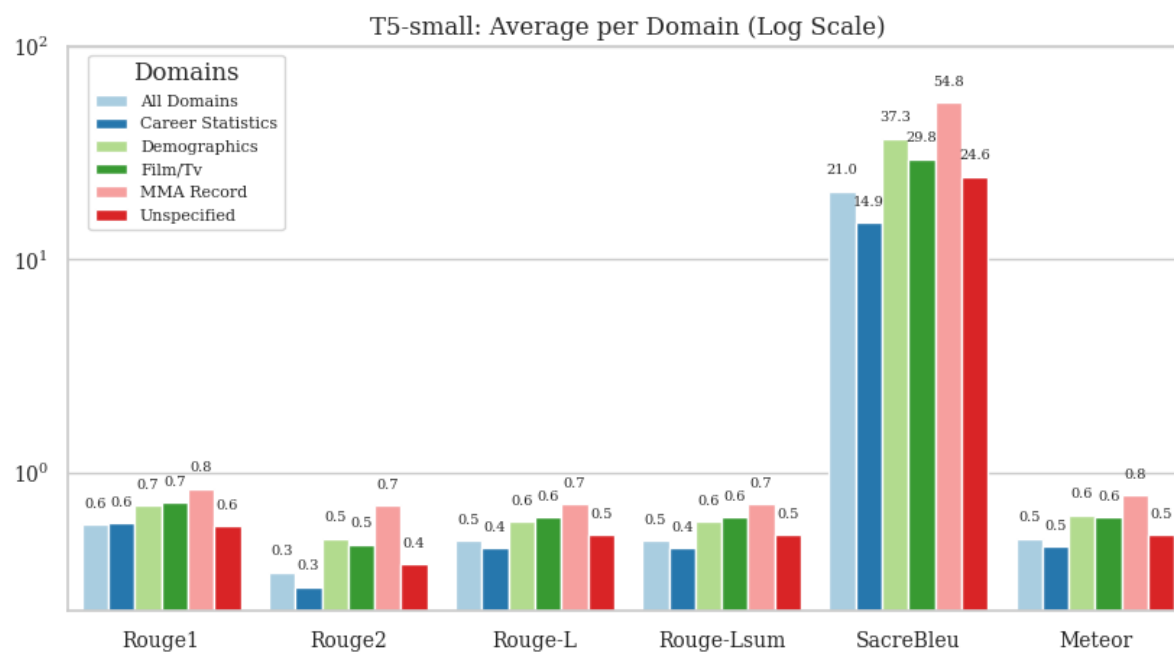
ToTTo

Evaluation: All – Top 5 Domains



Results

T5-small



Moderate to Good Performance

	Median					
Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.57	0.62	0.64	0.74	0.86	0.56
Rouge2	0.31	0.30	0.42	0.45	0.73	0.33
Rouge-L	0.45	0.44	0.48	0.61	0.70	0.48
Rouge-Lsum	0.45	0.44	0.48	0.61	0.70	0.48
SacreBleu	14.29	11.57	21.34	25.37	52.71	16.59
Meteor	0.48	0.46	0.61	0.62	0.80	0.51

	Standard Deviation					
Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.20	0.18	0.22	0.14	0.15	0.22
Rouge2	0.23	0.18	0.30	0.19	0.22	0.26
Rouge-L	0.21	0.15	0.28	0.17	0.21	0.23
Rouge-Lsum	0.21	0.15	0.28	0.17	0.21	0.23
SacreBleu	19.56	11.50	30.96	19.83	24.99	22.72
Meteor	0.22	0.19	0.25	0.19	0.18	0.24

Results

T5-small



Table	Generated Summary	Annotation
0	Daniel Henry Chamberlain served as the Governor of South Carolina from December 1, 1874.	Daniel Henry Chamberlin was the 76th Governor of South Carolina from 1874.
1	In 2016, Jodorowsky played Evelyn in Kids in Love.	Alma Jodorowsky had the role of Evelyn in the 2016 film Kids in Love.
2	In 2006, A. J. Hawk scored 119 touchdowns in 119 games.	In his rookie season, Hawk led with 119 total tackles.



76 Daniel Henry Chamberlain
December 1, 1874



2016 Kids in Love Evelyn

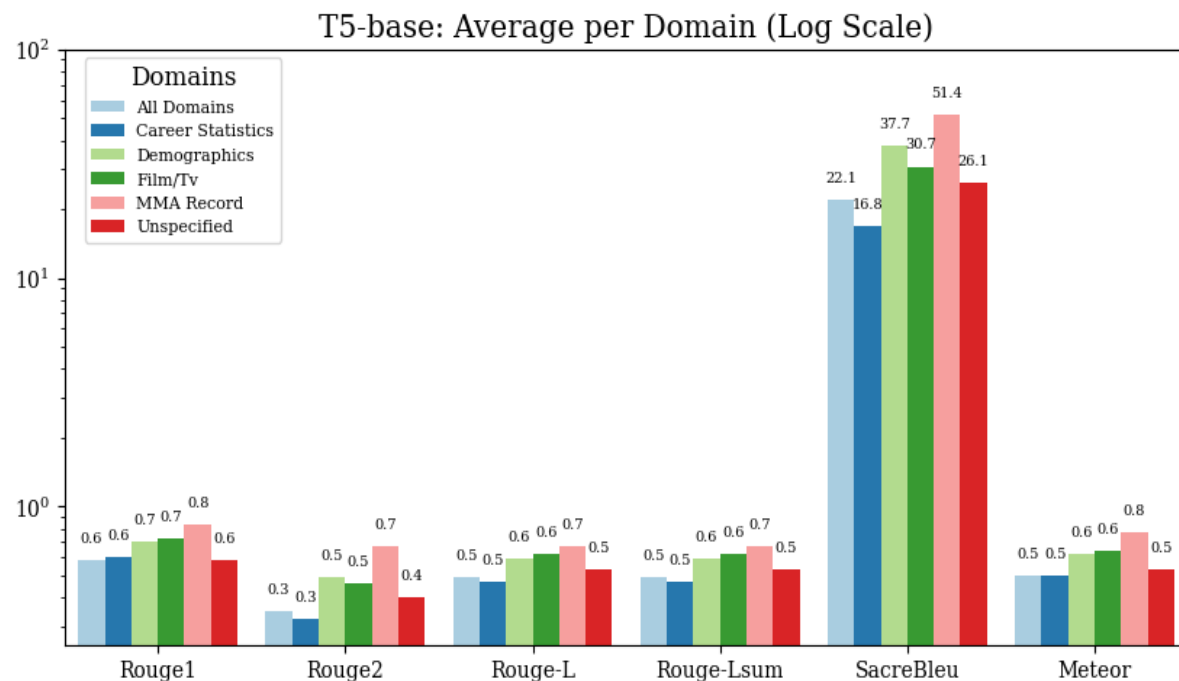


119

- Min Output Length: 5
- Max Output Length: 30

Results

T5-base



Moderate to Good Performance

	Median					
Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.59	0.63	0.64	0.74	0.87	0.57
Rouge2	0.32	0.31	0.42	0.45	0.73	0.36
Rouge-L	0.47	0.45	0.48	0.63	0.70	0.50
Rouge-Lsum	0.47	0.45	0.48	0.63	0.70	0.50
SacreBleu	15.01	12.82	21.99	26.60	53.75	18.14
Meteor	0.50	0.50	0.60	0.64	0.82	0.54

	Standard Deviation					
Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.21	0.17	0.22	0.13	0.17	0.23
Rouge2	0.23	0.19	0.30	0.19	0.22	0.26
Rouge-L	0.21	0.16	0.29	0.16	0.20	0.24
Rouge-Lsum	0.21	0.16	0.29	0.16	0.20	0.24
SacreBleu	20.19	13.46	31.28	19.43	24.13	22.83
Meteor	0.23	0.19	0.26	0.18	0.20	0.25

Results

T5-base



Table	Generated Summary	Annotation
0	Daniel Henry Chamberlain was the 76th Governor of South Carolina from December 1, 1874.	Daniel Henry Chamberlin was the 76th Governor of South Carolina from 1874.
1	In 2016, Alma Jodorowsky played the role of Evelyn in the film Kids in Love.	Alma Jodorowsky had the role of Evelyn in the 2016 film Kids in Love.
2	A. J. Hawk had 119 tackles.	In his rookie season, Hawk led with 119 total tackles.



76 Daniel Henry Chamberlain
December 1, 1874



2016 Kids in Love Evelyn

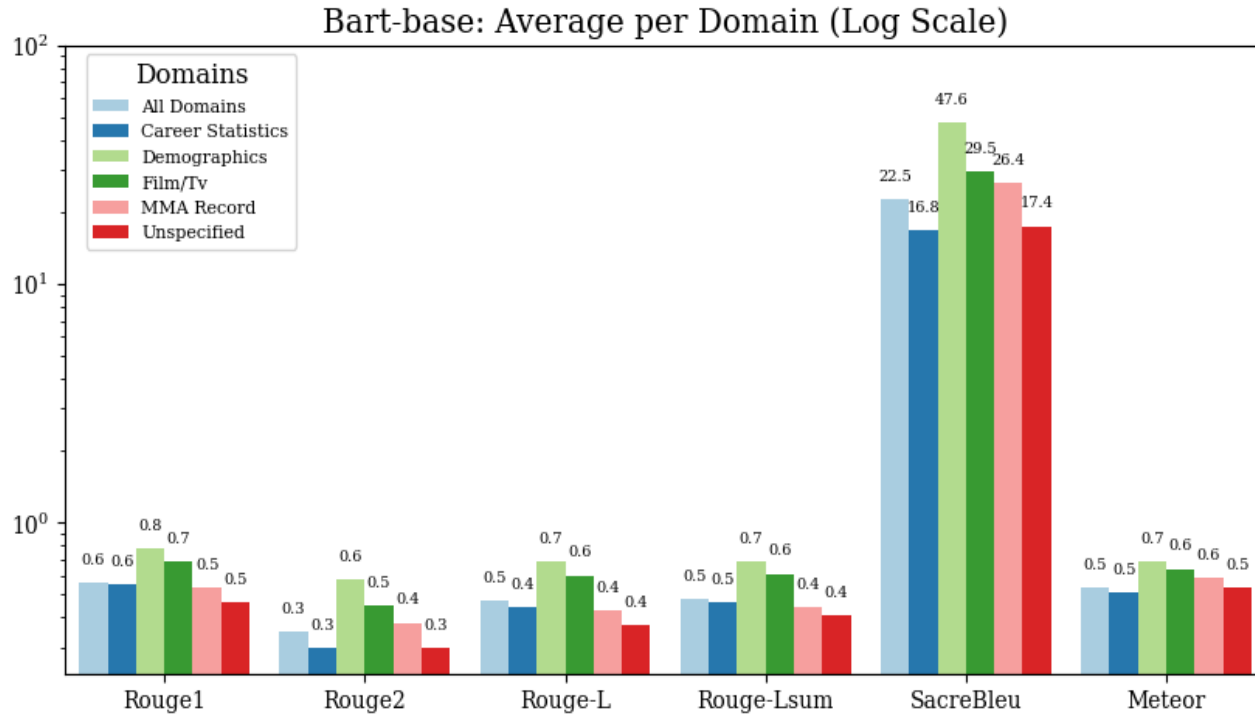


119

- Min Output Length: 5
- Max Output Length: 30

Results

Bart-base



Moderate to Good Performance

Median

Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.56	0.53	0.87	0.71	0.48	0.46
Rouge2	0.31	0.26	0.67	0.44	0.32	0.26
Rouge-L	0.44	0.40	0.80	0.59	0.37	0.33
Rouge-Lsum	0.46	0.42	0.80	0.60	0.37	0.39
SacreBleu	14.57	11.20	52.04	24.66	17.85	11.50
Meteor	0.53	0.51	0.78	0.64	0.62	0.54

Standard Deviation

Metric	All	Career Statistics	Demographics	Film/Tv	MMA Record	Unspecified
Rouge1	0.23	0.19	0.20	0.17	0.26	0.22
Rouge2	0.24	0.20	0.33	0.21	0.28	0.21
Rouge-L	0.23	0.18	0.28	0.19	0.22	0.22
Rouge-Lsum	0.23	0.18	0.28	0.18	0.22	0.22
SacreBleu	22.03	15.71	33.55	20.66	26.43	17.29
Meteor	0.22	0.19	0.26	0.18	0.26	0.21

Results

Bart-base



Table	Generated Summary	Annotation
0	Daniel Henry Chamberlain was the 76th Governor of South Carolina.	Daniel Henry Chamberlin was the 76th Governor of South Carolina from 1874.
1	Alma Jodorowsky played Evelyn in the 2016 Kids in Love.	Alma Jodorowsky had the role of Evelyn in the 2016 film Kids in Love.
2	A. J. Hawk finished the 2006 season with 119 tackles, nine sacks, and three forced fumbles.	In his rookie season, Hawk led with 119 total tackles.



76 Daniel Henry Chamberlain
December 1, 1874



2016 Kids in Love Evelyn



119

- Min Output Length: 5
- Max Output Length: 30

Results

QTSumm

T5-small

Results

Metric	Average	Median	Standard Deviation	Minimum	Maximum
Rouge1	0.5214	0.5245	0.1341	0	0.9655
Rouge2	0.2856	0.2716	0.1360	0	0.8889
Rouge-L	0.3984	0.3846	0.1374	0	0.9655
Rouge-Lsum	0.3983	0.3846	0.1374	0	0.9655
SacreBleu	16.1109	13.5828	11.3419	0	69.997
Meteor	0.398	0.3764	0.1479	0	0.9364

Moderate Performance

Results

T5-small



Table	Query	Generated Summary	Annotation
0	Who are the Catholic justices who were appointed by a President Bush and what states are they from?	John Roberts from Maryland, Samuel Alito from New Jersey, and Sonia Sotomayor from New York are the Catholic justices who were appointed by a President Bush. Roberts was appointed in 2005, and Alito was appointed as incumbent in 2005.	John Roberts from Maryland and Samuel Alito from New Jersey were appointed by a President Bush.
1	Summarize the result of the game between Army Cadets and Stevens Tech.	The Army Cadets faced off against Stevens Tech in the first game of the season . The result of the game was a win for the team, with 27 points and a record of 1 - 0.	The Army Cadets played Stevens Tech in the first game of their season on October 5th and won 27-0. This was a shutout win for the Cadets, as they were able to keep their opponents' score at zero while scoring 27 of their own points. Overall, it was a successful start to the 1912 season and they were able to gain momentum going into the next game against Rutgers.



Focus on: John Roberts, Maryland, 1955, living, 2005, incumbent, G. W. Bush, —, Samuel Alito, New Jersey, 1950, living, 2006, incumbent, G. W. Bush, —



Focus on: 1, Oct 5, Stevens Tech, Win, 27, 0, 1 - 0

- Min Output Length: 1
- Max Output Length: 150

Results

T5-base

Results

Metric	Average	Median	Standard Deviation	Minimum	Maximum
Rouge1	0.4358	0.4348	0.1189	0	0.8462
Rouge2	0.2144	0.2035	0.1090	0	0.7500
Rouge-L	0.3164	0.3060	0.1070	0	0.8462
Rouge-Lsum	0.3165	0.3061	0.1070	0	0.8462
SacreBleu	10.2954	7.9374	8.4059	0	69.3098
Meteor	0.3281	0.3059	0.1400	0	0.8452

Low Performance

Results

T5-base

Table	Query	Generated Summary	Annotation
0	Who are the Catholic justices who were appointed by a President Bush and what states are they from?	who are the Catholic justices who were appointed by a President Bush and what states are they from?	John Roberts from Maryland and Samuel Alito from New Jersey were appointed by a President Bush.
1	Summarize the result of the game between Army Cadets and Stevens Tech.	Army Cadets play Stevens Tech in a game on Oct 5 . the game was won by the army, with a score of 27 points and a draw of 1 - 0. The result was a win for the army cadets .	The Army Cadets played Stevens Tech in the first game of their season on October 5th and won 27-0. This was a shutout win for the Cadets, as they were able to keep their opponents' score at zero while scoring 27 of their own points. Overall, it was a successful start to the 1912 season and they were able to gain momentum going into the next game against Rutgers.



Focus on: John Roberts, Maryland, 1955, living, 2005, incumbent, G. W. Bush, —, Samuel Alito, New Jersey, 1950, living, 2006, incumbent, G. W. Bush, —



Focus on: 1, Oct 5, Stevens Tech, Win, 27, 0, 1 - 0

- Min Output Length: 1
- Max Output Length: 150

Results

Bart-base

Results

Metric	Average	Median	Standard Deviation	Minimum	Maximum
Rouge1	0.5153	0.5127	0.1376	0	0.9796
Rouge2	0.2783	0.2621	0.1387	0	0.9362
Rouge-L	0.3951	0.3739	0.1375	0	0.9796
Rouge-Lsum	0.3951	0.3750	0.1375	0	0.9796
SacreBleu	16.2992	13.9352	11.5794	0	89.854
Meteor	0.4000	0.3801	0.1512	0	0.9954

Moderate Performance

Results

Bart-base



Table	Query	Generated Summary	Annotation
0	Who are the Catholic justices who were appointed by a President Bush and what states are they from?	The Catholic Supreme Court appointed two Catholic justices who were appointed by a President Bush: John Roberts from Maryland and Samuel Alito from New Jersey. John Roberts was born in 1955 and died in 2005. He was incumbent of the court and did not manage to terminate his term. Samuel Alito, born in 1950, was incumbent for the duration of his tenure.	John Roberts from Maryland and Samuel Alito from New Jersey were appointed by a President Bush.
1	Summarize the result of the game between Army Cadets and Stevens Tech.	The Army Cadets faced off against Stevens Tech on October 5th. The Cadets won the game with a score of 27-0. This put them at a 1-0 record.	The Army Cadets played Stevens Tech in the first game of their season on October 5th and won 27-0. This was a shutout win for the Cadets, as they were able to keep their opponents' score at zero while scoring 27 of their own points. Overall, it was a successful start to the 1912 season and they were able to gain momentum going into the next game against Rutgers.



Focus on: John Roberts, Maryland, 1955, living, 2005, incumbent, G. W. Bush, —, Samuel Alito, New Jersey, 1950, living, 2006, incumbent, G. W. Bush, —



Focus on: 1, Oct 5, Stevens Tech, Win, 27, 0, 1 - 0

- Min Output Length: 1
- Max Output Length: 150

Results

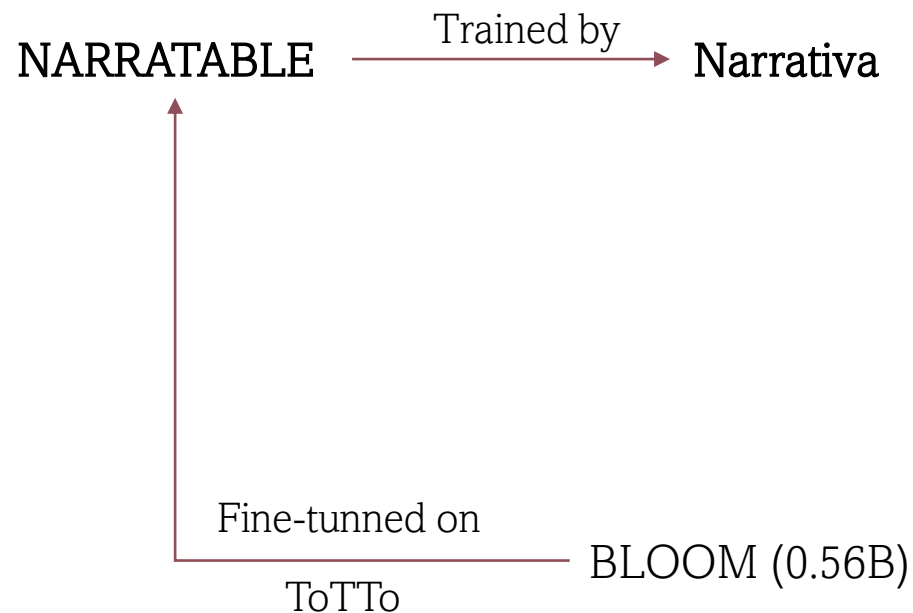
Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Results
7. **Comparative Data Analysis**
8. Conclusion and Future Work

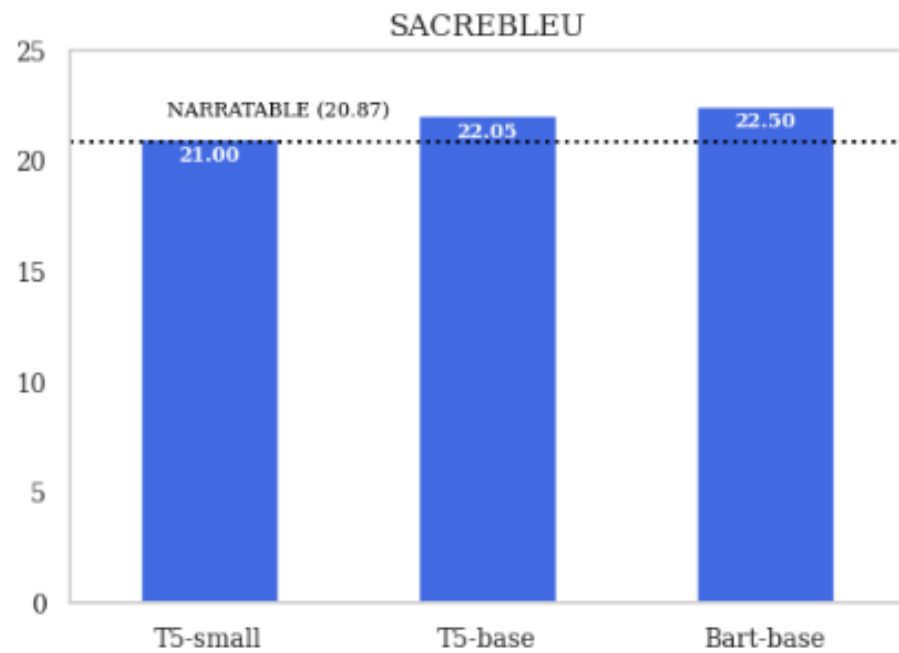
ToTTo

Benchmark

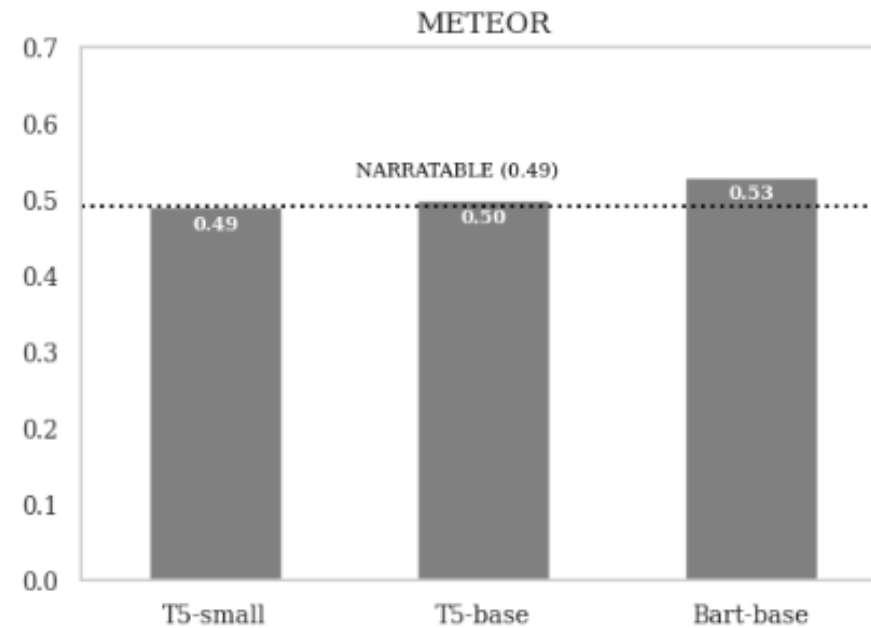
Table Template Example

[illegible]

SACREBLEU - METEOR



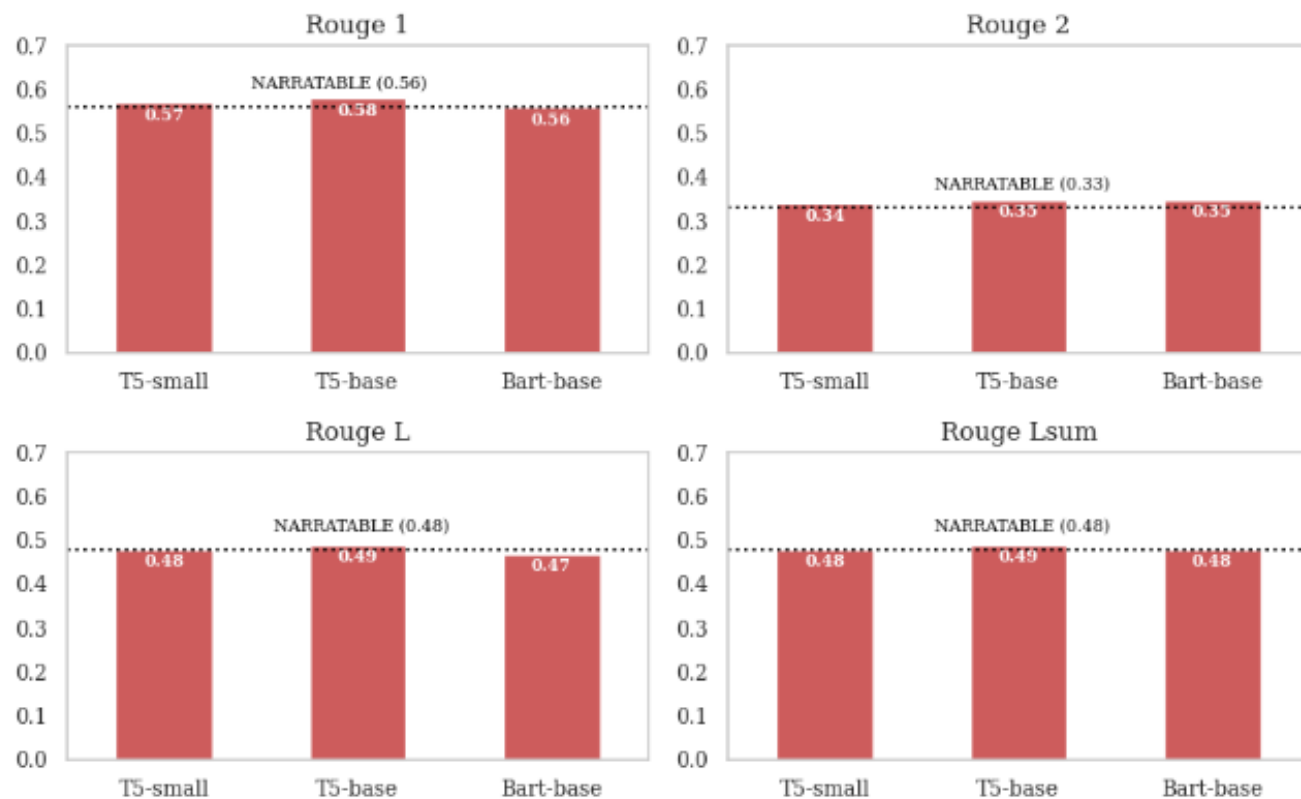
SacreBleu Range: 0 - 100



METEOR Range: 0 - 1

ROUGE

Rouge Metrics



Standard Deviation

Metric	T5-small	T5-base	Bart-base
Rouge1	0.20	0.21	0.23
Rouge2	0.23	0.23	0.24
Rouge-L	0.21	0.21	0.23
Rouge-Lsum	0.21	0.21	0.23
SacreBleu	19.56	20.19	22.03
Meteor	0.22	0.23	0.22

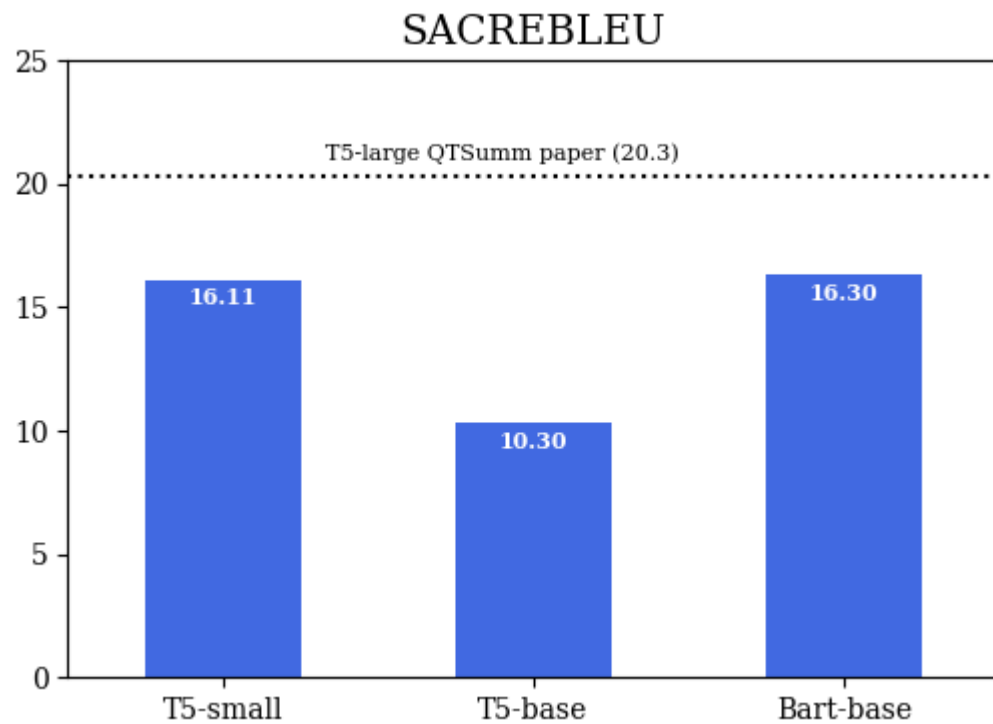
QTSumm

Benchmark

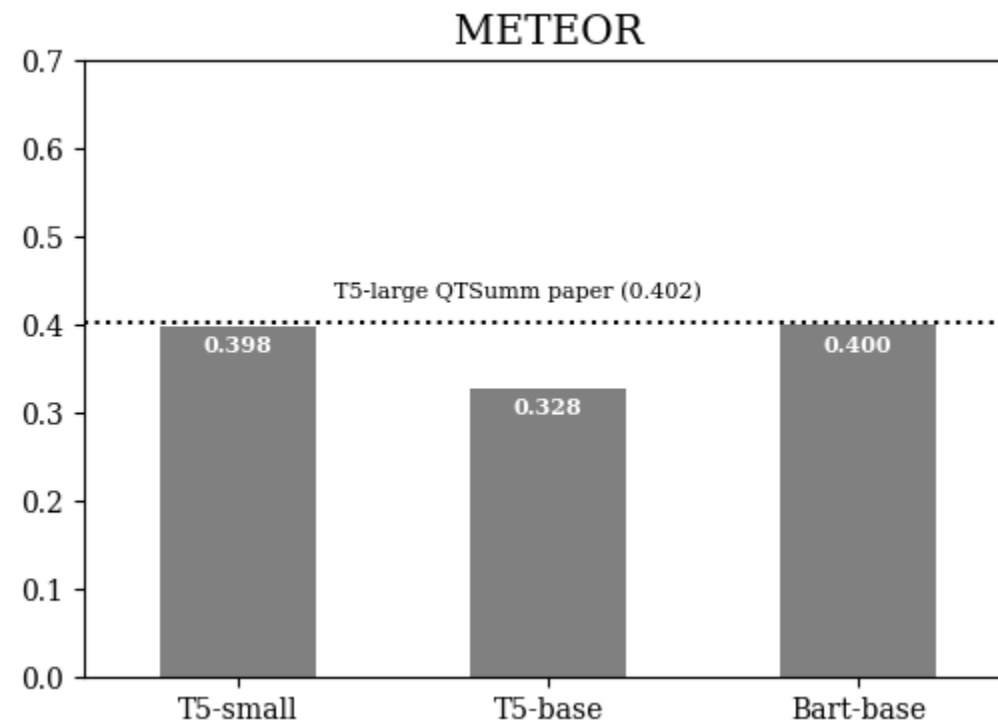
T5-large $\xrightarrow{\text{Fine-tuned by}}$ Yilun Zhao et al. $\xrightarrow{\text{On}}$ QTSumm

Input Data \longrightarrow $T = \overset{\text{Serialized Table}}{[HEADER] : h, [ROW]_1 : r1, \dots, [ROW]_n : rn} + \text{Query}$

SACREBLEU - METEOR



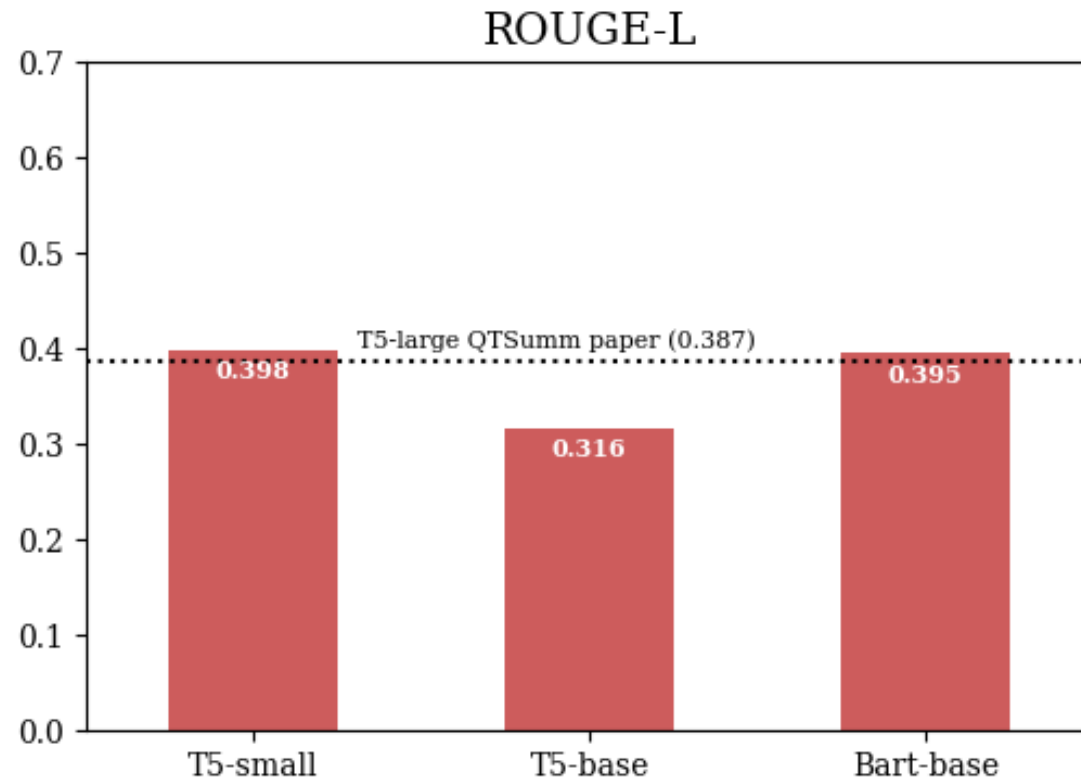
SacreBleu Range: 0 - 100



METEOR Range: 0 - 1

Comparative Data Analysis

ROUGE-L



Standard Deviation

Metric	T5-small	T5-base	Bart-base
Rouge1	0.1341	0.1189	0.1376
Rouge2	0.1360	0.1090	0.1387
Rouge-L	0.1374	0.1070	0.1375
Rouge-Lsum	0.1374	0.1070	0.1375
SacreBleu	11.3419	8.4059	11.5794
Meteor	0.1479	0.1400	0.1512

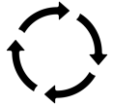
Outline

1. Introduction
2. Datasets
3. Preprocessing
4. Models
5. Evaluation Metrics
6. Results
7. Comparative Data Analysis
8. **Conclusions and Future Work**

Conclusions



Highest Performance on ToTTo & QTSumm – **All Test Set** : Bart-base



If **consistency** is a priority **T5-small** is a better option for **QTSumm**

↓ Standard Deviation – Similar Performance



Notable Performance - Top 5 ToTTo Domains

T5 Variations : MMA Record

Bart-base : Demographics



QTSumm is a more **challenging task** compared to ToTTo for all models of choice.

Future Work



Llama2 7B chat was fine-tuned on ToTTo using LORA.



Necessary computational resources for inference surpass the 40 GB of available GPU RAM.



No available results.



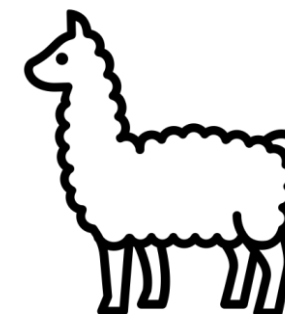
The next step : obtain enough GPU RAM to generate all test set summaries and publish the results.

Future Work

Llama 7B Chat
Summaries

Sample

Table	Generated Summary	Annotation
0	Daniel Henry Chamberlain was the 76th Governor of South Carolina.	Daniel Henry Chamberlin was the 76th Governor of South Carolina from 1874.
1	Sure! Here is the summary of the text in one sentence: In 2016, Alma Jodorowsky played Evelyn in the film Kids in Love.	Alma Jodorowsky had the role of Evelyn in the 2016 film Kids in Love.
2	In 2006, A.J. Hawk played in 16 games, combined tackles were 119, assisted tackles were 37, sacks were 3.5, forced fumbles were 1, fumble recoveries were 2, interceptions were 2, passes defended were 8, and longest interception return was 25.	In his rookie season, Hawk led with 119 total tackles.
3	Peter II the Simple (Pêr II) ruled from 1450 to 1457, succeeded by Arthur III the Justicier (Arzhur III) who ruled from 1457 to 1458.	At the very end of his life, Arthur III became duke of Brittany, succeeding Peter II.





Thank You!

Appendix

EDA of Summaries : ToTTo

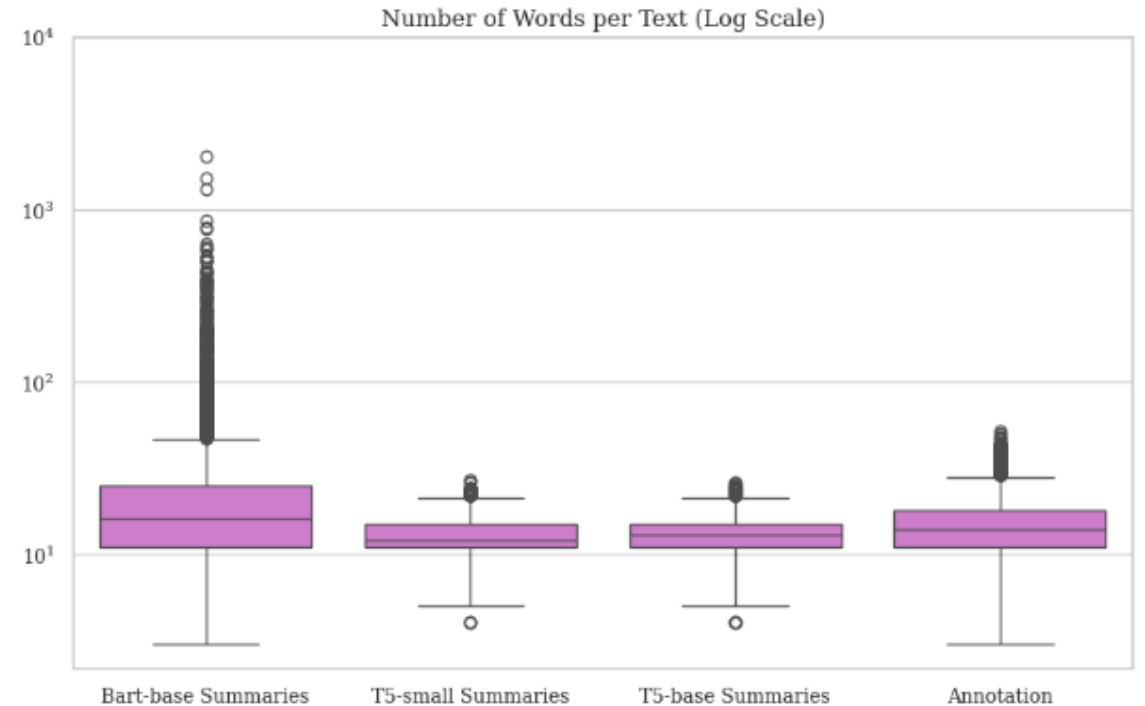
Number of Words per Text

Model	Average	Median	Standard Deviation	Minimum	Maximum
T5-small	13	12	3	4	27
T5-base	13	13	3	4	26
Bart-base	27	16	53	3	2029
Annotation	15	14	6	3	52

Flesch Reading Ease Score

Model	Average	Median	Standard Deviation
T5-small	76	77	17
T5-base	76	77	17
Bart-base	75	76	17
Annotation	74	75	18

70 - 79 : Fairly Easy to comprehend.



EDA of Summaries : QTSumm

Number of Words per Text

Model	Average	Median	Standard Deviation	Minimum	Maximum
T5-small	45	41	19	12	109
T5-base	45	44	13	13	96
Bart-base	48	44	21	12	117
Annotation	58	53	28	1	138

Flesch Reading Ease Score

Model	Average	Median	Standard Deviation
T5-small	72	74	15
T5-base	74	76	14
Bart-base	72	74	15
Annotation	58	53	28

50 – 59 : Fairly Difficult

70 - 79 : Fairly Easy

Number of Words per Text

