

# Multilingual News article Similarity based on BERT and Roberta



Department of Computer Science and Engineering  
National Institute of Technology, Silchar

**Presented By (Group No. 49):**

**Samudranil Dutta (1912014)**

**Rajdeep Paul (1912038)**

**Veeranjaneyulu Chowdary  
Battula (1912103)**

**Under the Supervision of : Dr. Partha Pakray**

# Outline

1. Introduction
2. Problem Statement
3. Motivation
4. Literature Survey
5. Objectives
6. Dataset Preparation
7. Methodology
8. Observation
9. Evaluation Results
10. System Prototype
11. Timeline
12. References

- Multilingual News Article Similarity is the task of identifying whether news articles address the same subject or not.
- Understanding which articles refer to the same story can not only improve applications like news aggregation but enable cross-linguistic analysis of media consumption and attention.
- This will cover measuring sentence and document similarity.

# PROBLEM STATEMENT

To calculate the similarity between two distinct news articles (English-Regional language) and develop a system for determining whether the two news articles are similar or not.

Measuring news article similarity enables:

- comparison of news sources coverage
- helps identifying the stories that dominate the media agenda
- tracks the diffusion of news items over time within a media ecosystem.

We found that many people had worked on this previously, but very few covered Indian languages. So we plan to work for English-Indian language pairs and prepare a standard dataset consisting of English-Indian language Pairs.

If news articles can be properly clustered, they can be used for a wide range of purposes, such as recommendation and displaying related articles.

# Objectives

- To translate the news article from regional language into english language.
- To identify the named entities.
- To calculate the similarity between the two distinct news articles.

# Literature Survey

S.No	Authors	Title of the paper	Findings
1	Xi Chen, Ali Zeynali, Chico Camargo, Fabian Flöck, Devin Gaffney, Przemyslaw Grabowicz, Scott Hale, David Jurgens, Mattia Samory	SemEval-2022 Task 8: Multilingual news article similarity	Xi et al. measured similarity on basis of following annotators: <b>GEO, ENT, TIME, NAR, STYLE, TONE</b> . Errors were particularly common when the news articles shared some similarity in terms of their geographic location, named entities, time, narratives, style, tone but were nonetheless dissimilar overall.
2	Shotaro Ishihara, Hono Shirai	Nikkei at SemEval-2022 Task 8: Exploring BERT-based Bi-Encoder Approach for Pairwise Multilingual News Article Similarity	The authors used BERT based Bi-Encoder approach for SemEval-2022 task 8. The experiments done in this paper showed that Bi-Encoder architecture worked better than Cross-Encoder.

# Literature Survey

S.No	Authors	Title of the paper	Findings
3	Ritika Singh, Satwinder Singh	Text Similarity Measures in News Articles by Vector Space Model Using NLP	They conducted a comparison of three different methods to estimate the semantic similarity among two news articles. The three methodologies are the Cosine similarity with tf-idf vectors, Jaccard similarity with tf-idf vectors, Bag of words Euclidean distance. All three of these methods showed promising results, but among these three methods, cosine similarity using tf-idf showed greater accuracy, recall and F-measure scores of 81.25%, 100% and 76.92%, respectively.
4	Hameed, Naamah Hussien, Adel M. Alimi, and Ahmed T. Sadiq	Short Text Semantic Similarity Measurement Approach Based on Semantic Network	The authors presented a method for the semantic similarity measurement between texts based on the semantic network. The threshold value has an important effect on the results. Choosing a higher threshold value leads to lower similarity values between the compared texts and vice versa.



# Dataset Preparation

- Prepared test dataset that contain a pair of news articles.
- Dataset prepared so far consist of 200 pairs each.
- The first dataset consist of English news and Bengali news as pairs.
- The second dataset consist of English news and Assamese news as pairs.
- The third dataset consist of English news and Telegu news as pairs.
- We have manually tagged it as similar or dissimilar.

# Dataset

Source	Number of Articles
The Print	80
Mint	70
India Today	60
The Economic Times	45
NDTV	90

Source	Number of Articles
The Times of India	75
Hindustan Times	65
CNN	35
Reuters	25
BBC	35
The Indian Express	20

TABLE-I : ENGLISH NEWS ARTICLES SOURCE

# Dataset

Source	Number of articles
Anandabazar Patrika	40
ABP Ananda	40
Hindustan Times Bangla	10
News18 Bangla	10
Zee 24Ghanta	20
Tv9 Bangla	30
Sangbad Pratidin	20
oneindia Bengali	30

**TABLE II : BENGALI NEWS ARTICLES SOURCE**

# Dataset

Source	Number of articles
Asomiya Pratidin	30
Niyomiya Barta	25
NorthEast now	25
Time8	25
DY365	25
News18 Assam	50
Gana Adhikar	20

**TABLE III : ASSAMESE NEWS ARTICLES SOURCE**

# Dataset

Source	Number of articles
TV9 telugu	50
Andhra Jyothi	50
Andhra Prabha	20
TV5	40
Sakshi	15
News 18 Telugu	25

**TABLE IV: TELUGU NEWS ARTICLES SOURCE**

We shall be doing three things:

## **Translation:**

- The regional language is translated to English language.
- It is achieved by using the microsoft bing translate api which converts the given text of a language into target language.

## **Named Entity Recognition:**

- We find out the named entities from the given text and classify them into different categories.
- Example of categories include Person, Organization, Place/Location, etc.
- Stanza and corenlp API are used for doing the same.
- Core NLP is able to recognize named entities better than stanza.

## Calculating Similarity:

- To find out the similarities between the two articles Cosine similarity is used.
- In case of Cosine similarity, the text is first converted to vectors and the similarity of two vectors are calculated by taking the dot product and dividing it by the magnitudes of each vector.
- Cosine similarity gives an output between 0 to 1. 1 means the documents are most similar and 0 means they are dissimilar to each other.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

## BERT:

- Bidirectional Encoder Representations from Transformers.
- BERT makes use of Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text.
- As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once. Therefore it is considered bidirectional, though it would be more accurate to say that it's non-directional. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).



## **RoBERTa:**

- Robustly optimized BERT approach RoBERTa, is a retraining of BERT with improved training methodology, 1000% more data and compute power.
- To improve the training procedure, RoBERTa removes the Next Sentence Prediction (NSP) task from BERT's pre-training and introduces dynamic masking so that the masked token changes during the training epochs.
- Additional data included from CommonCrawl News dataset (63 million articles, 76 GB), Web text corpus (38 GB) and Stories from Common Crawl (31 GB)

- We tested the bert and roberta model on the sample 200 test cases prepared each for English-Bengali, English-Assamese and English-Telugu pairs. The number of correct predictions obtained were used to calculate the threshold value.
- If the result obtained is below the threshold value then the articles will be considered as dissimilar. Else it will be considered as similar.
- To calculate the threshold value we plotted accuracy vs threshold value graph.

**Accuracy= Number of correct predictions/Total number of predictions**

# MANUAL TAGGING

Verdict	HE 1	HE 2	HE 3
Similar	122	122	120
Dissimilar	78	78	80

**TABLE IV HUMAN EVALUATOR (HE) STATISTICS FOR ENGLISH-BENGALI LANGUAGE PAIR**

# MANUAL TAGGING

Verdict	HE 1	HE 2	HE 3
Similar	119	125	119
Dissimilar	81	75	81

**TABLE V HUMAN EVALUATOR (HE) STATISTICS FOR ENGLISH-ASSAMESE LANGUAGE PAIR**

# MANUAL TAGGING

<b>Verdict</b>	<b>HE 1</b>	<b>HE 2</b>	<b>HE 3</b>
<b>Similar</b>	<b>120</b>	<b>118</b>	<b>120</b>
<b>Dissimilar</b>	<b>80</b>	<b>82</b>	<b>80</b>

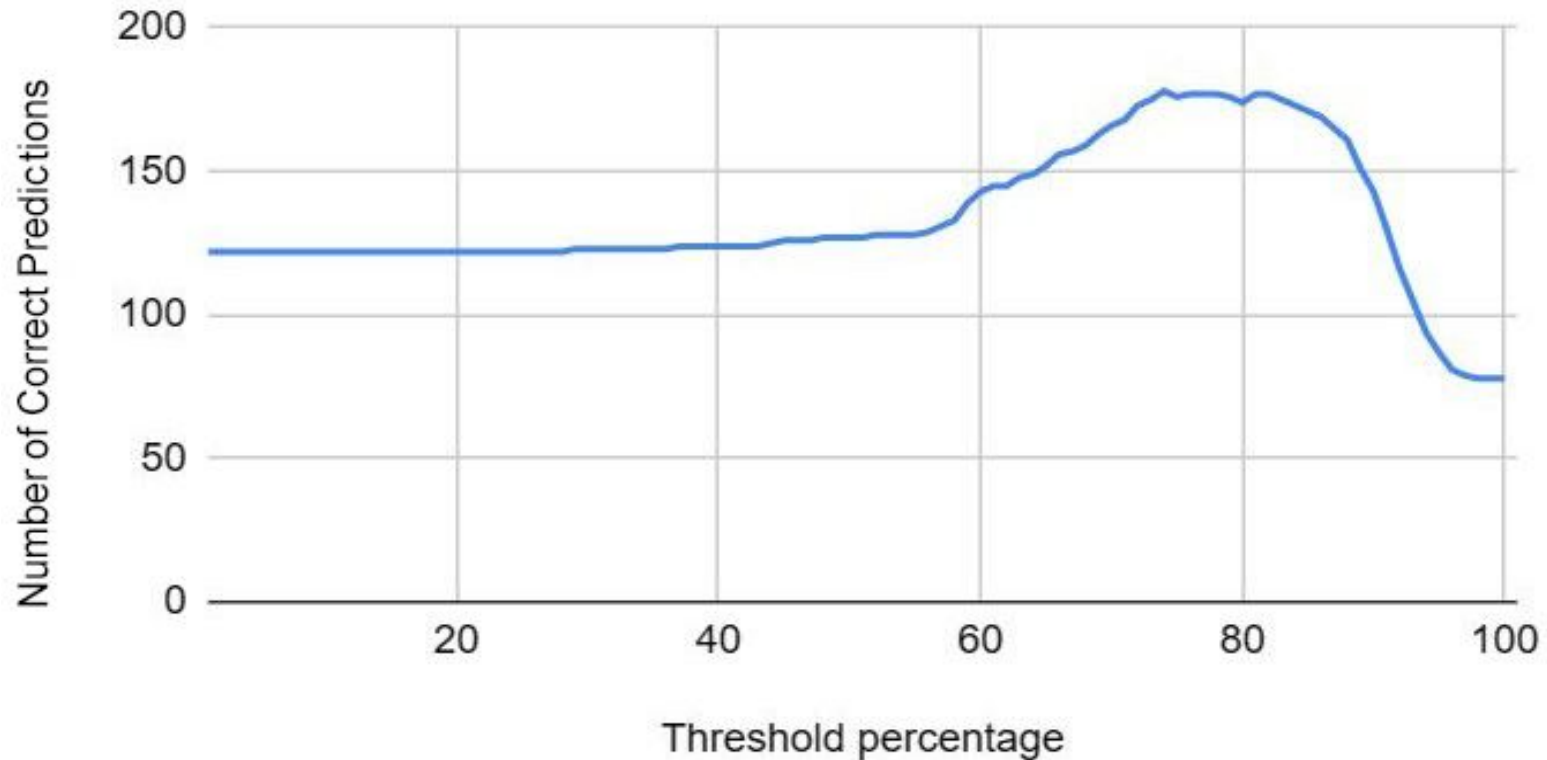
**TABLE VI HUMAN EVALUATOR (HE) STATISTICS FOR ENGLISH-TELEGU LANGUAGE PAIR**

# DATASET SUMMARY

Language pair	Similar	Dissimilar
English-Assamese	119	81
English Bengali	122	78
English Telugu	120	80

# Observation and Results

## Threshold value calculation for English-Bengali pair using BERT



# Observation and Results

- BERT model on English-Bengali pair

BERT	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	119 (TP)	19 (FP)
SYSTEM TAGGED DISSIMILAR	3 (FN)	59 (TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $178 / 200 = 0.89$

**PRECISION(P)** =  $TP / TP+FP = 119 / 119+19 = 0.862$

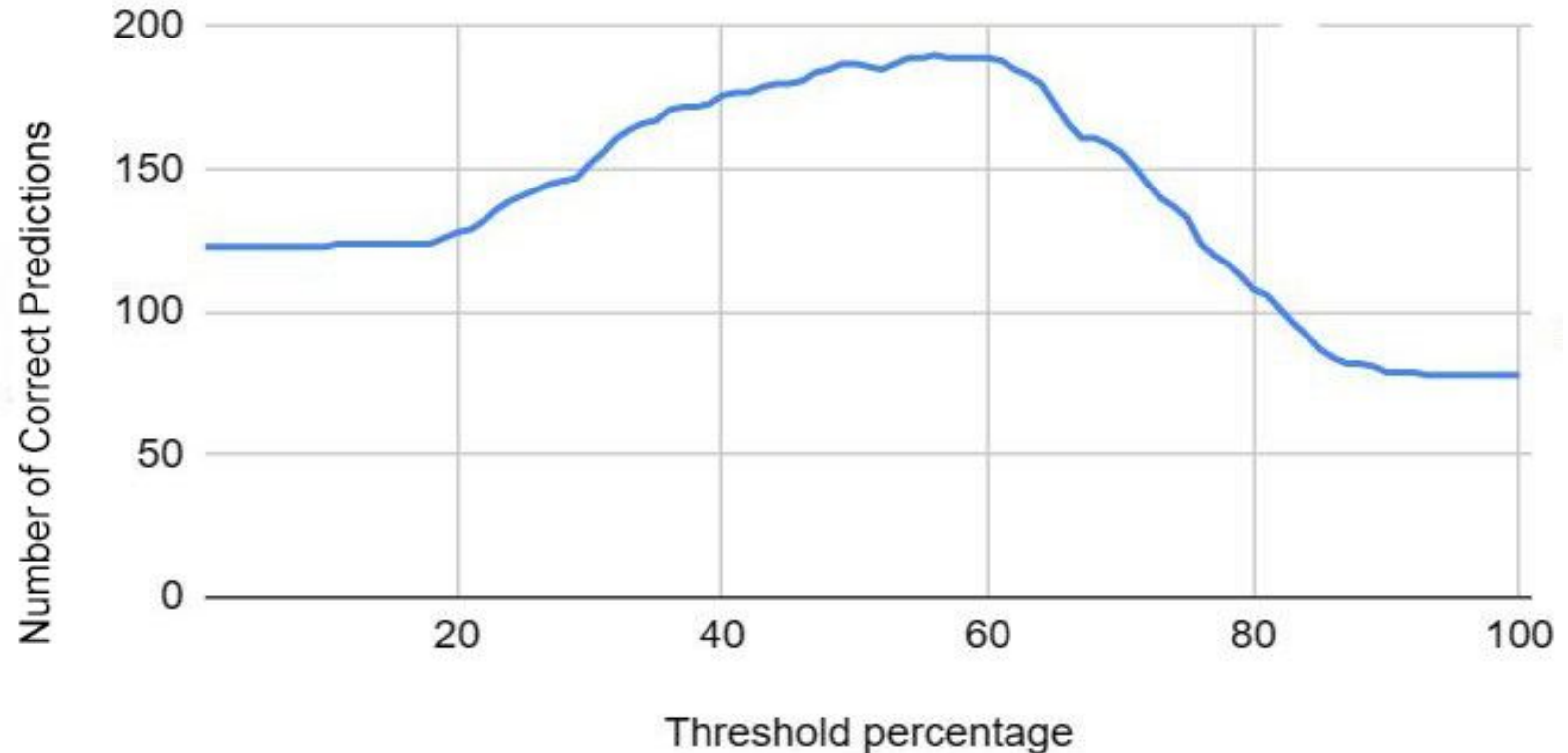
**RECALL(R)** =  $TP / TP+FN = 119 / 119+3 = 0.974$

**F – SCORE** =  $2PR / P+R = 0.914$



# Observation and Results

## Threshold value calculation for English-Bengali pair using RoBERTa



# Observation and Results

- RoBERTa model on English-Bengali pair

RoBERTa	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	117 (TP)	5 (FP)
SYSTEM TAGGED DISSIMILAR	5 (FN)	73 (TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $190 / 200 = 0.95$

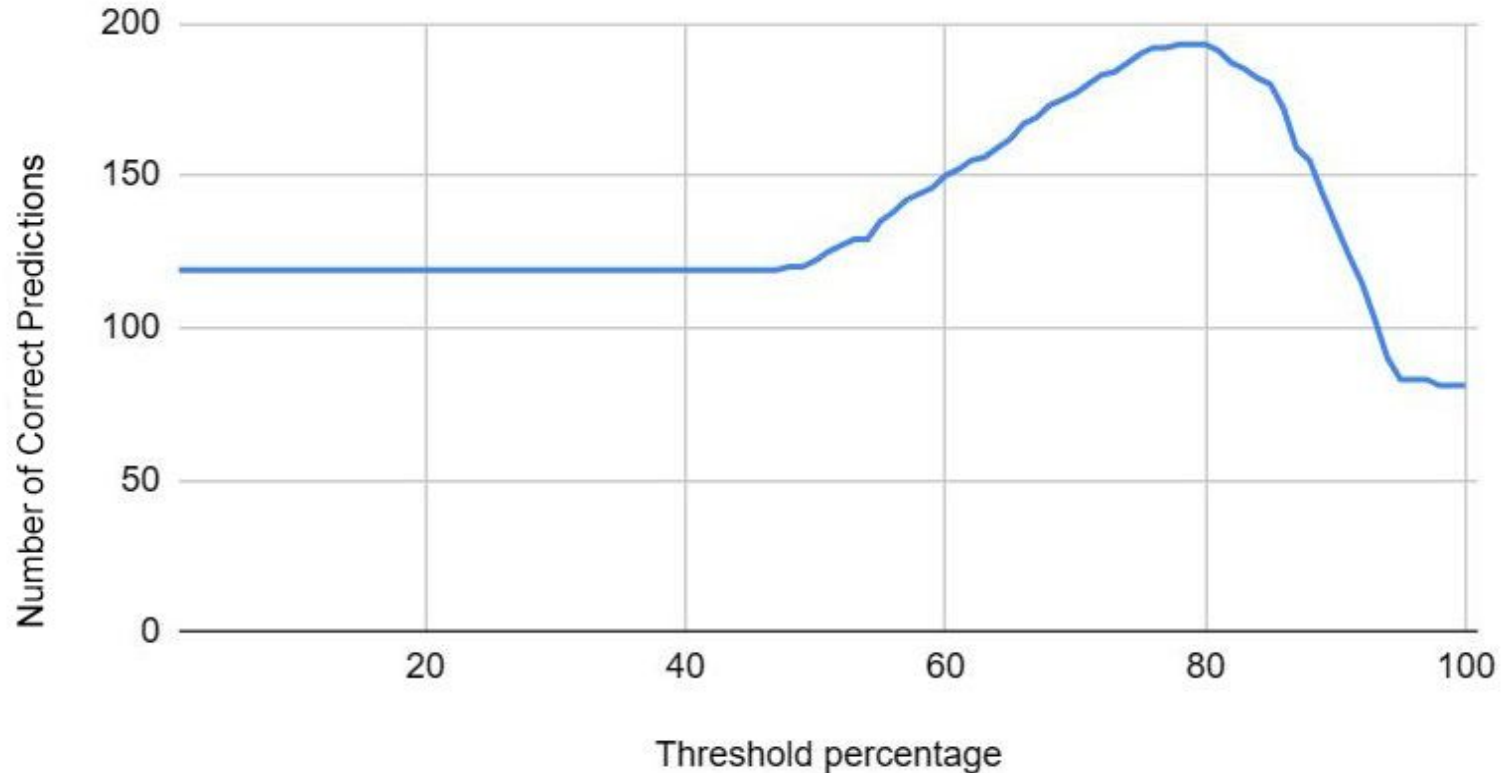
**PRECISION(P)** =  $TP / TP+FP = 117 / 117+5 = 0.959$

**RECALL(R)** =  $TP / TP+FN = 117 / 117+5 = 0.959$

**F – SCORE** =  $2PR / P+R = 0.959$

# Observation and Results

## Threshold value calculation for English-Assamese pair using BERT



# Observation and Results

- BERT model on English-Assamese pair

BERT	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	116 (TP)	4 (FP)
SYSTEM TAGGED DISSIMILAR	3 (FN)	77 (TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $193 / 200 = 0.965$

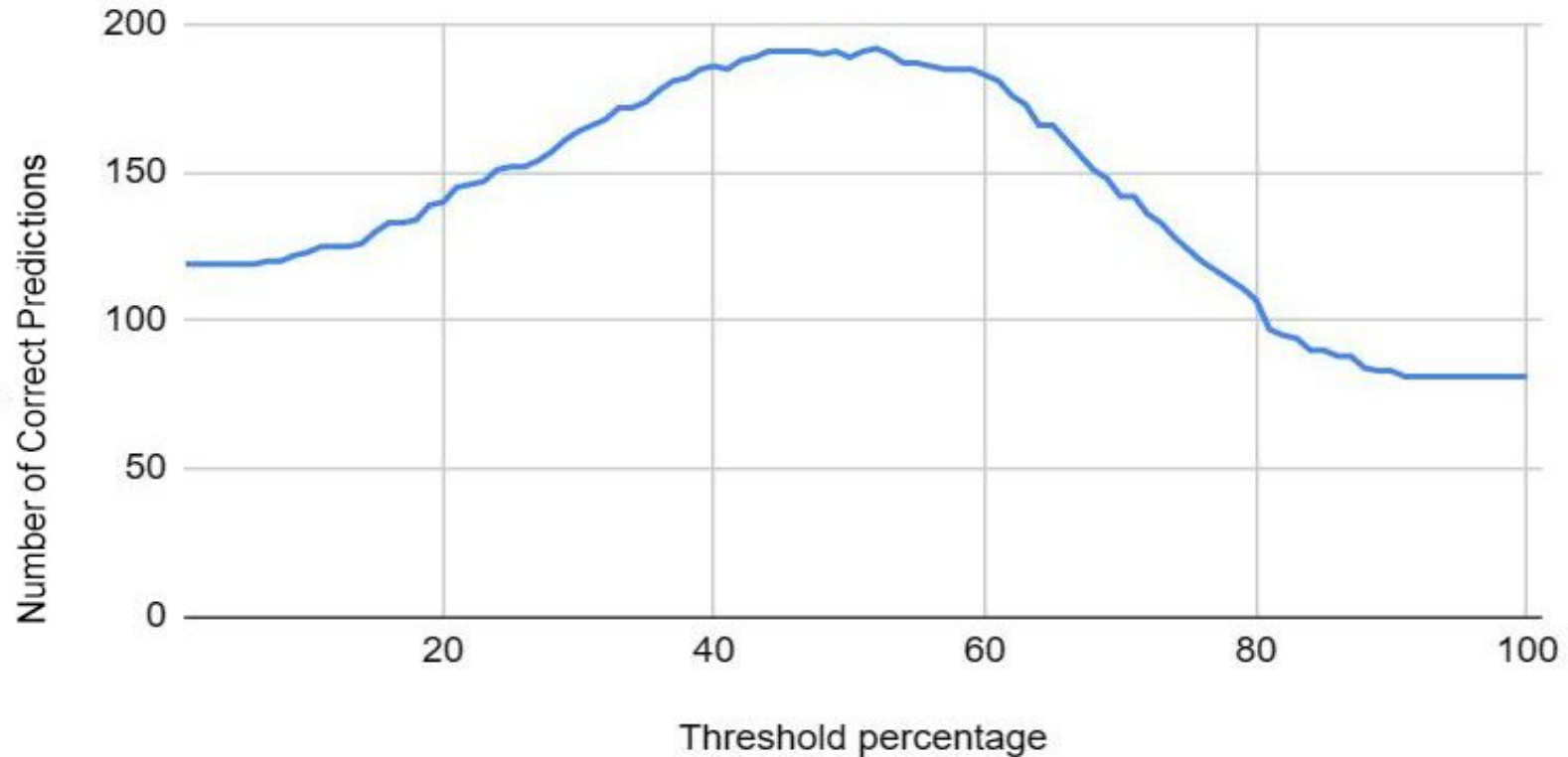
**PRECISION(P)** =  $TP / TP+FP = 116 / 116+4 = 0.966$

**RECALL(R)** =  $TP / TP+FN = 116 / 116+3 = 0.974$

**F - SCORE** =  $2PR / P+R = 0.97$

# Observation and Results

## Threshold value calculation for English-Assamese pair using RoBERTa



# Observation and Results

- RoBERTa model on English-Assamese pair

BERT	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	113 (TP)	2 (FP)
SYSTEM TAGGED DISSIMILAR	6 (FN)	79 (TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $192 / 200 = 0.96$

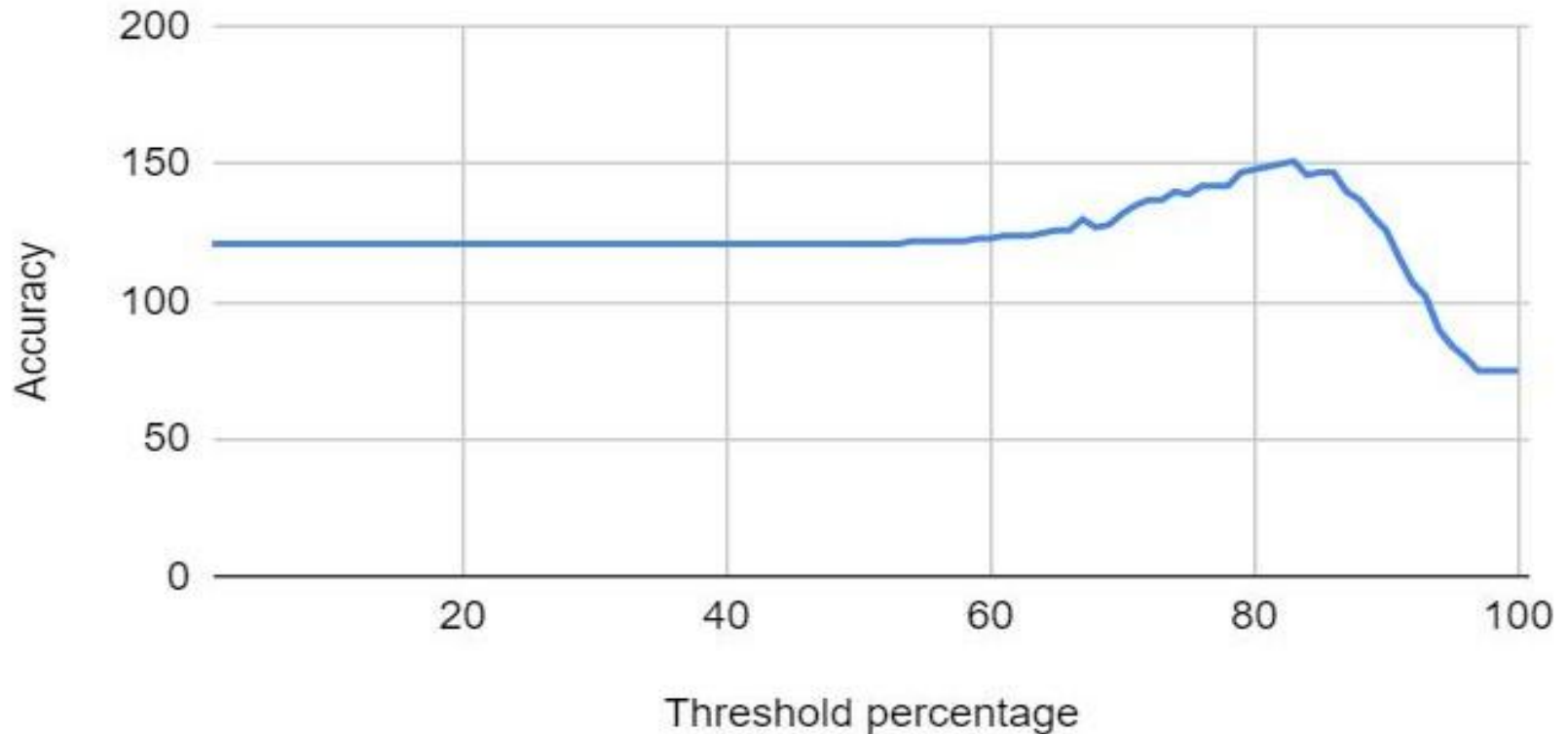
**PRECISION(P)** =  $TP / TP+FP = 113 / 113+2 = 0.982$

**RECALL(R)** =  $TP / TP+FN = 113 / 113+6 = 0.949$

**F – SCORE** =  $2PR / P+R = 0.965$

# Observation and Results

## Threshold value calculation for English-Telugu pair using BERT



# Observation and Results

- BERT model on English-Telugu pair

BERT	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	99(TP)	20 (FP)
SYSTEM TAGGED DISSIMILAR	25(FN)	56(TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $155 / 200 = 0.77$

**PRECISION(P)** =  $TP / TP+FP = 99 / 99+20 = 0.831$

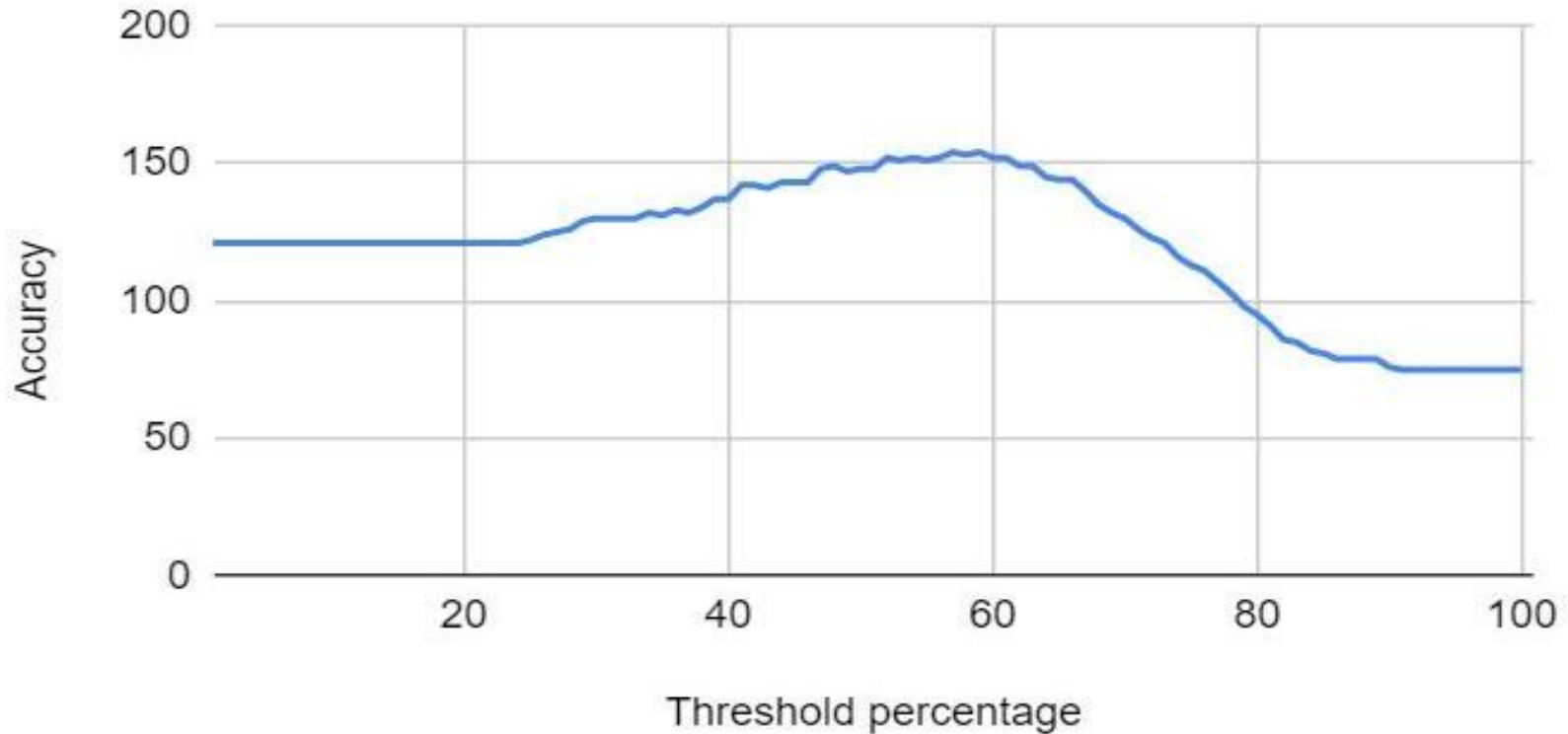
**RECALL(R)** =  $TP / TP+FN = 99 / 99+25 = 0.798$

**F - SCORE** =  $2PR / P+R = 0.813$



# Observation and Results

## Threshold value calculation for English-Telugu pair using RoBERTa



# Observation and Results

- RoBERTa model on English-Telugu pair

BERT	ACTUAL SIMILAR	ACTUAL DISSIMILAR
SYSTEM TAGGED SIMILAR	91(TP)	12 (FP)
SYSTEM TAGGED DISSIMILAR	30(FN)	67(TN)

**ACCURACY** = Correct Prediction / Total Predictions =  $158 / 200 = 0.79$

**PRECISION(P)** =  $TP / TP+FP = 91/91+12=0.883$

**RECALL(R)** =  $TP / TP+FN = 91/91+30=0.752$

**F – SCORE** =  $2PR / P+R = 0.812$

# Result

- **THRESHOLD VALUE :**

Model	Language Pair	Threshold (in %)	No. of correct predictions (200)
<b>BERT</b>	English-Assamese	78	193
<b>RoBERTa</b>	English-Assamese	52	192
<b>BERT</b>	English-Bengali	74	178
<b>RoBERTa</b>	English-Bengali	56	190
<b>BERT</b>	English-Telugu	83	155
<b>RoBERTa</b>	English-Telugu	57	158

# System Description

The front end is made using HTML, CSS, ReactJS.

The back end is made using flask (python). MongoDB is used as the database.

Features in our application:

1. **Signup**: Create a new account.
2. **Login**: For existing users.
3. **History**: View the operations performed (stores the last 25 operations performed).
4. **Upload**: Upload .xlsx file and automatically get the BERT and RoBERTa score calculated.
5. **File History**: View the files uploaded and can also be downloaded.
6. **Logout**: End of a session.

Paste English Text

Paste text of any other language

Step 1

TRANSLATE

Paste English Text

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped the head of his wife's lover with an axe after catching them together. Sundi has been arrested by the police and an investigation has been initiated into the matter.

Vishwanath Sundi had been suspicious about his wife having an illicit affair with Shyamlal Hembram, a young man from Segaisai village. On Friday night, his suspicion was proven right when he caught his wife with Hembram in a compromising position. Hembram had come to Lonjo to meet Sundi's wife.

Seeing the two of them together, a furious Vishwanath Sundi started thrashing his wife's lover. He then proceeded to drag Hembram and tied him to a tree near his house.

After tying his wife's lover, Sundi picked up an axe and chopped his head off. Shyamlal Hembram died on the spot.

Sonuva police station in-charge Sohan Lal arrived in Lonjo village on Saturday morning along with other police personnel and arrested Vishwanath Sundi. The police also recovered Shyamlal's body and the axe used to decapitate him. The body has been sent for an autopsy and an investigation has been initiated into the matter.



Step 1

Paste text of any other language

স্ত্রীর বিবাহবহির্ভূত সম্পর্ক রয়েছে, সন্দেহ করতেন স্বামী। তাকে তাকে ছিলেন তিনি। শুক্রবার রাতে প্রেমিকের সঙ্গে স্ত্রীকে ঘনিষ্ঠ অবস্থায় দেখে ফেলেন তিনি। রাগের বশে প্রেমিককে প্রথমে মারধর করেন। তার পর গাছের সঙ্গে বেঁধে কুড়ুল দিয়ে মাথা কেটে ফেলেন বলে অভিযোগ। ঘটনাটি ঘটেছে ঝাড়খণ্ডের লঞ্জো গ্রামে।

পুলিশ সূত্রে খবর, অভিযুক্তের নাম বিশ্বনাথ সুন্দি। তাঁর স্ত্রীর সঙ্গে পাশের গ্রামের এক যুবকের বিবাহবহির্ভূত সম্পর্ক গড়ে উঠেছিল। স্ত্রীর অন্য কোনও সম্পর্ক রয়েছে, সন্দেহ করতেন বিশ্বনাথ। সেই সন্দেহ সত্যি হয়। শুক্রবার বিশ্বনাথের স্ত্রীর সঙ্গে দেখা করতে এসেছিলেন প্রেমিক শ্যামলাল হেমব্রম। বাড়িতে ফিরতেই দু'জনকে ঘনিষ্ঠ অবস্থায় দেখতে পেয়ে মেজাজ হারিয়ে ফেলেন বিশ্বনাথ। অভিযোগ, এর পরই শ্যামলালকে ঘর থেকে টেনে বার করেন বিশ্বনাথ। তার পর বেধড়ক মারধর করেন। এখানেই থেমে থাকেননি বিশ্বনাথ। কাছেই একটি গাছের সঙ্গে শ্যামলালকে বাঁধেন। তার পর কুড়ুল নিয়ে এসে শ্যামলালের ধড় থেকে মস্তিষ্ক আলাদা করে দেন। ঘটনাস্থলেই মৃত্যু হয় শ্যামলালের।

খবর পেয়ে শনিবার সকালে লঞ্জো গ্রামে পৌঁছয় সোহানুভা থানার দায়িত্বপ্রাপ্ত আধিকারিক সোহন লাল। বিশ্বনাথকে বাড়ি থেকেই গ্রেফতার করা হয়। শ্যামলালের দেহ উদ্ধার করে ময়নাতদন্তের জন্য পাঠানো হয়। উদ্ধার করা হয়েছে খুনে ব্যবহৃত কুড়ুলটিও।



TRANSLATE

Paste English Text

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped the head of his wife's lover with an axe after catching them together. Sundi has been arrested by the police and an investigation has been initiated into the matter.

Vishwanath Sundi had been suspicious about his wife having an illicit affair with Shyamlal Hembram, a young man from Segaisai village. On Friday night, his suspicion was proven right when he caught his wife with Hembram in a compromising position. Hembram had come to Lonjo to meet Sundi's wife.

Seeing the two of them together, a furious Vishwanath Sundi started thrashing his wife's lover. He then proceeded to drag Hembram and tied him to a tree near his house.

After tying his wife's lover, Sundi picked up an axe and chopped his head off. Shyamlal Hembram died on the spot.

Sonuba police station in-charge Sohan Lal arrived in Lonjo village on Saturday morning along with other police personnel and arrested Vishwanath Sundi. The police also recovered Shyamlal's body and the axe used to decapitate him. The body has been sent for an autopsy and an investigation has been initiated into the matter.



Step 1

TRANSLATE

Step 2

EXTRACT NAMED ENTITIES

Paste text of any other language

The husband suspected that his wife had an extramarital affair. He was on the table. On Friday night, he saw his wife in a close relationship with her boyfriend. In a fit of rage, he first beat his lover. After that, he allegedly cut off his head with a knife tied to a tree. The incident took place in Jharkhand's Lanjo village.

According to the police, the accused has been identified as Vishwanath Sundi. His wife had an extramarital affair with a young man from a nearby village. Vishwanath suspected that his wife had some other relationship. That doubt is true. On Friday, His lover Shyamlal Hembram came to meet Vishwanath's wife. On returning home, Vishwanath lost his temper after seeing the two in close proximity. It is alleged that Vishwanath then dragged Shyamlal out of the house. After that, he was beaten up. Vishwanath did not stop here. He tied Shyamlal to a nearby tree. After that, he brought the axe and separated mundu from Shyamlal's torso. Shyamlal died on the spot.

Sohan Lal, officer-in-charge of Sonuba police station, reached Lonjo village on Saturday morning. Vishwanath was arrested from his home. Shyamlal's body was recovered and sent for autopsy. The axe used in the murder has also been recovered.



Paste English Text

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped his lover with an axe after catching them together. Sundi has been arrested. Investigation has been initiated into the matter.

Vishwanath Sundi had been suspicious about his wife having an illicit relationship with Hembram, a young man from Segaisai village. On Friday night, his suspicion was confirmed when he caught his wife with Hembram in a compromising position. He came to Lonjo to meet Sundi's wife.

Seeing the two of them together, a furious Vishwanath Sundi started beating his lover. He then proceeded to drag Hembram and tied him to a tree near the village.

After tying his wife's lover, Sundi picked up an axe and chopped his wife. Hembram died on the spot.

Sonuba police station in-charge Sohan Lal arrived in Lonjo village on Saturday along with other police personnel and arrested Vishwanath Sundi. The body of Shyamlal's wife and the axe used to decapitate him. The body has been recovered and an investigation has been initiated into the matter.

Parameters	English Language	Regional Language
DATE	<ul style="list-style-type: none"> <li>2022-09-09TNI</li> <li>2022-09-10TMO</li> </ul>	<ul style="list-style-type: none"> <li>2022-09-09TNI</li> <li>2022-09-09</li> <li>2022-09-10TMO</li> </ul>
LOCATION	<ul style="list-style-type: none"> <li>Lonjo</li> <li>Segaisai</li> <li>Lonjo</li> <li>Lonjo</li> </ul>	<ul style="list-style-type: none"> <li>Jharkhand</li> <li>Lanjo</li> <li>Sonubha</li> <li>Lonjo</li> </ul>
NUMBER	<ul style="list-style-type: none"> <li>2</li> </ul>	<ul style="list-style-type: none"> <li>2</li> </ul>
PERSON	<ul style="list-style-type: none"> <li>Vishwanath Sundi</li> <li>his</li> <li>Sundi</li> <li>Vishwanath Sundi</li> <li>Shyamlal Hembram</li> <li>his</li> <li>Hembram</li> <li>his</li> <li>he</li> <li>his</li> <li>Hembram</li> <li>Sundi</li> <li>Vishwanath Sundi</li> <li>his</li> <li>Hembram</li> <li>He</li> <li>him</li> </ul>	<ul style="list-style-type: none"> <li>his</li> <li>He</li> <li>he</li> <li>his</li> <li>her</li> <li>he</li> <li>his</li> <li>he</li> <li>his</li> <li>Vishwanath Sundi</li> <li>His</li> <li>Vishwanath</li> <li>his</li> <li>Shyamlal Hembram</li> <li>Vishwanath</li> <li>His</li> <li>Vishwanath</li> <li>his</li> <li>Vishwanath</li> </ul>

his wife had an extramarital affair. He was on the table. On Friday night, he was in a close relationship with her boyfriend. In a fit of rage, he first allegedly cut off his head with a knife tied to a tree. The incident took place in Lonjo village.

The accused has been identified as Vishwanath Sundi. His wife had a young man from a nearby village. Vishwanath suspected that his wife was having an affair. That doubt is true. On Friday, His lover Shyamlal Hembram was caught with Vishwanath's wife. On returning home, Vishwanath lost his temper and started beating her. It is alleged that Vishwanath then dragged Shyamlal Hembram to a nearby spot where he was beaten up. Vishwanath did not stop here. He tied Shyamlal Hembram to a tree after that, he brought the axe and separated mundu from Shyamlal Hembram and died on the spot.

Shyamlal Hembram, who was arrested from his home. Shyamlal's body was recovered and sent to the Sonubha police station, reached Lonjo village on Saturday. The body of the murder has also been recovered.





Paste English Text

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped the head of his wife's lover with an axe after catching them together. Sundi has been arrested by the police and an investigation has been initiated into the matter.

Vishwanath Sundi had been suspicious about his wife having an illicit affair with Shyamial Hembram, a young man from Segaisai village. On Friday night, his suspicion was proven right when he caught his wife with Hembram in a compromising position. Hembram had come to Lonjo to meet Sundi's wife.

Seeing the two of them together, a furious Vishwanath Sundi started thrashing his wife's lover. He then proceeded to drag Hembram and tied him to a tree near his house.

After tying his wife's lover, Sundi picked up an axe and chopped his head off. Shyamial Hembram died on the spot.

Sonuva police station in-charge Sohan Lal arrived in Lonjo village on Saturday morning along with other police personnel and arrested Vishwanath Sundi. The police also recovered Shyamial's body and the axe used to decapitate him. The body has been sent for an autopsy and an investigation has been initiated into the matter.



Paste text of any other language

The husband suspected that his wife had an extramarital affair. He was on the table. On Friday night, he saw his wife in a close relationship with her boyfriend. In a fit of rage, he first beat his lover. After that, he allegedly cut off his head with a knife tied to a tree. The incident took place in Jharkhand's Lonjo village.

According to the police, the accused has been identified as Vishwanath Sundi. His wife had an extramarital affair with a young man from a nearby village. Vishwanath suspected that his wife had some other relationship. That doubt is true. On Friday, His lover Shyamial Hembram came to meet Vishwanath's wife. On returning home, Vishwanath lost his temper after seeing the two in close proximity. It is alleged that Vishwanath then dragged Shyamial out of the house. After that, he was beaten up. Vishwanath did not stop [here](#). He tied Shyamial to a nearby tree. After that, he brought the axe and separated mundu from Shyamial's torso. Shyamial died on the spot.

Sohan Lal, officer-in-charge of Sonubha police station, reached Lonjo village on Saturday morning. Vishwanath was arrested from his home. Shyamial's body was recovered and sent for autopsy. The axe used in the murder has also been recovered.



Calculated Similarity

0.945249

Step 1

TRANSLATE

Step 2

EXTRACT NAMED ENTITIES

Loading.....

Step 3

CALCULATE BERT SIMILARITY

You can now see your last 25 records (operations performed).....

Prev

1 - 7

Next

### English News article

### Indian Regional Language News article

### Task Performed

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped the head of his wife's lover with an axe after catching them together. Sundi has been arrested by the police and an investigation has been initiated into the matter.

Vishwanath Sundi had been suspicious about his wife having an illicit affair with Shyamal Hembram, a young man from Segaisai village. On Friday night, his suspicion was proven right when he caught his wife with Hembram in a compromising position. Hembram had come to Lonjo to meet Sundi's wife.

Seeing the two of them together, a furious Vishwanath Sundi started thrashing his wife's lover. He then proceeded to drag Hembram and tied him to a tree near his house.

স্ত্রীর বিবাহবহির্ভূত সম্পর্ক রয়েছে, সন্দেহ করতেন স্বামী। তাকে তাকে ছিলেন তিনি। শুক্রবার রাতে প্রেমিকের সঙ্গে স্ত্রীকে ঘনিষ্ঠ অবস্থায় দেখে ফেলেন তিনি। রাগের বশে প্রেমিককে প্রথমে মারধর করেন। তার পর গাছের সঙ্গে বেঁধে কুড়ুল দিয়ে মাথা কেটে ফেলেন বলে অভিযোগ। ঘটনাটি ঘটেছে বাড়ুখণ্ডের লাঞ্জা গ্রামে।

পুলিশ সূত্রে খবর, অভিযুক্তের নাম বিশ্বনাথ সুন্ডি। তাঁর স্ত্রীর সঙ্গে পাশের গ্রামের এক যুবকের বিবাহবহির্ভূত সম্পর্ক গড়ে উঠেছিল। স্ত্রীর অন্য কোনও সম্পর্ক রয়েছে, সন্দেহ করতেন বিশ্বনাথ। সেই সন্দেহ সত্যি হয়। শুক্রবার বিশ্বনাথের স্ত্রীর সঙ্গে দেখা করতে এসেছিলেন প্রেমিক শ্যামলাল হেমব্রম। বাড়িতে ফিরতেই দু'জনকে ঘনিষ্ঠ অবস্থায় দেখতে পেয়ে মেজাজ হারিয়ে ফেলেন বিশ্বনাথ। অভিযোগ, এর পরই শ্যামলালকে ঘর থেকে টেনে বার করেন বিশ্বনাথ। তার পর বেধড়ক মারধর করেন। এখানেই থেমে থাকেননি বিশ্বনাথ। কাছেই একটি গাছের সঙ্গে শ্যামলালকে বাঁধেন।

BERT Similarity  
0.945249

Vishwanath Sundi, a resident of Jharkhand's Lonjo village chopped the head of his wife's lover with an axe after catching them together. Sundi has been arrested by the police and an investigation has been initiated into the matter.

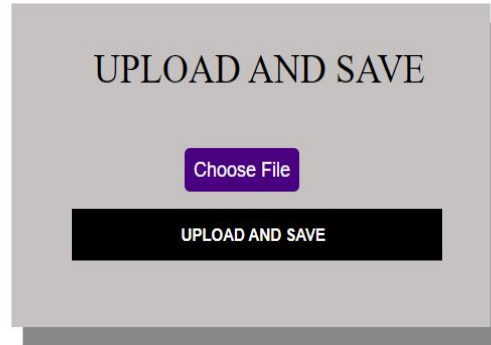
Vishwanath Sundi had been suspicious about his wife having an illicit affair with Shyamal Hembram, a young man from Segaisai village. On Friday night, his suspicion was proven right when he caught his wife with Hembram in a compromising position. Hembram had

স্ত্রীর বিবাহবহির্ভূত সম্পর্ক রয়েছে, সন্দেহ করতেন স্বামী। তাকে তাকে ছিলেন তিনি। শুক্রবার রাতে প্রেমিকের সঙ্গে স্ত্রীকে ঘনিষ্ঠ অবস্থায় দেখে ফেলেন তিনি। রাগের বশে প্রেমিককে প্রথমে মারধর করেন। তার পর গাছের সঙ্গে বেঁধে কুড়ুল দিয়ে মাথা কেটে ফেলেন বলে অভিযোগ। ঘটনাটি ঘটেছে বাড়ুখণ্ডের লাঞ্জা গ্রামে।

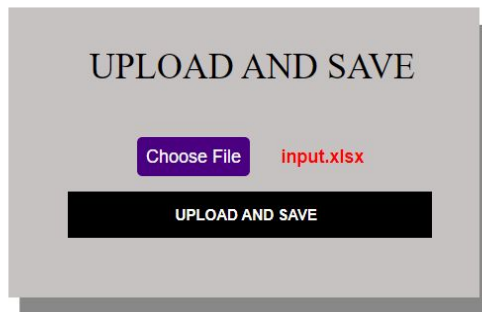
পুলিশ সূত্রে খবর, অভিযুক্তের নাম বিশ্বনাথ সুন্ডি। তাঁর স্ত্রীর সঙ্গে পাশের গ্রামের এক যুবকের বিবাহবহির্ভূত সম্পর্ক গড়ে উঠেছিল। স্ত্রীর অন্য কোনও সম্পর্ক রয়েছে, সন্দেহ করতেন বিশ্বনাথ।

Named Entity  
Recognition

Upload only .xlsx files in the format as shown in given [link!](#)



Upload only .xlsx files in the format as shown in given [link!](#)



	A	B
1	<b>Article1 (English)</b>	<b>Article2 (Regional Language)</b>
2	Delhi University's Hansraj College has stopped serving	দিল্লির হংসরাজ কলেজের হস্টেল ও ক্যান্টিনে আমিষ খাবার পরিবেশন
3	Vishwanath Sundi, a resident of Jharkhand's Lonjo village	দ্বীপ বিবাহবহির্ভূত সম্পর্ক রয়েছে, সন্দেহ করতেন স্বামী। তাকে তাকে ছিলেন
4	Foreign minister S Jaishankar said India's long-suffering	বার বার নিয়ন্ত্রণরেখায় (এলওসি) স্থিতিবস্থা নষ্ট করার চেষ্টা করেছে চীন।
5		
6		
7		
8		

Multilingual News Article Similarity

Upload only .xlsx files in the format as shown in g

UPLOAD AND SAVE

Choose File

UPLOAD AND SAVE

Downloads

What do you want to do with output.xlsx?

Open Save as

input.xlsx  
Removed

See more

	A	B	C	D	E
1	English	Regional Language	Translation	BERT	roBERTa
2	Delhi University's Hansraj C	দিল্লির হংসরাজ কলেজে	The hostels and canteens o	0.931448	0.833895
3	Vishwanath Sundi, a reside	স্ত্রীর বিবাহবাহির্ভূত সম্প	The husband suspected tha	0.945249	0.819937
4	Foreign minister S Jaishank	বার বার নিয়ন্ত্রণরেখায় (	China has repeatedly tried t	0.912754	0.755761
5					

Demo dataset16776400

Download

input16789269640706

Download

## Multilingual News Article Similarity

### Log In

**LOGIN**

[Create a new account](#)

## Multilingual News Article Similarity

### Sign Up

**Signup**

[Already have an account](#)

# References

- [1] Nikkei at SemEval-2022 Task 8: Exploring BERT-based Bi-Encoder Approach for Pairwise Multilingual News Article Similarity
- [2] SemEval-2022 Task 8: Multilingual news article similarity
- [3] The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents D Gunawan<sup>1</sup>, C A Sembiring and M A Budiman
- [4] Overview - CoreNLP ([stanfordnlp.github.io](https://stanfordnlp.github.io))
- [5] Microsoft Bing API translator
- [6] BERT Explained: A Complete Guide with Theory



Thank you!

