

# LEGAL ASSISTANCE SYSTEM FOR CRIMINAL VIOLENCE VICTIMS USING NATURAL LANGUAGE PROCESSING

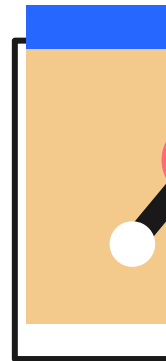
## **Guide:**

Dr. Gopika Rani N

Adarsh G	(20Z204)
Elanthamil R	(20Z215)
Jeevan Krishna K V	(20Z220)
Nirmal M	(20Z267)
Ajay Deepak P M	(21Z431)



# Table of contents

- Introduction
- Dataset
- Initial process
- BERT
- Working of BERT
- Hyper parameter Tuning
- Generating Word Embeddings
- L2 Normalization
- Cosine Similarity
- Streamlit Interface
- Results
- Conclusion and Future Work
- References





# Introduction


- Traditional legal processes often present challenges for individuals seeking assistance on appropriate legal actions, leading to confusion and inefficiencies.
  - Through the use of NLP, the system enables users to interact with it in a conversational manner, posing queries and receiving clear, concise responses.
  - Hence, the system strives to democratize legal knowledge, ensuring that everyone, regardless of their background, can navigate the intricacies of criminal violence cases with complete understanding.
- 
- 

# Literature Survey

Topic	Algorithm	Advantage	Disadvantage
LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors	Natural Language Processing (NLP) pipelines, including similarity score model (SIM), SIM×KEY, SIM×KEY×SYN	<ul style="list-style-type: none"><li>- High accuracy (88.24%) on testing datasets.</li><li>- Utilizes NLP techniques for legal guidance.</li><li>- Stratified 5-fold validation for robust evaluation.</li></ul>	<ul style="list-style-type: none"><li>- Limited to Thai Supreme Court cases related to sexual violence.</li><li>- Specific to legal guidance for sexual violence survivors.</li></ul>
ChEMBL Bot - A Chat Bot for ChEMBL database	Dialogflow for chatbot development	<ul style="list-style-type: none"><li>- Provides a conversational interface for accessing ChEMBL database resources.</li><li>- Utilizes natural language conversation for data retrieval.</li></ul>	<ul style="list-style-type: none"><li>- Dependency on external tools like Dialogflow.</li><li>- Limited to ChEMBL database context.</li></ul>

Research paper recommendation with topic analysis	Integrates topic analysis techniques into collaborative filtering methods	<ul style="list-style-type: none"><li>- Addresses the gap in finding relevant research papers.</li><li>- Incorporates thematic similarity based on topic analysis.</li></ul>	<ul style="list-style-type: none"><li>- May require significant computational resources for topic modeling.</li><li>- Effectiveness may depend on the quality of topic models.</li></ul>
A Movie Recommender System based on Topic Modeling using Machine Learning Methods	Latent Dirichlet Allocation (LDA) modeling for topic modeling	<ul style="list-style-type: none"><li>- Addresses the concern of inappropriate content for children and adolescents.</li><li>- Leverages LDA and age ratings for movie recommendations.</li></ul>	<ul style="list-style-type: none"><li>- Relies on the quality of age ratings and LDA modeling.</li><li>- Limited to movie recommendations.</li></ul>
Text-Based Recommendation System for E-Commerce Apparel Stores	Bag of Words (BoW), Inverse Document Frequency (IDF), Term Frequency-Inverse Document Frequency (TF-IDF) applied to product titles	<ul style="list-style-type: none"><li>- Utilizes three text-based similarity approaches for eCommerce apparel recommendations.</li><li>- Uses data from Amazon Product Advertising API.</li></ul>	<ul style="list-style-type: none"><li>- Limited to ladies' apparel recommendations.</li><li>- May face challenges with varied and evolving product titles.</li></ul>


# Dataset



## Indian Penal Code, 1860

IPC India, Act No. 45 OF 1860

[Home](#) [Prev](#) [Next](#)

Select Language 

You can ← S W I P E → on connected pages!

[IPC](#)

### Chapters in IPC

- 01 Introduction
- 02 General Explanations
- 03 Of Punishments
- 04 General Exceptions
- 05 Abetment
- 05a Criminal Conspiracy
- 06 Offences Against The State

# Dataset

	A1	A	B	
1			Section	Description
2	53		Punishments	The punishments to which offenders are liable under the provisions of this Code are:Death;Impris
3	53A		Construction of reference to transportation	Subject to the provisions of sub-section (2) and sub-section (3), any reference to "transportation f
4	54		Commutation of sentence of death	In every case in which sentence of death shall have been passed, the appropriate Government m
5	55		Commutation of sentence of imprisonment for life	In every case in which sentence of imprisonment for life shall have been passed, the appropriate
6	55A		Definition of "appropriate Government	In sections 54 and 55 the expression "appropriate Government" means,—in cases where the sen
7	56		Sentence of Europeans and Americans to penal servitude	Proviso as to sentence for term exceeding ten years but not for life – Rep. by the Criminal Law (F
8	57		Fractions of terms of punishment	In calculating fractions of terms of punishment, imprisonment for life shall be reckoned as equival
9	58		Offenders sentenced to transportation how dealt with until transported	Rep. by the Code of Criminal Procedure (Amendment) Act, 1955 (26 of 1955)
10	59		Transportation instead of imprisonment	Rep. by the Code of Criminal Procedure (Amendment) Act, 1955 (26 of 1955)
11	60		Sentence may be (in certain cases of imprisonment) wholly or partly rigorous or si	In every case in which an offender is punishable with imprisonment which may be of either descri
12	61		Sentence of forfeiture of property	Rep. by the Indian Penal Code (Amendment) Act. 1921 (16 of 1921)
13	62		Forfeiture of property in respect of offenders punishable with death, transportation	Rep. by the Indian Penal Code (Amendment) Act. 1921 (16 of 1921) section 4
14	63		Amount of fine	Where no sum is expressed to which a fine may extend, the amount of fine to which the offender
15	64		Sentence of imprisonment for non-payment of fine	In every case, of an offence punishable with imprisonment as well as fine, in which the offender is
16	65		Limit to imprisonment for non-payment of fine, when imprisonment and fine awarded	The term for which the court directs the offender to be imprisoned in default of payment of a fine s
17	66		Description of imprisonment for non-payment of fine.	The imprisonment which the Court imposes in default of payment of a fine may be of any descrip
18	67		Imprisonment for non-payment of fine, when offence punishable with fine only	If the offence be punishable with fine only, the imprisonment which the Court imposes in default o
19	68		Imprisonment to terminate on payment of fine	The imprisonment which is imposed in default of payment of a fine shall terminate whenever that
20	69		Termination of imprisonment on payment of proportional part of fine	If, before the expiration of the term of imprisonment fixed in default of payment, such a proportio
21	70		Fine leviable within six years, or during imprisonment – Death not to discharge pro	The fine, or any part thereof which remains unpaid, may be levied at any time within six years aft
22	71		Limit of punishment of offence made up of several offences	Where anything which is an offence is made up of parts, any of which parts is itself an offence, th



# Initial process

- The initial process of our legal assistance system involves obtaining user input and performing a series of essential steps to ensure accurate and effective assistance.
  - Tokenization
  - Semantic Analysis
  - Error Correction
  - POS Tagging
  - Vectorization

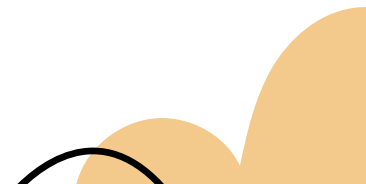


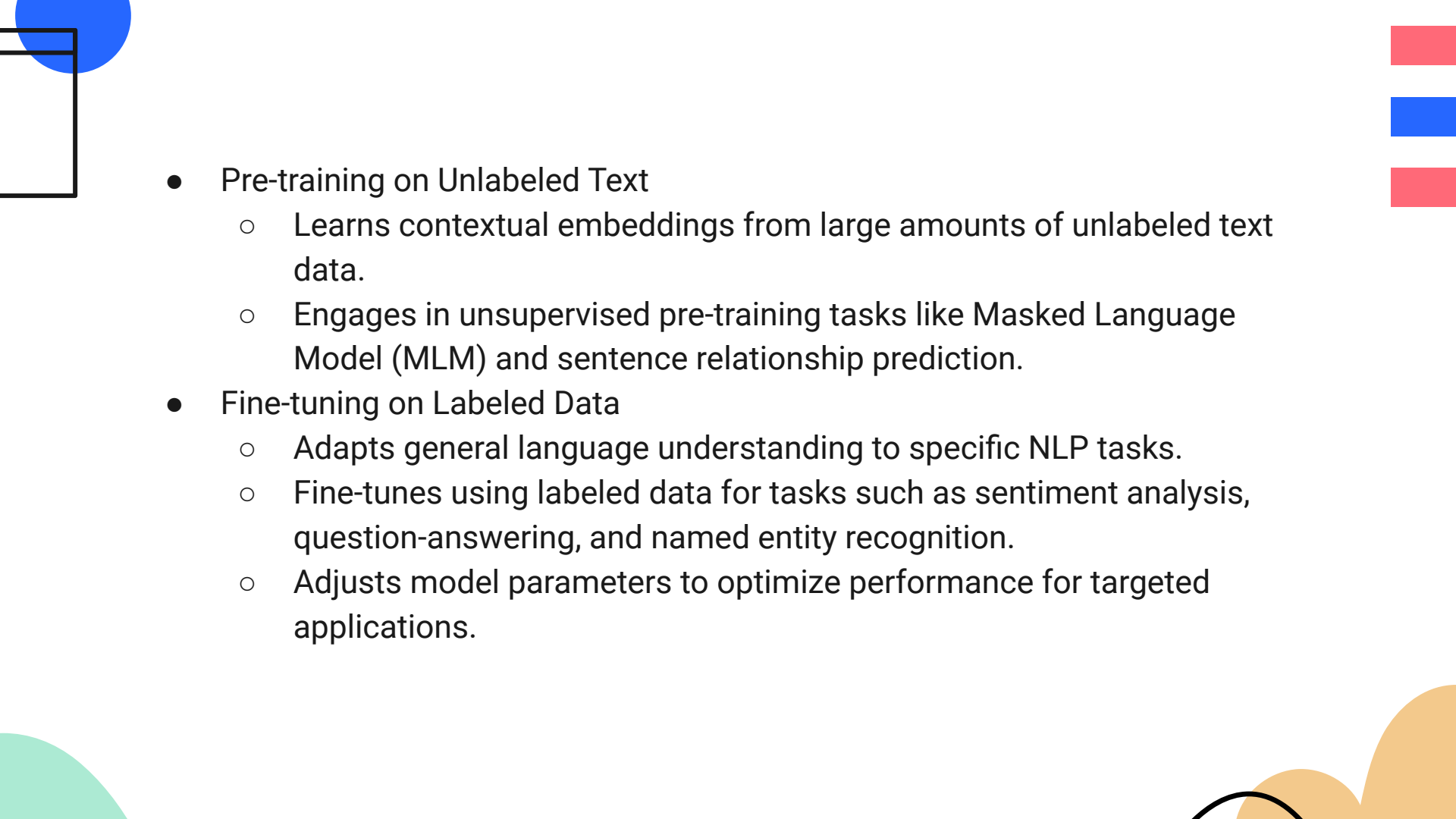


# BERT

- BERT stands for Bidirectional Encoder Representations from Transformers.
- Introduced by Google researchers in 2018.
- Revolutionized natural language processing (NLP) by enhancing contextual understanding.


## Key features

- Bi-directional approach: Considers both left and right contexts simultaneously.
  - Captures deeper contextual relationships within sentences.
- 

- 
- Pre-training on Unlabeled Text
    - Learns contextual embeddings from large amounts of unlabeled text data.
    - Engages in unsupervised pre-training tasks like Masked Language Model (MLM) and sentence relationship prediction.
  - Fine-tuning on Labeled Data
    - Adapts general language understanding to specific NLP tasks.
    - Fine-tunes using labeled data for tasks such as sentiment analysis, question-answering, and named entity recognition.
    - Adjusts model parameters to optimize performance for targeted applications.





# Working of BERT

- BERT is designed for generating a language model, utilizing only the encoder mechanism of the Transformer.
  - Input sequences of tokens are fed into the Transformer encoder.
  - Tokens are first converted into vectors and then processed within the neural network.
  - Output consists of contextualized representations for each input token.
- 

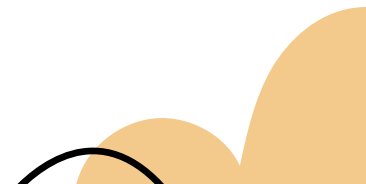


# Masked Language Model

- In pre-training, a portion of words in input sequences is masked.
  - BERT is trained to predict the original values of these masked words based on surrounding context.
  - Process involves masking words (about 15% of input) and replacing them with [MASK] symbols.
  - BERT guesses the hidden words by considering the surrounding context.
  - A classification layer is added atop the encoder output to predict masked words.
- 
- 




# Next Sentence Prediction

- BERT predicts if the second sentence follows the first in a pair.
  - [CLS] token output is transformed into a 2×1 shaped vector using a classification layer.
  - Probability of the second sentence following the first is calculated using SoftMax.
  - During training, BERT learns relationships between pairs of sentences.
  - 50% of input pairs have the second sentence following the first in the original document.
  - Input processing involves inserting [CLS] token at the start of the first sentence, [SEP] token at the end of each sentence, and adding sentence and positional embeddings.
- 



# Hyper parameter Tuning




- Hyperparameter tuning involves selecting the optimal set of hyperparameters for a machine learning model to maximize its performance on a specific dataset.
  - Hidden layers within the framework of the transformer architecture are designated for pre-training purposes.
  - The number of hidden layers in the BERT model is increased from 12 to 24.
  - The BERT model is structured such that each transformer encoder layer incorporates numerous attention heads.
  - The number of attention layers in the BERT model is increased from 12 to 24.
- 

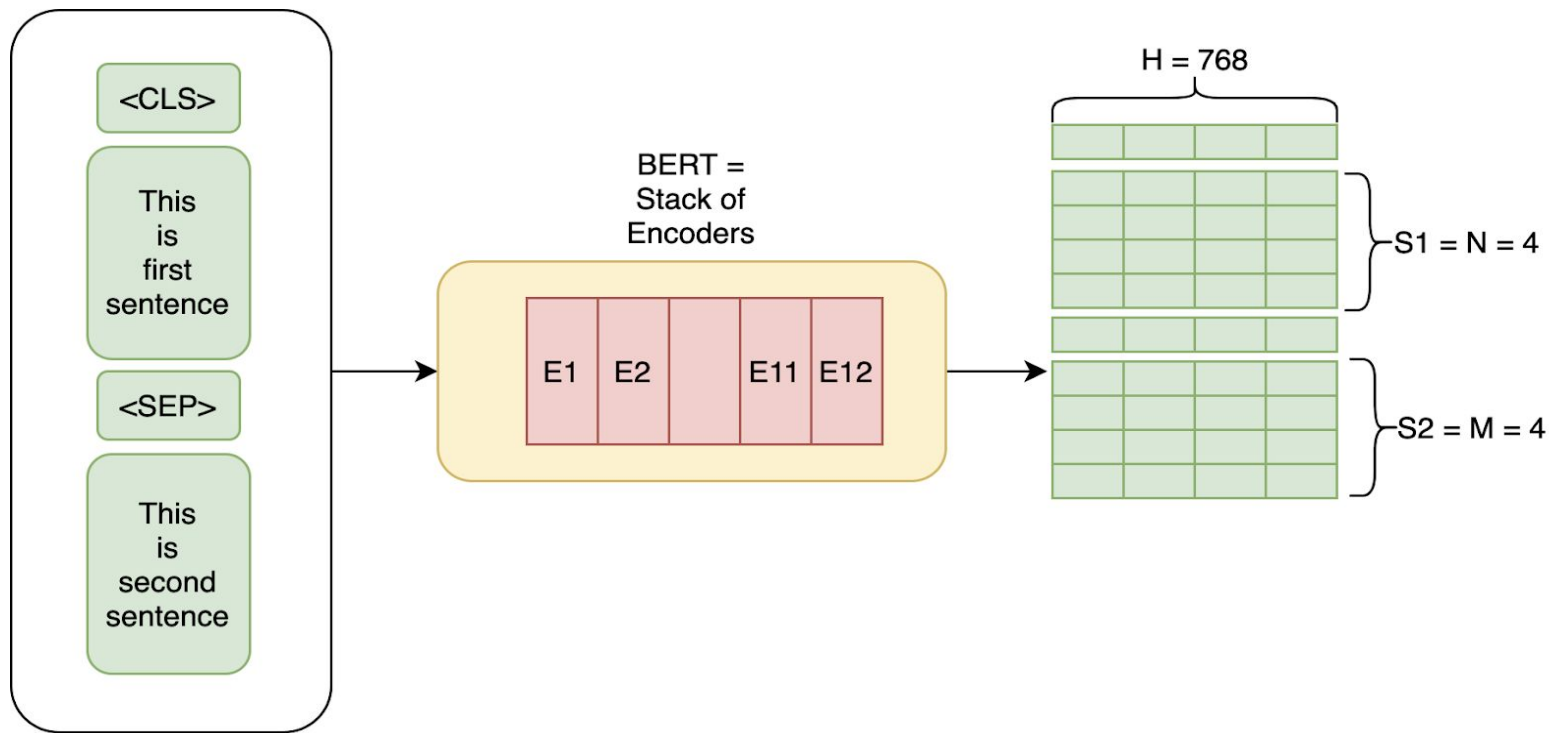


# Generating Word Embeddings



- BERT model is used to generate embeddings for all words in a sentence based on context.
  - This approach ensures BERT's tolerance to homonyms (same words with different meanings).
  - Word embeddings are extracted from the last layer of the BERT model.
  - Embeddings consist of 768 fields, providing unique representations for words based on sentence context.
  - Sentence embeddings are generated by averaging along each of the 768 fields of all word embeddings in a sentence.
  - This process ensures that each section has a unique vector representation, leveraging BERT's contextual understanding.
- 

# Generating Word Embeddings





# L2 Normalization

- The magnitude of the sentence embedding increases with the number of words in a sentence or section.
- This variation in magnitude can overshadow the uniqueness and context of the sentence.
- L2 Normalization ensures that the magnitude of vectors is equal to 1, regardless of sentence length.

$$x_{norm} = \frac{x}{\|x\|_2}$$

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_m^2}$$



# Cosine Similarity

- Cosine similarity is utilized to measure the similarity between user input and section data vectors.
- This method quantifies the similarity between two vectors by calculating the cosine of the angle between them.
- Cosine similarity provides a numerical score indicating the degree of similarity between vectors.



# Streamlit Interface

- Streamlit is a widely-used Python library for creating interactive web applications with ease and speed.

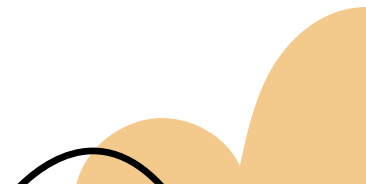
## Welcome to Legal Assistance System (LAS)

Your First Stop for any LEGAL advise

Hii !!!!! I am LAS short for Legal Advisory System. State the incident or problem you're facing. I will help you with a solution.

o Snatching - Section Number: 379A

Whoever, with the intention to commit theft, suddenly or quickly or forcibly seizes or secures or grabs or takes away from any person or from his possession any moveable property, and makes or attempts to make escape with such property, is said to commit snatching.



# Streamlit Interface

Hii !!!!! I am LAS short for Legal Advisory System. State the incident or problem you're facing. I will help you with a solution.

o Snatching - Section Number: 379A

Whoever, with the intention to commit theft, suddenly or quickly or forcibly seizes or secures or grabs or takes away from any person or from his possession any moveable property, and makes or attempts to make escape with such property, is said to commit snatching.

o Act causing slight harm - Section Number: 95

Nothing is an offence by reason that it causes, or that it is intended to cause, or that it is known to be likely to cause, any harm, if that harm is so slight that no person of ordinary sense and temper would complain of such harm.

o Assault - Section Number: 351

Whoever makes any gesture, or any preparation intending or knowing it to be likely that such gesture or preparation will cause any person present to apprehend that he who makes that gesture or preparation is about to use criminal force to that person, is said to commit an assault. Explanations Mere words do not amount to an assault. But the words which a person uses may give to his gestures or preparation such a meaning as may make those gestures or preparations amount to an assault.

# Streamlit Interface

Whoever makes any gesture, or any preparation intending or knowing it to be likely that such gesture or preparation will cause any person present to apprehend that he who makes that gesture or preparation is about to use criminal force to that person, is said to commit an assault. Explanations Mere words do not amount to an assault. But the words which a person uses may give to his gestures or preparation such a meaning as may make those gestures or preparations amount to an assault.

o Exclusion of acts which are offences independently of harm cause - Section Number: 91

The exceptions in sections 87, 88 and 89 do not extend to acts which are offences independently of any harm which they may cause, or be intended to cause, or be known to be likely to cause, to the person giving the consent, or on whose behalf the consent is given.

o Keeping lottery office - Section Number: 294A

Injuring or defiling place of worship, with intent to insult the religion of any class

Enter Response

I was attacked by a group of people. They threatened me to sell my prop

Enter

# Results

- The Legal Assistance System relies on a tuned BERT model with L2 normalization to provide accurate recommendations for user incidents.
- To measure the system's efficiency, the TOP-10 method is employed. This method calculates the accuracy of the model by averaging the related recommendations among the top 10 suggestions for 30 user incidents.


Model	Avg Top - 5 Score
BERT	0.16
BERT with L2	0.32
Hyperparameter tuning	0.24
Tuned BERT with L2 (proposed model)	0.56



# Results



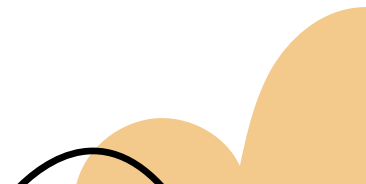
Model	Avg Top - 10 Score
BERT	0.12
BERT with L2	0.46
Hyperparameter tuning	0.32
Tuned BERT with L2 (proposed model)	0.62





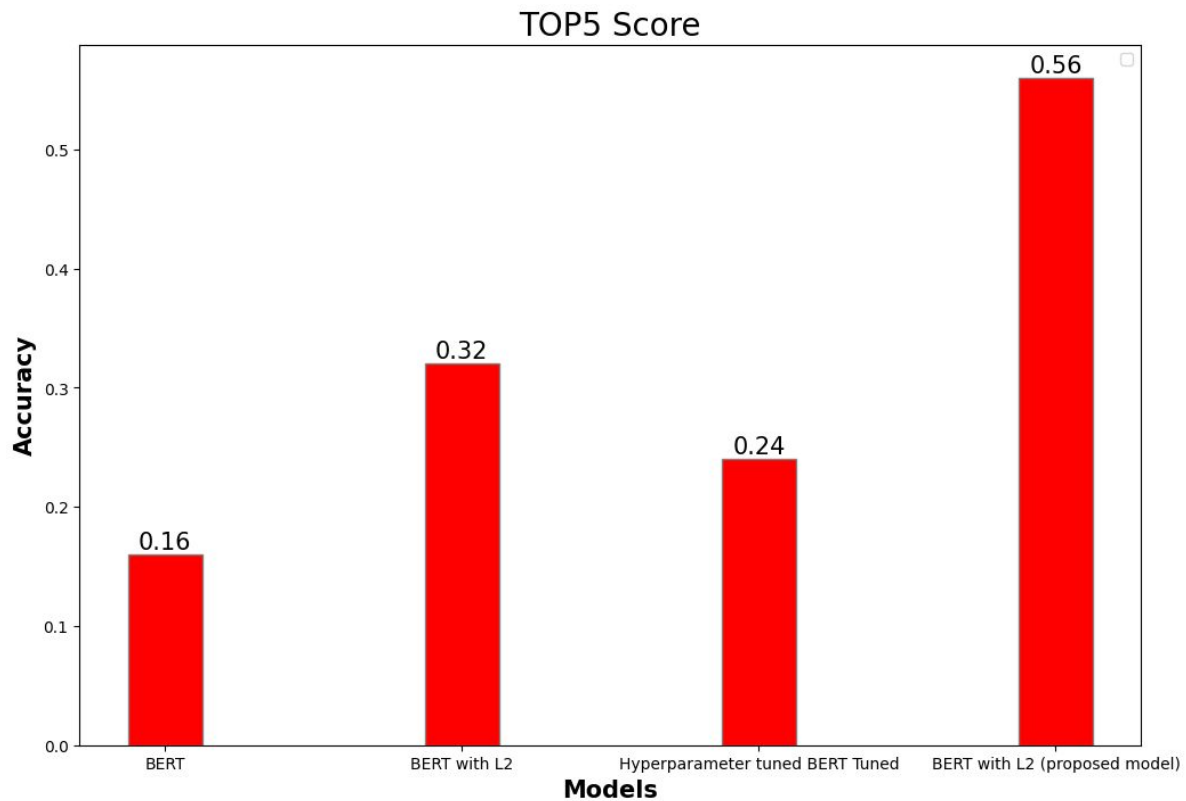
# Results

Some of the User Incident Queries used for the TOP - 10 testing are listed below:

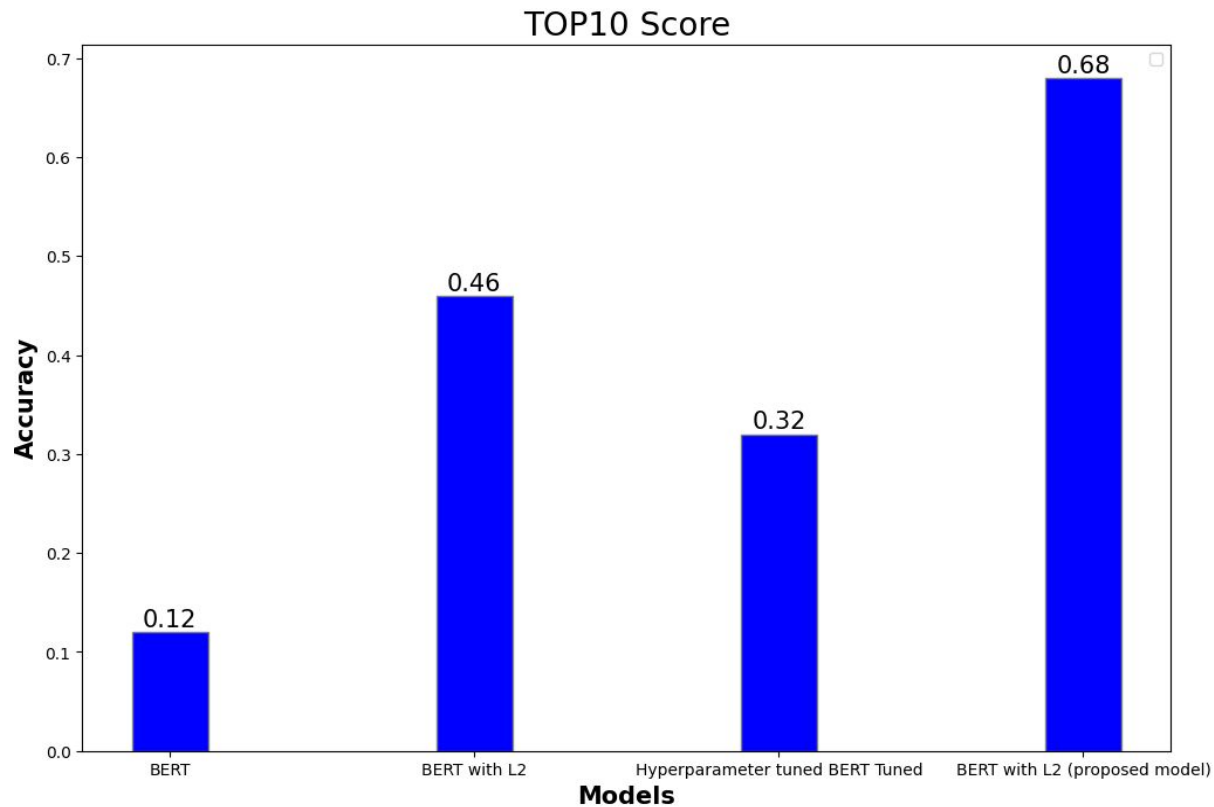
1. A group of individuals entered my house to rob, and during the robbery, they assaulted me.
  2. A government official misuses his position and stole public funds
  3. An individual forges documents to secure a loan from a bank
  4. During a religious procession, some individuals entered a store, assaulted the owner, and stole valuable items
  5. A candidate during an election campaign spreads false information about the opponent to sway voters
- 



# Results




# Results





# Conclusion & Future work




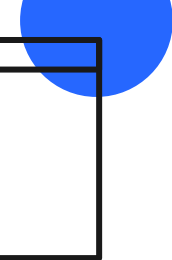

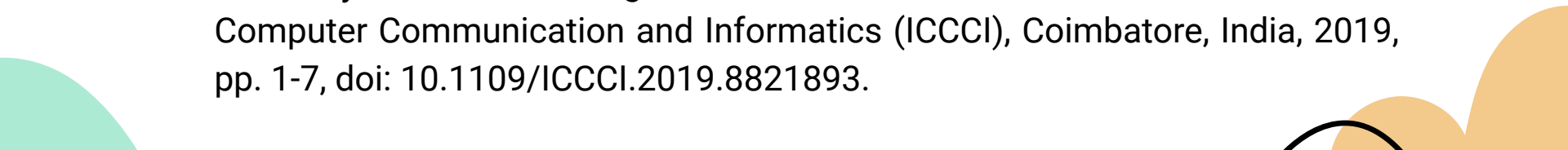
- The legal assistance system represents a solution to the challenges individuals face when seeking legal guidance through NLP
  - The system's ongoing learning and development guarantee that it stays current with the most recent legal norms and practices
  - By offering personalized guidance, the system empowers individuals to make informed decisions and take appropriate legal actions, contributing to enhanced efficiency in the legal process.
  - The efficiency of the model can be increased by using the BERT embeddings to train an ML model which is trained on a dataset that contains user incident embeddings and corresponding Penal Code embeddings
- 

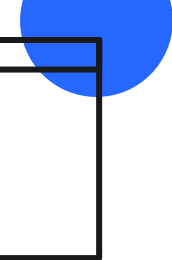

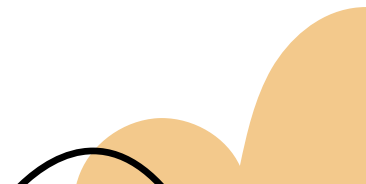


# References



- V. Socationurak et al., "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors," in IEEE Access, vol. 9, pp. 131440-131461, 2021, doi: 10.1109/ACCESS.2021.3113172.
  - V. Murali, R. J. Sarma, P. A. Sukanya, and P. Athri, "ChEMBL Bot - A Chat Bot for ChEMBL database," 2018 Fourteenth International Conference on Information Processing (ICINPRO), Bangalore, India, 2018, pp. 1-6, doi: 10.1109/ICINPRO43533.2018.9096710.
  - Chenguang Pan and Wenxin Li, "Research paper recommendation with topic analysis," 2010 International Conference On Computer Design and Applications, Qinhuangdao, China, 2010, pp. V4-264-V4-268, doi: 10.1109/ICCDA.2010.5541170.
- 

- 
- 
- Chenguang Pan and Wenxin Li, "Research paper recommendation with topic analysis," 2010 International Conference On Computer Design and Applications, Qinhuangdao, China, 2010, pp. V4-264-V4-268, doi: 10.1109/ICCDA.2010.5541170.
  - M. Kordabadi, A. Nazari, and M. Mansoorizadeh, "A Movie Recommender System based on Topic Modeling using Machine Learning Methods," International Journal of Web Research, vol.5, no.2, pp.19 -28, 2022, doi: 10.22133/ijwr.2022.370251.1139.
  - U. C. De, B. B. Dash, T. M. Behera, T. Samant and S. Banerjee, "Text-Based Recommendation System for E-Commerce Apparel Stores," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-4, doi: 10.1109/ICONAT57137.2023.10080436.
  - R. Shrivastava and D. S. Sisodia, "Product Recommendations Using Textual Similarity Based Learning Models," 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICCCI.2019.8821893.
- 

- 
- 
- S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," in IEEE Access, vol. 7, pp. 7011-7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
  - Tyagi, Nemika & Bhushan, Bharat. (2023). Natural Language Processing (NLP) Based Innovations for Smart Healthcare Applications in Healthcare 4.0. 10.1007/978-3-031-22922-0\_5.
  - H. Liu et al., "A Natural Language Processing Pipeline of Chinese Free-Text Radiology Reports for Liver Cancer Diagnosis," in IEEE Access, vol. 8, pp. 159110-159119, 2020, doi: 10.1109/ACCESS.2020.3020138.
  - M. Yang, Z. Meng and I. King, "FeatureNorm: L2 Feature Normalization for Dynamic Graph Embedding," 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 731-740, doi: 10.1109/ICDM50108.2020.00082.
- 

# Thank You!

