INDIAN PENAL CODE SECTION IDENTIFICATION USING SEQUENTIAL NEURAL NETWORKS

MINI PROJECT REPORT

Submitted by

SRIMATHI G	**210701259**
SRIMATHY R	**210701260**

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

ANNA UNIVERSITY:: CHENNAI 600 025

APRIL 2024

RAJALAKSHMI ENGINEERING COLLEGE,

CHENNAI

BONAFIDE CERTIFICATE

Certified that this Report titled "**Indian Penal Code Section Identification using Sequential Neural Networks**" is the bonafide work of **"SRIMATHI G (210701259), SRIMATHY R (210701260)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr.V. KARTHICK,M.E.,Ph.D.,
Assosiate Professor,
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai – 602015

Submitted to Mini Project Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

ABSTRACT

Identifying the Indian penal code for the crime committed requires in-depth knowledge in crime as well as remembering Indian penal code (IPC) sections which have high chances of human error.So the aim is to build a prediction model using feed forward neural network or Sequential neural network helps in giving judgements or the appropriate Indian Penal Code sections for the crime committed using the keywords of the crime which benefits everyone related to the case and can get an insight of the prediction over judgment. This model uses text analysis which involves tokenization regularization techniques and building a neural network using the preprocessed set. The output provides the IPC sections based on the keyword entered as input. The Sequential neural network (a type of deep learning) used here helps in finding the pattern between the keywords and tokenized words from preprocessed data from the dataset which contains punishment, offense, Indian penal code for the respective crimes and thereby building a precise recognition system. The same system using NLP provided an accuracy score of 67.66% but the proposed system using sequential neural networks provided an overall accuracy score of 78.333% which makes it more potential for commoners to identify sections under which complaints can be raised.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time.

We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr.V. Karthick,M.E.,Ph.D.,** Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

Srimathi G -210701259
Srimathy R -210701260

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SNN	Sequential Neural Network
FIR	First Information Report
IPC	Indian Penal Code

CHAPTER 1
INTRODUCTION

1.1 GENERAL

Traditional crime classification and assignment of appropriate IPC section are time consuming and prone to human error .Hence our system is built to overcome the above challenge by providing an accurate Indian Penal Code (IPC) section trained using a dataset which contains an official comprehensive framework of criminal offenses and their corresponding punishment using Sequential neural networks.

1.2 OBJECTIVE

The main aim of the system is to help commoners identify the IPC section for the crime using plain text format .

1.3 EXISTING SYSTEM

Existing System involves using natural language processing and incorporates techniques such as tokenization and stemming for accuracy to identify the crime section using the keywords in the FIR .The one disadvantage associated with this approach is the potential leakage of sensitive information contained within the FIR.

1.4 PROPOSED SYSTEM

Proposed system uses an additional layer of Sequential neural network .Our system also includes Preprocessing[9] techniques to ready text data for analysis by a Sequential neural network (SNN).These steps include tokenization, lowercasing, punctuation removal, stopwords elimination, stemming or lemmatization, padding, and vectorization. These preprocessing steps ensure data cleanliness, standardization, and suitability for SNN input, enhancing accuracy in classifying legal descriptions. Sequential neural network (SNN) enhances accuracy by leveraging its ability to recognize patterns in text data. It understands the hierarchical features and local

connectivity within the data, making them well-suited for tasks like identifying relevant sections of the Indian Penal Code (IPC) and associated punishments based on the nature of the offense. The input will be the description of the crime. Our system works with preprocessed data and gives you back the relevant section of the law and what punishment might be involved. It helps in quickly finding the right rules and consequences for any given situation and making the legal stuff easier to understand.

CHAPTER 2
LITERATURE SURVEY

On studying the paper found how machine learning and its techniques were incorporated in identification of Indian penal code sections for the crime using the keywords in FIR. To implement the model techniques of preprocessing like tokenization and stemming were used to increase the accuracy and to train the model natural language processing is defined to map the keywords with crime associated in the dataset.

A novel Particle Swarm-Cuckoo Search (PS-CS) optimization algorithm[3] is proposed to enhance the optimization of network parameters and enable deep neural network training for crime prediction. The PS-CS model surpasses traditional backpropagation algorithms in terms of convergence speed . The study evaluates the proposed PS-CS methodology against prevalent classification models, including conventional machine learning[4] techniques and advanced deep learning models.

The criminal offenses lead to certain punishment according to the Indian Penal Code (IPC). For particular crimes, particular sections are assigned to punish the criminal or convicts with jail terms and fines. On these pre-processed data sets, by applying a Naïve Bayesian algorithm[5] predict the crime type in reference to Indian penal code.

Rabia Musheer Aziz, Prajwal Sharma, Aftab Hussain et al, This paper presents a data-driven approach using regression models, notably random forest regression, to analyze Indian crime data and predict regional crime occurrences.

Prakash Aryan et al, This author uses the multiclass classification model that can determine the applicable Indian Penal Code according to the description provided by the user.

Ambrish Srivastav, Shaligram Prajapat et al, This research study focuses on the development and investigation of a Decision Support System (DSS) for Indian Penal Code (IPC) Sections using a combination of similarity algorithm and fuzzy logic.

Kavita Shirsat1,, Aditya Keni2,, Pooja Chavan3,, Manasi Gosavi4 et al,In this paper, we analyze the basic description of the case, and apply a machine learning model to predict the possible IPC Section that will be applicable based on the fact of the case

Dipti Pawade, Avani Sakhapara, Hussain Ratlamwala et al ,The author suggests using NLP with Word Mover's Distance to bridge the gap between citizens' lack of awareness and the legal system, accurately assigning IPC sections to crime descriptions and validating results with lawyers for accuracy assessment.

Rabia Musheer Aziz, Aftab Hussain, Prajwal Sharma et al , The study presents a new PS-CS algorithm for crime prediction, achieving 99.87% accuracy and offering a robust solution for controlling crime rates in India.

Stanley Yeo et al ,This paper highlights the enduring appeal of the Indian Penal Code (IPC) while acknowledging its shortcomings, advocating for courts to prioritize solutions rooted in the IPC's principles and legal history before turning to common law.

Data mining method is used for extracting knowledge from large datasets, particularly in fields like criminology, where understanding crime characteristics and relationships is vital. This paper discusses employing data mining techniques to extract meaningful information from police narrative reports, aiding law enforcement agencies in automatically inputting crime data into databases. Additionally, the paper applies Self-Organizing Maps (SOM)[6] clustering for crime analysis, utilizing the clustering for crime analysis, utilizing the clustering results for crime matching processes. Overall, the study demonstrates the utility of data mining[7] in enhancing crime analysis and supporting police forces with valuable insights gleaned from complex datasets.

CHAPTER 3
SYSTEM DESIGN

3.1 DEVELOPMENT ENVIRONMENT

3.1.1 HARDWARE SPECIFICATIONS

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

Table 3.1.1 Hardware Specifications

PROCESSOR	Intel Core i5
RAM	4GB or above (DDR4 RAM)
GPU	Intel Integrated Graphics
HARD DISK	6GB
PROCESSOR FREQUENCY	1.5 GHz or above

3.1.2 SOFTWARE SPECIFICATIONS

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre-installed and the languages needed to develop the project has been listed out below.

Table 3.1.2 Software Specifications

SOFTWARES USED	Google colab / Jupyter Notebook , Tensor Flow

3.2 SYSTEM DESIGN

3.2.1 ARCHITECTURE DIAGRAM

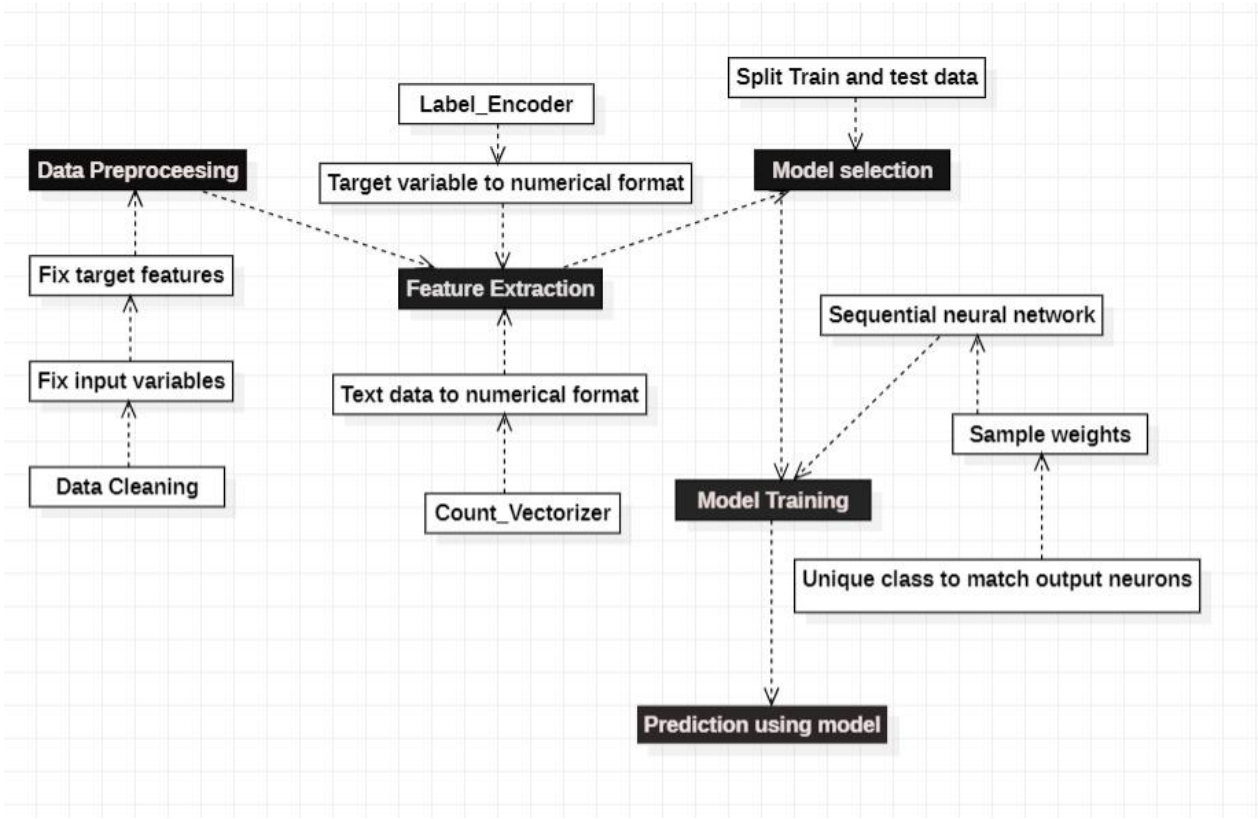

Fig 3.2.1 Architecture Diagram

CHAPTER 4
PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

4.1.1Data Collection:

Gathering information about crimes or offenses, along with corresponding Indian Penal Code (IPC) sections and punishments, from various legal sources for dataset. Dataset also includes a detailed description of crime.

4.1.2Data Preprocessing:

The collected data undergoes preprocessing steps to clean and standardize it for analysis. Textual descriptions are converted to lowercase, and punctuation is removed to ensure consistency. Text data is tokenized, splitting it into individual words or tokens for further processing. Stemming or lemmatization is applied to reduce words to their root forms, enhancing text uniformity. Text features, such as 'Offense' and 'Punishment', are combined into a single cohesive feature to provide comprehensive context.

4.1.3Text vectorization:

The preprocessed text data is converted into numerical vectors using Count Vectorizer. This conversion process allows the neural network to process text data effectively by representing words as numerical values .

4.1.4Model Training:

The vectorized text data is used to train a Sequential neural network model . During training, the model adjusts its internal parameters iteratively to minimize a defined loss function. Dropout layers are included to prevent overfitting, ensuring the model generalizes well to unseen data. The model is trained using a portion of the data, while performance is monitored using a separate validation set.

4.1.5Model Evaluation:

Once training is complete, the model's performance is evaluated using a separate test dataset. Performance metrics such as accuracy, precision, and recall are calculated to assess the model's effectiveness in classifying legal descriptions.

4.1.6Prediction and deployment:

The trained model is deployed to make predictions on new, unseen textual descriptions of offenses. Users input descriptions of crimes, and the model predicts the relevant IPC sections and associated punishments. These predictions provide valuable insights for legal interpretation and decision-making processes.

CHAPTER 5
IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

The result would be a user-friendly tool accessible to the general public. When a user enters keywords like "killed" or "robbery," the system would promptly display the corresponding sections of the Indian Penal Code (IPC) related to those keywords.

This empowers common people to quickly understand the legal implications of various crimes without needing specialized legal knowledge. Essentially, this project serves as a valuable resource for anyone seeking clarity on the legal aspects of specific actions described in simple terms they can understand.

```
# Extract relevant columns for input features and target variable
X = df[['Offense', 'Punishment']].astype(str).values  # Convert to string
y = df['Description'].values  # Target variable

# Convert labels to numerical format using LabelEncoder
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(y)

# Number of unique classes in your dataset
output_neurons = len(np.unique(y))
# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Combine 'Offense' and 'Punishment' into a single text feature
X_text_train = X_train[:, 0] + ' ' + X_train[:, 1]
X_text_test = X_test[:, 0] + ' ' + X_test[:, 1]

# Use CountVectorizer to convert text data to numerical format
vectorizer = CountVectorizer()
X_train_numerical = vectorizer.fit_transform(X_text_train)
X_test_numerical = vectorizer.transform(X_text_test)

# Compute sample weights for balancing classes
sample_weights = compute_sample_weight('balanced', y_train)

# Define the neural network model with adjustments
model = Sequential()
model.add(Dense(128, input_dim=X_train_numerical.shape[1], activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))
```

```
Epoch 857/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.7902 - accuracy: 0.7600WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 35ms/step - loss: 0.8548 - accuracy: 0.7394 - val_loss: 23.4388 - val_accuracy: 0.0000e+00
Epoch 858/1000
1/3 [=========>....................] - ETA: 0s - loss: 1.0120 - accuracy: 0.7200WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 36ms/step - loss: 0.8285 - accuracy: 0.7535 - val_loss: 23.4455 - val_accuracy: 0.0000e+00
Epoch 859/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.7691 - accuracy: 0.7700WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 42ms/step - loss: 0.8724 - accuracy: 0.7606 - val_loss: 23.4583 - val_accuracy: 0.0000e+00
Epoch 860/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.9790 - accuracy: 0.7200WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 34ms/step - loss: 0.8346 - accuracy: 0.7570 - val_loss: 23.4630 - val_accuracy: 0.0000e+00
Epoch 861/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.6898 - accuracy: 0.7900WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 41ms/step - loss: 0.8610 - accuracy: 0.7641 - val_loss: 23.4667 - val_accuracy: 0.0000e+00
Epoch 862/1000
1/3 [=========>....................] - ETA: 0s - loss: 1.0823 - accuracy: 0.6900WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 42ms/step - loss: 0.8621 - accuracy: 0.7430 - val_loss: 23.4435 - val_accuracy: 0.0000e+00
Epoch 863/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.8119 - accuracy: 0.7900WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 38ms/step - loss: 0.8238 - accuracy: 0.7852 - val_loss: 23.4403 - val_accuracy: 0.0000e+00
Epoch 864/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.6988 - accuracy: 0.7900WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
3/3 [==============================] - 0s 48ms/step - loss: 0.8001 - accuracy: 0.7676 - val_loss: 23.4421 - val_accuracy: 0.0000e+00
Epoch 865/1000
1/3 [=========>....................] - ETA: 0s - loss: 0.7656 - accuracy: 0.8200WARNING:tensorflow:`evaluate()` received a value for `sample_weight`,
```

5.2 OUTPUT SCREENSHOTS

```
# Your specific input
input_text = "illegal weapon"
```

```
1/1 [==============================] - 0s 33ms/step
['Description of IPC Section 148\n According to section 148 of Indian penal code, Whoever is guilty of rioting, being armed with a deadly weapon or with anything which, used as a weapon of offence, is likely to cause death, shall be punished with imprisonment of either description for a term which may extend to three years, or with fine, or with both.\n\n\n IPC 148 in Simple Words\n If someone takes part in a riot while being armed with a deadly weapon or anything likely to cause death, they can be punished with up to three years in jail, a fine, or both.']
```

```
# Your specific input
input_text = "illegal selling"
```

```
1/1 [==============================] - 0s 49ms/step
['Description of IPC Section 238\n According to section 238 of Indian penal code, Whoever imports into India, or exports there from, any counterfeit coin, which he knows or has reason to believe to be a counterfeit of Indian coin, shall be punished with Imprisonment for life, or with imprisonment of either description for a term which may extend to ten years, and shall also be liable to fine.\n\n\n IPC 238 in Simple Words\n In simple words, Section 238 of the Indian Penal Code states that if someone knowingly imports or exports any counterfeit Indian coin, they can be punished with either life imprisonment or imprisonment up to ten years, and fined.']
```

CHAPTER 6
CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

Thus Indian Penal Code (IPC) section identification using Sequential neural networks was built to ease the task of human involvement in manually providing a judgment and the commoners can acknowledge whether the justice for the victim and punishment for the criminal is under article 14 that all are equal before law .

6.2 FUTURE ENHANCEMENTS

The future of identifying IPC sections using Sequential Neural Networks (SNNs) promises significant advancements in crime detection[11] and law enforcement. As SNN architectures and training techniques continue to evolve, accuracy in IPC section identification from textual data is poised to improve, streamlining legal procedures and reducing manual effort. Multimodal approaches may combine text with other data sources for richer analysis, enabling real-time crime detection and swift law enforcement response. Customization to regional variations and integration with legal databases will further enhance effectiveness in diverse contexts, aiding decision-making processes for legal professionals. Overall, SNN-based IPC section identification holds great potential in terms of accuracy (78.33%) which is comparatively higher value when trained by other NLP techniques, thus enhancing public safety and justice administration.

.

REFERENCES

[1] M. R. Bigler and O. Baum, "Deep learning-based classification of the capillary ultrastructure in human skeletal muscles," *Front Mol Biosci*, vol. 11, p. 1363384, May 2024.

[2] K. D. Gaur, *Commentary on the Indian Penal Code: As Amended by the Criminal Law (Amendment) Act, 2018 (Act No. 22 of 2018)*. 2019.

[3] N. Dey, *Applications of Cuckoo Search Algorithm and its Variants*. Springer Nature, 2020.

[4] V. Chandra S. S. and A. Hareendran S., *MACHINE LEARNING: A PRACTITIONER'S APPROACH*. PHI Learning Pvt. Ltd., 2021.

[5] F. Sabry, *Naive Bayes Classifier: Fundamentals and Applications*. One Billion Knowledgeable, 2023.

[6] J. I. Mwasiagi, *Self Organizing Maps: Applications and Novel Algorithm Design*. BoD – Books on Demand, 2011.

[7] M. J. Zaki and W. Meira Jr, *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press, 2020.

[8] M. Podda *et al.*, "Classification of genomes with a bag-of-words approach and machine learning," *iScience*, vol. 27, no. 3, p. 109257, Mar. 2024.

[9] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer, 2014.

[10] S. Vajjala, B. Majumder, A. Gupta, and H. Surana, *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly Media, 2020.