An Analysis of Natural Language Text Relating to Thai Criminal Law

Weerayut Krungklang and Sukree Sinthupinyo

July 3, 2020

An Analysis of Natural Language Text Relating to Thai Criminal Law

Weerayut Krungklang
Department of Computer Engineering
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
6070971521@student.chula.ac.th

Sukree Sinthupinyo
Department of Computer Engineering
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
Sukree.S@chula.ac.th

***Abstract*—This paper analyses Thailand's criminal law enforcement in chapter 1, Offenses causing death section category section 288 and 289 of title 10 offenses affecting life and body under the Thai Criminal Code. The first part of this paper is using criminal law domain knowledge and supreme court judgment results, to be the initial domain information and result is the rules that humans can understand. The second part of this paper is bringing training data set from the final judgment to train with deep learning methods. Due to the training set which have severe imbalances, the Synthetic Minority Over-Sampling TEchnique (SMOTE) [1] is used to solve this problem. Models are trained on the training set using unidirectional Long Short-Term Memory (LSTM) [2] networks and bidirectional Long Short-Term Memory (BiLSTM) [3] are type of Recurrent Neural Networks (RNN) [2]. The word embeddings of the dataset can be learned while training a deep neural network. BiLSTM average F1 score is higher than LSTM. Pre-trained word embeddings are then used to make the average F1 score higher than before. Finally, using models to predict online crime news, the highest average probability of each model is selected by using Soft Voting as input to the rules. The test results compared with the predictions of our methods with the opinion of the lawyer, corresponding 76%.**

Keywords—Criminal law, Thai Supreme Court, Word embedding, Word2Vec, LSTM, BiLSTM, SMOTE, Pre-trained word embeddings, Deepcut, Decision tree, Soft Voting.

I. INTRODUCTION

Criminal laws maintain social harmony by punishing criminals. Consequently, an offense and the punishment are defined by the law [4]. Interpretations according to criminal law must be conducted strictly in accordance with the provisions. When the criminal offense occurs, the person related to the case may receive punishment advice to be made aware of the preliminary sentence. However, searching for legal information via the internet may be incorrect, fail to meet the objective, and generally just be a waste time. According to this limitation background, this paper is meant to provide an analysis of Natural Language Text relating to Thai Criminal Law. This will help to reduce misinterpretation from the law section, make it easy to access information, and to get correct and accurate results which can be referenced from reliable sources.

This paper is divided into two main parts. The first part is to extract rules from the decision tree using the CART algorithm for interpreting legal reasons to humans. The second part is a model for natural language processing through deep learning.

II. RELATED WORKS

A. Research Related to Thai Criminal Law

P. Osathitporn, N. Soonthornphisaj, and W. Vatanawood [4] proposed an acquisition schema of the Thai criminal code by using ontology. The result is a criminal structure of knowledge through ontology that uses content from the Thai Criminal Code to construct the SWRL rules. The Thai Criminal Code is divided into three parts. Part 1: General provisions, which determines criminal liability by focusing on the action of the individual which will be used for said offenses. Accordingly, knowledge base and related rules of each law element will be created. Part 2: Specific offenses will be provisions relating to twelve different offenses. The provisions of Title 10, Category 1; Crimes Against Life will be applied in two sections: namely section 288 and section 289. Part 3 is not in the scope of this paper. This research will focus on the liability structure identified in [4] as the main structure for building a knowledge base in criminal law. The liability structure can be divided into three main structures: actions are considered complete fault components, there is no law except for the offense, and the act has no law except for the punishment. The authors use the ontology mentioned above as a criminal liability structure in our paper.

S.Thammaboosadee, B.Watanapa and N.Charoenkitkarn [5] propose a framework multi-stage classifier for applying the facts to the fault components in each case. They then specify relevant laws according to the fault components. Finally, they determine offenses of criminal penalties in the offenses relating to life and body. This research is divided into three levels: fact level, case level, and legal level using an ensemble model of Artificial Neural Network and C4.5 decision tree, which is then mapped to the legal charge codes. This research uses dataset, which is selected from 150 criminal judgments. However, using data that has already been encrypted is different from our paper which uses input data as free text and 2,698 criminal judgments from the Supreme Court of Thailand Judgment Search System [6]. By filtering the categories of the criminal code, section 288 and 289 from 2500 B.C. to 2560 B.C. are used as raw data. These consist of section 288 in the amount of 2,078 and section 289 in the amount of 620 final judgments.

K. Kowsrihawat, P. Vateekul and P. Boonkwan [7] proposed predicting judicial decisions of criminal cases from the Thai Supreme Court using bidirectional GRU with an attention mechanism. Data entered to test the model are facts from the criminal case that occurred, and are relevant matters of law for their consideration. Using the TSCC data set (Thai Supreme Court cases) consists of judgments and laws which are classified into three categories including life and body, reputation, and property. These models give two results. The act of fact is either an offense or not according to the law specified. This research uses deep learning, which is akin to our paper. However, it is different in that the input data does not require any relevant laws in any way, but use only the facts as input and focus only on life's offenses.

B. *Techniques Involved in Research*

- Synthetic Minority Over-Sampling TEchniques [1]. Increases the number of minority class members by creating synthetic samples from the minor class. Since the training dataset of this paper have a relatively small amount of training set collected, it is necessary to use the sampling method to increase the minority of the class to equal the number of the majority class. This will solve the problem of the imbalanced class instead of creating copies in the training set.

- Long Short-Term Memory (LSTM) network [2]. When word order is important, RNN is used instead of the traditional Bag of Words approach. However, RNNs suffer from the problem of vanishing gradients that cannot deal with infinitely long recurrent and need to stop remembering. LSTM is one of the RNN [2] models that can learn sequential data and capture the context information effectively. Models are trained using Back-Propagation Through Time and have memory cells which are connected into layers to overcome the mentioned vanishing gradient problem.

- Bi-directional LSTM [3]. The concept of LSTM reads the context from left to right only, but in a very complex sentence may not be effective. Therefore, another LSTM that reads the context from right to left is used. The contexts are then concatenated as the input vector to solve the problem of forward information lacking less accuracy than the information behind.

- Extracting judgment rules [8]. This technique is used to classify factors that affect criminal liability and are extracted into rules. The authors then apply the 210 verdicts about life extracted into sub-data of each feature according to the criminal liability structure. Then extract the rules from the relationship of each tree's attribute to the form of "IF-THEN" to make it easier to interpret. Humans can understand cause and effect by using the CART algorithm [9] to classify punishment groups. Together with the Gini index is the classification criteria for each node. In this paper, the punishment is divided into seven groups. It can then be written as a trial rule with seven rules, namely A-G group. This will present the important characteristics which will determine the classification of the punishment. There are only four characteristics, which are intention, justification, causation and impunity.

- Word embedding. The process of representing words as dense vectors to convert words into vectors. This can also specify the dimensions as needed. In this paper, the 300 dimensions vector for representing each word is used. Word2Vec algorithm is one technique to learn word embedding. It was developed by Tomas Mikolov et al., in 2013 [10]. There are two main methods to implement the Word2Vec algorithm, including the Continuous Bag of Words model (CBOW) and the Skip-gram model. This paper used Skip-gram [11] which works well with a small amount of training set and rare words or phrases.

- Pre-trained word embeddings. Instead of training the embedding layer, this paper uses pre-trained word embeddings that has been trained in a large corpus. It saves time in training by using Fasttext that is an extension to Word2Vec proposed by Facebook in 2016 [12] and Thai2Vec [13] which is trained with Thai-Wikipedia data by ULMFit method in 300-dimension. In addition, to improving the efficiency of the model, the authors use the Gensim library implementation of the Word2Vec Technique for training 504 crime news (Crime2Vec).

- Regular Expression [14]. Before bringing the data into practice, the data must be cleaned first, such as the verbiage of personal names as well as specific locations which are not needed.

- Beautiful Soup4 [15]. Extracted text from HTML tags bring judgments to analysis to create judgment rules.

- Soft Voting [16]. The technique uses classifiers in different ways, but uses the same training data to select the highest average of each model passing inputs to judgment rule.

III. METHODOLOGY

This paper proposes the process of considering the facts from crime news that the offender must be punished with criminal penalties under the seven punishment groups. Overview of creating prediction rules, text classification models, and using models to predict crime news are shown in Figures 1,2 and 3.

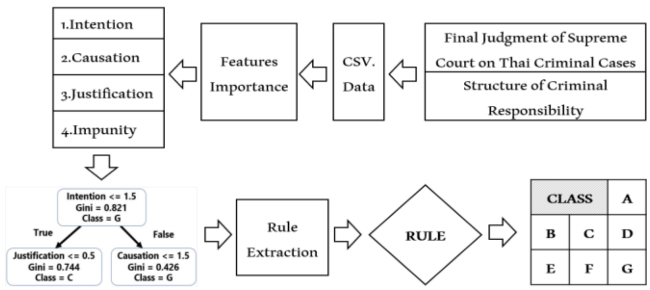

Fig. 1. An overview of rule model

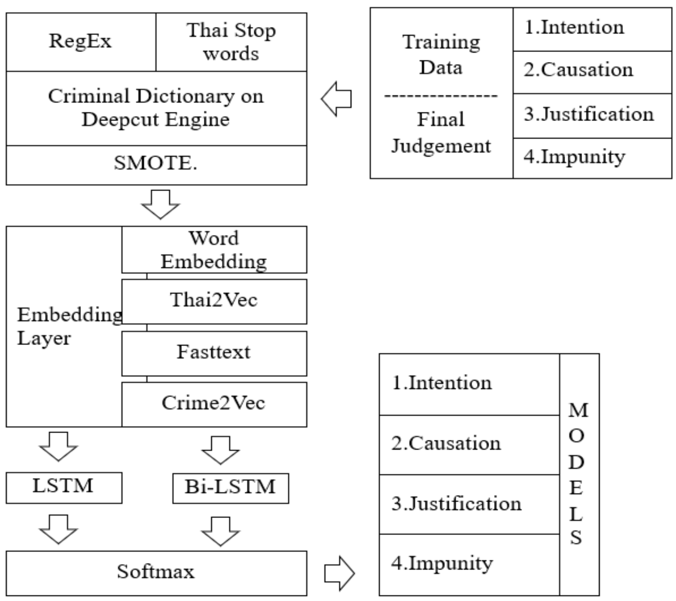

Fig. 2. The judgments model structure

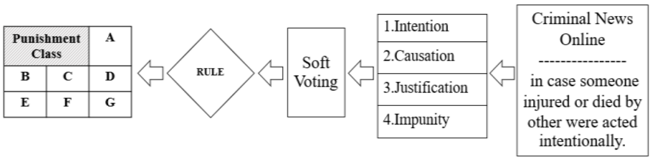

Fig. 3. Testing a model on crime news

A. *Creating Judgment Rule*

1) Data Cleansing: This section used Beautiful Soup4 with a regular expression to extract the Supreme Court verdicts in each case by searching the structure of HTML tags. The required attributes are the number of jurisdictions, and verdict of the Supreme Court consisting of facts, legal issues, and adjudication. Also used are related sections from the website of the Supreme Court [6] by filtering the categories of the criminal code sections 288 and 289, searching from 2500 B.C. to 2560 B.C. These are used as raw data by a total of 2,698 petitions, which consist of section 288 in the amount of 2,078 and section 289 in the amount of 620, as shown in Figure 4.

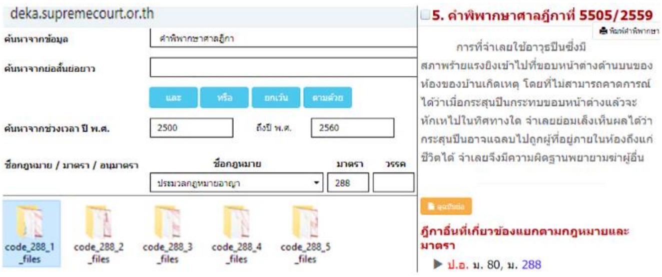

Fig. 4. Process of gathering final judgment related to criminal law

2) Data Preprocessing: The process analyses the content of judgment in each case. This determines the facts that constitute the element of criminal liability structure designed according to the ontology.

เมื่อพิจารณาพฤติการณ์ที่ จำเลยกับพวกตระเตรียมอาวุธมีดท่อนเหล็ก และท่อนไม้ ไว้เพื่อหาโอกาสทำร้ายผู้ตาย ด้วยสาเหตุที่มีเรื่องบาดหมางไม่พอใจกันมาก่อน เมื่อสบโอกาสจำเลยกับพวกก็ ร่วมกันใช้อาวุธดังกล่าวฟันและทำร้ายผู้ตาย ย่อมแสดงว่า จำเลยกับพวกมีเจตนาฆ่าผู้ตายโดยไตร่ตรองไว้ก่อน

Fig. 5. Extraction of facts from the final judgment

Figure 5 shows the offender had an altercation with the victim, in which he immediately attacked with a weapon. This premeditated attack displays planned elements which results in the victim's death. The offender is liable in group G, which is related to joint enterprise and premeditated killing cases. The authors used this process with every petition and news story with a total of 210 cases. This results in a CSV file containing 210 data rows and separating the properties that affect the consideration of penalties in a total of 7 columns. It is divided into seven groups of penalties, which are A – G as following: A is no offence being committed, B is attempt to cause death but there are mitigating circumstances, C is attempt to cause death, D is Murder but there are mitigating circumstances, E is Murder, F is Capital Murder but there are mitigating circumstances and G is Capital Murder aligned with Thailand's court punishment level. As shows in Table I.

TABLE I. CLASSIFICATION OF PUNISHMENT GROUP A - G

Actor Status	Victim	Causation	Justification	Impunity	Behavior	Intention	Punishment
Self	People	Injury	Not wrong	Except	Kill	None	A
Sup port	People	Injury	Wrong	Decrease	Kill	Regular	B
Self	People	Death	Wrong	Decrease	Kill	Regular	C
Self	People	Nothing	Wrong	Normal	Kill	Regular	D
Joint	People	Death	Wrong	Normal	Kill	Regular	E
Joint	People	Nothing	Wrong	Increase	Kill	Special	F
Joint	People	Death	Wrong	Increase	Kill	Special	G

The resulting CSV file is classified using the decision tree. The authors chose to use the CART algorithm [8] because it is a binary classification job that gives two target values. The true result will be on the left-hand side, and the false result will be on the right-hand side, according to the division conditions of the Gini index. All properties will be calculated, and when a decision tree is created, it is necessary to change from text to number in order with the Python language function to be suitable for processing. At this stage, the authors can find qualifications that affect the classification of penalties by finding the properties that are important for classification as shown in Figure 6 below.

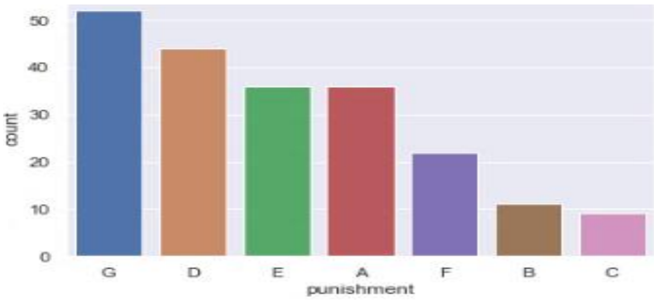

Fig. 6. Number of classes that identify the penalty group

3) Decision Tree Model Building [17]: The information is classified from the judgment sample according to the punishment group by the CART algorithm. This has a binary structure which devides each node into two groups based on true-false values. The true value is on the left node, and the false value is on the right node. The data is divided into two branches, training data and testing data with a ratio of 70:30. The Gini index is used to divide the characteristics according to (1). The evaluation result can be up to 100% by dividing each node. Next, the model is constructed and evaluated with cross-validation (5-fold) [18] to solve the problem of over-fitting which gives 99.48% accuracy. GridSearchCV is used to fine-tune the parameters, increasing the accuracy value to 99.5%.

$$I_G = \sum_{i=1}^{c} p(i) * \left(1 - p(i)\right) \qquad (1)$$

In (1) represents Gini impunity, where *C* is the number of classes, and *p*(*i*) is the probability of randomly picking an element of class *i*.

An important attribute for use in modeling obtained from tuning parameters with scikit-learn libraries is then selected. The result is four important features that affect accuracy consisting of intention, justification, causation, and impunity, as shown in Figure 7.

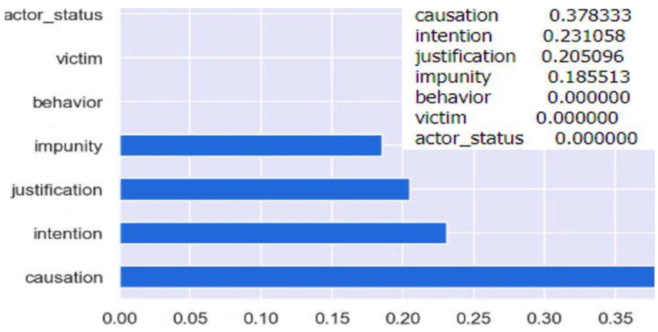

Fig. 7. Important features that affect the criminal trial

The decision tree graph [19] in considering punishments for human interpretations then displayed, as shown in Figure 8.

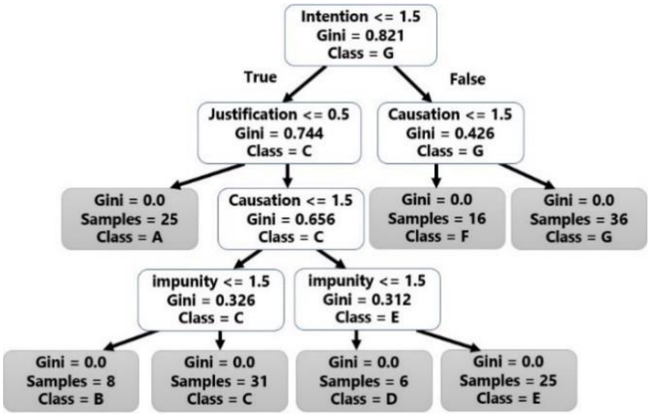

Fig. 8. Decision tree graph

4) Rule Extraction: The process for extraction of rules from decision trees according to the previous clause can be classified into seven rules. The authors created the IF-THEN function to make it easier to understand, as shown in Figure 9.

Rule 1: IF intention <=1.5 AND justification <=0.5 THEN Class = A

Rule 2: IF intention <=1.5 AND justification > 0.5 AND causation <= 1.5 AND impunity <= 1.5 THEN Class = B

Rule 3: IF intention <=1.5 AND justification > 0.5 AND causation > 1.5 AND impunity <= 1.5 THEN Class = C

Rule 4: IF intention <=1.5 AND justification > 0.5 AND causation <= 1.5 AND impunity > 1.5 THEN Class = D

Rule 5: IF intention <=1.5 AND justification > 0.5 AND causation > 1.5 AND impunity > 1.5 THEN Class = E

Rule 6: IF intention >1.5 AND causation <= 1.5 THEN Class = F

Rule 7: IF intention >1.5 AND causation > 1.5 THEN Class = G

Fig. 9. Rule extraction from decision trees

B. *Natural Language Classifier*

1) Data cleansing: The authors use regular expressions with Deepcut [20], which is the Thai word segmentation wrapping tool. The tool has an F1 score of 98.1% from the best data set of NECTEC 2009. The words will then be handled continuously, and the text will not affect consideration, or vocabulary building. ThaiStopWords will delete the information, as shown in Table II.

TABLE II. CLEAND DATA

Original Data	Cleaned Data
~~เมื่อวันที่ 5 มี.ค. พ.ต.อ.สัญญา พาตินธุ ผกก.สภ. ปากน้ำหลังสวน จ.ชุมพร~~ พร้อมด้วย ~~ร.ต.อ.วิฑูรย์ เผือกวารี รอง สว.(สอบสวน)~~ เจ้าของคดี นำกำลังควบคุมตัว ~~นายชู หรือนายตะวัน~~ แรงงานต่างด้าวชาวมอญ ไปทำแผนประกอบคำรับสารภาพภายในสวนปาล์มพื้นที่~~หมู่1ต.บางน้ำจืด อ.หลังสวน~~ หลัง~~นายชู~~ก่อเหตุข่มขืนแล้วฆ่ายาย~~วัย 72 ปี~~ โดยมีบรรดาทางญาติผู้ตายเดินทางมาดูหน้า~~นายชู~~เป็นจำนวนมาก	พร้อมด้วยเจ้าของคดีนำกำลังควบคุมตัวแรงงานต่างด้าวชาวมอญไปทำแผนประกอบคำรับสารภาพภายในสวนปาล์มพื้นที่หลังก่อเหตุข่มขืนแล้วฆ่ายายโดยมีบรรดาทางญาติผู้ตายเดินทางมาดูหน้าเป็นจำนวนมาก

2) Data Preprocessing: All four important attributes are then labeled by hand. Clean data is then used to label class assignments, according to each attribute as the data set seen in Table III. Then the authors have added a criminal custom dictionary to the word segmentation in legal domains. This was done using the Deepcut engine for creating word-level features and handling out-of-vocabulary, as shown in Table IV. Training data set in this paper is imbalanced. This problem is addressed by synthesizing the minor samples class to increase the size of every class except for the major class. Using SMOTE to increase the small amount of data in every class to be equal to the largest class, as shown in Table V. Finally, the authors have separated the data into 2 sets. The first set is for training and the second is for a test set with a ratio of 80:20. During training the model, the training set for data validation 10% are separated. This is for the accuracy rate validation and tuning model parameters. According to this ratio the results show an accuracy rate higher than the others.

TABLE III. INTENTION ATTRIBUTE WHICH IS A DATASET

Categories	Part of the final judgment content
none	จำเลยพาพ่อไปหาหมอ ผู้ตายพูดว่าสงสัยใกล้ตายแล้ว
regular	ใช้อาวุธปืนยิงบริเวณห้องนอน
regular	ใช้มีดทำครัวที่พกติดตัวมาฟันจนล้มลง และใช้มีดฟันซ้ำ
special	วางแผนและตระเตรียมอาวุธไปพร้อม
special	จำเลยขับรถขวาง พวกของจำเลยที่ซ้อนท้ายชักอาวุธปืนออกมา

TABLE IV. THAI WORD SEGMENTATION

<table>
<tr><th>Cleaned Data</th><th>Word Segmentation</th></tr>
<tr><td>พร้อมด้วยเจ้าของคดีนำกำลังควบคุมตัวแรงงานต่างด้าวชาวมอญไปทำแผนประกอบคำรับสารภาพภายในสวนปาล์มพื้นที่หลังก่อเหตุข่มขืนแล้วฆ่ายายโดยมีบรรดาทางญาติผู้ตายเดินทางมาดูหน้าเป็นจำนวนมาก</td><td>พร้อม|ด้วย|เจ้าของ|คดี|นำกำลัง|ควบคุมตัว|แรงงาน|ต่างด้าว|ชาวมอญ|ไป|ทำแผน|ประกอบคำรับสารภาพ|ภายใน|สวนปาล์ม|พื้นที่|หลังก่อเหตุ|ข่มขืน|แล้วฆ่า|ยาย|โดย|มี|บรรดา|ทาง|ญาติ|ผู้ตาย|เดินทาง|มา|ดู|หน้า|เป็น|จำนวนมาก</td></tr>
</table>

TABLE V. THE NUMBER OF SAMPLES FOR EACH ATTRIBUTE

Attributes	Categories	Number of Example	Original Dataset			Balanced Dataset (SMOTE)		
			Training set	*Validation set*	*Test set*	*Training set*	*Validation set*	*Test set*
Intention	Regular	261	188	21	52	188	21	52
	Special	168	121	13	34	188	21	52
	None	134	96	11	27	188	21	52
Causation	Death	214	154	17	43	154	17	43
	Injury	114	82	9	23	154	17	43
	Nothing	22	16	2	4	154	17	43
Impunity	Normal	145	104	12	29	104	12	29
	Except	139	100	11	28	104	12	29
	Increase	97	70	8	19	104	12	29
	Decrease	84	60	7	17	104	12	29
Justification	Wrong	142	102	11	28	102	11	28
	Not wrong	134	96	11	27	102	11	28

3) Pre-trained Word Embeddings: This paper uses the Gensim library in Word2Vec with 2,676 unique words from 504 crime news. The authors set the following key parameters: sg = 1 (Skip-gram), size = 300 (The number of dimensions in each represented word), Window = 5 (The maximum context location at which the words need to be predicted, meaning five words behind and five words ahead). These parameters are related to the language model. After training the model on the crime dataset, it became Crime2Vec for use as pretrained weight embedding. The t-SNE graph of eleven words crime similar in two dimensions is shown in Figure 10.

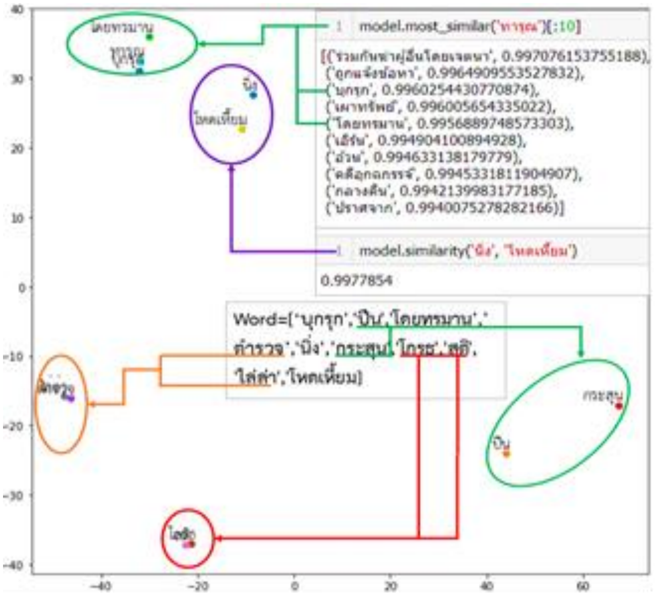

Fig. 10. Crime2Vec using Skip-gram in two-dimensional space

4) Define the LSTM and BiLSTM model: The Keras deep learning library on training data is used to fit all models. By using a word embedding from final judgment learning and train Word2Vec embedding with crime news. In addition, the embedding in a very large corpus (e.g., Fasttext with 2,000,000 embeddings and Thai2Vec with 60,002 embeddings) is pre-trained using the language model from ThaiWikiPedia and Word2Vec on 504 crime news in 300-dimension (2,676 embeddings) called Crime2Vec to initialize the first embedding layer.

C. Evaluation Models

This section uses the confusion matrix [21] to describe the performance of the multi-classification model, e.g., intention model. This shows that the BiLSTM+Fasttext feature achieves the highest F1 score and highest accuracy when compared to other features as shown in Figure 11. The accuracy and F1 score are shown in Table VI.

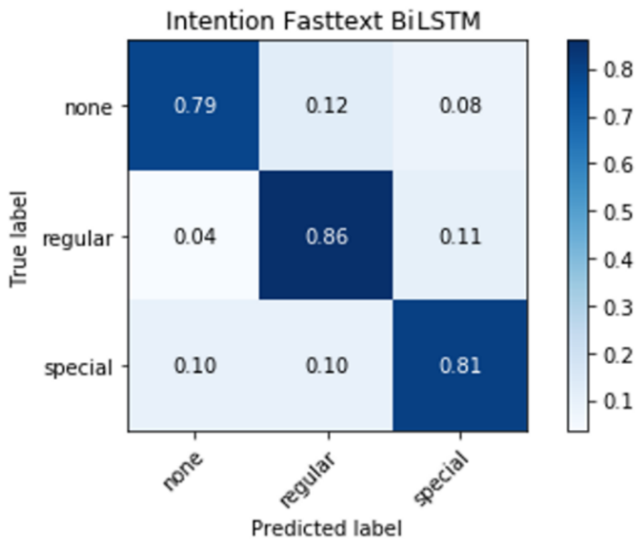

Fig. 11. Confusion matrix of intention model

TABLE VI. THE PERFORMANCE OF ALL MODELS BY FEATURES

Features	Models							
	Intention		*Causation*		*Impunity*		*Justification*	
	A	*F*	*A*	*F*	*A*	*F*	*A*	*F*
LSTM	78.98	78.88	62.02	62.40	75.86	75.98	80.70	80.68
LSTM+ Fasttext	78.98	78.47	75.97	75.50	73.28	72.63	94.74	94.71
LSTM+ Thai2Vec	68.79	68.75	62.02	60.69	33.62	33.40	78.95	78.94
LSTM+ Crime2Vec	75.16	75.01	71.32	71.32	68.97	68.79	92.98	92.87
BiLSTM	81.53	81.09	75.97	75.79	76.72	76.48	91.23	91.13
BiLSTM+ Fasttext	82.17	82.08	77.52	76.91	78.45	78.46	98.25	98.24
BiLSTM+ Thai2Vec	66.88	67.36	74.42	71.30	54.31	52.01	87.72	87.58
BiLSTM+ Crime2Vec	75.80	75.79	76.74	75.81	76.72	76.33	98.25	98.24

[a] Note. A = (% Accuracy), F = (% F1 Score macro avg)

IV. RESULT AND DISCUSSION

Table VI shows the accuracy and F1 score of all four models. Three of the models (Intention, Causation, and Impunity) are multi-classifications, whereas the Justification model is binary-classifications, combining BiLSTM and Fasttext feature. These models achieve the highest F1 score of 82.08%, 76.91%, 78.46%, and 98.24%, respectively, and reach the highest accuracy of 82.17%, 77.52%, 78.45% and 98.25% on the test set, respectively. Even though Crime2vec has the smallest number of words, this feature provides slightly less accurate than Fasttext, but higher accuracy than Thai2Vec. However, after considering the confusion matrix table, the authors cannot use only one model. Therefore, the authors evaluate all models by using 100 online crime news from the Daily News website. The authors can then predict the group of punishment and use Soft Voting to find the highest average probability value of each model according to (2). Comparison between our models and legal experts found that it matches 76% of lawyers' expectations. The average probability of each model is used as input to the rule, creating punishment group, prediction as shown in Figure 12.

$$\gamma = argmax\left(\frac{1}{n}\sum_{Classifier}(p_1, p_2, p_3, \ldots p_n)\right) \quad (2)$$

Equation (2) is a Soft Voting function. Where γ is the value of input for each model rule, n is the number of classifiers, and p is the probability from Softmax.

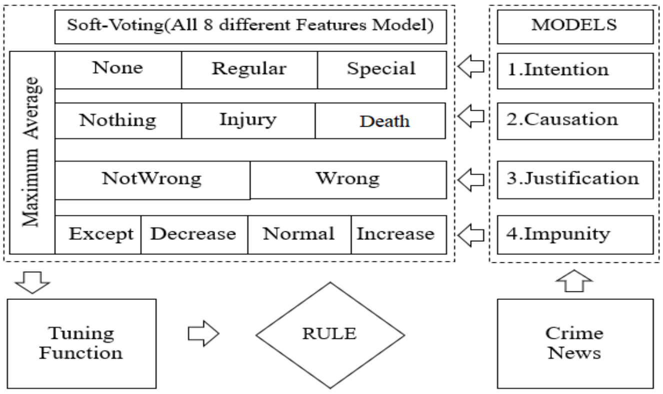

Fig. 12. The Soft Voting method

V. CONCLUSTION

In this paper, the authors have designed the model into two parts. The first part is the rule which was created from the decision tree. The second part is a deep learning classification with LSTM and BiLSTM. In addition, using pre-trained weight embedding (Fasttext, Thai2Vec, and Crime2Vec) to classify crime news content from which group of facts were punished within a specific scope. In the model training, the data is used according to the judgment of the Supreme Court and the penal code. In this paper, crime news from online sources was tested with all four models. Soft Voting is then used to select the maximum average representation of each model for entering as input for the rule created. The final evaluation is done by comparing the opinion of the lawyer. It is found that our prediction methods correspond to 76% of lawyers' opinions. In the future, the authors will bring the Transformer model applied with a corpus in criminal law and also increase the scope of the paper to cover carelessness until causing death to others.

REFERENCES

[1] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer.“SMOTE Synthetic Minority Over-Sampling Technique.” Journal of Artificial Intelligent Research, pp.321-357, 2002.

[2] Sherstinsky, Alex. (2018). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network.

[3] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019.

[4] P. Osathitporn, N. Soonthornphisaj and W. Vatanawood, "A scheme of criminal law knowledge acquisition using ontology," 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, 2017, pp. 29-34.

[5] Thammaboosadee, Sotarat & Watanapa, Bunthit & Charoenkitkarn, Nipon. (2012). A Framework of Multi-Stage Classifier for Identifying Criminal Law Sentences. Procedia Computer Science. 13. 53–59. 10.1016/j.procs.2012.09.113.

[6] Information Technology and Communication Center in Thai Supreme Court. “Thai Supreme Court Judgement Search System,” [Online]. Available: http://deka.supremecourt.or.th/

[7] K. Kowsrihawat, P. Vateekul and P. Boonkwan, "Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism," 2018 5th Asian Conference on Defense Technology (ACDT), Hanoi, 2018, pp. 50-55.

[8] López, Griselda & De Oña, Juan & Joaquín, Abellán. (2012). Using Decision Trees to Extract Decision Rules from Police Reports on Road Accidents. Procedia - Social and Behavioral Sciences. 53. 106–114. 10.1016/j.sbspro.2012.09.864.

[9] Breiman, L, Friedman, J H, Olshen, R A, and Stone, C J, 1984, Classification and regression trees: Wadsworth, Inc.

[10] Mikolov, Tomas & Corrado, G.s & Chen, Kai & Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space. 1-12.

[11] Mikolov, Tomas & Sutskever, Ilya & Chen, Kai & Corrado, G.s & Dean, Jeffrey. (2013). Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems. 26.

[12] Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of Tricks for Efficient Text Classification. ArXiv, abs/1607.01759.

[13] State-of-the-Art language modeling and text classification in Thai language. Available from: https://github.com/cstorm125/thai2fit.

[14] Fabian Beck, Stefan Gulan, Benjamin Biegel, Sebastian Baltes, and Daniel Weiskopf. 2014. RegViz: visual debugging of regular expressions. In Companion Proceedings of the 36th International Conference on Software Engineering. Association for Computing Machinery, New York, NY, USA, 504–507.

[15] Pratiksha Ashiwal, S.R.Tandan, Priyanka Tripathi, Rohit Miri. Web Information Retrieval Using Python and BeautifulSoup. International Journal for Research in Applied Science and Engineering Technology (INRASET).Volume4 Issue VI, 2016.

[16] R. Islam and M. A. Shahjalal, “Soft voting-ased ensemble approach to predict early stage DRC violations”, 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 2019.

[17] Rokach, L. and O.Z. Maimon, Data Mining with Decision Trees: Theory and Applications 2nd Edition. Vol. 81. 2014: World Scientific.

[18] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," 2016 IEEE 6th International Conference on Advanced Computing (IACC).

[19] Ellson J., Gansner E., Koutsofios L., North S.C., Woodhull G. Graphviz— Open Source Graph Drawing Tools. In: Mutzel P., Jünger M., Leipert S. (eds) Graph Drawing. GD 2001. Lecture Notes in Computer Science, vol 2265. Springer, Berlin, Heidelberg, 2002

[20] Rakpong Kittinaradorn et al. (2019, September 23). DeepCut: A Thai word tokenization library using Deep Neural Network. Zenodo.

[21] Ting K.M. Confusion Matrix. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining Second Edition. Springer, Boston, MA, 2017.