Original Article

Role Of Machine Learning In Machine Translation

[1]ARCHANA SACHINDEO MAURYA , [2]PROMILA BAHADUR , [3]DIVAKAR YADAV

[1]Research Scholar, Institute of Technology (IoT), Department of Computer Sciences & Information Systems, Shri Ramswaroop Memorial University, Lucknow
[2]Associate Professor, Institute of Engineering & Technology (IET), Lucknow
[3]Professor, Institute of Engineering & Technology (IET), Lucknow
DOI: 10.47750/pnr.2023.14.S01.155

Abstract

To enhance the quality of machine translation (MT), the most significant way is to train the system. Machine Learning (ML) provides an opportunity to train the system by using a training data set. Further, assessment is done using a test data set. This way training and assessment can be carried out in a given system. The training further helps the system make appropriate predictions. To provide training and testing data ML has techniques like supervised, unsupervised, and knowledge-based. MT has various challenges. Including ambiguity. In this paper, we have done an analysis of various ML techniques for Word Sense Disambiguation (WSD). The accuracy of all the analyzed algorithms ranges from sixty-eight to eight-four percent. We have also introduced a hybrid model named AmbiF to resolve WSD. This model has exhibited a higher accuracy percentage in comparison to the other analyzed techniques. The percentage of accuracy reported is eighty-five percent under the F-score value.

K**eywords:** AmbiF, ambiguity, knowledge-based, machine learning, machine translation, supervised approach, unsupervised approach.

1 INTRODUCTION

For many years, machine translation (MT) has been a broad area of study. The translation is the process of translating meaning of a text into the resulting equivalent meaning that transfers the same message into another language. In the translation process, the text that is to be translated is called the source text, and the text after translation is called the target text [1]. This translation process occurs between two languages. The language of the source text is called the source language, and of the target text is called the target language. Machine Translation is the automatic process of conversion of text from the source language to the target language using computers. The automatic machine translation process is shown in Figure 1.

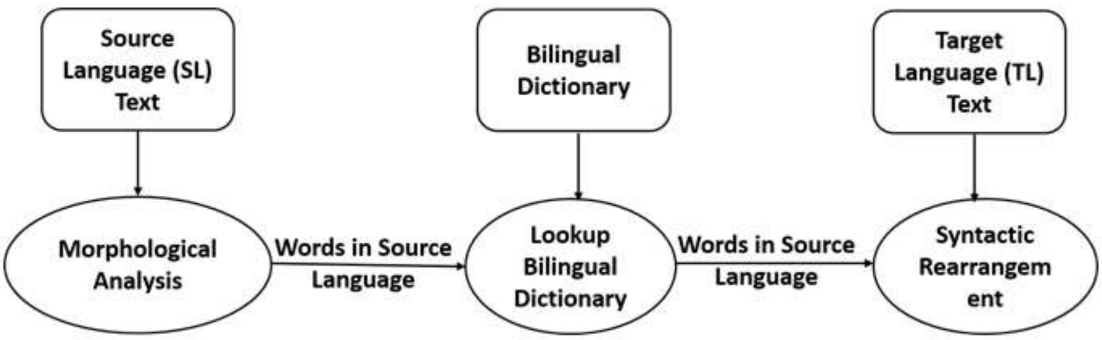


Figure 1: Process of Automatic Machine Translation

The automatic MT uses the bilingual dictionary for establishing fixing points for closing the gap between two languages [2]. India is a multilingual country, and only approximately three percent of people can understand the English language [3]. Therefore, we need a MT system to translate any information in the English language into any other local target language. There are various approaches to MT, and on the basis of these approaches, MT is classified into four categories. These four categories are depicted in Figure 2.

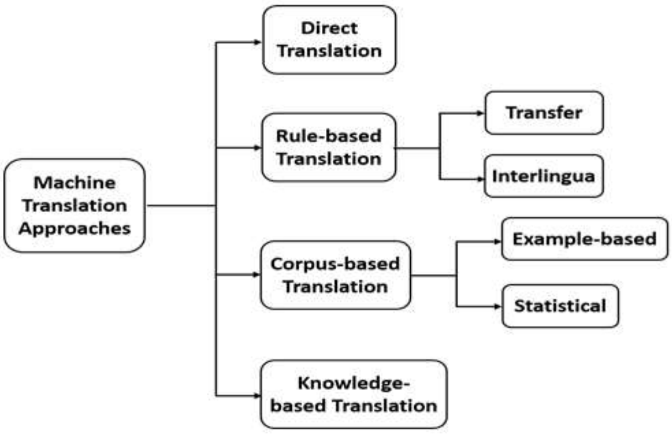


Figure 2: Machine Translation Approaches

Traditionally, MT had two different approaches: the approach based on rules (RBMT, Rule-Based Machine Translation) and the corpus-based approach. However, neural MT and hybrid MT have emerged in recent years [4]. Each MT approach has its own benefits and drawbacks. Today, the hybrid approach, and neural machine translation approach are widely used for testing different combinations.

Arthur Samuel defined the term “ML”, and he described ML as “a field of study that trains computers to enable an automatic and improving learning process from experience without any human intervention, i.e., the capability to learn without being explicitly programmed” [5].

1.1 Role of ML Approaches in MT

Machine Learning (ML) plays an important role in many real-life applications, like photo tagging applications, web search engines, spam detectors, machine translation, etc. ML enables computers to learn by providing training. There are various ML approaches that play an important role in improving the quality of the translated text. Blending ML with MT, enhances the quality of MT. ML addresses the grey areas of MT, one of which is ambiguity. ML approaches are classified into the following categories: supervised, unsupervised, and reinforcement learning. This classification is shown in Figure 3.

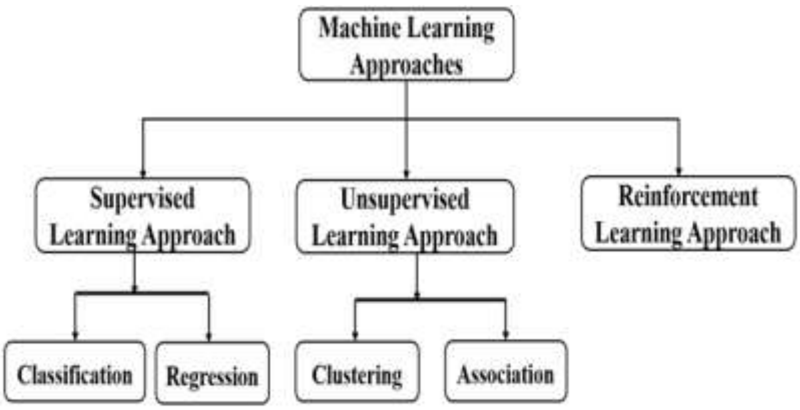


Figure 3: Classification of Machine Learning Approaches

These ML approaches are used in MT for automatic translation, and using these learning approaches, we can improve the quality and percentage of correctly translated text. According to Figure. 3, we can see that there are three types of learning approaches, and these approaches are:

1. **Supervised Learning Approach:** Supervised learning methods can be applied to well-labeled data to train the system. After the training process, the new set of data is applied to the system, which then generates results for the new data. Classification and regression are the two methods in the supervised learning approach.
    - **Classification:** By using supervised learning classification algorithms such as Naive Bayes, K Nearest Neighbor, Support Vector Machines, Logistic Regression, etc., we can classify the text into two or more classes.
    - **Regression:** This technique is used for the prediction of new data and also shows the connection among both dependent and independent variables. Example: linear regression and logistic regression.
2. **Unsupervised Learning Approach:** In the unsupervised learning approach, there is no need for a training corpus, which also significant amount of calculation time [6]. With this method, we may quickly obtain the computer's unlabeled data. Clustering and association are the two methods in this approach.
    - **Clustering:** In clustering, we can divide the data points into two or more groups on the basis of their group features.
    - **Association:** The association rule finds the relationship between various variables in a data set.
3. **Reinforcement Learning Approach:** This is a feedback-based learning approach in which an agent learns from experiences and then performs the action in the environment to see the results. In this method, there is no labelled data. This learning process depends on the hit-and-trial method. The main application areas of reinforcement learning are robotics, gaming, chemistry equations, manufacturing, etc.

2 ORGANIZATION OF PAPER

Our paper contains the following sections:

3 ROLE OF DATA SETS FOR TRAINING AND IN ML PROCESSES

There is a significant role for data sets in ML processes. The data set is divided into two sets: one is the training data set, and another is the test data set. The training data set contains information related to the domain, while the sentences in the test data set are given to the system to test the accuracy of the output result. The accuracy of the prediction output is directly proportional to the training provided to the model. The better trained model gives a better result.

To test the different ML approaches, we have taken a training data set of four sentences that contains an ambiguous word "block." This ambiguous words falls into several part of speech categories. He has checked the efficiency of each algorithm under the three-fold cross validation method. The data set is partitioned into three equal sections for the three fold cross-validation test procedure and in each iteration two parts of the given data set are used for the training purpose and the third part is used to predict the test data. After that, we applied various ML approaches to see the result and check the accuracy of the algorithm. Example: Consider the following sentences shown in Table 1.

Table 1: Data set for training purposes with POS categories of ambiguous words

Training sentence no.	Training sentences	POS Category
1	He tried to block me from joining the meeting.	Verb
2	Look at the block of ice.	Noun
3	She drove the car around the block to charge its batteries.	Noun
4	The block of stone had to be lifted into position with a system of pulleys.	Noun
5	There is a good deli on this block.	Noun
6	My friend circled the block looking for a place to park.	Noun
7	Child has a red colour block.	Noun
8	When a glacier meets the sea, a large block of ice may break off and float away as icebergs.	Noun
9	You can block unwanted numbers	Verb

Here we are trying to train the system for the given ambiguous word "**block.**" The idea is to train the system and obtain a correct prediction of POS for the test data provided. As shown in Table 1, the ambiguous word "**block"**, here belongs to different POS like noun and verb for different sentences.

The information provided by words in the vicinity of the ambiguous word will help train the system and obtain accurate predictions for test sentences. A working model for the prediction of test data is shown in Figure 4.

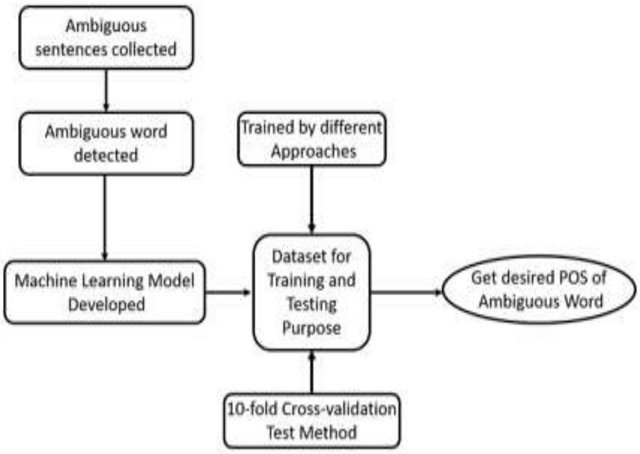


Figure 4.Working Model of Machine Learning Prediction

3.1 Format of the data set

For the preparation of the data set, first of all, we have extracted the neighboring words around the target ambiguous word "**block**" for the fixed size of the window, i.e., 3. It means three neighbor words are on the left and three words are on the right of the target ambiguous words. We have also stored the position-specific information (POS) of these neighboring words. These words are classified as features and will be processed further. The features of the test word "**block**" from the sentences in Table 1 are shown in Table 2.

Table. 2: Sample Format of Data set with neighboring attributes and POS category

Sentences	Ambiguous Words	w-3	POS(w-3)	w-2	POS(w-2)	w-1	POS(w-1)	w	POS(w)	w+1	POS(w+1)	w+2	POS(w+2)	w+3	POS(w+3)	Category
He tried to block me from joining the meeting	block	nil	nil	nil	nil	tried	VB	block	VB	joining	VB	meeting	NN	nil	nil	Verb
Look at the block of ice	block	nil	nil	nil	nil	look	VB	block	NN	ice	NN	nil	nil	nil	nil	Noun
She drove the car round the block to charge its batteries.	block	nil	nil	block	NN	round	NN	block	NN	charge	VB	batteries	NN	nil	nil	Noun
The block of stone had to be lifted into position with a system of pulleys.	block	nil	nil	nil	nil	nil	nil	block	NN	stone	DET	nil	nil	nil	nil	Noun
There is a good deli on this block.	block	nil	nil	good	ADJ	deli	NN	block	NN	nil	nil	nil	nil	nil	nil	Noun
My friend circled the block looking for a place to park	block	nil	nil	tom	NN	circled	VB	block	NN	looking	VB	place	NN	park	NN	Noun
Tom circled the block looking for a place to park his car	block	nil	nil	Tom	NN	circled	VB	block	NN	looking	VB	nil	nil	nil	nil	Noun
When a glacier meets the sea, large block of ice may break off and float away as icebergs.	block	meets	VB	sea	NN	large	ADJ	block	NN	ice	NN	break	VB	off	prep	Noun
You can block unwanted numbers	block	nil	nil	nil	nil	nil	nil	block	VB	unwanted	ADJ	numbers	NN	nil	nil	Verb

4 SUPERVISED MACHINE LEARNING ALGORITHMS

There are many ML algorithms that can help computers perform correct segmentation based on the training data and extracted features. In the process of WSD, these algorithms are helpful in the selection of the accurate meaning of an ambiguous word [7]. Commonly used supervised ML approaches are shown in Figure 5.

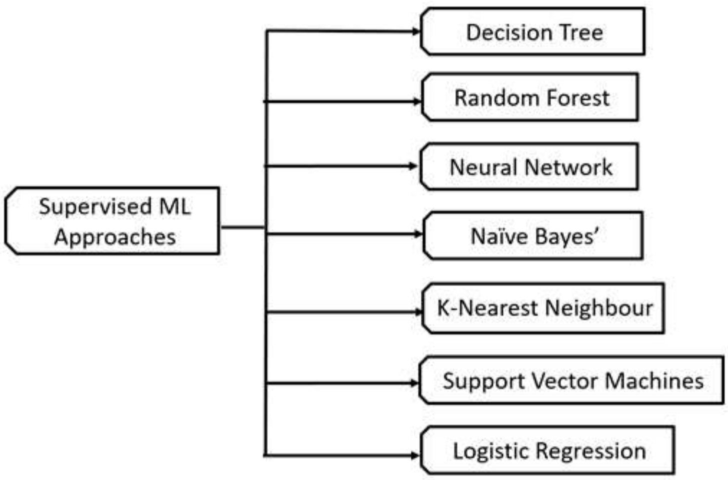


Figure 5: Various commonly used Supervised ML Approaches

4.1 Decision Tree

Decision tree learning is a tree-based ML method, and it is one of the most important and widely used supervised learning algorithms. These algorithms provide great accuracy and stability for the prediction based on test data. These trees can be used to take decisions and implement systems based on these decisions [8]. A person, for example, can play indoor games depending on the weather. The visual representation of this example in the form of a decision tree is shown in Figure 6.

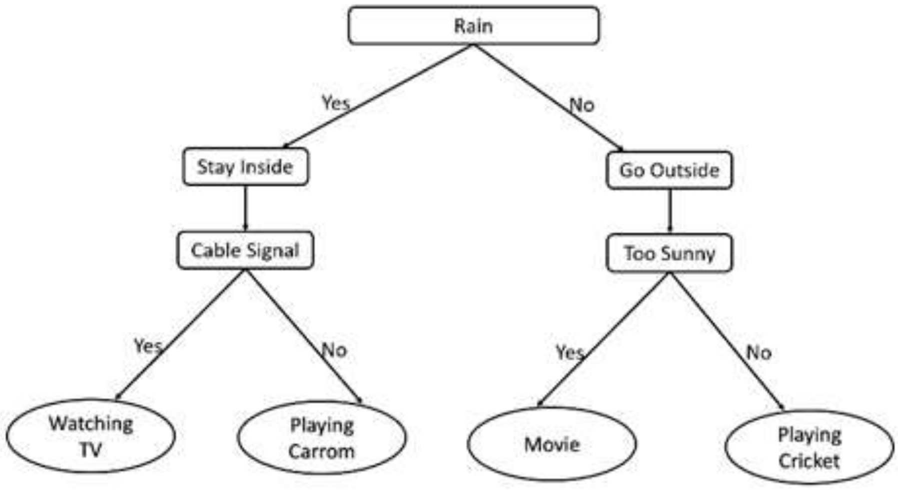


Figure 6: The visual representation of the Decision Tree

The output generated for the decision tree approach for the test data set is shown in Figure. 7.

=== Predictions on test data ===

inst#	actual	predicted	error	prediction
1	2:Noun	2:Noun		0.667
2	2:Noun	2:Noun		0.667
3	2:Noun	2:Noun		0.667
1	2:Noun	2:Noun		0.833
2	2:Noun	2:Noun		0.833
3	1:Verb	2:Noun	+	0.833
1	2:Noun	2:Noun		0.833
2	2:Noun	2:Noun		0.833
3	1:Verb	2:Noun	+	0.833

Figure 7: Predictions made using the decision tree approach on the test data set

4.2 Random Forest

Random Forest (RF) is also one of the most powerful and widely used methods for data exploration, data modeling, and predictive modeling. This classifier is a grouping algorithm; in this type of algorithm, various decision trees are combined with each other. It is called an ensemble technique. An ensemble decision tree will have a low variance and a high accuracy value in comparison to a single decision tree. Figure 8 displays a graphic representation of the Random Forest method.

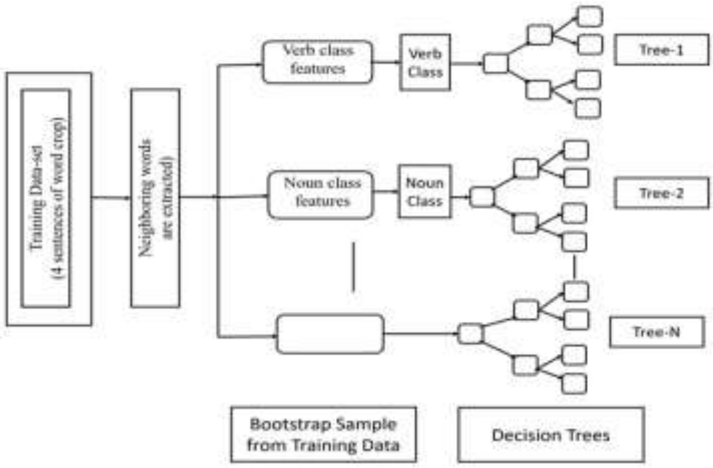


Figure 8: Visual representation of the Random Forest Algorithm

A random forest can be built by using the decision trees for the same data set, but the trees cannot be correlated. The result of this algorithm will be a tree constructed from the results of separate decision trees [9]. This algorithm reduces the chance of overflow, and the accuracy is much higher than a single decision tree. In this algorithm, the decision trees work in parallel, and the problem of bottlenecking doesn't happen. The output generated for the decision tree approach for the test data set is shown in Figure 9.

=== Predictions on test data ===

inst#	actual	predicted	error	prediction
1	2:Noun	2:Noun		0.723
2	2:Noun	2:Noun		0.762
3	2:Noun	2:Noun		0.742
1	2:Noun	2:Noun		0.883
2	2:Noun	2:Noun		0.683
3	1:Verb	2:Noun	+	0.773
1	2:Noun	2:Noun		0.89
2	2:Noun	2:Noun		0.858
3	1:Verb	2:Noun	+	0.843

Figure 9: Predictions made on the set of test data using the Random Forest method

4.3 Neural Network

To process the data in the neural network approach, artificial neurons are used [10]. Artificial Neural Networks (ANNs) are one of the most important and significant features of neural networks because they have the ability to learn like the human brain [11]. ANN is made up of the processing units that are called neurons. A neuron has input units called dendrites and output units called synapses or axons. An ANN is composed of three layers: an input layer, a hidden layer, and an output layer. The first layer is the input layer, which contains neurons, and it transfers information to the middle hidden layer. The hidden layer, after doing some computations on the received information, transfers the calculated information to the output layer. The formula below can be used to determine an artificial neural network's input for its general model [12].

$\mathrm{Yin} = \mathrm{x1.w1} + \mathrm{x2.w} \ldots\ldots\ldots \mathrm{xm.wm}$............ (1)

i.e. total input $\mathrm{yin} = \sum_{i}^{m} \mathrm{xm.wm}$ .... (2)

After applying the stimulation function to the entire entered value, the outcome may be calculated.

$y = F(yim)$…………… (3)

The prediction output predicted by the neural network approach for the test data set is shown in Figure 10.

=== Predictions on test data ===

inst#	actual	predicted	error	prediction
1	2:Noun	2:Noun		0.849
2	2:Noun	2:Noun		0.993
3	2:Noun	2:Noun		0.99
1	2:Noun	2:Noun		0.926
2	2:Noun	2:Noun		0.601
3	1:Verb	2:Noun	+	0.553
1	2:Noun	2:Noun		0.989
2	2:Noun	2:Noun		0.894
3	1:Verb	2:Noun	+	0.543

Figure 10: Predictions made using the neural network approach on the test data set

4.4 Naïve Bayes'

This algorithm is a probability-based supervised machine learning algorithm that works on Bayes' theorem [13]. This algorithm is very popular for solving segmentation problems. There are many variables present in the training data set, and these variables are independent of each other. These independent variables are called "features." To compute the likelihood of certain features in a given class, this approach uses the Bayes' theorem [14-16]. The Bayes' Theorem is widely applied in the field of machine learning.

Suppose there are two events X and Y, then the Bayes' theorem formula can be given by the following equation:

$P(\frac{X}{Y}) = \frac{P(Y/X)*P(X)}{P(Y)}$…………… (4)

Here,

X and Y are the two events.

P(X) is the prior probability before the event is occurred

P(Y) represents the marginal probability

P(Y | X) denotes the likelihood probability of event Y after occurring evidence of event X.

P(X | Y) represents the posterior probability of event X after the event Y is occurred.

Bayes' rule can be used to find features of conditional probability in a particular class [17-19]. The conditional probabilities of each term's value and features in a given sentence are calculated with the help of this algorithm. The best outcome will be produced at the maximum value.

The accuracy of correctly classified sentences in the test data set is shown in Figure 11.

=== Predictions on test data ===

inst#	actual	predicted	error	prediction
1	2:Noun	1:Verb	+	0.674
2	2:Noun	2:Noun		0.885
3	2:Noun	2:Noun		0.921
1	2:Noun	2:Noun		0.695
2	2:Noun	1:Verb	+	0.725
3	1:Verb	1:Verb		0.913
1	2:Noun	1:Verb	+	0.985
2	2:Noun	2:Noun		0.837
3	1:Verb	1:Verb		0.904

Figure 11: Predictions made using the Naïve Bayes' approach on the test data set

4.5 K-nearest Neighbors (KNN)

This is a non-parametric approach to grouping methods. This is a basic but active approach in various situations [20]. KNN is a very popular method for classification techniques, and it is commonly used for unlabeled data. This algorithm works on the votes of its k neighbors [21].

This algorithm works on two important concepts:

i. The first concept is based on the calculation of distances in the testing and training data sets for the two similar features. In this concept, the test data is used to determine which category the neighbor belongs to before calculating the value of k [22].

ii. In the second concept, first of all, select the value of K. The value of k determines the total number of neighbors used in the calculation process.

The value of k in this algorithm should not be very small or very large. This condition is called over and under-fitting, respectively. The Euclidean distance between two points can be calculated using the formula below:

$= \sqrt{((P_2 - P_1)^2 + (Q_2 - Q_1)^2)}$………. (5)

Here,

P1 and Q1 are the coordinates of the test word.

P2 and Q2 are the coordinates of the matching feature.

The algorithm can be applied to the test data, and the output of the prediction is shown in Figure 12.

```
=== Predictions on test data ===

    inst#     actual  predicted error prediction
        1     2:Noun     2:Noun       0.667
        2     2:Noun     2:Noun       0.667
        3     2:Noun     2:Noun       0.667
        1     2:Noun     2:Noun       0.833
        2     2:Noun     2:Noun       0.833
        3     1:Verb     2:Noun   +   0.833
        1     2:Noun     2:Noun       0.833
        2     2:Noun     2:Noun       0.833
        3     1:Verb     2:Noun   +   0.833
```

Figure 12: Predictions made using the K-Nearest Neighbors approach on the test data set

4.6 Support Vector Machine (SVM)

SVM is the widely known supervised model of machine learning that works on the classification problem. The primary objective of the support vector machine is finding the largest margin between the hyperplane for the two specified classes in the training data set. In this algorithm, it is not acceptable for the hyperplane to be closer to data points from the other class. This prerequisite is also required for more accurate generalizations [23]. The hyperplane, which is removed from the data points in each category, will be selected. Support vectors are those points that lie nearest to the margin of the classifier [24]. The predictions made using the support vector machine approach are shown in Figure 13.

```
=== Predictions on test data ===

    inst#     actual  predicted error prediction
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       1
        3     2:Noun     2:Noun       1
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       1
        3     1:Verb     2:Noun   +   1
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       1
        3     1:Verb     2:Noun   +   1
```

Figure 13: Predictions made using the Support Vector Machine approach on the test data set

4.7 Logistic Regression

This is a probability-based text classification method. This is a predictive algorithm that is applied for the prediction of test data. The approach is used for the unconditional data sets, and the output will be in binary form. The classification problem that is based on the binary output is called a "binary classification problem" [25]. The output generated for the logistic regression approach is shown in Figure 14.

```
=== Predictions on test data ===

    inst#     actual  predicted error prediction
        1     2:Noun     1:Verb   +   0.504
        2     2:Noun     2:Noun       1
        3     2:Noun     2:Noun       1
        1     2:Noun     2:Noun       1
        2     2:Noun     1:Verb   +   1
        3     1:Verb     1:Verb       1
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       0.84
        3     1:Verb     1:Verb       1
```

Figure 14: Predictions made using the logistic regression approach on the test data set

5 DEVELOPED MODEL AMBIF

The AmbiF model is a hybrid model that was created to improve the efficiency of ML algorithms. In this model, we are using the combination of the following three ML algorithms:

i. Naïve Bayes algorithm
ii. Support Vector Machine (SVM) algorithm
iii. Decision Tree algorithm

We have also done the evaluation of the data set in Table 1 in this AmbiF model. The output generated for the prediction of test data is shown in Figure 15.

```
=== Predictions on test data ===

    inst#     actual  predicted error prediction
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       1
        3     2:Noun     2:Noun       1
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       1
        3     1:Verb     1:Verb       0.76
        1     2:Noun     2:Noun       1
        2     2:Noun     2:Noun       0.952
        3     1:Verb     2:Noun   +   1
```

Figure 15.Predictions made using the AmbiF model on the test data set

5.1 Testing Method

All the testing was done using the ten-fold cross-validation method, and we choose k = 5 which means that the number of folds used to split the data set is five. This testing method is more common for the prediction of the class prediction of the test data set [26, 27]. The prediction of the test data involves the following steps:

i. The data set is prepared in ARFF or CSV format.
ii. Identification of attributes and classes for each ambiguous word.
iii. Divide the data set into five equal parts after shuffling.
iv. In each iteration, four parts of the training data set are used, and a learning model is developed.
v. On the basis of this learning model, the fifth part is tested, and the final prediction for the class is generated.

5.2 Proposed Approach

We have presented a hybrid model for achieving better accuracy in the prediction of the POS of ambiguous words. The name of the model is given as AmbiF. The purpose of the AmbiF model is to improve the accuracy of previously analyzed ML algorithms [28]. In this model, we have combined three ML approaches, i.e., support vector machines, decision trees, and Naïve Bayes'. The model AmbiF is used to improve the accuracy of the supervised and ML approaches. The AmbiF model is shown in Figure 16.

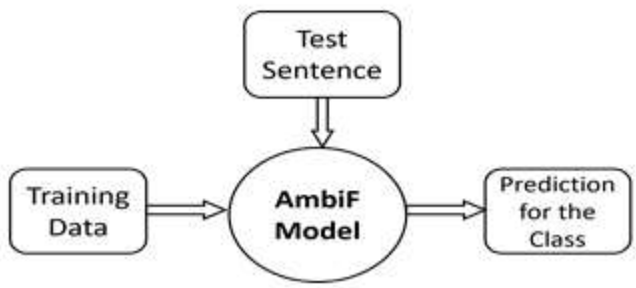


Figure 16: Developed AmbiF model

As shown in Figure.16 the AmbiF model is trained and tested for the prediction of a given test data set.

5.3 AmbiF Model

The AmbiF model has been developed to remove the POS ambiguity. It also helps in the correct categorization of ambiguous words on the basis of the training data set. In this model, we have combined the three supervised ML approaches. Naïve Bayes'. Support Vector Machine (SVM), and Decision Tree are three of these approaches. In Figure 17, the whole AmbiF model is displayed.

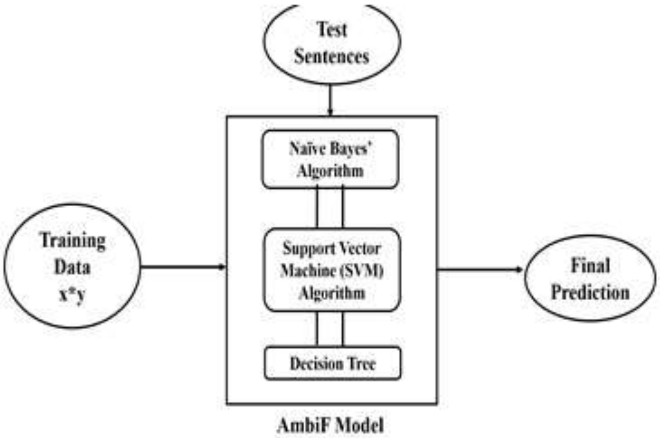
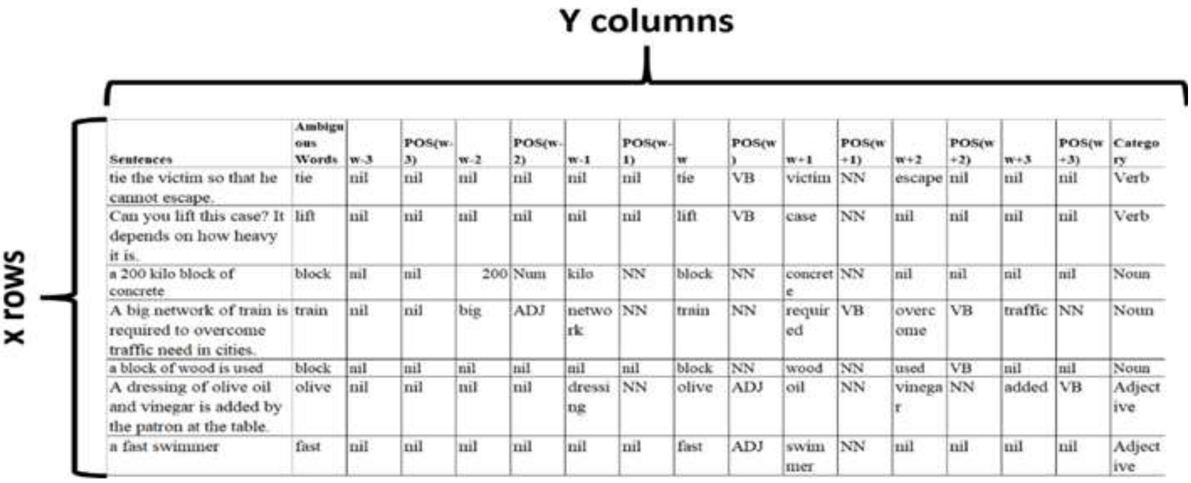


Figure 17: The AmbiF Model

In the AmbiF model shown in Figure 18, the data set for training purposes is defined by the x*y. It means the training data set contains a certain number of rows i.e. x and a certain number of columns, i.e. y. We have prepared a training data set of about 2000 sentences. Here we have defined 17 attributes, which consist of words that are next-door neighbors to the existing ambiguous words. Therefore, in our training data set, x = 2000 and y = 17.

Figure 18 depicts the sample data set, which shows x and y.



Sentences	Ambiguous Words	w-3	POS(w-3)	w-2	POS(w-2)	w-1	POS(w-1)	w	POS(w)	w+1	POS(w+1)	w+2	POS(w+2)	w+3	POS(w+3)	Category
tie the victim so that he cannot escape.	tie	nil	nil	nil	nil	nil	nil	tie	VB	victim	NN	escape	nil	nil	nil	Verb
Can you lift this case? It depends on how heavy it is.	lift	nil	nil	nil	nil	nil	nil	lift	VB	case	NN	nil	nil	nil	nil	Verb
a 200 kilo block of concrete	block	nil	nil	200	Num	kilo	NN	block	NN	concrete	NN	nil	nil	nil	nil	Noun
A big network of train is required to overcome traffic need in cities.	train	nil	nil	big	ADJ	network	NN	train	NN	required	VB	overcome	VB	traffic	NN	Noun
a block of wood is used	block	nil	nil	nil	nil	nil	nil	block	NN	wood	NN	used	VB	nil	nil	Noun
A dressing of olive oil and vinegar is added by the patron at the table.	olive	nil	nil	nil	nil	dressing	NN	olive	ADJ	oil	NN	vinegar	NN	added	VB	Adjective
a fast swimmer	fast	nil	nil	nil	nil	nil	nil	fast	ADJ	swimmer	NN	nil	nil	nil	nil	Adjective

Figure 18. The screen shot of the sample data set showing rows (x) and columns (y)

6 DISCUSSION AND RESULT ANALYSIS

A data set of 2000 sentences containing ambiguous words has been prepared. The ambiguous words have many parts-of-speech, i.e., verb, noun, adjective, and adverb. This data set is evaluated under the ten-fold cross-validation testing method for k = 10. Different supervised ML algorithms, namely, Decision Tree, Neural Network, Random Forest, K-Nearest Neighbor, Naïve Bayes', Logistic Regression, and Support Vector Machines, are tested. The developed ML model AmbiF is used to evaluate all algorithms. AmbiF stands for "**Ambiguity Free.**" This tool is developed for the detection and resolution of POS ambiguity. This plugin is implemented in the EtranS software to provide solutions for ambiguous words. The result is generated on the basis of the prediction reported on the correct POS. The output of the various ML algorithms is given in Table 3.

Table 4: Output generated for the various ML approaches

S. No.	Algorithm	No. of Correct Classifications	Accuracy (Percentage)

1	Logistic Regression	1370	68
2	Random Forest	1407	70
3	IBK	1469	73
4	Neural Network	1520	76
5	Naïve Bayes' Classifier	1525	76
6	SMO	1665	83
7	Decision Tree (J48)	1692	84
8	Hybrid Approach	1696	85

From Table 3, it is clear that the correctly classified percentage and F-score value of the AmbiF model have improved in comparison to the other algorithms except the KNN algorithm. The AmbiF model's accuracy is reported to be eighty-five percent. Thus, developed AmbiF model could be a better approach to resolving ambiguity for existing MT models such as EtranS. The comparative accuracy chart with respect to the number of correctly classified ambiguous words is shown in Figure 19.

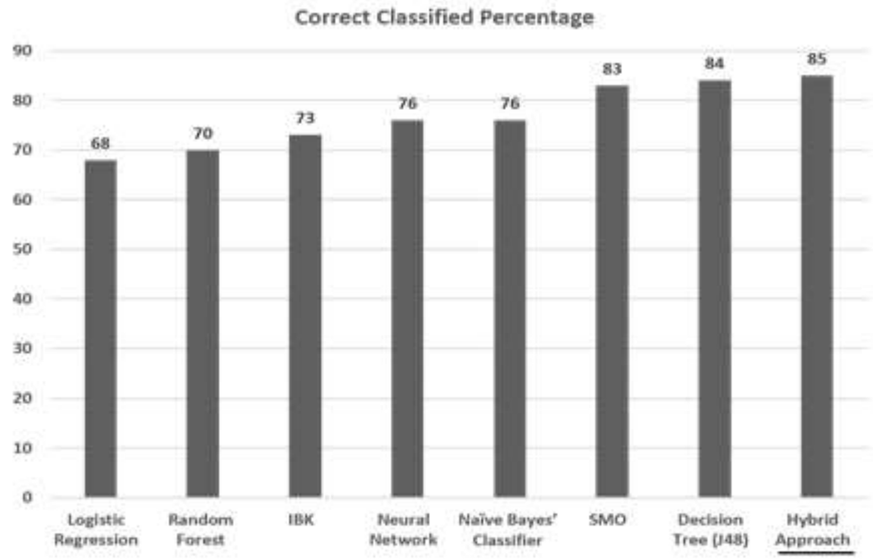


Figure 19: Predictions made using the AmbiF model on the test data set

7 CONTRIBUTION

The community and substantial effect of the given work are as follows:

1. A hybrid model named AmbiF is presented for ambiguous word prediction on the basis of their parts of speech.
2. The proof of concept is given with the help of the AmbiF model.
3. This helps provide ambiguity-free translation with better automated categorization of the ambiguous words.

8 CONCLUSION AND FUTURE SCOPE

In this paper, the significance of ML in MT is discussed in relation to solving the ambiguity problem. To achieve an ambiguity free translation, we have trained the system by using various ML algorithms. In this paper, we have evaluated the performance of various supervised learning approaches, namely Decision Tree, Random Forest, Neural Network, Naïve Bayes', K-Nearest Neighbor, Support Vector Machines, and Logistic Regression, The test was carried out on developed hybrid model, named AmbiF model, using a dataset of 2000 sentences. The parameters are accuracy and efficiency on the pre-processed data set. The performance of ML algorithms was reported sixty-eight to eighty-four percent for all the algorithms.

We have combined the widely used three supervised ML algorithms i.e., Support Vector Machine, Decision Tree, and, Naïve Bayes approach. The AmbiF model has an accuracy rate of eighty-five percent. The performance of all the algorithms is evaluated on the basis of correctly classified sentences.

We are extending our research on improving the efficiency of the system by increasing the size of the training data sets and the window size of the neighboring words.

DECLARATIONS

Conflict of interest statement: The authors declare that there is no conflict of interest.

Author's Contribution: All authors equally contributed to the preparation of this manuscript and approved the final.

Ethical Approval and Consent to participate- Not Applicable

Consent for publication- Not applicable

Human and Animal Ethics- Not applicable

Availability of supporting data-This is a review paper. There is no need of dataset.

FUNDING

Council of Science & Technology sponsored project No CST/D 2475 (Design and Development of Software Tools and Technologies for Translation of Natural Languages). Part of the work presented in the paper was carried out under this research project.

REFERENCES

1. Poibeau, T. (2017). Machine translation. MIT Press.
2. Duan, X., Ji, B., Jia, H., Tan, M., Zhang, M., Chen, B.,... & Zhang, Y. (2020). Bilingual dictionary based neural machine translation without using parallel sentences. arXiv preprint arXiv:2007.02671.
3. Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.
4. Sinha, R. M. K., & Jain, A. (2003). AnglaHindi: an English to Hindi machine-aided translation system. In Proceedings of Machine Translation Summit IX: System Presentations.
5. Rivera-Trigueros, I. (2021). Machine translation systems and quality assessment: a systematic review. Language Resources and Evaluation, 1-27
6. Samuel, Arthur L. (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 44: 206 226. CiteSeerX 10.1.1.368.2254.doi:10.1147/rd.441.0206
7. Rakesh Kumar, Dr. Ravinder Khanna, "Natural Language Engineering: The Study of Word Sense Disambiguation in Punjabi", Research Cell: An International Journal of Engineering Sciences ISSN: 2229-6913 Issue July 2011, Vol. 1, pp. 230-238.
8. Maurya, A. S., & Bahadur, P. (2022). A Detailed Analysis of Word Sense Disambiguation Algorithms and Approaches for Indian Languages. In Proceedings of Second Doctoral Symposium on Computational Intelligence (pp. 693-710). Springer, Singapore.
9. Cuong Anh Le and Akira Shimazu, "High WSD accuracy using Naive Bayesian classifier with rich features", PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo, pp. 105-114.
10. Venkatesan, N., & Priya, G. (2015). A study of random forest algorithm with implementation using weka. International journal of innovative research in computer science and engineering, 1(6), 156-162.
11. G.E.a. D. Radev. "Lexrank: Graph Based Lexical," Artificial Intelligence Research, vol 22 pp.457-479, 2004.
12. Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), 717-727.
13. Haroon, R. P. (2010, December). Malayalam word sense disambiguation. In 2010 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-4). IEEE.
14. Rish, I., 2001. An empirical study of the naive Bayes classifier, in: IJCAI 2001 workshop on empirical methods in artificial intelligence, IBM. pp. 41–46
15. Szepesvári, C. (2010). Algorithms for reinforcement learning. Synthesis lectures on artificial intelligence and machine learning, 4(1), 1-103.
16. http://www.cs.upc.edu/~escudero/wsd/00-ecai.pdf
17. Nyein Thwet Thwet Aung, Khin Mar Soe, Ni Lar Thein, "A Word Sense Disambiguation System Using Naïve Bayesian Algorithm

for Myanmar Language”, International Journal of Scientific & Engineering Research Volume 2, Issue 9, September-2011, pp. 1-7.

18. https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/
19. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science, 3-21.
20. Nyein Thwet Thwet Aung, Khin Mar Soe, Ni Lar Thein, “A Word Sense Disambiguation System Using Naïve Bayesian Algorithm for Myanmar Language”, International Journal of Scientific & Engineering Research Volume 2, Issue 9, September-2011, pp. 1-7.
21. Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 986-996). Springer, Berlin, Heidelberg.
22. Wang, L. (2019, December). Research and implementation of machine learning classifier based on knn. In IOP Conference Series: Materials Science and Engineering (Vol. 677, No. 5, p. 052038). IOP Publishing.
23. Cuong Anh Le and Akira Shimazu, “High WSD accuracy using Naive Bayesian classifier with rich features”, PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo, pp. 105-114.
24. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html
25. S.N. Jeyanthi, “Efficient Classification Algorithms using SVMs for Large Data sets,” A Project Report Submitted in partial fulfillment of the requirements for the Degree of Master of Technology in Computational Science, Supercomputer Education and Research Center, IISC, BANGALORE, INDIA, Jun. 2007
26. https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_quick_guide.htm#
27. Liu, Y., Wang, X., Wang, L., &Lv, Z. (2019). A Bayesian collocation method for static analysis of structures with unknown-but-bounded uncertainties. Computer Methods in Applied Mechanics and Engineering, 346, 727-745.
28. Berrar, D. (2019). Cross-Validation.