

# Named Entity Recognition Using Machine Learning Approaches

A.Salini<sup>1</sup>, U.Jeyapriya<sup>2</sup>

PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India<sup>1</sup>

Asst. Professor, Department of Computer Science, Stella Maris College, Chennai, India<sup>2</sup>

**ABSTRACT:** The term Named-entity recognition is a subtask of information extraction that seeks to locate and classify the named entities from the given text. This paper surveys about the different approaches and techniques that are used for Named Entity Recognition, which identifies the named entities from the given data. It explains the different machine learning models that are used in the Named Entity Recognition. The performances of various Named Entity Recognition systems that are developed using machine learning models are reviewed. It also enumerates the applications of Named Entity Recognition in various fields and the challenges faced by them.

**KEYWORDS:** Named Entity Recognition(NER),Machine Learning(ML),Hidden Markov Model (HMM), Conditional Random Field (CRF), Support Vector Machine (SVM), Maximum Entropy Model (ME),Decision Tree (DT), Parts of Speech (POS).

## I. INTRODUCTION

A large amount of information is being generated rapidly on the web every day and social networking applications are one of the main reasons for the increase of the web data. All these data are usually unstructured in nature. Extracting relevant information from these unstructured data gives rise to a new field called Information Extraction (IE). Information Extraction is a task of automatically extracting structured information like entities and the relations among entities from unstructured data and/or semi-structured documents [25]. Identifying these entities is one of the important sub tasks of Information Extraction and it is known as Named Entity Recognition (NER). Message Understanding Conference (MUC) 6 has defined Named Entity Recognition as the task of identifying the entities like name, location, time, currency, percentage expressions and organization from a given text. After the MUC 6, the interest in Named Entity Recognition has increased and huge efforts have been devoted to its research. MUC 7 has made an attempt to enhance the task of Named Entity Recognition by including the time and date entities and it has introduced the Named Entity Recognition for multiple languages. NER is also being used in the biomedical field to recognize the biomedical entities like gene, protein, DNA and RNA from the given text. An enormous amount of resources are created to assess the performance of the biomedical NER. In the earlier days, NER systems were focusing mainly on the handcrafted rules, where all these rules were coded by the developers. Even today this method is still one of the best ways to extract the named entities but these handcrafted rules suffered from portability. To overcome this difficulty they began to use the machine learning models to replace the hand-crafted rule-based systems. The Hidden Markov Models and Maximum Entropy machine learning models were first deployed in parallel with the handcrafted rule based systems and these models were later suppressed by the conditional model popularly known as Conditional Random Field. In the comparison between the machine learning models, there is no particular winner found. A model which is advantageous in one view then that model lags behind from another viewpoint [24]. It is not possible to apply the best approaches in all cases because each approach needs different technical requirements. This paper is organized into 6 sections. Section 1 contains the introduction to NER. Section 2 presents the various machine learning models used in NER. The related works done using NER is discussed in section3. Section 4 contains the applications of NER. The challenges faced by Named Entity Recognition are described in section 5.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

## II. MACHINE LEARNING APPROACHES

Statistical models are used by the machine learning approaches which focus on automatically recognizing the sequence labelling algorithm or patterns and make decisions based on the observed data. In machine learning approaches the model must be trained using the annotations which are present on the documents that are annotated, these documents are created by the people who have a lot of experience or expertise people in that field, the time taken for training the model depends on the complexity of the model and the annotated documents that are available [10]. The second step in machine learning after training the model is to test the system with raw data and the system annotates the raw data based on the previous experience learned from the annotated data during training. In Named Entity Recognition machine learning approaches focuses on extracting entities and classifying them from the given data. Different features are applied to the text and these representations of text are used by the machine learning models to recognize the features present in the training data, which helps to automatically generate patterns for identifying the information which is similar in the unseen data.

### 2.1 SUPERVISED MACHINE LEARNING

In supervised learning, labelled training data is needed to develop a statistical model because the labelled example helps the system to take the right decision [15]. It is not possible for a supervised learning method to achieve a high performance without large quantities of training data. In supervised learning, if the amount of training data is very small it may lead to data sparseness problem. The different supervised learning algorithms are Hidden Markov Model (HMM), Conditional Random Fields (CRF), Support Vector Machine (SVM), Maximum Entropy Model (ME) and Decision Trees.

#### 2.1.1 Hidden Markov Model

Hidden Markov Model is a generative model and it has proved to be very successful in various sequence labelling tasks such as chunking, Named Entity Recognition, POS tagging and speech recognition. The HMM model is based on the Markov chain and each label is represented as states. Since HMM model defines the joint probability for the observation symbols and state transition there is a probability associated with each transition. Hence, this model can predict the Named entity of the next word when the Named Entity class of the previous word is given [22] [11].

#### 2.1.2 Conditional Random Field

Conditional Random Field (CRF) is a discriminative probability model. It is mainly applied in machine learning and pattern recognition. It is similar to the Maximum Entropy Model and it overcomes the label bias problem faced by the ME model. CRF is an undirected graphical model (random fields). It is used to identify the relationship between the observed data and it generates structured predictions based on the observed data. A discrete classifier predicts a label for entities only without considering the context or the neighbouring entities, but a CRF model will take the context into account [15] [13].

#### 2.1.3 Maximum Entropy Model

Maximum Entropy is a conditional probabilistic sequence model. This model is very flexible and it can handle the dependency between multiple features that are extracted from a single word. The probability conversion that is leaving any state must sum to one, therefore the Maximum Entropy model is influenced towards the state that contains the least number of outgoing transitions. The Label Bias problem can be solved by changing the state-transition structure [13].

#### 2.1.4 Support Vector Machine

Support Vector Machine is a machine learning approach that is used for solving the two-class pattern recognition problems. The main goal of the SVM is to identify whether an entity belongs to the target class or not. A hyperplane is generated while training using the SVM model. The hyperplane helps to categorize the observed data into positive and the negative class that is present on the opposite sides of the hyperplane. This model computes the distance of every vector from the hyperplane which is known as margin [22] [8].

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

## 2.1.5 Decision Tree

Decision Tree is one of the popular and powerful tools that is used for categorizing and forecast. Rules are used for artificial intelligence and neural network in a decision tree. Decision Tree is a classifier which is in the form of a tree structure where each node represents a leaf node. It is an inductive approach that helps to acquire knowledge on classification [13].

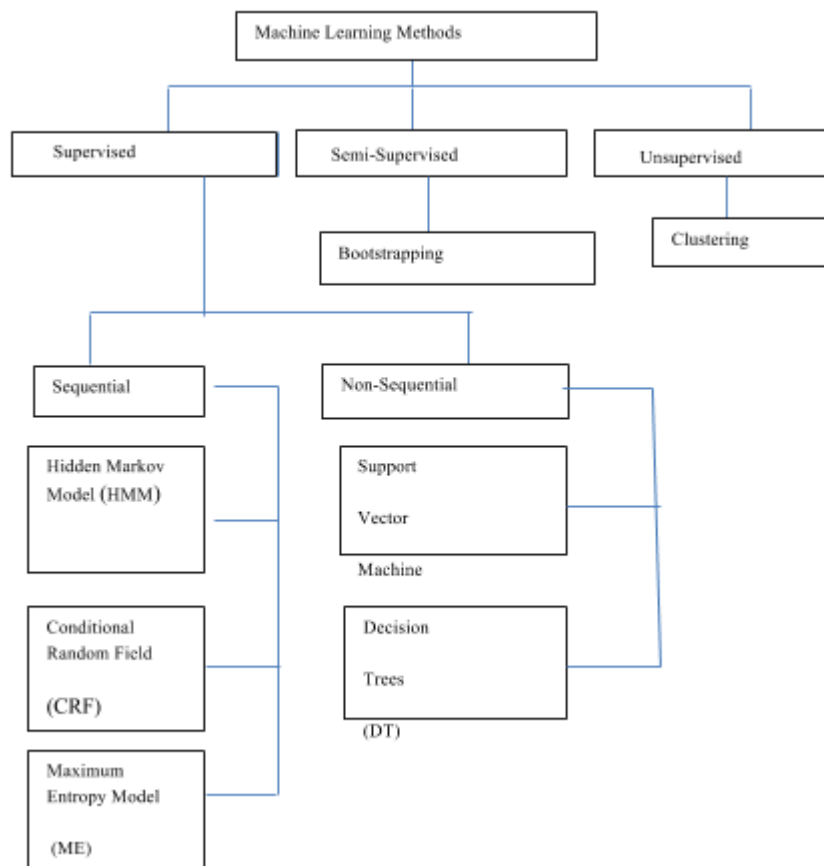


Fig-1: Machine Learning Approaches

## 2.2 SEMI SUPERVISED LEARNING

In semi supervised learning the model is trained using a small amount of labelled data and then the model is tested using another set of unlabelled data. Therefore, after testing the model with unlabelled data the performance of the model is improved by iterating over the predictions which were done by the previous models [13]. For example, a NER system which is developed to identify the “drug names” will be asking for a few drug names as an example, and then the NER systems will be trying to find these drug names along with the other drug names that occur in the similar context. These steps are repeated several times so that the system can identify numerous drug names.

## 2.3 UNSUPERVISED LEARNING

The main disadvantage of supervised learning is that it requires a large amount of annotated data but many languages are not having any annotated data. Therefore, to overcome this difficulty unsupervised technique has been proposed in Named Entity Recognition. Clustering is one of the typical approaches to unsupervised learning. In clustering, we can

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

identify the data from the clustered group on the basis of similar context. The main disadvantage of unsupervised learning is that it requires a lot of features and it is fully dependent on the lexical patterns and the statistics on a large annotated data [13]

S.No	Machine Learning Models	Implemented On The Following Fields	Advantage	Disadvantage
1)	HIDDEN MARKOV MODEL (HMM)	HMM model was implemented in biomedical field to extract the entities like protein, DNA, RNA, cell-type and gene from GENIA corpus. It was also used to extract entities like name, location and organization from the MUC.	-In HMM model the future observations are taken into account. -The theory of HMM is very elegant and easy to understand. Hence, it is very easy to analyse and implement it.	-In HMM model there is no dependency among the words in the input sentence. Due to the joint probability definition in HMM, a lot of parameters has to be evaluated. Therefore it requires a large data set for training.
2)	Conditional Random Field (CRF)	CRF model was implemented in biomedical and also in agricultural domain to identify the crop names, weather, location etc. It is also used to identify the name, location and Organization.	-The CRF model overcomes the label bias problem faced by the Maximum Entropy model. -The CRF model is very adaptable or flexible with respect to feature selection. -In CRF model it is not mandatory for features to be conditionally independent.	- The computation time is very long.
3)	MAXIMUM ENTROPY MODEL (ME)	This model was implemented in biomedical field and to identify the entities like name, person and organization.	-In Maximum Entropy model, it is not mandatory for features to be conditionally independent. -Every state contains an exponential model which takes the observation feature as input and outputs the conditional probability of the possible next state.	-Maximum Entropy model suffers from label bias problem. -In Maximum Entropy Model, the future observations are not taken into account.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

4)	SUPPORT VECTOR MACHINE (SVM)	This model is used in biomedical NER to extract entities like gene, protein, DNA and RNA.	-Since SVM supports Kernel functions, it can learn various combinations of the given features with less computational complexity.	-The dependencies such as state to state and the feature to feature dependencies are not considered by the Support Vector Machine.[22]
----	------------------------------	---	---	--

**Table-1:Comparative Analysis of Machine Learning Models in Named Entity Recognition**

### III. RELATED WORKS

N. Kanya and T. Ravi [16], the author has proposed a named entity recognition system which aims to extract significant information from the biomedical archives. The author demonstrates the extraction of entities in the datasets like PUBMED and MEDLINE database using various machine learning algorithms like SVM and feature space based CRF. Three corpora (Breast cancer corpus, Lung cancer corpus, and Thyroid cancer corpus) were retrieved from the PUBMED corpus Precision, Recall and f-score are calculated for these corpora. The results obtained from the named entity recognition systems using SVM and feature space based CRF are as follows. The obtained results prove that the outcome of the feature space based CRF is better than SVM.

Alireza Mansouri, Lilly Suriani Affendey, Ali Mamat [5], this paper explains about the various named entity recognition methods like rule based, machine learning based and hybrid Named entity recognition for extracting entities like Persons, locations, and Organizations from the MUC and the review of these methods. Results of this experiment reveal that Hand-made rule NER has a higher precision and recall for a specific domain, but it still suffers from portability. Performance of the machine learning based NER depends on the data training value. The performance increases when we train the system with the huge amount of data. Authors of this paper have proposed a new Fuzzy Support vector machine for Named entity recognition. Future work of this paper includes implementing different models for achieving high results by defining a rich feature set.

Lin Yao Hong Liu, Yi Liu, Xinxin Li and Muhammad Waqas Anwar [14], the author explains how the feature information generated by the neural networks from the unlabelled biomedical text files are represented as word vectors. This paper presents a biomedical named entity recognition which is based on deep neural network architecture. It uses a skip-gram neural network language model for training the word representations with unlabelled training data. The performance of the neural network model has achieved close to the other state of art models.

Shipra Dingare, MalvinaNissim Jenny Finkel, Christopher Manning and Claire Grover [3], the author presents a Named Entity Recognition based on maximum entropy system for extracting entities present in the biomedical text and reviews its performance. This paper demonstrates the extraction of the various biomedical entities from the Bio Creative and Bio NLP dataset of MEDLINE. This paper suggests that Biomedical NER is difficult than the NER that is used in the newswire domain for extracting entities like Persons, locations, and organizations, because of the complexity that is present in the biomedical domain that affects the performance of the system. The challenge for NER in biomedical domain is the names are very long and complex and even the data annotation is difficult.

Ahmed Sultan Al-HegamiAmeen Mohammed Farea Othman FuadTarboshBagash [20], this paper highlights the importance of NER in the biomedical domain and it aims to develop a named entity recognition model for extracting biomedical text files. This paper proposes a NER system based on Naïve Bayes, K-Nearest Neighbour, and decision tree classifiers. The three classifiers are tested on a different subset of features and the values are compared. The results demonstrate that the K-Nearest Neighbour classifier performs better than the other models for biomedical texts when it is trained with a set of rich suitable features.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

Natalia Ponomareva, Paolo Rosso, FerranPla, Antonio Molina [4], this paper focuses on the named entity recognition system which aims to extract significant information from the biomedical archives. It demonstrates the extraction of entities in the JNLPBA version of GENIA corpus using various machine learning algorithms like HMM (Hidden Markov Models) and CRF (Conditional Random Fields). This paper addresses the various Biomedical Named Entity Recognizer that could be built based on the HMM and CRF model. They have enhanced the HMM model by specializing its state and the CRF model is enhanced by using POS tag information. The results of the two models for extracting the biomedical entities like protein and cell-type are compared based on the precision, recall and f-score values. The final results prove that CRF model has a better performance when we compare the F-score and the precision values but the recall value is low when compared to the HMM model, therefore this paper concludes that the combinations of both the models would improve the performance of the system.

GuoDong, Zhou Jian Su [1], this paper focuses on the named entity recognition system which aims to extract and classify name, time and numerical quantities. It demonstrates the extraction of entities in the MUC-6 and MUC-7 using HMM (Hidden Markov Models) and HMM-based chunk tagger. Features like capitalization, semantic features, gazetteer feature and context features were used for improving the performance of the system. This model yields a very high performance when compared to other machine learning models. It is concluded that we can increase the performance of the system a little more by using name alias algorithm which is more effective.

Malarkodi C. S, Elisabeth Lex, SobhaLalitha Devi [17], the authors of this paper presents the Named Entity Recognition System that is used to extract the entities like crop names, fertilizers, climate, location in the agricultural domain, So far they have not developed any Named Entity Recognizers for extracting entities in the agricultural field. This was a first attempt for using Named Entity Recognition in agricultural domain and due to that reason no annotated corpus and tag set are available in this domain. Therefore, they have developed an NE tag set containing 19 fine grain tags which include all the important key terms that are present in the agricultural domain. They have developed a corpus which contains one lakh word forms. The dataset used to extract entities were collected from the Wikipedia articles, and they have developed a Named Entity Recognizer based on CRF (Conditional Random Field) model which uses features like POS and chunking information. The system has achieved 76% accuracy by using the basic features like POS and chunk information and by adding some heuristic rules they have increased the performance of the system.

Sujan Kumar Saha, Sudeshna Sarkar, PabitraMitra [7], this paper focuses on the named entity recognition system which aims to extract entities like proteins, cell-line, DNA, RNA and cell-type from the biomedical archives. It demonstrates the extraction of biomedical entities from the JNLPBA 2004 dataset of GENIA corpus version 3.02. This paper presents the word clustering and selection based feature reduction methods used by the Named Entity Recognizer that is built based on the Maximum Entropy Model. The results obtained from the proposed system prove that the performance of the system is superior to other systems that do not use domain knowledge.

Jie Zhang, Dan Shen, GuodongZhoua, Jian Sua, Chew-Lim Tanb [2], the authors of this paper presents a Biomedical Named Entity Recognition which is based on Hidden Markov Model for extracting biomedical entities present in the GENIA version 1.1 and 3.02. Features like orthographic, part-of-speech, morphological and semantic trigger features were added to the HMM model using the back-off modelling for improving the performance of the system. This paper also proposes a Named Entity Recognizer which uses methods for identifying or recognizing the abbreviations and the cascaded named entity recognition.

Shaodian Zhang, NoemieElhadad [12], this paper presents the unsupervised Biomedical Named Entity Recognition system which is used for extracting biomedical and clinical texts. This paper proposes a biomedical NER for extracting entities from the i2b2 and the GENIA corpus. It provides a step by step procedure for solving the challenges in entity detection and entity classification without depending on annotated data, hand-crafted rules.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

TitleOf The Paper	Year	Dataset	Model	Feature	Precision	Recall	F-Score
Enhancing HMM-based biomedical named entity recognition by studying special phenomena [2].	2004	GENIA version	Hidden Markov Model	Orthographic, part-of-speech, morphological and semantic trigger	67.7	65.3	66.5
		3.02 version 1.1			63.8	61.3	62.5
A system for identifying named entities in biomedical text: how results from two evaluations reflect on both the system and the evaluation [3].	2005	MEDLINE	Bio NLP		71.62	68.6	70.1
			Bio Creative		82.8	83.5	83.2
Conditional Random Fields vs. Hidden Markov Models in a biomedical Named Entity Recognition task [4]	2007	JNLPBA version of GENIA corpus	Hidden Markov Model	Specialization of states	65.55	70.45	64.48
			Conditional Random Fields	POS tag	74.7	65.35	67.7

**Table-2:Named Entity Recognition in Biomedical Field**

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

TitleOf The Paper	Year	Dataset	Model	Feature	Precision	Recall	F-Score
Feature selection techniques for maximum entropy based biomedical named entity recognition.[7]	2009	JNLPBA 2004 dataset of GENIA corpus version 3.02	Maximum Entropy Model	NE Tag, Special Char, Normalization, Affix, POS, Trigger words	64.76	66.85	65.79
Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts.[12]	2013	GENIA corpus	Un-supervised Machine Learning Approach		37.0	42.3	39.5
Biomedical Named Entity Recognition based on Deep Neural Network” International Journal of Hybrid Information.[14]	2015	GENIA CORPUS	Deep Neural Network	Word Vectors	66.54	76.13	71.01
Named Entity Recognition From Biomedical Text -An Information Extraction Task[16]	2016	PUBMED	SVM		87.12	88.82	87.67
			CRF		91.66	95.61	93.61
A Biomedical Named Entity Recognition Using Machine Learning Classifiers and Rich Feature Set.[20]	2017		Decision Trees	prefix, suffix, previous tag, previous word	93.20	59	64.34
			Naïve Bayes		92.36	60	69.95
			KNN		92.40	57	64.77



## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

Title Of The Paper	Year	Dataset	Model	Feature	Precision	Recall	F-Score
Named Entity Recognition for the Agricultural Domain [17]	2016	Articles from Wikipedia Related to agriculture	Conditional Random Field	POS and chunking information	83.24	83.13	83.18

**Table-3: Named Entity Recognition in Agricultural Field**

Title Of The Paper	Year	Dataset	Model	Feature	Precision	Recall	F-Score
Named Entity Recognition using an HMM-based Chunk Tagger.[1]	2002	MUC-6	Hidden Markov Model	Capitalization, semantic features, gazetteer feature and context features	96.3	96.9	96.6
		MUC-7			93.7	94.5	94.1
Named Entity Recognition Approaches [5]	2008	MUC	SVM		89.57	83.46	86.40
			Maximum Entropy		43.70	60.89	50.88
			MENE		89	96	92.20

**Table-4: Named Entity Recognition to Identify Person Name, Location and Organization**

### IV. APPLICATIONS OF NAMED ENTITY RECOGNITION

Named Entity Recognition is used in different fields, some major applications of NER are mentioned below:

#### 4.1 NER IN SOCIAL MEDIA:

Social media refers to the means of interaction through which millions of people create share and exchange information or ideas which lead to generate a large volume of data and many organizations are reported to watch these data to know

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

about the users view or suggestions about an organization. Extracting valuable information from the social media is done by using parts of speech tagging and entity chunking [21]. There are almost more than 250 million people using twitter and given the fact that more than 450 million tweets are being sent per day, and all these data are informal in nature. Hence, classifying these named entities is a difficult task in social media [21]. There is an increase in the usage of Named Entity Recognition in social media analysis as people are using social media like Facebook and twitter for sharing their opinions which are valuable information for many organizations or business purpose.

## 4.2 NER IN HEALTHCARE:

In health care Named Entity Recognition is used for extracting the patient details like patient names, disease and the treatment given to the patients from the Electronic Medical Records. The information that is extracted from these medical health records may be used by the researchers in the medical domain for their future study [21]. The Medical Named Entity Recognition which is a sub task of information extraction is used to identify medical entities like disorders and the Medicine names. There is a lot of complicated terms present in the medical field. Hence it is difficult to identify the medical terms.

## 4.3 NER IN BIOMEDICAL FIELD:

Due to the increase in the growth of biomedical data, a lot of demand is arising to extract the biomedical entities. In Biomedical field Named Entity Recognition is used to find entities like genes, proteins, viruses, DNAs, and RNAs. The challenges of NER that is present in biomedical domain are, it contains very long and complex names and also the data annotations are very difficult [3]. Due to the difficulty and complexity that exists in the biomedical field, the performance of the Named Entity Recognition systems in this field is very low compared to the performance of NER systems in other fields.

## V. CHALLENGES OF NAMED ENTITY RECOGNITION

Some of the challenges faced by Named Entity Recognition are described below:

- It is difficult for recognizing words that have different meanings when it is used in different context. For example, the word "Harry Potter" refers to a person in the 1<sup>st</sup> example and in the 2<sup>nd</sup> example, the word "Harry Potter" refers to the Harry Potter novel.

Eg1. "**Harry Potter** is a brave boy."

Eg2. "*Harry Potter* is written by the author J. K. Rowling."

To overcome this difficulty we must assume many examples containing the word "HARRY POTTER" in different contexts and the system should be trained with a lot of sentences.

- It is difficult for identifying the person or the location names which are not frequently used.
- For machine learning approaches a lot of annotated data is required for training the models but in some fields like agriculture, there is not much annotated data available which reduces the performance of the system.
- Selecting the features is a very difficult in Named Entity Recognition because selecting a wrong feature may reduce the performance of the system.
- Sometimes the performance of the NER system may be high on a small training dataset but the performance might reduce to a greater extent when it is applied to a larger dataset

## VI. CONCLUSION

This paper surveys the different machine learning approaches that are used in various fields of Named Entity Recognition in detail. Among the various machine learning approaches of NER that has been surveyed, it is very difficult to state which model is best suited for NER because each model has its own advantages and limitations. Based on the results it has been proved that the selection of features is playing a very important role in the performance of the machine learning models. Even though machine learning models obtain good results but the performance of the rule-based approach is very high when compared to the machine learning approach. The main disadvantage of the rule-based approach is that it is not portable. Therefore, the advantages present in the machine learning approaches and the

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

high performance of the rule-based approach in NER are combined and it is being used as the Hybrid approach in several Named Entity Recognition systems.

### REFERENCES

- [1] GuoDong, Zhou Jian Su, "Named Entity Recognition using an HMM-based Chunk Tagger", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 473-480.
- [2] Jie Zhang, Dan Shen, Guodong Zhoua, Jian Sua, Chew-Lim Tanb, "Enhancing HMM-based biomedical named entity recognition by studying special phenomena", Journal of Biomedical Informatics (2004).
- [3] ShipraDingare, MalvinaNissim Jenny Finkel, Christopher Manning and Claire Grover, "A system for identifying named entities in biomedical text: how results from two evaluations reflect on both the system and the evaluations", Comparative and Functional Genomics 2005; .DOI: 10.1002/cfg.457.
- [4] Natalia Ponomareva, Paolo Rosso, FerranPla, Antonio Molina, "Conditional Random Fields vs. Hidden Markov Models in a biomedical Named Entity Recognition task", Int. Conf. Recent Advances in Natural Language Processing, RANLP-2007.
- [5] Alireza Mansouri, Lilly SurianiAffendey, Ali Mamat. "Named Entity Recognition Approaches", IJCSNS International Journal of Computer Science and Network Security, February 2008.
- [6] Dingcheng Li, Karin Kipper-Schuler, GuerganaSavova. "Conditional Random Fields and Support Vector Machines for Disorder Named Entity Recognition in Clinical Texts", 2008 Association for Computational Linguistics.
- [7] Sujan Kumar Saha, Sudeshna Sarkar, PabitraMitra, "Feature selection techniques for maximum entropy based biomedical named entity recognition", Journal of Biomedical Informatics 42 (2009) doi:10.1016/j.jbi.2008.12.01.
- [8] Asif Ekbal, SivajiBandyopadhyay, "Named Entity Recognition using Support Vector Machine: A Language Independent Approach", World Academy of Science, Engineering and Technology 2010.
- [9] Shilpi Srivastava, MukundSanglikar, D.C Kothari. "Named Entity Recognition System for Hindi Language: A Hybrid Approach", International Journal of Computational Linguistics (IJCL) 2011.
- [10] David Campos, Sérgio Matos and José Luís Oliveira, "Biomedical Named Entity Recognition: A Survey of Machine-Learning Tools", 2012 DOI: 10.5772/5106.
- [11] SudhaMorwal, NusratJahan, Deepti Chopra, "Named Entity Recognition using Hidden Markov Model (HMM)", International Journal on Natural Language Computing (IJNLC), December 2012.
- [12] Shaodian Zhang, NoemieElhadad, "Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts", Journal of Biomedical Informatics (2013).
- [13] Daljit Kaur, Ashish Verma, "Survey on Name Entity Recognition Used Machine Learning Algorithm", International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5875-5879.
- [14] Lin Yao Hong Liu, Yi Liu, Xinxin Li and Muhammad Waqas Anwar. " Biomedical Named Entity Recognition based on Deep Neutral Network", International Journal of Hybrid Information Technology (2015), pp.279-288.
- [15] Ms Maithilee, L. Patwar, Mrs M. A. Potey, "Approaches to Named Entity Recognition: A Survey", International Journal of Innovative Research in Computer and Communication Engineering", December 2015.
- [16] N. Kanya and T. Ravi,"Named Entity Recognition from Biomedical Text -An Information Extraction Task", ICTACT Journal on Soft Computing, July 2016, ISSN: 2229-6956.
- [17] Malarkodi C. S, Elisabeth Lex, SobhaLalitha Devi, "Named Entity Recognition for the Agricultural Domain", Research in Computing Science (2016).
- [18] Wahab Khan, Ali Daud, Jamal A. Nasir, TehminaAmjad "A survey on the state-of-the-art machine learning models in the context of NLP" Kuwait J. Sci. pp. 95-113, 2016.
- [19] Alshaikhdeeb, KamsuriahAhma,"Biomedical Named Entity Recognition: A Review", *International Journal OF Advanced Manufacturing Technology* (2016) ISSN: 2088-5334.
- [20] Ahmed Sultan Al-HegamiAmeen Mohammed Farea Othman FuadTarboshBagash, "A Biomedical Named Entity Recognition Using Machine Learning Classifiers and Rich Feature Set", IJCSNS International Journal of Computer Science and Network Security, January 2017.
- [21] Sanjana Kamath, RupaliWagh, "Named Entity Recognition Approaches and Challenges", International Journal of Advanced Research in Computer and Communication Engineering Vol. 6, February 2017.
- [22] Shrutika Kale, SharvariGovilkar. "Survey of Named Entity Recognition Techniques for Various Indian Regional Languages", International Journal of Computer Applications (0975 – 8887) Volume 164 – No 4, April 2017.
- [23] Prof.VikasBalasahebBurgute, Prof.A.K.Gupta, "Named Entity Recognition using Tweet Segmentation", International Research Journal of Engineering and Technology (IRJET) ISSN: 2395, July -2017, p-ISSN: 2395-0072.
- [24] Rohit Kumar, PriyaBatta, DeepshikhaChhabra and LokeshPawar,"Improved Speech Recognition, Classification, Extraction using Conditional Random Field with Kernel Approach", International Journal of Computational Intelligence Research (2017) ISSN 0973-1873.
- [25] About Information Extraction," Available: [https://en.wikipedia.org/wiki/Information\\_extraction](https://en.wikipedia.org/wiki/Information_extraction).