IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 1, February 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Prediction of Loan Status using Clustering Technique in Machine Learning

[1] M Ashish Naidu; [2] K Radha

[1] III.BTECH-CSE,GITAM University
Hyderabad,Telangana,India

[2] Asst.Professor ,CSE,GITAM UNIVERSITY
Hyderabad,Telanagna,India

Abstract - Nowadays, large amount of data is available everywhere. Therefore, it is very important to analyse this data in order to extract some useful information and to develop an algorithm based on this analysis. This can be achieved through data mining and machine learning. Machine learning is an integral part of artificial intelligence, which is used to design algorithms based on the data trends and historical relationships between data. Machine learning is used in various fields such as bioinformatics, intrusion detection, Information retrieval, game playing, marketing, malware detection, image de-convolution and so on. This paper presents the Necessity of Machine Learning and Role of Natural Language Processing in Machine Learning, Classification of Machine Learning Algorithms, Role of Machine Learning in Deep Learning and its Applications, A Comparison of Machine-Learning Classifiers, Multimodal Machine Learning, Deep Learning Based Natural Language Processing, Natural Language Processing Generation, Integration of Predictive Intelligence with Social Media Data. Implemented the Classification and Clustering Techniques and compared the results with EM Clustering algorithm ,ZeroR algorithm and Random Forest Algorithm for the given attributes. It has taken less time when compared to Random Forest and EM Clustering Algorithm.

.

Keywords **-** *Machine Learning, Deep Learning, Artificial Intelligence, Malware Detection, ZeroR, EM, Random Forest algorithm, Classification, Clustering*

1. Introduction

Machine Learning can be learned from the experience and examples with the knowledge of Programming. Instead of writing the Source Code, we will feed to the generic algorithm; it constructs the logic based on the data given. Consider a Classification algorithm. It keeps the data into distinct groups. It determines the hand written alphabets used to classify the emails into spam and not-spam. There are various examples for machine learning such as Medical Diagnosis (it diagnose the patient suffering from disease or not, Email Filtering (it classify the mails into spam and not spam) and Face Detection (it identifies the faces in the form of images).

1.1 Necessity of Machine Learning and Role of Natural Language Processing in Machine Learning

Machine Learning is an area which is raised out of Artificial Intelligence. By applying this AI, we can build the better intelligent machines. We were unable to write a source code in a complex manner Except for the few tasks like to determine the shortest path among the X and Y points, which gives so many issues. Hence, there is realization that, we can only able to get this problem to let the machine learn from itself. Hence Machine Learning was developed as a new ability to the computers. While using the Machine Learning, we don't even realise that machine learning exists in various segments of technology. Determining the patterns in data on planet earth, which is possible only for human brains. As data is increasing and generating in massive manner, it is consuming more time to . Hence, Machine Learning came into the picture that to assist the users with massive volumes of data in a less time. As Machine Learning assists to analyse the big data , it makes the task of Data Scientists easier in an automated process which gains an equal recognition and importance.
The mechanisms which we are using for Data mining exists around many years. But these data mining algorithms is not that much efficient and effective to run the algorithms. If we run the Deep Learning with better data access, it generates the output in a such a manner that which will leads to dramatics breakthroughs in terms of Machine Learning. Machine Learning Algorithms can be Classifies

into three ways such as Supervised Learning, Unsupervised Learning and Reinforcement Learning. Mostly, Machine Learning uses the Supervised Learning algorithms only. From the given previous examples, System is trying to learn. The systems is able to determine the patterns directly from given examples in an unsupervised learning. Consider the Mapping Function for a supervised learning, both the input variables and output variables such as X,Y can be expressed as Y=f(X). Supervised Learning problems are divided into two ways such as Regression and Classification. In the Classification algorithm output variable is considered as Category such as spam, no spam or black or white. Regression algorithm considers the output variable as Real Value such as Height or Rupees. Unsupervised Learning algorithms can be classified into two parts such as Association and Clustering. In Association Rule algorithm we will find the rules which describe the larger portions of data. For Example, if the Customer is buying the Computer means he tends to buy the Antivirus Software also. In Clustering Algorithm, we will find the similar kind of groups such as grouping the Customers by their buying behaviour. In Reinforcement Learning algorithm, machine is trained to make the particular decisions. The machine is continuously trains itself by using the trial and error method.

1.2. Natural Language Generation

It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation.
It involves −

- **Text planning** − It includes retrieving the relevant content from knowledge base.
- **Sentence planning** − It includes choosing required words, forming meaningful phrases, setting tone of the sentence.
- **Text Realization** − It is mapping sentence plan into sentence structure.

1. **Natural Language Processing**: It can be applied into various areas like Machine Translation, Email Spam detection, Information Extraction, Summarization, Question Answering etc.

Natural language processing (NLP) analyzes human language computationally and gained a enormous attention. Machine translation, email spam detection, information extraction, summarization, medical, and question answering are few of the applications of NLP.

The paper distinguishes four phases by discussing different levels of NLP and components of **N**atural **L**anguage **G**eneration (NLG) followed by presenting the history and evolution of NLP, state of the art presenting the various applications of NLP and current trends and challenges [7]. Natural language processing is the subfield of machine learning which deals with the interaction between machine and natural language of humans. A recent advance in machine learning has enabled computers to understand and process human languages. Deep learning is been used for text language translation, semantic understanding and text summarization.

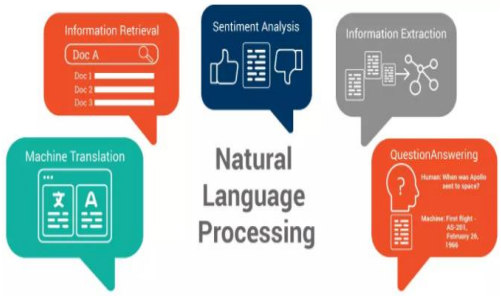

Fig.1. Natural Language Processing Applications

The most well-known applications for NLP include:

(a) .Named Entity Recognition: The main task of Named Entity Recognition is to categorize named entities into predefined categories. Jack, Alibaba, China are categorized as Name, Organization and country.

(b) Sentiment Analysis:

It deciphers the emotions expressed by authors. This analysis is significantly been used to the customer's feelings towards a product.

(c) Speech recognition:

Speech recognition has many applications such as home automation, virtual assistance, hands free computing etc. Neural networks are widely used in this area.

(d) Parts of Speech tagging:

Natural language processing libraries can determine the parts of speech of each word and can build the relationships between subject and verb.

(e) Spell checking:

Neural networks are incorporated into text editors to check the text documents containing spelling mistakes.

(f) Information extraction:
It summarizes the larger paragraphs into shorter text, which can be considered as writing an abstract about a document.

(g) Machine translation :
Machine translation is to translate texts into another language (or to any specific languages) automatically.

2. Database:

Database is an essential component used in machine learning. To build a machine learning system the data must be collected in the form of datasets or data must be generated manually. All the datasets that are used for machine learning are combined together to form the database. Mostly, datasets are classified into three categories.

I) Train dataset: Train dataset is portion of the dataset that is used to train the machine learning system the important features of data which is a compulsion to any of the machine learning model.
II) Test dataset: machine learning Model's performance is tested is using test dataset.
III) Validate dataset: Validate dataset is used to compare models to pick out the optimal one by trimming models' coefficients. It provides an
unbiased evaluation of a model fit on the training dataset while tuning the model's hyper parameters .Validate dataset is different from train dataset, and it cannot be used in training section or over fitting may occurred and adversely affect new data generation. Generally the ratio between these three datasets is in 50:25:25(train: test: validate). In some models as the validation and test data are combined and ratio is considered to be 70:30(train: test + validate).

3. Computer vision:

Computer vision is a subfield in artificial intelligence which deals with analysing and understanding multimedia data like figure or videos.

Fig.2. Computer Vision

The problems in computer vision include
Image classification: Image classification is a computer vision fundamental task that teaches computer to recognize and classify certain images.Target detection: All the targets that are to be detected by teaching a model intially.The found targets are circled out by using rectangles. For instance, target detection can be used to configure the traffic recognition in a self-driving car. Model can detect every predefined matter and highlight them out. Image segmentation: The important goal of segmentation is to simplify or change the image representation that is easier and user friendly to analyse. Segmentation is performed to partition an image into multiple segments also known as super-pixels. Significance test: Once sample data has been gathered through an observational study or experiment, statistical inference assess evidence in favour or some claim about the population from which the sample has been drawn from. The methods of inference used to support or reject claims based on sample data are known as tests of significance.

2. Comparison of Machine-Learning Classifiers

A. Supervised Learning

[4] Supervised learning as the name indicates a presence of supervisor as teacher. Supervised learning is a learning in which we teach or train the machine using data which is well labelled that means some data is already tagged with correct answer. After that, machine is provided with new set of examples so that supervised learning algorithm analyses the training data and produces an correct outcome from labelled data. The supervised machine learning algorithms are those algorithms in which external assistance is required. The input dataset is divided into train and test dataset. Prediction or classification is performed on the train dataset output . Algorithms learn some kind of patterns from the training dataset and are applied to the test dataset for prediction or classification. Three most famous supervised machine learning algorithms have been discussed here.

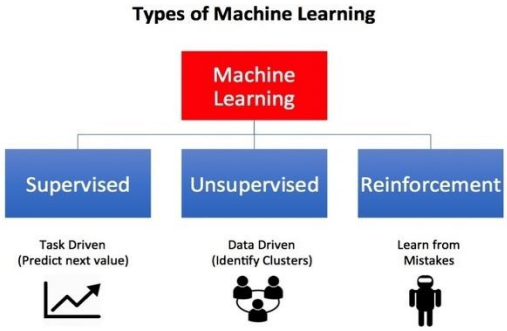

Fig.3. Classification of Machine Learning Algorithms

1. Decision Tree: Decision tree is the most powerful tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.It groups attributes by sorting them based on their values. Each tree consists of nodes and branches. Each nodes represents attributes in a group that is to be classified and each branch represents a value that the node can take.

2) Naïve Bayes: Naïve Bayes also known as Bayesian network mainly targets the text classification industry. It is significantly used for classification and prediction. The conditional probability determines the architecture of Naïve Bayes.

3) Support Vector Machine: Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). It is used for classification. Margin calculation is the principle behind . It basically, draw margins between the classes. It is used in separation of classes.

B. Unsupervised Learning

The unsupervised learning algorithm learns from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. Clustering and feature reduction use unsupervised learning. Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

1) K-Means Clustering: K distinct clusters are built in k-Means clustering algorithm. Clustering or grouping is a type of unsupervised learning technique that when initiates, creates groups automatically. The items which possess similar characteristics are kept in the same cluster.

2) Principal Component Analysis

The main idea of principal component analysis is to reduce the dimensionality of a data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent.

B. Semi - Supervised Learning

Power of both supervised and unsupervised learning is incorporated. where the unlabelled data is already present and getting the labelled data is a tedious process, it can give better results. There are many categories of semi-supervised learning. Some of which are discussed below:

1) Generative Models*:* Generative models are semi-supervised learning method assumes a structure like $p(x,y) = p(y)p(x|y)$ where $p(x|y)$ is a mixed distribution
e.g. Gaussian mixture models. Within the unlabelled data, the mixed components can be identifiable. One labelled example per component is enough to confirm the mixture Distribution.

2) Self-Training: labelled data is used for training the model. The classifier is then fed with unlabelled data. Unlabelled points and the predicted labels are integrated together in the training set. This procedure is further repeated further. Since the classifier is learning itself, hence the name self-training.

3) Transductive SVM: Transductive support vector machine is an extension of SVM. The labelled and unlabelled data both are considered in TVSM. Labelling, the unlabelled data such a way that the margin is maximum between the labelled and unlabelled data. Finding an exact solution by TSVM is a NP-hard problem.

D. Reinforcement Learning

Decisions based on which actions to take such that the outcome is more positive are considered in Reinforcement learning. The learner has no knowledge which actions to take until it's been given a situation. The action performed by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome.

Multitask Learning

The goal of Multitask learning is helping other learners to perform better. When multitask learning algorithms are applied on a task, it remembers the procedure how it solved the problem or how it reaches to the particular conclusion. The algorithm uses these steps to find the solution of other similar problem or task. This helping of one algorithm to another can also be termed as inductive transfer mechanism. If the learners share their experience with each other, the learners can learn concurrently rather than individually and can be much faster.

F. Ensemble Learning

When various individual learners are combined to form only one learner then that particular type of learning is called ensemble learning. The individual learner may be Naïve Bayes, decision tree, neural network, etc. Ensemble learning is a hot topic since 1990s. It has been observed that, a collection of learners is almost always better at

doing a particular job rather than individual learners . Two popular Ensemble learning techniques are given below:

1) Boosting: Boosting is a technique in ensemble learning which is used to decrease bias and variance. Boosting creates a collection of weak learners and convert them to one strong learner. A weak learner is a classifier which is barely correlated with true classification. On the other hand, a strong learner is a type of classifier which is strongly correlated with true classification.

2) Bagging*:* Bagging or bootstrap aggregating is applied where the accuracy and stability of a machine learning algorithm needs to be increased. It is applicable in classification and regression. Bagging also decreases variance and helps in handling over fitting.

G. Neural Network Learning

The neural network (or artificial neural network or ANN) is derived from the biological concept of neurons. A neuron is a cell like structure in a brain. To understand neural network, one must understand how a neuron works. A neuron has mainly four parts. They are dendrites, nucleus, soma and axon.

Supervised Neural Network: In the supervised neural network, the output of the corresponding input is already known. The output which is predicted of the neural network is compared with the actual output. The parameters are changed, based on parameters, and then fed into the neural network again.

H. Instance-Based Learning

In instance-based learning, the learner learns a particular type of pattern. It tries to apply the same pattern to the newly fed data. Hence, the name instance-based. It is a type of lazy learner which waits for the test data to arrive and then act on it together with training data. The complexity of the learning algorithm increases with the size of the data. Given below is a well-

Known example of instance-based learning which k-nearest neighbour is.

1) K-Nearest Neighbour:
In k-nearest neighbour (or KNN), the training data (which is well-labelled) is fed into the learner. When the test data is introduced to the learner, it compares both the data. *k* most correlated data is taken from training set. The majority of *k* is taken which serves as the new class for the test data. The pseudo code for KNN [1].

Machine learning algorithms have given results very well on noisy datasets that are based on financial data [4]. Although other studies have suggested that the performance of classifiers appears to be highly dependent on the nature of the problem and the dataset, the learning algorithms had almost predictive accuracy and all of them performed by a substantial margin simple manager selection rules that are typical of the ways in which money managers and mutual funds are selected by investors.

Naïve Bayes, neural network and decision tree learners each had approximately the same result and all performed very well and by a substantial margin a simple scoring model that is typically on its way in which money managers and mutual funds are selected by investors. These algorithms have worked really well on financial classification process. The conclusion of these tests that machine learning can be used as investing decision making and decision support aid to improve the selection of money managers. Applying different machine learning mechanisms on a simple voting mechanism which is related to the investing decision making can be more accurate than applying a single method. By applying these methods can make a manager more successful. If the machine learning techniques predict the results negatively then it will make the manager to get losses and also may feel regret. So, machine learners should not only focus on eradicating the error or negative results but also must focus on minimizing the costs of the investing manager. It should also focus on the manger's performance. The data set which has words as inputs can be replaced with numerical digits can give a precise result.

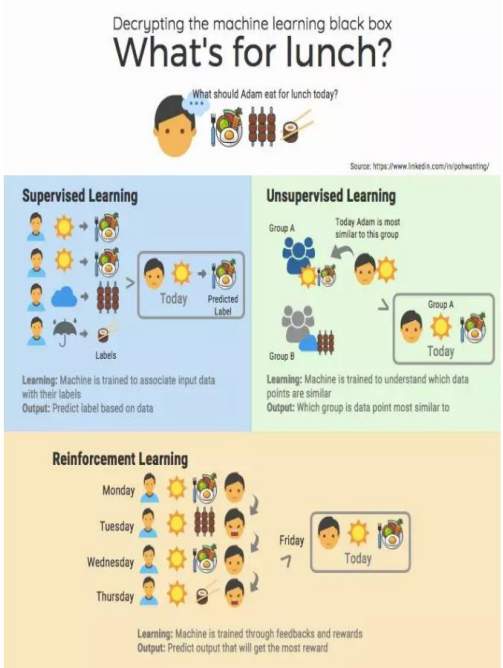

Fig.4. Classification of Machine Learning Algorithms

6. Reinforcement learning

Reinforcement learning deals with optimizing in order to obtain maximum profit in a particular situation. It is employed by various software and machines to find the best possible behaviour or path it should take in a specific situation. They are goal-oriented algorithms, which learn how to attain a complex objective or maximize along a particular dimension over many steps.

7. Neural Network

The idea of ANNs is based on the belief that working of human brain by making the right connections, can be imitated using silicon and wires as living **neurons** and **dendrites**. The human brain is composed of 86 billion nerve cells called neurons. They are connected to other thousand cells by Axons**.** Stimuli from external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural network. A neuron can then send the message to other neuron to handle the issue or does not send it forward. ANNs are composed of multiple nodes, which imitate biological neurons of human brain.

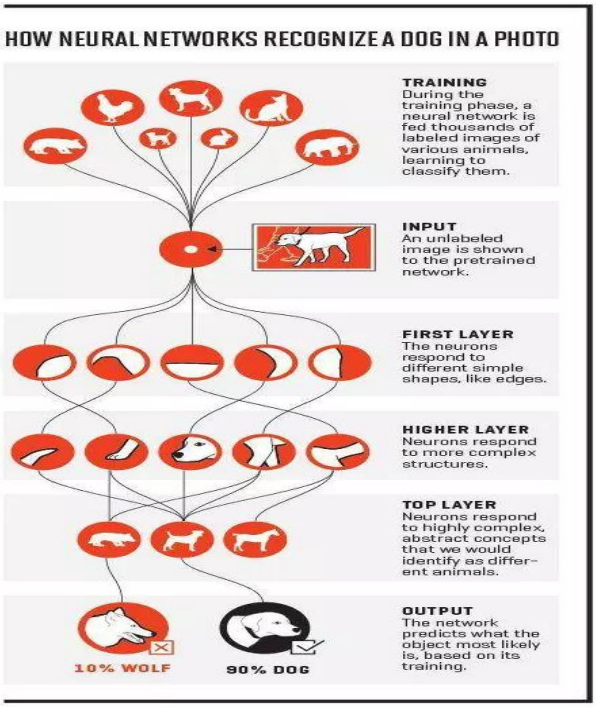

Fig.5. Neural Networks Layers

The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value.Each link is associated with weight. Neural networks have many variations Convolutional Neural Network- it made great breakthroughs in computer vision and is used to analyze visual imagery. Recurrent neural network- The output obtained out of a perceptron is feeded back into the same network to build a recurrent neural network. Recurrent Neural network are significantly used in natural language processing.Fully connected network- it's the easiest model used for process static/tabular data.

8. Over fitting

Production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably is known as over fitting. Deviation would occur when insufficient data is fed to the model.

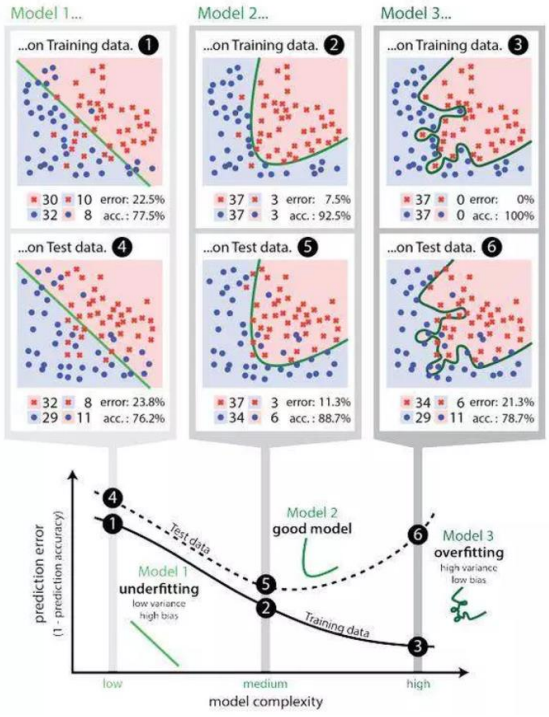

Fig.6. Over fitting

This is a critical problem, yet a common problem. In over fitting, model will take random noises as data input, and take it as an important signal to fit in and this is why model might behave worse in new data (there are deviations in random noises, too). This commonly occurs in complicated

models such as neural networks or acceleration gradient models.

3. Role of Machine Learning in Deep Learning and its Applications

Since 2006 [2], big data has been the greatest emerging technologies of all time in recognising objects, image segmentation and machine translation. During these days data driven machine health monitoring is gaining popularity due to the wide spread deployment of low cost sensors and their connection to the internet. Deep learning provides some tools which will be able to process and analyses those machine data. The applications of deep learning in machine health monitoring systems are reviewed mainly from the following aspects: Auto-encoder and its variants, Restricted Boltzmann Machines and its variants including Deep Belief Network and Deep Boltzmann Machines, Convolutional Neural Networks and Recurrent Neural Networks. The internet of things and data driven techniques are revolutinalizing by enabling the computer networks to gather the data and turn the big machinery data into actionable information. The top down and bottom up approach will help in detection and prediction of failures and from a data we use various machine learning techniques as a solution for the data. As the hottest subfield of machine learning, deep learning is able to act as a bridge connecting big machinery data and intelligent machine health monitoring. As Big data is a branch of machine learning, it tries to model hierarchal representation behind data and predict the patterns using stacking multiple layers of data in hierarchal manner. Big data has already been kept in practice in the fields like face recognition, computer vision, voice recognition etc. Considering the capacity of deep learning to address large scale data and learn multi-level representations. Applications of deep learning in machine health monitoring the conventional MLP has been applied in the field of machine health monitoring for many years. The deep learning techniques have recently been applied to a large number of machine health monitoring systems.

The layer-by-layer pre-training of DNN based on AE or RBM can facilitate the sets of DNN and change its differentiating power to recognize machinery data. Learning critical representations for a person's image is a very hard task. Most of the modern techniques learn deep representations in classification tasks, which essentially reduces the empirical classification risk on the train set. To get a discriminative result they use a deep representation learning procedure named Part Loss Networks, to mimize both representation and classification risks. The representation learning risk is evaluated by the proposed part loss, which automatically detects human body parts, and computes the person classification loss on each part separately. Experimental results on three person ReID datasets, i.e., Market1501, CUHK03, VIPeR, show that our representation outperforms existing deep representations. Person part loss firstly generates K body parts in an unsupervised way, then optimizes the classification loss for each part separately. In this way, part loss network learns discriminative representations for different parts. In that way loss network learns discriminative representations of different parts.

3.1. Multimodal Machine Learning

A human being can experience the surrounding environment in multiple ways like- seeing,hearing,feeling the texture, smelling and tasting different materials.So,modailty also refers to something which happens or is experienced and a testing problem is characterised as multimodal when it also experiences things in the same way as a human. Multimodal machine's main aim is to build them in such a way that they can sense the environment and related to the multi modalities. They go beyond the typical early and late fusion differentiation and identify different challenges that are faced by multimodal machine learning, namely: representation, translation, alignment, and learning [6].

3.2. Deep Learning Based Natural Language Processing

Deep learning methods have multiple processing layers to learn hierarichal representation of data, and have given state of the art results in many dimensions of data sets. Recently a variety of models have been developed in the field of natural language processing. It offers a way to enhance the large data and computation with different methods. Supervised learning is one of the most used techniques for learning deep learning research In NLP. All the other schemas are in their developing state but we can expect the Deep learning based use of the unlabelled data. We can expect the more advancement in this model designs .So, at the end we can expect more deep learning models whose internal memory is linked to the external memory and Coupling symbolic and sub-symbolic AI will be the first stepping stone in the path from NLP for better understanding of Natural languages. It no only requires that but also relying on machine learning can make a good guess based on the sub symbolic methods encode correlation and decision making. To use Noam Chomsky's rule "we have to be more theoritical,but not to collect some data and doing statistical analysis on them". We have to be more theoritical than analytical [8].

4. Integration of Predictive Intelligence with Social Media Data

A key analytical task across many domains is model building and exploration for predictive analysis. Data is collected, parsed and analyzed for relationships, and features are selected and mapped to estimate the response of a system under exploration. As social media data has grown more abundant, data can be captured that may potentially represent behavioral patterns in society. In turn, this unstructured social media data can be parsed and integrated as a key factor for predictive intelligence. In this paper, we present a framework for the development of predictive models utilizing social media data. We combine feature selection mechanisms, similarity comparisons and model cross-validation through a variety of interactive visualizations to support analysts in model building and prediction. In order to explore how predictions might be performed in such a framework, we present results from a user study focusing on social media data as a predictor for movie box-office success [10].

4.1 Natural Language Processing Generation

The main functionality of Natural language processing came is to ease the user's work and to satisfy the wish to communicate with the computer in natural language. Since all the users may not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it. Natural Language Processing basically can be classified into two parts i.e. *Natural Language Understanding* and *Natural Language Generation* which evolves the task to understand and generate the text. The 'levels of language' are one of the most explanatory method for representing the Natural Language processing which helps to generate the NLP text by realizing Content Planning, Sentence Planning and Surface Realization phases SVM(support vector machine), J48 (decision tree), Bayesian with Adaboost, Naïve Bayes with Adaboost). The concerned classifiers are tested and evaluated on metric (such as F-measure (accuracy), False Positive Rate, and training time). SVM is the best classifier to be used. It has the high accuracy and the low false positive rate. However, training time of SVM to build the model is high, but as the results on other parameters are positive, the time does not pose such an issue.

5. Methodology

Machine Learning algorithms determine the for the given loan amount what is the loan amount term period is given by the Banks and the Customers loan status with the bank data set. Bank data set contains for Classification Algorithm Correctly classified instances are 422 for ZeroR algorithm. compared to Random Forest Tree with ZeroR algorithm is taking less time and it is forming 4 clusters for EM Clustering algorithm. To pre-process and analyse the data we have used WEKA tool and implemented EM Clustering Algorithm, ZeroR and Random Forest algorithm.

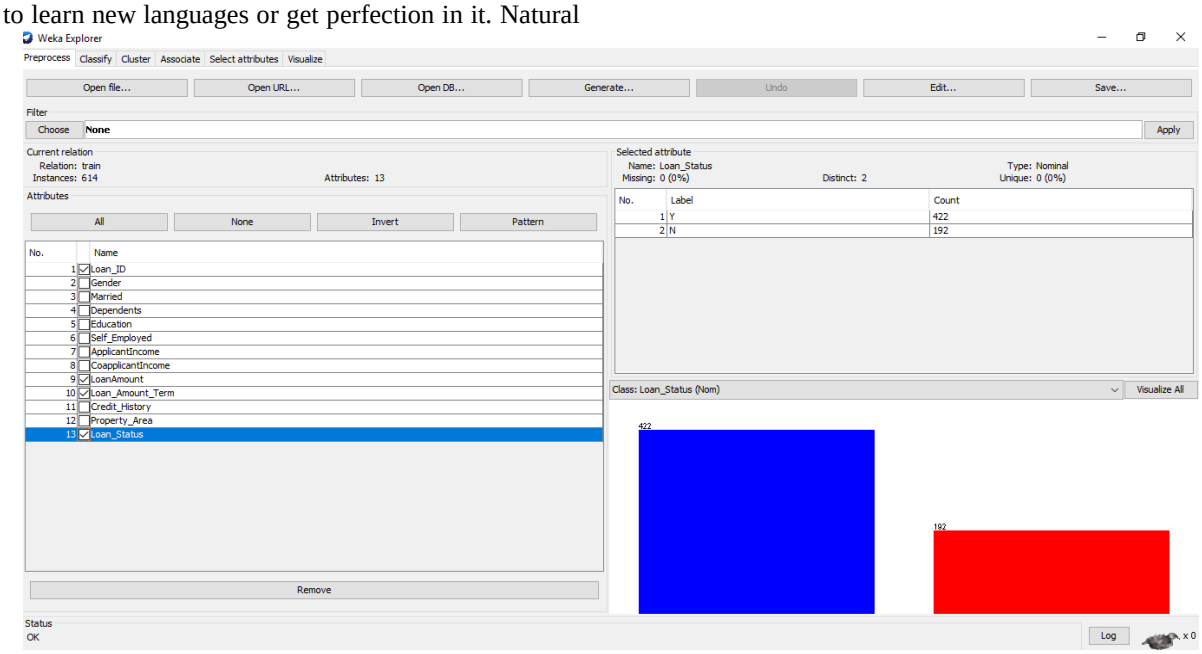

Fig.7.Preprocessing of Bank Data set .CSV File

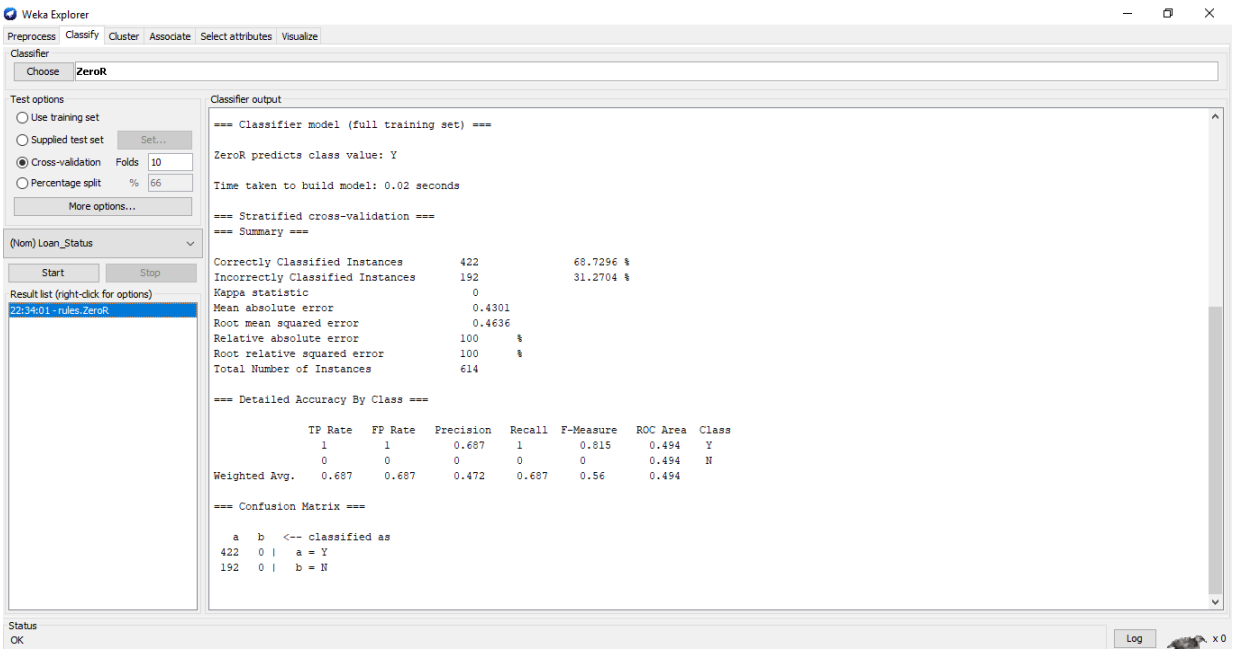

Fig.8. Classification Algorithm-ZeroR

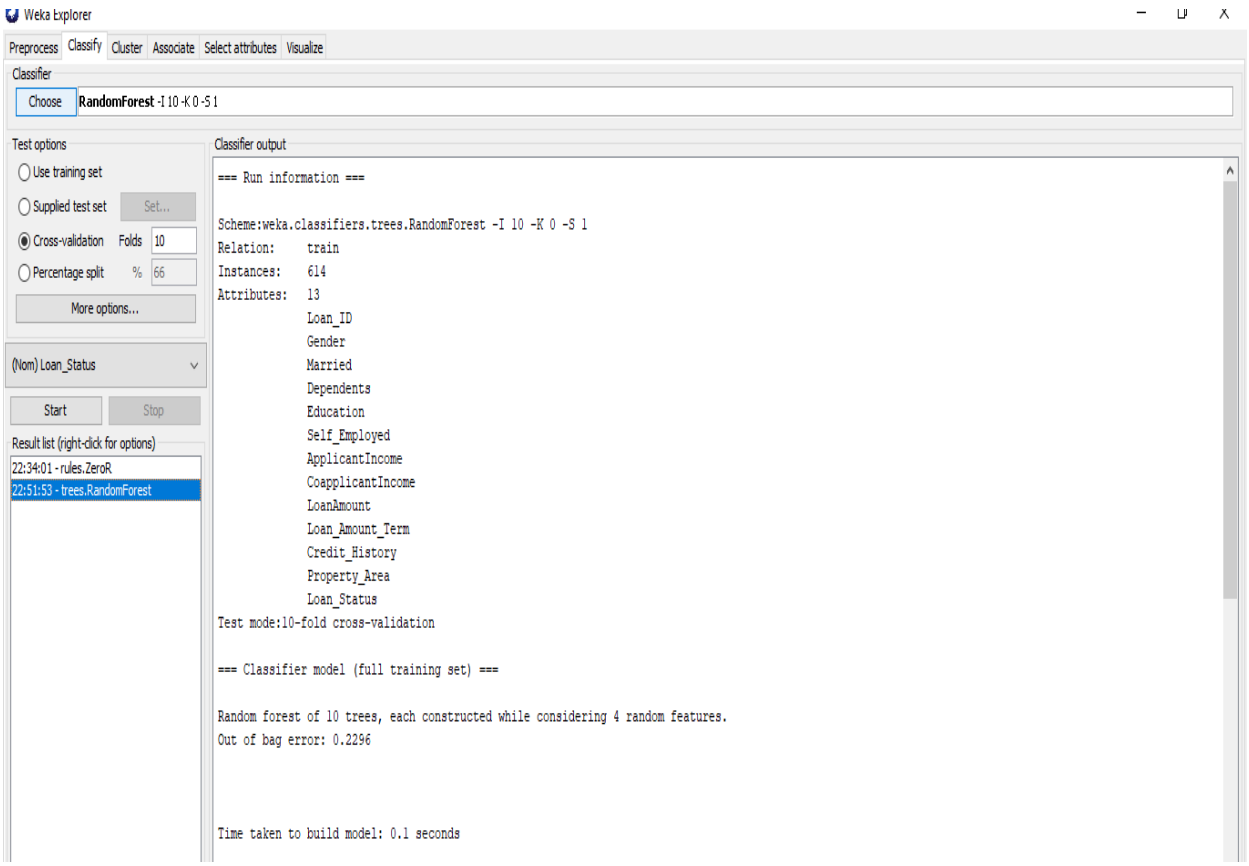

Fig.9. Random Forest Algorithm

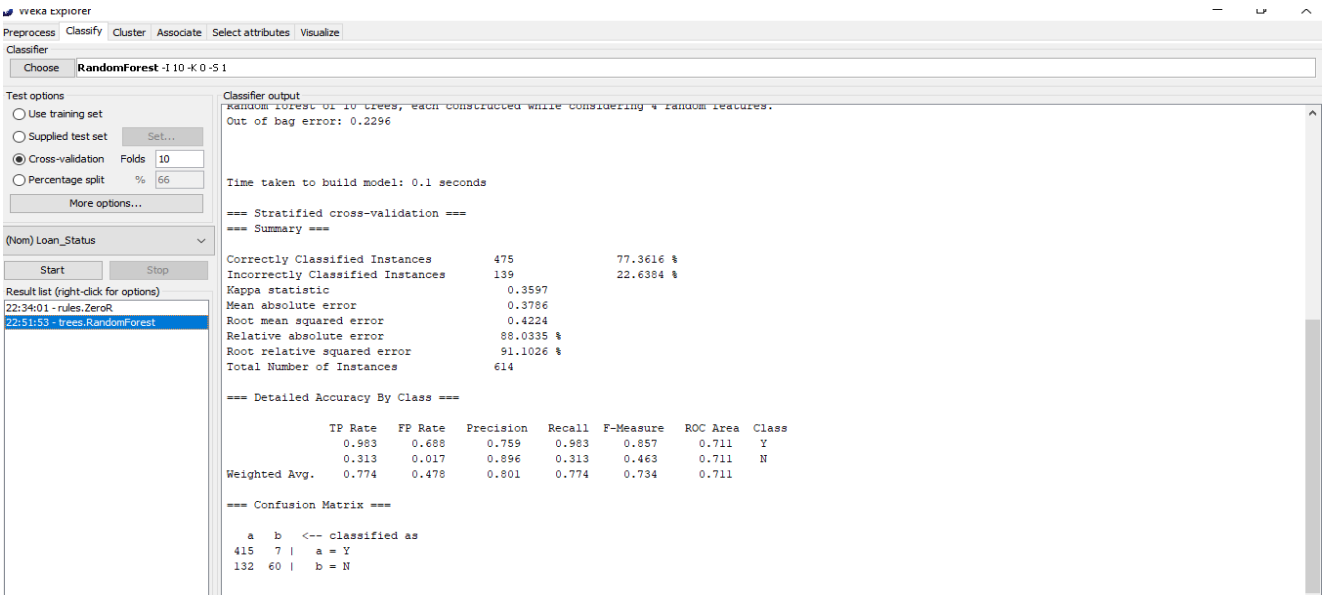

Fig.10. Random Forest

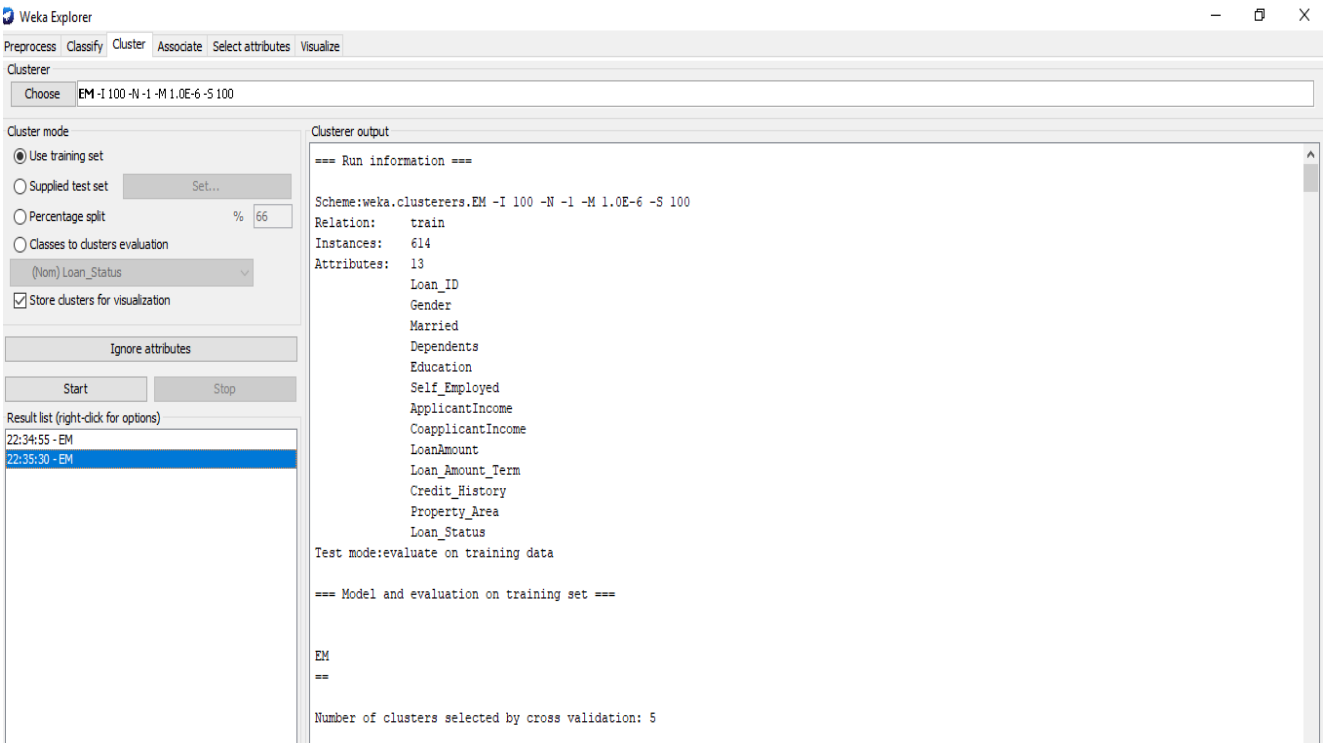

Fig.11.Clustering-EM

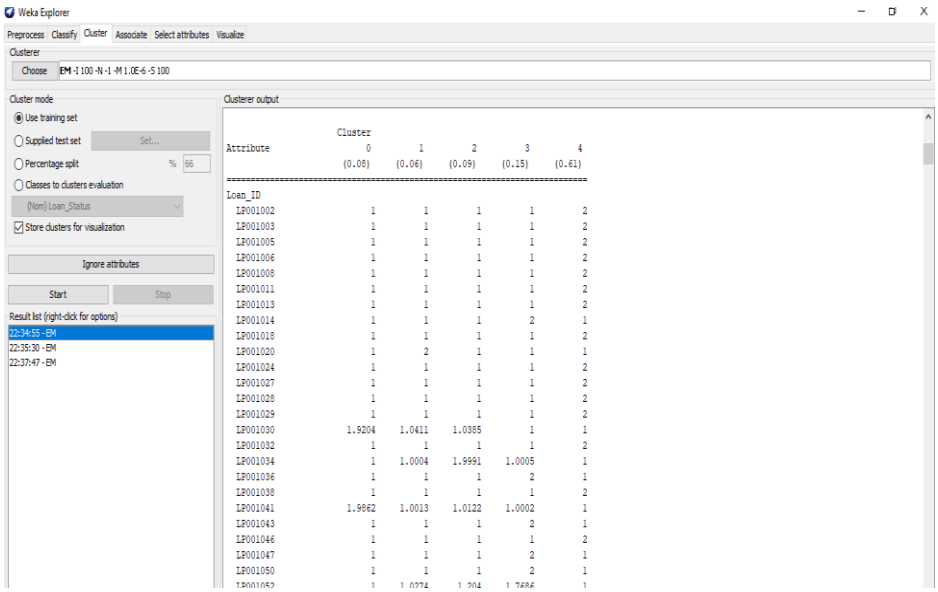

Fig.12.Clustering-EM

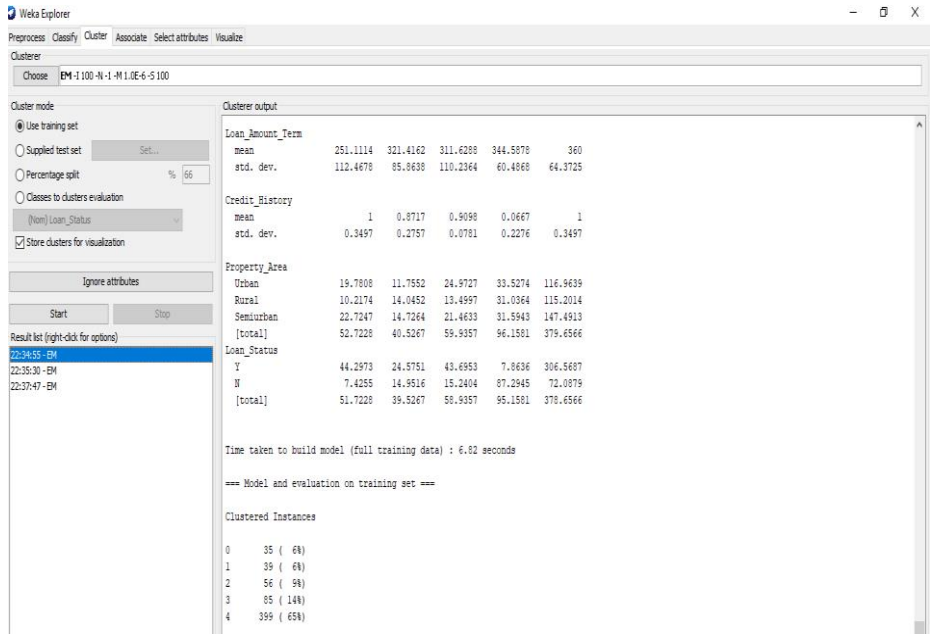

Fig.13.Clustering-EM

6. Conclusions

Due to large amount of data is available everywhere. Therefore, it is very important to analyse this data in order to extract some useful information and to develop an algorithm based on this analysis. Classification of Machine Learning Algorithms, Role of Machine Learning in Deep Learning and its Applications, A Comparison of Machine-Learning Classifiers, Multimodal Machine Learning, Deep Learning Based Natural Language Processing, Natural Language Processing Generation, Integration of Predictive Intelligence with Social Media Data. Implemented the Classification and Clustering Techniques and compared the results with EM Clustering algorithm ,ZeroR algorithm and RandomForest Algorithm for the given attributes. It has taken less time when compared to Random Forest and EM Clustering Algorithm. Finally , ZeroR is the Best algorithm Compared to EM and Random Forest Algorithm.

.

References

[1] https://www.octoparse.com/blog/8-keywords-must-know-for-machine-learning.

[2] Rui Zhao, et al,"Deep learning and its applications to machine health Monitoring", Mechanical Systems and Signal Processing 115 (2019) 213–237,elsevier.

[3] Hantao Yao, Shiliang Zhang,"Deep Representation Learning with Part Loss for Person Re-Identification". 4 Jul 2017

[4] Robert S. Ludwig," A Comparison of Machine-Learning Classifiers For Selecting Money Managers", Intelligent Systems in Accounting Finance & Management 13(3):151-164 · July 2005 , DOI: 10.1002/isaf.262

[5] REFRENCE: Marinka Zitnik,"Machine Learning for Integrating Data in Biology and Medicine:Principles, Practice, and Opportunities", *10 Oct 2018,* Information Fusion 50 (2019) 71-91, DOI:

[6] Tadas Baltruˇsaitis," Multimodal Machine Learning:A Survey and Taxonomy", IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 41 , Issue: 2 , Feb. 1 2019).

[7] Diksha Khurana ," Natural Language Processing: State of The Art, Current Trends and Challenges", August 2017

[8] Tom Young," Recent Trends in Deep Learning Based Natural Language Processing", 9 Aug 2017.

[9] Shrawan Kumar Trivedi," A Study of Machine Learning Classifiers for Spam Detection", 2016 4th International Symposium on Computational and Business Intelligence, 5-7 Sept. 2016, DOI: 10.1109/ISCBI.2016.7743279

[10] REFRENCE: Yafeng Lu, Robert Kr¨uger, "Integrating Predictive Analytics and Social Media", Proc. In 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), 25-31 Oct. 2014, DOI: 10.1109/VAST.2014.7042495, 978-1-4799-6227-3

First Author M Ashish Naidu,Currently pursuing III B.Tech Computer Science at GITAM University,Hyderabad. My Research areas are Data Mining,Information Retrieval Systems,Big Data Analytics,Machine Learning.

Second Author K Radha working as an Asst Professor at GITAM University,Hyderabad. She has Completed M.Tech(CSE) at JNTUH,Pursuing PhD at KL University,Vijayawada.She has 12 years of Teaching Experience and 1Year Industrial Experience.She has published numerous research papers and presented at Various conferences.She is a Member of IAENG.