

# [Winter it.it’s a feed forward neural network](https://assignbuster.com/winter-itits-a-feed-forward-neural-network/)

Winter Internship ReportDesigning a Part Of Speech TaggerBySudhanshu Srivastava1506041NIT Patna, BiharUnder the Guidance ofDr. A.

K. SinghDepartment of Computer Science & EngineeringINDIAN INSTITUTE OF TECHNOLOGY (BANARAS HINDU UNIVERSITY)VARANASI – 221005Artificial IntelligenceIt could be taken as the superset of machine learning which itself is a superset ofdeep learning. On a frank scale, it could be said as the Technology which gives amachine human like computational approach. Natural language processingA branch of Artificial Intelligence which deals with the way of communicatingwith a machine/intelligent system with any natural language like English orHindi.

Machine learningGiving a computer the ability to learn without being explicitly programmed onthat very interest. Basically, training a system on the past so that it could predictthe output of present/future. It has two Sub branches –o Supervised Learningo Unsupervised learningMachine learning is the superset of Deep learning. Deep learningThe machines generate their features by themselves, basically forming Algorithmsto mimic human brain.

It is implemented through neural networks which has a basic unit calledperceptron which is the functional unit of the neural networks. The basic Structure of a perceptron. At first the weights are randomly assigned tothe inputs. Back propagation methodCompares the output with the given output and changes the weightcorrespondingly.

Multiple neural network with several hidden layers constitute of deep networkFeed forward networksNetworks that are not cyclic in nature, i. e. the outputs are independent of eachother. Convolutional neural networkHere, a neuron in a layer is only connected to a small region of the layer before it. It’s a feed forward neural network inspired from the visual cortex. Recurrent neural networksThe neural network in which the present output depends on the previous outputs(Could be understood as an analogy to Dynamic programming).

Basic structure of a RNNThere are some limitations with RNNVanishing gradient problemWhen the change in weight is very very small i. e(;;;; 1), it corresponds to(de/dw);;; 1. The new weight is almost equal to the old one. This is removed by using another neural network known as LONG SHORT TERMMEMORY NETWORKS(LSTMs)Long short term memory networks(lstm)RNN equipped with long term dependencies. WORD2VECA model that predicts between a center word and context words in terms of wordvectors. It comprises of two models:? Skip – Gram model? Continuous Bag of words modelTaskDesigning a Part of Speech tagger.

DatasetA merged Bhojpuri dataset containing of sentences of Bhojpuri and thecorresponding labels to the words. A sample of the dataset. Tools used? Python 3? Keras? Tensor Flow BackendAfter having a thorough understanding of the above listed topics. I have firsttaken the Word2vec Embeddings of the words with their correspondingsentences. So, I have extracted a sentence and then created the vector word by word. Theimplementation could be taken as a 2D array with sentences and words. The very same I have done with the labels, I have created a 2D array of thecorresponding words in the sentences. A dictionary is being used to map the words and the corresponding labels. For the label Vector Part, The total different tags were used to create the one hot vector, The total numberof different labels are 29 in number and namely are:'NNP', 'NN', 'PSP','NST','VM','JJ','RB','RP','CC','VAUX','SYM','RDP','QC','PRP','QF','NEG','DEM','RDP','WQ','INJ','CL','ECH','UT','INTF','UNK','NP','VGF','CCP','BLK'Another dictionary is used to map the labels to the vectors. Now, we have to take a sample test data, train the lstm model on that and thenpredict it on test values. We have encoded the test vector and labels of the test dataset as well which wehave used as the validation data. A sequential model has been taken and as the size of the sentence with maximumwords came out to be 226Lstm was trained with an input shape of 226\*100 as the vector size is 100 and themaximum size is 226 with the return sequences as True. 29 was passed to the Dense function as there are 29 different tags. After being trained in lstm attention mechanism is applied.