

SOFT COMPUTING
DIGITAL ASSIGNMENT-1
NAME-ANISH SINGH WALIA
REG-No:15BIT0116
FACULTY: PROF LAKSHMI PRIYA GG

PROJECT TITLE-HUMAN RESOURCE ANALYTICS

1)Dataset name-Human resource dataset. The dataset has 10 attributes(variables) and 14999 observations(instances).

```
data.frame': 14999 obs. of 10 variables:
  ATTRIBUTES      TYPE  INSTANCES
 $ satisfaction_level : num  0.38 0.8 0.11 0.72 0.37 0.41 0.1 0.92 0.89 0.4
2 ...
 $ last_evaluation    : num  0.53 0.86 0.88 0.87 0.52 0.5 0.77 0.85 1 0.53
...
 $ number_project     : int   2 5 7 5 2 2 6 5 5 2 ...
 $ average_monthly_hours : int  157 262 272 223 159 153 247 259 224 142 ...
 $ time_spend_company  : int   3 6 4 5 3 3 4 5 5 3 ...
 $ Work_accident       : int   0 0 0 0 0 0 0 0 0 0 ...
 $ left               : Factor w/ 2 levels 1 1 1 1 1 1 1 1 1 1 ...
 $ promotion_last_5years: int   0 0 0 0 0 0 0 0 0 0 ...
 $ sales              : Factor w/ 10 levels "accounting","hr",...: 8 8 8 8 8
8 8 8 8 8 ...
 $ salary             : Factor w/ 3 levels "high","low","medium": 2 3 3 2 2
2 2 2 2 2 ...
```

Sample Instances:

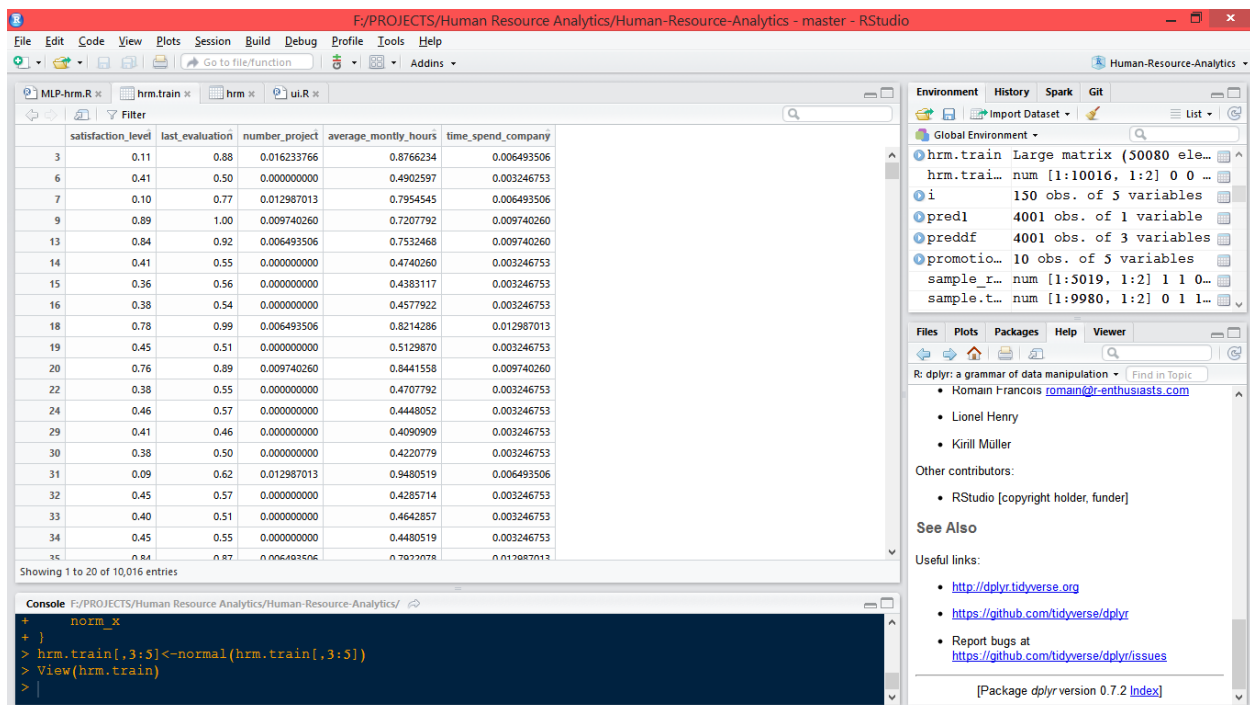
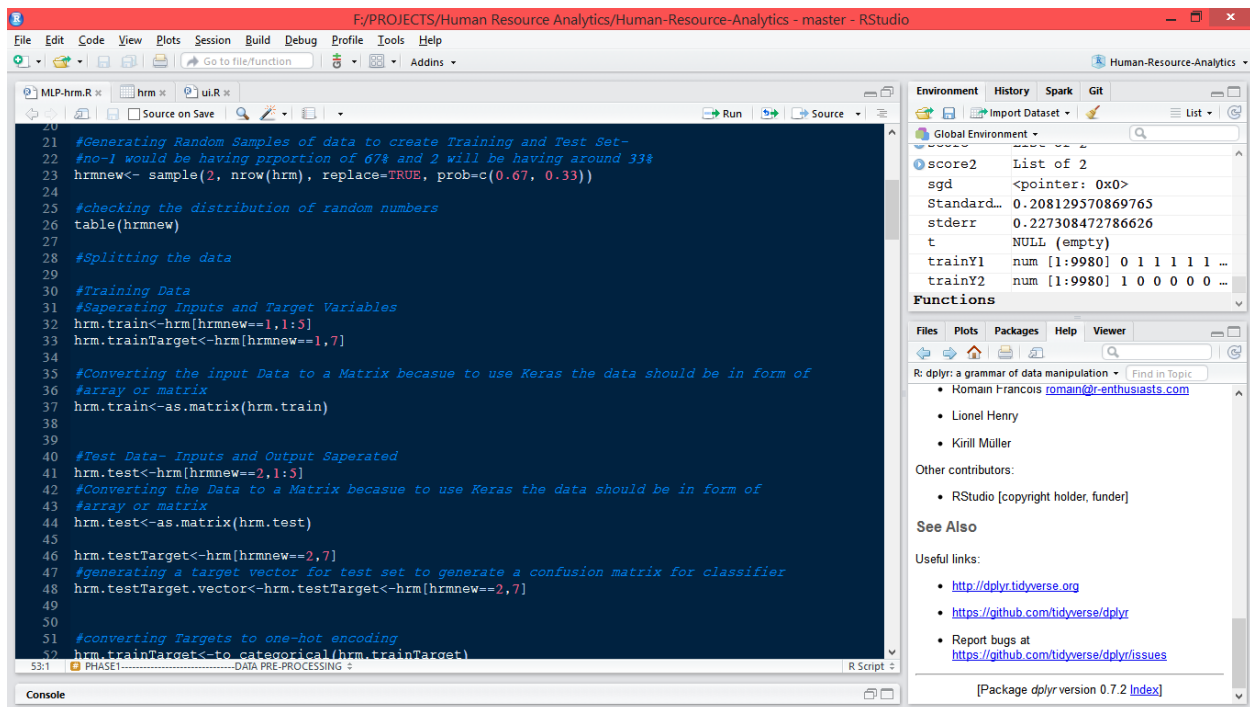
	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spent_company	Work_accident	left	promotion_last_5years	sales	salary
1	0.38	0.53	2	157	3	0	1	0	sales	low
2	0.80	0.86	5	262	6	0	1	0	sales	medium
3	0.11	0.88	7	272	4	0	1	0	sales	medium
4	0.72	0.87	5	223	5	0	1	0	sales	low
5	0.37	0.52	2	159	3	0	1	0	sales	low
6	0.41	0.50	2	153	3	0	1	0	sales	low
7	0.10	0.77	6	247	4	0	1	0	sales	low
8	0.92	0.85	5	259	5	0	1	0	sales	low
9	0.89	1.00	5	224	5	0	1	0	sales	low
10	0.42	0.53	2	142	3	0	1	0	sales	low
11	0.45	0.54	2	135	3	0	1	0	sales	low
12	0.11	0.81	6	305	4	0	1	0	sales	low
13	0.84	0.92	4	234	5	0	1	0	sales	low
14	0.41	0.55	2	148	3	0	1	0	sales	low
15	0.36	0.56	2	137	3	0	1	0	sales	low
16	0.38	0.54	2	143	3	0	1	0	sales	low
17	0.45	0.47	2	160	3	0	1	0	sales	low
18	0.78	0.99	4	255	6	0	1	0	sales	low
19	0.45	0.51	2	160	3	1	1	1	sales	low
20	0.76	0.89	5	262	5	0	1	0	sales	low
21	0.11	0.83	6	282	4	0	1	0	sales	low
22	0.38	0.55	2	147	3	0	1	0	sales	low
23	0.09	0.95	6	304	4	0	1	0	sales	low
24	0.46	0.57	2	139	3	0	1	0	sales	low

DATA PRE-PROCESSING

- 1) **Data preprocessing** is done so as to make data suitable to feed it to a neural network.
- 2) Random sampling of data is done- i.e it is shuffled.
- 3) Normalization operations and scaling numeric variables to range [0,1] to reduce the computational complexity for the model.

```
#Min-max normalization function in R
normal<-function(x)
+ {
+   norm_x<-(x-min(x))/(max(x)-min(x))
+   norm_x
+ }
```

- 4) Separating Training and test data and selecting 5 best variables which explain and summarize most information in dataset.
- 5) Converting the data into **matrix** form so as to feed it to a neural network.



Normalized training data-i.e between range [0,1]

3) Training a Multi-layer Deep Neural network

Problem Statement- Predicting whether an employee will *leave the company* or not based on five inputs after basic data analytics and exploring the dataset.

MLP architecture-

Training inputs instances: 9980 observations

Layers in Neural network- 4 layers

Input Units – 5 input neurons as we have 5 input variables

Hidden layer 1 – 32 hidden neurons

Hidden layer 2 – 10 hidden neurons

Output Units — 2 output neurons (as we have 2 class labels)

The activation function used in order to learn non-linear interactions by the *hidden layers* are **ReLU**-rectified linear units and this function is of form- **$\max(0, x)$** , i.e. it converts negative values to 0.

Diagram of the architecture--

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 32)	192
dense_2 (Dense)	(None, 10)	330
dense_3 (Dense)	(None, 2)	22

Total params: 544

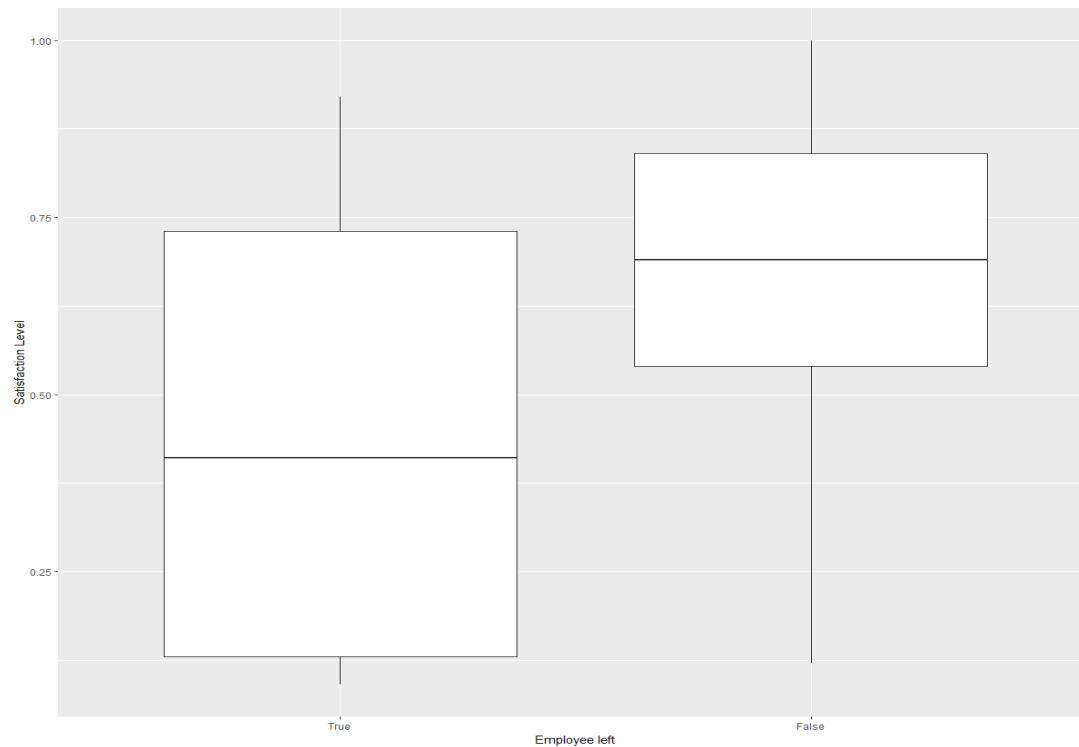
Trainable params: 544

Non-trainable params: 0

Hence the number of trainable parameters are **544 i.e weights and bias** as they are learned and tuned during the learning process of the model from training data.

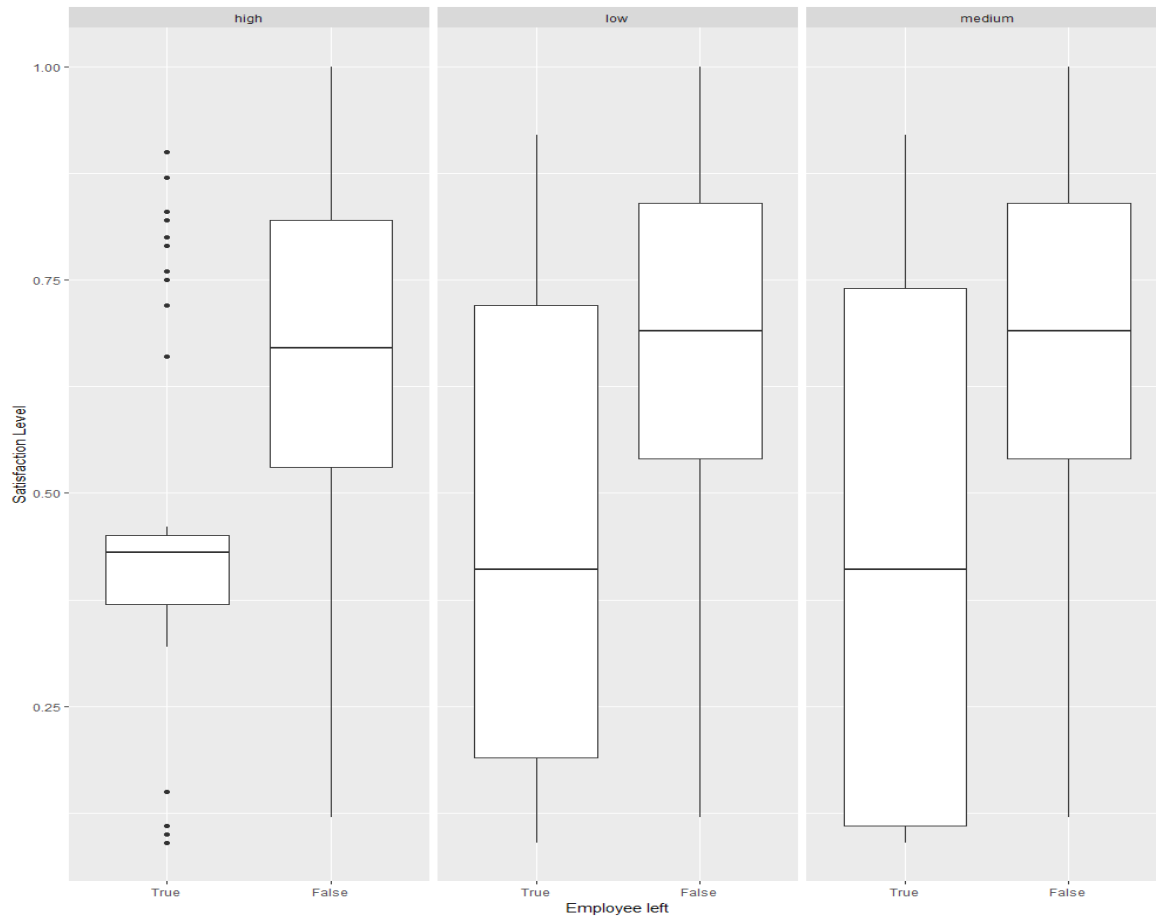
4) **DESCRIPTIVE ANALYSIS** of dataset

BOX PLOTS obtained while analyzing the dataset.

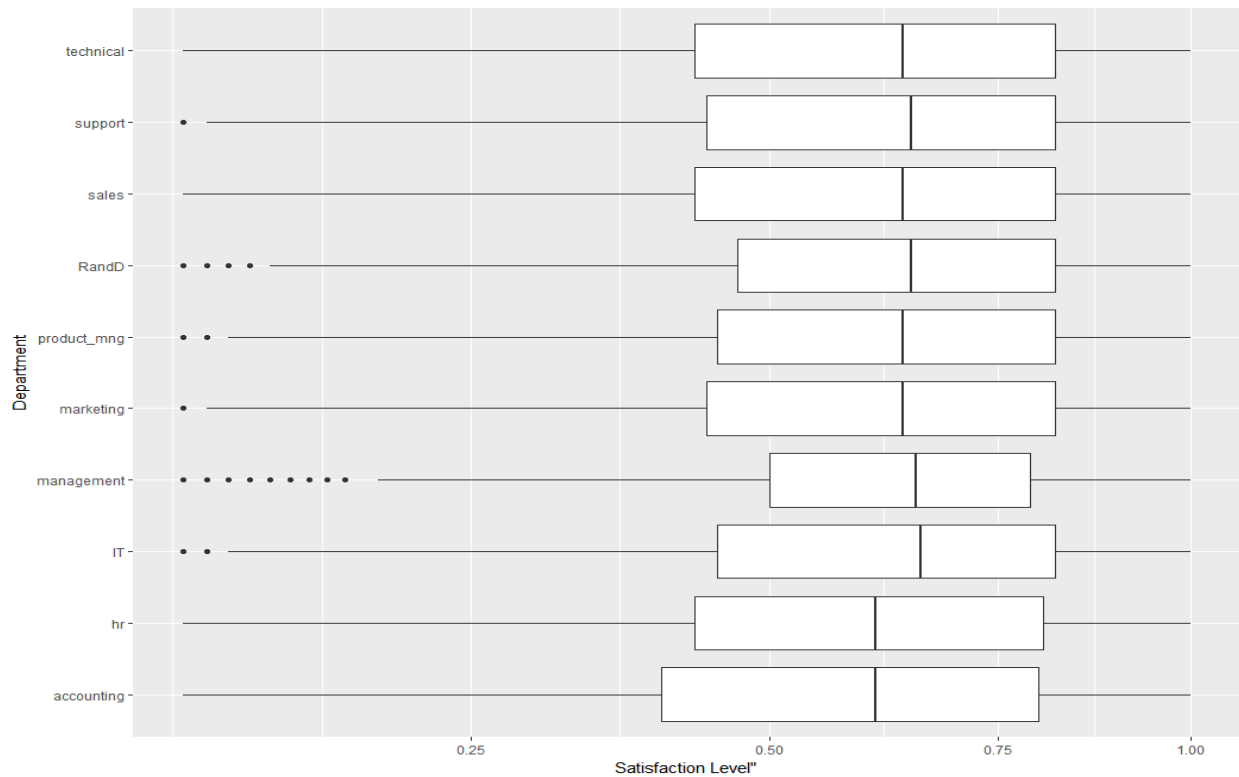


The plot above shows the boxplot which explains the statistical parameters such as median, quantiles and percentiles of a categorical variable and a numeric variable.

In the above plot we can see that employee who left have lower median satisfaction levels of around 0.35. Employee who didn't leave the company have higher median satisfaction levels of around 0.60.



This boxplot shows the multi-variate analysis of three different variables- Employee_left, Satisfaction level and salary.



Boxplot of satisfaction level of employees department wise

One can notice that the highest satisfaction level is for employees in IT, management and R & D departments.