

Topic: Neural Network

Instructions

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

Name: Nukala Ayyappa Bharthwaj

Batch Id: DSWDMCOS 21012022

Topic: Neural Network

1. Business Problem

1.1. Objective

1.2. Constraints (if any)

2. Work on each feature of the dataset to create a data dictionary as displayed in the below image:

Name of Feature	Description	Type	Relevance
ID	Customer ID	Quantitative, Nominal	Irrelevant, ID does not provide useful information

2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.

Using R and Python codes perform:

3. Data Pre-processing

3.1 Data Cleaning, Feature Engineering, etc.

3.2 Outlier Imputation if applicable.

4. Exploratory Data Analysis (EDA):

4.1. Summary

4.2. Univariate analysis

4.3. Bivariate analysis

5. Model Building:

5.1 Perform Artificial Neural Network on the given datasets.

5.2 Use TensorFlow keras to build your model in Python and use Neural net package in R

5.3 Briefly explain the output in the documentation for each step in your own words.

5.4 Use different activation functions to get the best model.

6. Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

The assignment should be submitted in the following format:

- R code
- Python code
- Code Modularization should be maintained
- Documentation of the model building (elaborating on steps mentioned above)

Problem Statement:-

- 1.) WE have Dataset which contains the details of 50 startup's . Predicts the profit of a new Startup based on certain features. To Venture Capitalists this could be a boon as to whether they should invest in a particular Startup or not. So Build a Neural Network model to predict profit and which startup's end up performing better. By seeing that if they spent more money on marketing or was it their stellar R&D department which led them to this huge profit and in turn huge fame and success.

	R.D.Spend	Administration	Marketing.Spend	State	Profit
1	165349.20	136897.80	471784.10	New York	192261.83
2	162597.70	151377.59	443898.53	California	191792.06
3	153441.51	101145.55	407934.54	Florida	191050.39
4	144372.41	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94
6	131876.90	99814.71	362861.36	New York	156991.12
7	134615.46	147198.87	127716.82	California	156122.51
8	130298.13	145530.06	323876.68	Florida	155752.60
9	120542.52	148718.95	311613.29	New York	152211.77
10	123334.88	108679.17	304981.62	California	149759.96
11	101913.08	110594.11	229160.95	Florida	146121.95
12	100671.96	91790.61	249744.55	California	144259.40
13	93863.75	127320.38	249839.44	Florida	141585.52
14	91992.39	135495.07	252664.93	California	134307.35
15	119943.24	156547.42	256512.92	Florida	132602.65
16	114523.61	122616.84	261776.23	New York	129917.04

Sol:

Business Objective: To predict the profits of the startups company with other factors by using Neural Networks model.

Constraints: Lack of analysis of the startups company data.

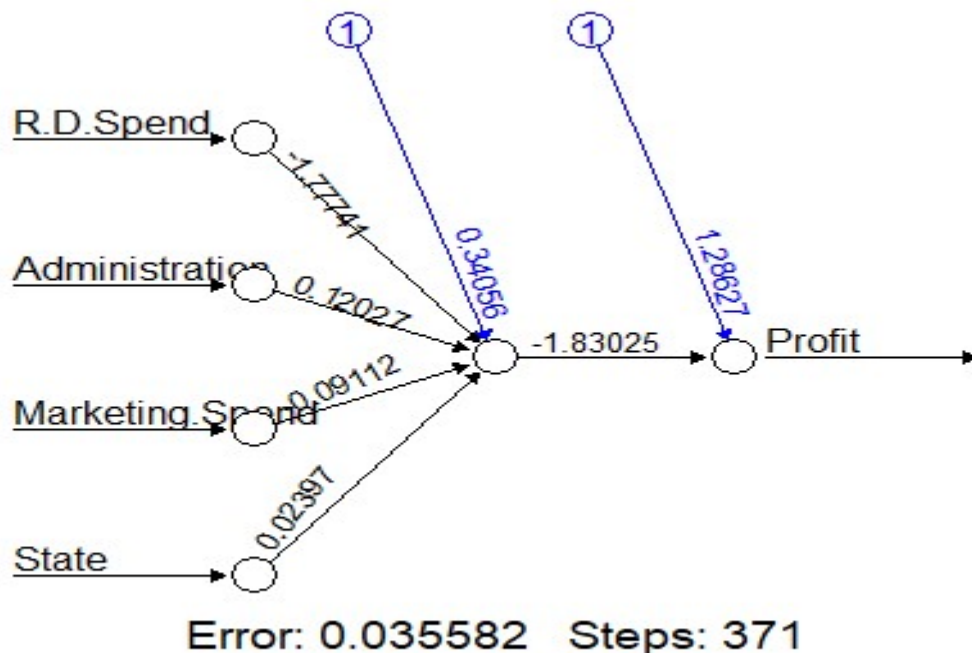
Data Types: All the given data and its types are as follows:

Name of feature	Description	Data Type	Relevance
R&D spend	Amount spent on R&D	Ratio	Relevant
Administration	Amount spend on Administration	Ratio	Relevant
Marketing Spend	Amount spend on marketing	Ratio	Relevant
State	State of the start-up company	Nominal	Relevant
Profit	Profit amount of the company	Ratio	Relevant

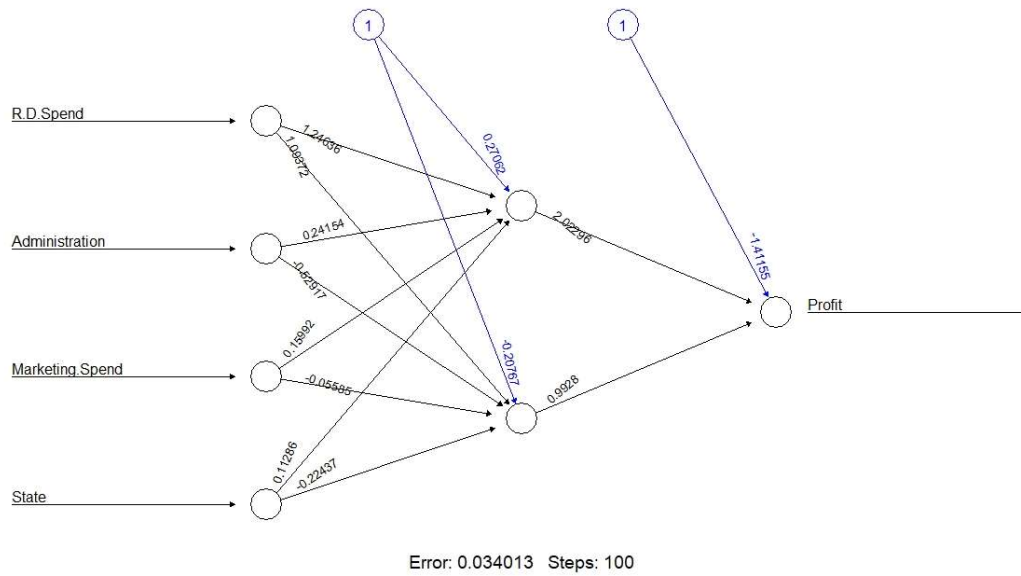
Data Cleaning: Since the states column is in Non-Numeric format the same is converted to numeric data for doing the analysis.

Exploratory Data Analysis: Normalization on the data is applied in order to use the data for doing the analysis.

Neural-Network: After applying the normalization on the data used for doing the neural network model by using neuralnet in R and tensorflow package in Python. Initially neural network is applied without any hidden layers and the accuracy of the model is 95.66% and the network plot of the model is as follows:



After applying neural network with the hidden layers and the accuracy of the model is 96.39% and the network is as follows:



Problem Statement:-

- 2.) We have Dataset listed 517 fires from the Montesano natural park in Portugal. For each incident, weekday, month, coordinates, and the burnt area are recorded, as well as several meteorological data such as rain, temperature, humidity, and wind. Predict the burned area of forest fires with Neural Networks. The model is then used to predict the burnt area based on the current date and the coordinates, where the fire is spotted. This prediction can be used for calculating the forces sent to the incident.

#	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	dayfri	daymon	daysat	daysun	c
1	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00	1	0	0	0	0
2	oct	tue	90.6	35.4	689.1	6.7	18.0	33	0.9	0.0	0.00	0	0	0	0	0
3	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00	0	0	1	0	0
4	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00	1	0	0	0	0
5	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00	0	0	0	0	1
6	aug	sun	92.3	85.3	488.0	14.7	22.2	29	5.4	0.0	0.00	0	0	0	0	1
7	aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0.0	0.00	0	1	0	0	0
8	aug	mon	91.5	145.4	608.2	10.7	8.0	86	2.2	0.0	0.00	0	1	0	0	0
9	sep	tue	91.0	129.5	692.6	7.0	13.1	63	5.4	0.0	0.00	0	0	0	0	0
10	sep	sat	92.5	88.0	696.6	7.1	22.8	40	4.0	0.0	0.00	0	0	1	0	0
11	sep	sat	92.5	88.0	696.6	7.1	17.8	51	7.2	0.0	0.00	0	0	1	0	0
12	sep	sat	92.8	73.2	713.0	22.6	19.3	38	4.0	0.0	0.00	0	0	1	0	0
13	aug	fri	63.5	70.8	665.3	0.8	17.0	72	6.7	0.0	0.00	1	0	0	0	0
14	sep	mon	90.9	126.5	696.5	7.0	21.3	42	2.2	0.0	0.00	0	1	0	0	0
15	sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0.0	0.00	0	0	0	0	0
16	sep	fri	93.3	141.2	713.9	13.9	22.9	44	5.4	0.0	0.00	1	0	0	0	0

Sol:

Business Objective: To predict the forest area with other factors using Neural Network model.

Constraints: Lack of analysis of the previous forest fire data.

Data Types: Given data and its types are as follows:

Name of feature	Description	Data Type	Relevance
Month	Month when fire occurred	Nominal	Relevant
Day	Day when fire occurred	Nominal	Relevant
FFMC	Fine Fuel Moisture Code for moisture content in fire data	Ratio	Relevant
DMC	Duff Moisture Code for moisture content in fire data	Ratio	Relevant
DC	third moisture index for moisture content in fire data	Ratio	Relevant
ISI	Initial Spread Index of the fire data	Ratio	Relevant
Temp	Temperature of the area	Ratio	Relevant
RH	Relative humidity of the fire area	Ratio	Relevant
Wind	Wind speed of the fire area	Ratio	Relevant
Rain	Rain forecast of the fire area	Ratio	Relevant
Area	Area of the fired region	Ratio	Relevant
Day type	All the days are converted	Nominal	Relevant

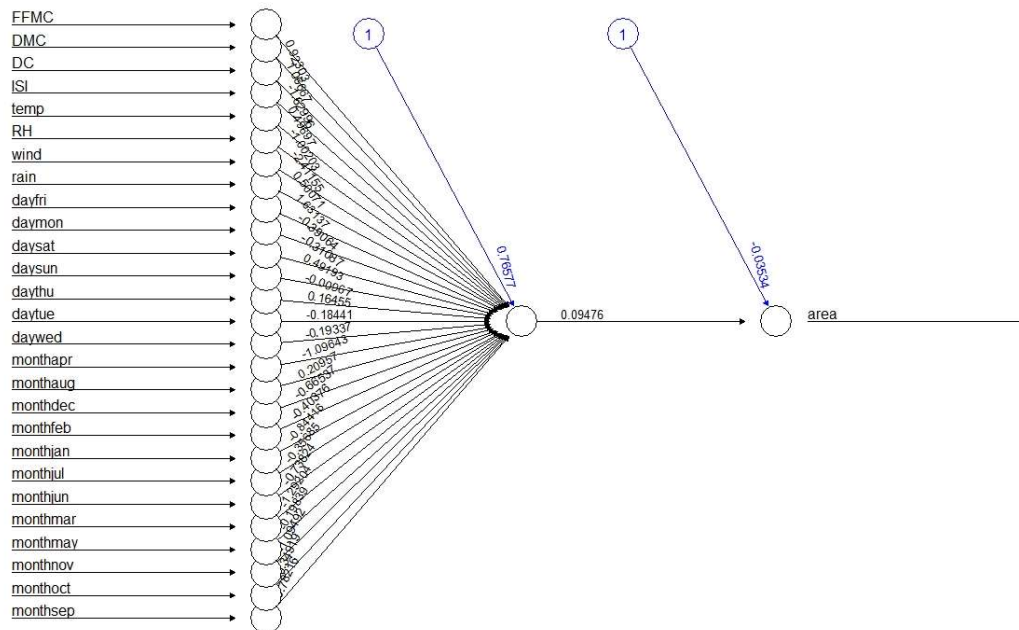
	into 1 or 0 based on type of day		
Month type	All the months are converted into 1 or 0 based on type of months	Nominal	Relevant

Data Pre-Processing:

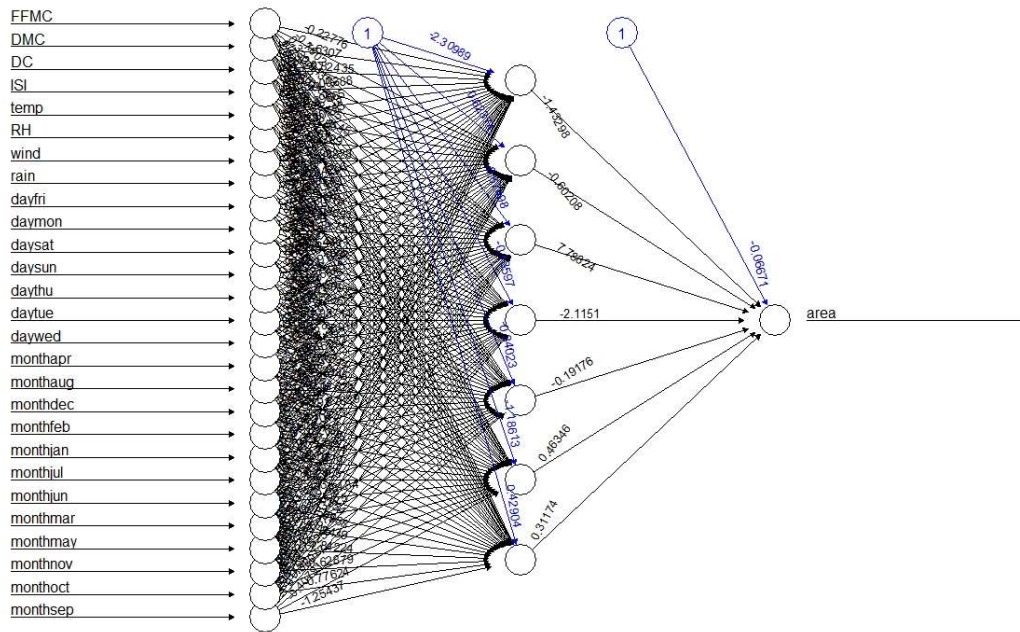
Since first two columns are repeated which are month and the day the same are removed for doing the analysis on the data.

EDA: after observing the data all the variables are of different quantity so normalization on the data is applied so that it will be used for doing the analysis.

Neural-Network: After applying the normalization on the data used for doing the neural network model by using neuralnet in R and tensorflow package in Python. Initially neural network is applied without any hidden layers and the accuracy of the model is 80.46% and the network plot of the model is as follows:



After applying neural network with the hidden layers and the accuracy of the model is 96.39% and the network is as follows:



Problem Statement:-

- 3.) The following dataset consists of 1030 instances with 9 attributes and has no missing values. There are 8 input variables and 1 output variable. Seven input variables represent the amount of raw material (measured in kg/m^3) and one represents Age (in Days). The target variable is Concrete Compressive Strength measured in (MPa — Mega Pascal). Build Neural network model to predict the compressive strength.

	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
1	141.3	212.0	0.0	203.5	0.0	971.8	748.5	28	29.89
2	168.9	42.2	124.3	158.3	10.8	1080.8	796.2	14	23.51
3	250.0	0.0	95.7	187.4	5.5	956.9	861.2	28	29.22
4	266.0	114.0	0.0	228.0	0.0	932.0	670.0	28	45.85
5	154.8	183.4	0.0	193.3	9.1	1047.4	696.7	28	18.29
6	255.0	0.0	0.0	192.0	0.0	889.8	945.0	90	21.86
7	166.8	250.2	0.0	203.5	0.0	975.6	692.6	7	15.75
8	251.4	0.0	118.3	188.5	6.4	1028.4	757.7	56	36.64
9	296.0	0.0	0.0	192.0	0.0	1085.0	765.0	28	21.65
10	155.0	184.0	143.0	194.0	9.0	880.0	699.0	28	28.99
11	151.8	178.1	138.7	167.5	18.3	944.0	694.6	28	36.35
12	173.0	116.0	0.0	192.0	0.0	946.8	856.8	3	6.94
13	385.0	0.0	0.0	186.0	0.0	966.0	763.0	14	27.92
14	237.5	237.5	0.0	228.0	0.0	932.0	594.0	7	26.26
15	167.0	187.0	195.0	185.0	7.0	898.0	636.0	28	23.89
16	213.8	98.1	24.5	181.7	6.7	1066.0	785.5	100	49.97
17	237.5	237.5	0.0	228.0	0.0	932.0	594.0	28	30.08

Sol:

Business Objective: To predict the strength of concrete with other factors using Neural Network model.

Constraints: Lack of analysis of the previous concrete data.

Data Types: Given data and its types are as follows:

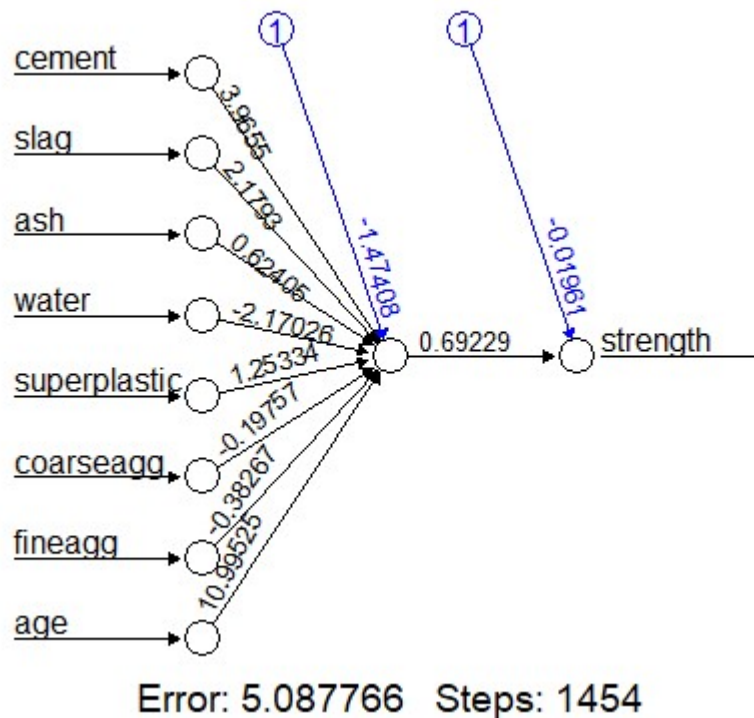
Name of feature	Description	Data Type	Relevance
cement	Quantity of the cement	Ratio	Relevant
slag	Slag quantity in concrete	Ratio	Relevant
ash	Ash qty in concrete	Ratio	Relevant
water	Water concentration in concrete	Ratio	Relevant
superplastic	Sulphur plastic qty in concrete	Ratio	Relevant
coarseagg	Coasre agg in concrete	Ratio	Relevant
fineagg	Fineagg in concrete	Ratio	Relevant
age	Age of the concrete	Ratio	Relevant
strength	Strength value of the concrete	Ratio	Relevant

Data Pre-Processing:

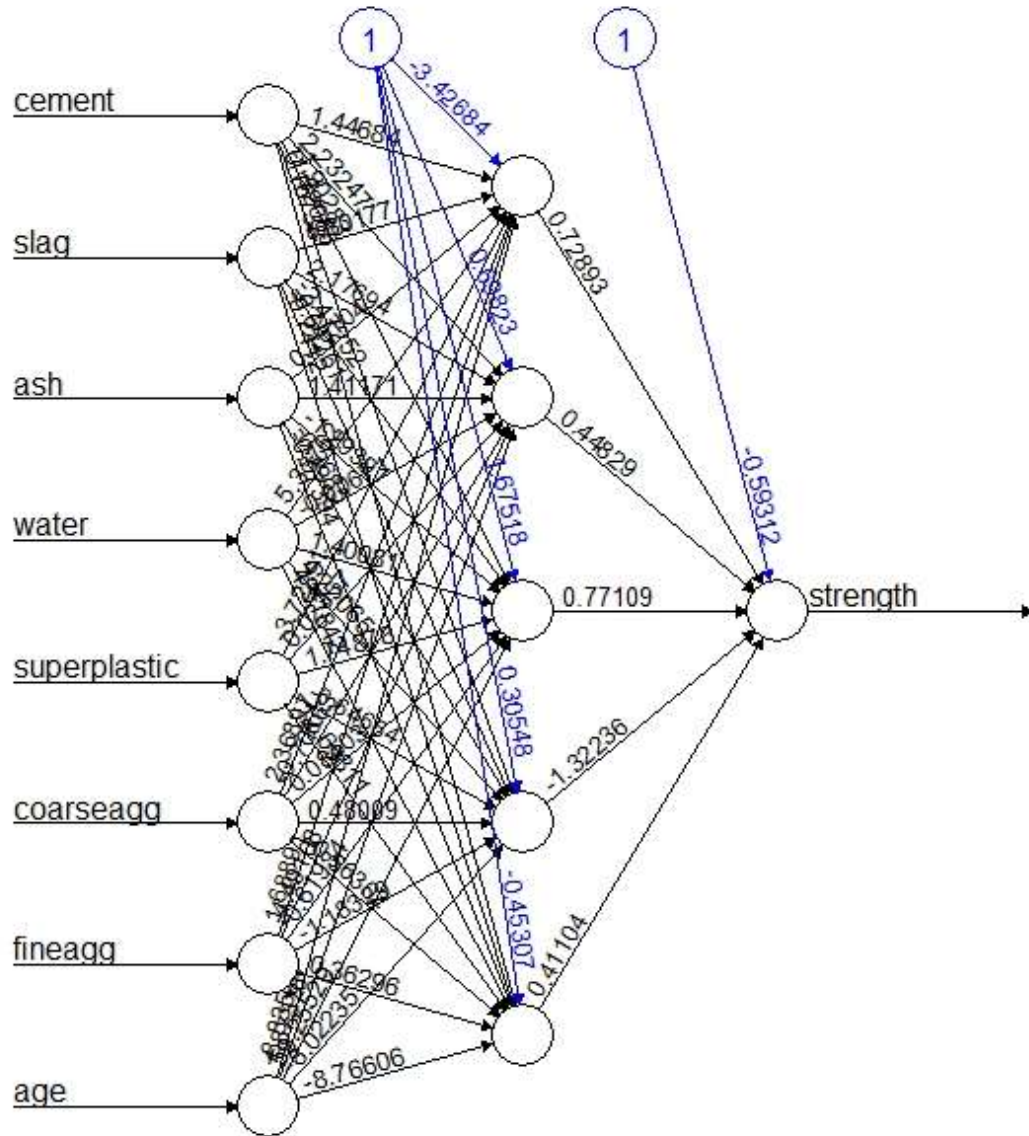
All the data is used for doing the analysis.

EDA: after observing the data all the variables are of different quantity so normalization on the data is applied so that it will be used for doing the analysis.

Neural-Network: After applying the normalization on the data used for doing the neural network model by using neuralnet in R and tensorflow package in Python. Initially neural network is applied without any hidden layers and the accuracy of the model is 80.5% and the network plot of the model is as follows:



After applying neural network with the hidden layers and the accuracy of the model is 92.53% and the network is as follows:



Error: 1.863766 Steps: 17177

Problem Statement: -

RPL Banking and Financing company wants to study the behavior patterns of their customers, so that they can efficiently provide their services and solve the problem of churn and also which would help the business to reduce the loss by giving out loan to customers who cannot repay on time. They have historical data of their customers, build an Artificial Neural network model to predict what kind of customers existed in their business over the time period.

RPL_Bank

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfPr	HasCrCard	IsActiveM	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.9	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.6	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.6	1
4	15701354	Boni	699	France	Female	39	1	0	2	0	0	93826.63	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.8	1	1	1	79084.1	0
6	15574012	Chu	645	Spain	Male	44	8	113755.8	2	1	0	149756.7	1
7	15592531	Bartlett	822	France	Male	50	7	0	2	1	1	10062.8	0
8	15656148	Obinna	376	Germany	Female	29	4	115046.7	4	1	0	119346.9	1
9	15792365	He	501	France	Male	44	4	142051.1	2	0	1	74940.5	0
10	15592389	H?	684	France	Male	27	2	134603.9	1	1	1	71725.73	0
11	15767821	Bearce	528	France	Male	31	6	102016.7	2	0	0	80181.12	0
12	15737173	Andrews	497	Spain	Male	24	3	0	2	1	0	76390.01	0
13	15632264	Kay	476	France	Female	34	10	0	2	1	0	26260.98	0
14	15691483	Chin	549	France	Female	25	5	0	2	0	0	190857.8	0
15	15600882	Scott	635	Spain	Female	35	7	0	2	1	1	65951.65	0
16	15643966	Goforth	616	Germany	Male	45	3	143129.4	2	0	1	64327.26	0
17	15737452	Romeo	653	Germany	Male	58	1	132602.9	1	1	0	5097.67	1
18	15788218	Henderso	549	Spain	Female	24	9	0	2	1	1	14406.41	0
19	15661507	Muldrow	587	Spain	Male	45	6	0	1	0	0	158684.8	0
20	15568982	Hao	726	France	Female	24	6	0	2	1	1	54724.03	0
21	15577657	McDonald	732	France	Male	41	8	0	2	1	1	170886.2	0
22	15597945	Dellucci	636	Spain	Female	32	8	0	2	1	0	138555.5	0
23	15699309	Gerasimos	510	Spain	Female	38	4	0	1	1	0	118913.5	1

Business Objective: To predict the kind of customer with other factors using Neural Network model.

Constraints: Lack of analysis of the previous customer data.

Data Types: Given data and its types are as follows:

Name of feature	Description	Data Type	Relevance
RowNumber	Row number of the observation	Ordinal	Irrelevant
CustomerId	Customer Id	Ordinal	Irrelevant
Surname	Surname of the customer	Nominal	Irrelevant
CreditScore	Credit score of the customer	Ratio	Relevant
Geography	Location of the customer	Nominal	Relevant
Gender	Gender of the customer	Nominal	Relevant
Age	Age of the customer	Ratio	Relevant
Tenure	Tenure for the customer	Ratio	Relevant
Balance	Balance of the account	Ratio	Relevant
NumOfProducts	Number of products customer bought	Nominal	Relevant
HasCrCard	Whether customer has credit card or not	Nominal	Relevant
IsActiveMember	Whether the customer is active member or not	Nominal	Relevant
EstimatedSalary	Estimated salary of the customer	Ratio	Relevant
Exited	Whether the customer is exited or not	Nominal	Relevant

Data Pre-Processing:

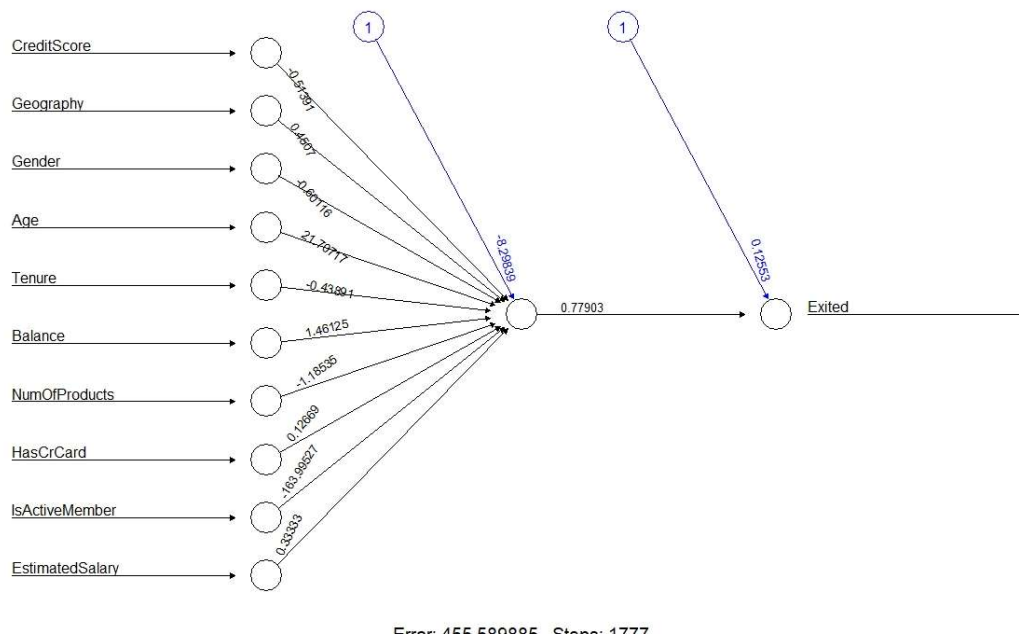
First three columns in the given data which are rownumber, customerid and surname are not used for doing the analysis so the same is deleted from the main data.

Geography and gender columns are given in categorical format so the same is converted

into numeric data so that it will be used for doing the analysis.

EDA: after observing the data all the variables are of different quantity so normalization on the data is applied so that it will be used for doing the analysis.

Neural-Network: After applying the normalization on the data used for doing the neural network model by using neuralnet in R and tensorflow package in Python. Initially neural network is applied without any hidden layers and the accuracy of the model is 82.5% and the network plot of the model is as follows:



After applying neural network with the hidden layers and the accuracy of the model is 90.2% and the network is as follows:

