# A picture containing shape, arrow Description automatically generatedTopic: 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. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**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:**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

**3.2 Outlier Imputation if applicable.**

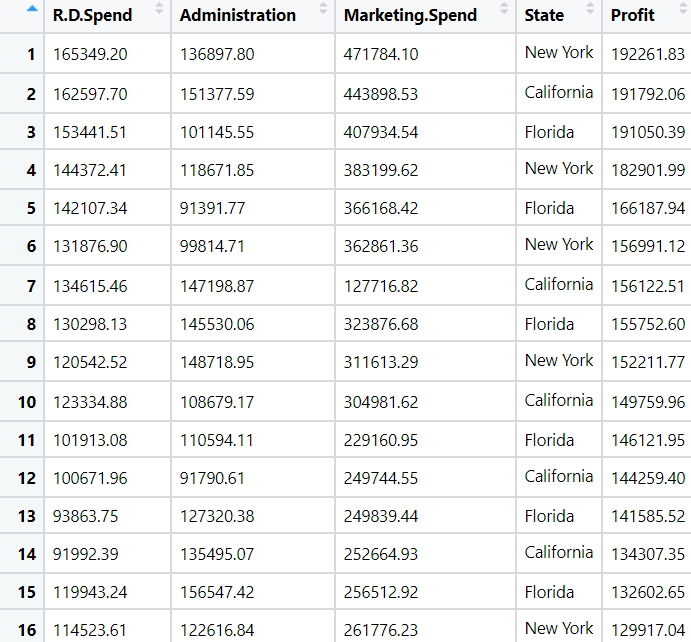
1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building:**
   1. **Perform Artificial Neural Network on the given datasets.**
   2. **Use TensorFlow keras to build your model in Python and use Neural net package in R**
   3. **Briefly explain the output in the documentation for each step in your own words.**
   4. **Use different activation functions to get the best model.**
3. **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.

**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.

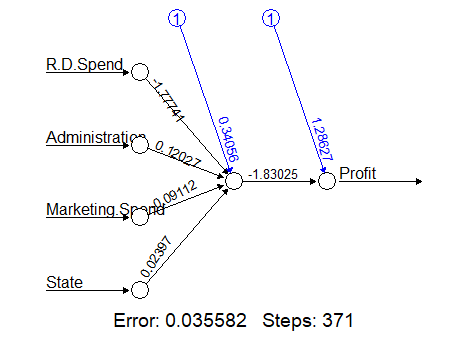
**Data Types:** All the given data and its types are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

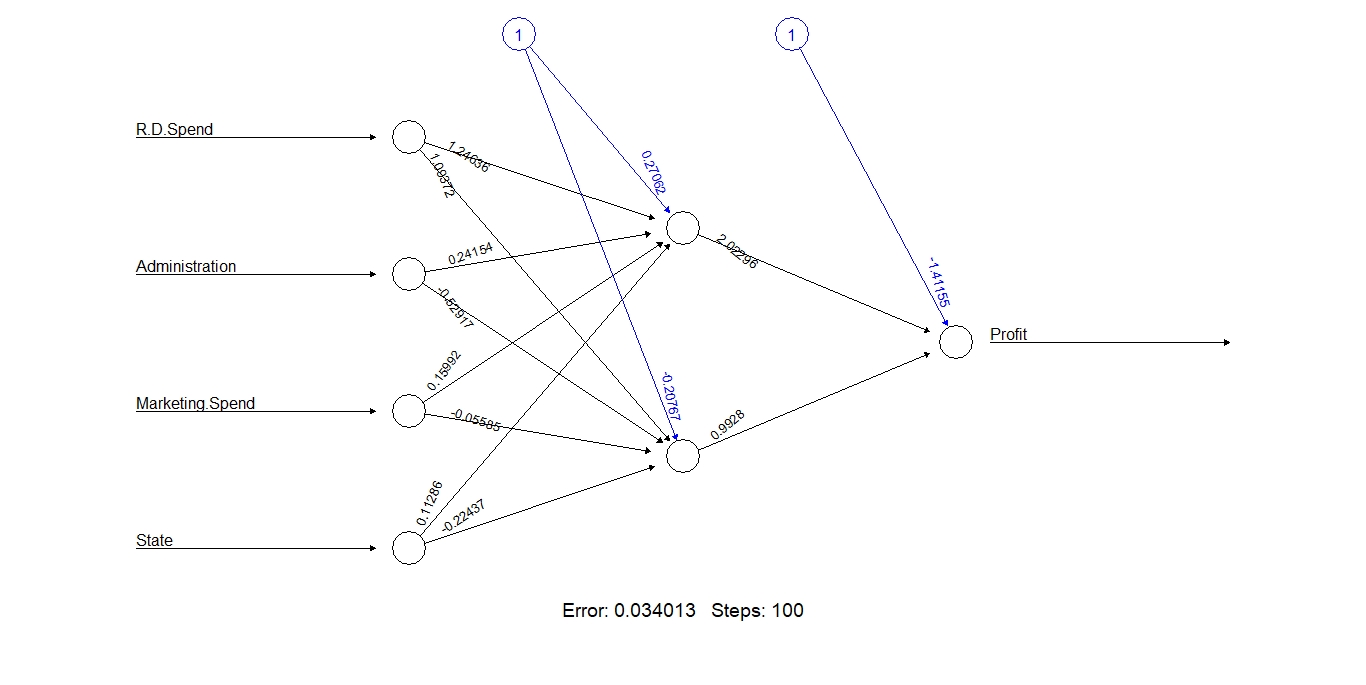
**Data Cleaning:** Since the states column is in Non-Numeric format the same is converted to numeric data for doing the analysis.

**Exploratory Data Analysis:** Normalization on thedata 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 tenserflow 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:

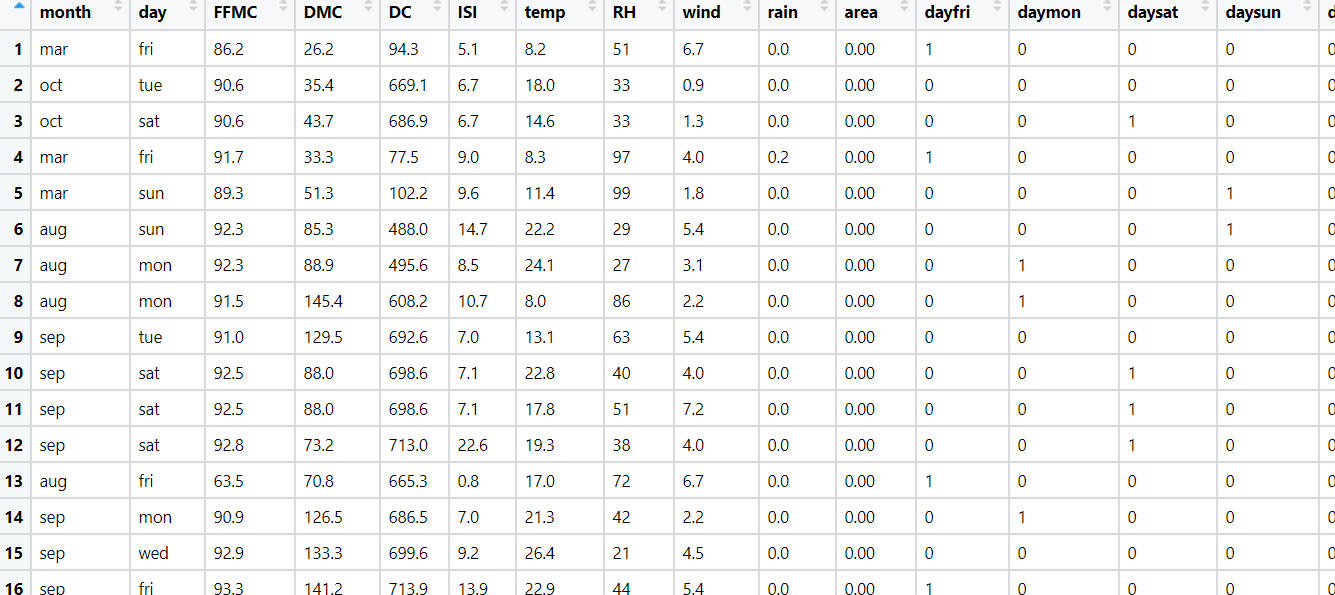


After applying neural network with the hidden layers and the accuracy of the model is 96.39% and the network is as follows:



**Problem Statement:-**

1. W 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.



**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:

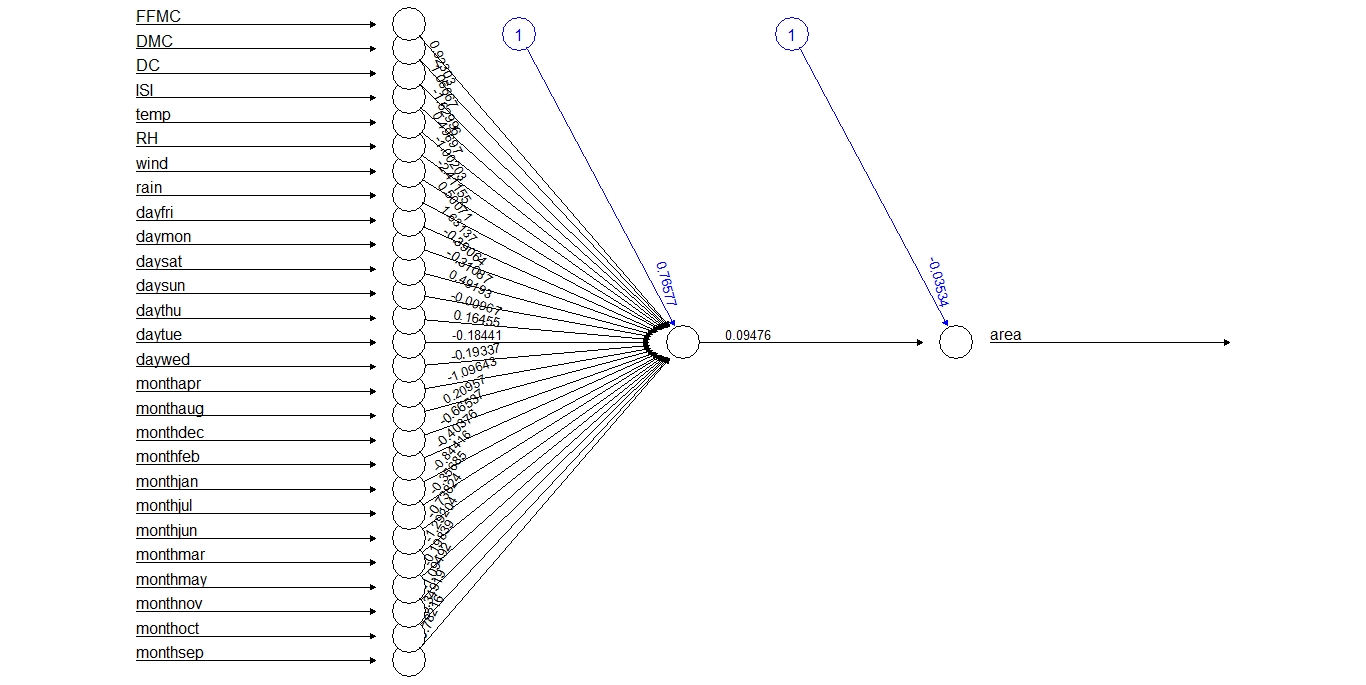
|  |  |  |  |
| --- | --- | --- | --- |
| Name of feature | Description | Data Type | Relevance |
| Month | Month when fire occurred | Nominal | Relevant |
| Day | Day when fire occured | Nominal | Relevant |
| FFMC | Fine Fuel Moisture Code for moisture content in fire data | Ratio | Relevant |
| DMC | 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 into 1 or 0 based on type of day | Nominal | Relevant |
| Month type | All the months are converted into 1 or 0 based on type of months | Nominal | Relevant |

**Data Pre-Processing:**

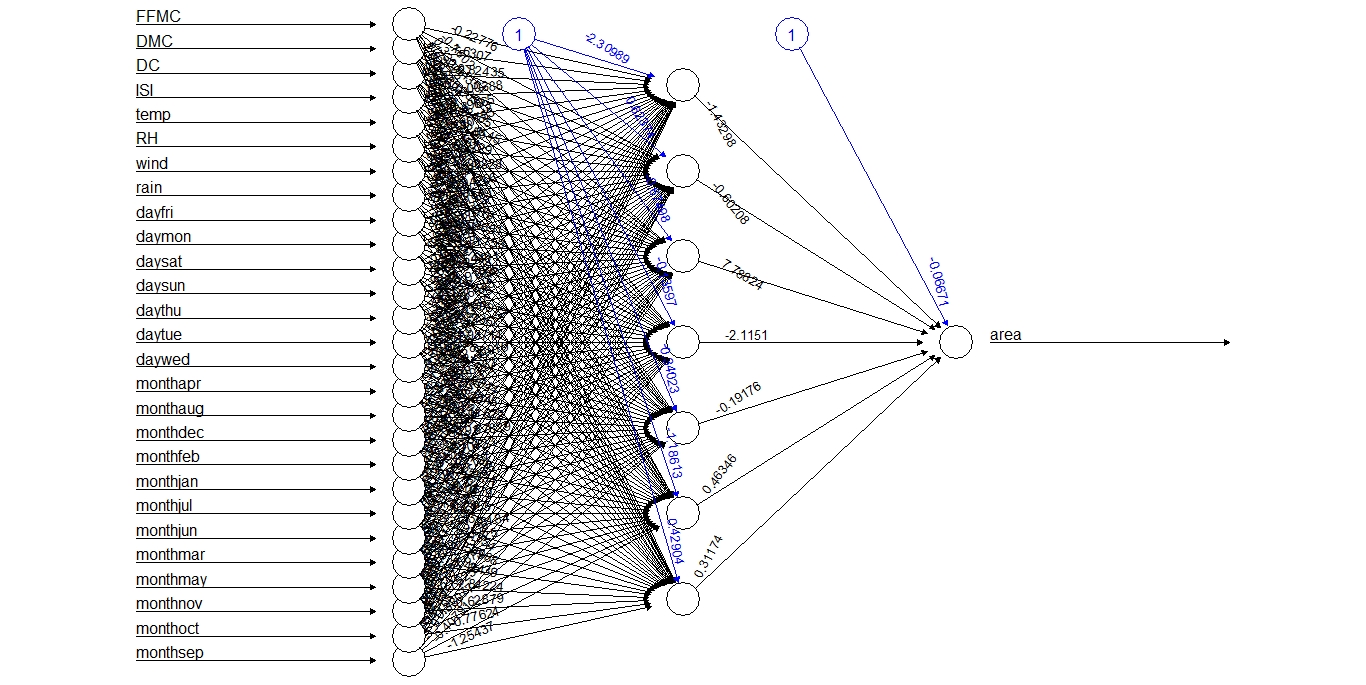
Since first two columns are repeated which are month and the day the same are removed for doing the analysis on the data.

**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 tenserflow 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:

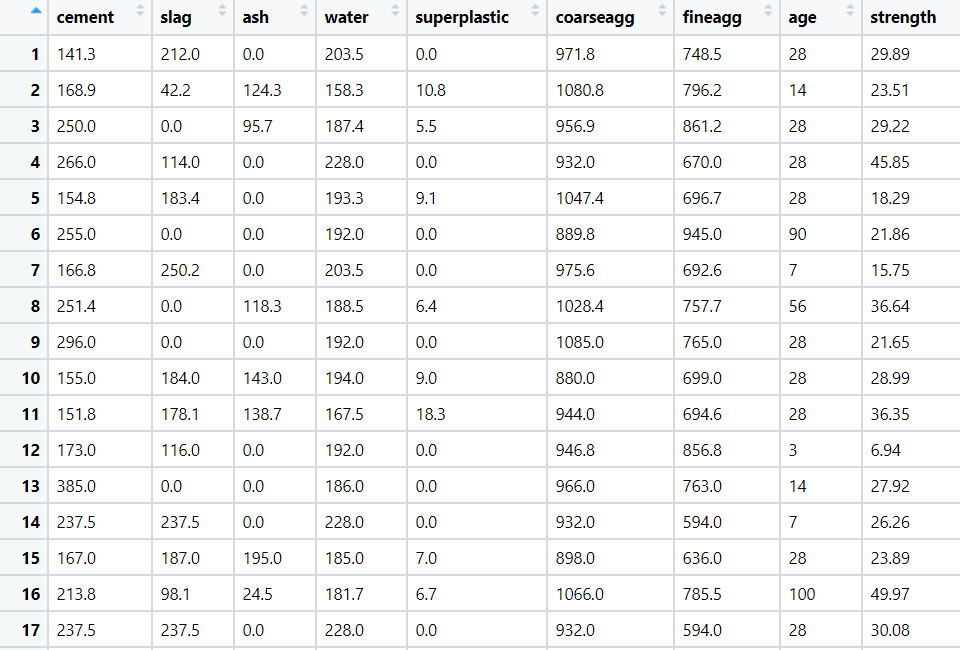


After applying neural network with the hidden layers and the accuracy of the model is 96.39% and the network is as follows:



**Problem Statement:-**

1. 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/m³) 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.



**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:

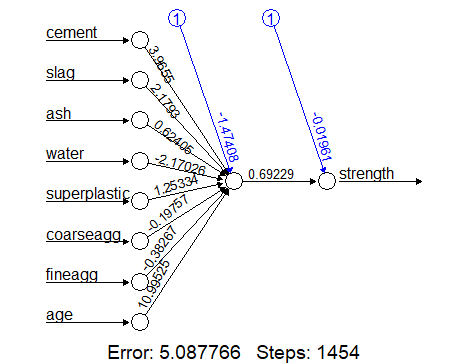
|  |  |  |  |
| --- | --- | --- | --- |
| Name of feature | Description | Data Type | Relevance |
| cement | Quantity of the cement | Ratio | Relevant |
| slag | Slag quantity in concrete | Ratio | Relevant |
| ash | Ash qty in concrete | Ratio | Relevant |
| water | Water concentration in concrete | 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:**

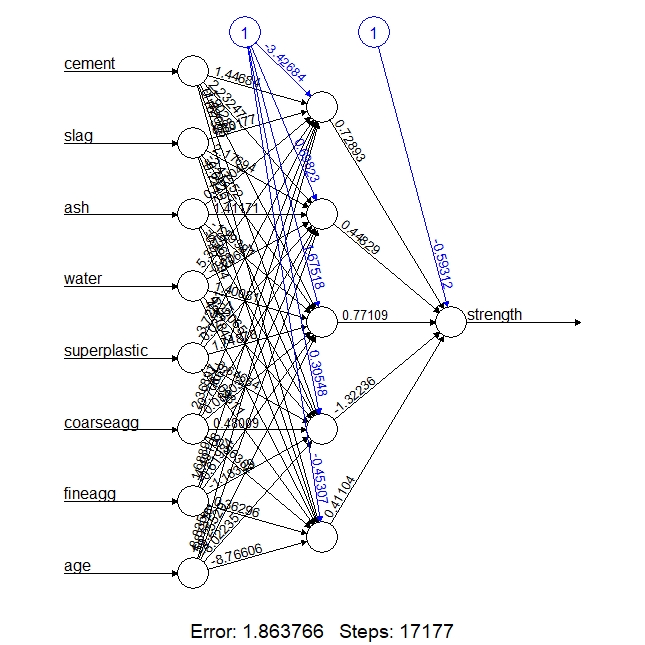
All the data is used for doing the analysis.

**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 tenserflow 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:

****

**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

A picture containing shape, arrow

Description automatically generated

Table

Description automatically generated

**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 |

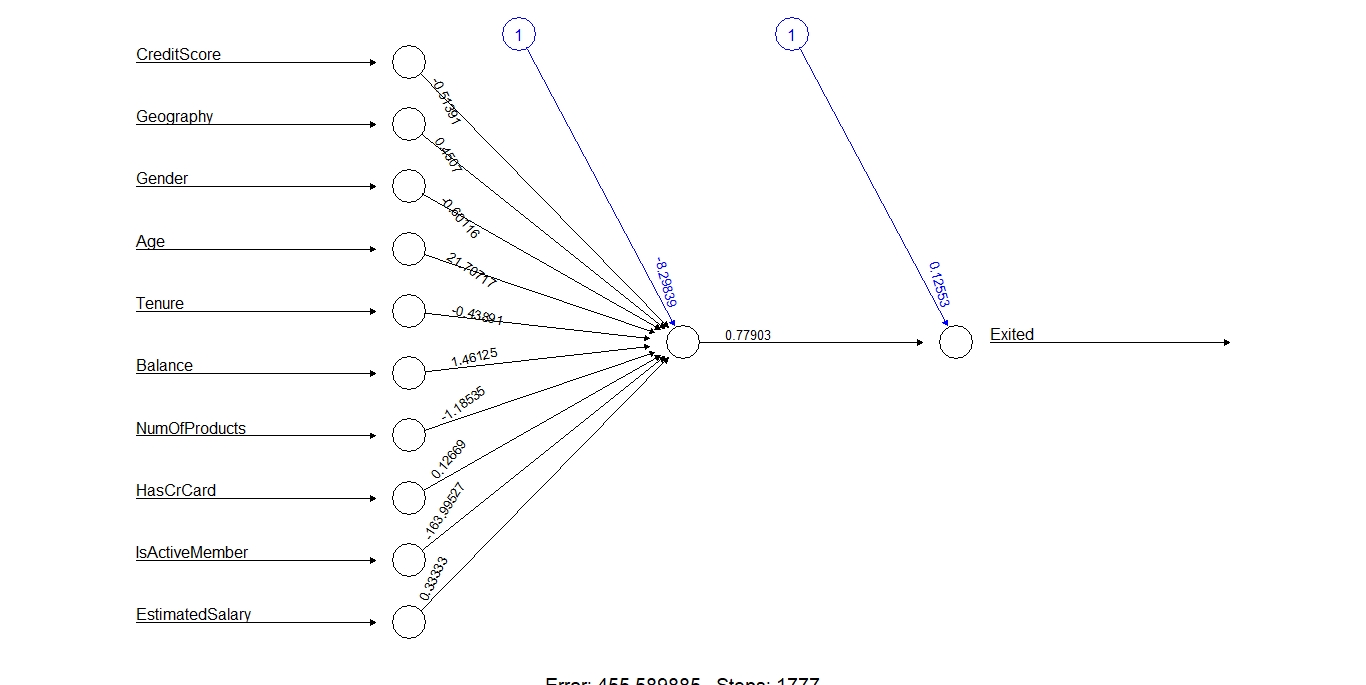
**Data Pre-Processing:**

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 tenserflow 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:

