# Artificial Neural Networks

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: DHEERAJ MISHRA Batch ID:**  DS\_01072021

**Topic: Artificial Neural Networks**

**Grading Guidelines:**

**1. An assignment submission is considered complete only when correct and executable code(s) are submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline.
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of problems are completed.

(OR)

* + All assignments are submitted after the deadline.
* **Grade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of the problems are completed.

(OR)

* + Less than 80% of problems in the assignments are submitted after the deadline.
* **Grade D: (>= 60 and < 70):**
  + Assignments submitted after the deadline and with 50% or less problems.
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of problems in the assignments are submitted after the deadline.

(OR)

* + Less than 30% of problems in the assignments are submitted before the deadline.
* **Grade F: (< 50):** No submission (or) malpractice.

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
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. And if not relevant, provide reasons and a description of the feature.**

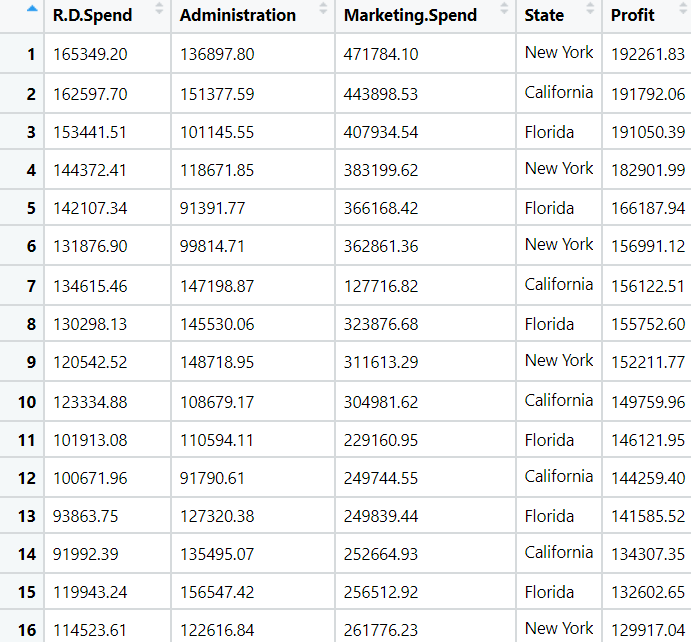
1. **Data Pre-processing**

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

**3.2 Outlier Treatment if applicable.**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary.**
   2. **Univariate analysis.**
   3. **Bivariate analysis.**
2. **Model Building:**
   1. **Build an Artificial Neural Network model on the given datasets.**
   2. **Use TensorFlow and Keras packages.**
   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. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

1. We have a dataset which contains the details of 50 startups. Build an ANN model to predict the profit of a new startup based on certain features.





1. BUSINESS OBJECTIVE:-

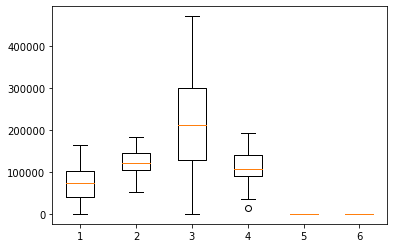
Maximize relationship between dependent variable

1. DATA UNDERSTANDING:-

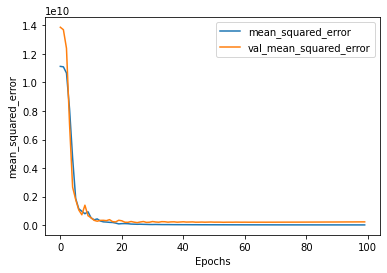
|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | DESCRIPTION | TYPE | RELEVANCE |
| R &D Spend | Research and Development team | Continuous | Relevant |
| Administration | Administrative team | Continuous | Relevant |
| Marketing.spend | Marketing spend | Continuous | Relevant |
| State | State name | Discrete | Relevant |
| Profit | Total profit | Continuous | Relevant |

1. DATA CLEANSING :-
2. Dataset consists of 5 colums and 50 rows
3. Duplicate row does not exists
4. All data types are of form int64 and object
5. No null values found in each column
6. From describe function mean , median and standard deviation obtained
7. Outliers present and retained
8. Corelation coefficients calculated
9. Some are positively skewed and some are negatively skewed
10. Scaling is done by standard scalar
11. Splitting data to train and test
12. EDA:-

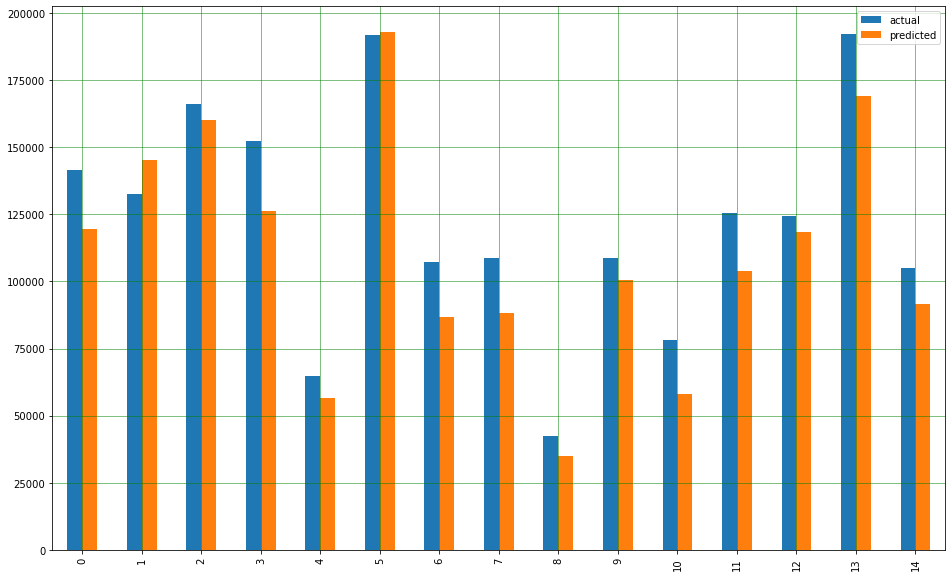
From box plot



1. MODEL BUILDING:-
2. Model builded
3. Splitting data to test = 30% and train = 70%
4. Hidden layers five
5. Activation function relu
6. History plot



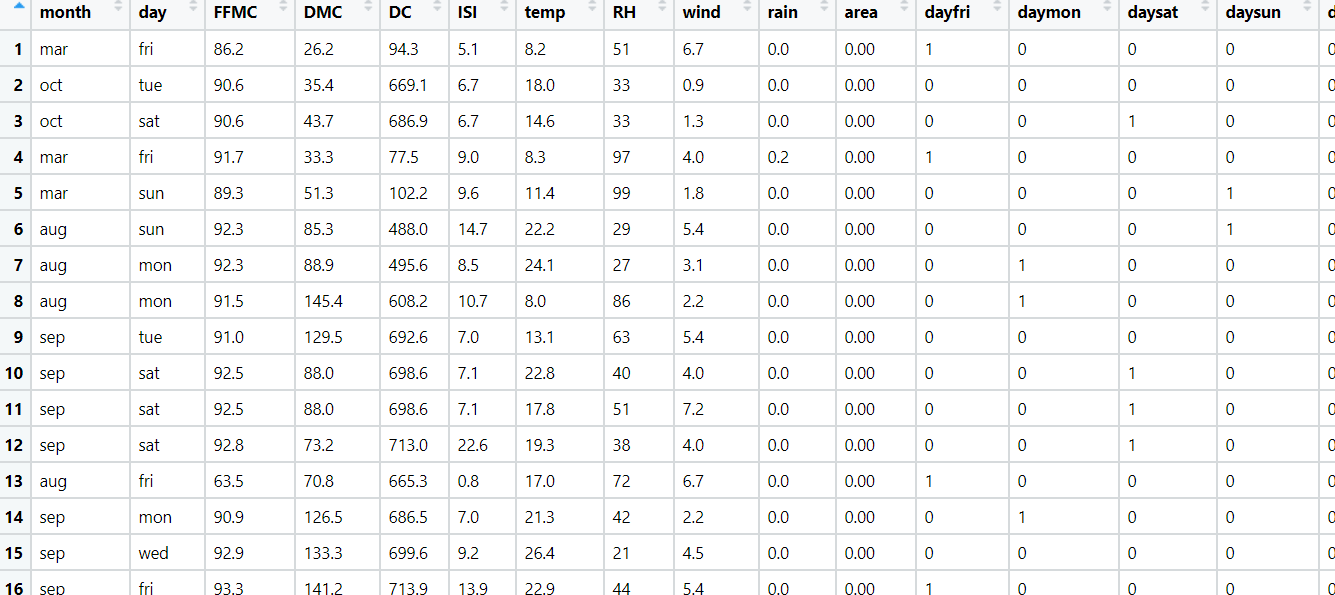
1. Test RMSE = 60595.65
2. Train RMSE = 54189.14
3. Bar plot for actual value and predicted value



OUTPUT:-

1. From history plot rmse is constant after 10 epochs
2. From bar plot actual values and predicted values are similar
3. It is right fit model
4. BENEFITS :-

From above information we can predict for profit against all input variables

1. We have a dataset about 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 burnt area of forest fires with the help of an Artificial Neural Network model.

A picture containing shape, arrow

Description automatically generated

1. BUSINESS OBJECTIVE:-

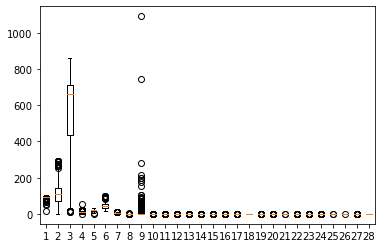
Maximize relationship between dependent variable

1. DATA UNDERSTANDING:-

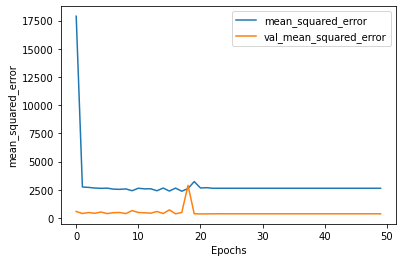
|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | DESCRIPTION | TYPE | RELEVANCE |
| month | Month of year | Text | Not Relevant |
| day | Day of week | Text | Not Relevant |
| FFMC | Fine Fuel Moisture Code | Continuous | Relevant |
| DMC | Duff Moisture Code | Continuous | Relevant |
| DC | Drought Code | Continuous | Relevant |
| ISI | Initial Spread Index | Continuous | Relevant |
| temp | temperature noon | Continuous | Relevant |
| RH | Relative Humidity | Continuous | Relevant |
| wind | Wind speed | Continuous | Relevant |
| rain | total day | Continuous | Relevant |
| area | Area burnt | Continuous | Relevant |

1. DATA CLEANSING :-
2. Dataset consists of 30 colums and 517 rows
3. Dropping nominal colums
4. Duplicate row exists and removed
5. All data types are of form float64 and object
6. No null values found in each column
7. From describe function mean , median and standard deviation obtained
8. Outliers present and retained
9. Corelation coefficients calculated
10. Some are positively skewed and some are negatively skewed
11. Scaling is done by standard scalar
12. Splitting data to train and test
13. EDA:-

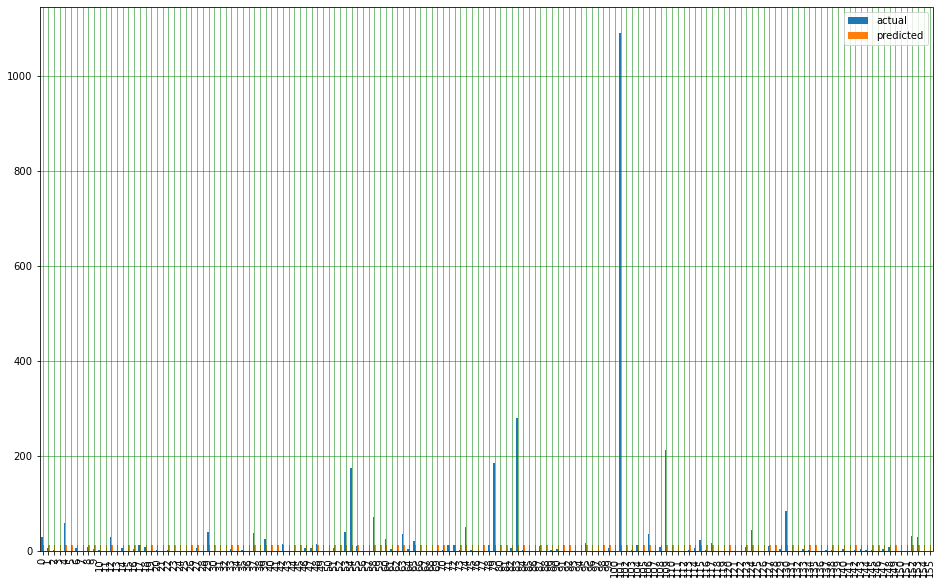
From box plot



1. MODEL BUILDING:-
2. Model builded
3. Splitting data to test = 30% and train = 70%
4. Hidden layers five
5. Activation function relu
6. History plot



1. Test RMSE = 93.5447
2. Train RMSE = 44.86142
3. Bar plot for actual value and predicted value

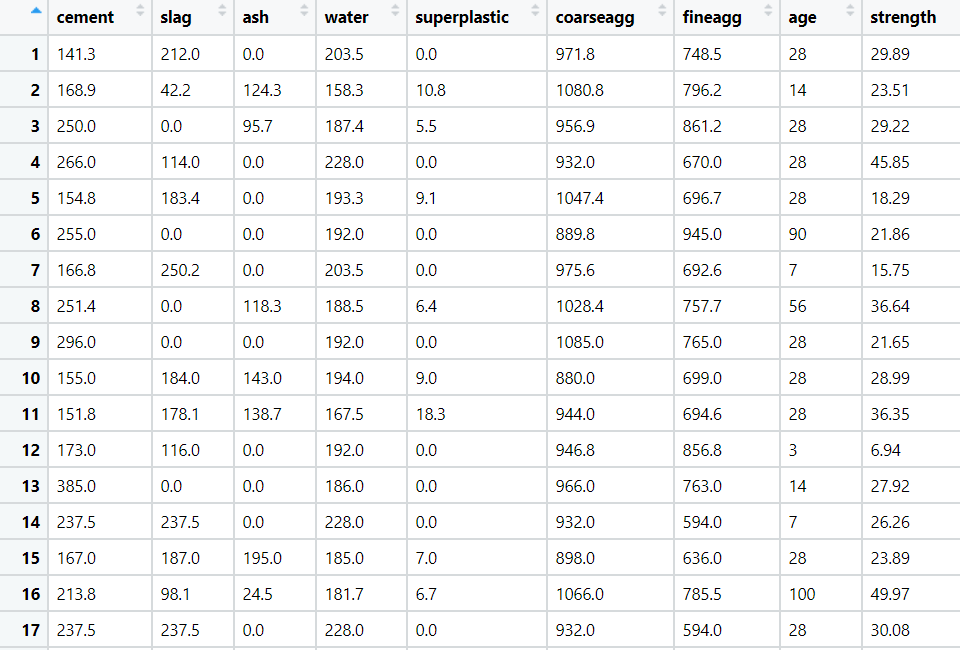


OUTPUT:-

1. From history plot rmse is constant after 20 epochs
2. From bar plot actual values and predicted values are quite similar
3. It is right fit model
4. BENEFITS :-

From above information we can predict for burnt area against all input variables

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 (Mega Pascal). Build a Neural network model to predict the compressive strength.



1. BUSINESS OBJECTIVE:-

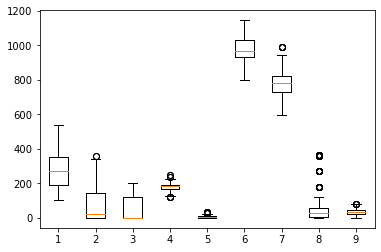
Maximize relationship between dependent variable

1. DATA UNDERSTANDING:-

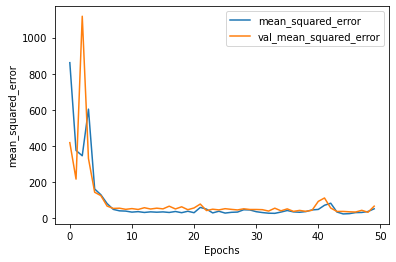
|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | DESCRIPTION | TYPE | RELEVANCE |
| cement | Cement weight | Continuous | Relevant |
| slag | Slag in cement | Continuous | Relevant |
| ash | Ash in cement | Continuous | Relevant |
| Water | Amount of water | Continuous | Relevant |
| Super plastic | Super plastic quality | Continuous | Relevant |
| coarseagg | Coarse aggregate | Continuous | Relevant |
| fineagg | Fine aggregate | Continuous | Relevant |
| age | Age of cement | Discrete | Relevant |
| strength | Strength of cement | Continuous | Relevant |

1. DATA CLEANSING :-
2. Dataset consists of 9 colums and 1030 rows
3. Duplicate row does not exists
4. All data types are of form float64 and int64
5. No null values found in each column
6. From describe function mean , median and standard deviation obtained
7. Outliers present and retained
8. Corelation coefficients calculated
9. Some are positively skewed and some are negatively skewed
10. Scaling is done by standard scalar
11. Splitting data to train and test
12. EDA:-

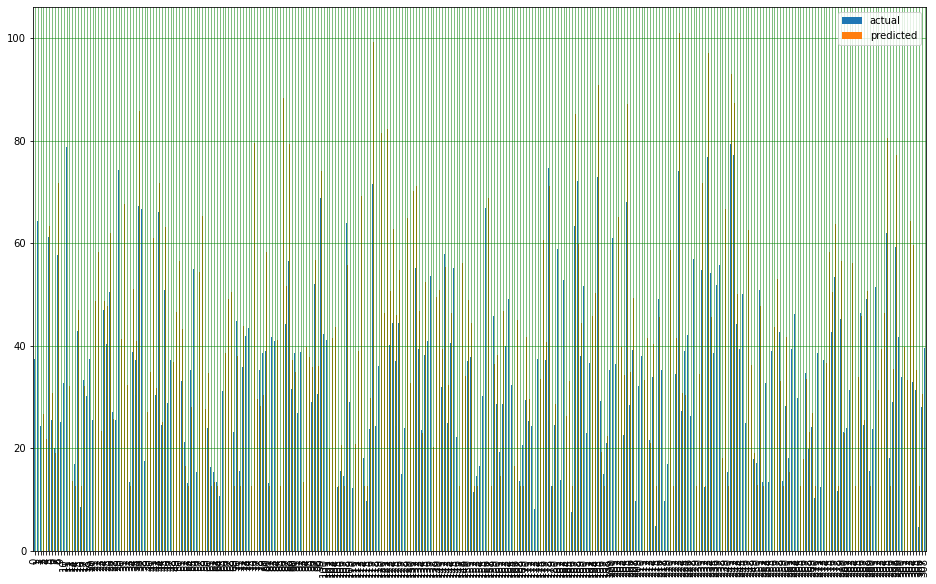
From box plot



1. MODEL BUILDING:-
2. Model builded
3. Splitting data to test = 30% and train = 70%
4. Hidden layers five
5. Activation function relu
6. History plot



1. Test RMSE = 27.011051
2. Train RMSE = 26.046236
3. Bar plot for actual value and predicted value



OUTPUT:-

1. From history plot rmse is constant after 5 epochs
2. From bar plot actual values and predicted values are quite similar
3. It is right fit model
4. BENEFITS :-

From above information we can predict for strength of cement against all input variables

1. 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. They have historical data of their customers. Build an Artificial Neural Network with Exited as the target variable.

Table

Description automatically generated

1. BUSINESS OBJECTIVE:-

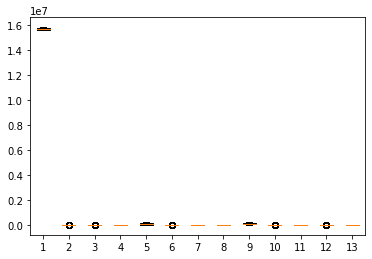
Maximize relationship between dependent variable

1. DATA UNDERSTANDING:-

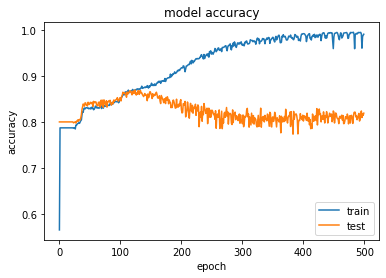
|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | DESCRIPTION | TYPE | RELEVANCE |
| RowNumber | Row no | Discrete | Not Relevant |
| CustomerId | Customer id | Discrete | Not Relevant |
| Surname | Surname | Text | Not Relevant |
| CreditScore | Credit score | Discrete | Relevant |
| Geography | Location | Text | Relevant |
| Gender | Gender | Text | Relevant |
| Age | Age | Discrete | Relevant |
| Tenure | Tenure of customer | Discrete | Relevant |
| Balance | Balance | Continuous | Relevant |
| NumOfProducts | No of products | Discrete | Relevant |
| HasCrCard | Has credit card | Binary | Relevant |
| IsActiveMember | Active member or not | Binary | Relevant |
| EstimatedSalary | Estimated salary | Continuous | Relevant |
| Excited | Excited or not | Binary | Relevant |

1. DATA CLEANSING :-
2. Dataset consists of 14 colums and 10000 rows
3. Dropping nominal columns
4. Duplicate row does not exists
5. All data types are of form float64 , int64 and object
6. No null values found in each column
7. From describe function mean , median and standard deviation obtained
8. Outliers present and retained
9. Corelation coefficients calculated
10. Some are positively skewed and some are negatively skewed
11. Scaling is done by standard scalar
12. Splitting data to train and test
13. EDA:-

From box plot



1. MODEL BUILDING:-
2. Model builded
3. Splitting data to test = 30% and train = 70%
4. Hidden layers five
5. Activation function sigmoid
6. History plot



1. Test RMSE = 0.5352471
2. Train RMSE = 0.5632835
3. Bar plot for actual value and predicted value

OUTPUT:-

1. From history plot rmse is constant after 150 epochs
2. It is right fit model
3. BENEFITS :-

From above information we can predict for excited against all input variables