# 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: Nithin Dsouza**

**Batch Id: 05012021-10AM**

**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.**

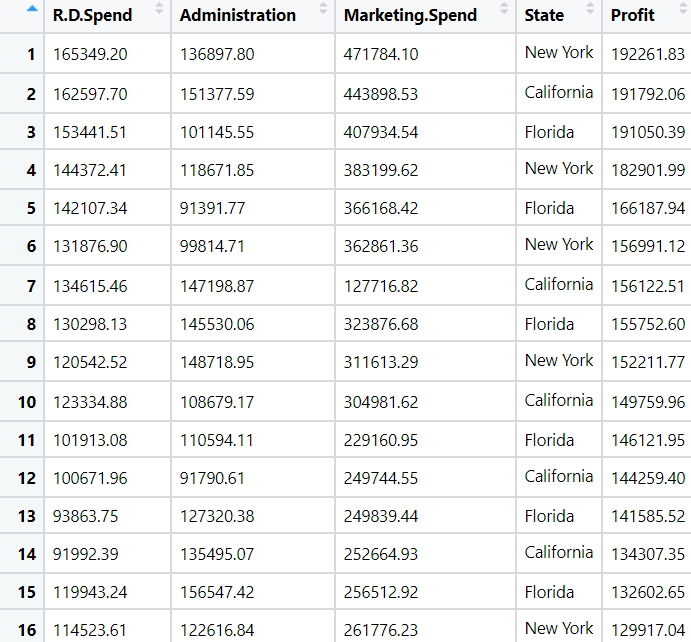
1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building:**
   1. **Perform Artificial Neural Network on the given datasets.**
   2. **Use TensorFlow keras to build your model in Python and use Neural net package in R**
   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.

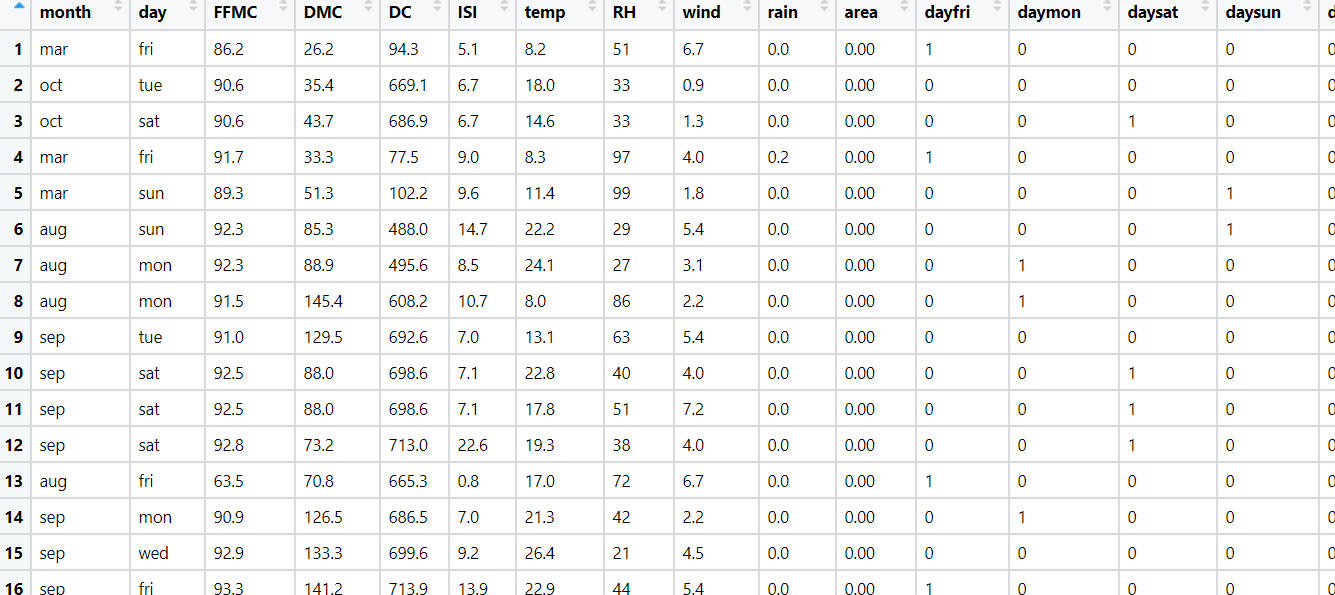
|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Type | Relevance |
| R&D Spend | Research and development spend | Continuous, Ratio | It provides useful information |
| Administration | Administration Spend | Continuous, Ratio | It provides useful information |
| Marketing Spend | Marketing Spend | Continuous, Ratio | It provides useful information |
| State | State Name | Discrete, Nominal | It is not useful in analysis |
| Profit | Profit of company | Continuous, Ratio | It provides useful information |

**Business Objective**: Maximize accuracy between independent input variables and dependent output variable

**Business Constraint**: Minimize correlation problem

**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.



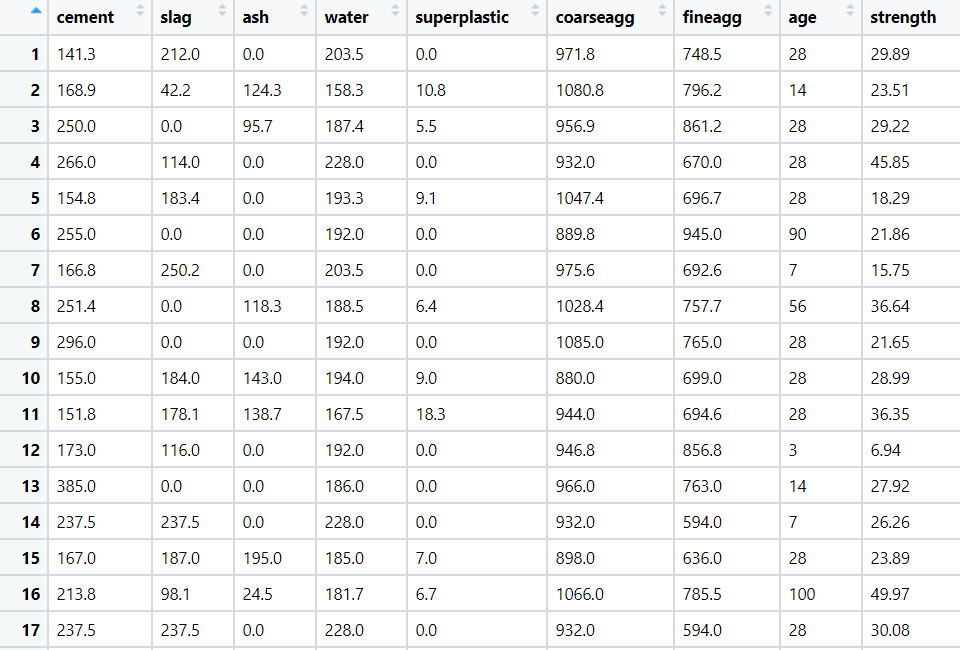
|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Type | Relevance |
| Month | Month Name | Discrete, Ordinal | Not useful |
| Day | Day in a week | Discrete, Ordinal | Not useful |
| FFMC | Full-fledged money changer | Continuous, Ratios | It’s useful information |
| DMC | Delhi medical council | Continuous, Ratios | It’s useful information |
| DC | Defective Comics | Continuous, Ratios | It’s useful information |
| ISI | Indian statistical institution | Continuous, Ratios | It’s useful information |
| Temp | Temperature | Continuous, Ratios | It’s useful information |
| RH | Relative Humidity | Continuous, Ratios | It’s useful information |
| Wind | Wind blow rate | Continuous, Ratios | It’s useful information |
| Rain | Rainfall Rate | Continuous, Ratios | It’s useful information |
| Area | Area of land | Continuous, Ratios | It’s useful information |
| Dayfri | Friday | Discrete, Ordinal | It’s useful information |
| Daymon | Monday | Discrete, Ordinal | It’s useful information |
| Daysat | Saturday | Discrete, Ordinal | It’s useful information |
| Daysun | Sunday | Discrete, Ordinal | It’s useful information |
| Daythu | Thursday | Discrete, Ordinal | Discrete, Ordinal |
| Daytue | Tuesday | Discrete, Ordinal | Discrete, Ordinal |
| Daywed | Wednesday | Discrete, Ordinal | Discrete, Ordinal |
| Monthapr | April month | Discrete, Ordinal | Discrete, Ordinal |
| Monthaug | August month | Discrete, Ordinal | Discrete, Ordinal |
| Monthdec | December month | Discrete, Ordinal | Discrete, Ordinal |
| Monthfeb | February month | Discrete, Ordinal | Discrete, Ordinal |
| Monthjan | January Month | Discrete, Ordinal | Discrete, Ordinal |
| Monthjul | July Month | Discrete, Ordinal | Discrete, Ordinal |
| Monthjun | June month | Discrete, Ordinal | Discrete, Ordinal |
| monthmar | Month March | Discrete, Ordinal | Discrete, Ordinal |
| Monthmay | Month may | Discrete, Ordinal | Discrete, Ordinal |
| Monthnov | Month November | Discrete, Ordinal | Discrete, Ordinal |
| Monthoct | Month October | Discrete, Ordinal | Discrete, Ordinal |
| monthsep | Month September | Discrete, Ordinal | Discrete, Ordinal |

**Business Objective**: Maximize accuracy between independent input variables and dependent output variable

**Business Constraint**: Minimize correlation problem

**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.



|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Type | Relevance |
| Cement | Cement Ratio | Continuous, Ratio | It is useful information |
| Slag | Slag Ratio | Continuous, Ratio | It is useful information |
| Ash | Ash Ratio | Continuous, Ratio | It is useful information |
| Water | Amount of water | Continuous, Ratio | It is useful information |
| Superplastic | Plastic content | Continuous, Ratio | It is useful information |
| Coarseagg | Construct component | Continuous, Ratio | It is useful information |
| fineagg | Construct component | Continuous, Ratio | It is useful information |
| Age | Age factor | Continuous, Ratio | It is useful information |
| strength | Strength of building | Continuous, Ratio | It is useful information |

**Business Objective**: Maximize accuracy between independent input variables and dependent output variable

**Business Constraint**: Minimize correlation problem

**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

Table

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Type | Relevance |
| CustomerId | Identification card | Discrete, Nominal | Not useful |
| Surname | Surname of a person | Discrete, Nominal | Not useful |
| CreditScore | Credit score of person | Continuous, Ratios | Useful information |
| Geography | Geography details | Discrete, Nominal | Not useful |
| Gender | Gender of a person | Discrete, Nominal | Not useful |
| Age | Age of a person | Discrete, Nominal | Not useful |
| Tenure | Set of conditions | Continuous, Ratios | Useful information |
| Balance | Account Balance | Continuous, Ratios | Useful information |
| NumOfProducts | Number of products | Continuous, Ratios | Useful information |
| HasCrCard | Number of HasCrCard | Continuous, Ratios | Useful information |
| IsActiveMember | Checking status of a person | Continuous, Ratios | Useful information |
| EstimatedSalary | Estimated salary | Continuous, Ratios | Useful information |
| Exited | Kind of customers existed | Continuous, Ratios | Useful information |

**Business Objective**: Maximize accuracy between independent input variables and dependent output variable

**Business Constraint**: Minimize correlation problem

A picture containing shape, arrow

Description automatically generated