

TOOLS AND TECHNOLOGIES

In the Diabetes Analysis using Pima Indian's Dataset project, the following tools and technologies will be used:

1. **Libraries for Visualising the data** like ggplot() , ggally() and other plotting techniques.
2. **Packages for convert wide data to narrow data like** , reshape2() and other packages.
3. **Statistical models:** The project will employ statistical models such as linear regression, logistic regression, and decision trees to build a predictive model for diabetes.
4. **Programming language:** The project will be implemented using the R programming language, which is widely used for statistical analysis in field of data science.
5. **R studio** – to perform R language actions for easy purposes.
6. **Microsoft Excel** – to store the dataset .

Used packages installation guide,

In R programming contains several plotting packages to visualise the data , that packages are installed using syntax of

Packages.install(“packages_name”)

Versions of packages knowing by using **utils** library

- **Version of dplyr**

```
{r}  
library(utils)  
packageVersion("dplyr")  
  
[1] '1.0.10'
```

The **dplyr** package is part of the larger tidyverse ecosystem, which includes a number of other packages for data analysis, visualization, and modeling.

- **Version of ggplot2()**

```
{r}  
library(utils)  
packageVersion("ggplot2")  
  
[1] '3.4.0'
```

ggplot2 is a popular data visualization library for R. It provides a high-level interface for creating static, animated, and interactive visualizations. The library is based on the Grammar of Graphics, a systematic approach to describing the components of a graphic, making it easy to create complex visualizations using simple commands.

- **Version of GGally for ggcorr()**

```
{r}  
library(utils)  
packageVersion("GGally")  
  
[1] '2.1.2'
```

Ggally is a package in R that extends the functionality of the popular **ggplot2** package by providing easy-to-use functions for creating complex and informative graphics. **ggcorr()** is particularly useful for creating correlation plots that visualize the relationship between multiple variables in a dataset.

- **Version of reshape2() for melt()**

```
{r}  
library(utils)  
packageVersion("reshape2")  
  
[1] '1.4.4'
```

The **melt()** function can be used to convert wide data into long format, making it easier to perform certain types of data analysis and visualization.

- **Version of Stats() for lm() , residuals() and fitted()**

```
{r}  
library(utils)  
packageVersion("stats")
```

```
[1] '4.2.2'
```

The **stats** library is part of the base R installation and provides basic statistical functions, including linear regression. The **lm** function can be used to fit linear regression models, and the residuals and fitted values can be obtained from the **residuals** and **fitted** methods, respectively.

- **Version of rpart()**

```
{r}  
library(utils)  
packageVersion("rpart")
```

```
[1] '4.1.19'
```

rpart is a package in R for creating classification and regression trees (CART). CART is a type of decision tree that is commonly used for predictive modeling and machine learning. The **rpart** package provides functions for building and plotting CART models in R.

- **Version of Hmisc()**

```
{r}  
library(utils)  
packageVersion("Hmisc")
```

```
[1] '4.7.2'
```

Hmisc provides functions for creating high-quality graphics, including scatterplots, histograms, and box plots.

Hmisc provides functions for fitting regression models and calculating residuals and fitted values.

These tools and technologies have been selected for their ease of use, versatility, and suitability for the task at hand. The goal is to use the best available tools and technologies to gain insights into the risk factors associated with diabetes in the Pima Indian community and to provide recommendations for reducing its impact