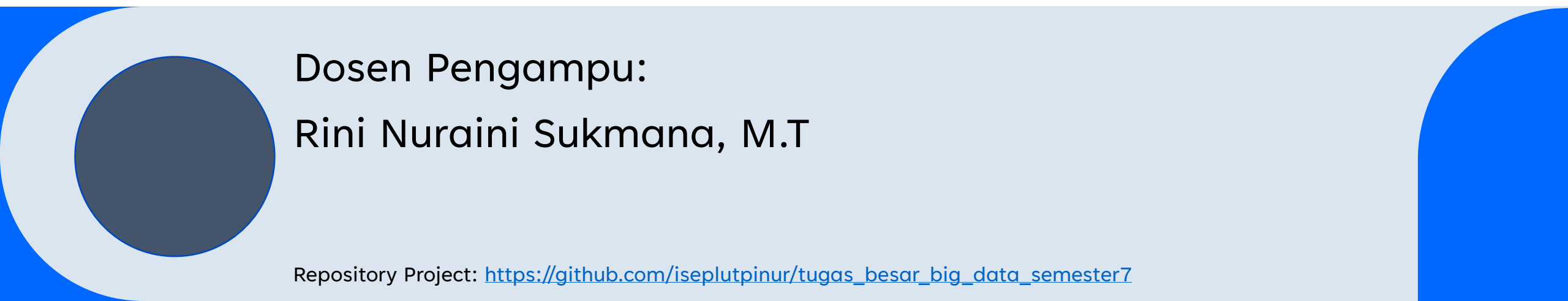


# ***EKSPLORASI, VISUALISASI DAN KLASIFIKASI DATA MENGGUNAKAN METODE DECISION TREE STUDI KASUS KORBAN TENGGEAMNYA KAPAL TITANIC MENGGUNAKAN BAHASA R***

Kelompok:

1. Isep lutpi nur (2113191079)
2. Farhan azis (2113191097)
3. M. Taufiq hidayatuloh (2113191036)
4. Dara atria ferliandini (2113191098)



Dosen Pengampu:  
Rini Nuraini Sukmana, M.T

# Agenda

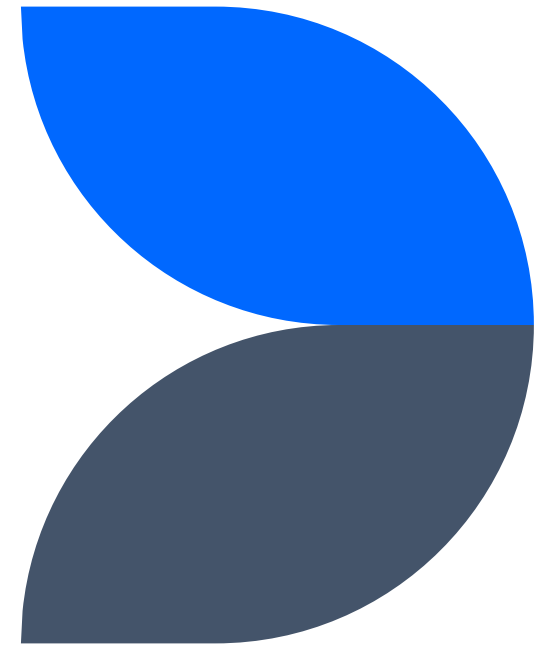
Latar Belakang

Tools Yang Digunakan

Implementasi

Kesimpulan dan hasil

# Latar Belakang



# Latar Belakang

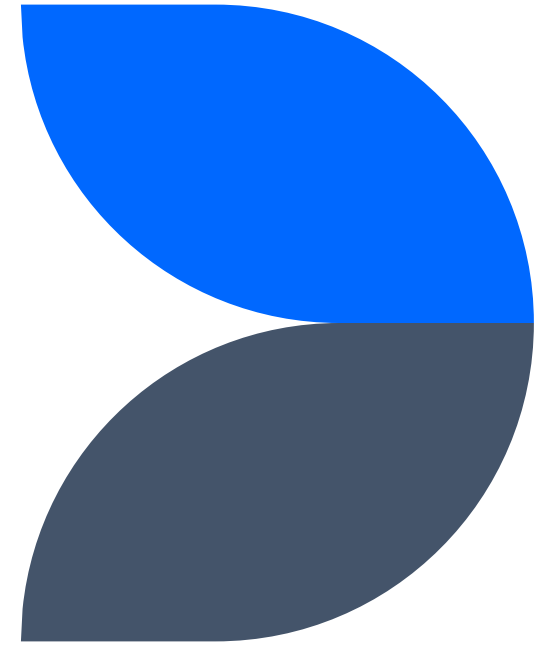
Bencana adalah musibah yang menimpa setiap makhluk hidup dunia, termasuk manusia. Salah satu bencana, fenomena kapal tengelam yang menarik perhatian dunia adalah peristiwa tenggelamnya kapal titanic. Dari data penumpang yang ada akan di ekspolorasi, visualisasi dan klasifikasi menggunakan metode Decision tree untuk memodelkan kemungkinan resiko terjadi hal yang sama.

# Kapal Titanic

**RMS *Titanic*** adalah sebuah kapal penumpang super Britania Raya yang tenggelam di Samudra Atlantik Utara pada tanggal 15 April 1912 setelah menabrak sebuah gunung es pada pelayaran perdananya dari Southampton, Inggris ke New York City. Tenggelamnya Titanic mengakibatkan kematian sebanyak 1.514 orang dalam salah satu bencana maritim masa damai paling mematikan sepanjang sejarah. Titanic merupakan kapal terbesar di dunia pada pelayaran perdananya. Satu dari tiga kapal samudra kelas Olympic dioperasikan oleh White Star Line. Kapal ini dibangun pada 1909 sampai 1911 oleh galangan kapal Harland and Wolff di Belfast. Kapal ini sanggup mengangkut 2.224 penumpang.



# Tools Yang Digunakan



# Bahasa R

R adalah sebuah program komputasi statistika dan grafis (R Core Team 2022). Saat ini R sudah dikenal luas sebagai salah satu *powerful software* untuk analisis data dan *Data Science*. Tentu saja selain R masih banyak *software* lain yang juga sering digunakan untuk analisis data, misalnya Python. R dibuat dengan tujuan awal untuk komputasi statistika dan grafis.

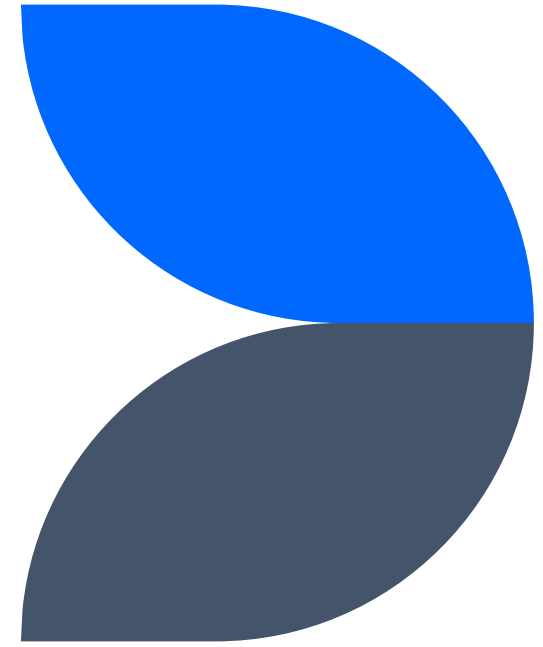


# R Studio

RStudio adalah *Integrated Development Environment* (IDE) untuk R yang banyak digunakan hingga saat ini. Dapat dikatakan bahwa hampir semua pengguna R yang sudah mengetahui RStudio akan lebih memilih menggunakan R melalui RStudio dibandingkan dengan menggunakan R GUI.



# Implementasi



# Eksplorasi Data


## Dataset

Data yang kami miliki bersumber dari website

Kaggle. Link data:

<https://www.kaggle.com/datasets/brendan45774/test-file>

## Dimensi data

```
R 4.2.2 . ~/   
> dim(titanic)  
[1] 891 12  
> |
```

Mempunyai 12  
variable dan 891  
jumlah data

## Struktur Data

```
> str(titanic) # struktur data
'data.frame':  891 obs. of  12 variables:
 $ PassengerId: int  1 2 3 4 5 6 7 8 9 10 ...
 $ survived   : int  0 1 1 1 0 0 0 0 1 1 ...
 $ Pclass     : int  3 1 3 1 3 3 1 3 3 2 ...
 $ Name       : chr   "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs
Thayer)" "Heikkinen, Miss. Laina" "Futrelle, Mrs. Jacques Heath (Lily May Peel)" ...
 $ Sex        : chr   "male" "female" "female" "female" ...
 $ Age        : num  22 38 26 35 35 NA 54 2 27 14 ...
 $ Sibsp      : int  1 1 0 1 0 0 0 3 0 1 ...
 $ Parch      : int  0 0 0 0 0 0 0 1 2 0 ...
 $ Ticket     : chr   "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803" ...
 $ Fare       : num  7.25 71.28 7.92 53.1 8.05 ...
 $ Cabin      : chr   "" "C85" "" "C123" ...
 $ Embarked   : chr   "S" "C" "S" "S" ...
```

## Ringkasan Data

```
> summary(titanic) # ringkasan data
```

PassengerId	Survived	Pclass	Name
Min. : 1.0	Min. :0.0000	Min. :1.000	Length:891
1st Qu.:223.5	1st Qu.:0.0000	1st Qu.:2.000	Class :character
Median :446.0	Median :0.0000	Median :3.000	Mode :character
Mean :446.0	Mean :0.3838	Mean :2.309	
3rd Qu.:668.5	3rd Qu.:1.0000	3rd Qu.:3.000	
Max. :891.0	Max. :1.0000	Max. :3.000	

Sex	Age	Sibsp	Parch
Length:891	Min. : 0.42	Min. :0.000	Min. :0.0000
Class :character	1st Qu.:20.12	1st Qu.:0.000	1st Qu.:0.0000
Mode :character	Median :28.00	Median :0.000	Median :0.0000
	Mean :29.70	Mean :0.523	Mean :0.3816
	3rd Qu.:38.00	3rd Qu.:1.000	3rd Qu.:0.0000
	Max. :80.00	Max. :8.000	Max. :6.0000
	NA's :177		

Ticket	Fare	Cabin	Embarked
Length:891	Min. : 0.00	Length:891	Length:891
Class :character	1st Qu.: 7.91	Class :character	Class :character
Mode :character	Median :14.45	Mode :character	Mode :character
	Mean :32.20		
	3rd Qu.:31.00		
	Max. :512.33		

## Data Paling Atas

```
> head(titanic) # data paling atas
```

	x	pclass	survived	name	sex	age	sibsp	parch	ticket
330	330	2	1	Angle, Mrs. William A (Florence 'Mary' Agnes Hughes)	female	36	1	0	226875
1047	1047	3	0	Naidenoff, Mr. Penko	male	22	0	0	349206
1145	1145	3	0	Rice, Master. Eugene	male	2	4	1	382652
1289	1289	3	0	wiklund, Mr. Jakob Alfred	male	18	1	0	3101267
692	692	3	0	Brocklebank, Mr. William Alfred	male	35	0	0	364512
1088	1088	3	0	Olsson, Mr. Nils Johan Goransson	male	28	0	0	347464
	fare	cabin	embarked	home.dest					
330	26	?	S	warwick, England					
1047	7.8958	?	S	?					
1145	29.125	?	Q	?					
1289	6.4958	?	S	?					
692	8.05	?	S	Broomfield, Chelmsford, England					
1088	7.8542	?	S	?					

```
> |
```

## Data Paling Bawah

```
> tail(titanic) # data paling bawah
```

	PassengerId	Survived	Pclass	Name	Sex	Age
886	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39
887	887	0	2	Montvila, Rev. Juozas	male	27
888	888	1	1	Graham, Miss. Margaret Edith	female	19
889	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NA
890	890	1	1	Behr, Mr. Karl Howell	male	26
891	891	0	3	Dooley, Mr. Patrick	male	32

	sibsp	Parch	Ticket	Fare	Cabin	Embarked
886	0	5	382652	29.125		Q
887	0	0	211536	13.000		S
888	0	0	112053	30.000	B42	S
889	1	2	w./c. 6607	23.450		S
890	0	0	111369	30.000	C148	C
891	0	0	370376	7.750		Q

# VISUALISAI DATA

**Mengubah Kolom yang class menjadi sebuah factor**

Untuk keterbacaan contoh nya di data Survied (Orang yang bertahan hidup) hanya nilai 1 untuk selamat dan 0 yang meninggal. Maka di ganti 1 menjadi survived dan 0 menjadi died. Sebelumnya variable dataset utama yaitu di beri nama titanic.

# Mengubah Kolom yang class menjadi sebuah factor

Sebelum

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
1	1	0	3 Braund, Mr. Owen Harris	male	22.00	1	0
2	2	1	1 Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.00	1	0
3	3	1	3 Heikkinen, Miss. Laina	female	26.00	0	0
4	4	1	1 Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.00	1	0
5	5	0	3 Allen, Mr. William Henry	male	35.00	0	0
6	6	0	3 Moran, Mr. James	male	NA	0	0
7	7	0	1 McCarthy, Mr. Timothy J	male	54.00	0	0
8	8	0	3 Palsson, Master. Gosta Leonard	male	2.00	3	1
9	9	1	3 Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.00	0	2
10	10	1	2 Nasser, Mrs. Nicholas (Adele Achem)	female	14.00	1	0
11	11	1	3 Sandstrom, Miss. Marguerite Rut	female	4.00	1	1

Sesudah

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp
1	died	3	Braund, Mr. Owen Harris	male	22.00	1
2	survived	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.00	1
3	survived	3	Heikkinen, Miss. Laina	female	26.00	0
4	survived	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.00	1
5	died	3	Allen, Mr. William Henry	male	35.00	0
6	died	3	Moran, Mr. James	male	NA	0
7	died	1	McCarthy, Mr. Timothy J	male	54.00	0
8	died	3	Palsson, Master. Gosta Leonard	male	2.00	3
9	survived	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.00	0
10	survived	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.00	1
11	survived	3	Sandstrom, Miss. Marguerite Rut	female	4.00	1

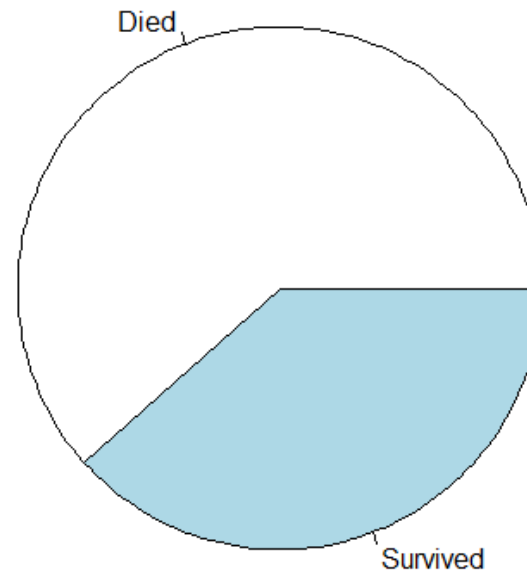


## Distribusi class menggunakan pie chart

Untuk melihat perbandingan class(Data bertahan hidup/survived) menggunakan pie chart.

```
> survivedTable = table(titanic$survived)  
> survivedTable
```

```
   died survived  
    549     342
```



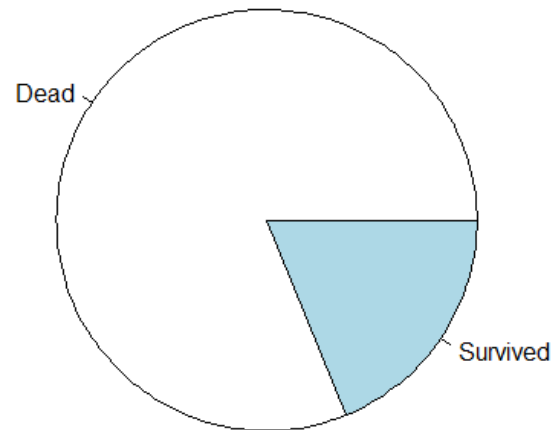
## Perbandingan korban berdasarkan jenis kelamin menggunakan pie chart

Data di bagi menjadi berdasarkan jenis kelamin kemudian dilihat perbandingan nya melalui pie chart.

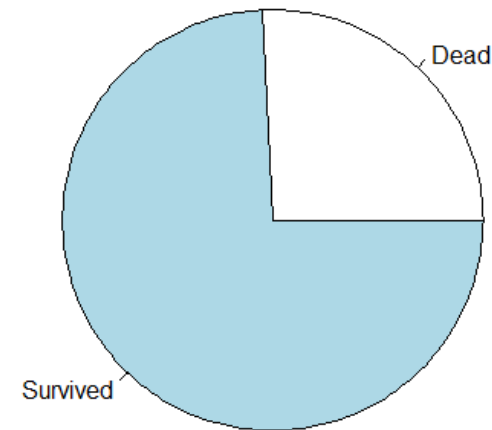
```
> survivedTable = table(titanic$Survived)
> survivedTable
```

died	survived
549	342

Perbandingan Korban Penumpang Pria



Perbandingan Korban Penumpang Wanita

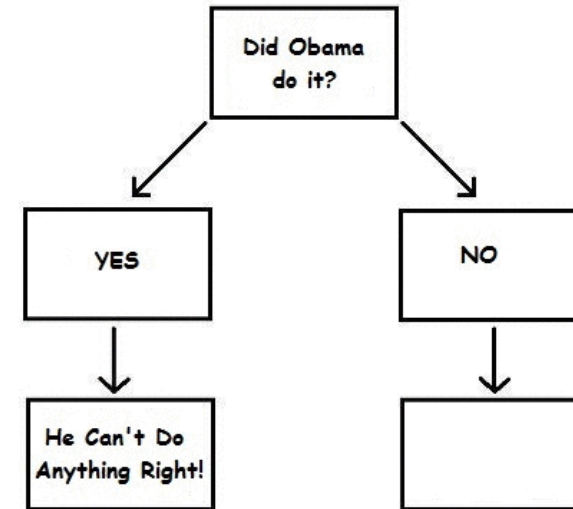


# KLASIFIKASI DATA

## Decision Tree

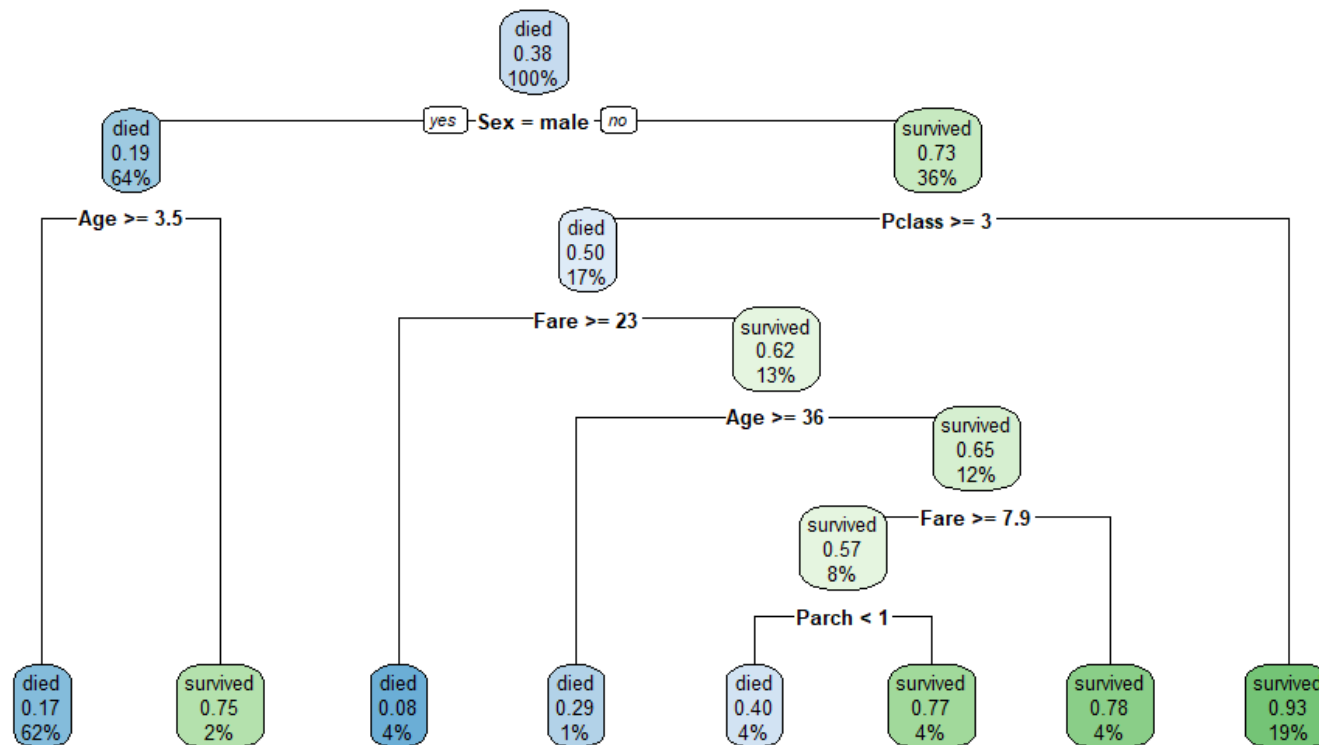
Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko.

Republican Decision Tree



## Decision Tree

### Pohon keputusan



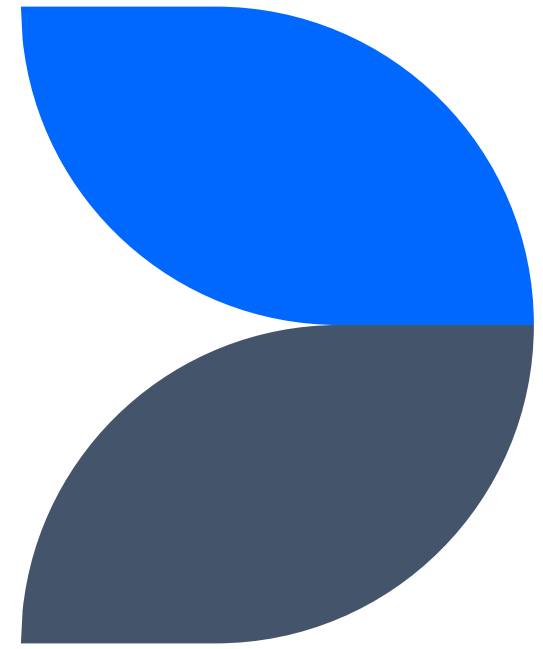
## Confusion Matrix

```
> cm <- table(predict(tree), trainData$Survived)
> print(cm)
```

	died	survived
died	343	70
survived	42	171

result_accuracy	0.821086261980831
-----------------	-------------------

# Kesimpulan Dan Saran



# Kesimpulan dan hasil

Dari hasil visualisasi data terdapat 549 orang meninggal dan 342 orang selamat dari data yang ada, kemudian dari perbandingan menurut jenis kelamin bahwa penumpang Wanita lebih banyak yang selamat dibandingkan penumpang pria, kemudian pada tahap klasifikasi data menggunakan metode Decision tree terdapat terbentuk pola model dari data yang ada dengan tingkat akurasi sekitar 82,10%.



**Terima Kasih**