

**PRAKTIKUM DATA MINING DAN DATA WAREHOUSING**

**MODUL 7**



**Nama : Nicky Julyatrika Sari**

**Nim : L200200101**

**Kelas : D**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA  
TAHUN 2022/2023**

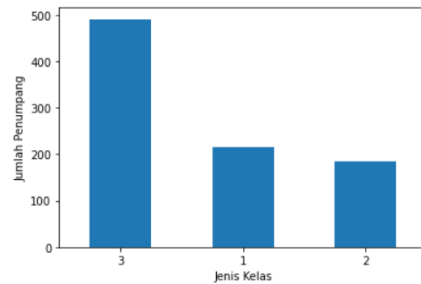
1. Lakukan kembali langkah 13 pada prosedur praktikum untuk melihat data atribut lainnya dengan grafik batang

- a. Pclass

In [5]: #L200200101\_NickyJs

```
dt = train_data['Pclass'].value_counts()
plt = dt.plot(kind = 'bar', rot=0)
plt.set_xlabel('Jenis Kelas')
plt.set_ylabel('Jumlah Penumpang')
```

Out[5]: Text(0, 0.5, 'Jumlah Penumpang')

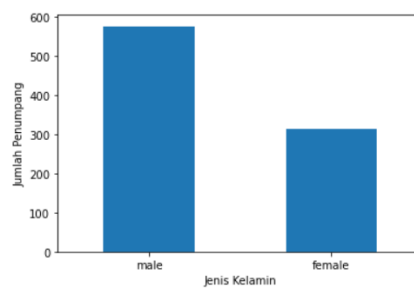


- b. Sex

In [2]: #L200200101\_NickyJs

```
dt = train_data['Sex'].value_counts()
plt = dt.plot(kind = 'bar', rot=0)
plt.set_xlabel('Jenis Kelamin')
plt.set_ylabel('Jumlah Penumpang')
```

Out[2]: Text(0, 0.5, 'Jumlah Penumpang')

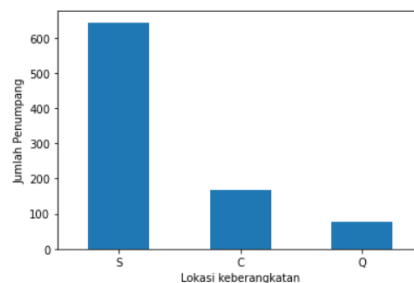


- c. Embarked

In [4]: #L200200101\_NickyJs

```
dt = train_data['Embarked'].value_counts()
plt = dt.plot(kind = 'bar', rot=0)
plt.set_xlabel('Lokasi keberangkatan')
plt.set_ylabel('Jumlah Penumpang')
```

Out[4]: Text(0, 0.5, 'Jumlah Penumpang')



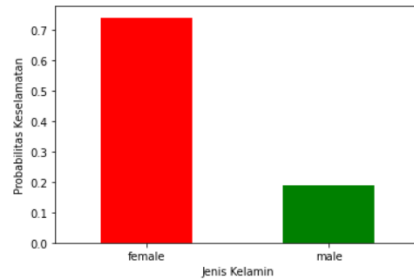
2. Lakukan kembali langkah 14 pada prosedur praktikum untuk melihat probabilitas keselamatan (Survived) berdasarkan :

- a. Jenis kelamin (Sex)

In [5]: #L200200101\_NickyJs

```
dt = train_data[['Sex', 'Survived']].groupby(
    'Sex').mean().Survived
plt = dt.plot(kind = 'bar', rot = 0, color = ['red', 'green', 'blue'])
plt.set_xlabel('Jenis Kelamin')
plt.set_ylabel('Probabilitas Keselamatan')
```

Out[5]: Text(0, 0.5, 'Probabilitas Keselamatan')

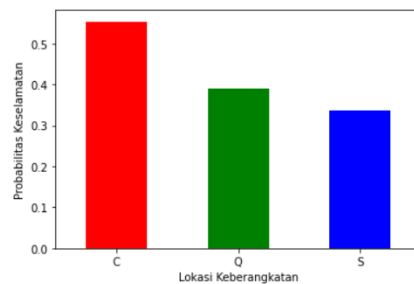


## b. Lokasi keberangkatan (Embarked)

In [6]: #L200200101\_NickyJs

```
dt = train_data[['Embarked', 'Survived']].groupby(
    'Embarked').mean().Survived
plt = dt.plot(kind = 'bar', rot = 0, color = ['red', 'green', 'blue'])
plt.set_xlabel('Lokasi Keberangkatan')
plt.set_ylabel('Probabilitas Keselamatan')
```

Out[6]: Text(0, 0.5, 'Probabilitas Keselamatan')

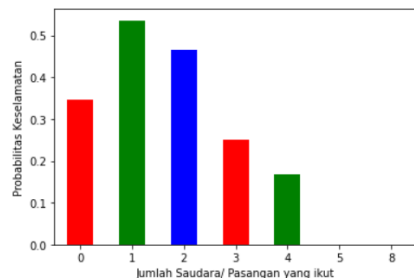


## c. Jumlah saudara/pasangan yang ikut (SibSp)

In [7]: #L200200101\_NickyJs

```
dt = train_data[['SibSp', 'Survived']].groupby(
    'SibSp').mean().Survived
plt = dt.plot(kind = 'bar', rot = 0, color = ['red', 'green', 'blue'])
plt.set_xlabel('Jumlah Saudara/ Pasangan yang ikut')
plt.set_ylabel('Probabilitas Keselamatan')
```

Out[7]: Text(0, 0.5, 'Probabilitas Keselamatan')

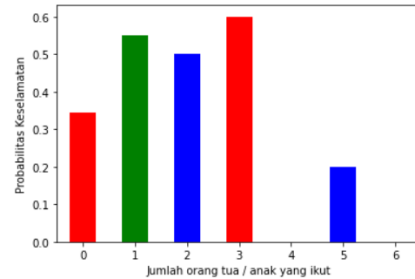


## d. Jumlah orang tua/anak yang ikut (Parch)

```
In [8]: #L200200101_NickyJs

dt = train_data[['Parch', 'Survived']].groupby(
    'Parch').mean().Survived
plt = dt.plot(kind = 'bar', rot = 0, color = ['red', 'green', 'blue'])
plt.set_xlabel('Jumlah orang tua / anak yang ikut')
plt.set_ylabel('Probabilitas Keselamatan')
```

Out[8]: Text(0, 0.5, 'Probabilitas Keselamatan')



3. Ulangi kembali langkah 15 pada prosedur praktikum untuk melihat multidimensi terhadap atribut

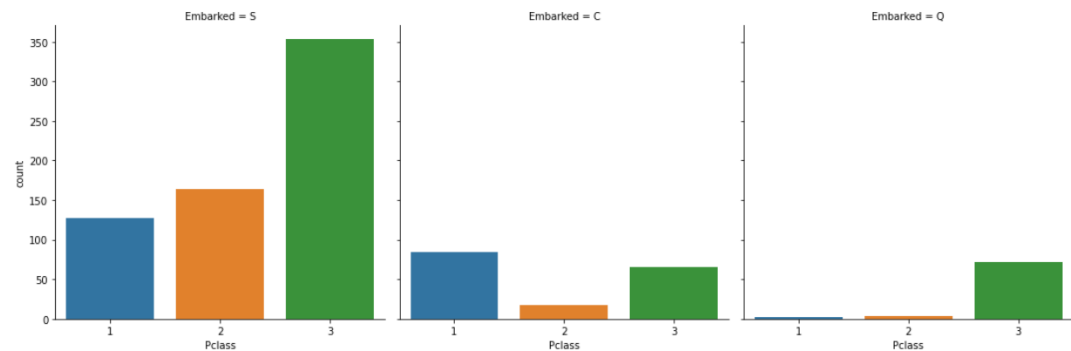
a. Kelas penumpang (Pclass) vs lokasi keberangkatan (Embarked)

```
In [5]: #L200200101_NickyJs

sns.catplot('Pclass', col = 'Embarked', data = train_data, kind = 'count')
```

C:\Users\Lenovo\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(

Out[5]: <seaborn.axisgrid.FacetGrid at 0x18c8b62bdf0>



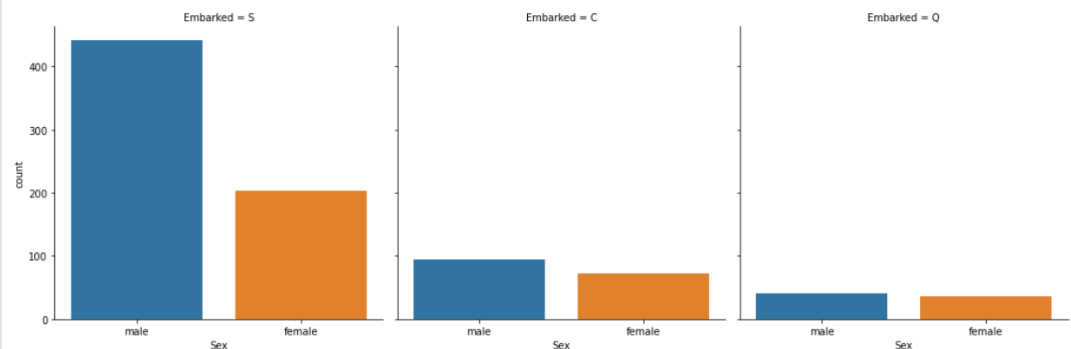
b. Jenis kelamin (Sex) vs lokasi keberangkatan (Embarked)

```
In [4]: #L200200101_NickyJs

sns.catplot('Sex', col = 'Embarked', data = train_data, kind = 'count')
```

C:\Users\Lenovo\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(

Out[4]: <seaborn.axisgrid.FacetGrid at 0x18c8aeca70>



4. Ubahlah data sebutan/gelar penumpang (Title) menjadi data angka dengan ketentuan sebagai berikut: Master: 0, Miss: 1, Mr: 2, Mrs: 3, Others: 4

```
In [11]: #L200200101_NickyJs

train_data['Title'] = train_data['Name'].str.extract('([A-Za-z]+)\.', expand=False)
train_data = train_data.drop(columns='Name')

train_data['Title'] = train_data['Title'].replace(['Dr', 'Rev', 'Col', 'Major', 'Countess', 'Sir', 'Jonkheer', 'Lady', 'Capt', 'Don'], 'other')
train_data['Title'] = train_data['Title'].replace('Ms', 'Miss')
train_data['Title'] = train_data['Title'].replace('Mme', 'Mrs')
train_data['Title'] = train_data['Title'].replace('Mlle', 'Miss')
```

```
In [12]: #L200200101_NickyJs
```

```
train_data.head()
```

```
Out[12]:
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S	Mr
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C	Mrs
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	Miss
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S	Mrs
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S	Mr

```
In [13]: #L200200101_NickyJs
```

```
train_data['Title'].unique()
```

```
Out[13]: array(['Mr', 'Mrs', 'Miss', 'Master', 'others'], dtype=object)
```

```
In [14]: #L200200101_NickyJs
```

```
train_data['Title'] = train_data['Title'].map({'Master':0, 'Miss':1, 'Mr':2, 'Mrs':3, 'others':4})
train_data.head()
```

```
Out[14]:
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S	2
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C	3
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	1
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S	3
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S	2

5. Carilah nilai korelasi antar atribut termasuk atribut Title setelah diubah menjadi data angka dengan menggunakan heatmap

```
In [15]: #L200200101_NickyJs
```

```
corr_matrix = train_data.corr()
```

```
In [16]: #L200200101_NickyJs
```

```
import matplotlib.pyplot as plt
corr_matrix = train_data.corr()
plt.figure(figsize=(9,8))
sns.heatmap(data = corr_matrix, cmap='BrBG', annot=True, linewidths= 0.2)
```

Out[16]: <AxesSubplot:>

