In [1]:

```
import warnings
warnings.filterwarnings('ignore')
warnings.filterwarnings('ignore', category = DeprecationWarni
ng)
```

In [2]:

```
import numpy as np
import pandas as pa
import seaborn as sb
```

In [3]:

```
import matplotlib.pyplot as plt
%matplotlib inline
import pylab as plot
params ={'axes.labelsize':"large",
         'xtick.labelsize': 'x-large',
         'legend.fontsize':20,
         'figure.dpi':150,
         'figure.figsize':[25,7]
}
plot.rcParams.update(params)
```

In [5]:

```
data = pa.read_csv(r'C:\Users\SACHIN K M\Desktop\python\data\
datasets\train.csv')
```

In [6]:

```
data.head()
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 175
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	1138
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450

In [7]:

```
print(data.shape)
```

```
(891, 12)
```

In [8]:

```
data.describe()
```

Out[8]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057

min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000

In [9]:

```
data['Age'] = data['Age'].fillna(data['Age'].median())
```

In [10]:

```
data.describe()
```

Out[10]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.361582	0.523008	0.381594
std	257.353842	0.486592	0.836071	13.019697	1.102743	0.806057
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000

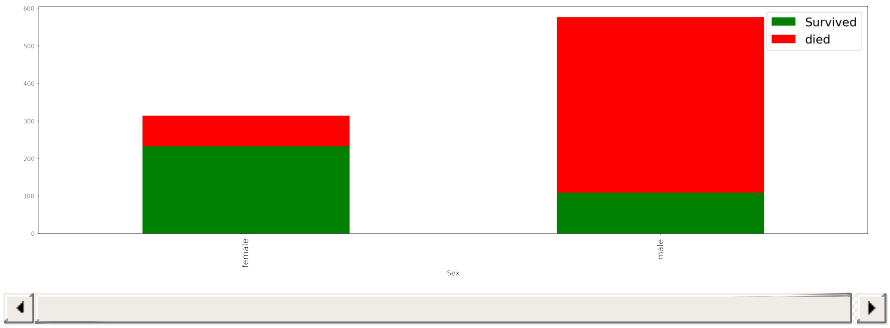
In [11]:

```
data['died']= 1-data['Survived']
```

In [12]:

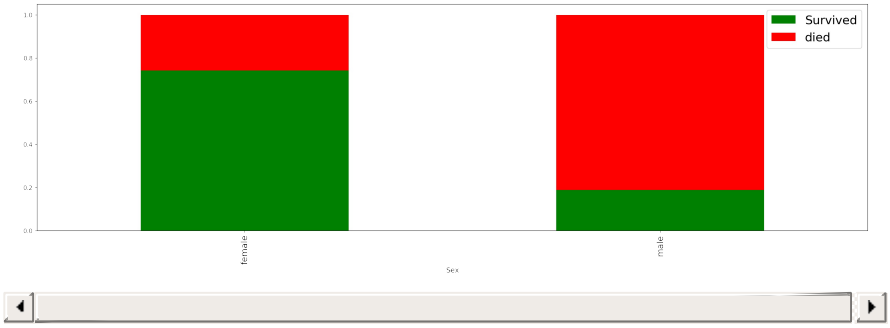
```
data.groupby('Sex').agg('sum')[['Survived','died']].plot(kind
```

```
 ='bar', figsize = (25,7),
                                                    stack
ed = True, color=['g', 'r']);
```



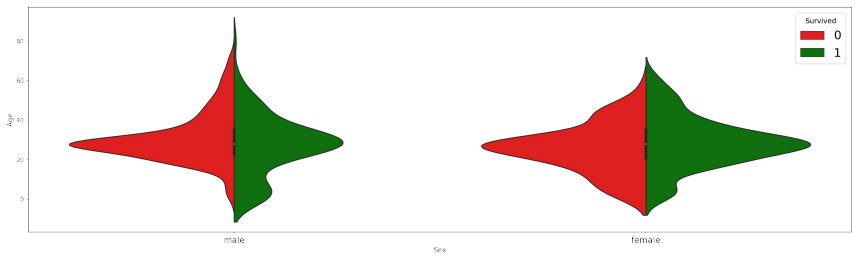
In [13]:

```
data.groupby('Sex').agg('mean')[['Survived','died']].plot(kin
d ='bar', figsize = (25,7),
                                                    stack
ed = True, color=['g', 'r']);
```



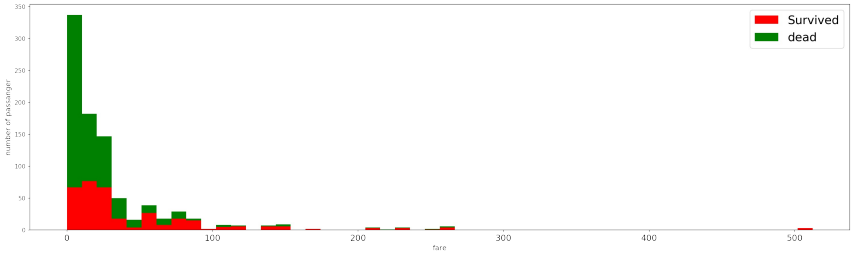
In [14]:

```
sb.violinplot(x = 'Sex', y = 'Age',
              hue = 'Survived', data = data,
               split = True,
              palette={0:'r', 1:'g'});
```



In [15]:

```
plt.hist([data[data['Survived']==1]['Fare'],data[data['Survived']==0]['Fare']],
         stacked=True, color=['r','g'],
         bins= 50, label= ['Survived', 'dead'])
plt.xlabel('fare')
plt.ylabel('number of passanger')
plt.legend();
```



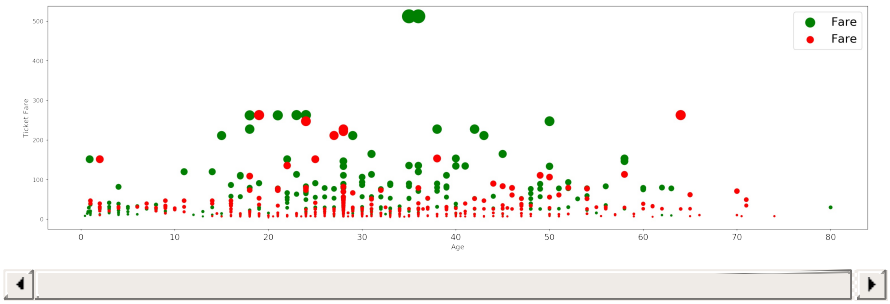
In [16]:

```
ax=plt.subplot()

ax.scatter(data[data['Survived']==1]['Age'],data[data['Survived']==1]['Fare'],
           c= 'green', s= data[data['Survived']==1]['Fare'])
ax.scatter(data[data['Survived']==0]['Age'],data[data['Survived']==0]['Fare'],
           c = 'red', s = data[data['Survived']==0]['Fare']);
plt.xlabel('Age')
plt.ylabel('Ticket Fare')
plt.legend()
```

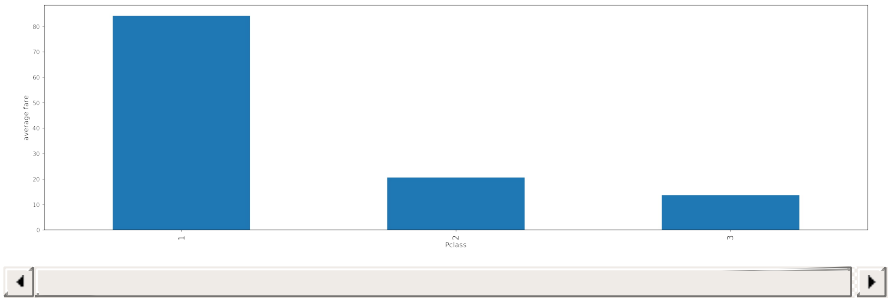
Out[16]:

```
<matplotlib.legend.Legend at 0x17b942c9400>
```



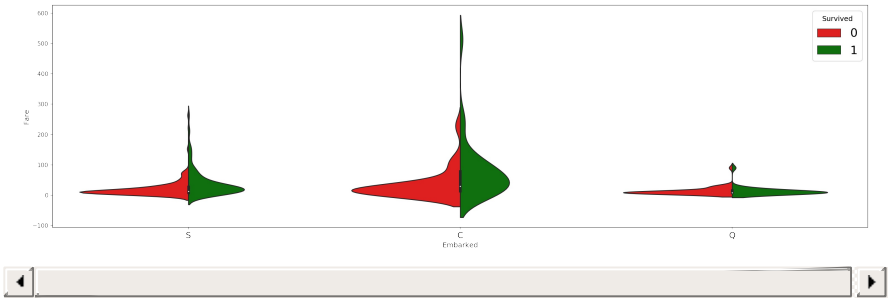
In [17]:

```
ax = plt.subplot()
ax.set_ylabel('average fare')
data.groupby('Pclass').mean()['Fare'].plot(kind='bar', figsiz
e=(25,7), ax=ax);
```



In [18]:

```
sb.violinplot(x='Embarked', y='Fare', hue= 'Survived', data=d
ata, split= True, palette= {0:'r', 1:'g'});
```



In [19]:

```
def status(feature):
    print('processing', feature , ': ok')
```

In [22]:

```
def get_combined_data():
    train = pa.read_csv(r'C:\Users\SACHIN K M\Desktop\python\
data\datasets\train.csv')
    test = pa.read_csv(r'C:\Users\SACHIN K M\Desktop\python\d
ata\datasets\test.csv')
    targets = train.Survived
    train.drop(['Survived'], 1, inplace=True)
    combined = train.append(test)
    combined.reset_index(inplace = True)
    combined.drop(['index', 'PassengerId'], inplace=True, axi
s = 1)

    return combined
```

In [23]:

```
combined = get_combined_data()
```

In [24]:

```
print (combined.shape)
```

```
(1309, 10)
```

In [25]:

```
combined.head()
```

Out[25]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	1	Cumings, Mrs. John Bradley	female	38.0	1	0	PC 17599	71.2833	C85	

		(Florence Briggs Th...							
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [26]:

```
title = set()
for name in data['Name']:
    title.add(name.split(',')[1].split('.')[0].strip())
```

In [27]:

```
print (title)
```

```
{'Major', 'Rev', 'Col', 'Sir', 'Mme', 'Ms', 'D
on', 'Capt', 'Jonkheer', 'Mr', 'Mrs', 'Mlle',
'Miss', 'Master', 'Lady', 'the Countess', 'Dr'
}
```

In [28]:

```
Title_Dictionary = {
    "Capt": "Officer",
    "Col": "Officer",
    "Major": "Officer",
    "Jonkheer": "Royalty",
    "Don": "Royalty",
```

```
    "Sir" : "Royalty",
    "Dr": "Officer",
    "Rev": "Officer",
    "the Countess":"Royalty",
    "Mme": "Mrs",
    "Mlle": "Miss",
    "Ms": "Mrs",
    "Mr" : "Mr",
    "Mrs" : "Mrs",
    "Miss" : "Miss",
    "Master" : "Master",
    "Lady" : "Royalty"
}

def get_titles():
    # we extract the title from each name
    combined['Title'] = combined['Name'].map(lambda name:name
.split(',')[1].split('.')[0].strip())

    # a map of more aggregated title
    # we map each title
    combined['Title'] = combined.Title.map(Title_Dictionary)
    status('Title')
    return combined
```

In [29]:

```
combined = get_titles()
```

```
processing Title : ok
```

In [30]:

```
combined.head()
```

Out[30]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	

In [31]:

```
combined[combined['Title'].isnull()]
```

Out[31]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarke
1305	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9	C105	

In [32]:

```
combined.iloc[:891].Age.isnull().sum()
```

Out[32]:

177

In [33]:

```
combined.iloc[891:].Age.isnull().sum()
```

Out[33]:

86

In [34]:

```
grouped_train = combined.iloc[:891].groupby(['Sex', 'Pclass',
 'Title'])
grouped_median_train = grouped_train.median()
grouped_median_train = grouped_median_train.reset_index()[['S
ex', 'Pclass', 'Title', 'Age']]
```

In [35]:

```
grouped_median_train.head()
```

Out[35]:

	Sex	Pclass	Title	Age
0	female	1	Miss	30.0
1	female	1	Mrs	40.0
2	female	1	Officer	49.0
3	female	1	Royalty	40.5
4	female	2	Miss	24.0

In [36]:

```
grouped_median_train
```

Out[36]:

	Sex	Pclass	Title	Age
0	female	1	Miss	30.0
1	female	1	Mrs	40.0
2	female	1	Officer	49.0
3	female	1	Royalty	40.5
4	female	2	Miss	24.0
5	female	2	Mrs	31.5
6	female	3	Miss	18.0
7	female	3	Mrs	31.0
8	male	1	Master	4.0
9	male	1	Mr	40.0
10	male	1	Officer	51.0
11	male	1	Royalty	40.0
12	male	2	Master	1.0
13	male	2	Mr	31.0
14	male	2	Officer	46.5
15	male	3	Master	4.0
16	male	3	Mr	26.0

In [37]:

```
def fill_age(row):
    condition = (
    (grouped_median_train['Sex'] == row['Sex'])&
        (grouped_median_train['Title'] == row['Title'])&
        (grouped_median_train['Pclass'] == row['Pclass'])
    )
```

```
    return grouped_median_train[condition]['Age'].values[0]


def process_age():
    global combined
    combined ['Age'] = combined.apply(lambda row:fill_age(row
) if np.isnan(row['Age']) else row['Age'], axis= 1)
    status('age')
    return combined
```

In [38]:

```
combined = process_age()
```

```
processing age : ok
```

In [39]:

```
combined.head()
```

Out[39]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	1	Futrelle, Mrs. Jacques Heath (Lily May	female	35.0	1	0	113803	53.1000	C123	

		Peel)							
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [40]:

```
combined.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 11 columns):
Pclass      1309 non-null int64
Name        1309 non-null object
Sex         1309 non-null object
Age         1309 non-null float64
SibSp       1309 non-null int64
Parch       1309 non-null int64
Ticket      1309 non-null object
Fare        1308 non-null float64
Cabin       295 non-null object
Embarked    1307 non-null object
Title       1308 non-null object
dtypes: float64(2), int64(3), object(6)
memory usage: 112.6+ KB
```

In [41]:

```
combined.describe()
```

Out[41]:

	Pclass	Age	SibSp	Parch	Fare
count	1309.000000	1309.000000	1309.000000	1309.000000	1308.000000
mean	2.294882	29.213270	0.498854	0.385027	33.295479
std	0.837836	13.400994	1.041658	0.865560	51.758668

min	1.000000	0.170000	0.000000	0.000000	0.000000
25%	2.000000	21.000000	0.000000	0.000000	7.895800
50%	3.000000	26.000000	0.000000	0.000000	14.454200
75%	3.000000	36.500000	1.000000	0.000000	31.275000
max	3.000000	80.000000	8.000000	9.000000	512.329200

In [42]:

```
def process_name():
    global combined

    Titles_dummies = pa.get_dummies(combined['Title'], prefix
='Title')
    combined = pa.concat([combined, Titles_dummies], axis = 1
)

    combined.drop('Title', axis = 1, inplace = True)

    status('names')
    return combined
```

In [43]:

```
combined = process_name()
```

```
processing names : ok
```

In [44]:

```
combined.head()
```

Out[44]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	

1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [45]:

```
def process_fares():
    global combined
    combined.Fare.fillna(combined.iloc[:891].Fare.mean(), inp
lace=True)
    status('Fare')
    return combined
```

In [46]:

```
combined = process_fares()
```

```
processing Fare : ok
```

In [47]:

```
combined[combined['Embarked'].isnull()]
```

Out[47]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
61	1	Icard, Miss. Amelie	female	38.0	0	0	113572	80.0	B28	NaN
829	1	Stone, Mrs. George Nelson (Martha Evelyn)	female	62.0	0	0	113572	80.0	B28	NaN

In [48]:

```
def process_embarked():
    global combined
    combined.Embarked.fillna('S', inplace = True)
    embarked_dummies = pa.get_dummies(combined['Embarked'], p
refix= 'Embarked')
    combined = pa.concat([combined, embarked_dummies], axis =
 1)
    combined.drop('Embarked', axis = 1, inplace = True)
    status('Embarked')

    return combined
```

In [49]:

```
combined = process_embarked()
```

```
processing Embarked : ok
```

In [50]:

```
combined.head()
```

Out[50]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Title

0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [51]:

```
train_cabin, test_cabin = set(), set()

for c in combined.iloc[:891]['Cabin']:
    try:
        train_cabin.add(c[0])
    except:
        train_cabin.add('u')

for c in combined.iloc[891:]['Cabin']:
    try:
        test_cabin.add(c[0])
    except:
        test_cabin.add('u')
```

In [52]:

```
print (train_cabin)
```

```
{'G', 'A', 'u', 'B', 'E', 'F', 'C', 'T', 'D'}
```

In [53]:

```
print (test_cabin)
```

```
{'G', 'A', 'B', 'u', 'E', 'F', 'C', 'D'}
```

In [54]:

```
def process_cabin():
    global combined
    combined.Cabin.fillna('u', inplace= True)
    combined['Cabin'] = combined['Cabin'].map(lambda c: c[0])
    cabin_dummies = pa.get_dummies(combined['Cabin'], prefix
= 'Cabin')
    combined = pa.concat([combined, cabin_dummies], axis = 1)

    combined.drop('Cabin', axis = 1, inplace = True)
    status('Cabin')
    return combined
```

In [55]:

```
combined = process_cabin()
```

```
processing Cabin : ok
```

In [56]:

```
combined.head()
```

Out[56]:

	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Title_Master

0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

5 rows × 26 columns

In [57]:

```
def precessing_sex():
    global combined
    combined['Sex'] = combined['Sex'].map({'male':1, 'female':0})
    status('Sex')
    return combined
```

In [58]:

```
combined = precessing_sex()
```

```
processing Sex : ok
```

In [59]:

```
def processing_pclass():
    global combined
    pclass_dummies = pa.get_dummies(combined['Pclass'], prefi
x = "Pclass")
    combined = pa.concat([combined, pclass_dummies], axis = 1
)
    combined.drop ('Pclass', axis = 1, inplace= True)
    status('Pclass')
    return combined
```

In [60]:

```
combined = processing_pclass()
```

```
processing Pclass : ok
```

In [61]:

```
combined.head()
```

Out[61]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Title_Master	Title_Mis
0	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	0	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	0	
2	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	0	
3	Futrelle, Mrs. Jacques Heath (Lily May	0	35.0	1	0	113803	53.1000	0	

	Peel)								
4	Allen, Mr. William Henry	1	35.0	0	0	373450	8.0500		0

5 rows × 28 columns

In [62]:

```
def cleanticket(ticket):
    ticket = ticket.replace(',', '')
    ticket = ticket.replace('/', '')
    ticket = ticket.split()
    ticket = map(lambda t : t.strip(), ticket)
    ticket = list(filter(lambda t: not t.isdigit(), ticket))
    if len(ticket) > 0:
        return ticket[0]
    else:
        return 'xxx'
```

In [63]:

```
tickets = set()
for t in combined['Ticket']:
    tickets.add(cleanticket(t))
```

In [64]:

```
print (len(tickets))
```

```
49
```

In [65]:

```
def processingticket():
    global combined

    def cleanTicket(ticket):
        ticket = ticket.replace('.', '')
```

```
        ticket = ticket.replace('/', '')
        ticket = ticket.split()
        ticket = map(lambda t : t.strip(), ticket)
        ticket = list(filter(lambda t : not t.isdigit(), tick
et))
        if len(ticket) > 0:
            return ticket[0]
        else:
            return 'XXX'



    combined['Ticket'] = combined['Ticket'].map(cleanticket)
    tickets_dummies= pa.get_dummies(combined['Ticket'], prefi
x='Ticket')
    combined = pa.concat([combined, tickets_dummies], axis =1
)
    combined.drop('Ticket', inplace=True, axis =1)
    status('Ticket')
    return combined
```

In [66]:

```
combined = processingticket()
```

```
processing Ticket : ok
```

In [67]:

```
def processing_family():
    global combined
    combined['Familysize'] = combined['Parch'] + combined['Si
bSp'] + 1
    combined['Singleton'] = combined['Familysize'].map(lambda
 s:1 if s==1 else 0)
    combined['SmallFamily'] = combined['Familysize'].map(lamb
da s:1 if 2<= s <=4 else 0)
```

```
    combined['LargeFamily'] = combined['Familysize'].map(lambda s:1 if s<=5 else 0)
    status ('Family')
    return combined
```

In [68]:

```
combined = processing_family()
```

```
processing Family : ok
```

In [69]:

```
print(combined.shape)
```

```
(1309, 80)
```

In [70]:

```
combined.head()
```

Out[70]:

	Name	Sex	Age	SibSp	Parch	Fare	Title_Master	Title_Miss	Title_Mr
0	Braund, Mr. Owen Harris	1	22.0	1	0	7.2500	0	0	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	71.2833	0	0	
2	Heikkinen, Miss. Laina	0	26.0	0	0	7.9250	0	1	
3	Futrelle, Mrs. Jacques Heath (Lily May	0	35.0	1	0	53.1000	0	0	

	Peel)								
4	Allen, Mr. William Henry	1	35.0	0	0	8.0500		0	0

5 rows × 80 columns

In [71]:

```
combined.drop('Name', axis = 1, inplace= True)
```

In [72]:

```
combined.drop('Ticket', axis=1, inplace=True)
```

```
---------------------------------------------
-----------------------------
KeyError                                  Trac
eback (most recent call last)
<ipython-input-72-720bdfc8b339> in <module>
----> 1 combined.drop('Ticket', axis=1, inplac
e=True)

~\Anaconda3\lib\site-packages\pandas\core\fram
e.py in drop(self, labels, axis, index, column
s, level, inplace, errors)
   3938
     index=index, columns=columns,
   3939
     level=level, inplace=inplace,
-> 3940
     errors=errors)
   3941
   3942     @rewrite_axis_style_signature('map
per', [('copy', True),

~\Anaconda3\lib\site-packages\pandas\core\gene
ric.py in drop(self, labels, axis, index, colu
```

```
mns, level, inplace, errors)
   3778         for axis, labels in axes.items
():
   3779             if labels is not None:
-> 3780                 obj = obj._drop_axis(l
abels, axis, level=level, errors=errors)
   3781
   3782         if inplace:

~\Anaconda3\lib\site-packages\pandas\core\gene
ric.py in _drop_axis(self, labels, axis, level
, errors)
   3810                 new_axis = axis.drop(l
abels, level=level, errors=errors)
   3811             else:
-> 3812                 new_axis = axis.drop(l
abels, errors=errors)
   3813             result = self.reindex(**{a
xis_name: new_axis})
   3814

~\Anaconda3\lib\site-packages\pandas\core\inde
xes\base.py in drop(self, labels, errors)
   4963             if errors != 'ignore':
   4964                 raise KeyError(
-> 4965                     '{} not found in a
xis'.format(labels[mask]))
   4966             indexer = indexer[~mask]
   4967         return self.delete(indexer)

KeyError: "['Ticket'] not found in axis"
```

In [73]:

```
combined
```

Out[73]:

	Sex	Age	SibSp	Parch	Fare	Title_Master	Title_Miss	Title_Mr	Title_
0	1	22.0	1	0	7.2500	0	0	1	
1	0	38.0	1	0	71.2833	0	0	0	
2	0	26.0	0	0	7.9250	0	1	0	
3	0	35.0	1	0	53.1000	0	0	0	
4	1	35.0	0	0	8.0500	0	0	1	
5	1	26.0	0	0	8.4583	0	0	1	
6	1	54.0	0	0	51.8625	0	0	1	
7	1	2.0	3	1	21.0750	1	0	0	
8	0	27.0	0	2	11.1333	0	0	0	
9	0	14.0	1	0	30.0708	0	0	0	
10	0	4.0	1	1	16.7000	0	1	0	
11	0	58.0	0	0	26.5500	0	1	0	
12	1	20.0	0	0	8.0500	0	0	1	
13	1	39.0	1	5	31.2750	0	0	1	
14	0	14.0	0	0	7.8542	0	1	0	
15	0	55.0	0	0	16.0000	0	0	0	
16	1	2.0	4	1	29.1250	1	0	0	
17	1	31.0	0	0	13.0000	0	0	1	
18	0	31.0	1	0	18.0000	0	0	0	
19	0	31.0	0	0	7.2250	0	0	0	
20	1	35.0	0	0	26.0000	0	0	1	
21	1	34.0	0	0	13.0000	0	0	1	
22	0	15.0	0	0	8.0292	0	1	0	
23	1	28.0	0	0	35.5000	0	0	1	
24	0	8.0	3	1	21.0750	0	1	0	

25	0	38.0	1	5	31.3875	0	0	0
26	1	26.0	0	0	7.2250	0	0	1
27	1	19.0	3	2	263.0000	0	0	1
28	0	18.0	0	0	7.8792	0	1	0
29	1	26.0	0	0	7.8958	0	0	1
...	...	...	...	...	...	...	...	...
1279	1	21.0	0	0	7.7500	0	0	1
1280	1	6.0	3	1	21.0750	1	0	0
1281	1	23.0	0	0	93.5000	0	0	1
1282	0	51.0	0	1	39.4000	0	0	0
1283	1	13.0	0	2	20.2500	1	0	0
1284	1	47.0	0	0	10.5000	0	0	1
1285	1	29.0	3	1	22.0250	0	0	1
1286	0	18.0	1	0	60.0000	0	0	0
1287	1	24.0	0	0	7.2500	0	0	1
1288	0	48.0	1	1	79.2000	0	0	0
1289	1	22.0	0	0	7.7750	0	0	1
1290	1	31.0	0	0	7.7333	0	0	1
1291	0	30.0	0	0	164.8667	0	1	0
1292	1	38.0	1	0	21.0000	0	0	1
1293	0	22.0	0	1	59.4000	0	1	0
1294	1	17.0	0	0	47.1000	0	0	1
1295	1	43.0	1	0	27.7208	0	0	1
1296	1	20.0	0	0	13.8625	0	0	1
1297	1	23.0	1	0	10.5000	0	0	1
1298	1	50.0	1	1	211.5000	0	0	1
1299	0	18.0	0	0	7.7208	0	1	0
1300	0	3.0	1	1	13.7750	0	1	0

1301	0	18.0	0	0	7.7500	0	1	0
1302	0	37.0	1	0	90.0000	0	0	0
1303	0	28.0	0	0	7.7750	0	1	0
1304	1	26.0	0	0	8.0500	0	0	1
1305	0	39.0	0	0	108.9000	0	0	0
1306	1	38.5	0	0	7.2500	0	0	1
1307	1	26.0	0	0	8.0500	0	0	1
1308	1	4.0	1	1	22.3583	1	0	0

1309 rows × 79 columns

In [74]:

```
from sklearn.pipeline import make_pipeline
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble.gradient_boosting import GradientBoosti
ngClassifier
from sklearn.feature_selection import SelectKBest
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_score
from sklearn.feature_selection import SelectFromModel
from sklearn.linear_model import LogisticRegression, Logistic
RegressionCV
```

In [75]:

```
def compute_score(clf, X, y, scoring='accuracy'):
    xval = cross_val_score(clf, X, y, cv=5, scoring=scoring)
    return (np.mean(xval))
```

In [78]:

```
def recover_train_test_target():
    global combined
```

```
    a = pa.read_csv(r'C:\Users\SACHIN K M\Desktop\python\data
\datasets\train.csv', usecols=['Survived'])['Survived'].value
s
    train = combined.iloc[:891]
    test = combined.iloc[891:]

    return (train, test, a)
```

In [79]:

```
train, test, a = recover_train_test_target()
```

In [80]:

```
clf = RandomForestClassifier(n_estimators = 50, max_features=
'sqrt')
clf = clf.fit(train, a)
```

In [81]:

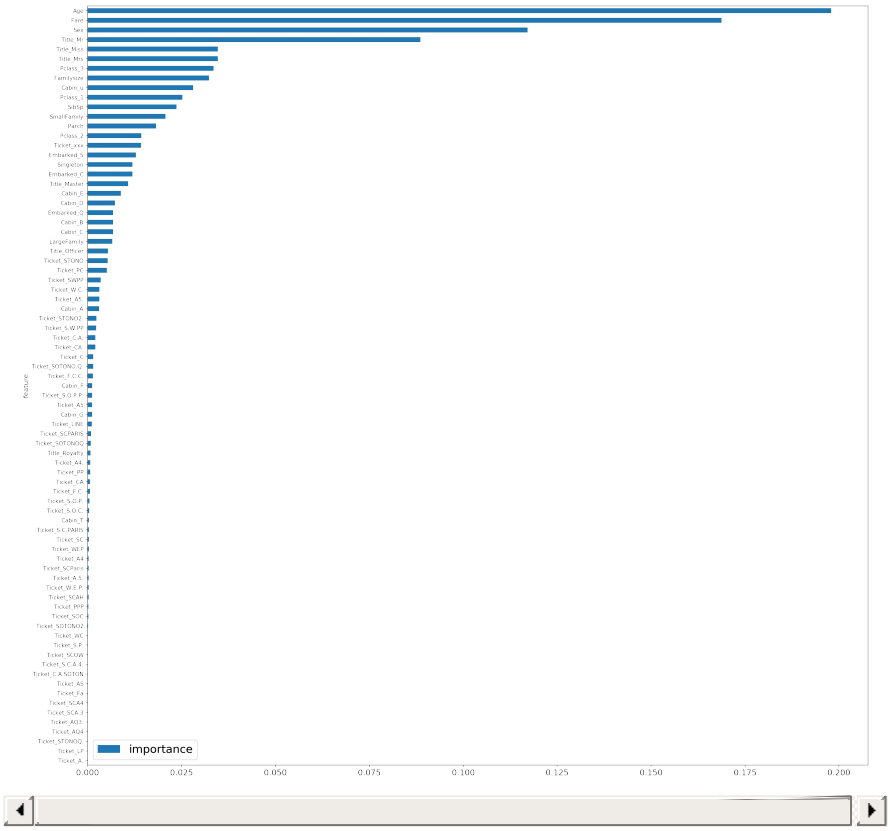
```
features = pa.DataFrame()
features['feature'] = train.columns
features['importance'] = clf.feature_importances_
features.sort_values(by=['importance'], ascending=True, inpla
ce=True)
features.set_index('feature', inplace = True)
```

In [82]:

```
features.plot(kind='barh', figsize=(25,25))
```

Out[82]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x17
b943d97b8>
```



In [83]:

```
model = SelectFromModel(clf, prefit = True)
train_reduced = model.transform(train)
print (train_reduced.shape)
```

```
(891, 16)
```

In [84]:

```
test_reduced = model.transform(test)
print (test_reduced.shape)
```

```
(418, 16)
```

In [85]:

```
logreg = LogisticRegression()
logreg_cv = LogisticRegressionCV()
rf = RandomForestClassifier()
```

```
gboost = GradientBoostingClassifier()

models = [logreg, logreg_cv, rf, gboost]
```

In [86]:

```
for model in models:
    print('Cross-validation of : {0}'.format(model.__class__)
)
    score = compute_score(clf=model, X= train_reduced, y = a,
 scoring = 'accuracy')
    print ('CV_score= {0}'.format(score))
    print ('****')
```

```
Cross-validation of : <class 'sklearn.linear_m
odel.logistic.LogisticRegression'>
CV_score= 0.8159478357762484
****
Cross-validation of : <class 'sklearn.linear_m
odel.logistic.LogisticRegressionCV'>
```

```
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
```

```
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
```

```
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
```

```
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
```

```
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
```

```
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
```

```
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
```

```
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
```

```
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
```

```
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
```

```
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
```

```
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
```

```
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
C:\Users\SACHIN K M\Anaconda3\lib\site-package
s\sklearn\linear_model\logistic.py:758: Conver
genceWarning: lbfgs failed to converge. Increa
se the number of iterations.
  "of iterations.", ConvergenceWarning)
```

```
CV_score= 0.8148368653377009
****
Cross-validation of : <class 'sklearn.ensemble
.forest.RandomForestClassifier'>
CV_score= 0.790148723999366
****
Cross-validation of : <class 'sklearn.ensemble
.gradient_boosting.GradientBoostingClassifier'
>
CV_score= 0.8293995255670021
```

****

In [87]:

```
run_gs = False

if run_gs:
    parameter_grid = {
        'max_depth : [4, 6, 8]'
        'n_estimator' : [50, 10],
        'max_features' : ['sqrt', 'auto', 'log2'],
        'min_samples_split': [2, 3, 10],
        'min_sample_leaf' : [1, 3, 10],
        'bootstrap' : [True, False],
    }


    forest = RandomForestClassifier()
    cross_validation = StratifiedKFold(n_splits = 5)


    grid_serarch= GridSearchCV(forest,
                               scoring='accuracy',
                               param_grid= parameter_grid,
                               cv = cross_validation,
                               verbose=1
                               )
    grid_search.fit(train, a)
    model = grid_search
    parameters = grid_search.best_params_

    print ('Best score : {}'.format(grid_search.best_score_))
    print ('Best parameters : {}'.format(grid_search.best_par
ams_))

else:
    parameters = {'bootstrap': False, 'min_samples_leaf': 3,
```

```
'n_estimators':50,
                 'min_samples_split':10, 'max_features': 'sqr
t', 'max_depth':6}


    model = RandomForestClassifier(**parameters)
    model.fit(train, a)
```

In [89]:

```
output = model.predict(test).astype(int)
df_output = pa.DataFrame()
aux = pa.read_csv(r'C:\Users\SACHIN K M\Desktop\python\data\d
atasets\test.csv')
df_output['PassengerId'] = aux['PassengerId']
df_output['Survived'] = output
df_output[['PassengerId', 'Survived']].to_csv(r'C:\Users\SACH
IN K M\Desktop\python\data\datasets\result_titanic.csv', inde
x=False)
```

In [ ]:

```

```