



Session 23

Assignment 1 Question

Session 23: Assignment 1

Table of Contents

1. Introduction

2. Problem Statement

3. Output

1. Introduction

This assignment will help you to consolidate the concepts learnt in the session.

2. Problem Statement

Predicting Survival in the Titanic Data Set

We will be using a decision tree to make predictions about the Titanic data set from Kaggle. This data set provides information on the Titanic passengers and can be used to predict whether a passenger survived or not.

Loading Data and modules

```
import numpy as np
```

```
import pandas as pd
```

```
import seaborn as sb
```

```
import matplotlib.pyplot as plt
```

```
import sklearn
```

```
from pandas import Series, DataFrame
```

```
from pylab import rcParams
```

```
from sklearn import preprocessing
```

```
from sklearn.linear_model import LogisticRegression
```

```
from sklearn.cross_validation import train_test_split
```

```
from sklearn import metrics
```

```
from sklearn.metrics import classification_report
```

Url= <https://raw.githubusercontent.com/BigDataGal/Python-for-Data-Science/master/titanic-train.csv>

```
titanic = pd.read_csv(url)
```

```
titanic.columns =  
['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked']
```

You use only Pclass, Sex, Age, SibSp (Siblings aboard), Parch (Parents/children aboard), and Fare to predict whether a passenger survived.

NOTE: The solution shared through Github should contain the source code used and the screenshot of the output.

3. Output

N/A