# Predicting Survival on the Titanic Using a Decision Tree

**Swetha Kolalapudi**
CO-FOUNDER, LOONYCORN

www.loonycorn.com

# Overview

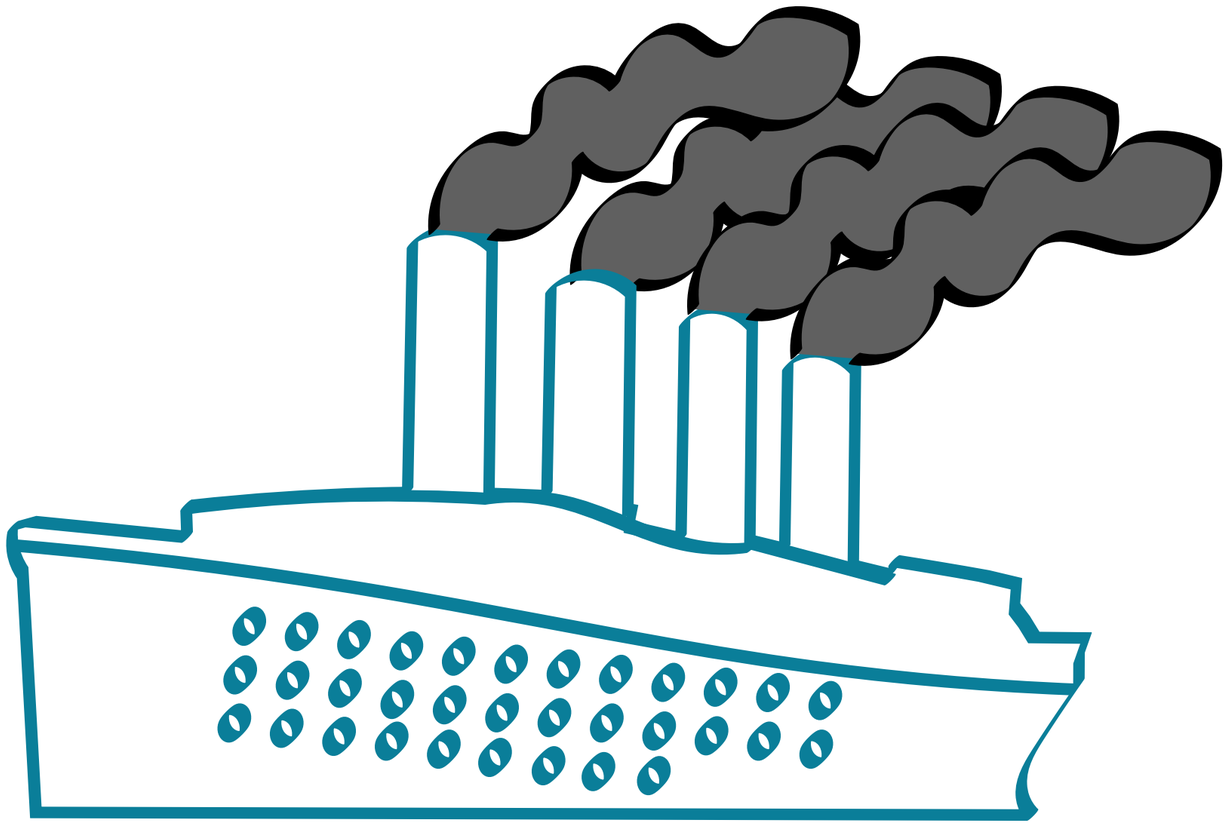
Recognize that predicting survival on the Titanic is a classification problem

Select relevant features

Build a decision tree

Understand the different parameters which can be used to control the tree

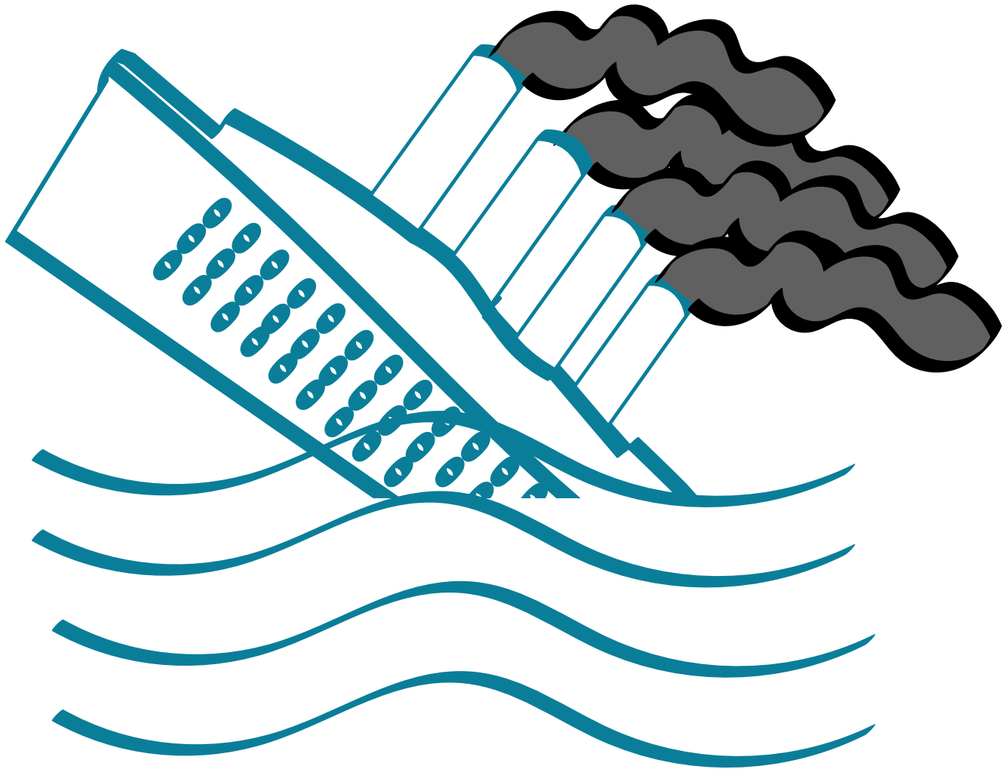
- The world’s biggest and fastest ocean liner
- Hit an iceberg on it’s maiden voyage
- Only 700 of the ~2500 passengers and crew survived



# Survival on the Titanic

**Given a passenger/crew member**

**or**



Classification Problem

**Passenger**

**Problem Instance**

Classification Problem

**Survived, Did not survive**

**Labels**

Classification Problem

**Solve this problem by building a decision tree**

# Classification Problem

**Choose attributes/ features which can divide the passengers into homogenous subsets**

**Age, Passenger class, Gender etc**

# Demo

**Download a dataset with details of the Titanic's passengers and crew**

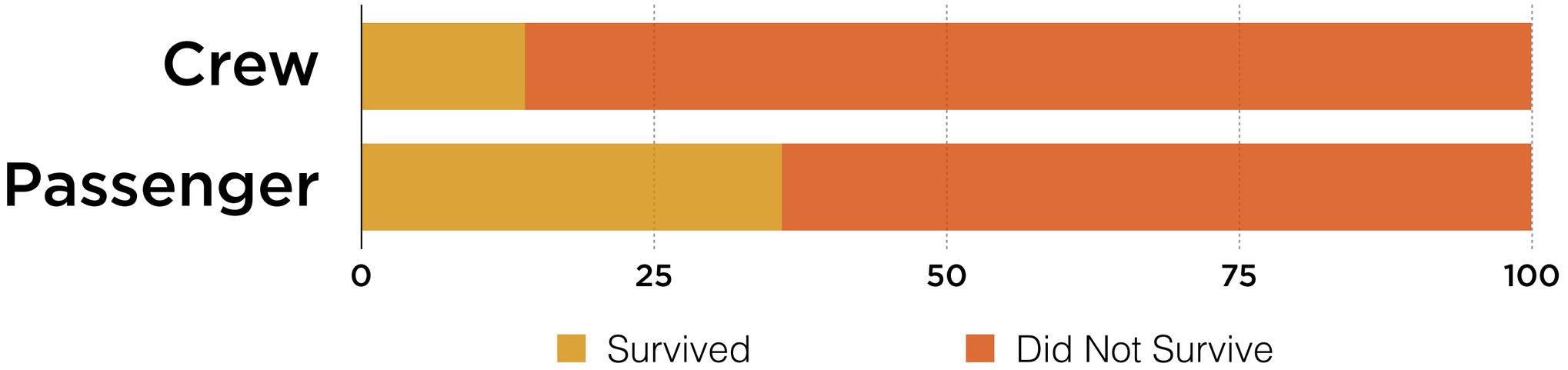
**Extract relevant features from the dataset**

# Demo

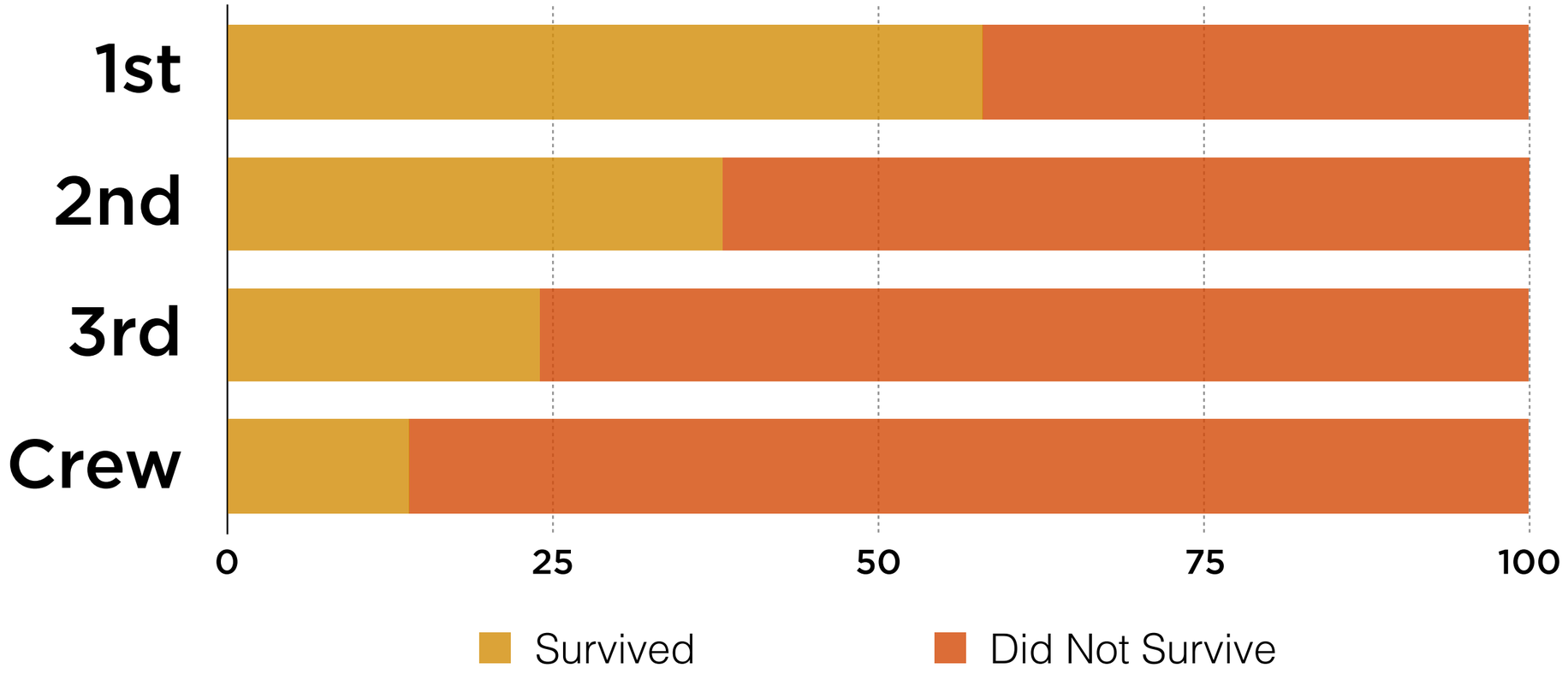
**Summarize survival rate by different features**

**Check which features might be more important than others**

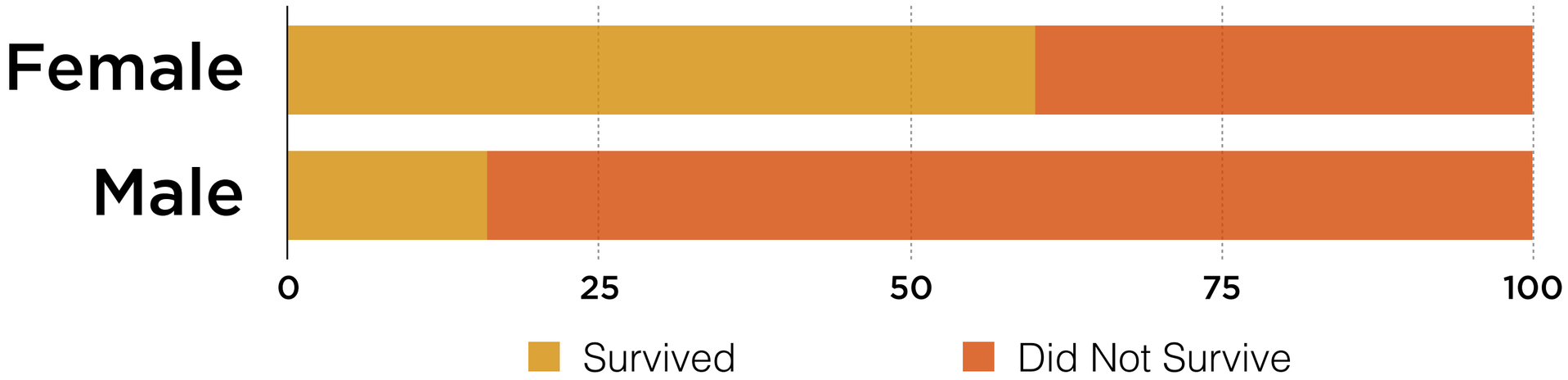
# Survival on the Titanic



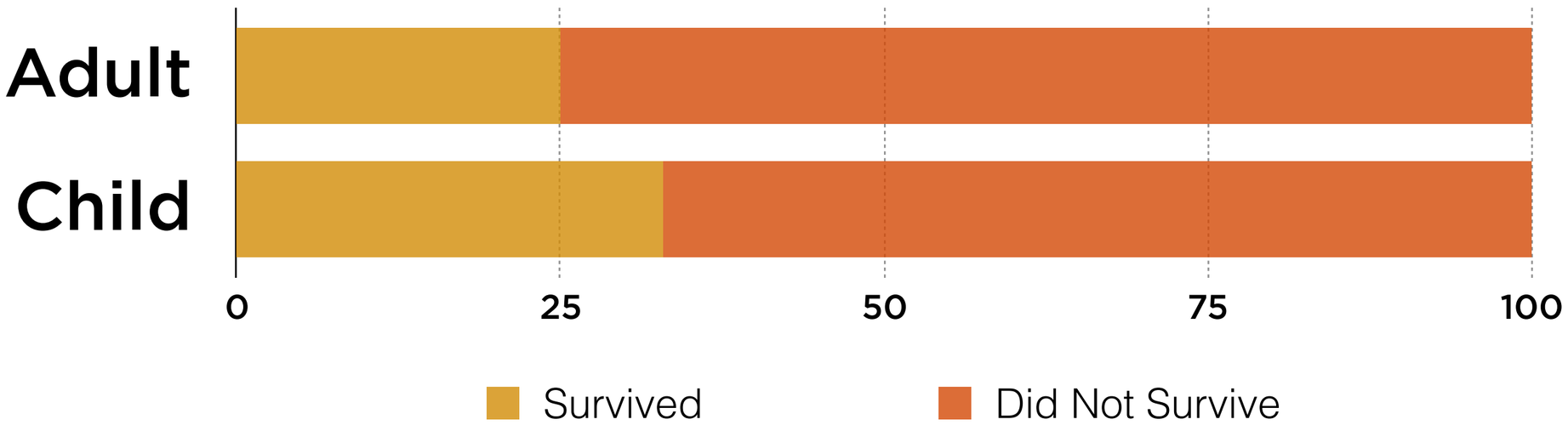
# Survival on the Titanic

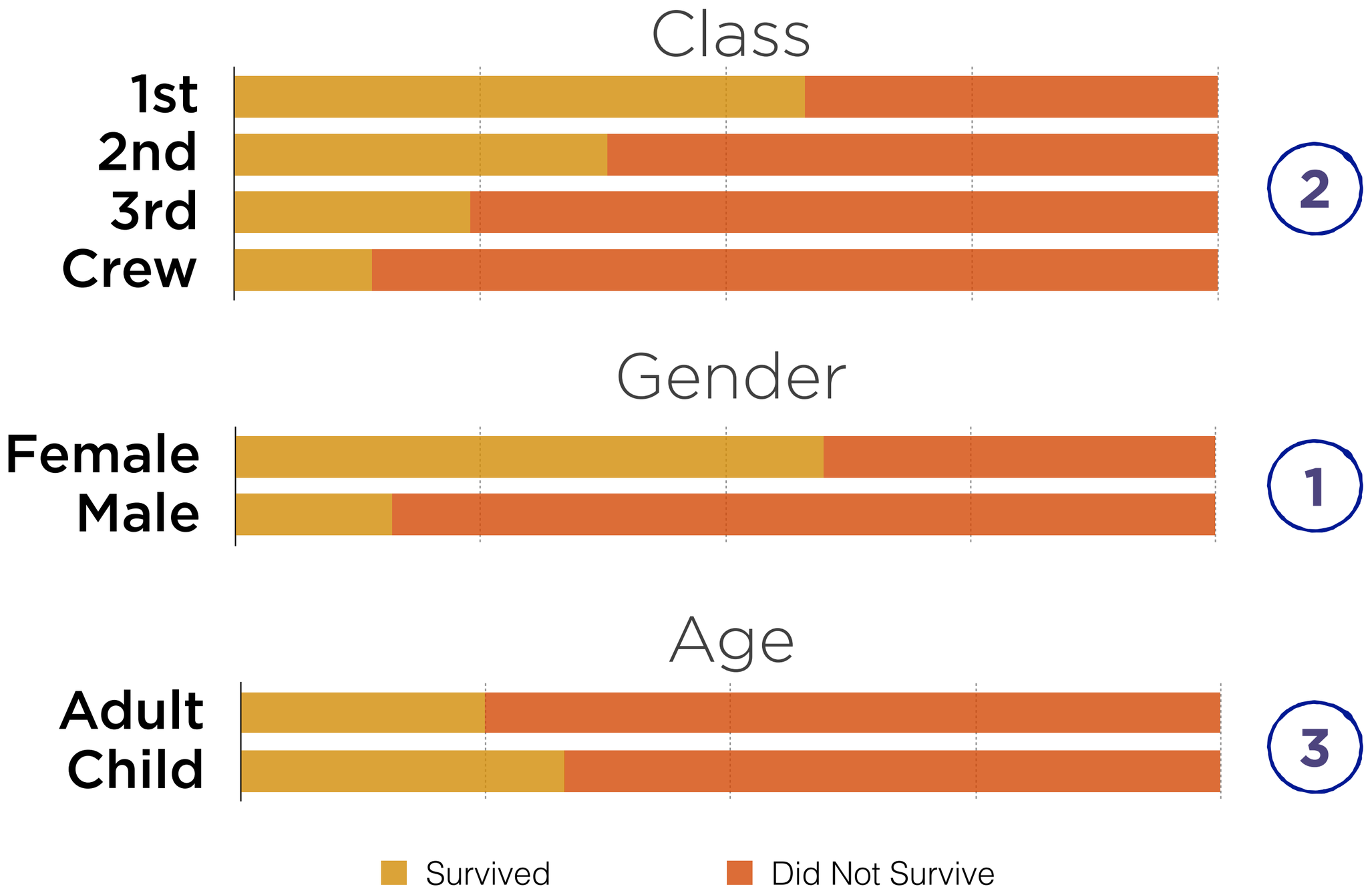


# Survival on the Titanic



# Survival on the Titanic




Class
1st
2nd
3rd
Crew
2
Gender
Female
Male
1
Age
Adult
Child
3
Survived
Did Not Survive

# Demo

**Build a decision tree**

# Demo

**Visualize the decision tree using Graphviz**

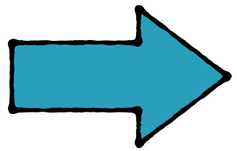
# Demo

**Understand the different parameters used to control the decision tree**

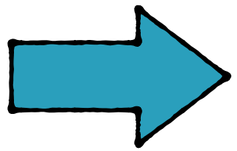
# Building a Decision Tree

**Training Data**

| Features | Label |
| --- | --- |
| | |
| | |
| | |
| | |



Machine Learning Algorithm

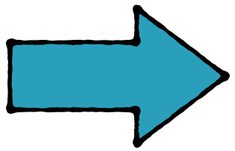


**Decision Tree**

# 1. Split the Training Data

**Training Data**

| Features | Label |
|---|---|
| | |
| | |
| | |
| | |



**Train**

| Features | Label |
|---|---|
| | |
| | |
| | |

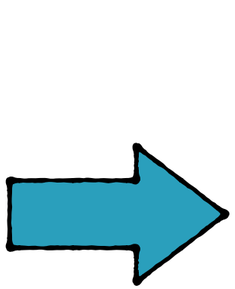
**Test**

| Features | Label |
|---|---|
| | |
| | |

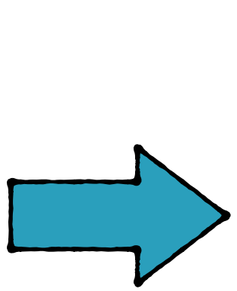
# 2. Build the Decision Tree

**Train**

| Features | Label |
| --- | --- |
| | |
| | |
| | |

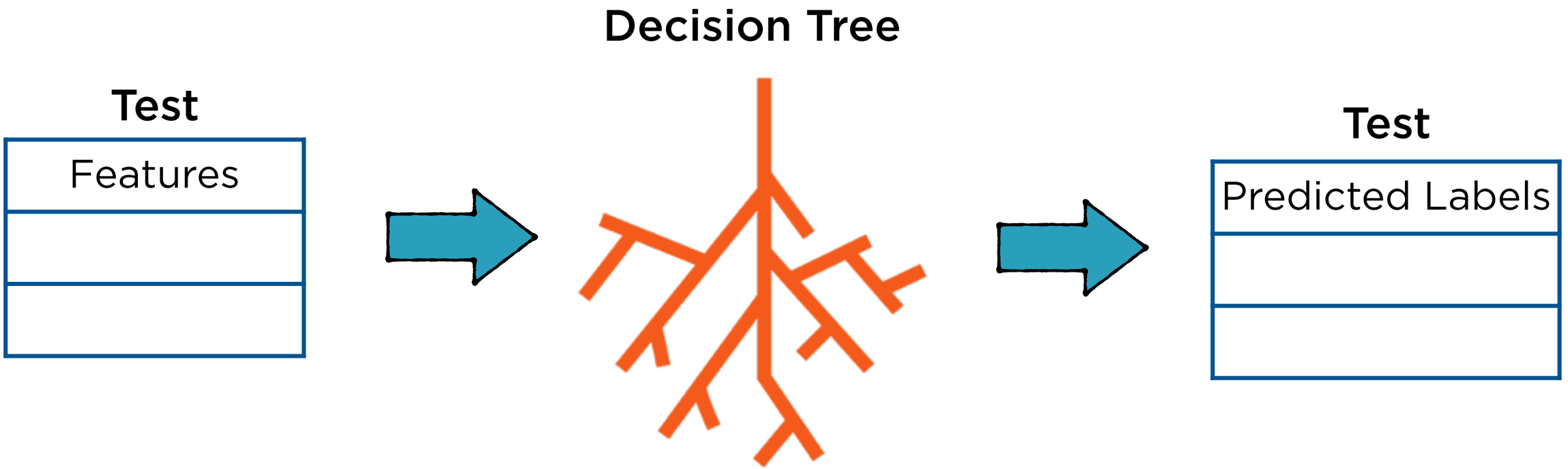


Machine Learning Algorithm



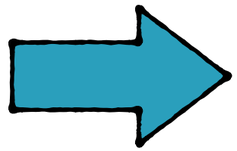
**Decision Tree**

# 3. Test the Decision Tree



# Map Categories to Numbers

| Gender |
| --- |
| Female |
| Male |



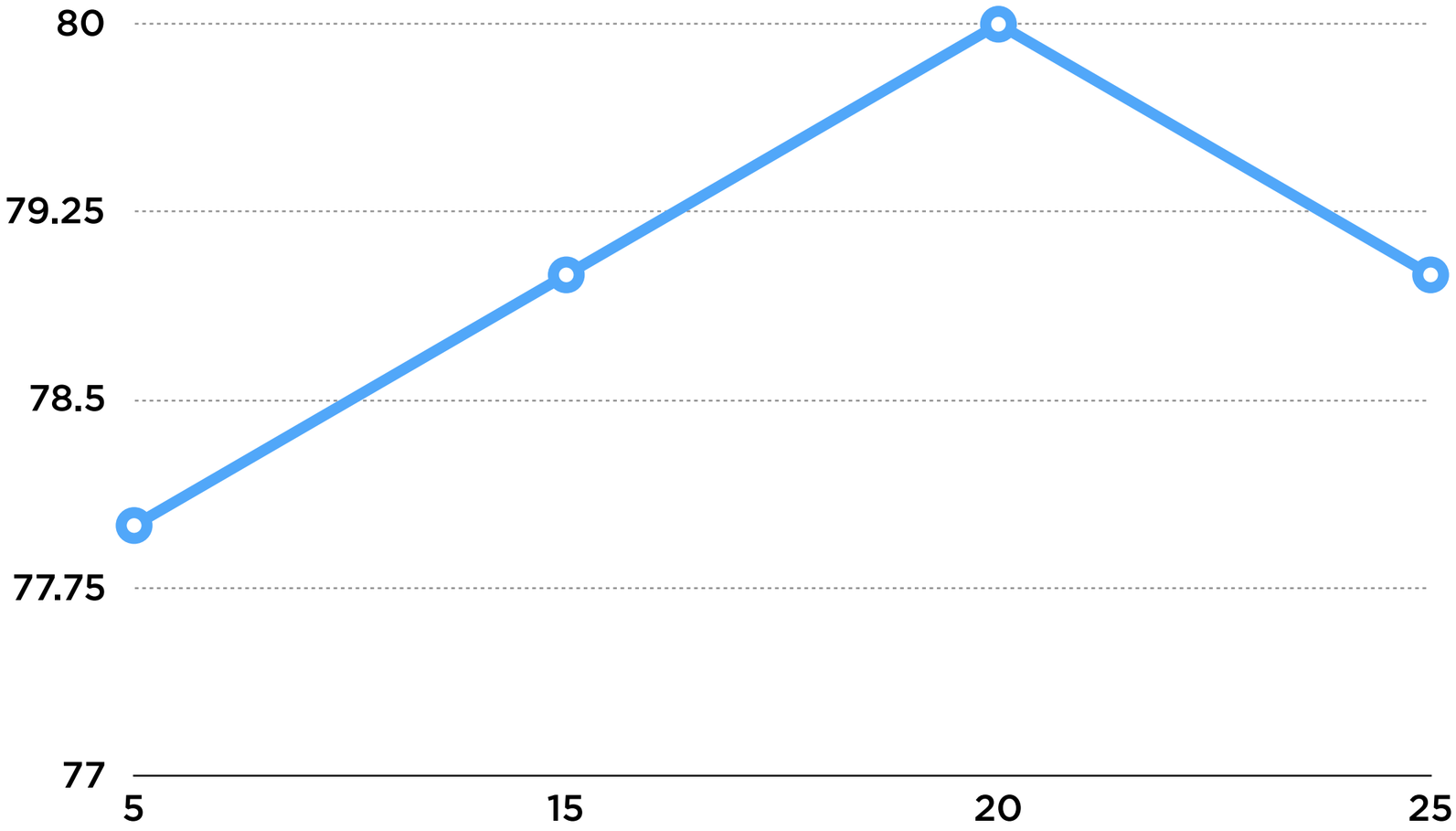
| Gender |
| --- |
| 0 |
| 1 |

# Demo

**Measure the accuracy of the decision tree**

**Vary different parameters of the decision tree to check impact on accuracy**

# Prediction Accuracy



# Summary

Recognize that predicting survival on the Titanic is a classification problem

Select relevant features

Build a decision tree

Understand the different parameters which can be used to control the tree

Recognize the problem of overfitting