Predicting Survival on the Titanic Using Random Forests

Swetha Kolalapudi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Understanding the Random Forests technique

Use the Random Forests technique to solve the Titanic problem

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

Random Forests

- An ensemble learning technique
- Builds an ensemble of decision trees

Random Forests

Models built using different

Training Sets

Each tree built from a different subset of the training set

Features

Each tree built using a different subset of features

Random Forests

Bagging

Each tree built from a different subset of the training set

Random Subspace

Each tree built using a different subset of features

Random Forests

Bagging

Each tree built from a different subset of the training set

Random Features Subspace

Each tree built using a different subset of features

Bagging

Training Data

Jane	Lawrence
Maria	Sam
Eliza	Elliot
Ellen	Tom
Teri	Jack

Bagging

Training Data

Jane	Lawrence
Maria	Sam
Eliza	Elliot
Ellen	Tom
Teri	Jack

Random Forests

Tree 1

Bagging

Training Data

Jane	Lawrence
Maria	Sam
Eliza	Elliot
Ellen	Tom
Teri	Jack

Random Forests

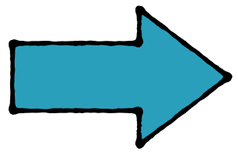
Tree 1

Tree 2

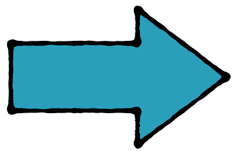
Bagging

Training Data

Jane	Lawrence
Maria	Sam
Eliza	Elliot
Ellen	Tom
Teri	Jack



Random Forests



Tree 1

Tree 2

Tree 3

Bagging

Tree 1

Tree 2

Tree 3

Each training set is a randomly generated subset of the original training set

Bagging

Bootstrap Aggregating

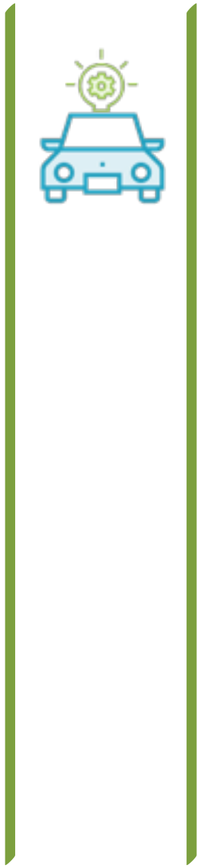
Bagging

Bootstrap Sampling

A statistical technique to select samples from a dataset

Bootstrap Sampling

A person is studying how fast cars are traveling at an intersection



Bootstrap Sampling

The person randomly selects some cars and measure their speed

Bootstrap Sampling

Every car has an equal probability of being picked

Cars which passed by might pass by again

Bootstrap Sampling

- Every data point has an equal probability of being picked

- A data point can be picked for a training set more than once

Random Forests

Bagging

Each tree built from a different subset of the training set

Random Features Subspace

Each tree built using a different subset of features

Random Forests

Training Bagging Sets

Each tree built from a different subset of the training set

Random Subspace

Each tree built using a different subset of features

Random Subspace

Training Data

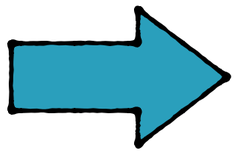
Vowel ending	Vowel beginning	Begin with K	End with N

Random Subspace

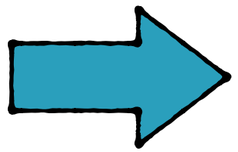
Tree 1

Training Data

Vowel ending		Begin with K	



Random Forests



Random Subspace

Training Data

	Vowel beginning		End with N

Random Forests

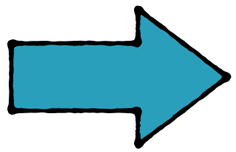
Tree 1

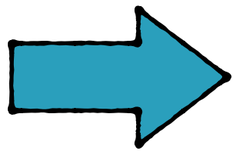
Tree 2

Random Subspace

Training Data

Vowel ending	Vowel beginning		





Tree 1

Tree 2

Tree 3

Random Forests

Bagging

Each tree built from a different subset of the training set

Random Subspace

Each tree built using a different subset of features

Demo

Use Random Forests to solve the Titanic problem

Summary

Understanding the Random Forests technique

Use the Random Forests technique to solve the Titanic problem

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