НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Кафедра автоматизації проектування енергетичних процесів і систем

Лабораторна робота № 7

З дисципліни: «Big Data та аналіз даних»

**Виконав:**

студент 5 курсу

групи ТР-02мп, ТЕФ

Круглий Д.В.

**Перевірив:**

Матичин І.І.

Київ – 2021

**Постановка задачі**

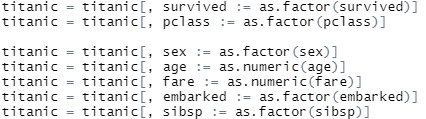
1. Імпортувати дані з файлу titanic\_data.csv
2. Перетворити змінні pclass і survived на тип фактор та перетворити інші змінні з символьного типу на числовий в разі потреби.
3. Усунути неіснуючі значення NA.
4. Застосувати випадковий ліс для прогнозування значення змінної survived.
5. Оцінити точність та порівняти її з точністю, отриманою за допомогою дерева рішень.

**Хід роботи**

1. Зчитуємо дані з файлу *tatanic\_data.csv* до змінної ***tatanic***.



1. Перетворюємо змінні pclass і survived на тип фактор, а також деякі інші змінні на числовий та тип фактор.

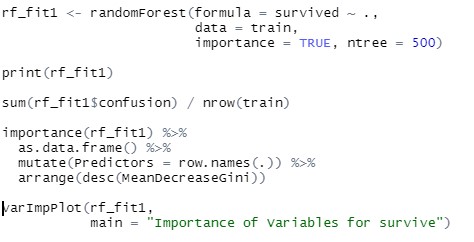


1. Усуваємо неіснуючі значення NA.

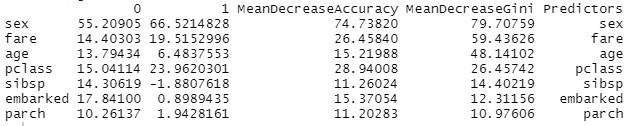


1. Розбиваємо дані на навчальну та тестову вибірки. І застосовуємо випадковий ліс для прогнозування значення змінної survived.

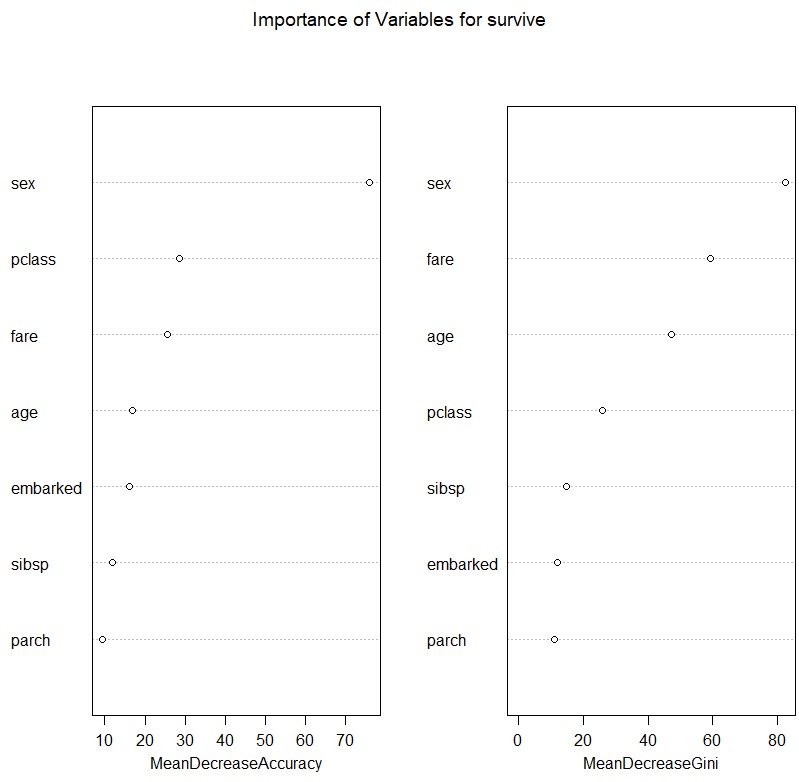




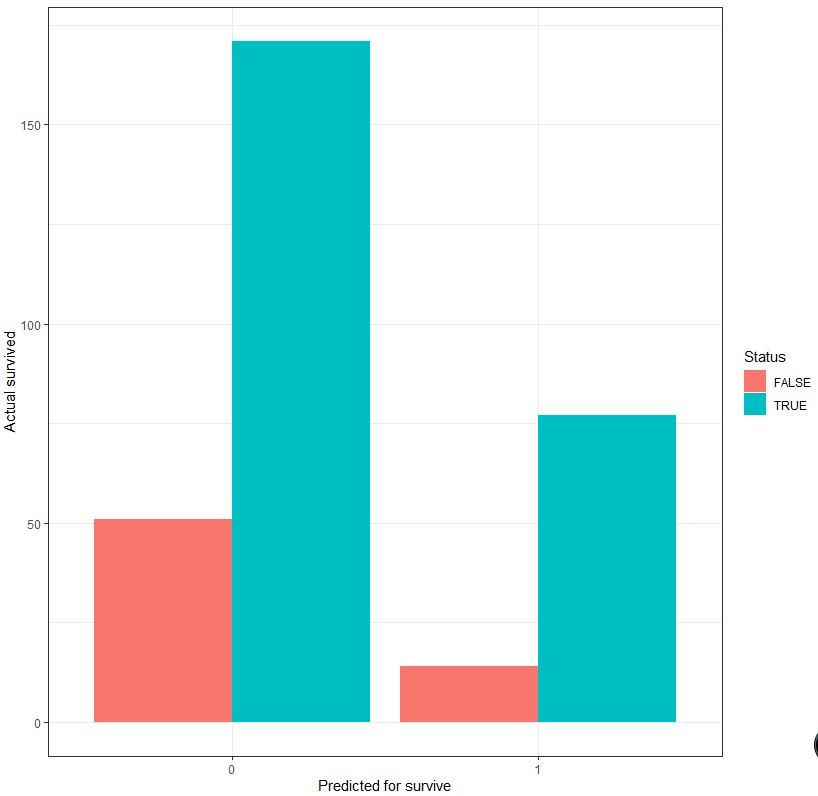
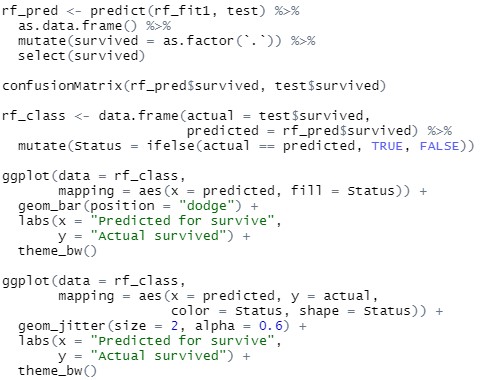
Маємо результати, вплив змінних для на значення змінною ***survive***.

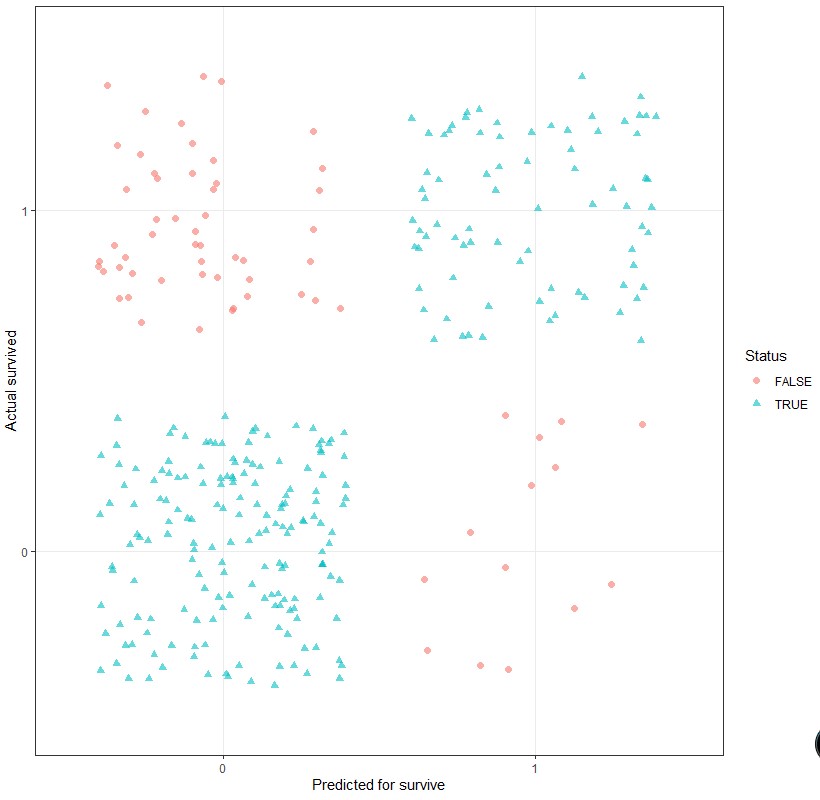


Показуємо це графіками.



1. Оцінюємо точність і порівнюємо її з точністю отриманою за допомогою дерева рішень. Візуалізовуємо порівняння за допомогою графіків.





**Висновок:** виконавши дану лабораторну роботу, я навчився використовувати випадковий ліс для прогнозування значення обраних параметрів. Окрім того, оцінювати та порівнювати точність різних моделей і будувати відповідні графіки для візуалізації порівняння.