Изготвили:

Сияна славова, 24963

Иван Капукаранов, 24958

Проект за курса „Откриване на знания в данни“

Задача: wine quality classification

# Декларация за липса на плагиатство

1. Тази курсова работа е моя работа, като всички изречения, илюстрации и програми от други хора са изрично цитирани.

2. Тази курсова работа или нейна версия не са представени в друг университет или друга учебна институция.

3. Разбирам, че ако се установи плагиатство в работата ми ще получа оценка “Слаб”.

Иван Капукаранов, ФН: 24958, 1 курс, ИИОЗ

Сияна Славова, ФН: 24963, 1 курс, ИИОЗ

Съдържание

[Декларация за липса на плагиатство 1](#_Toc454991495)

[Мотивация, Задача на курсовата работа 3](#_Toc454991496)

[Съществуващо решение 3](#_Toc454991497)

[Избрано решение 3](#_Toc454991498)

[Програмна реализация 3](#_Toc454991499)

[Предварителна обработка 3](#_Toc454991500)

[Данни 3](#_Toc454991501)

[Реализация 3](#_Toc454991502)

[Резултати от експерименти 3](#_Toc454991503)

[Заключение и бъдещо развитие 3](#_Toc454991504)

[Разпределение на задачите 3](#_Toc454991505)

[Литература и използвани източници 3](#_Toc454991506)

# Мотивация, Задача на курсовата работа

# Съществуващо решение

В намереното [съществуващо решение](http://www.r-bloggers.com/predicting-wine-quality-using-random-forests/) е използван алгоритъмът Random forest. То използва само White Wine dataset.

Като предварителна обработка качеството на виното се разпределя в три категории:

* Лошо (с качеството, по – малко от 6)
* Нормално (с качество, което е равно на 6)
* Добро (с качество, по – голямо от 6)

Тази предварителна обработка е направена поради липсата на най – много обучаващи данни с качество 6 и доста по – малко в другите категории.

В това решение ефективността на алгоритъма е измерена като accuracy, която е приблизително 71,5%.

# Избрано решение

# Програмна реализация

## Предварителна обработка

## Данни

Реално качеството има стойности между 3 и 9 и в двата сета.

## Red wine

## White wine

## Реализация

# Резултати от експерименти

# Заключение и бъдещо развитие

# Разпределение на задачите

# Литература и използвани източници

* Wine dataset:
  + <https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/>
* Съществуващо решение:
  + <http://www.r-bloggers.com/predicting-wine-quality-using-random-forests/>
* Random forest algorithm:
  + <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
* RSNNS – neural networks algorithm
  + <https://cran.r-project.org/web/packages/RSNNS/RSNNS.pdf>
  + <https://www.jstatsoft.org/article/view/v046i07>