МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Кафедра «Вычислительная техника»

Дисциплина «Интеллектуальные САПР»

**Лабораторная работа №5**

**«Алгоритмы кластеризации данных»**

Выполнил:

студент группы ИВТАПбд-41

Кондратьев П. С.

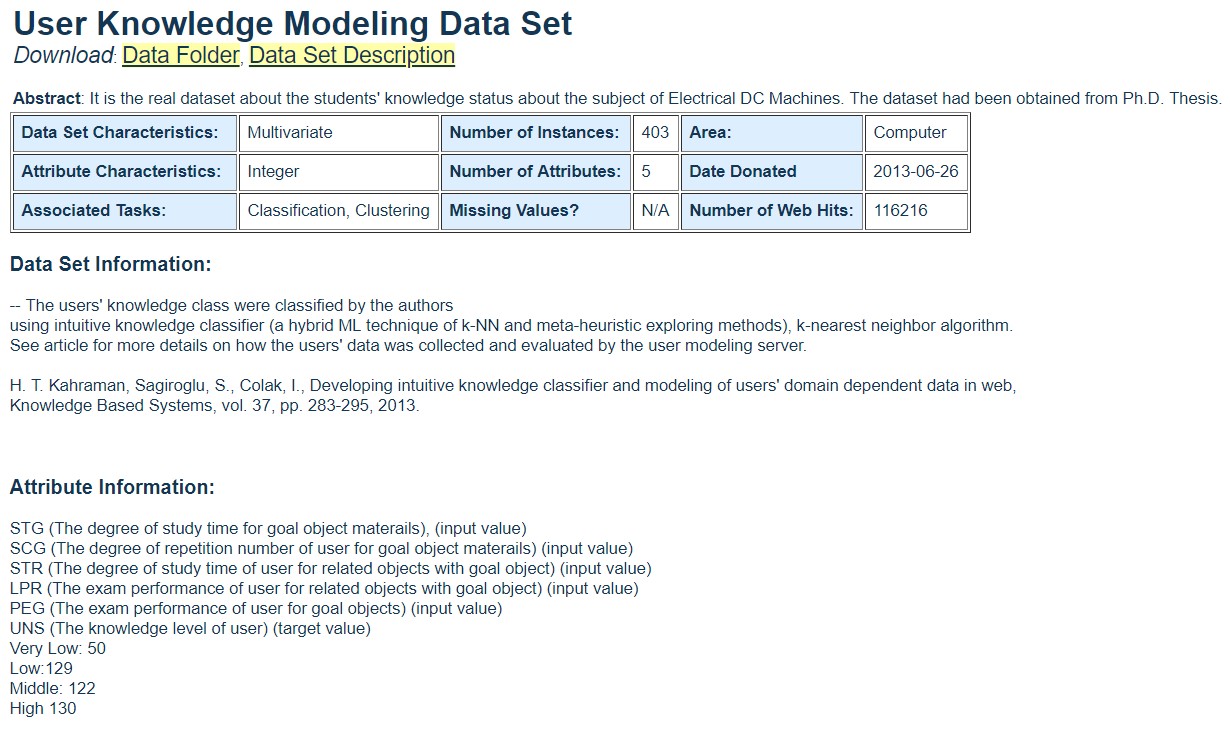
Проверил:

к.т.н. Святов К. В.

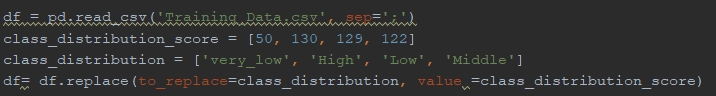
Ульяновск, 2019

Задание:

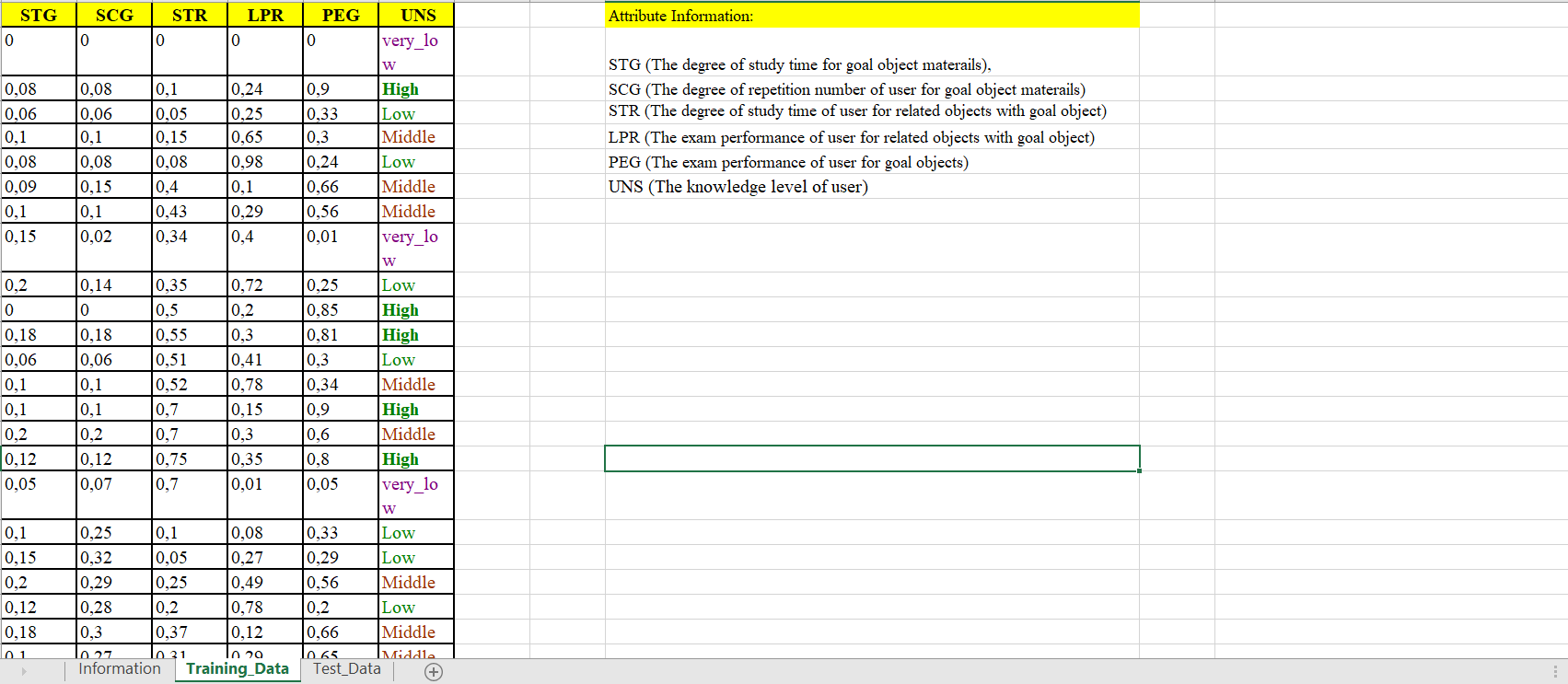
Data Set



1. Произвести масштабирование признаков (scaling).
2. С использованием библиотеки scikit-learn написать программу с использованием алгоритмов кластеризации данных, позволяющую разделить исходную выборку на классы, соответствующие предложенной вариантом задаче (<http://scikit-learn.org/stable/modules/clustering.html>).



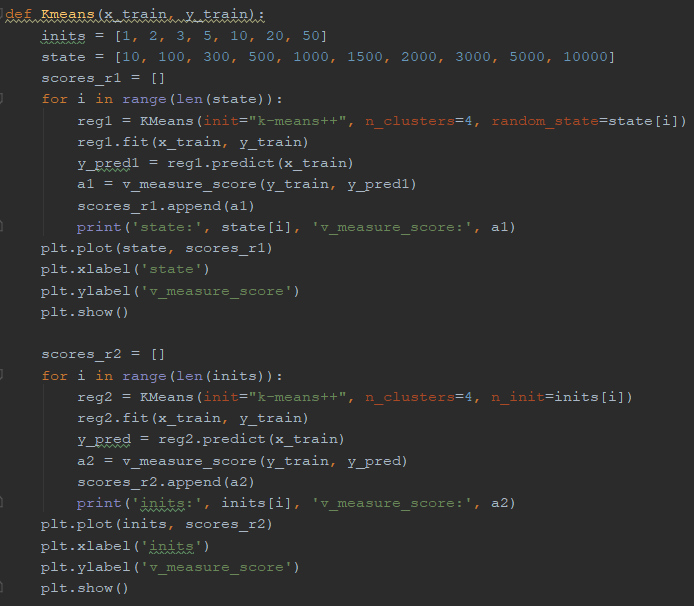
Набор данных представляет из себя файла: Data\_User\_Modeling\_Dataset Hamdi Tolga KAHRAMAN.xls, в этот файл разделен на исходную выборку на обучающую и тестовую.

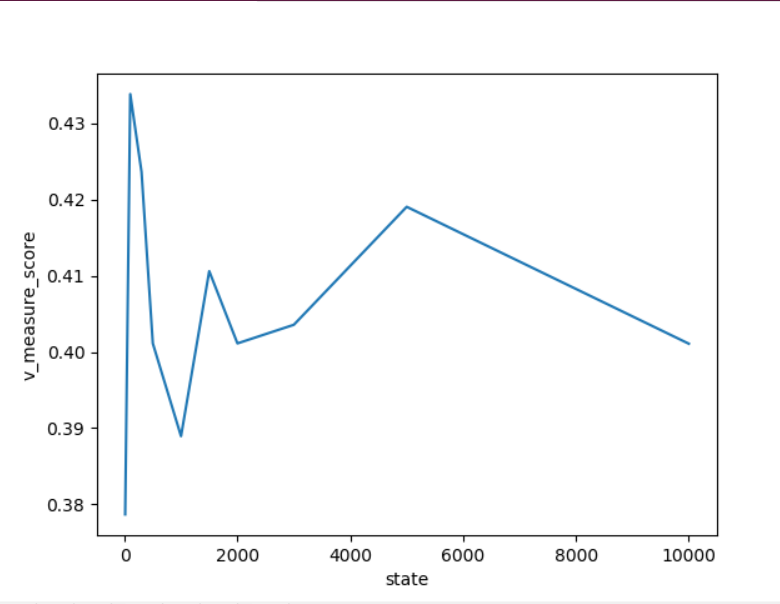


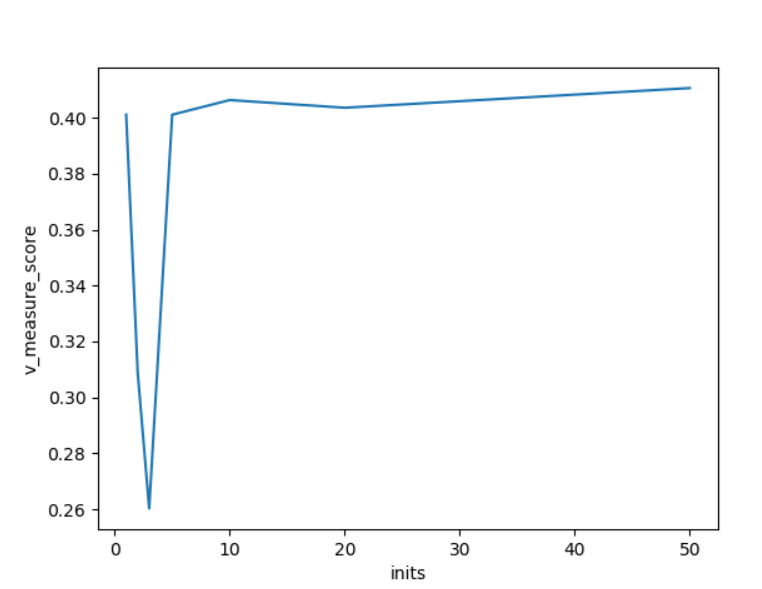
Так как в Training\_Data содержит 259 строк и Test\_Data 146, было решено объединить файлы т.к. объём данных мал для определить наилучший алгоритм кластеризации.

1. Провести эксперименты и определить наилучший алгоритм кластеризации, параметры алгоритма. Необходимо использовать не менее 3-х алгоритмов. Данные экспериментов необходимо представить в отчете (графики, ход проведения эксперимента, выводы).

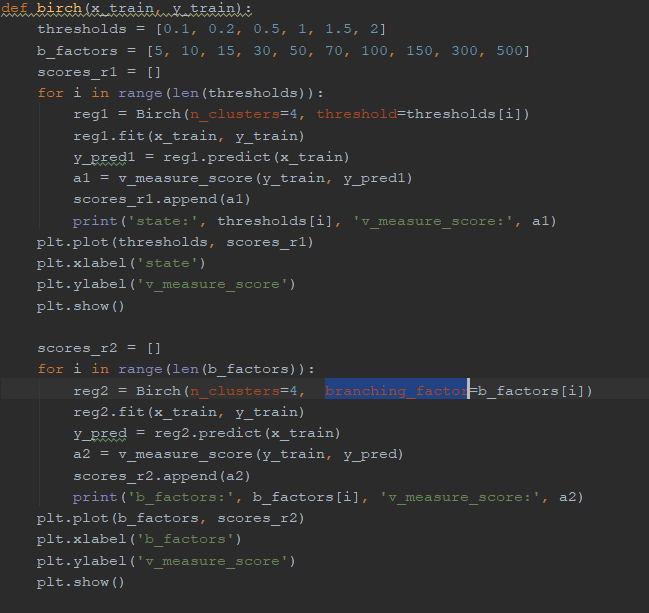
Реализован Kmeans с изменяющими параметрами random\_state и n\_init.

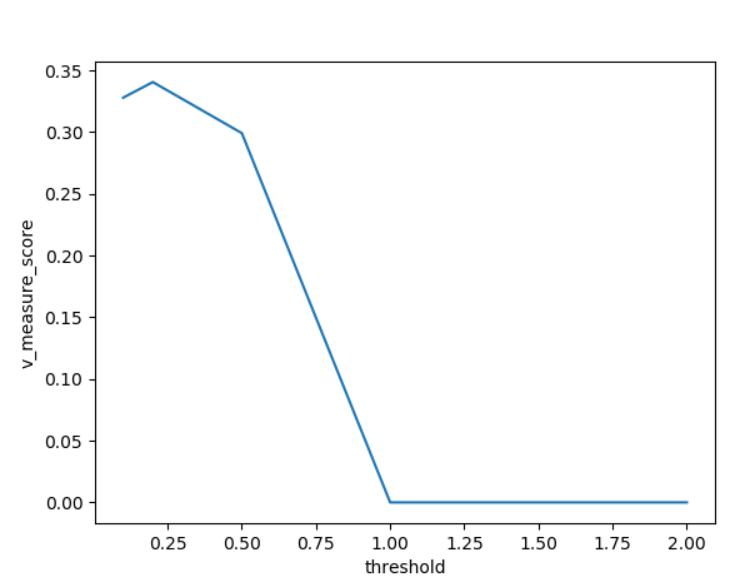


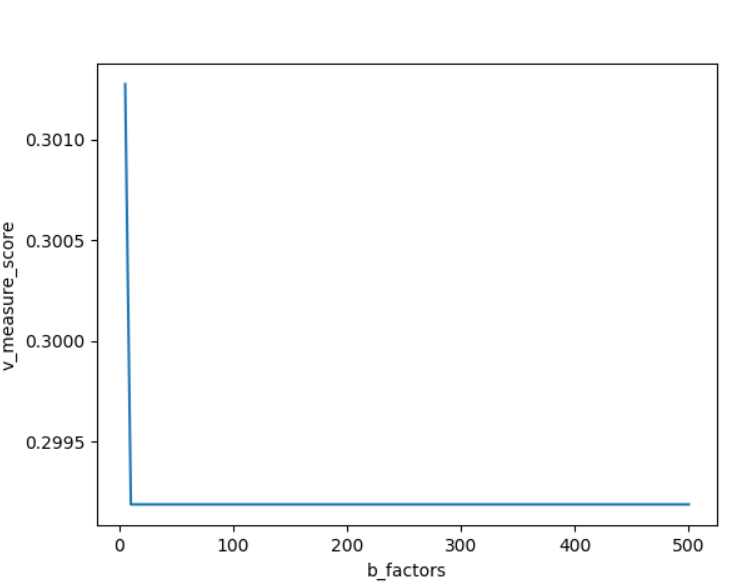




Реализован birch с изменяющими параметрами threshold и branching\_factor.

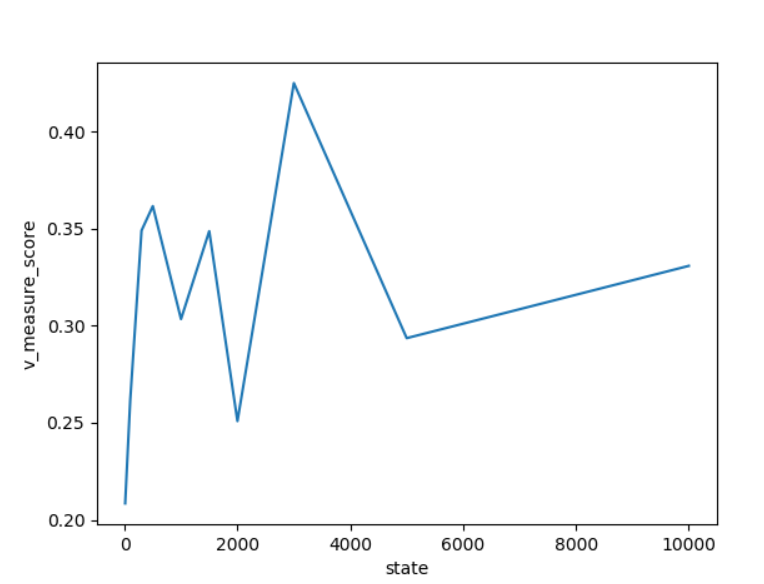


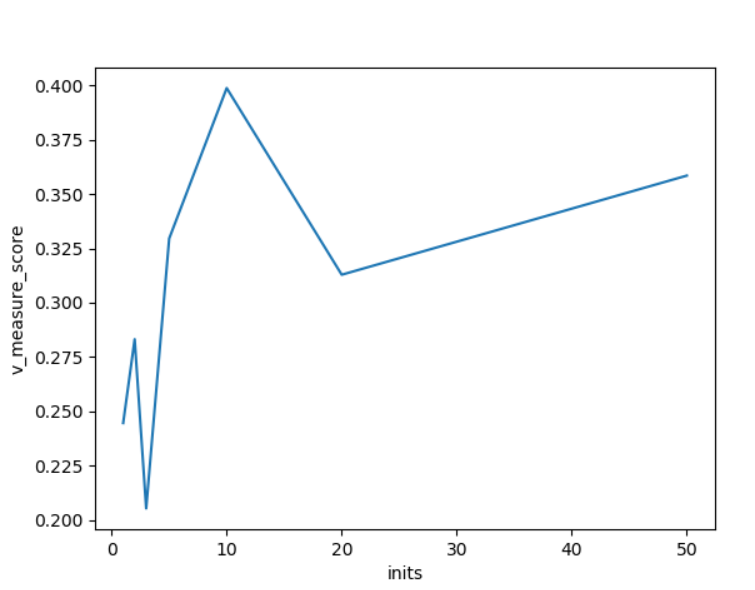




Реализован birch с изменяющими параметрами random\_state и n\_init.







Вывод: в ходе работы были изучены 3 модели алгоритмов кластеризации. Проверены экспериментально наилучшие алгоритмы кластеризации (параметры алгоритма).