Министерство цифрового развития, связи и массовых коммуникаций РФ

Федеральное государственное образовательное учреждение высшего профессионального образования

«Поволжский государственный университет телекоммуникаций и информатики»

Основы сбора и обработки больших данных

Лабораторная работа №   4

Предобработка данных в текстовом редакторе и редакторе таблиц

Выполнил студент группы МОИС-91: \_\_\_\_\_\_\_\_\_

Принял к.т.н., доцент кафедры ИСТ: Козлов В.В.

Самара 2021

**Цель работы:** получить базовые навыки для работы с большими данными на одном компьютере.

**Часть 1.**

Запустил IDLE, вставил в него приведённый в методичке код и запустил.

Изображение выглядит как текст

Автоматически созданное описание

**Изображение выглядит как текст

Автоматически созданное описание**

Изменил код так, чтобы выводились не только прогнозы, но и значения целевой и независимой переменных. Поэкспериментировал со значениями learning\_rate и max\_epoch, а также поменял значения целевой переменной и независимых переменных.

Изображение выглядит как текст

Автоматически созданное описание

Первая итерация (0.01, 1):

Изображение выглядит как текст

Автоматически созданное описание

Вторая итерация (0.05, 2):

Изображение выглядит как текст

Автоматически созданное описание

Третья итерация (0.1, 3):

Изображение выглядит как текст

Автоматически созданное описание

Четвёртая итерация (0.15, 4):

Изображение выглядит как текст

Автоматически созданное описание

Пятая итерация (0.2, 4):

Изображение выглядит как текст

Автоматически созданное описание

Шестая итерация (0.2, 5):

Изображение выглядит как текст

Автоматически созданное описание

Седьмая итерация (0.25, 5):

**Изображение выглядит как текст

Автоматически созданное описание**

**Восьмая** итерация (0.3, 5)::

**Изображение выглядит как текст

Автоматически созданное описание**

**Девятая** итерация (0.35, 5):

**Изображение выглядит как текст

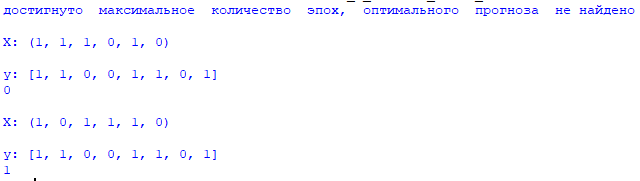
Автоматически созданное описание**

**Десятая** итерация (0.4, 5):

**Изображение выглядит как текст

Автоматически созданное описание**

**Одиннадцатая** итерация (0.4, 4):

****

**Двенадцатая** итерация (0.5, 4):

**Изображение выглядит как текст

Автоматически созданное описание**

**Тринадцатая** итерация (0.55, 4):

**Изображение выглядит как текст

Автоматически созданное описание**

**Четырнадцатая** итерация (0.55, 3):

**Изображение выглядит как текст

Автоматически созданное описание**

**Пятнадцатая** итерация (0.6, 3):

**Изображение выглядит как текст

Автоматически созданное описание**

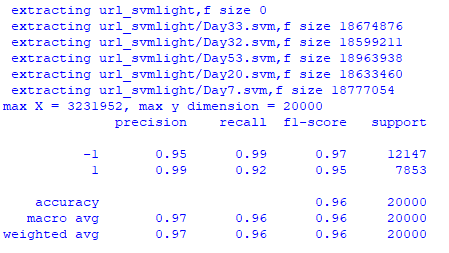
**Из полученных результатов можно сделать вывод, что слишком малое значение max\_epochs помешать обучению, так же себя ведёт learning\_rate. Корректируя эти параметры, можно добиться успешного обучения.**

**Часть 2.**

**Создал новый файл в IDLE, скопировал в него код и запустил. Файл переместил по пути, который указан изначально.**

Изображение выглядит как текст

Автоматически созданное описание

****

Изменил код так, чтобы он мог считать все файлы. Перенёс полученный данные в excel, а затем построил гистограмму.



В результате исследования мы видим, что значения метрик precision, recall,  f1-score и support в классе -1 равны соответственно 0,95, 0,99, 0,97, 12147. В классе 1 значения понизились до 0,99, 0,92, 0,95, 7853.

**Вывод:** В проделанной работе получил навыки для работы с большими данными на одном компьютере.