МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

**Обнинский институт атомной энергетики –**

филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ»

**(ИАТЭ НИЯУ МИФИ)**

Отделение интеллектуальных кибернетических систем

ЛАБОРАТОРНАЯ РАБОТА №4

«Машинное обучение »

по дисциплине

«Большие данные»

Выполнил студент 1 курса

группы ИВТ-М20

Лискунов Р. Г.

Проверил:

кандидат технических наук

Грицюк С. В.

Обнинск, 2020

# Цель работы

Определить простую задачу машинного обучения и решить ее. Точность на испытательном наборе должна быть не менее 60%.

# Краткая теория

Apache Spark MLlib используется для создания приложения машинного обучения. Приложение выполняет прогнозный анализ на открытом наборе данных. MLlib — это основная библиотека Spark, которая предоставляет множество служебных программ, полезных для задач машинного обучения, таких как:

1. Классификация;
2. Регрессия;
3. Кластеризация;
4. Моделирование сингулярного разложения и анализа по методу главных компонент;
5. Проверки гипотез и статистической выборки.

## Общие сведения о классификации и логистической регрессии

Классификация — это распространенная задача машинного обучения, которая представляет собой процесс сортировки входных данных по категориям. Это задание алгоритма классификации, позволяющее определить, как назначить "метки" входным данным, которые вы предоставляете. Например, можно представить алгоритм машинного обучения, который принимает в качестве входных данных данные о акции. Затем делит биржевую акцию на две категории: акции, которые следует продавать и акции, которые следует хранить.

Логистическая регрессия — один из алгоритмов классификации. API Spark для логистической регрессии подходит для задач двоичной классификации или разделения входных данных на две группы. Дополнительные сведения о логистической регрессии см. в статье Википедии.

В целом, процесс логистической регрессии создает логистическую функцию. Используйте функцию для прогнозирования вероятности того, что входной вектор принадлежит одной группе или другой.

# Ход работы

# Листинг кода

# Вывод

В ходе данной лабораторной работы я изучил особенности и возможности способов работы со Spark ML. Для этого я использовал набор данных популярных детских имен.