**Разработка программного модуля**

Блок-схема алгоритма получения набора данных с равномерным распределением значений по классам и заданным процентным соотношением относительно исходного набора данных *ГУИР.400201.2 БС*.

Данный алгоритм позволяет гарантировать равномерное обучение классификатора благодаря получению сбалансированного набора данных.

Алгоритм получает на вход значение типа integer, обозначающее желаемый объем результирующего набора данных относительно исходного (от 1% до 100%) .

Результатом работы алгоритма является искомый набор данных в виде объекта класса SupervisedDataSet из библиотеки PyBrain, который содержит ассоциативный массив, состоящий из массива входных данных «input» и массива выходных данных «target».

Словесное описание алгоритма:

1. Проверить переданное в функцию значение процентов на корректность (оно не должно быть меньше 0 и больше 100)
2. Если передан параметр 100%
   1. Вызвать функцию получения полного набора данных и вернуть полученное значение
3. Получить возможные классы выходных значений
4. Рассчитать общее число значений и число значений на каждый класс в результирующем наборе данных
5. Для каждого класса выходных данных из полученных ранее
   1. Получить частичный набор данных, состоящий только из значений текущего класса
   2. Пока не достигнуто необходимое число значений в классе
      1. Сгенерировать случайное число – номер случайного набора значений в наборе данных текущего класса
      2. Если полученное значение ранее не добавлялось в результирующий набор данных
         1. Добавить номер значение в список уже сгенерированных номеров
         2. Получить значение под текущим номером из набора данных текущего класса
         3. Добавить полученное значение в результирующий набор данных
6. Вернуть результирующий набор данных

Данный алгоритм получения набора данных с равномерным распределением значений по классам используется для всех алгоритмов классификации, присутствующих в приложении. Именно его использование позволяет избежать известной ошибки слишком узкого обучения классификаторов, в результате которой классификатор не имеет возможности точно определить класс данных из тестовой выборки.