Метод фильтрации спам-сообщений электронной почты на основе нейронной сети и метода опорных векторов

Студент: Яковлев Денис Владиславович

Группа: ИУ7-81Б

Научный руководитель: Русакова Зинаида Николаевна

Москва, 2024

Актуальность работы

Проблемы, создаваемые рассылками спама:

- трата трафика «в пустую»;
- угроза безопасности системы, попадающей под рассылку;
- трата рабочего времени.

Цели и задачи

Цель работы – разработка и программная реализация метода фильтрации спам-сообщений электронной почты на основе нейронной сети и метода опорных векторов.

Задачи:

- проведен обзор и сравнительный анализ основных методов классификации;
- разработан метода фильтрации спам-сообщений;
- осуществлена программная реализация разработанного метода;
- проведено исследование и сравнение разработанного ПО с существующими аналогами.

Известные решения

Вид / Критерий	Фильтрация вручную	Автоматизированная фильтрация
Требование постоянного корректирования пользователем	+	-
Масштабируемость	-	+
Своевременность	-	+

Сравнительный анализ методов классификации

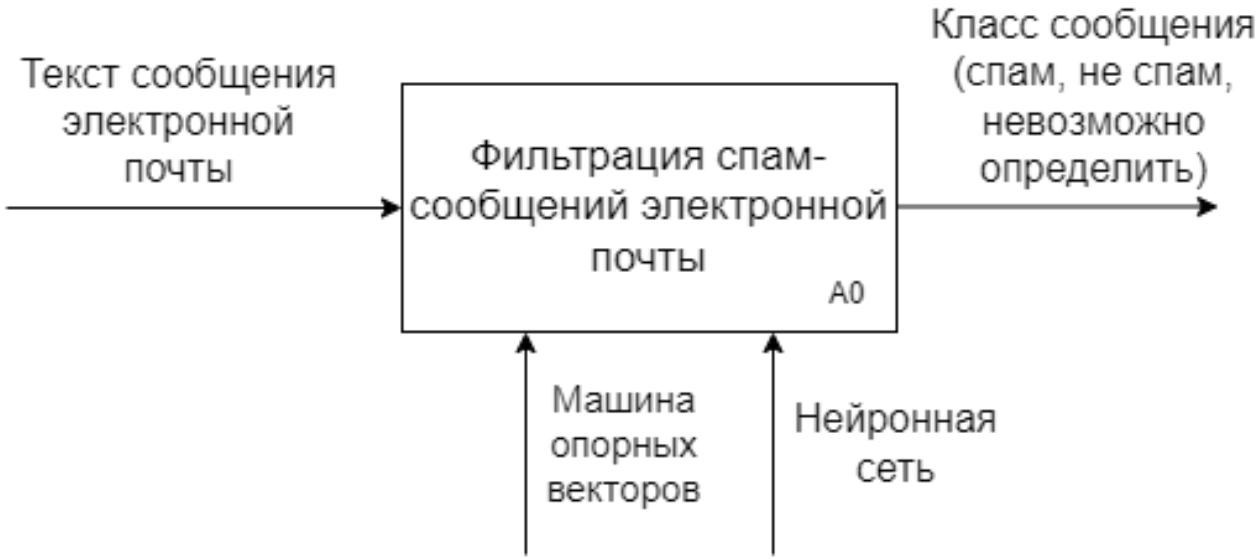
1. К1 – Устойчивость к выбросам.
2. К2 – Требование к тщательной настройке параметров.
3. К3 – Возможность переобучения.
4. К4 – Возможность прогнозирования непрерывных функций.
5. К5 – Формальность вывода.
6. К6 – Возможность работы с разреженными данными

	К1	К2	К3	К4	К5	К6
Логистическая регрессия	Нет	Да	Да	Нет	Да	Да
Случайный лес	Да	Нет	Да	Да	Нет	Да
Метод опорных векторов	Нет	Нет	Нет	Да	Да	Да

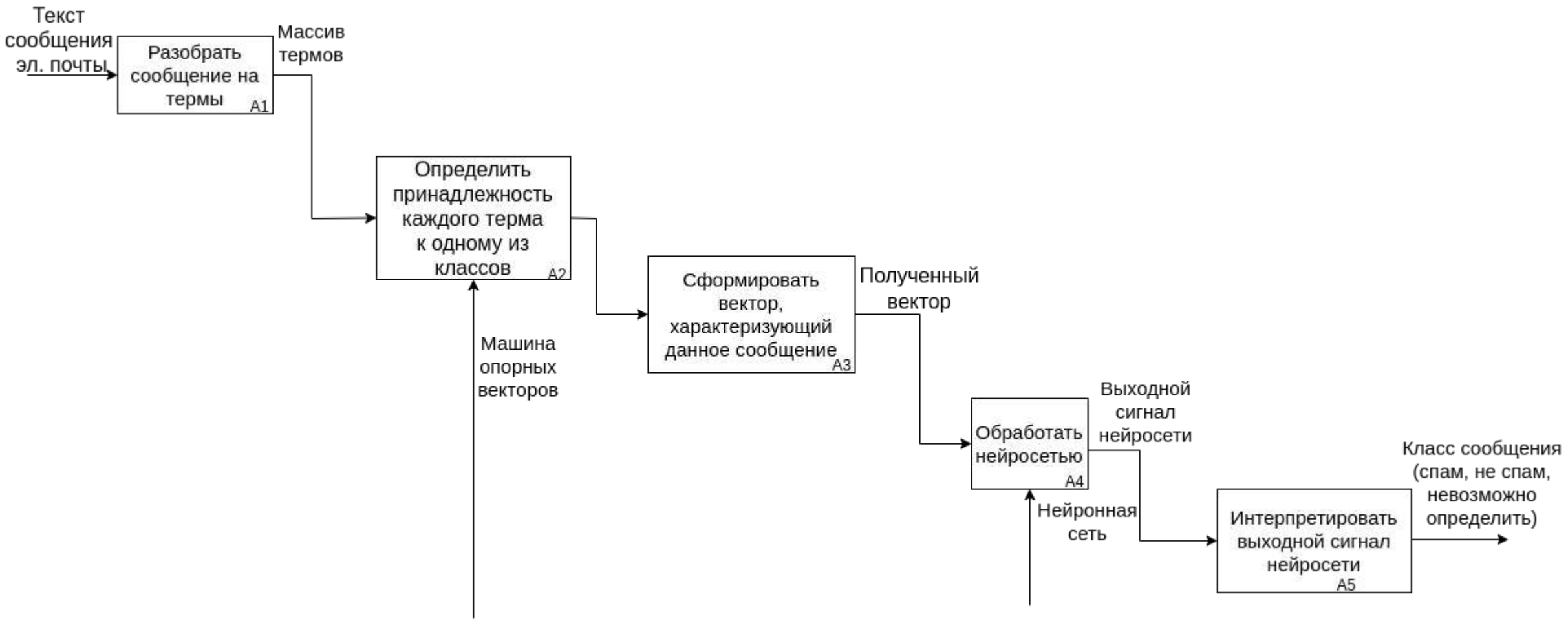
Постановка задачи

Ограничения, накладываемые на метод:

- метод должен иметь конечное время работы
- метод должен уметь обрабатывать любые сообщения

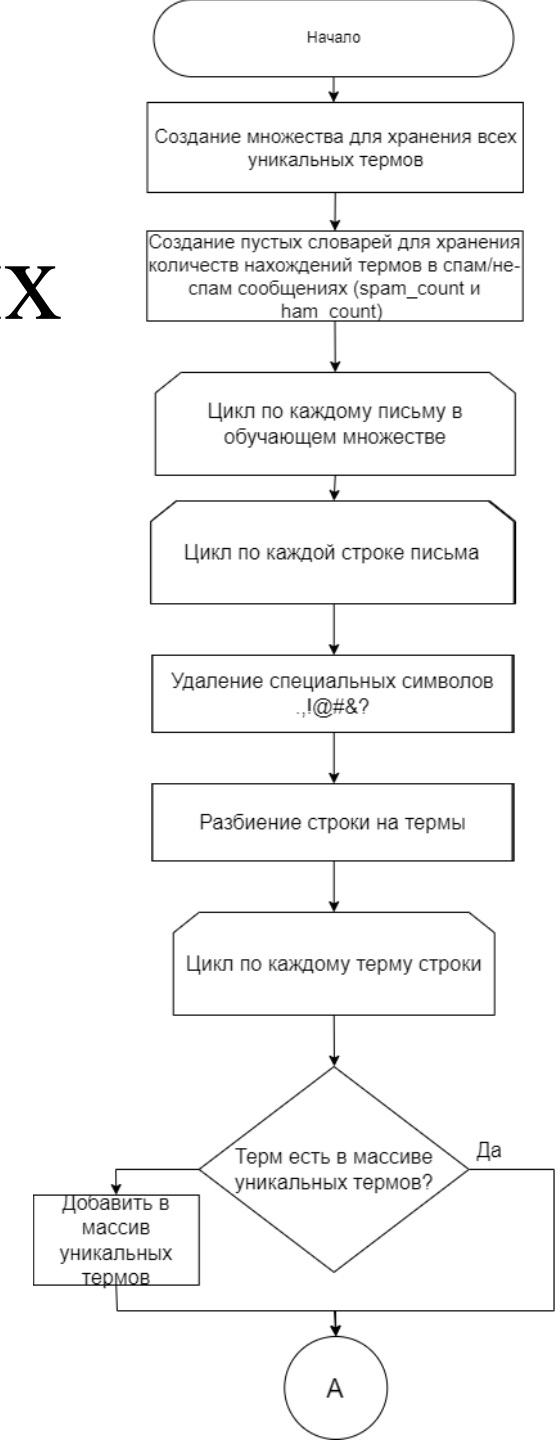


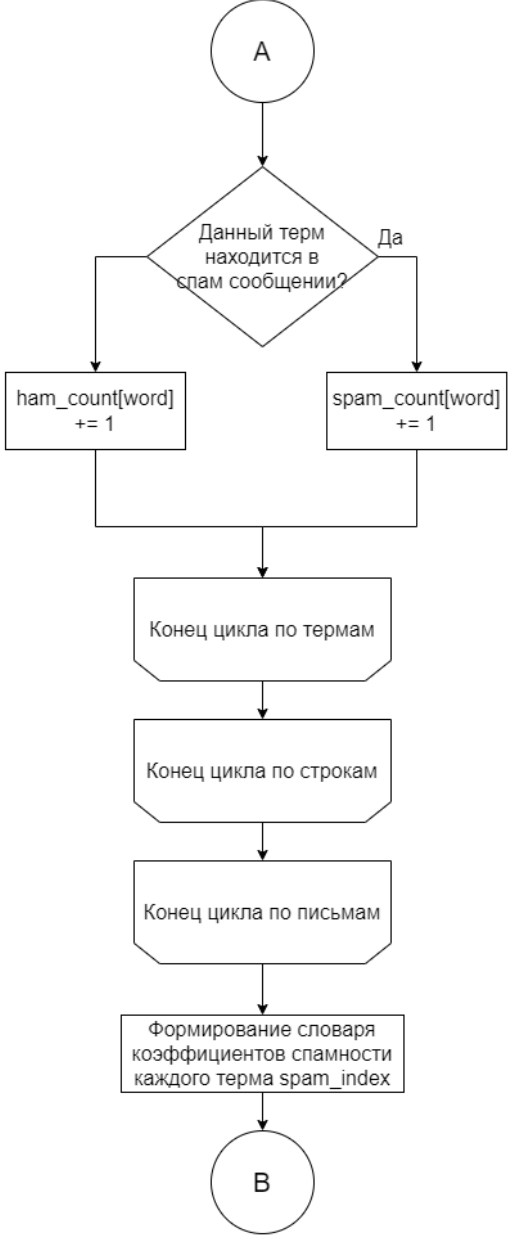
Функциональная схема метода

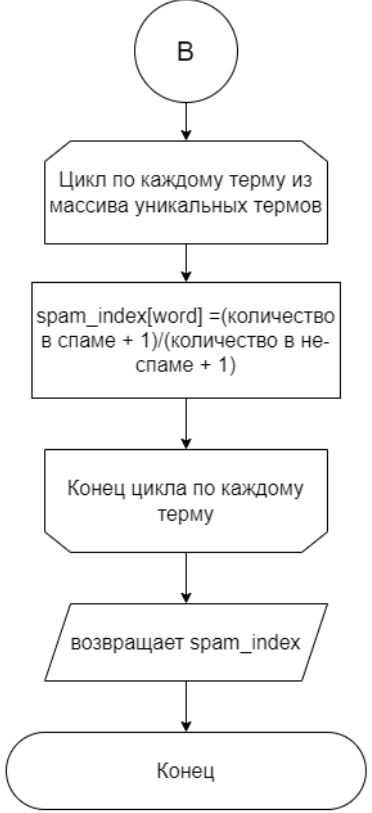


Обработка входных данных

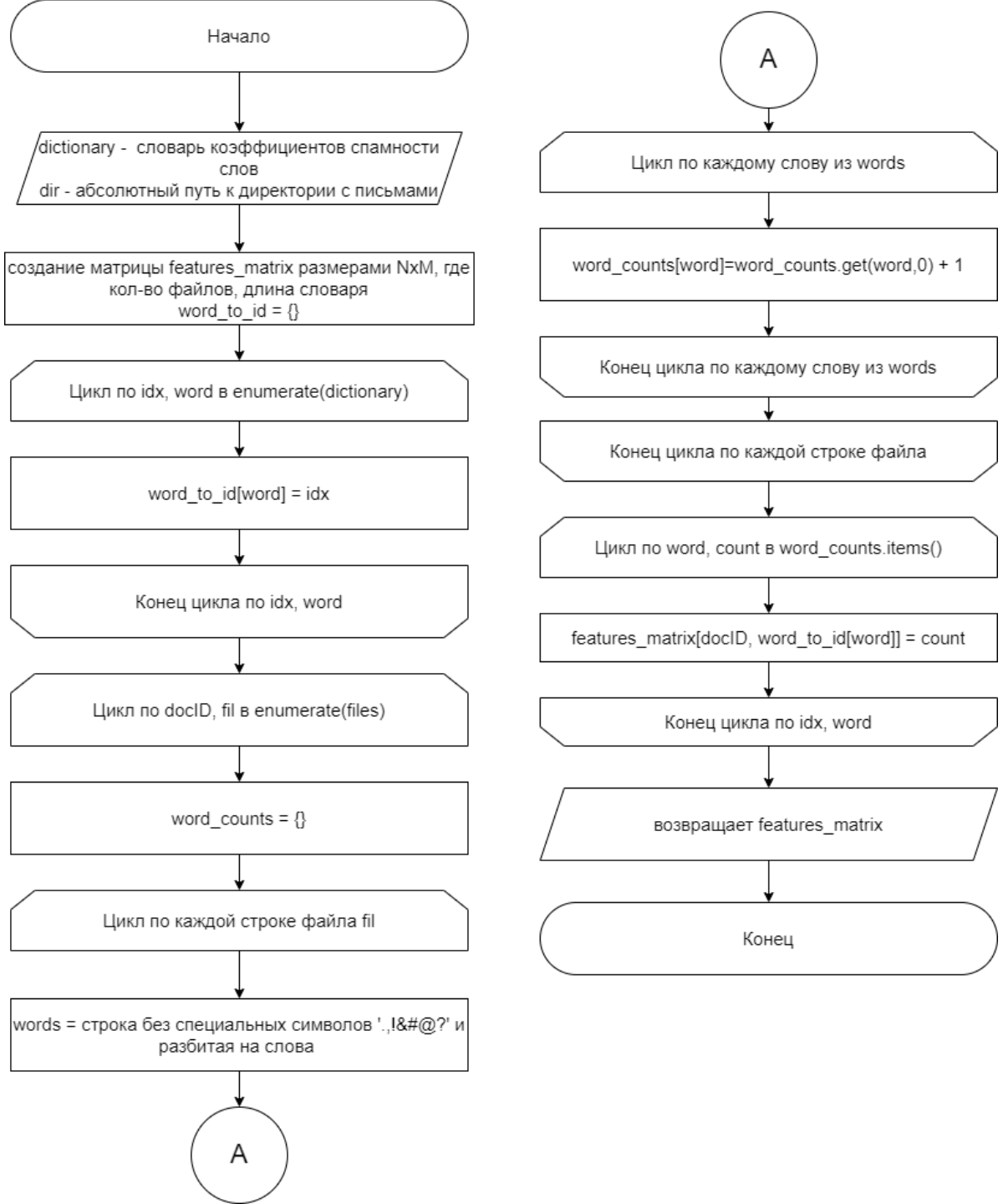
Здесь и далее:
— @ — оператор умножения матриц;
— {} — оператор создания пустого словаря.





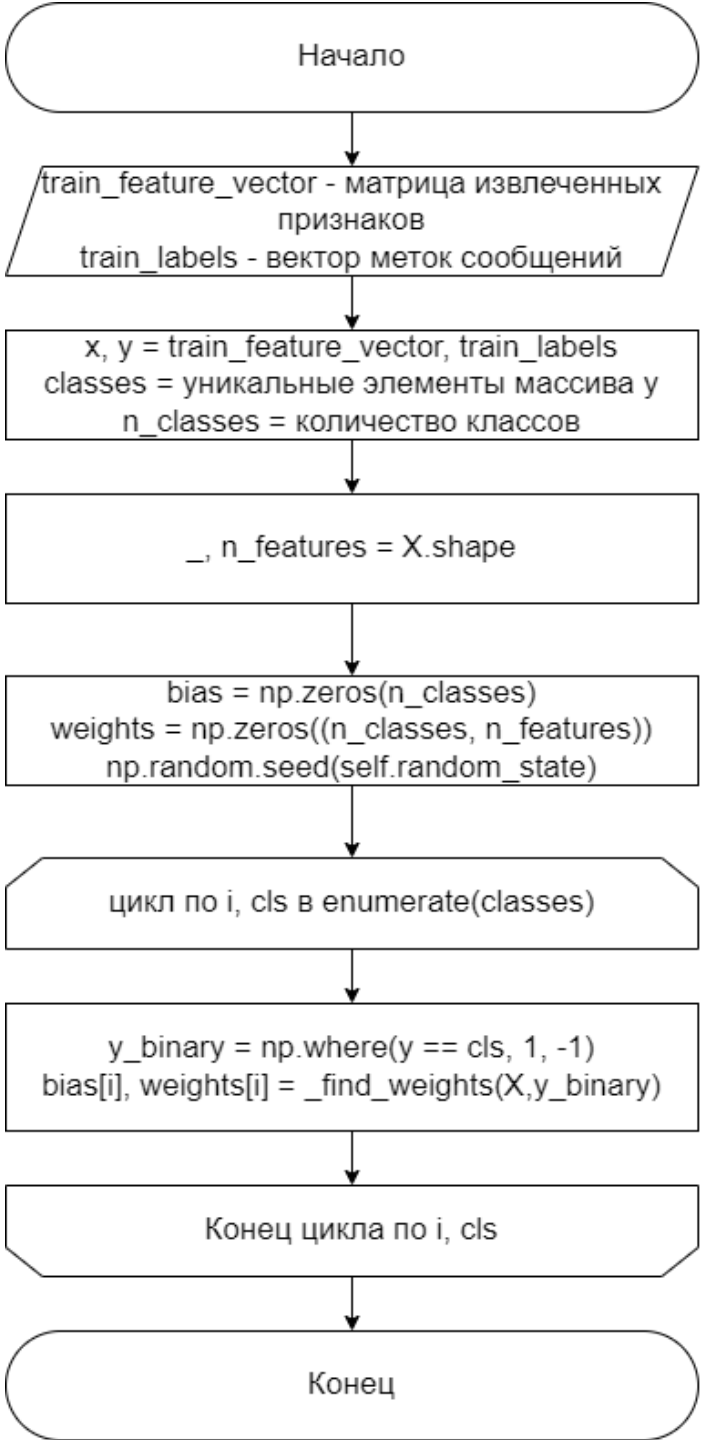


Формирование матрицы признаков

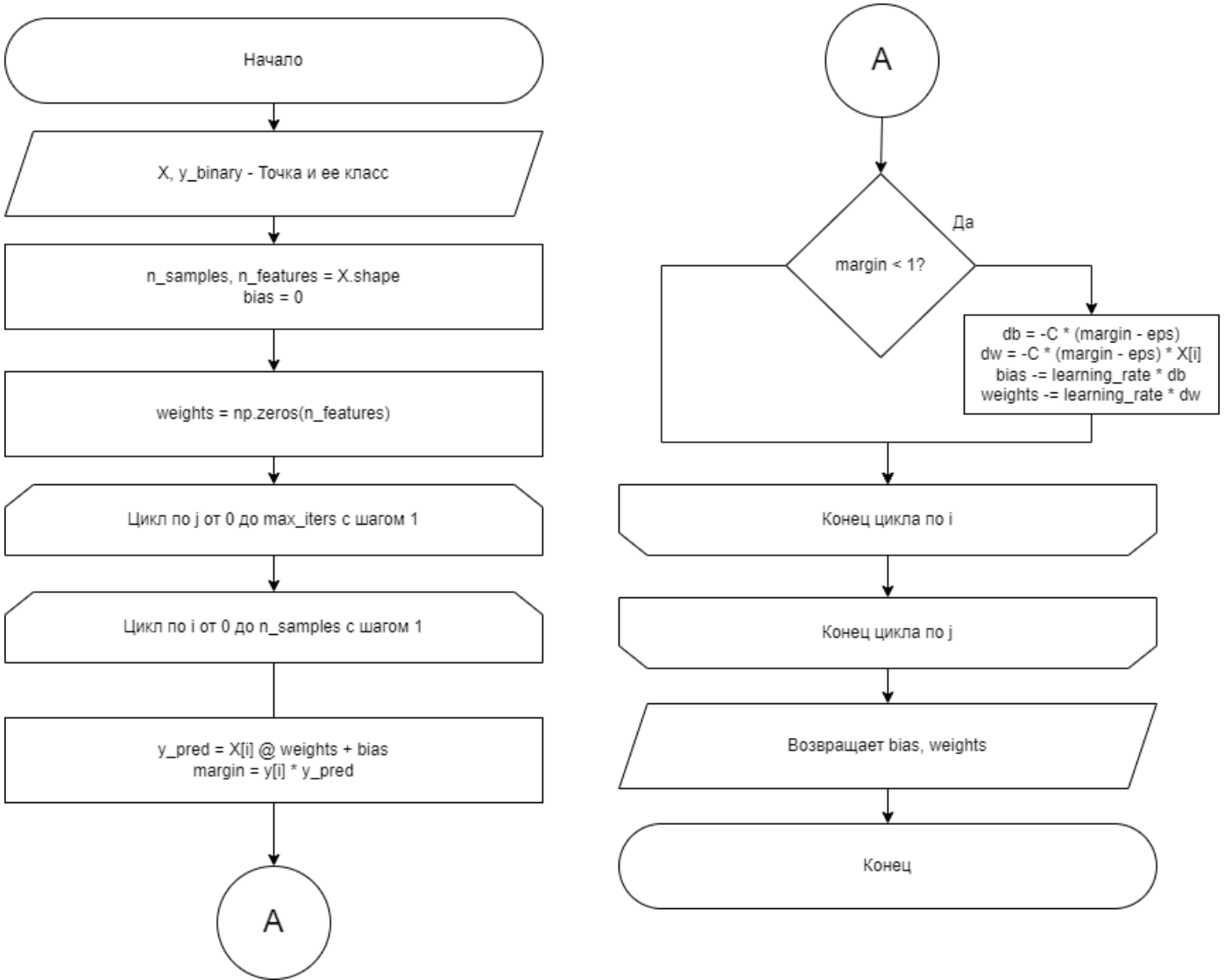


Обучение машины опорных векторов

Было выбрано линейное ядро ввиду линейной разделимости данных.

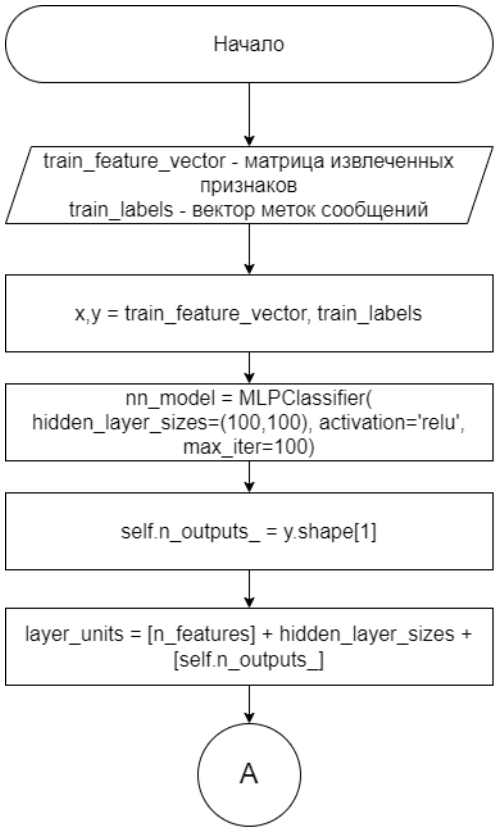


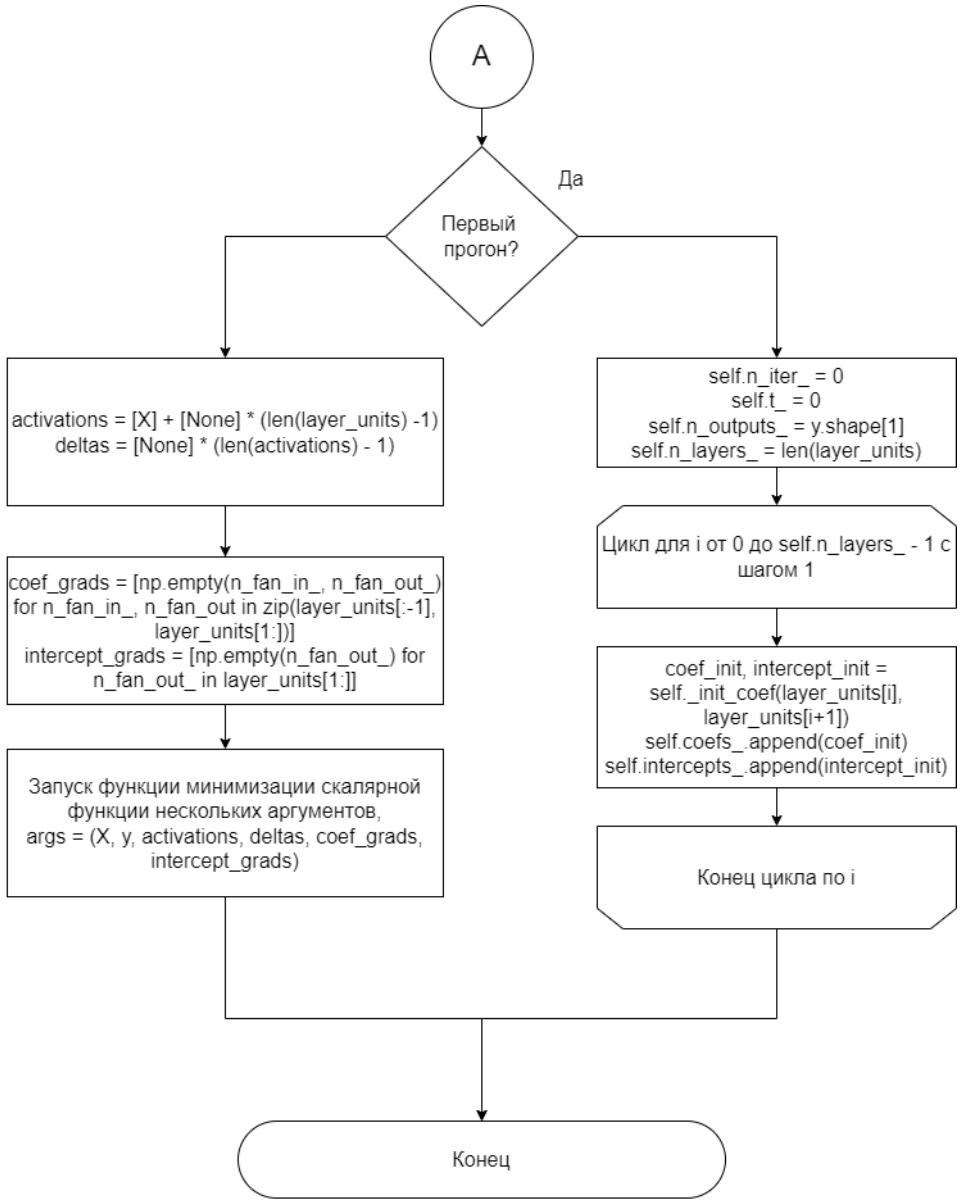
Функция нахождения коэффициентов гиперплоскости для SVM



Обучение нейросети

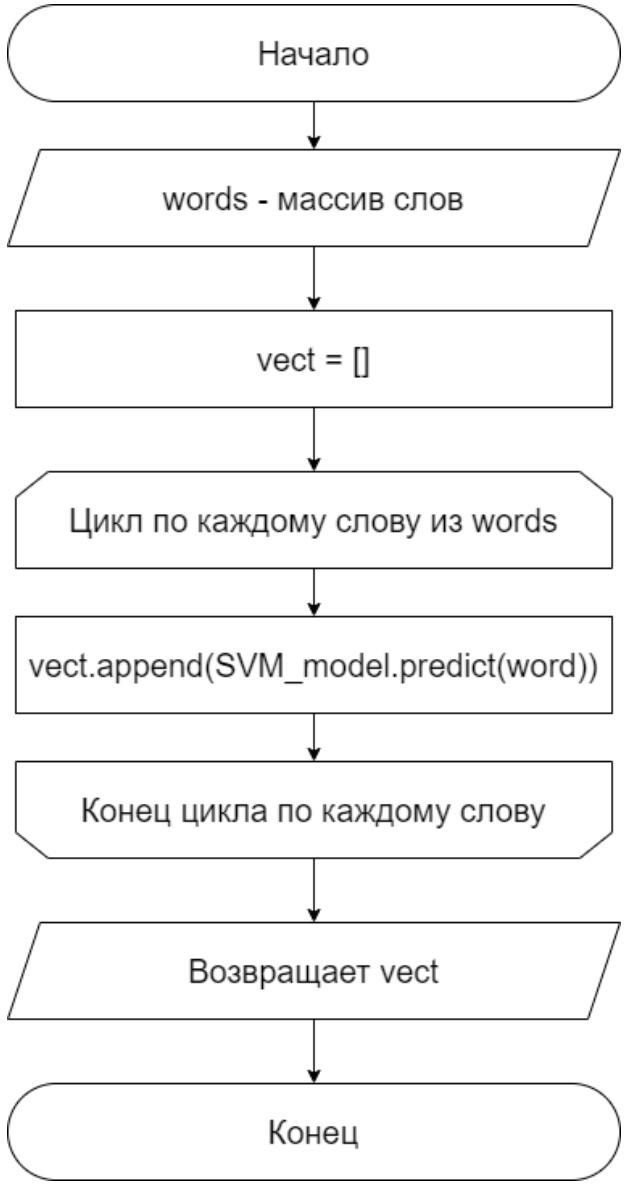
В качестве архитектуры нейронной сети был выбран многослойный перцептрон Румельхарта ввиду того, что он может работать с векторными представлениями текстовых данных.
На входном слое 32 нейрона, при поступлении вектора размером больше 32, вектор разбивается на куски по 32 элемента с перекрытием в 16 элементов, 2 скрытых слоя по 100 нейронов в каждом, количество выходов равно 1





Построение вектора репутационных коэффициентов

Под репутационным коэффициентом здесь понимается соотношение количества появлений слова в спам сообщениях и количества появлений в легитимных сообщениях



Обработка входного вектора нейросетью

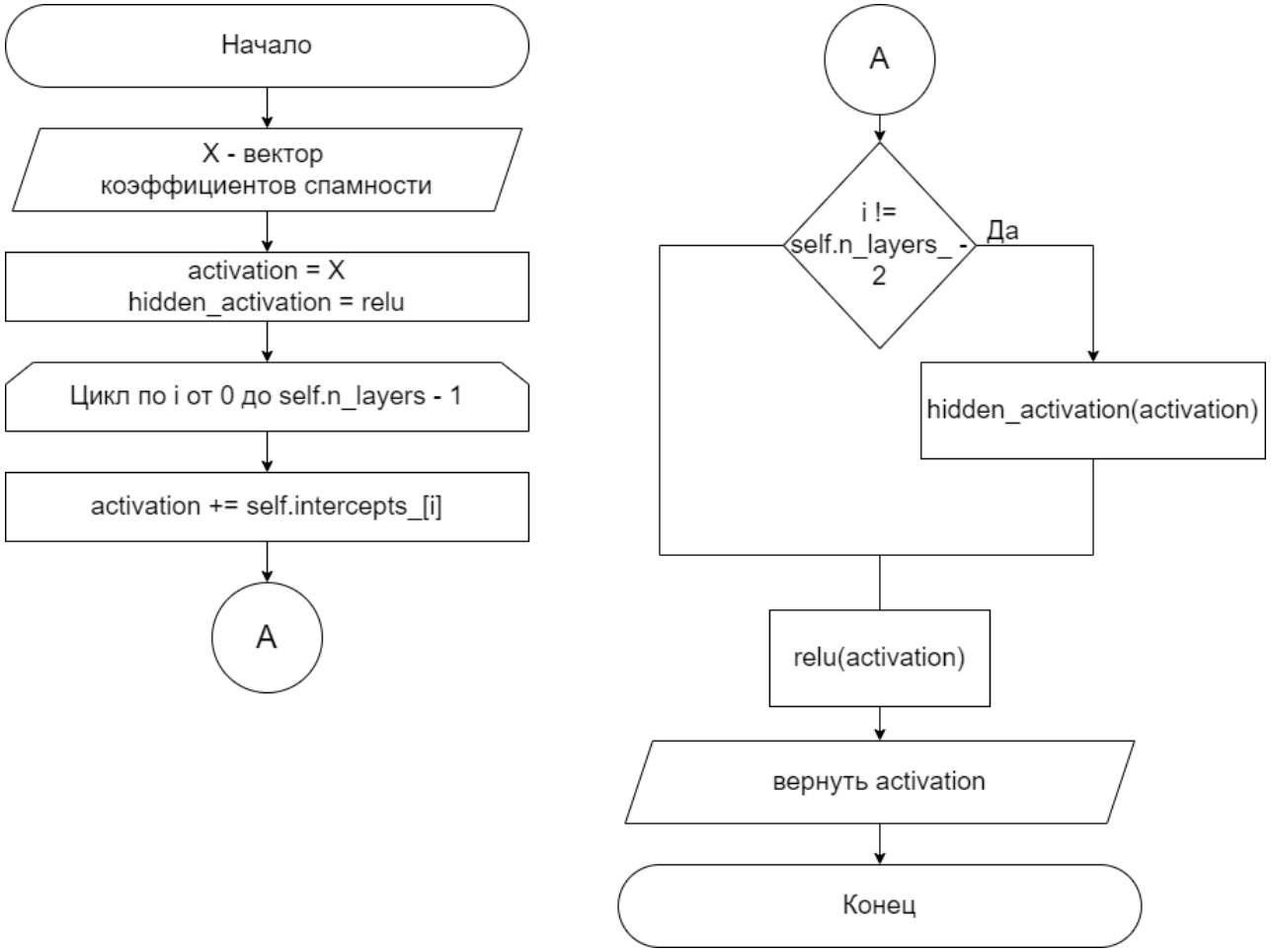
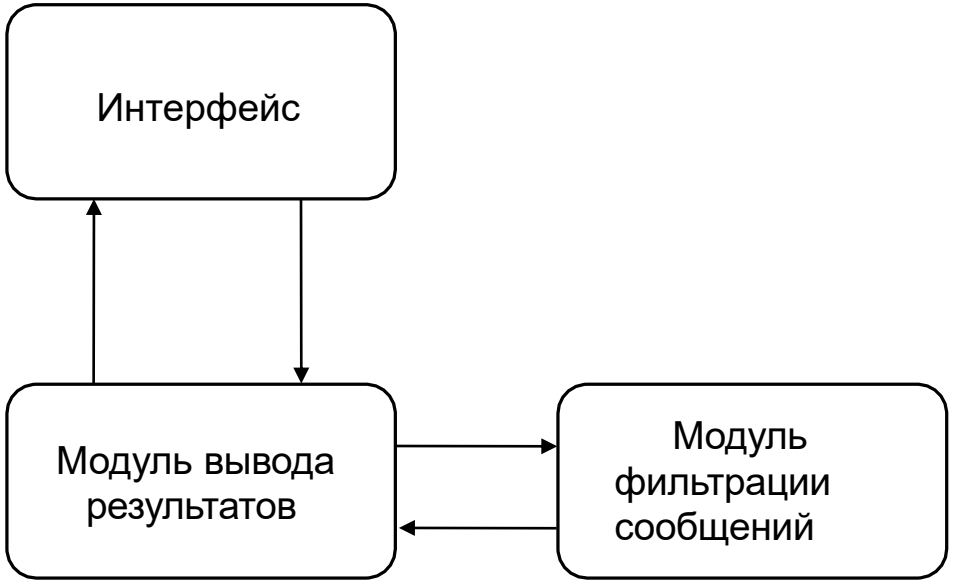
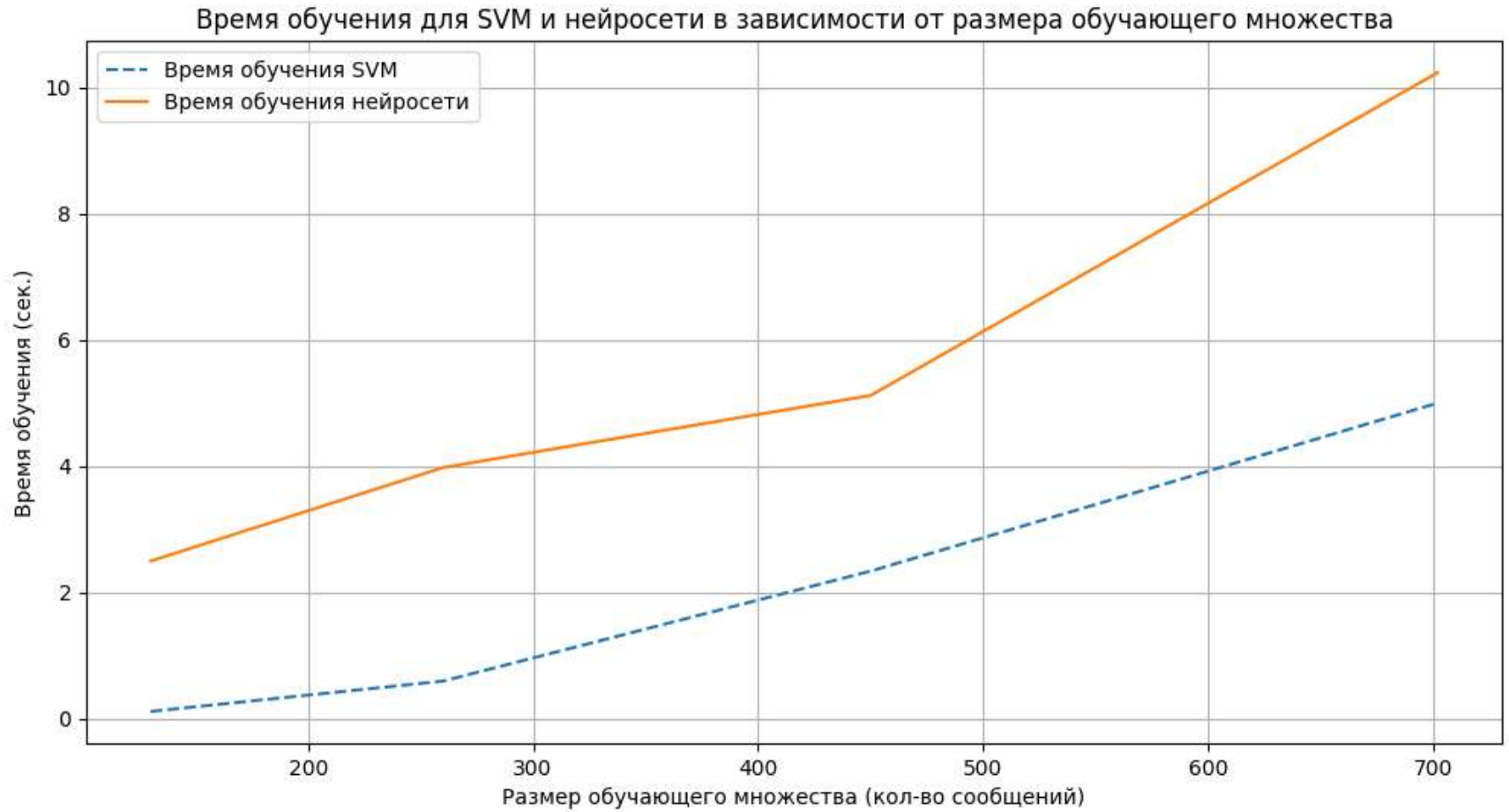


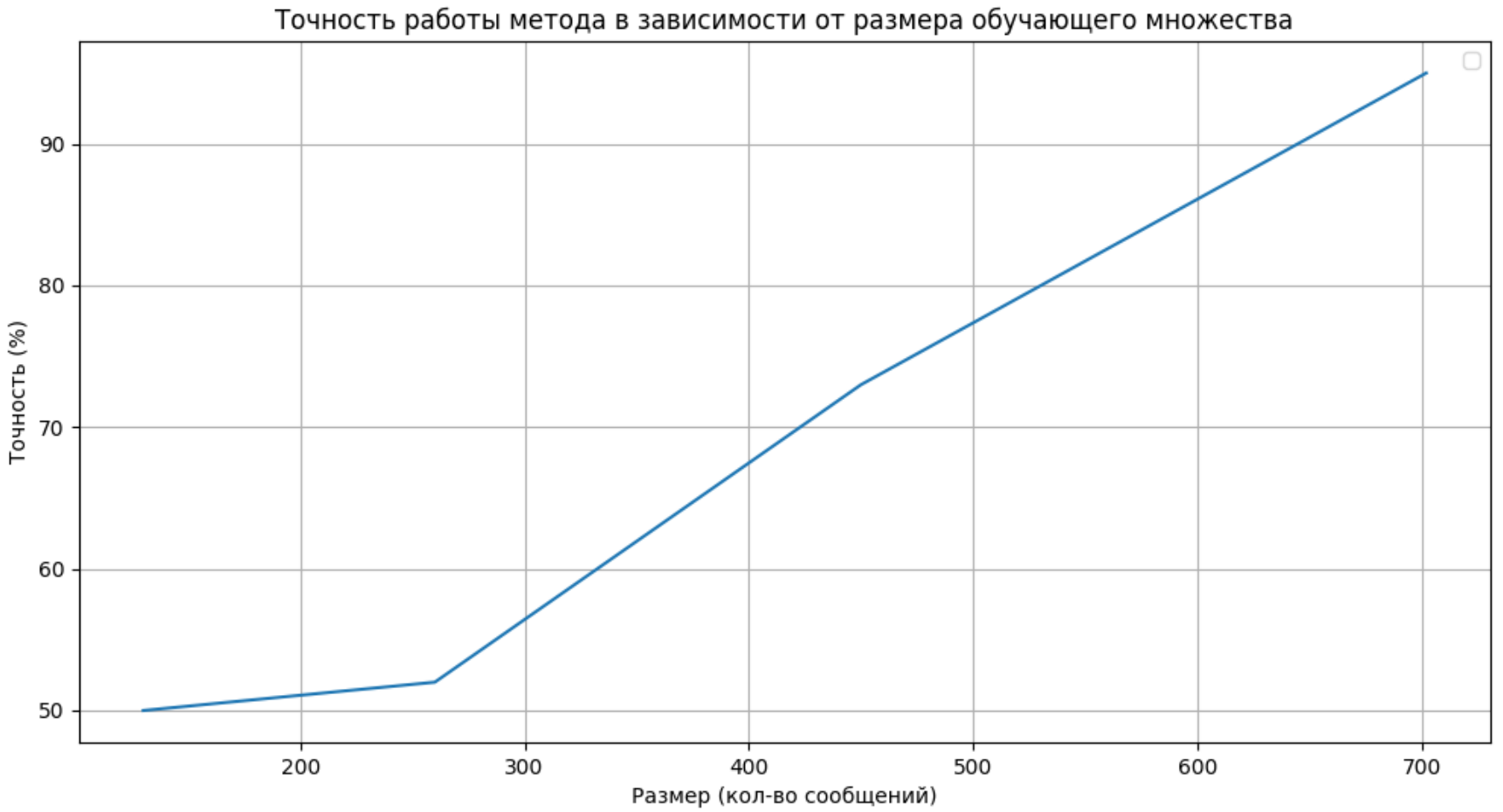
Схема взаимодействия модулей



Зависимость времени обучения и точности от размера обучающей выборки



Зависимость точности разработанного метода от размера обучающей выборки



Сравнение разработанного метода с известными аналогами

К1 — Использование правил для фильтрации.
К2 — Использование сигнатурного анализа.
К3 — Использование репутационных методов
К4 — Использование машинного обучения.

Спам-фильтр	Точность	К1	К2	К3	К4
Разработанный метод	95%	Нет	Нет	Нет	Да
Microsoft 365 Defender	97%	Да	Нет	Да	Да
Proofpoint	95%	Да	Нет	Да	Да
Fortinet FortiMail	96%	Да	Да	Да	Да
Trend Micro	98%	Да	Да	Да	Да
Symantec	96%	Да	Да	Нет	Да
Sophos	95%	Да	Да	Нет	Да
Barracuda Essentials	94%	Да	Нет	Да	Да
Cisco Secure Email	93%	Да	Нет	Да	Да
Zoho Mail	92%	Да	Нет	Да	Да

Заключение

Достигнута цель работы: разработан метод фильтрации спам-сообщений электронной почты на основе нейронной сети и метода опорных векторов.

Были решены все поставленные задачи:

- проведен обзор и сравнительный анализ основных методов классификации;
- разработан метода фильтрации спам-сообщений;
- осуществлена программная реализация разработанного метода;
- проведено исследование и сравнение разработанного ПО с существующими аналогами.