# АЛГОРИТМ ОБНАРУЖЕНИЯ СПАМА ДЛЯ КОМПЛЕКСНОЙ ЗАЩИТЫ ЭЛЕКТРОННОЙ ПОЧТЫ

Т.Ю. Голиков
Научный руководитель – Тонкович И.Н.
канд.хим.наук, доцент
**Белорусский государственный университет информатики и радиоэлектроники**

С развитием информационных технологий люди по всему миру получили возможность коммуницировать между собой. Различные компании начали использовать возможности информационно-коммуникационных технологий для продвижения своих собственных товаров и услуг. Некоторые фирмы рекламируют свой товар только тем клиентам, которые согласились на данный тип рекламы, и в таком случае фирма устраивает рассылки. Однако есть компании, которые отправляют весь материал, без согласия всем возможным пользователям. Данные сообщения могут содержать как простую рекламу, так и вредоносные программы, цель которых кража конфиденциальных данных или получение контроля над устройством.

Главная проблема для обнаружения спама – позволить программе видеть разницу между рабочей рассылкой и спамом. Для решения данной проблемы предлагается алгоритм определения аномального поведения. В основу алгоритма положена регрессионная модель дерева принятия решений.

Первым вопросом в дереве решений является вопрос, было ли предыдущее сообщение расценено пользователем, как спам. Поэтому в случае, если сообщение не было спамом, проще всего предположить, что спамом оно будет только в том случае, если оно было отправлено не только одному пользователю, что может быть следующим вопросом в дереве принятия решений.

Вторым шагом является проверка содержимого сообщения. На сегодняшний день многие компании используют проверку действий пользователя, как история запросов, интересующие темы и другие. Система сканирует каждое слово, полученное в сообщении, а также в заголовке. Если сообщение содержит в себе темы, интересующие пользователя, тогда оно может попасть к нему. В противном случае оно отправляется на проверку с использованием наивного байесовского классификатора [1].

Различные алгоритмы машинного обучения могут обнаруживать спам, но более привлекательным является наивный байесовский алгоритм, основанный на теореме Байеса.

Принимая во внимание предложенный алгоритм определения аномального поведения, можно сделать вывод, что проверка спама для каждого пользователя будет улучшаться с каждым последующим сообщением. Однако учитывая алгоритм работы байесовского классификатора, нет гарантии отсутствия ошибок. Следовательно, предложенный алгоритм нужно улучшать, чтобы не оперировать понятием вероятность.

*Библиографический список*

1. Попов, В. А. Теория вероятности / В. А. Попов. – Казань : ФГАОУВПО «Казанский (Приволжский) федеральный университет», 2013. – 49 с.