

АВТОМАТИЗАЦІЯ ДОКУМЕНТООБІГУ ПОШТОВОГО ВІДДІЛЕННЯ. ЧАСТИНА 2. ФУНКЦІЇ НА СТОРОНІ СЕРВЕРА

¹ Вінницький національний технічний університет;

Анотація

Запропоновано метод по автоматизації документообігу поштового відділення, а саме функції на стороні сервера.

Ключові слова: пошта, повідомлення, спам, автоматизація.

Abstract

A method for automating the circulation of the post office, namely the functions on the server side..

Keywords: mail, messages, spam, automation.

Вступ

З появою та розвитком мережі Інтернет великої популярності набула електронна пошта. Швидкість передачі електронних повідомлень зробили її незамінною спочатку у бізнес-сфері, а згодом і серед звичайних користувачів. Щороку кількість зареєстрованих електронних скриньок зростає, але із зростанням популярності пошти збільшується і масштаб проблем, що з'явилися з її появою.

Однією з основних проблем пошти є спам-повідомлення. За статистикою у 2019 році бізнес-користувач пошти отримує в середньому 96 електронних повідомлень на день, близько 20% з яких є рекламою або спамом [1]. За останні 4 роки кількість спаму виросла майже на 10%. Очевидно, що в майбутньому кількість повідомлень буде лише зростати, але, разом з тим, вже зараз існує потреба у їх класифікації для фільтрування від спаму.

Всі повідомлення, що минувають спам-фільтр, потрапляють до однієї купи. Користувачу пошти важко знайти необхідне електронне повідомлення серед десятків інших, оскільки відсутня класифікація їх за категоріями.

На сьогоднішній день існують методи для фільтрації та класифікації поштових повідомлень за категоріями. Усі вони є недостатньо ефективними, оскільки мають досить низьку точність через використання єдиного класифікатора для фільтрації та класифікації за категоріями.

Отже, підвищення точності автоматизованої класифікації поштових повідомлень є актуальною задачею.

Результати дослідження

Запропоновано комбінований метод автоматизованої класифікації текстового вмісту поштових повідомлень з метою підвищення точності класифікації, а також надання можливості класифікації повідомлення за декількома категоріями.

Проаналізовано існуючі методи класифікації текстового вмісту повідомлень, визначено їх основні переваги та недоліки для класифікації поштових повідомлень.

Даний метод був реалізований у вигляді веб-додатку пошти з можливістю групування повідомлень за категоріями. Обґрунтований вибір технологій, обраних для розробки веб-додатку. Описана архітектура розробленого веб-додатку, а також структура його серверної та клієнтської частин.

. Хоча в комбінований методі досить повільний час тренування, можливість зберігати моделі позбавляє необхідності виконувати тренування моделі при кожному запуску сервера. На рис. 4.1 зображений графік залежності часу підготовки моделей від розміру вибірки. З ростом кількості повідомлень у вибірці час завантаження і десеріалізації моделі майже не змінюється

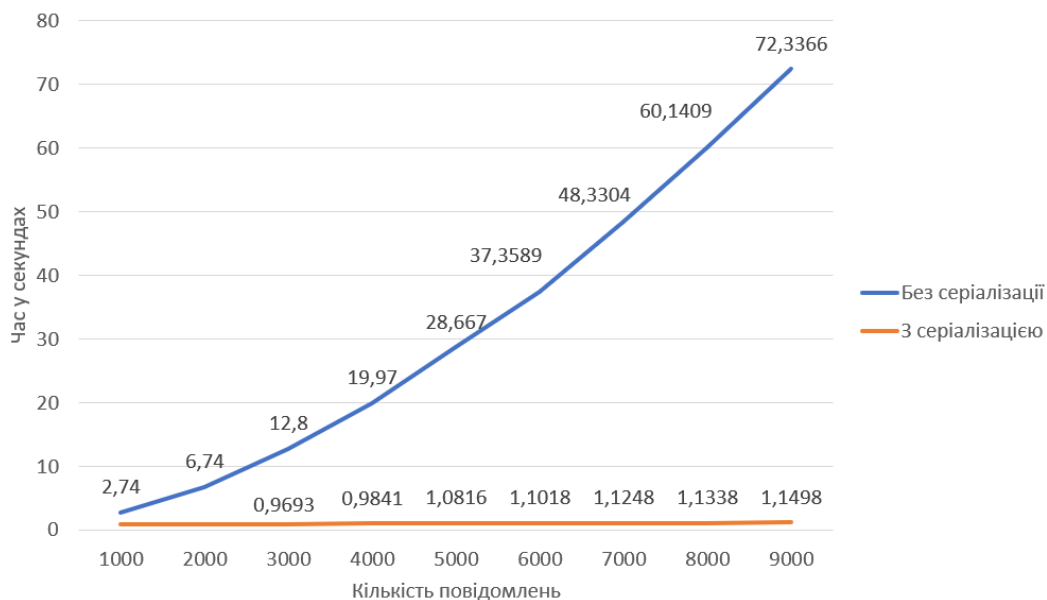


Рис. 1. Графік залежності часу підготовки моделей від розміру вибірки

За результатами тесту комбінованої класифікації визначено, що розроблений комбінований метод автоматизованої класифікації текстового вмісту поштових повідомлень має набагато кращу точність класифікації, в порівнянні з іншими методами, в яких не розділені етапи фільтрації та класифікації за категоріями.

Хоча в комбінований методі досить повільний час тренування, можливість зберігати моделі позбавляє необхідності виконувати тренування моделі при кожному запуску сервера. На рис. 4.1 зображений графік залежності часу підготовки моделей від розміру вибірки. З ростом кількості повідомлень у вибірці час завантаження і десеріалізації моделі майже не змінюється.

Висновки

Проведено порівняння результатів класифікації комбінованим методом автоматизованої класифікації текстового вмісту поштових повідомлень з іншими методами машинного навчання. Визначено, що реалізований метод надає більшу точність класифікації повідомлень ніж інші методи, використані для порівняння.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Email Statistics Report, 2015-2019. [Електронний ресурс]. – Режим доступу: <https://bit.ly/2PvB971>. Дата доступу – Жовтень 2019.
2. О задачах классификации. [Електронний ресурс]. – Режим доступу: <http://bazhenov.me/blog/2012/06/05/classification.html>. Дата доступу – Жовтень 2019.
3. Muller, Andreas. Introduction to Machine Learning with Python / A. C. Muller, S. Guido. – O'Reilly Media, Inc., 2017. – 400 p. – ISBN: 978- 1449369415.

Осипенко Ігор Олегович — студент групи 2АКІТ-18Б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: viskiscosta@gmail.com

Науковий керівник: **Ковтун Вячеслав Васильович** — д-р техн. наук, доцент, Вінницький національний технічний університет, м. Вінниця

Oyipenko Igor Olegovich. — Department of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: viskiscosta@gmail.com

Supervisor: **Kovtun Vyacheslav Vasilyevich.** — Dr. Tech. Sciences, Associate Professor, Vinnytsia National Technical University, Vinnytsia