Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____Комп'ютерних наук_____
(повна назва)

Кафедра _____Програмної інженерії_____
(повна назва)

АТЕСТАЦІЙНА РОБОТА
Пояснювальна записка

_____другий (магістерський)_____
(рівень вищої освіти)

Дослідження методів розпізнавання спаму у вхідних sms-повідомленнях для iOS
(тема)

Виконав: студент 2 курсу, групи ІПЗм-18-1
спеціальності 121- Інженерія програмного забезпечення
(код і повна назва спеціальності)
Освітньо-наукової програми
Інженерія програмного забезпечення_____
(повна назва освітньої програми)

Єрьоменко М. О.
(прізвище, ініціали)
Керівник _____проф. д.т.н. Єрохін А. Л._____
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри, проф. _____ З.В.Дудар

2020 р.

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Факультет_Комп'ютерних наук_____
Кафедра_Програмної інженерії_____
Рівень вищої освіти – другий (магістерський)

Спеціальність _121 – Інженерія програмного забезпечення_____
(код і повна назва)
Тип програми освітньо-наукова програма_____
Освітня програма_Інженерія програмного забезпечення_____

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«_____»_____ 20 ___ р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

студентові Єрьоменко Максиму Олесандровичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів розпізнавання спаму у вхідних sms-повідомленнях для iOS
затверджена наказом університету від "_27_" __03___2020_ р. № 473 СТ__
заповнюється вручну після отримання наказу
2. Термін подання студентом роботи до екзаменаційної комісії 20 травня 2020 р.
3. Вихідні дані до роботи *алгоритми класифікації текстів, пояснювальна записка. Використовувати macOS, середовище об'єктно-орієнтованого проектування Xocde.*
4. Перелік питань, що потрібно опрацювати в роботі *мета роботи, аналіз проблемної галузі і постановка задачі, огляд методів класифікації sms-спаму, етапи аналізу, існуючі рішення і бібліотеки.*_____
_____
_____

5 Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім’я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Спецрозділ	проф. д. т. н. Єрохін А. Л.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка *
1	Аналіз предметної галузі	30.03 – 5.04.2020	
2	Огляд існуючих методів	6.04 – 14.04.2020	
3	Реалізація та тестування програмного продукту	15.04 – 21.04.2020	
4	Підготовка пояснювальної записки	22.04 – 26.04.2020	
5	Спецчастина	27.04 – 1.05.2020	
6	Підготовка презентації та доповіді	2.05 – 4.05.2020	
7	Попередній захист	5.05.2020	
8	Нормоконтроль, рецензування	6.05 – 11.05.2020	
9	Занесення диплома в електронний архів	14.05 – 17.05.2020	
10	Допуск до захисту у зав. кафедри	17.05.2020	
* заповнюється вручну після виконання чергового пункту			

Дата видачі завдання _29__ ____03_____ 2020 р.

Студент _____
(підпис)

Керівник роботи _____ проф д. т. н. Єрохін А. Л.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Атестаційна робота магістра містить: 59 с., 3 рис., 1 табл., 24 джер., 13 формул.

СПАМ, SMS, ПОВІДОМЛЕННЯ, КЛАСИФІКАЦІЯ, iOS, МОБІЛЬНИЙ ДОДАТОК, ФІШИНГ, MACHINE LEARNING.

Метою роботи є дослідження методів класифікації SMS-спаму та розробка програмного додатку для фільтрації SMS-спаму на iOS пристроях.

Для розробки iOS додатка використана мова програмування Swift 5.2, середовище розробки Xcode 11.4, iMessage Filter Extension та бібліотека CoreML.

Для реалізації методів класифікації тексту використана мова програмування Python, web-додаток Jupiter Notebook та інструмент scikit-learn.

SPAM, SMS, MESSAGE, CLASSIFICATION, iOS, MOBILE APP, PHISHING, MACHINE LEARNING.

The purpose of this work is to analyze the classification methods of SMS-spam and to develop a software application for SMS-spam filtering on iOS devices.

Swift 5.2, Xcode 11.4, iMessage Filter Extension, and CoreML have been used to develop the iOS app.

Python programming language, the Jupiter Notebook web application, and the scikit-learn tool were used to implement the text classification methods.

ЗМІСТ

ВСТУП

На сьогоднішній день людство не може обійтися без комп'ютерних та мережевих технологій. Засоби зв'язку вторглися в більшість сфер життєдіяльності суспільства, починаючи з допомоги в освіті і закінчуючи рекламою і просуванням будь-якої платної продукції. Звідси виникла проблема отримання небажаних повідомлень, іншими словами, спаму.

Слово «спам» походить від назви марки консервів «SPAM» виробництва американської компанії Hormel Foods. Під час Другої світової війни цей продукт використовувався в якості продукту харчування американських солдатів, але, коли війна закінчилася, залишилися великі запаси продукції, і, щоб позбутися від них, компанія стала вести дуже активну рекламу. З тих пір слово «спам» прижилося як назва рекламної розсилки [1].

Комерційні організації роблять масову розсилку, в тому числі і людям, які не хотіли б отримувати подібні повідомлення. Іноді вони навіть становлять небезпеку, тому що можуть містити комп'ютерні віруси, шахрайські посилання. Крім того, широкомовні розсилання тягнуть за собою великі витрати ресурсів сервера, віднімають час користувача, що витрачається на прочитання і сортування подібних листів.

Виділяють різні види спаму:

– фішинг – спроба дізнатися секретні дані, такі як паролі, номери банківських карт та інше;

– реклама;

– антиреклама – інформація, спрямована на зменшення інтересу користувача до продукції будь-якої компанії, до відомої особистості або гучній події;

– «Нігерійські листи» – носять шахрайський характер, в яких йдеться про нібито отриманому спадщині і прохання одержувача надіслати трохи грошей для оформлення документів. Таким чином порушник закону виманює гроші у обманутої людини;

За способами поширення спам класифікується:

– відправляється на електронну пошту – як правило, це спам у вигляді «Нігерійських листів» або реклами;

– посилається у вигляді SMS по мобільній мережі – зазвичай реклама або фішинг;

– відправляється користувачам соціальних мереж.

За статистикою частка спаму в електронній пошті становить близько 60% в SMS-повідомленнях – 15%. Такий стан справ є приводом для розвитку технологій фільтрації повідомлень.

Зростання користувачів мобільних телефонів привів до різкого збільшення кількості небажаних SMS-повідомлень. Незважаючи на те, що для електронної пошти існує багато різних фільтрів, боротьбі зі спамом на мобільних телефонах приділяється не так багато уваги. Таким чином актуальними будуть дослідження, які пов‘язані з проблемою SMS-спаму, оглядом існуючих рішень, їх недоліки та пошук нового підходу для пошуку небажаних SMS-повідомлень.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Класифікація спаму

Термін SPAM з'явився ще в 1937 р – так називалися свинячі консерви компанії Hornel Foods (Shoulder of Pork and HAM – «свинячі лопатки і окіст»). Отруєння цим продуктом і подальші скандали практично назавжди пов'язали слово «спам» з безсоромним і безвідповідальним рекламуванням часто не цілком доброякісного товару. В даний час спам можна визначити як нелегальну масову розсилку по електронній пошті рекламних й інших не цікавлять одержувача матеріалів. В наш час, практично за 10-15 років, спам перетворився з легкого дратівного фактору в одну з найсерйозніших загроз інформаційної безпеки. Непрохані поштові повідомлення переповнюють індивідуальні поштові скриньки користувачів і паралізують роботу корпоративних серверів. Найчастіше користувачі просто не звертають уваги на мережеву рекламу, щодня видаляючи такі повідомлення зі своїх поштових скриньок. Насправді згубність таких розсилок полягає в тому, що вартість їх для спамера практично незначна, при цьому вони дорого обходяться всім іншим – як одержувачу спаму, так і його провайдеру. Велика кількість рекламної кореспонденції може привести до зайвого навантаження на канали і поштові сервери провайдера, через що звичайна пошта, яку, можливо, дуже чекають одержувачі, буде проходити значно повільніше. Спамер практично не платить значних сум за те, що передає свої повідомлення. За все розплачується одержувач спаму, який оплачує своєму провайдеру час в мережі Інтернет, що витрачається на отримання непроханої кореспонденції з поштового сервера. При цьому в високорозвинених країнах відсоток спаму в електронній пошті і збитки від нього істотно вище, ніж в країнах, що тільки приступили до освоєння мережі Інтернет.

Для «зомбування» домашніх комп'ютерів з метою підготовки їх для розсилки спаму, використовуються комп'ютерні віруси (зазвичай «поштові черв'яки»), такі як Sobig, Sinit, Fizzer, MyDoom і ін. Інфіковані комп'ютери (таких може виявитися сотні тисяч) утворюють так звану бот-мережу, через яку здійснюється розсилка спаму; при цьому власники комп'ютерів можуть і не підозрювати про використання останніх в цих операціях спамерів.

Абоненти операторів мобільного зв'язку стають для спамерів набагато більш привабливою аудиторією, ніж навіть користувачі Інтернету завдяки тому, що спамери отримують доступ до Інтернет-сайтів операторів і розсилають через них нелегальні оголошення і SMS-повідомлення.

Відомі такі способи розсилки мобільного спаму:

– злом SMS-шлюзів операторів в Інтернеті;

– розсилка SMS вручну з мобільного телефону, номер якого не визначається;

– повідомлення самих операторів, що «агресивно» рекламують свої послуги.

Найбільш поширені види спаму в даний час:

– реклама (в тому числі реклама незаконної продукції);

– так звані «Нігерійські листи» для виманювання різними способами грошових коштів;

– фішинг – виманювання в одержувача номера його кредитних карток та паролів доступу до систем онлайнових платежів;

– розсилка листів релігійного змісту; масова розсилка листів для виведення з ладу поштової системи;

– розсилка листів, що містять комп'ютерні віруси або жалісну історію про необхідність надання термінової допомоги та деякі інші.

Для боротьби з тим чи іншим явищем, спочатку потрібно розібратися в тому, що це таке і як працює. Знаючи механізми роботи спаму на мобільних мережах,

можна пропонувати відповідні засоби для боротьби з ним. Отже, почнемо з сценаріїв спаму та існуючих рекламних розсилок.

Рекламна розсилка оператора – абонент забув відмовитися отримувати рекламні повідомлення. Оператор, як правило, висилає три-чотири повідомлення на місяць. Це воля абонента відмовитися від їх отримання.

При оплаті послуг зв'язку в терміналах, часто приходять повідомлення про поповнення рахунків з рекламою, це також варіант реклами, на яку погоджується абонент.

Цей варіант реклами як правило не дратує людей, так як міститься в повідомленні про поповнення рахунку. Але законодавчо можна обмежити такі розсилки.

Спам в чистому вигляді виглядає наступним образом: хтось невідомий, з незнайомого номера надсилає абоненту повідомлення, – «Створюємо сайти швидко і недорого. Пишіть на xxxxxx ». Варіантів такого спаму може існувати безліч, але технічно відрізняються способи його розсилки, можна привести два самих поширених.

Непрофесійні спамери: кампанія людей, що відкрила «Рекламне агентство», купує SIM-карти з яких комп'ютер відправляє рекламні SMS. Вартість SMS незрівнянно низька з вартістю «білої» реклами. В інтернеті можна знайти безліч варіантів таких послуг, вони коштують недорого, відправка від 10.000 SMS і більше. В умовах договору з абонентами зазначено, що SIM-карти не можуть використовуватися для бізнес-цілей, тому, як правило, оператори швидко блокують такі сімки. Зазвичай вдається відправити від 500 до 3000 SMS, в залежності від оператора. Потім SIM-карту блокують, але наявність величезної кількості SIM-карт, які можна оформляти хоч щогодини, роблять такий спосіб простим і доступним.

Професійні спамери не використовують такі методи, так як їх завдання отримати доступ до мільйонів номерів і робити розсилку щохвилини і по дуже

великій базі. Вони використовують професійне програмне забезпечення, збирають свої бази, а також можуть використовувати автоматичний перебір номерних ємностей, щоб визначати «живі» номера. Цей процес підтримує базу в актуальному стані. Професійні спамери зазвичай перебувають у країнах Африки, мають у своєму розпорядженні технічні засоби та купують трафік (SMS-повідомлення) великим оптом, щоб знизити вартість контакту. Якщо на 10 відправлених SMS хоча б одна дійшла до адресата, то вони вже знаходяться в «плюсі». Оскільки відправлення SMS повідомлень автоматизовано, це поліпшує використання SMS-шлюзів, які дозволяють змінювати службові заголовки SMS. Наприклад, писати замість номера що завгодно. У користувача може створитися враження, що це SMS від реального сервісу або людини в записній книжці. Шахраї активно використовують SMS-шлюзи та вибирають нейтральні адресати: «мати» або «батько».

У мережах українських операторів близько 90 відсотків спаму розсилається професійно, тобто приходить з закордонних мереж. Часто такий спам не має однакових заголовків, що ускладнює його фільтрацію.

На сьогоднішній день спамери використовують розсилку в різних месенджерах. Досить лише набрати в пошуковому рядку «Viber спам», «WhatsApp спам» або «Telegram спам» та отримати безліч засобів автоматичної розсилки [2].

1.2 Існуючі методи виявлення спаму в SMS розсилках

Можна виділити наступні способи боротьби зі SMS-спамом для мобільних пристроїв:

– програмно-апаратний;

– звернення до розповсюджувача реклами;

– звернення до оператора зв'язку.

З усіх перерахованих вище способів для небажаних повідомлень в месенджерах підходить тільки програмно-апаратний. Звертатися до операторів зв'язку не має сенсу, так як трафік в таких месенджерах передається по виділеним шифрованим каналам. В даному випадку можна тільки поскаржитися в службу підтримки месенджера з скаргою на небажане повідомлення. Звернення до розповсюджувача реклами теж не принесе бажаного результату [2].

Якщо мобільний пристрій підтримує установку сторонніх додатків, то існує можливість встановити програму, яка буде фільтрувати SMS за певними ознаками, відсікаючи повідомлення, які схожі на спам.

Для операційних систем мобільних пристроїв створено безліч відповідних додатків. Але при цьому треба мати на увазі, що більшість додатків мають досить обмежений функціонал, і сучасні механізми фільтрації, які застосовуються в тій же електронній пошті, вони не використовують.

Абонент може спробувати зателефонувати розповсюджувачу реклами та попросити його припинити розсилку повідомлень на його адресу. В більшості випадків цей варіант спрацює, тому абонента запитають номер телефону і виключать його з розсилки.

Але в тих випадках, коли спам з рекламою приходить по кілька разів на дню цей варіант не підходить, тому що абонент ризикує витрачати весь свій час на такі дзвінки. При цьому абонент повинен дзвонити в рекламні компанії і повідомляти свій номер телефону для виключення його з якоїсь «бази даних» замість того, щоб перш ніж направляти SMS з рекламою у нього запитували про це дозвіл.

Навряд чи такий спосіб можна назвати скільки-небудь ефективним, крім того він не підійде для повідомлень, які розповсюджуються в месенджерах [2].

Для боротьби з SMS-спамом можна звернутися до оператора мобільного зв'язку. У операторів створені спеціальні номери і адреси електронної пошти для обробки таких претензій:

При дзвінку оператору call-центру абоненту необхідно назвати свої дані, і викласти суть справи: приходять SMS з рекламою, яку він не замовляв. Відразу необхідно сказати, що рекламодавцеві клієнт свого номер не залишав.

1.3 Аналіз існуючих програмних рішень для виявлення спаму

Більшість з існуючих програм для боротьби зі спамом фільтрують повідомлення, що приходять в поштову скриньку. Це зручно з двох причин. По-перше, за допомогою цих програм можна не перекачувати з сервера непотрібні листи. По-друге, вони дозволяють організувати сортування решти кореспонденції.

Треба відзначити, що можливості сортування повідомлень зараз закладені в більшості поштових програм. Головна відмінність антиспамерських програм від сортування полягає в тому, що в програму закладена база даних поштових адресів спамерських компаній. Це дозволяє в більшості випадків з високою ймовірністю відокремити спам від істинної пошти. У антиспамерських програмах закладена велика кількість зумовлених фільтрів, відсіваючих пошту зі спамерським змістом. Крім того, антиспамерські фільтри зручні для тих, хто користується поштовим клієнтом без можливостей фільтрації.

Антиспам додатки для мобільних пристроїв можна розділити на категорії декількома способами:

За реалізацією:

– антиспам функціонал, реалізований в комплексному антивірусному засобі (як правило, це додатки з багатим функціоналом);

– окремий додаток антиспам, з функціями блокування вхідних дзвінків / SMS;

– антиспам функціонал, реалізований в складі різних месенджерів.

За параметрами, що блокуються:

– за номером телефону;

– за номером телефону і ключовими словами;

– за номером телефону і хешу текстового повідомлення.

Блокування за номером телефону – напевно найпростіший спосіб блокувати спам, однак багато рішень не вміють блокувати спам, який прийшов ні з цифрового номера, а з текстового.

Блокування по хешу текстового повідомлення – цей варіант більш кращий, тому що однотипний спам може надходити з різних номерів.

У даній класифікації відсутній пункт «по контенту», хоча всі актуальні на даний момент месенджери реалізовані у вигляді діалогу, і MMS-повідомлення відображаються в діалозі, відповідно реалізація блокування по контенту – це не життєво необхідний, але бажаний пункт даної класифікації.

За наявністю і розташуванням глобального чорного списку (номерів, ключових слів) для аналізу контенту:

– бази немає (наповнюється користувачем / зберігається на телефоні / НЕ завантажується на сервер);

– база в телефоні;

– база в телефоні і на віддаленому сервері;

Антиспам рішення з розміщенням бази в телефоні – найпростіший варіант, але як правило ці рішення будуються на тому, що користувач сам "створює" базу номерів або ключових слів, тобто сам навчає програму.

Плюсами даного рішення є простота реалізації подібних утиліт, а мінусом є тривалість навчання такої програми.

Рішення ж з базою в телефоні і на віддаленому сервері є найбільш переважними, тому що користувачеві не потрібно навчати програму.

Розглянемо декілька програмних рішень з боротьби з SMS-спаму.

Mobile Security & Antivirus (Avast): В даному додатку антиспам представлений таким чином: При отриманні вхідного повідомлення (SMS) проводиться його сканування на наявність вірусів, ознак "фішінга", шкідливих і неправильно введених url-адрес, так само дозволяє блокувати вхідні повідомлення з невідомих номерів. В ході перевірки вхідних повідомлень на пристрій посилалися повідомлення з різними, в тому числі і неправильними URL, додаток лише написав наступне: "SMS відскановано", відповідно ніяких варіантів, що робити з потенційно небезпечними повідомленнями користувачеві не пропонувалося, що в свою чергу ставить під сумнів ефективність спам-фільтра в даному додатку. Так само є можливість створення профілів фільтрації. Заблоковані повідомлення можна потім подивитися в історії. До чорного списку можна помістити всі приховані номери, а також всі невідомі номера (не з контактів), можливість запровадити не числовий номер відсутній. Також до недоліків можна віднести те, що заявлена обробка неправильно введених URL – не працює. При надходженні SMS з абсолютно різним вмістом (фішингове посилання, невірний URL) антиспам пропускає це, при цьому додаток навіть не пропонує користувачеві відзначити SMS спамом.

ALYac Android. В даному додатку реалізовані чорні списки по номеру і по ключовими словами. Інтерфейс інтуїтивно зрозумілий. Однак, все що може цей додаток - вивести повідомлення про те, що прийшло SMS з номера з чорного списку або містить ключове слово з чорного списку. В цілому, це просте, інтуїтивно зрозуміле рішення, просто реалізоване. Але на відміну від попереднього додатку не може нічого крім блокування за параметрами з чорного списку. Звичайно,

користувач може сам занести все те, що він не хоче бачити, але відсутність, наприклад, бази фішингових сайтів або прикладів реклами – це безперечно недолік.

MobileHeal Pro. В даному додатку реалізовані чорні списки по номеру і по ключовими словами. Але додаток не виводить навіть повідомлення про те, що прийшло SMS з ключовим словом з чорного списку. До переваг цього додатка можна віднести простоту реалізації антиспам частини програми. А до недоліків - відсутність глобального чорного списку.

Dr. WEB. В даному додатку реалізовані чорні списки по номеру і по ключовими словами. Повідомлення містить ключове слово з чорного списку, або що прийшло з номера з чорного списку не влучає у контент провайдер content: // sms / inbox.

Заблоковані повідомлення можна подивитися в меню програми. Також в додатку є додатково реалізований функціонал "Cloud Checker", який блокує доступ до різних потенційно небезпечних ресурсів, відповідно якщо прийде SMS-повідомлення з шкідливим посиланням – Dr.WEB НЕ дасть перейти по ньому. До очевидних переваг цього додатка можна віднести наявність додатково реалізованого функціоналу "Cloud Checker", який здатний перевіряти посилання в повідомленнях.

Kaspersky Internet Security. В даному додатку реалізований чорний список за номером і по ключовими словами. Однак на відміну від інших додатків тут є блокування вхідних SMS від нечислових номерів (що містять літерні символи).

Повідомлення містить ключове слово з чорного списку, або прийшло з номера з чорного списку не влучає у контент-провайдер content://sms/inbox, при цьому не відображається повідомлення, що прийшла SMS і вона заблокована. Заблоковані повідомлення можна подивитися в звітах в меню налаштування. До переваг цього додатка можна віднести простоту реалізації антиспам частини програми. До недоліків – відсутність глобального чорного списку номерів / хеш повідомлень / посилань.

Очевидним лідером з розглянутих засобів є Dr. Web. Однак, варто відзначити, що у жодному з додатків в даній категорії по факту не має глобального чорного списку номерів / хеш повідомлень / посилань, і користувачеві необхідно витрачати час на "навчання" програми. Однак у Dr.Web є додатковий функціонал "Cloud Checker", який є чимось на зразок чорного списку шкідливих посилань. Хоча це і не повноцінний глобальний чорний список, такий функціонал додатково огороджує користувача від потенційно небезпечного вмісту SMS-повідомлень. Відсутність чорного списку хеш повідомлень обумовлено саме відсутністю глобального чорного списку.

Антиспам функціонал, реалізований в окремому додатку. У даній категорії були розглянуті найбільш популярні додатки, мають тільки антиспам функціонал.

Jalapeno Antispam. В даному додатку реалізований чорний список за номером і по хешу повідомлення. Цей додаток переглядає всі вхідні SMS-повідомлення. Небажані SMS-повідомлення у вхідних необхідно позначити спамом, після цього ні заблоковане повідомлення, ні будь-які інші від цього користувача більше не будуть приходити. Так само в Jalapeno Antispam реалізований механізм, який поміщає повідомлення і його відправника в глобальний чорний список, при цьому період навчання програми скорочується. Здійснюється це наступним чином: поскаржившись на SMS-повідомлення, передається хеш-сума від тексту SMS, по якій неможливо відновити вихідне повідомлення, що забезпечує конфіденційність інформації.

The Call (Anti Spam). В даному додатку реалізований чорний список за номером і по ключовими словами. Додаток більш профільовано на блокування спам-дзвінків.

При першому запуску програми завантажується база даних розміром 4.86Мб.

При всіх перевагах цього додатку інформація про те, як поповнюється база спаму немає, тобто незрозуміло що зберігається і як SMS та номери телефонів

передаються до бази даних. До переваг даного рішення можна віднести завантаження чорного списку спаму при першому запуску додатку. Недоліками даного рішення є не цілком зручний і не дуже зрозумілий інтерфейс, а також довга первісна настройка.

Spam Blocker (AforApps). В даному додатку реалізований чорний список за номером і по ключовими словами. Додаток має інтуїтивно зрозумілий інтерфейс. Перевагами даної програми є простота реалізації. Недоліки: програма має занадто мало критеріїв фільтрації.

Spam Blocker. В даному додатку реалізований чорний список за номером. Простий додаток. Ніяких повідомлень в разі надходження спаму НЕ відображається. Перевагою даного рішення є простота реалізації додатку. Недоліки даного рішення: програма має занадто мало критеріїв фільтрації.

SMS Антиспам online. В даному додатку реалізований чорний список за номером і, по ключовими словами, а також реалізований глобальний чорний список за вмістом.

При на ходженні SMS цифровий відбиток відправника і тіла повідомлення йде в центральну базу даних на сервері. Після перевірки пристрій отримує відповідь - оцінки інших користувачів. Від їх співвідношення спам / не спам робиться висновок про бажаність повідомлення. Якщо про SMS нічого невідомо, то з'являється повідомлення і діалогове вікно. Користувач також може дати свою оцінку SMS-повідомленню: спам або не спам. У цей момент запит відправляється на сервер, і одночасно відправник SMS додається в offline-список заборони чи дозволів. Можна і просто пропустити SMS у вхідні – в цьому випадку ніяких правил створюватися і відправлятися не буде. Для наповнення бази використовуються хеші (SHA-256) вмісту SMS-повідомлень, що забезпечує конфіденційність інформації користувача. При установці не потрібно напрацьовувати свою базу бажаних і небажаних

відправників. Також є можливість блокування не цифрових відправників, блокування всіх відправників, яких немає в списку контактів.

До переваг даного рішення можна віднести: реалізацію глобальної бази спаму, і постійна робота з нею, настройка критеріїв визначення спаму. Недоліками даного рішення є: неповна працездатність на iOS.

Антиспам функціонал, реалізований в складі різних месенджерів. У даній категорії були розглянуті найбільш популярні месенджери, включають в себе антиспам функціонал. Варто відзначити, що переважна більшість месенджерів мають досить бідний антиспам функціонал (наприклад, тільки чорний список номерів). Най головною перевагою даної категорії є те, що вони замінюють собою стандартний месенджер. Загальним недоліком є те, що ні в одному з них не реалізований глобальний чорний список.

GO SMS Pro. Цей додаток є заміною стандартному додатку для прийому і відправки SMS. В даному додатку антиспам реалізований за допомогою чорного списку ключових слів і чорного / білого списку за номером. Присутня можливість додати номери в чорний список або додати ключові слова, по яким буде класифікуватися спам. Також можна залишити або відключити оповіщення про наявність SMS-спам. З прийнятих SMS можна також відзначити будь-який повідомлення як спам – номер відправника відправиться в чорний список і далі всі SMS з даного номера будуть блокуватися. В принципі в цьому додатку антиспам функція вторинна, але при цьому вона чудово працює на будь-якої версії.

SMS Blocker Clean Inbox. Цей додаток також є повноцінною заміною стандартному SMS-месенджер, і одна з основних його функцій – блокування спаму. Антиспам реалізований за допомогою створення чорного і білий списку за номером і за ключовими словам. При запуску програма просить вибрати країну і далі в списку входять - спам повідомлення.

В налаштуваннях також дозволяє додавати номери в чорний список, а також вказувати серії номерів (наприклад, всі що починаються на +1800, або всі закінчуються на 120) або дозволяє додавати слова / фрази, за якими повідомлення буде класифікуватися як спам. Точно такі ж способи є і для білого списку. Також присутня функція блокування MMS повідомлень (1 номер безкоштовно, далі – преміум). Також при покупці преміум відкриваються функції в вигляді швидкого видалення спам-повідомлення в 1 клік, автовідповідь на спам повідомлення, блокування невідомих та видалення реклами). Переваги даного рішення: можливість тонкої настройки програми, блокування MMS з застосуванням чорного списку за номерами, простота реалізації антиспам функціонала додатку. Недоліки даного рішення: не реалізований глобальний чорний список.

Postman. Антиспам функціонал в даному месенджері реалізований за допомогою потужного унікального фільтру (Powerful & unique spam detection filter). Відразу ж після установки програми, вона автоматично визначило SMS-спам (однотипні повідомлення з одного номера). Налаштувати можна тільки білий список. Такий підхід не можна назвати гнучким, тому що багато користувачів скаржаться на те, що програма автоматично блокує дописи, а часом пропускає типовий спам. До переваг даного рішення можна віднести: застосування унікального фільтру для визначення спаму. Недоліками даного рішення є: застосування унікального фільтру для визначення спаму, неможливість тонкої настройки антиспам функціоналу; відсутність реалізації глобального чорного списку.

Viber, WhatsApp, Telegram. Фільтрація спаму в даних месенджерах реалізована чорним списком номерів. Варто також відзначити наступне: якщо повідомлення приходить від користувача, якого немає в контактах користувача, для перегляду повідомлення йому доводиться додавати користувача в свій контакт-лист. Такий метод фільтрації не дуже зручний, з іншого боку користувач може ігнорувати повідомлення, що прийшли від абонента не з його контактів.

Актуальність розробки принципово нового механізму захисту вбудованих комунікаційних додатків обумовлена тим, що на сьогоднішній день подібного універсального рішення просто не існує. Існуючі рішення не здатні забезпечувати превентивний захист від небажаних повідомлень, так само вони не використовують технології статистичної фільтрації, які широко і успішно застосовуються для фільтрації спаму в електронній пошті.

В якості власного персоналізованого механізму захисту вбудованих комунікаційних додатків на платформі iOS від спаму, буде реалізований модуль, який розробники комунікаційних додатків зможуть використовувати для фільтрації небажаних повідомлень.

1.4 Постановка задач дослідження

Проведений аналіз показав, що SMS-спам є великою проблемою і розробка підходів для виявлення та блокування спаму є актуальними.

Метою роботи є дослідження методів класифікації SMS-спаму та розробка програмного додатку для фільтрації SMS-спаму на iOS пристроях.

Об‘єктом дослідження є процес виявлення та блокування SMS-спаму.

Предметом дослідження є методи та засоби проектування та розробки програмного забезпечення для блокування SMS-спаму.

Для досягнення поставленої мети необхідно вирішити такі задачі:

– скористатись результатами аналізу предметної галузі та проаналізувати моделі методів класифікації спаму для sms-повідомлень;

– сформувати набір sms-повідомлень українською мовою, що будуть використані при тренуванні машинних моделей;

– розробити машині моделі;

– розробити iOS додаток та інтегрувати машині моделі;

– оцінити точність методів класифікації та порівняти результати.

Машині моделі можуть бути використані як у розробленому iOS додатку, так і у інших програмних рішеннях.

Створений iOS додаток дозволить переглядати вхідні sms-повідомлення та користувач зможе перейти до результатів класифікації описаними методами. Також користувач зможе ввести текст власноруч та подивитись результати класифікації.

Розробленою базою sms-повідомлень українською мовою зможуть скористатись при інших дослідженнях, наприклад для порівняння з іншими методами

Після розробки програмних рішень буде проведена апробація, порівняння та обґрунтування отриманих результатів.

2 РОЗРОБКА ТА ДОСЛІДЖЕННЯ МОДЕЛЕЙ РОЗПІЗНАННЯ SMS СПАМУ

2.1 Опис методів аналізу даних

Найстаріший спосіб аналізу даних – ручний аналіз, що виконується без використання засобів обчислювальної техніки. Цей метод трудомісткий і неприйнятний у випадках, коли необхідно аналізувати з високою швидкістю значну кількість інформації.

Інший підхід полягає в написанні правил і регулярних виразів, за якими можна віднести аналізовану інформацію до тієї чи іншої категорії. Наприклад, одне з таких правил може виглядати наступним чином: «якщо текст містить слова похідна і рівняння, то віднести його до категорії математика». Спеціаліст, знайомий з предметною областю і який володіє навиком написання регулярних виразів, може скласти ряд правил, які потім автоматично застосовуються до вхідних документів для їх класифікації [7]. Цей підхід кращий за попередній, оскільки процес класифікації автоматизується і, отже, кількість оброблюваної інформації практично не обмежена. Однак створення і підтримання правил в актуальному стані вимагає постійних зусиль фахівця.

При машинному аналізі інформації набір правил і загальний критерій прийняття рішення текстового класифікатора обчислюється автоматично, навчаючи класифікатор стандартними загальноприйнятими словами, фразами або кількісною оцінкою. При такому підході необхідна ручна розмітка та первісна впорядкованість інформації. Термін розмітка означає присвоєння документу (або окремої інформації) класу, рангу або важливості. Розмітка більш просте завдання, ніж написання правил. Крім того, розмітка може бути проведена в звичайному режимі використання системи.

Наприклад, в програмі електронної пошти може існувати можливість позначати листи як спам [8], тим самим формуючи навчальну множину для класифікатора – фільтра небажаної пошти. Таким чином, класифікація текстів, заснована на машинному навчанні, є прикладом навчання з вчителем, де в ролі вчителя виступає людина, що задає набір класів і розмічає навчальну множину [9]. Розглянемо кілька підходів реалізації методу навчання класифікатора.

2.2 Аналіз моделі байєсівського класифікатора

Головне призначення ймовірнісної моделі – визначення ймовірностей настання деяких подій. Тому в основі імовірнісних моделей лежить теорія ймовірності і використання її базових елементів, таких як теорема Байеса [10]. Основою для імовірнісного методу навчання класифікатора є наївна байєсівська модель. Нехай документи розбиті на кілька класів $c_1,...,c^k$, C – загальна безліч класів. Суть її полягає у тому, що ймовірність того, що документ d потрапить в клас c, записується як $P(c|d)$:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \tag{2.1}$$

де $P(d|c)$ – ймовірність зустріти документ d серед всіх документів класу c,

$P(c)$ – безумовна ймовірність зустріти документ класу c в корпусі документів,

$P(d)$ – безумовна ймовірність документа наявності d в корпусі документів.

Щоб оцінити умовну ймовірність $P(d|c) = P(t_1, t_2, ..., t_n |c)$, де t_k – терм з документа d, n – загальна кількість термів в документі (включаючи повторення),

необхідно ввести припущення про умовну незалежності термів і про незалежність позицій термів.

Іншими словами, ми нехтуємо тим фактом, що в тексті на природній мові поява одного слова часто тісно пов'язане з появою інших слів (найімовірніше, що слово інтеграл зустрінеться в одному тексті зі словом рівняння, ніж зі словом бактерія) і, що ймовірність зустріти одне і те ж слово різна для різних позицій в тексті. Саме через ці спрощень розглянута модель природної мови називається наївною [11].

Таким чином, імовірнісні моделі надають зручні засоби прогнозування настання різних подій.

Оскільки мета класифікації – знайти найкращий клас для даного документа, то завдання наївною байєсівської класифікації полягає в знаходженні найбільш ймовірного класу c_m, який розраховується за формулою [12]:

$$c_m = \underset{c \epsilon C}{\operatorname{argmax}} P(c|d) \tag{2.2}$$

де c – клас,

d – документ,

$argmax$ – елемент, на якому досягається максимум.

Обчислити значення цієї ймовірності безпосередньо неможливо, оскільки для цього потрібно, щоб навчальна множина містила всі (або майже всі) можливі комбінації класів і документів. Однак, використовуючи формулу Байеса, можна переписати вираз для $P\ (c \mid d)$, у вигляді:

$$c_m = \underset{c \epsilon C}{\operatorname{argmax}} \frac{P(d|c)P(c)}{P(d)} = \underset{c \epsilon C}{\operatorname{argmax}} P(d|c)P(c) \tag{2.3}$$

де $P(d|c)$ – ймовірність зустріти документ d серед документів класу c,

$P(c)$ – ймовірність того, що зустрінеться клас c, незалежно від розглянутого документа,

$P(d)$ знаменник опущений, тому що не залежить від c і, отже, не впливає на знаходження максимуму.

Використовуючи навчальну множину, ймовірність $P(c)$ можна оцінити за формулою:

$$P(c) = Nc/N \tag{2.4}$$

де Nc – кількість документів з навчальної множини в класі c,

N – загальна кількість документів в навчальній множині.

Використовуючи правило множення ймовірностей незалежних подій [13], можна записати наступну формулу:

$$P(d|c) = P(t_1, t_2, \dots, t_n|c) = P(t_1|c)\, P(t_2|c) \dots P\,(t_n \mid c) = \prod_{k=1}^{n} P(t_k|c) \tag{2.5}$$

Оцінка ймовірностей $P(t|c)$ за допомогою навчальної множини буде розраховуватися за формулою:

$$P(t|c) = T_{ct}/T_c \tag{2.6}$$

де T_c – загальна кількість термів в документах класу c,

T_{ct} – кількість входжень терма t у всіх документах класу c (на будь-яких позиціях). При підрахунку враховуються всі повторні входження.

Після того, як класифікатор «навчений», тобто знайдені величини $P(t|c)$ і $P(c)$, можна відшукати клас документа за допомогою співвідношення:

$$c_m = \underset{c \epsilon C}{\operatorname{argmax}} P(d|c)P(c) = \underset{c \epsilon C}{\operatorname{argmax}} P(c) \prod_{k=1}^{n} P(t_k|c) \quad (2.7)$$

Щоб уникнути в останній формулі переповнення знизу через велике число малих співмножників, на практиці замість добутку зазвичай використовують суму логарифмів. Логарифмування не впливає на знаходження максимуму, так як логарифм є монотонно зростаючою функцією. Тому в більшості реалізацій використовують формулу:

$$c_m = \underset{c \epsilon C}{\operatorname{argmax}} \left[logP(c) + \sum_{k=1}^{n} logP(t_k|c) \right] \quad (2.8)$$

Ця формула має просту інтерпретацію. Шанси класифікувати документ часто зустрічається класом вище, і доданок $\log P(c)$ вносить в загальну суму відповідний внесок. Величини $\log P(t_k|c)$ тим більше, ніж важливіше терм t для ідентифікації класу c, і, відповідно, тим вагомішим їх внесок в загальну суму [14].

2.3 Аналіз моделі ЛСА

Метод латентно-семантичного аналізу (ЛСА) дозволяє виявляти значення слів з урахуванням контексту їх використання шляхом обробки великого обсягу текстів.

Модель представлення тексту, що використовується в латентно-семантичному аналізі, багато в чому схожа з сприйняттям тексту людиною. Наприклад, за допомогою цього методу можна оцінити текст на відповідність заданій темі [15].

В якості вихідної інформації використовується терм-документна матриця.

Терм-документна матриця – це математична матриця, що описує частоту термінів, які зустрічаються в колекції документів.

Рядки відповідають документам в колекції, а стовпці відповідають термінам. До матриці застосовується сингулярне розкладання.

Сингулярне розкладання – це математична операція, що розкладає матрицю на 3 складових. Сингулярне розкладання можна представити у вигляді формули:

$$A=USV^T \tag{2.9}$$

де A – вихідна матриця,

U і V^T– ортогональні матриці,

S – діагональна матриця, значення на діагоналі якої називаються сингулярними коефіцієнтами матриці.

Сингулярне розкладання дозволяє виділити ключові складові вихідної матриці. Основна ідея ЛСА полягає в тому, що якщо в якості матриці використовувалася терм-документна матриця, то матриця, яка містить тільки перші лінійно незалежні компоненти, відображає основну структуру різних залежностей, присутніх у вихідній матриці. Структура залежностей визначається ваговими функціями термів [16].

2.4 Аналіз моделі SVM

Метод SVM будує гіперплощину (або набір гіперплощин) у просторі високої або нескінченної вимірності, що може бути використовувана для завдань класифікації.

Дан набір з n точок вигляду $\overrightarrow{x_1}, y_1, \ldots \overrightarrow{x_n}, y_n$ де y_i є або 1, або -1, і кожен з них вказує клас, до якого належить точка $\overrightarrow{x_i}$. Кожен x_i є p-вимірним дійсним вектором. Треба знайти максимально розділову гіперплощину, яка відділяє групу точок $\overrightarrow{x_i}$ для яких $y_i = 1$ від групи точок, для яких $y_i = -1$, і визначається таким чином, що відстань між цією гіперплощиною та найближчою точкою $\overrightarrow{x_i}$ з кожної з груп є максимальною.

Будь-яка гіперплощина може бути записана як множина точок $\overrightarrow{x_i}$, що задовольняють $\vec{w} * \vec{x} - b = 0$, де $\vec{w}$ є вектором нормалі до цієї гіперплощини.

Якщо тренувальні дані є лінійно роздільними, то ми можемо обрати дві паралельні гіперплощини, які розділяють два класи даних таким чином, що відстань між ними є якомога більшою. Область, обмежена цими двома гіперплощинами, називається *розділенням*, а максимально розділова гіперплощина є гіперплощиною, яка лежить посередині між цими двома. Ці гіперплощини може бути описано рівняннями $\vec{w} * \vec{x} - b = 1$ та $\vec{w} * \vec{x} - b = -1$.

Це можна сформулювати у задачу оптимізації, де потрібно мінімізувати $\|\vec{w}\|$ за умови $y_i(\vec{w} * \overrightarrow{x_i} - b) \geq 1$ для i = 1, … n. Отже, $\vec{w}$ та b, що розв'язують задачу, визначають класифікатор $\vec{x} \to sgn(\vec{w} * \vec{x} - b)$, а максимальна розділова гіперплощина повністю визначається $\overrightarrow{x_i}$ які лежать найближче до неї. Ці $\overrightarrow{x_i}$ називають опорними векторами.

2.5 Аналіз моделі з CoreML та MLTextClassifier

Текстовий класифікатор MLTextClassifier з пакету Apple CoreML використовується для підготовки моделі машинного навчання, яку можна включити у iOS додаток, щоб класифікувати текстові данні. Модель вчиться пов'язувати мітки з особливостями вхідного тексту, якими можуть бути речення, абзаци чи навіть цілі документи.

Після підготовки класифікатор тексту зберігається як модель CoreML до файлу, щоб використовувати разом з класом NLModel пакета Natural Language.

Достовірно неможливо описати точний алгоритм класифікації, оскільки цей інструмент є приватним, але можливо вказати безліч параметрів тренування та класифікації. До таких параметрів відносяться алгоритм тренування та мова.

Мова вказується одна з 56 доступних або універсальна.

Доступні тренувальні алгоритми це метод максимальної ентропії (MaxEnt) чи метод умовного випадкового поля (CRF).

2.5.1 Аналіз методу максимальної ентропії

Метод максимальної ентропії знаходить залежності між кожним елементом вхідних даних, що суттєво відрізняє його від моделі наївного байєсівського класифікатора.

Ймовірність елементу даних x відноситись до мітки y визначається за формулою 2.10.

$$P(y|x) = \frac{1}{Z(x)} \exp(\sum_{i=1}^{m} w_i f_i(x, y)) \quad (2.10)$$

де f_i – функція залежності, що враховує взаємозв'язки між даними та міткою,

w_i – вектор ваг,

x – вектор даних,

y – вектор міток,

m – кількість елементів,

$\sum_{i=1}^{m} w_i f_i(x, y)$ – підсумовування всіх результатів функцій залежностей,

$Z(x)$ – функція, що допомагає нормалізувати ймовірність, та визначається як:

$$Z(x) = \sum_{y'} \exp(\sum_{i=1}^{m} w_i f_i(x, y)) \quad (2.11)$$

Модель максимальної ентропії використовує такий ж самий підхід як логарифмічно-лінійна модель, але не враховує послідовність вхідного вектора даних.

2.5.2 Аналіз методу умовного випадкового поля

Метод умовного випадкового поля (CRF) моделює залежність між кожним елементом та всіма вхідними послідовностями. На відміну від метода максимальної ентропії, CRF долає проблему *label bias problem* (ситуація, коли перевага отримує елемент з меншою кількістю переходів) за допомогою глобального нормалізатора, а також розглядає усю послідовність міток замість однієї.

$$P(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{j=1}^{n} \sum_{i=1}^{m} w_i f_i(y_{j-1}, y_j, x, j)\right) \quad (2.12)$$

де f_i – функція залежності, що враховує взаємозв'язки між даними та міткою,

w_i – вектор ваг,

x – вектор даних,

y – вектор міток,

n – кількість елементів,

m – кількість міток,

$\sum_{i=1}^{m} w_i f_i(y_{j-1}, y_j, x, j)$ – підсумовування всіх результатів функцій залежностей,

$Z(x)$ – *глобальний* нормалізатор, що враховує суму всіх можливих послідовностей мітки y та визначається як:

$$Z(x) = \sum_{y \in Y} \exp\left(\sum_{j=1}^{n} \sum_{i=1}^{m} w_i f_i\left(y_{j-1}, y_j, x, j\right)\right) \quad (2.13)$$

де f_i – функція залежності, що враховує взаємозв'язки між даними та міткою,

$\sum_{i=1}^{m} w_i f_i(y_{j-1}, y_j, x, j)$ – сума всіх можливих послідовностей міток.

Недоліком методу умовного випадкового поля є обчислювальна складність аналізу навчальної вибірки, що ускладнює постійне оновлення моделі до нових навчальних даних.

2.6 Реалізація машинних моделей

Для програмної реалізації машинних моделей була обрана мова програмування Python та пакет інструментів scikit-learn, за допомогою якого реалізовані методи наївного байєсівського класифікатора, максимальної ентропії та умовного випадкового поля у якості машинних моделей.

Основний спосіб застосування scikit-learn – створення та тренування класифікатора, а також обробку та токенізацію тексту датасета.

Токенізація – це поділ текстового матеріалу на невеликі частини – токени. До токенів відносяться слова та знаки пунктуації. Після чого необхідно представити текст у вигляді масиву значущих слів [17]. Тоді необхідно провести чистку на предмет знаків пунктуації та не значимих слів (наприклад, прийменників). Це робиться за допомогою передачі бібліотеці списку стоп-слів, які автоматично виключаються з розгляду, що істотно підвищує продуктивність методів [18].

Аналіз текстів може здійснюватися як без словника (з урахуванням будь-яких слів зустрінутих в тексті), так і зі словником (з урахуванням тільки слів, присутньому в сконфігурованому словнику). Крім застосування словників, при аналізі текстів не використовуються граматичні особливості тої чи іншої людської мови. Виняток граматичного аналізу дозволяє забезпечити високу швидкість класифікації при великому обсязі вхідних даних. Однак відсутність граматичного аналізу не дозволяє використовувати певні чинники, пов'язані з синтаксисом і морфологією текстів українською або іншими мовами. У поточній версії системи в якості особливостей виступають слова.

Для навчання класифікатора був підготовлен тренувальний набір даних, що включає sms-повідомлення отримані реальними користувачами в Україні. Обсяг складає більше 20000 екземплярів.

2.7 Порівняння моделей

Після програмної реалізації моделей класифікаторів спаму, для кожного з них була проведена оцінка точності розпізнання.

Точність розпізнання визначається як кількість правильних відповідей поділене на загальну кількість питань.

Таблиця 1 – Порівняння розроблених класифікаторів за точністю

Назва класифікатора	Точність
Наївний байєсівський класифікатор	89%
SVM	93%
CoreML + метод максимальної ентропії (MaxEntropy)	94%
CoreML + метод умовного випадкового поля (CRF)	96%

Наївний байєсівський класифікатор з точністю розпізнання 89% показав досить непоганий результат, як на метод, котрий нехтує тим фактом, що в тексті поява одного слова часто тісно пов'язане з появою інших слів.

Інші методи показують високу та досить схожу точність розпізнання – 93-96%. Це пов’язано с тим, що невідмінно від інших текстів, SMS – це зазвичай коротке повідомлення, де кожне слово може суттєво впливати на результат передбачення. Це спостереження також підтверджують методи, що можуть аналізувати взаємозв’язки між словами. Також варто зазначити, що усі методи проводили тренування на досить великому наборі даних (понад 20000 sms-повідомлень), що дозволяє методам CRF та MaxEntropy показувати кращі результати. У той же час, метод наївного байєсівського класифікатора та SVM починають показувати зазначені результати на менших обсягах тренувальних даних (починаючи з 500 повідомлень).

3 РОЗРОБКА IOS ДОДАТКУ ДЛЯ ВИЯВЛЕННЯ СПАМУ В SMS

3.1 Обґрунтування вибору технологій

Для реалізації iOS додатку була використана мова програмування Swift 5.2, середовище розробки Xcode 11.4, iMessage Filter Extension та бібліотека CoreML, оскільки усі перераховані технології є єдиними сучасними та офіційним інструментами, що надаються компанією Apple для розробки iOS-додатків з машинним навчанням.

3.2 Діаграма прецедентів

Для iOS додатку була розроблена діаграма прецедентів (рисунок 3.1).

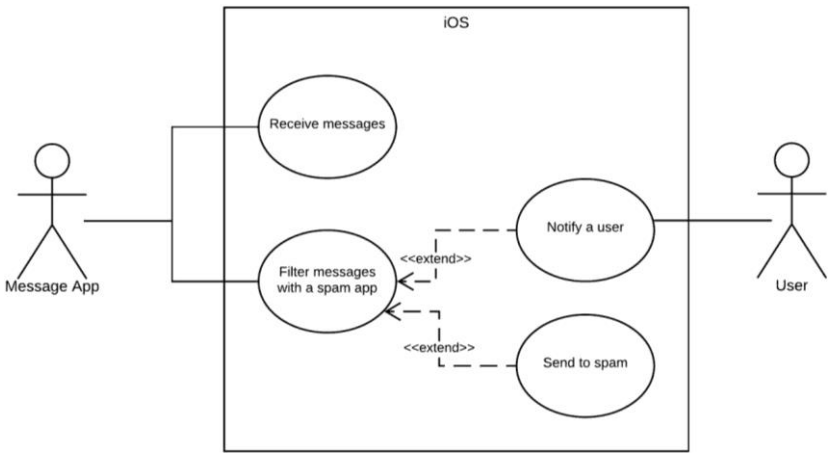

Рисунок 3.1 – Діаграма прецедентів

Це дозволяє продемонструвати як користувач взаємодіє з додатком.

На діаграмі представлено 2 актори: додаток «повідомлення» (message app) та користувача (user). Message app має 2 прецедента: отримання повідомлень та фільтрація повідомлень. Якщо повідомлення є спамом – воно буде направлено до папки спам, якщо ні – нотифікація буде відправлена до користувача, щоб привернути увагу.

3.3 Інструкція користувача

Інструкція користувача описує функціональні можливості та інтерфейс користувача. На рисунку 3.2 відображений інтерфейс додатку.

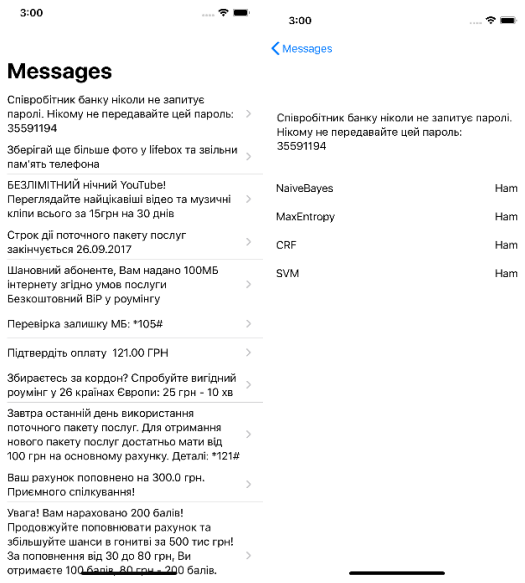

Рисунок 3.2 – Інтерфейс iOS додатку

Список повідомлень (зліва) відображає поточні повідомлення. При натисканні комірки відкривається екран з цим повідомленням та результатами класифікації, де

Spam значить, що повідомлення є спамом, а Ham навпаки. У лівому верхньому куті розташована кнопка, щоб повернутися до попереднього екрану.

3.4 Діаграма класів

Діаграма класів відображає основні елементи (класи) розробленого додатку та їх відношення.

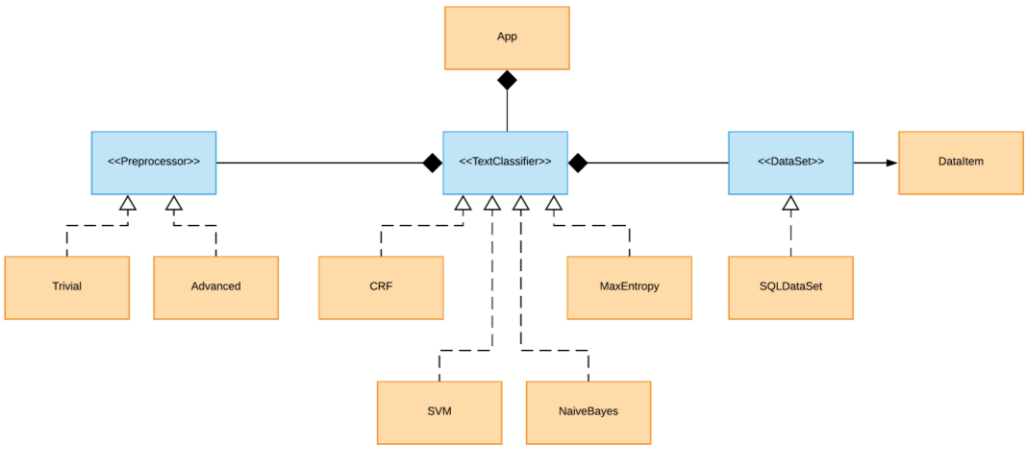

Рисунок 3.3 – Діаграма класів

TextClassifier – це основний протокол, що використовується додатком (App), та задає абстракцію. За ним може скриватися різні конкретні класифікатори. Зокрема, метод умовного випадкового поля (CRF), SVM, наївний байєсівський класифікатор (NaiveBayes), та метод максимальної ентропії (MaxEntropy).

Для роботи текстового класифікатору потрібні данні (DataSet) та препроцессор тексту (Preprocessor). Це теж протоколи, що дозволяють мати декілька реалізацій.

Зокрема простий препроцесор (Trivial), що поділяє текст на слова, та продвинутий (Advanced), що видаляє дати, номера телефонів, email та інші не суттєві данні.

Данні представлені протоколом DataSet, що повертає DataItem для кожного елементу. В поточній реалізації існує тільки DataSet, що загружає дані з SQL сховища (SQLDataSet). Однак структура системи дозволяє розширити джерела даних.

ВИСНОВКИ

На сьогоднішній день частка sms-спаму складає близько 15% від загальної кількості спам-повідомлень та все ще залишається одним з головних та дешевих каналів для масових розсилок. Окрім небажаного рекламного контенту спам може призвести до зараження пристрою, втрати персональних даних та матеріальних збитків [24].

Хоча деякі країни поступово роблять законодавчі обмеження для масових розсилок, загальна кількість повідомлень залишається на зазначеному рівні, що робить актуальним дослідження та розробку сучасних засобів фільтрації спаму.

В атестаційній роботі проведено аналіз предметної області та сформована постановка задач дослідження. Досліджені методи класифікації такі як наївний байєсівський класифікатор, SVM, метод максимальної ентропії та метод умовного випадкового поля. Створена база sms-повідомлень українською мовою. Розроблені машинні моделі за допомогою Python та інструменту scikit-learn. Проведене порівняння розроблених класифікаторів за точністю розпізнання. Розроблено додаток та інтегровані машинні моделі для фільтрації вхідних sms-повідомлень на платформі iOS. Для iOS додатку сформована діаграма класів, діаграма прецедентів та інструкція користувача, а також обґрунтовано вибір використаних технологій.

Усі методи показують досить високу точність розпізнання: від мінімальних 89% для байєсівського класифікатора до максимальних 96% за допомогою методу умовного випадкового поля. Метод SVM та наївний байєсівський класифікатор доцільно використовувати, коли обсяг тренувального набору даних невеликий (до 500 повідомлень). Натомість, методи CRF та MaxEntropy дозволяють отримати точність розпізнання вище, але потребують більшого обсягу тренувальних даних (від 20000 повідомлень).

У якості пропозиції для подальшого дослідження пропонується дослідити розроблені класифікатори з іншими мовами: англійська, російська та транслітерація, що є найпоширенішими мовами sms в Україні.

У додатку пропонується реалізувати функцію накопичування спам повідомлень за певний період та відображення їх у якості випусків, щоб потурбувати користувача лише раз та надати йому змогу відмітити повідомлення як спам чи не спам. Ці відповіді можливо повторно використати, щоб скорегувати ваги машинної моделі. Ця пропозиція дозволить формувати персональні фільтри кожному користувачу.

Оскільки жоден з методів не забезпечує 100% фільтрації спам повідомлень, доцільно аналізувати не тільки текст повідомлення, але й номер телефону з якого надходить sms. Таким чином можливо створити базу даних компаній-спамерів та автоматично її поповнювати номерами спамерів. Для уникнення некоректних скарг, наприклад, коли конкуренти навмисно скаржаться на інших компаній, необхідно розробити систему перевірки даних скарг, а також можливість для компаній видаляти себе з даної бази у разі підтвердження ними неправильного додавання.

Крім того, однією з функцій додатка може стати автоматична заявка про видалення номера абонента з бази даних оператора або позначки про те, що абонент не хоче отримувати ніякі рекламні пропозиції та дзвінки.

ПЕРЕЛІК ДЖЕРЕЛ ТА ПОСИЛАНЬ

1. Yerokhin A., Nechyporenko A., Babii A., Turuta O., Mahdalina I. Usage of phase space diagram to finding significant features of rhinomanometric signals // Computer Science & Information Technologies - CSIT'2016. Lviv, Ukraine: 2016. C. 70 - 72. DOI: 10.1109/STC-CSIT.2016.7589871.

2. Yerokhin A., Nechyporenko A., Babii A., Turuta O. A new intelligence-based approach for rhinomanometric data processing // Proc. of 2016 IEEE 36th International Conference on Electronics and Nanotechnology - ELNANO 2016. : 19-21 April 2016. – C.198-201. DOI: 10.1109/ELNANO.2016.7493047.

3. Yerokhin A., Semenets V., Nechyporenko A., Babii A., Turuta O. F-transform 3D point cloud filtering algorithm // Proc. of the 2th IEEE International Conference on Data Stream Mining & Processing : 21-25 August 2018, Lviv, Ukraine. - C.524-527. DOI: 10.1109/DSMP.2018.8478581.

4. Клименко А. П. Борьба со спамом: история и методы [Електронний ресурс] // МФТИ. URL: https://mipt.ru/dmcp/student/diff_articles/no_spam.php (дата звернення: 22.02.2020).

5. СМС спам – как бороться с рекламными SMS и звонками // Адвокатское Бюро Шмелёва [Електронний ресурс]. URL: http://www.advocatshmelev.narod.ru/sms-spam.html (дата звернення: 15.02.2020).

6. Коновальчук А. Оценка эффективности спама // Trustlink [Електронний ресурс] // URL: https://www.trustlink.ru/subscribe/show/48/ (дата звернення: 15.02.2020).

7. An evaluation of naïve bayesian anti-spam filtering techniques // Utah State University [Електронний ресурс] // URL:

http://digital.cs.usu.edu/~erbacher/publications/Bayes-Vikas2.pdf (дата звернення: 15.02.2020).

8. СМС не надо! // Smsnenado.ru [Електронний ресурс] // URL: https://smsnenado.ru/ (дата звернення:15.02.2020).

9. Методики "зашумления" текста // Forum.antichat.ru [Електронний ресурс] // URL: http://forum.antichat.ru/threads/286921/ (дата звернення: 15.02.2020).

10. Бурлаков М. Е. Применение в задаче классификации SMS сообщений оптимизированного наивного байесовского классификатора // Известия Минского научного центра академии наук. 2016. №. 4-4. С. 18.

11. Павлов А., Добров Б. Обнаружение поискового спама в Вебе на основе анализа разнообразия текстов // Труды Института системного программирования РАН. 2011. Т. 21. С. 277—296.

12. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. // Verlag : Springer, 2013. – 746 с.

13. Байєсьвська фільтрація спаму // Вікіпедія вільна енциклопедія [Електронний ресурс] // URL: https://ru.wikipedia.org/wiki/Байесовская_фильтрация_спама (дата звернення:11.03.2020).

14. Bayesian poisoning // Wikipedia The Free Encyclopedia [Електронний ресурс] // URL: https://en.wikipedia.org/wiki/Bayesian_poisoning (дата звернення 14.04.2020).

15. Bansal S. Comparison between the probabilistic and vector space model for spam filtering // International Journal of Computational Intelligence Techniques, 2012. Т. 3. С. 82.

16. Гмурман В. Е. Теория вероятностей и математическая статистика. — Москва: Высшая школа, 2013. — 479 с.

17. Segaran T. Programming collective intelligence. — LA: O’REILLY, 2012. — 368 с.

18. Пархоменко П. А., Григорьев А. А., Астраханцев Н. А. Обзор и экспериментальное сравнение методов кластеризации текстов // Труды Института системного программирования РАН. 2017. Т. 29. №. 2. С. 34–41.

19. Батура Т. В. Методы автоматической классификации текстов // Программные продукты и системы. 2017. Т. 30. №. 1. С. 121–122.

20. Landauer T. K., Foltz P., Laham D. An introduction to latent semantic analysis discours processes. 2015. Т. 25. С. 259–284.

21. Епрев А. С. Тематическая классификация документов по степени близости термов // Математические структуры и моделирование. 2014. № 20. С. 93–96.

22. Осипова Ю. А., Лавров Д. Н. Применение кластерного анализа методом k-средних для классификации текстов научной направленности // Математические структуры и моделирование. 2017. № 3(43). С. 108–121.

23. Костылев А. В., Лавров Д. Н., Гуц А. К. Идентификация суицидальных групп и нарушителей авторских прав в социальных сетях // Математические структуры и моделирование. 2017. № 3 (43). С. 150–168.

24. Єрьоменко М. О. Аналіз існуючих програмних рішень для виявлення спаму // Інформаційні технології: наука, техніка, технологія, освіта, здоров’я: тези доповідей XXVIII міжнародної науково-практичної конференції MicroCAD-2020. – Харків: НТУ «ХПІ» С. 118.