

Автоматизована перевірка ліцензійної відповідності для macOS

Літвінчук Захар

Науковий керівник - Франків О. О.

Структура macOS застосунку

Основні складові пакету застосунку (Bundle):

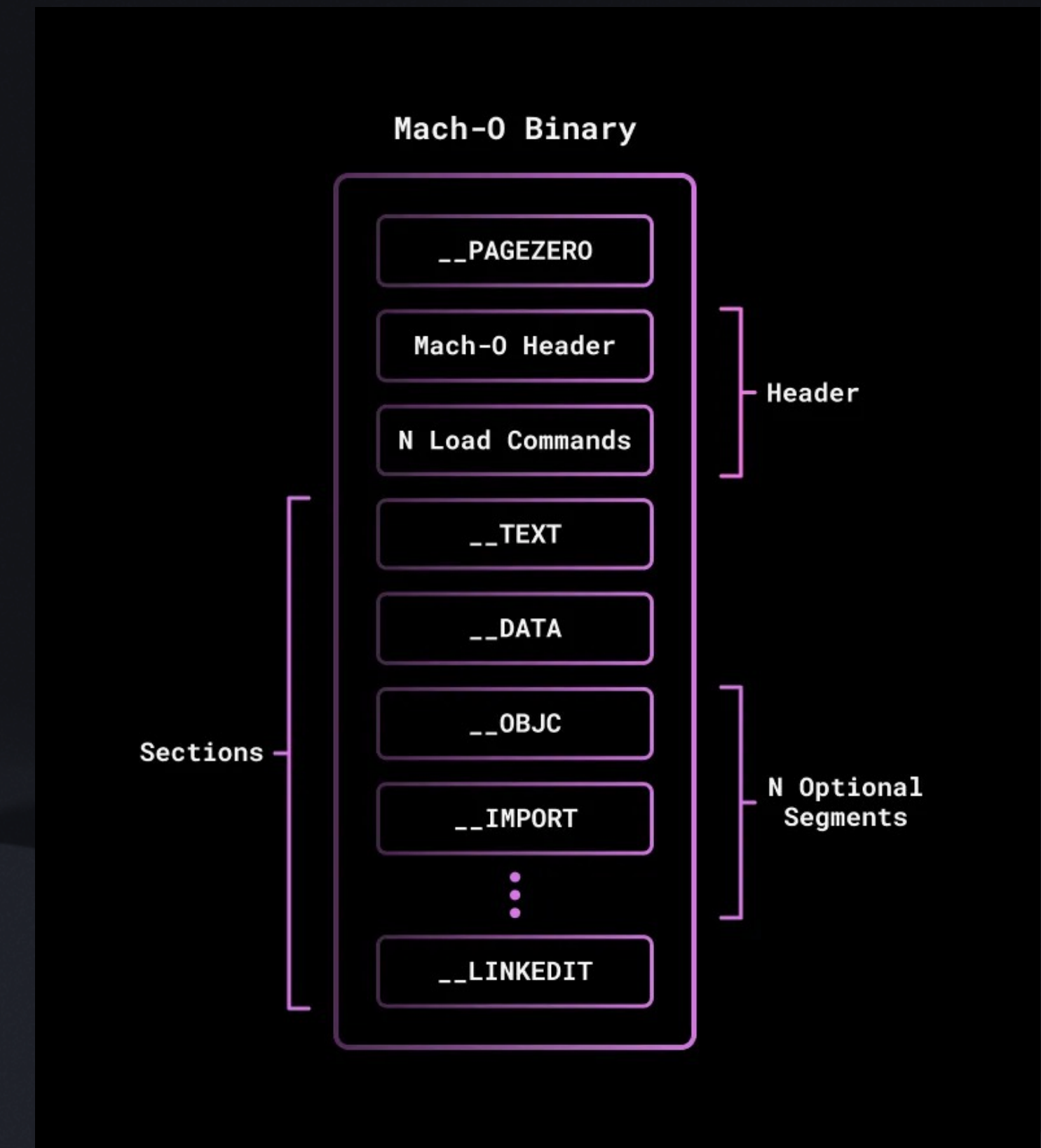
- Виконуваний файл формату Mach-O
- Info.plist
- MAS receipt
- Файли ресурсів
- Динамічні фреймворки та бібліотеки

Основні складові пісочниці застосунку (Sandbox):

- UserDefaults
- Логи
- Файли створені розробниками

Mach-O

- Заголовок з метаданими
- Секції для різних типів коду та даних
- Сегменти, які об'єднують 1 або більше секцій
- Команди завантаження



<https://medium.com/tokopedia-engineering/a-curious-case-of-mach-o-executable-26d5ecadd995>

Способи інтеграції платежів

- StoreKit
- Фреймворки-обгортки для StoreKit



Mac App Store

Способи інтеграції платежів

- StoreKit
- Фреймворки обгортки для StoreKit
- SDK платіжних сервісів
- RestAPI платіжних сервісів
- Платіжні веб-сторінки



MacUpdate

Алгоритм для класифікації



Технічний аналіз

- Аналіз виконуваного файлу:
 - Платіжні фреймворки
 - Класи та функції, які використовуються для здійснення покупок
 - Адреси ендпоінтів та веб-сторінок платіжних сервісів
- Аналіз receipt файлу:
 - Здійснені покупки
 - Активні/неактивні підписки

Мовний аналіз

- Полягає в пошуку ключових слів, що стосуються покупок
- Текст попередньо обробляється за допомогою лематизації та розділення слів (за потреби)
- Текстовий контент складається як з локальних компонентів, так і з інформації з інтернету

Мовний аналіз (локальні компоненти)

Компоненти пакунку програми, що містять текстовий контент:

- Info.plist
- Локалізації (зберігаються в ресурсах)

Компоненти пісочниці, що містять текстовий контент:

- UserDefaults
- Логи

Мовний аналіз (інтернет-контент)

Для всіх застосунків:

- Опис
- Відгуки користувачів

У випадку, якщо застосунок з Mac App Store додатково:

- Політика конфіденційності (Privacy Policy)
- Умови користування (Terms of Use)

Способи збору текстового інтернет-контенту

- iTunesSearchAPI для застосунків з Mac App Store
- Метод web scraping для застосунків з MacUpdate:
 - Статичний контент (опис застосунку)
 - Динамічний контент (відгуки користувачів)

Агрегація результатів аналізу

Технічний аналіз:

- Окремий стовпець з можливими значеннями 0 або 1 для всіх платіжних фреймворків, їх класів, та функцій
- Для платіжних ендпоінтів та покликань на сторінки оплати в стовпець заноситься їх знайдена кількість
- Для receipt файлу використовується кількість записів про здійснені покупки чи підписки

Агрегація результатів аналізу

Мовний аналіз:

- Для кожного слова, пов'язаного з покупками, виділяється окремий стовпець, який вказує на його наявність чи відсутність
- Однакові слова, вжиті в різних місцях, також відокремлюються окремими стовпцями

Тренування класифікатора

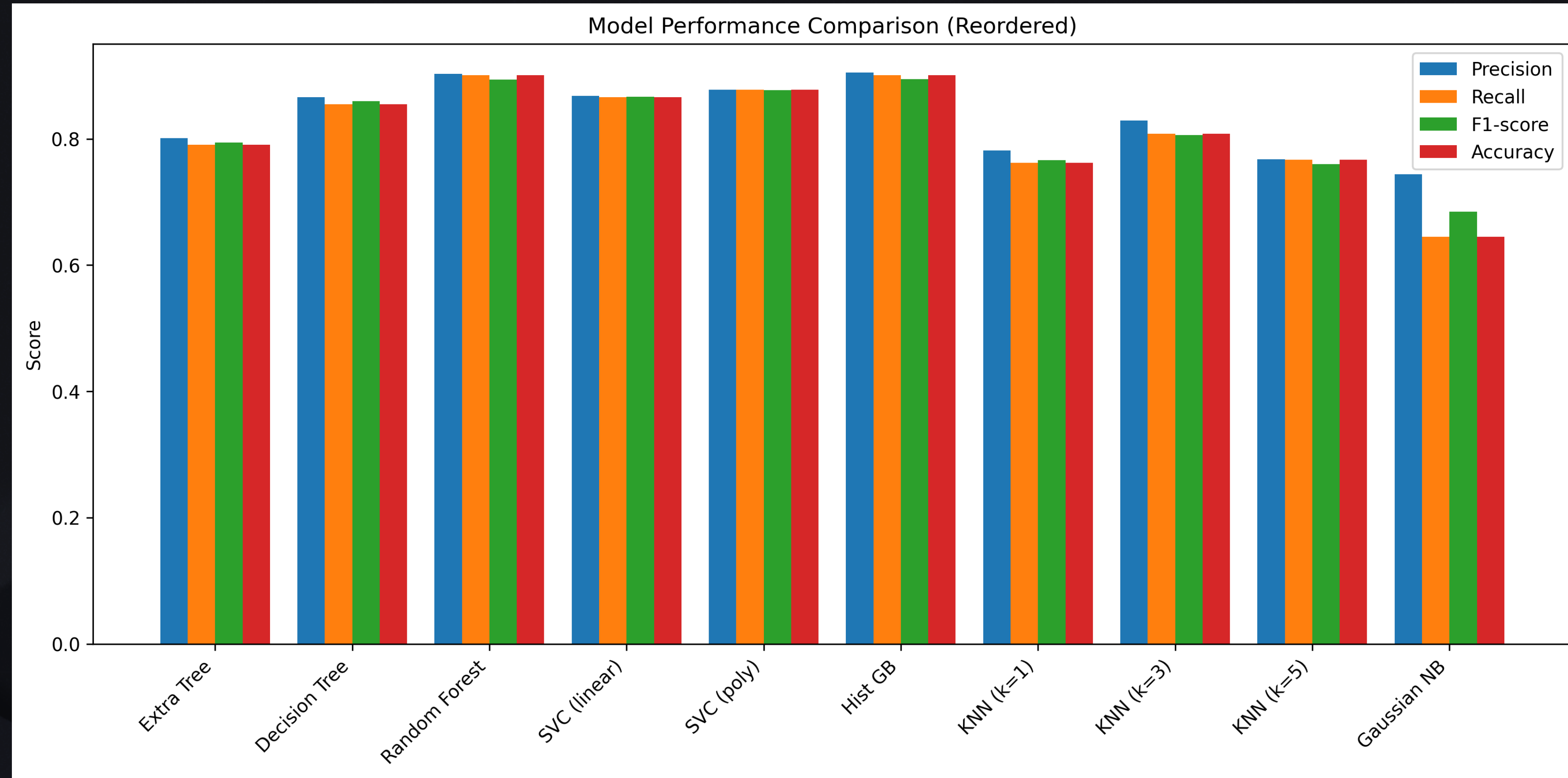
Для тренування використовувались такі алгоритми:

- Decision trees: Decision tree, Extra Randomized Trees та Random Forest
- K-Nearest Neighbor (KNN): $k = 1$, $k = 3$, $k = 5$
- Naive Bayes: Gaussian Naive Bayes
- Support Vector Machines (SVM): Support Vector Classifier with linear and polynomial kernel
- Gradient Boosting: Hist Gradient Boosting Classifier

Порівняння результатів

Метрики для порівняння

- Accuracy
- Precision
- Recall
- F1-Score



Порівняння результатів

Найбільш точними виявились алгоритми Random Forest та Hist Gradient Boosting, вони досягли майже ідентичних результатів за вибраними метриками:

- Random Forest - Accuracy: 0.901, Precision: 0.905, Recall: 0.901, F1-score: 0.895
- Hist Gradient Boosting - Accuracy: 0.901, Precision: 0.903, Recall: 0.901, F1-score: 0.894

Інтерфейс командного рядка

Основні команди:

- analyze з параметрами `—all_data`, `—classification`, `—result_file`
- train з параметрами `—apps_data`, `—classification_data`, `—model_path`
- predict з параметрами `—app_path`, `—url`, `—model_path`, `—percents`

Висновки

- Дослідження всіх компонентів macOS програми
- Виокремлення способів інтеграції покупок в застосунок
- Використання статичного та зовнішнього аналізу для виявлення покупок
- Ознайомлення з алгоритмами машинного навчання з учителем та їх прикладне використання у класифікації
- Розробка інтерфейсу командного рядка

Дякую за увагу!