**Специфікація програмного засобу DomainSense**  
**Версія 1.0**

**1. Загальна інформація**

**Назва продукту:** DomainSense  
**Тип:** Програмний засіб для аналізу доменних імен з використанням методів Big Data та машинного навчання  
**Призначення:** Система призначена для автоматичного аналізу доменних імен з метою виявлення потенційно зловмисних (абузних) доменів на основі WHOIS-даних, DNS-відповідей та структурних характеристик домену.  
**Цільова аудиторія:** Аналітики кібербезпеки, реєстратори доменних імен, автоматизовані системи моніторингу безпеки.

**2. Основні функціональні можливості**

* Прийом користувацького вводу (імені домену)
* Витяг технічної інформації про домен:
  + WHOIS (дата створення, дата закінчення, реєстратор, NS-сервери тощо)
  + DNS-записи типу A, MX, TXT
  + Ознаки наявності SPF, DKIM, DMARC
* Обчислення структурних характеристик домену (довжина, частка голосних, спеціальних символів, кількість дефісів тощо)
* Автоматичне формування ознак для передбачення
* Класифікація домену як **"абузний" / "неабузний"**
* Оцінка впевненості моделі (наприклад, 92% впевненості в тому, що домен зловмисний)
* Пояснення причин оцінки (відсутність MX-запису, занадто довга назва, підозрілі патерни тощо)
* Інтерфейс користувача (веб UI) для зручної взаємодії

**3. Архітектура та технології**

**Frontend:**

* Написаний на TypeScript із використанням React
* Стилізація через Tailwind CSS
* Забезпечує форму для вводу домену та вивід результату

**Backend:**

* Python + FastAPI
* Основна логіка реалізована у вигляді окремих модулів:
  + domain\_info.py — збір технічної інформації
  + domain\_analyzer.py — побудова ознак
  + risk\_evaluator.py — передбачення ризику за допомогою ML
* Модель: RandomForestClassifier, натренована на позначеному датасеті з понад 100 ознаками

**Зберігання моделі:**

* У форматі .pkl через joblib

**4. Вимоги до середовища**

* Python 3.10+
* Пакети: fastapi, scikit-learn, dnspython, whois, joblib, uvicorn, pydantic, numpy, pandas
* Для запуску UI: Node.js 18+, npm, бібліотеки React та Tailwind CSS

**5. Розгортання та запуск**

1. **Frontend:**
   * Встановити залежності: npm install
   * Запустити локально: npm run dev
2. **Backend:**
   * Встановити віртуальне середовище
   * Встановити залежності: pip install -r requirements.txt
   * Запустити API сервер: uvicorn app:app --reload

**6. Перспективи розвитку**

* Додавання підтримки ще більшого обсягу ознак (наприклад, історичні IP-адреси, SSL-сертифікати)
* Використання глибокого навчання для підвищення точності
* Інтеграція з зовнішніми базами даних загроз
* Створення API для інтеграції з реєстраторами або платформами моніторингу безпеки

**7. Джерело та ліцензія**

Програмний код проєкту DomainSense опубліковано у відкритому доступі на GitHub:  
https://github.com/USERNAME/DomainSense *(посилання буде додано пізніше)*  
Ліцензія: MIT

**DomainSense — Software Specification**  
**Version 1.0**

**1. General Overview**

**Product name:** DomainSense  
**Type:** Analytical tool for domain name evaluation using Big Data and Machine Learning  
**Purpose:** The system is designed to automatically analyze domain names and detect potentially malicious (abusive) domains based on WHOIS data, DNS responses, and domain structure.  
**Target users:** Cybersecurity analysts, domain registrars, automated security monitoring systems.

**2. Key Features**

* Accept user input (domain name)
* Extract technical information from the domain:
  + WHOIS (creation/expiration date, registrar, name servers, etc.)
  + DNS records (A, MX, TXT)
  + Indicators for SPF, DKIM, and DMARC presence
* Compute structural features of the domain (length, vowel ratio, hyphen count, presence of special characters, etc.)
* Automatically generate features for ML-based evaluation
* Classify the domain as **"malicious" or "benign"**
* Display confidence score (e.g., 92% confidence the domain is malicious)
* Provide brief explanation of key signals behind the classification
* User-friendly web interface for convenient interaction

**3. Architecture & Technologies**

**Frontend:**

* Built with TypeScript and React
* Styled using Tailwind CSS
* Provides input field for domain names and output of analysis results

**Backend:**

* Implemented using Python + FastAPI
* Core logic is structured into separate modules:
  + domain\_info.py — fetches domain data (WHOIS, DNS)
  + domain\_analyzer.py — constructs feature vector
  + risk\_evaluator.py — applies trained ML model for prediction
* Classification model: RandomForestClassifier trained on a labeled dataset with 100+ features

**Model persistence:**

* Stored in .pkl format via joblib

**4. System Requirements**

* Python 3.10+
* Required Python packages: fastapi, scikit-learn, dnspython, whois, joblib, uvicorn, pydantic, numpy, pandas
* For UI: Node.js 18+, npm, React, Tailwind CSS

**5. Deployment & Usage**

**Frontend:**

1. Install dependencies: npm install
2. Run locally: npm run dev

**Backend:**

1. Set up virtual environment
2. Install dependencies: pip install -r requirements.txt
3. Run the API server: uvicorn app:app --reload

**6. Future Improvements**

* Add support for additional features (e.g., historical IPs, SSL certificate analysis)
* Introduce deep learning for improved classification
* Integrate with external threat intelligence databases
* Provide a public API for registrars and monitoring systems

**7. Repository & License**

The complete source code is available on GitHub:  
https://github.com/USERNAME/DomainSense *(URL will be finalized later)*  
License: MIT