МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет (национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

Разработка веб-приложения для бинарной классификации вредоносных команд по метрике MITRE с использованием алгоритмов машинного обучения

Научный руководитель:
ст. преподаватель кафедры СП
К.Ю. Никольская

Автор:
студент группы КЭ-403
М.Д. Григорьев

Челябинск, 2024 г.

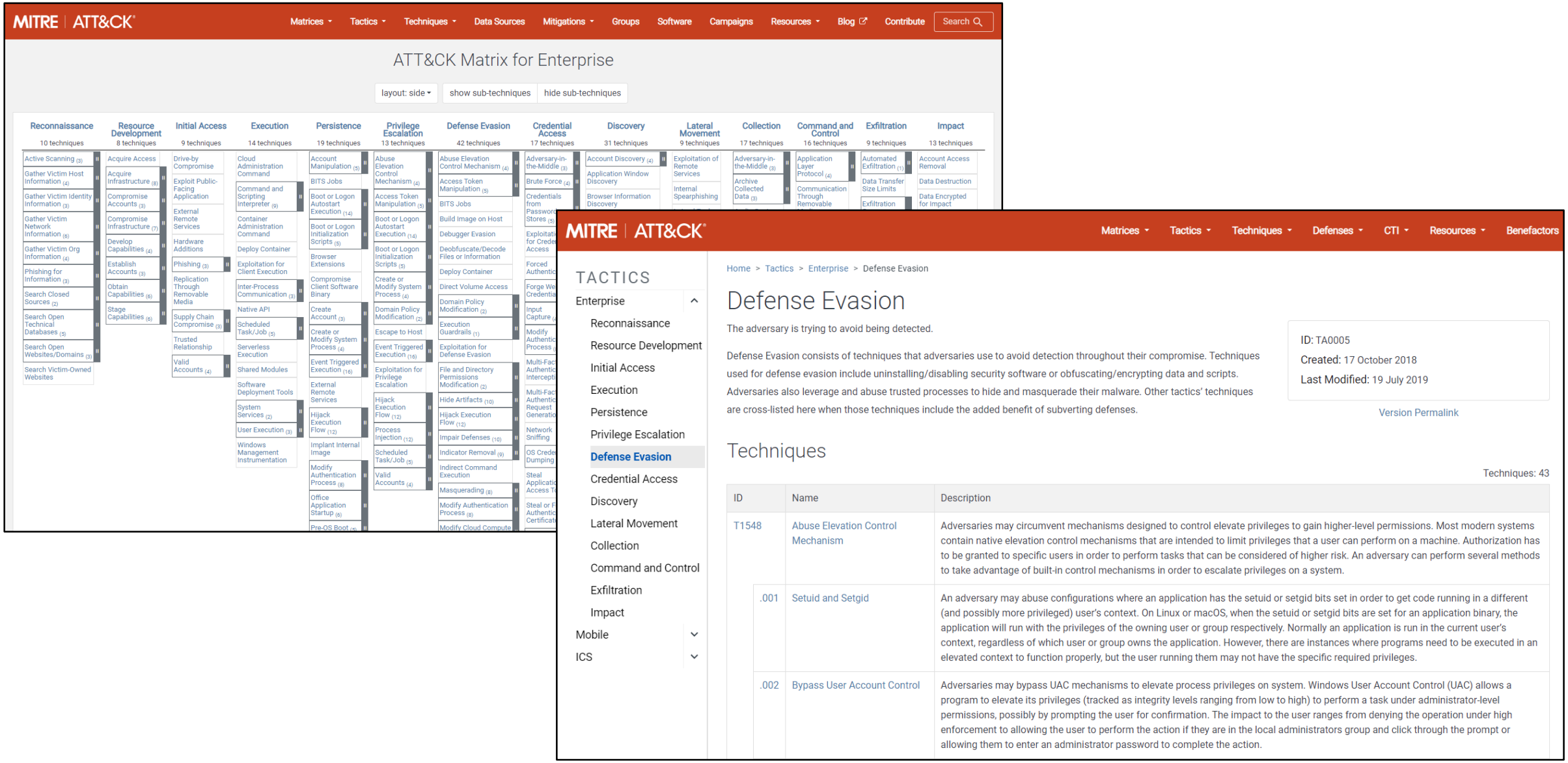
ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Цель: разработать веб-приложение для бинарной классификации вредоносных команд по метрике MITRE с использованием алгоритмов машинного обучения

Задачи:

1. Провести обзор научной литературы
2. Подготовить обучающий набор данных
3. Реализовать выбранные методы машинного обучения
4. Разработать веб-приложение для классификации вредоносных команд по метрике MITRE
5. Провести тестирование разработанного веб-приложения

MITRE ATT&CK



https://attack.mitre.org/

ДИАГРАММА ВАРИАНТОВ ИСПОЛЬЗОВАНИЯ

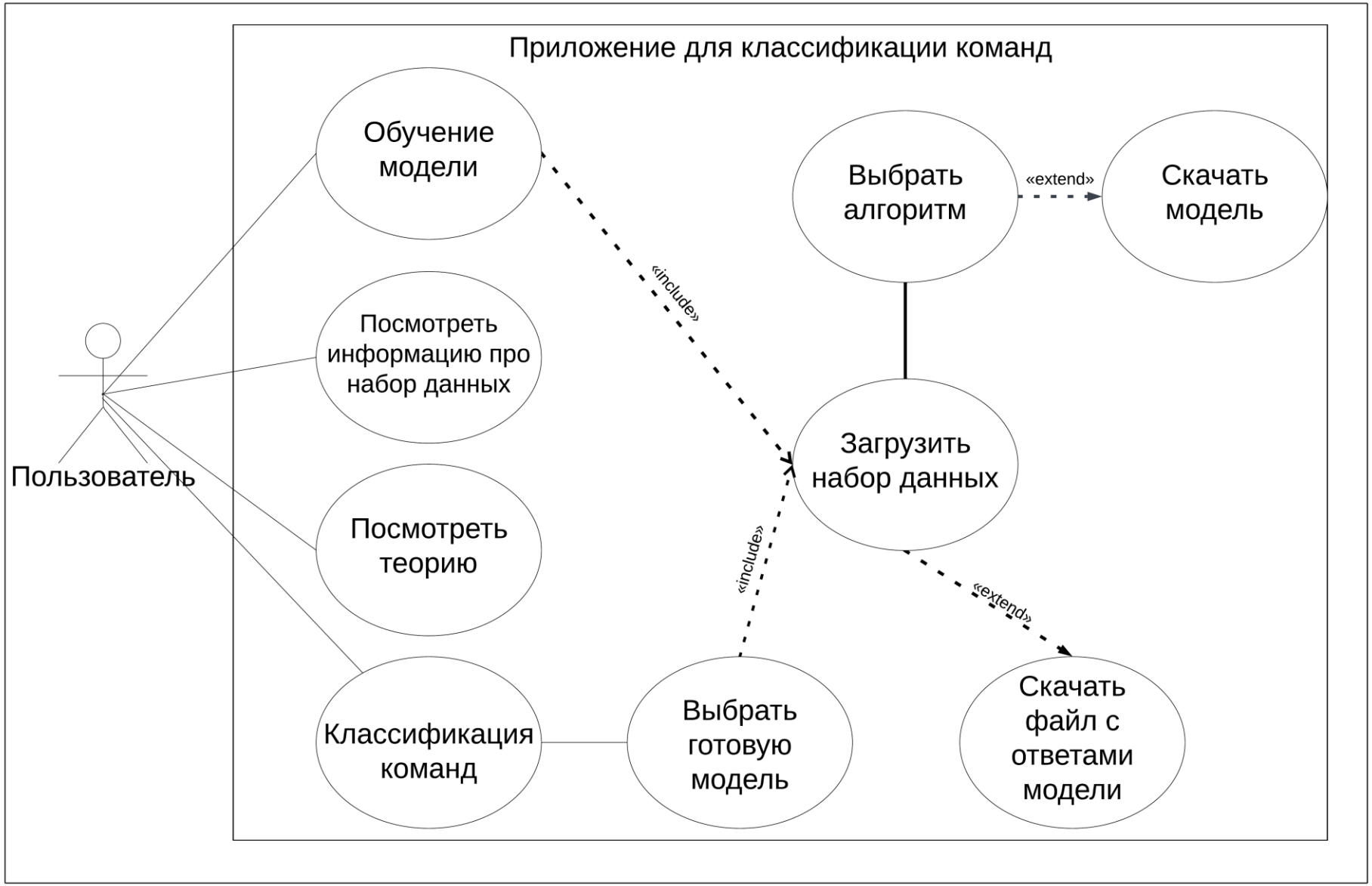
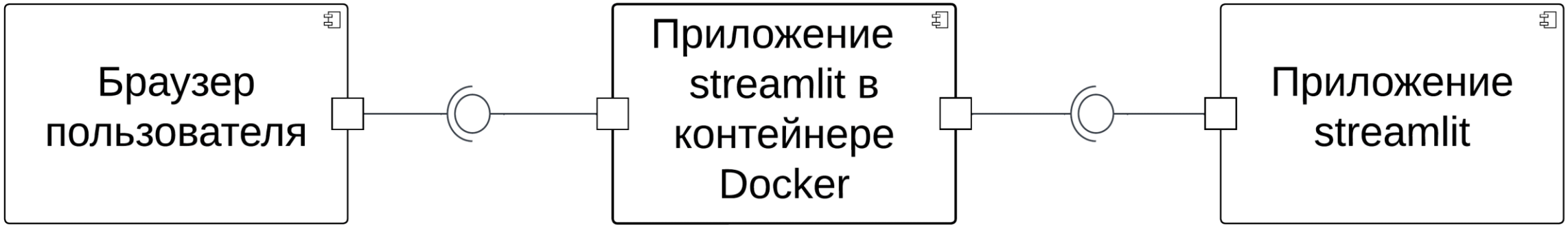


ДИАГРАММА КОМПОНЕНТОВ



СРЕДСТВА РАЗРАБОТКИ

- **Язык программирования**: Python 3.10.6
- **Редактор исходного кода:** VSCode 1.78.2
- **Среда разработки модели машинного обучения:** Jyputer Notebook
- **Библиотеки**: scikit-learn 1.2.2, pandas 1.5.0, NumPy 1.21.6, matplotlib 3.7.1, streamlit 1.35.0

НАБОР ДАННЫХ

- Количество классов команд: 2
- Общее количество записей: 1 742
- Количество видов команд: 14 (Reconnaissance, Resource Development, Initial Access, Execution, Persistence, Privilege Escalation, Defense Evasion, Credential Access, Discovery, Lateral Movement, Collection, Command and Control, Exfiltration, Impact, Safe)

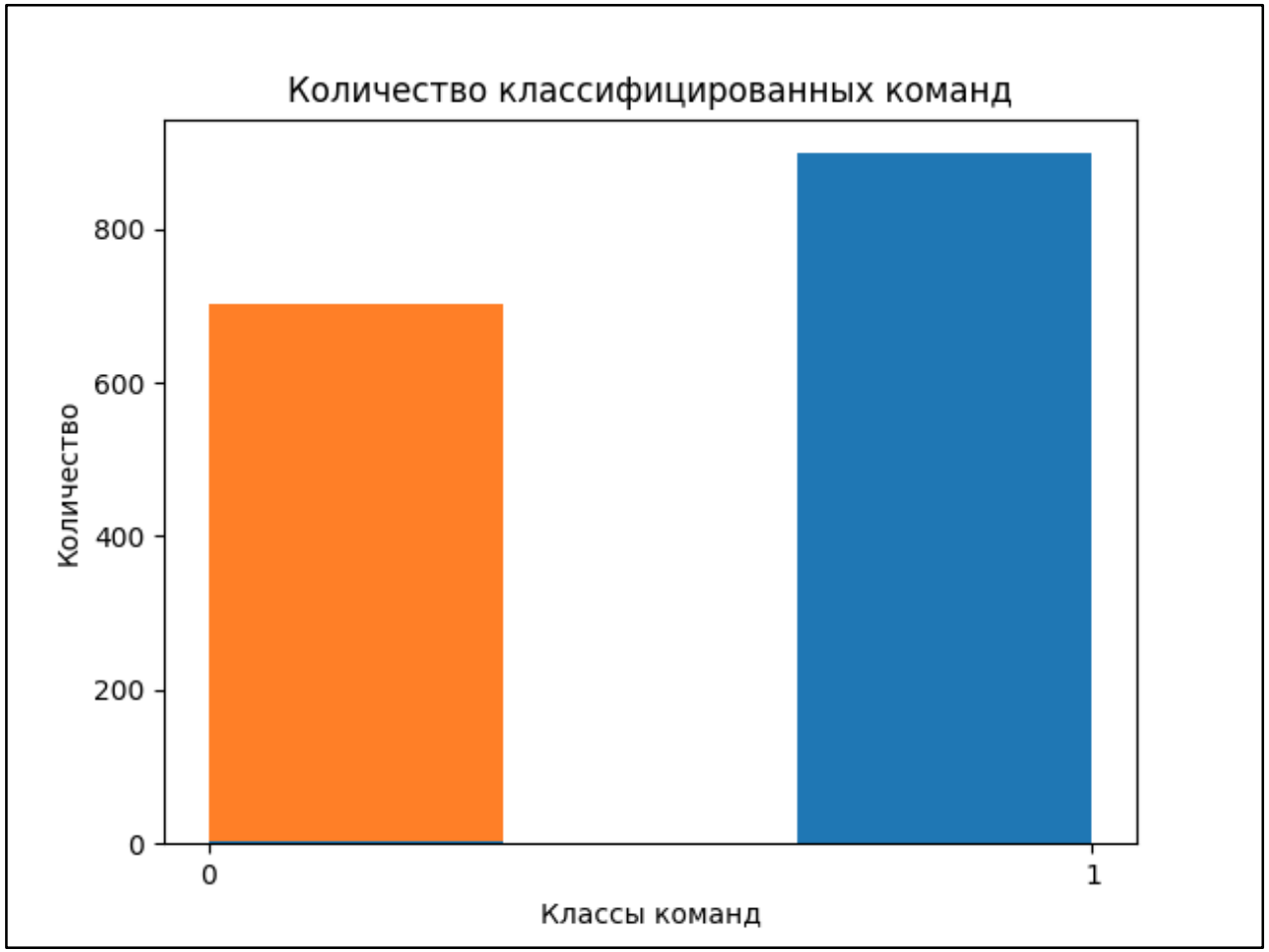
Вид команды	Команды MITRE
Execution	$assembly = [Ref].Assembly.GetType(('{0}{1}i{2}' -f $a,$b,$u))
Command and Control	$field.SetValue($null,$true)
Discovery	$ping = New-Object System.Net.Networkinformation.Ping
Privilege Escalation	$computer = "<hostname>"
Credential Access	$cred = New-Object System.management.Automation.PSCredential($user, $pass)

ПРЕДОБРАБОТКА

1. Разметка набора данных
2. Чистка данных

Итог:

- Количество записей: 1 600



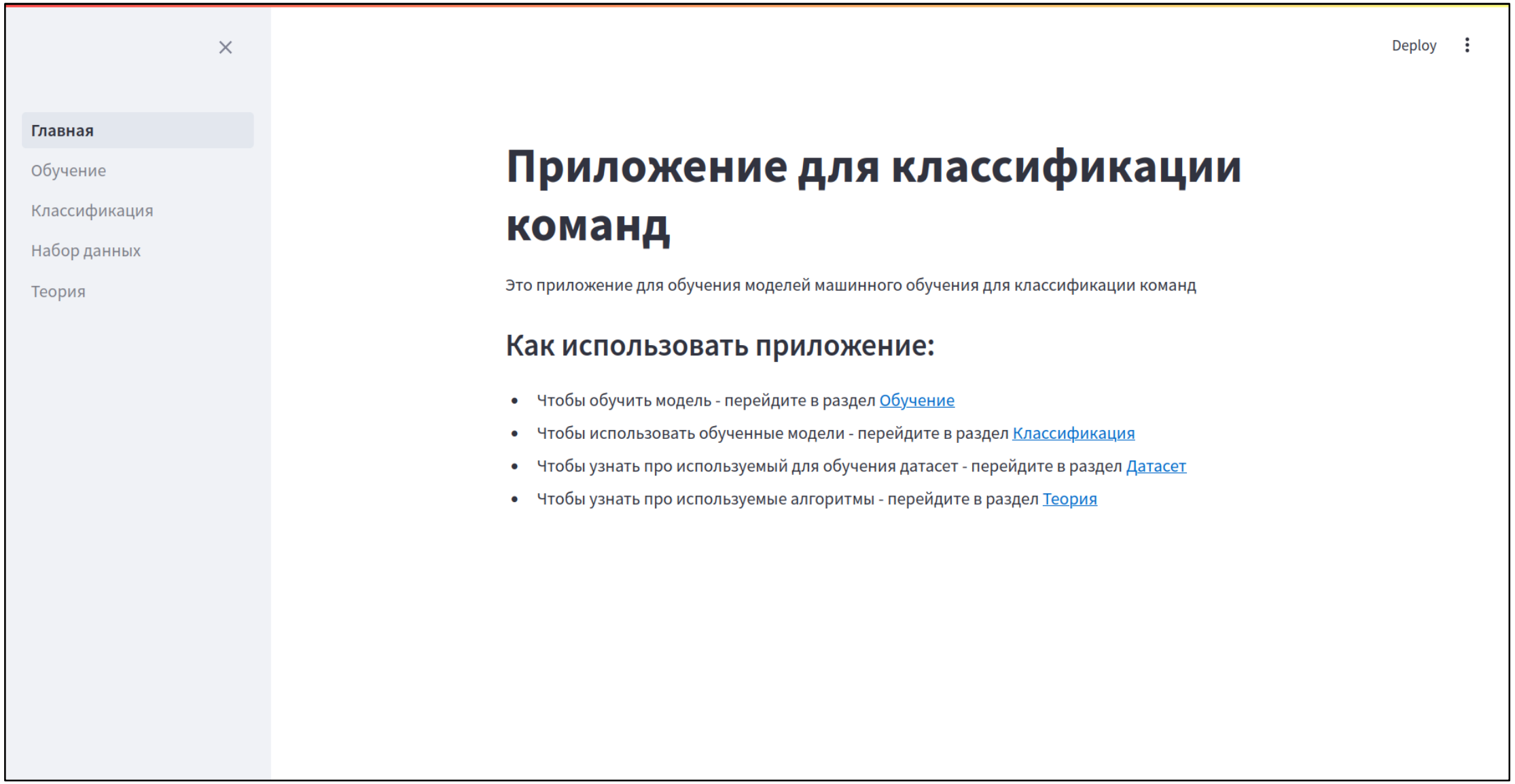
АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

1. Наивный Байес (Naïve Bayes)
2. Логистическая Регрессия (Logistic Regression)
3. Дерево Решений (Decision Tree)
4. Случайный Лес (Random Forest)
5. Бустинг (Boosting)
6. Метод опорных векторов (Support Vector Machines)
7. K-ближайших соседей (K-Nearest Neighbors)

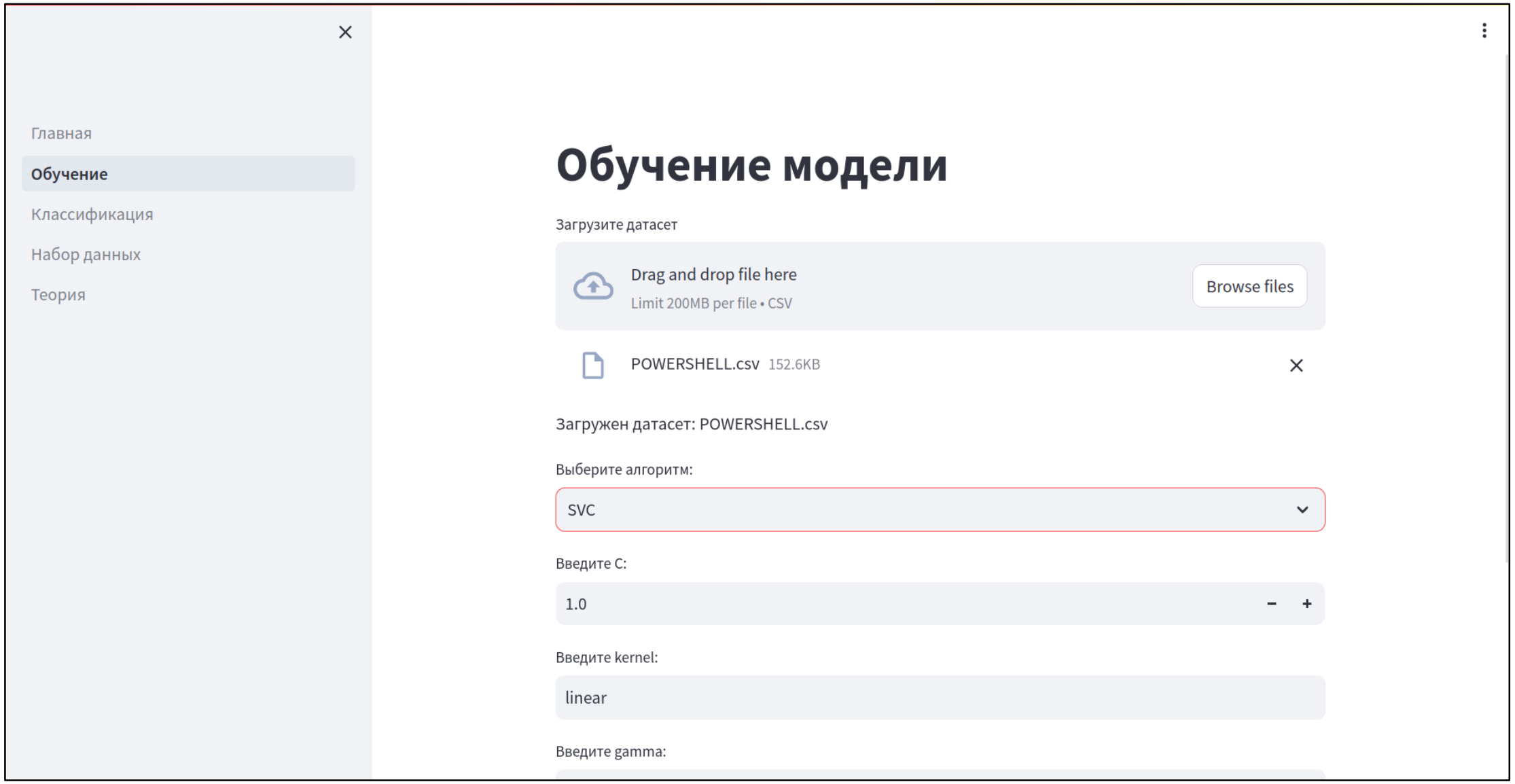
ОБУЧЕНИЕ МОДЕЛЕЙ

- Для обучения было использовано 80% от всей выборки (1 280 записей из 1 600)
- Для тестирования было использовано 20% от всей выборки (320 записей из 1600)
- Обучение производилось с помощью локальных возможностей (CPU: AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx @ 1.40GHz 2.10 GHz)

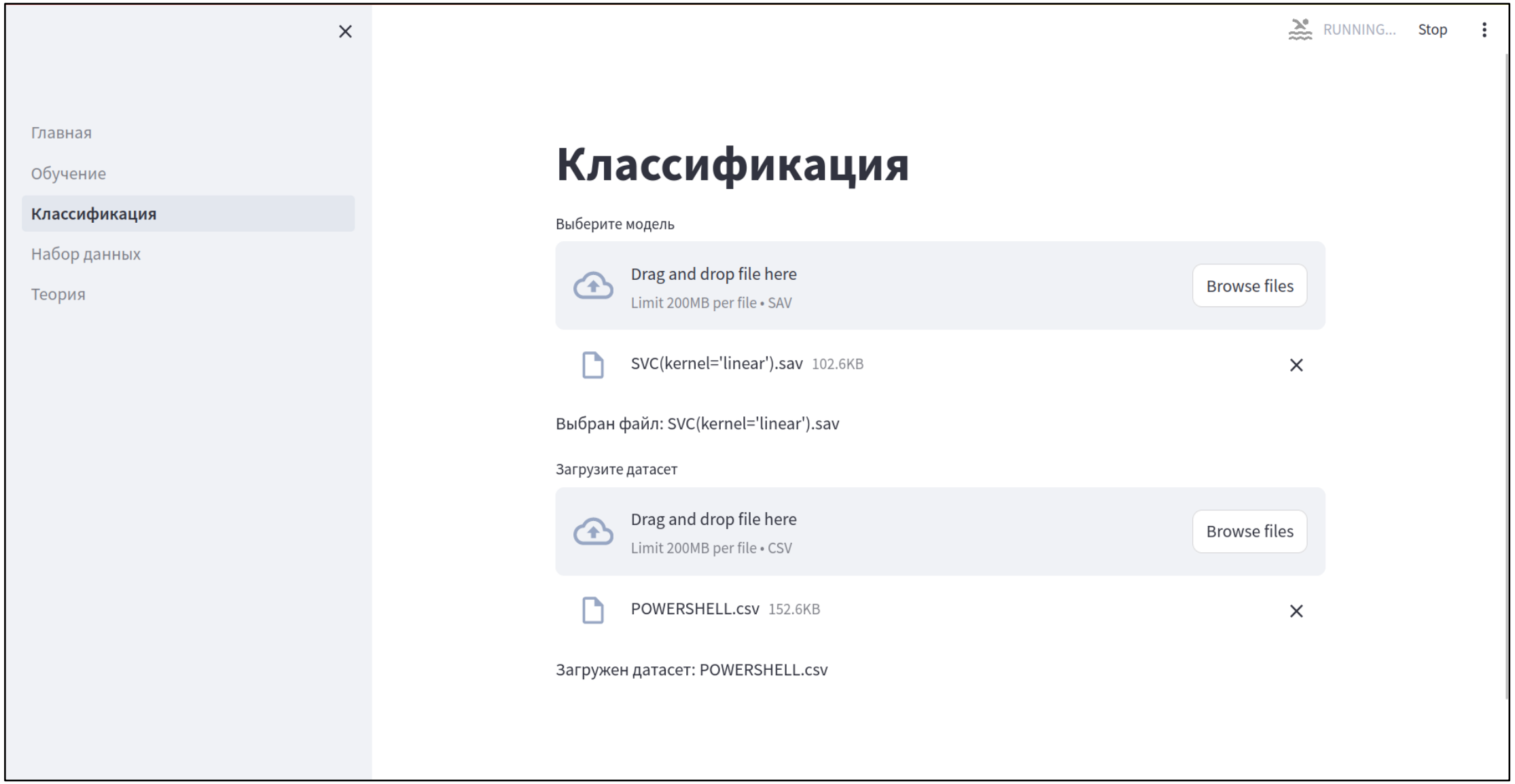
ГЛАВНАЯ СТРАНИЦА



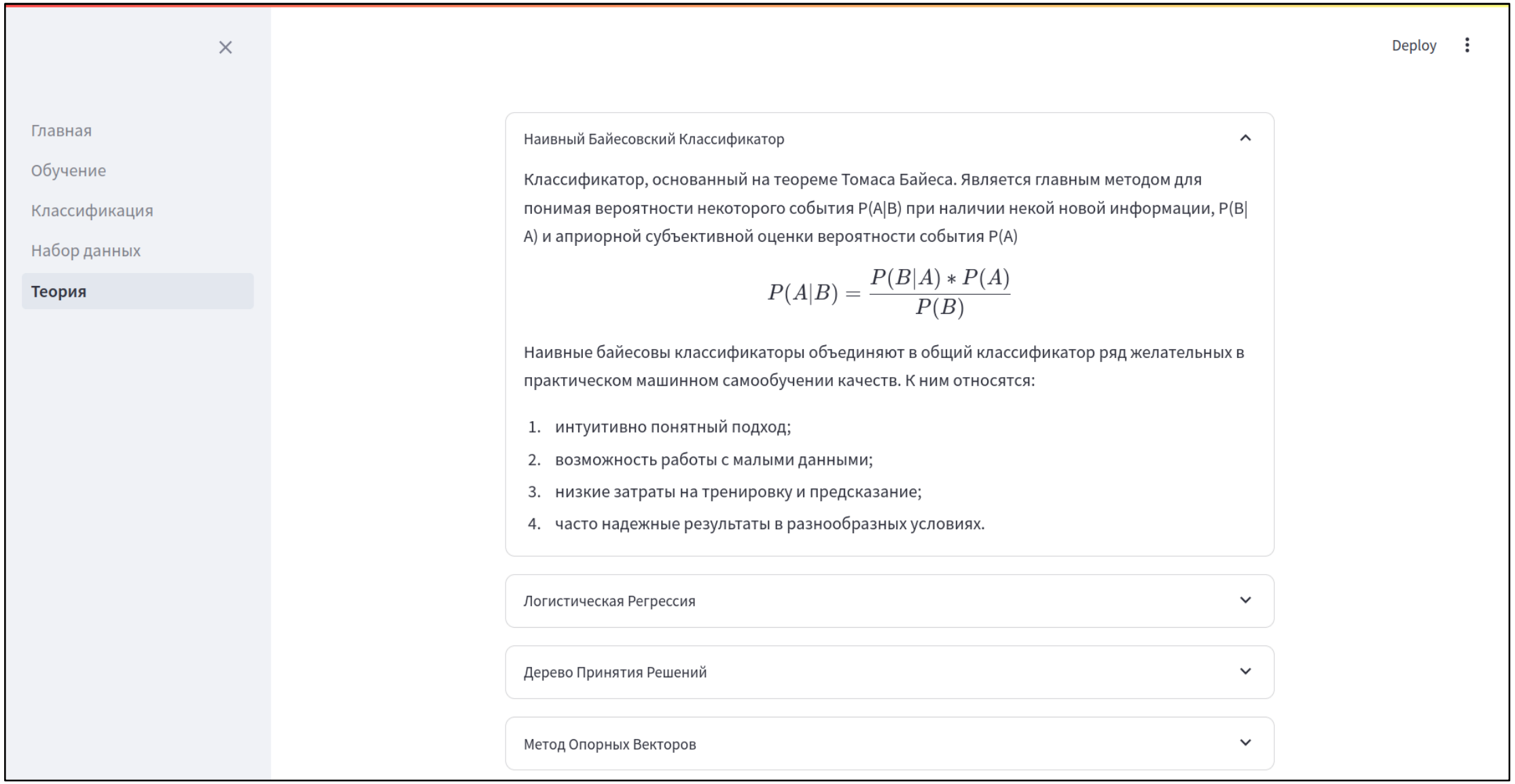
СТРАНИЦА ОБУЧЕНИЕ



СТРАНИЦА КЛАССИФИКАЦИЯ



СТРАНИЦА ТЕОРИЯ



АКТ О ВНЕДРЕНИИ

АКТ
о внедрении научно-технической продукции

Данный акт удостоверяет, что в ООО «Р-Вижн» внедрен в опытную эксплуатацию приложение для бинарной классификации вредоносных команд по метрике MITRE с использованием алгоритмов машинного обучения, разработанный студентом группы КЭ-303 Григорьевым Максимом Дмитриевичем, научный руководитель – старший преподаватель кафедры системного программирования ФГАОУ ВО «ЮУрГУ (НИУ)» Никольская Ксения Юрьевна.

Приложение для бинарной классификации вредоносных команд по метрике MITRE с использованием алгоритмов машинного обучения используется в коммерческих целях.

Акт подписал
Генеральный директор
Бондаренко А.В.
31.05.2023 г.
МП

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Проведен обзор научной литературы
2. Подготовлен обучающий набор данных
3. Реализованы выбранные методы машинного обучения
4. Разработано веб-приложение для классификации вредоносных команд по метрике MITRE
5. Проведено тестирование разработанного веб-приложения

АНАЛИЗ ЛИТЕРАТУРЫ

Название	Авторы	Набор данных	Алгоритм	Точность
Data Mining Applied to Intrusion Detection: MITRE Experiences	Bloedorn, E. E., Talbot, L. M., & DeBarr, D. D. (n.d.)	Собран вручную	Random Forest	91%
Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix	Wenjun Xiong, Emeline Legrand, Oscar Åberg, Robert Lagerström	база знаний MITRE ATT&CK	Decision Tree	97%
A Machine Learning Approach to Dataset Imputation for Software Vulnerabilities	Shahin Rostami, Agnieszka Kleszcz, Daniel Dimanov, Vasilios Katos	ENISA	Logistic Regression	99,88%
A Novel Enhanced Naïve Bayes Posterior Probability (ENBPP) Using Machine Learning: Cyber Threat Analysis	Ayan Sentuna, Abeer Alsadoon, P. W. C. Prasad, Maha Saadeh, Omar Hisham Alsadoon	Собран вручную	Naïve Bayes	92-96%

ФУНКЦИОНАЛЬНОЕ ТЕСТИРОВАНИЕ СТРАНИЦЫ КЛАССИФИКАЦИЯ

Название теста	Шаги	Ожидаемый результат	Тест пройден?
Загрузка готовой модели машинного обучения.	1. Нажать на кнопку «Загрузить файл». 2. Выбрать файл.	Программа должна вывести на экран название файла.	Да
Загрузка набора данных	1. Нажать на кнопку «загрузить файл». 2. Выбрать файл в формате .csv.	Программа должна вывести название файла.	Да
Вывод результата классификации модели.	1. После загрузки набора данных автоматически запуститься классификация команд.	Результат классификации будет выведен в виде таблицы, где правильные ответы выделены зеленым а неправильные – красным.	Да
Скачать результат классификации.	1. Нажать на кнопку «Скачать файл».	Файл с результатами классификации в формате .csv должен появиться в загрузках браузера.	Да

ФУНКЦИОНАЛЬНОЕ ТЕСТИРОВАНИЕ СТРАНИЦЫ ОБУЧЕНИЕ

Название теста	Шаги	Ожидаемый результат	Тест пройден?
Загрузка набора данных	1. Нажать на кнопку «загрузить файл». 2. Выбрать файл в формате .csv.	Программа должна вывести название файла.	Да
Выбор алгоритма и гиперпараметров модели	1. Выбрать алгоритм из списка. 2. Настроить значения гиперпараметров.	Значения гиперпараметров должны примениться к модели.	Да
Начать обучение модели	1. Нажать на кнопку «начать обучение».	Программа должна вывести на экран слово «Готово!» и точность получившейся модели.	Да
Скачать получившеюся модели	1. Нажать на кнопку «Скачать модель».	Файл скаченной модели должен появиться в загрузках браузера.	Да

A/B ТЕСТИРОВАНИЕ

Количество данных	Алгоритм	Score
300	GaussianNB	0,98
	LogisticRegression	0,97
	DecisionTreeClassifier	0,96
	RandomForestClassifier	0,96
	XGBClassifier	0,97
	SVC	**0,98**
	KNN	0,97
600	GaussianNB	0,96
	LogisticRegression	0,98
	DecisionTreeClassifier	0,98
	RandomForestClassifier	**0,99**
	XGBClassifier	0,97
	SVC	0,98
	KNN	0,98
900	GaussianNB	0,98
	LogisticRegression	**0,99**
	DecisionTreeClassifier	0,98
	RandomForestClassifier	0,97
	XGBClassifier	0,98
	SVC	0,96
	KNN	0,97

Количество данных	Алгоритм	Score
1200	GaussianNB	0,96
	LogisticRegression	0,98
	DecisionTreeClassifier	0,98
	RandomForestClassifier	0,97
	XGBClassifier	0,98
	SVC	0,96
	KNN	**0,99**
1500	GaussianNB	0,98
	LogisticRegression	0,96
	DecisionTreeClassifier	0,97
	RandomForestClassifier	0,95
	XGBClassifier	**0,99**
	SVC	0,97
	KNN	0,97

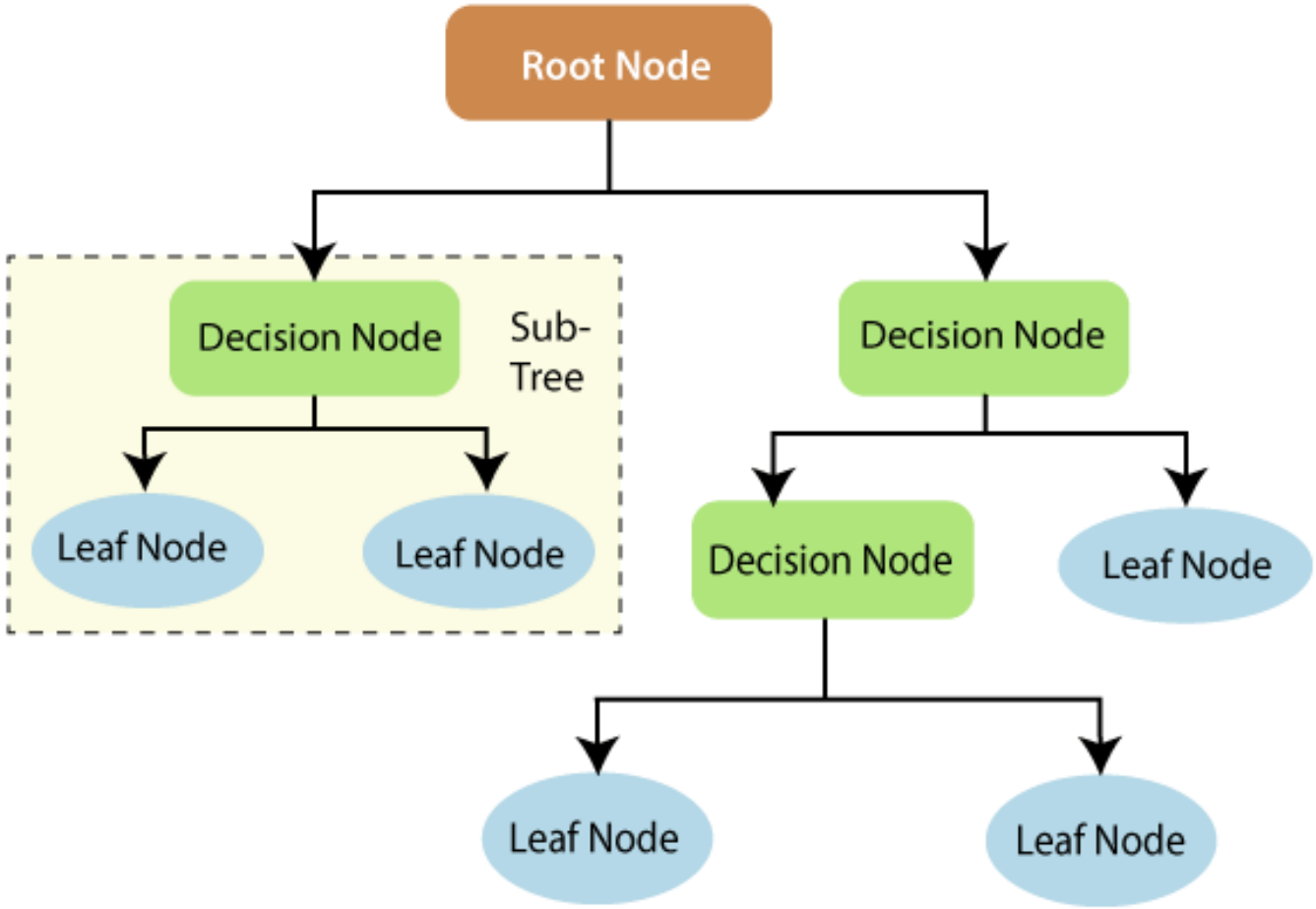
МЕТРИКИ НА ТЕСТОВЫХ ДАННЫХ

Модель	Accuracy	Recall	Precision	F1
Gaussian Naïve Bayes	0,9812500	0,9944751	0,9884393	0,9890109
Logistic Regression	0,9945234	0,9967534	0,9912876	**0,9915966**
Random Forest Classifier	0,9836862	0,9822995	0,9916362	0,9864222
Decision Tree Classifier	0,9687500	0,9833333	0,9719101	0,9708222
XGBoost	0,9631820	0,9787213	0,9681148	0,9723599
SVM	0,9884122	0,9886734	0,9881825	0,9812521
KNN	0,9887631	0,9812552	0,9823677	0,9887721

NAIVE BAYES

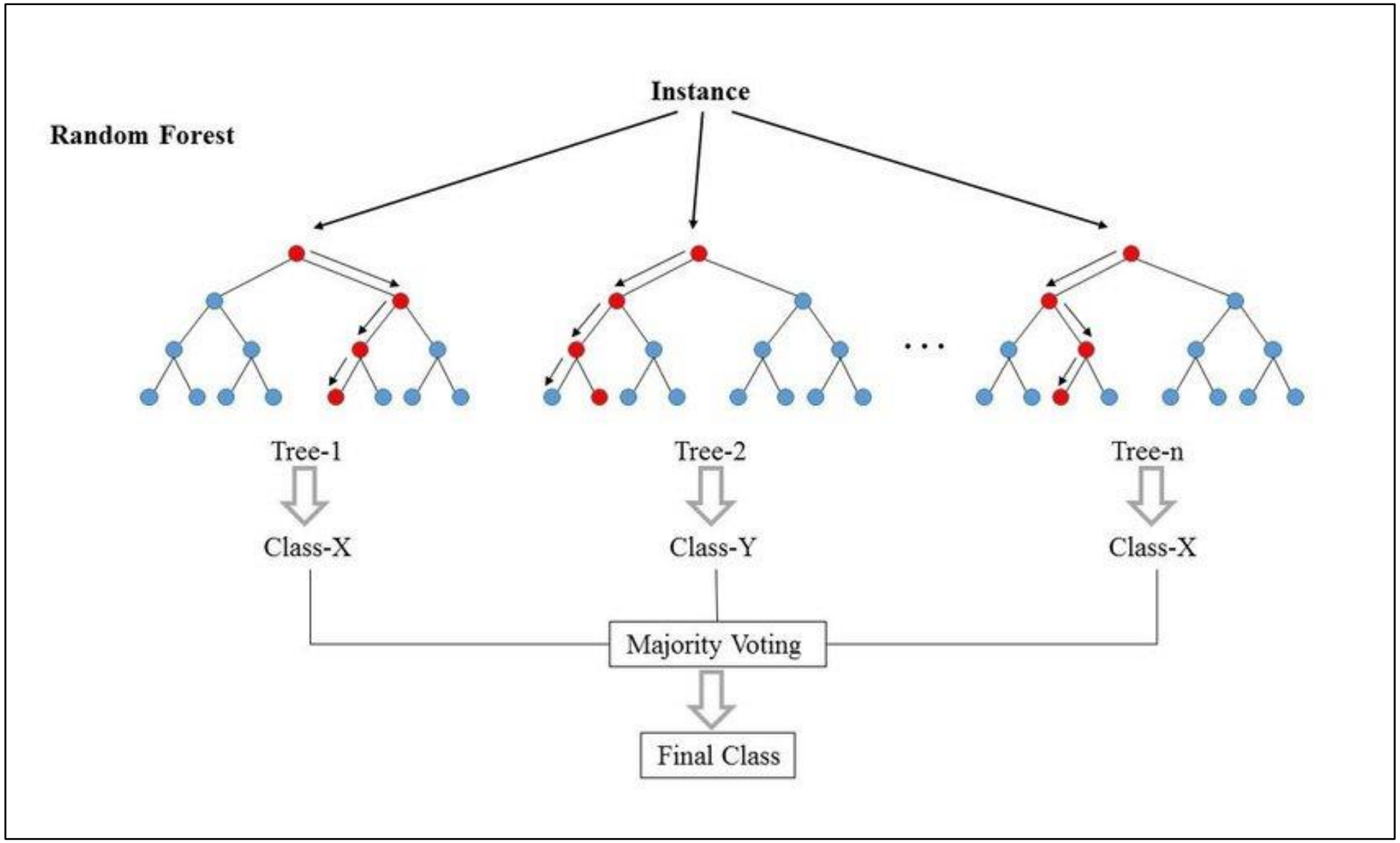
$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

DECISION TREE



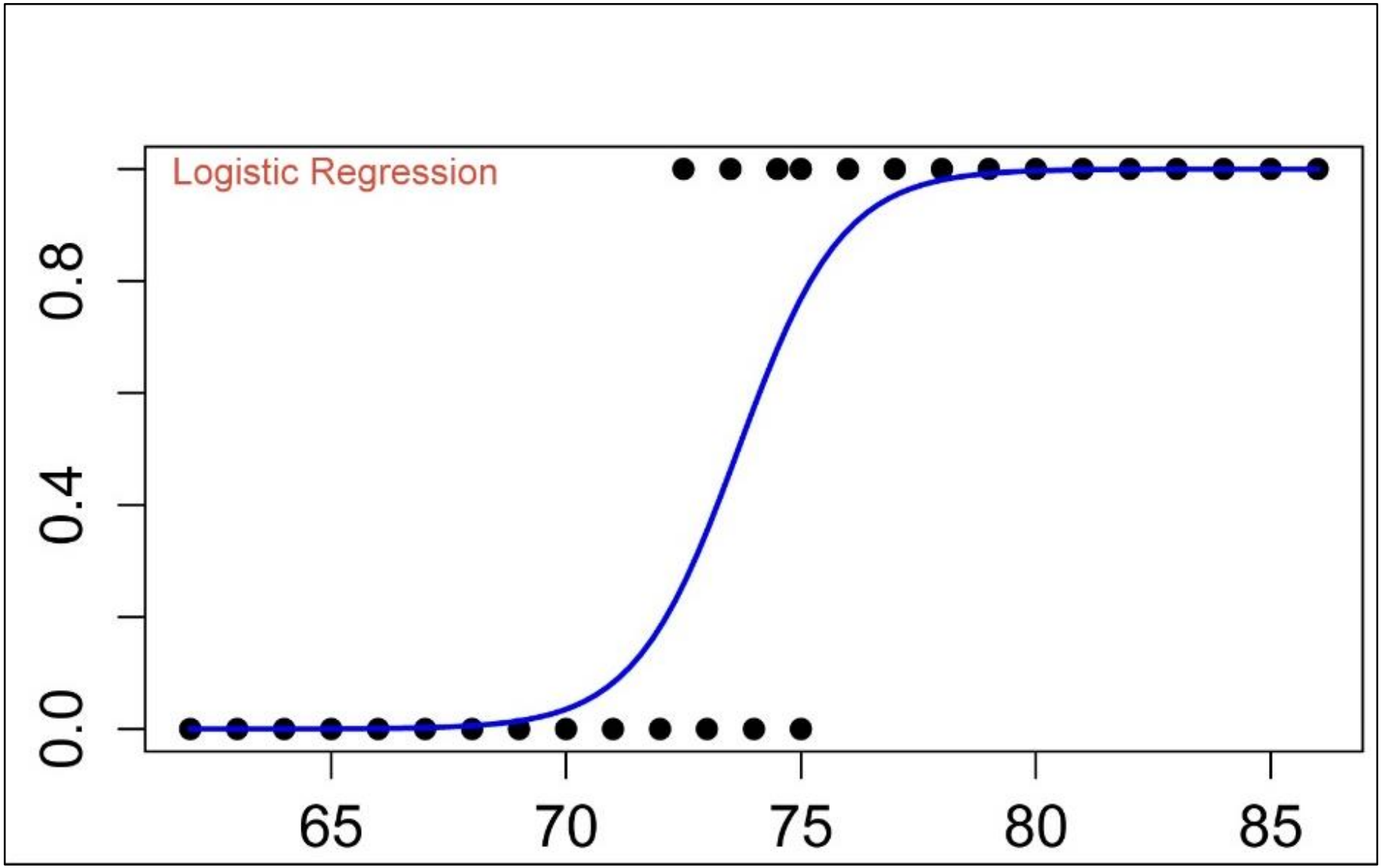
https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

RANDOM FOREST



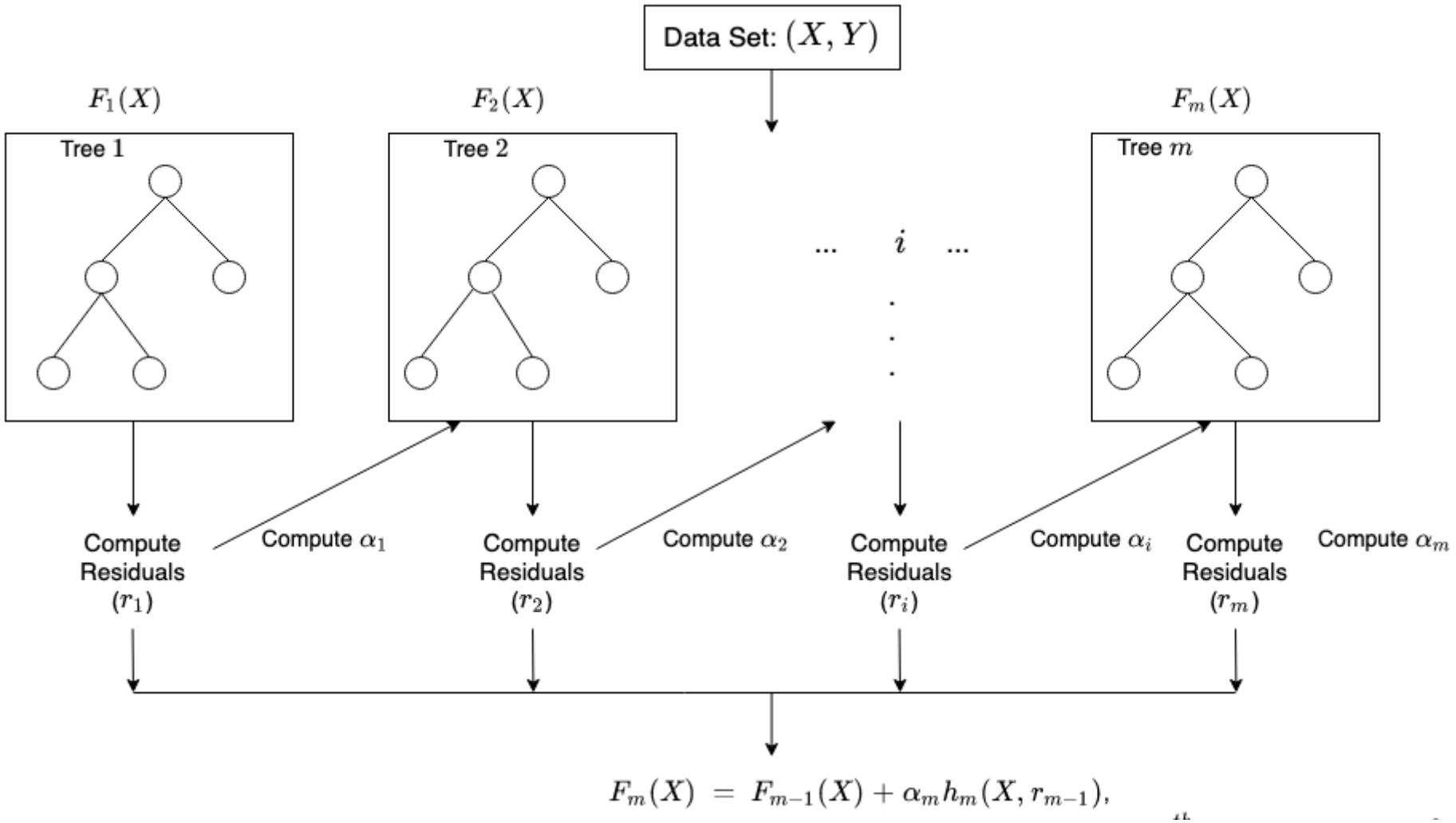
https://www.researchgate.net/figure/Classification-process-based-on-the-Random-Forest-algorithm-2_fig1_324517994

LOGISTIC REGRESSION



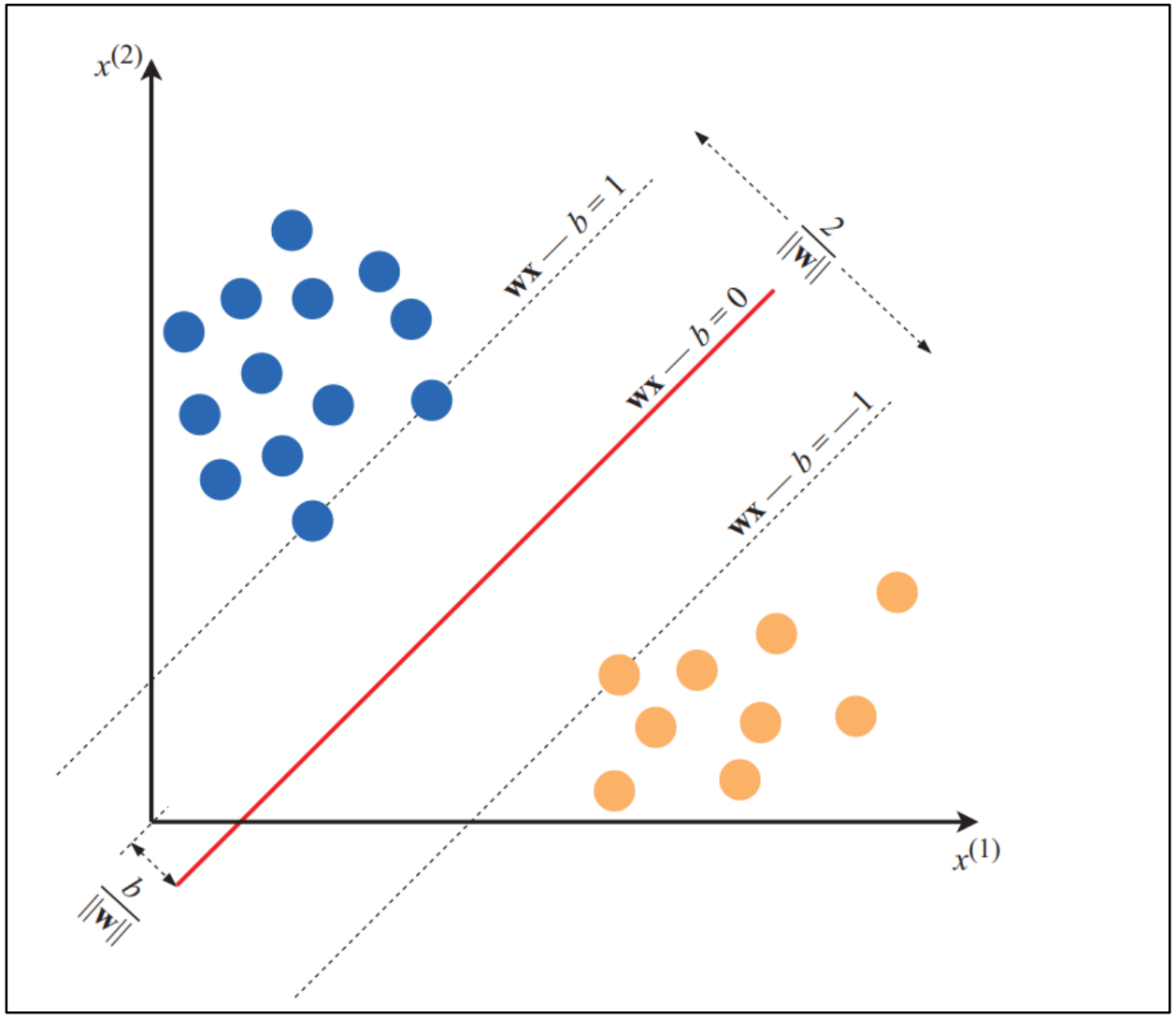
https://stepik.org/lesson/414553/step/1?auth=login&unit=914632

XGBOOST



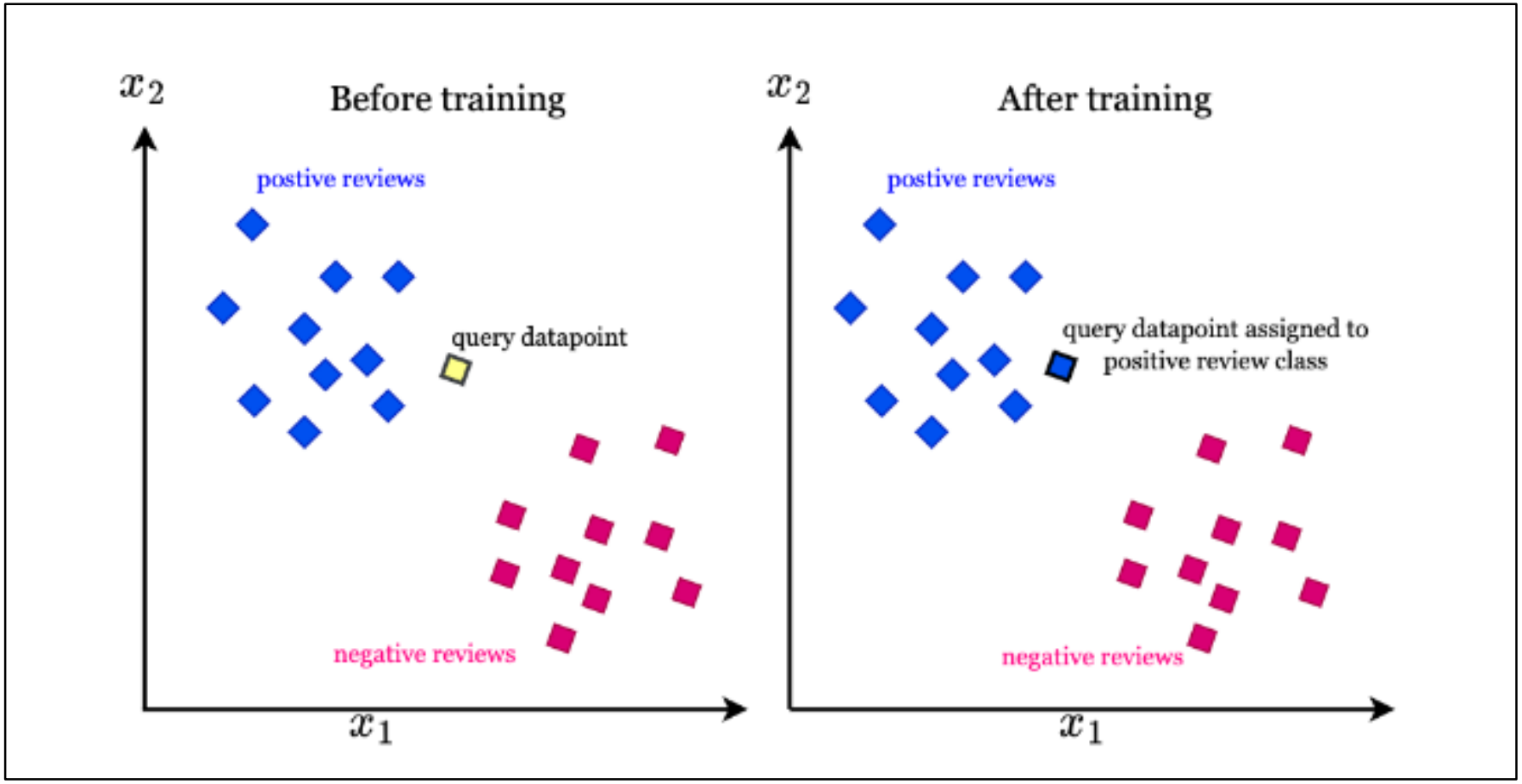
https://aws.amazon.com/ru/what-is/boosting/

Support Vector Machines



Бурков А. Машинное обучение без лишних слов. // под ред. – СПб: Питер, 2020. – 192 с.

K-Nearest Neighbors



https://intuitivetutorial.com/2023/04/07/k-nearest-neighbors-algorithm/