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 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-303
М.Д. Григорьев

Челябинск, 2023 г.

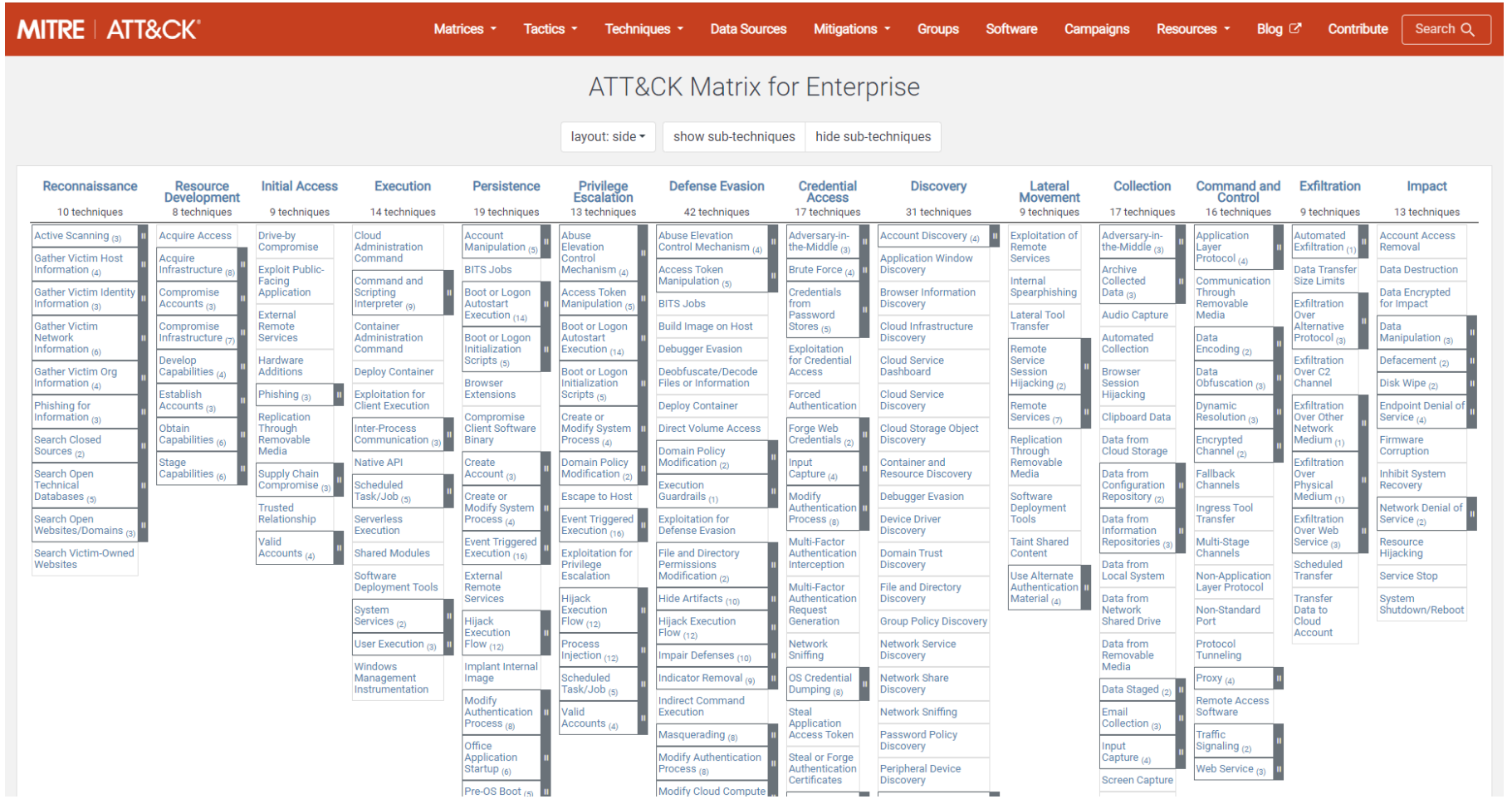
ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Цель: разработать 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 приложение классификации вредоносных команд по по метрике MITRE
5. Провести тестирование разработанного приложения

MITRE ATT&CK



https://attack.mitre.org/

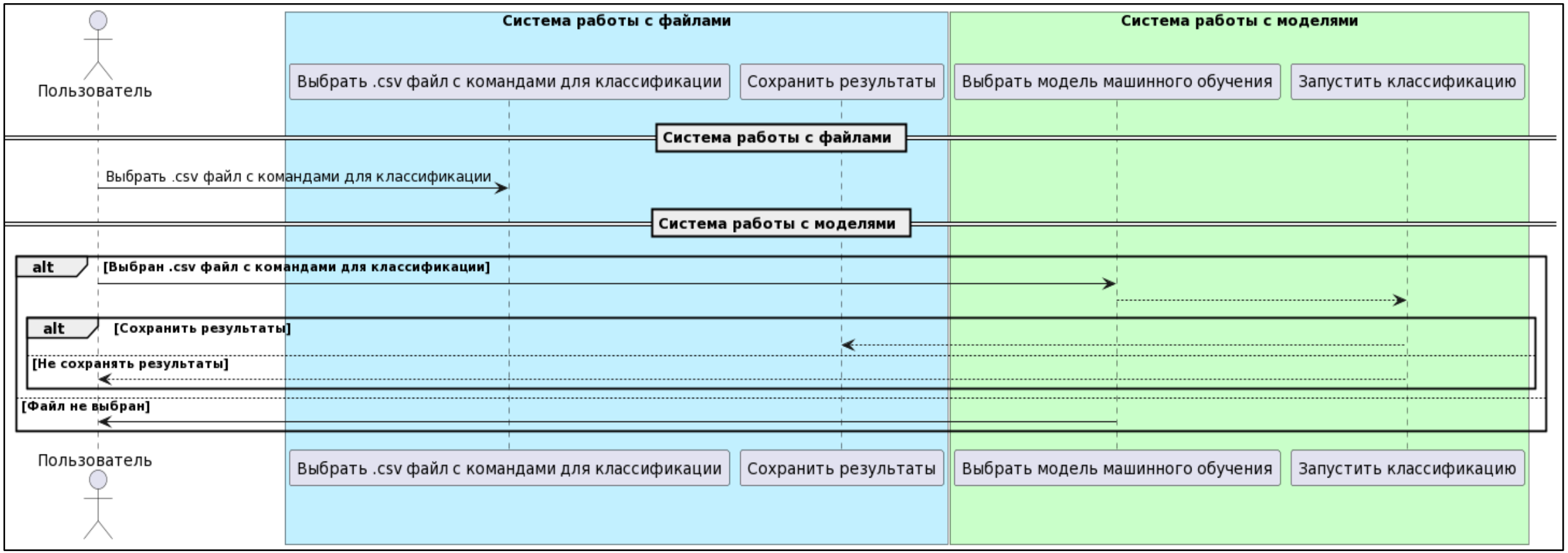
АНАЛИЗ ЛИТЕРАТУРЫ

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%

ВАРИАНТЫ ИСПОЛЬЗОВАНИЯ

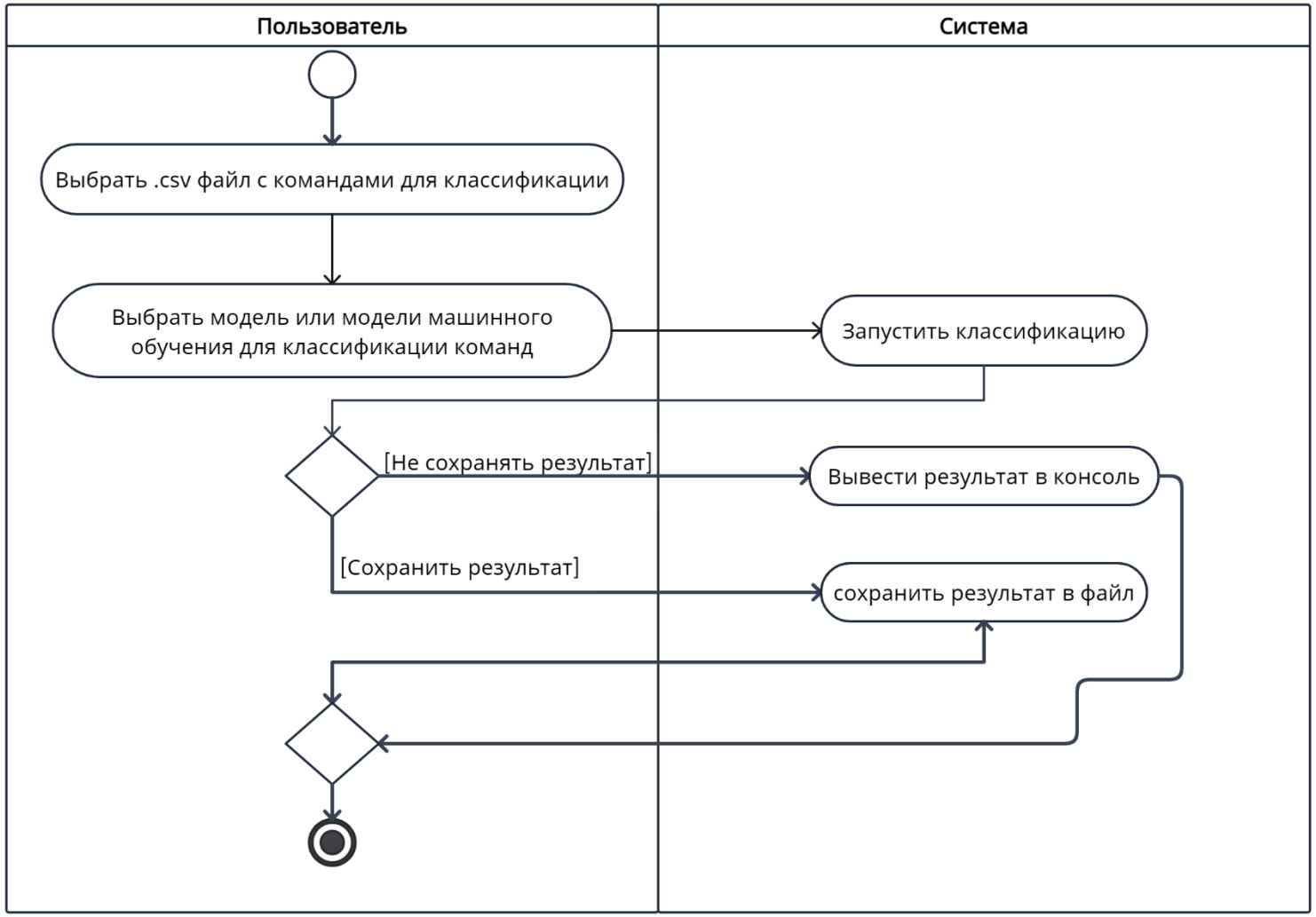
https://github.com/MyPowerIsDivine/CourseWork

ДИАГРАММА ПОСЛЕДОВАТЕЛЬНОСТИ



https://github.com/MyPowerIsDivine/CourseWork

ДИАГРАММА ДЕЯТЕЛЬНОСТИ РАБОТЫ С ПРИЛОЖЕНИЕМ



https://github.com/MyPowerIsDivine/CourseWork

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

https://github.com/MyPowerIsDivine/CourseWork

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)

https://github.com/MyPowerIsDivine/CourseWork

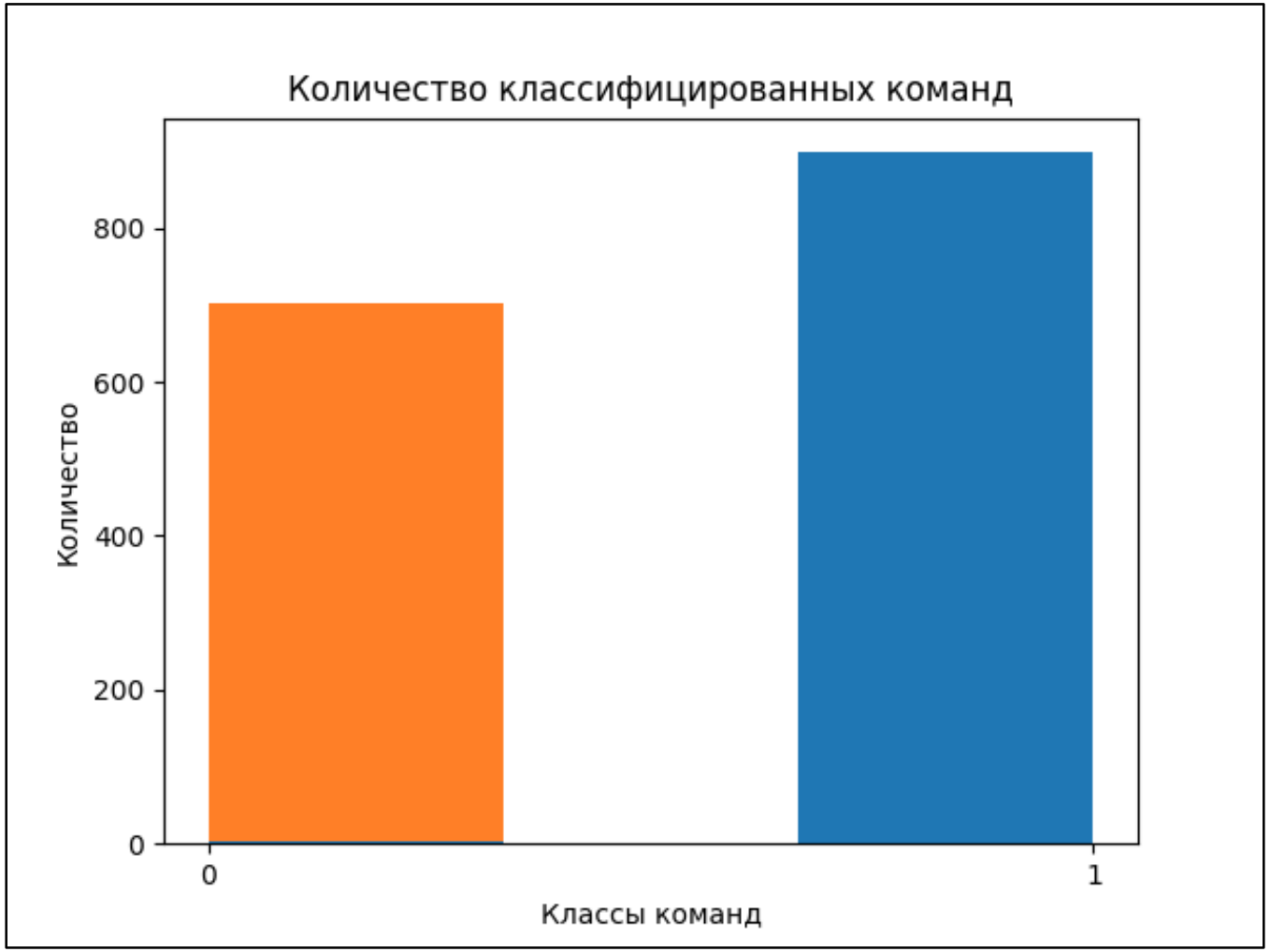
ПРЕДОБРАБОТКА

1. Проверка на наличие пустых значений и оценка типов признаков
2. Трансформация признаков

Итог:

- Количество записей: 1 600

РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ ПОСЛЕ ПРЕДОБРАБОТКИ



https://github.com/MyPowerIsDivine/CourseWork

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

1. Naïve Bayes
2. Logistic Regression
3. Decision Tree
4. Random Forest

https://github.com/MyPowerIsDivine/CourseWork

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)

https://github.com/MyPowerIsDivine/CourseWork

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

https://github.com/MyPowerIsDivine/CourseWork

РАБОТА ПРИЛОЖЕНИЯ (1)

```
divine@divinelaptop:~/PycharmProjects/DataAnalytics/CourseWork/dist/mitre_main$ ./mitre_main
Выберите действие:
1 --> Загрузить файл
2 --> Выбрать алгоритм
3 --> Выход
1
Введите абсолютный путь до файла: /home/divine/PycharmProjects/DataAnalytics/POWERSHELL.csv
                                    command_clear  malicious
513  C:\Windows\system32\CompatTelRunner.exe -m:inv...          0
794  taskeng.exe {00833C53-8F14-4E72-BB18-7DBAF7D0E...          0
731  c:\windows\system32\svchost.exe -k netsvcs -p ...          0
518  C:\Windows\system32\DllHost.exe /Processid:{C1...          0
947  C:\Windows\system32\wevtutil.exe cl Microsoft-...          1
Выберите действие:
1 --> Загрузить файл
2 --> Выбрать алгоритм
3 --> Выход
2
Введите количество алгоритмов: 2
Введите название алгоритма:
1 --> Байесовский классификатор
2 --> Логистическая регрессия
3 --> Бэггинг
4 --> Дерево решений
1
Введите название алгоритма:
1 --> Байесовский классификатор
2 --> Логистическая регрессия
3 --> Бэггинг
4 --> Дерево решений
2
Сохранить результаты?
1 --> Yes
2 --> No
```

https://github.com/MyPowerIsDivine/CourseWork

РАБОТА ПРИЛОЖЕНИЯ (2)

```
Сохранить результаты?
1 --> Yes
2 --> No
2
                                                Command  Naive Bayes  Logistic Regression  True Answer
604                  C:\Windows\system32\lpremove.exe              1                    0            0
1081  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1303  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
815                              taskhostw.exe TpmTasks            1                    0            0
522   C:\Windows\system32\MpSigStub.exe /stub 1.1.15...            0                    0            0
414            C:\Windows\SysWOW64\runonce.exe /Run6432            1                    0            0
1290  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
277   C:\Program Files\Common Files\Microsoft Shared...            1                    0            0
1739  reg  query "HKLM\SOFTWARE\Wow6432Node\Microsof...            1                    1            0
703   c:\windows\system32\svchost.exe -k dcomlaunch ...            0                    1            0
1601  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    0            1
464   C:\Windows\System32\svchost.exe -k LocalSystem...            0                    0            0
757   git pack-objects --all-progress-implied --revs...            0                    1            0
778                 onsent.exe 968 310 0000000003BD7EE0            1                    1            0
432   C:\Windows\System32\Upfc.exe /launchtype boot ...            0                    0            0
1349  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1092  C:\Windows\system32\wevtutil.exe  cl Microsoft...            0                    1            1
881                   C:\Windows\system32\taskmgr.exe /4            1                    0            1
1656  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    0            1
441    C:\Windows\System32\ie4uinit.exe -ClearIconCache            0                    0            0
408   C:\Windows\SysWOW64\SearchProtocolHost.exe Glo...            1                    1            0
474            C:\Windows\System32\svchost.exe -k swprv            0                    1            0
936   C:\Windows\system32\wevtutil.exe cl Microsoft-...            1                    1            1
1129  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
315   C:\ProgramData\Microsoft\Windows Defender\plat...            0                    1            0
774                              git.exe  remote update            1                    1            0
725   c:\windows\system32\svchost.exe -k netsvcs -p ...            0                    0            0
1220  C:\Windows\system32\wevtutil.exe  cl Microsoft...            0                    1            1
1075  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    0            1
1566  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1719  C:\Users\AlBungstein\AppData\Local\Microsoft\O...            1                    0            0
727   c:\windows\system32\svchost.exe -k netsvcs -p ...            1                    0            0
418   C:\Windows\System32\DataExchangeHost.exe -Embe...            1                    1            0
782   reg  query "HKLM\SOFTWARE\Wow6432Node\Microsof...            1                    1            0
413   C:\Windows\SysWOW64\rundll32.exe C:\Windows\Sy...            0                    0            0
961   C:\Windows\system32\wevtutil.exe cl Microsoft-...            1                    0            1
1697                        Get-NetConnectionProfile |            1                    1            1
1530  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1562  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    0            1
1670  Backup-SqlDatabase -ServerINstance YComputer\I...            1                    0            0
1232  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1340  C:\Windows\system32\wevtutil.exe  cl Microsoft...            0                    1            1
1611  C:\Windows\system32\wevtutil.exe  cl Microsoft...            0                    1            1
1642  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    0            1
687   c:\users\amber.turing\appdata\local\google\chr...            1                    0            0
1664  powershell -noProfile -nonInteractive -Windows...            1                    1            0
1312  C:\Windows\system32\wevtutil.exe  cl Microsoft...            1                    1            1
1429  netsh interface ipv6 delete address interface=...            1                    1            1
```

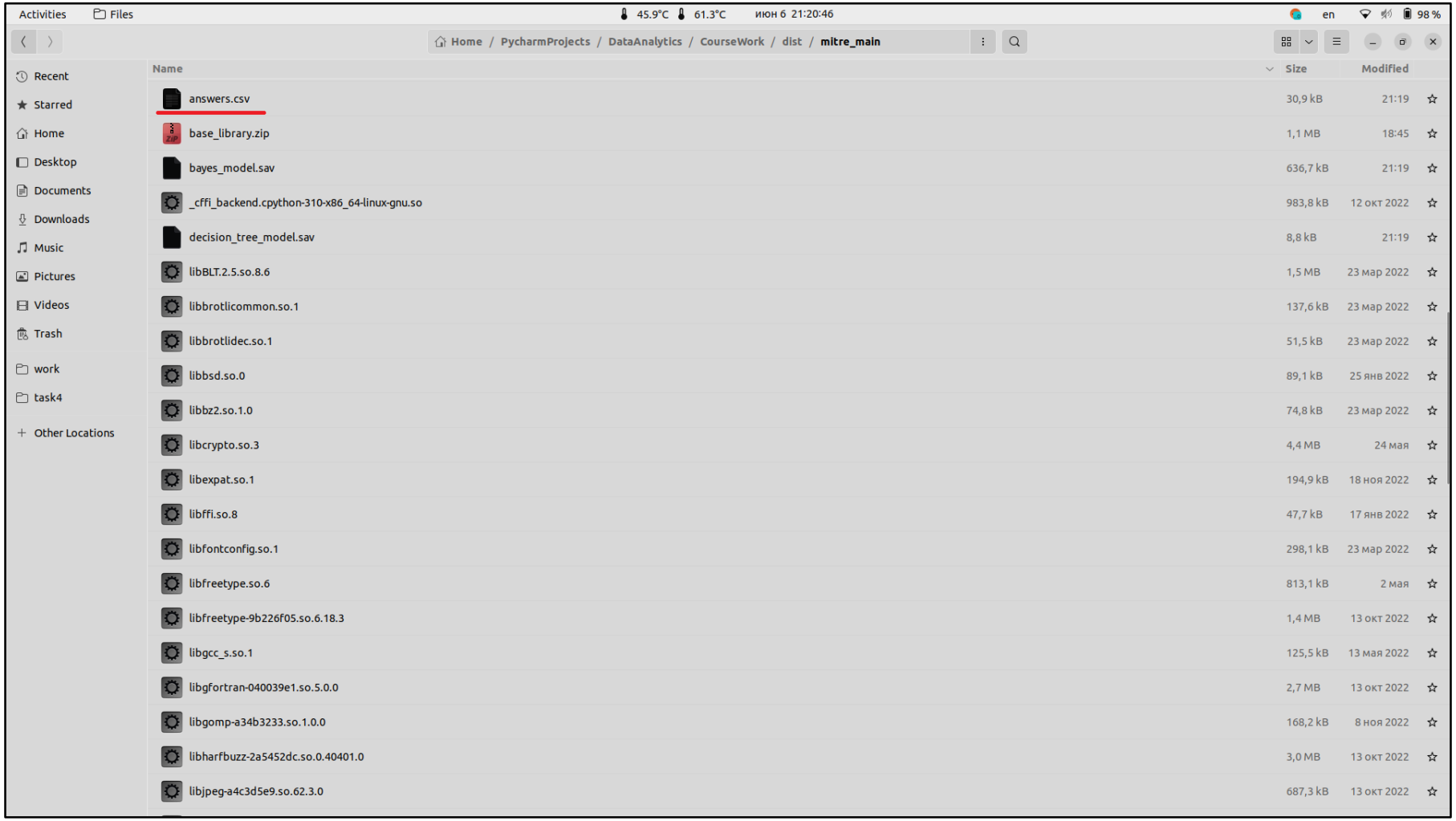
https://github.com/MyPowerIsDivine/CourseWork

РАБОТА ПРИЛОЖЕНИЯ (3)

```
(DataAnalytics) divine@divinelaptop:~/PycharmProjects/DataAnalytics/CourseWork/dist/mitre_main$ ./mitre_main
Выберите действие:
1 --> Загрузить файл
2 --> Выбрать алгоритм
3 --> Выход
1
Введите абсолютный путь до файла: /home/divine/PycharmProjects/DataAnalytics/POWERSHELL.csv
                                        command_clear  malicious
189   $cred=new-object system.management.automation....          1
1074  C:\Windows\system32\wevtutil.exe  cl Microsoft...          1
1099  C:\Windows\system32\wevtutil.exe  cl Microsoft...          1
1489  C:\Windows\system32\wevtutil.exe  cl "Microsof...          1
1522  C:\Windows\system32\wevtutil.exe  cl Microsoft...          1
Выберите действие:
1 --> Загрузить файл
2 --> Выбрать алгоритм
3 --> Выход
2
Введите количество алгоритмов: 2
Введите название алгоритма:
1 --> Байесовский классификатор
2 --> Логистическая регрессия
3 --> Бэггинг
4 --> Дерево решений
1
Введите название алгоритма:
1 --> Байесовский классификатор
2 --> Логистическая регрессия
3 --> Бэггинг
4 --> Дерево решений
2
Сохранить результаты?
1 --> Да
2 --> Нет
1
```

https://github.com/MyPowerIsDivine/CourseWork

РАБОТА ПРИЛОЖЕНИЯ (4)



https://github.com/MyPowerIsDivine/CourseWork

ФУНКЦИОНАЛЬНОЕ ТЕСТИРОВАНИЕ

№	Название теста	Шаги	Ожидаемый результат	Тест пройден?
1	Выбор файла	1. В меню ввести цифру 1. 2. Ввести абсолютный путь до файла.	Программа должна вывести первые несколько строчек файла.	Да
2	Выбор одной модели	1. В меню ввести цифру 2. 2. Ввести цифру 1. 3. С помощью цифр 1-4 выбрать алгоритм.	Программа должна вывести результаты в виде таблицы	Да
3	Выбор двух и более моделей	1. В меню ввести цифру 2. 2. Ввести цифру от 2 до 4. 3. С помощью цифр 1-4 поочередно выбрать алгоритмы.	Программа должна вывести результаты в виде таблицы	Да
4	Ввод неверных параметров	1. В меню выбора интерфейса ввести значение не из списка доступных интерфейсов.	Программа должна вывести сообщение об ошибке «Выберете заново».	Да

https://github.com/MyPowerIsDivine/CourseWork

A/B ТЕСТИРОВАНИЕ

Количество данных	Алгоритм	Score
300	GaussianNB	0,98
	LogisticRegression	1,0
	DecisionTreeClassifier	0,98
	RandomForestClassifier	1,0
600	GaussianNB	0,96
	LogisticRegression	0,98
	DecisionTreeClassifier	0,98
	RandomForestClassifier	0,99
900	GaussianNB	0,98
	LogisticRegression	0,99
	DecisionTreeClassifier	0,98
	RandomForestClassifier	0,99
1 200	GaussianNB	0,99
	LogisticRegression	0,99
	DecisionTreeClassifier	0,98
	RandomForestClassifier	0,99
1 500	GaussianNB	0,98
	LogisticRegression	1,0
	DecisionTreeClassifier	0,97
	RandomForestClassifier	0,99

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 приложение классификации вредоносных команд по метрике MITRE
5. Проведено тестирование разработанного приложения

ВИД КОМАНД

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)

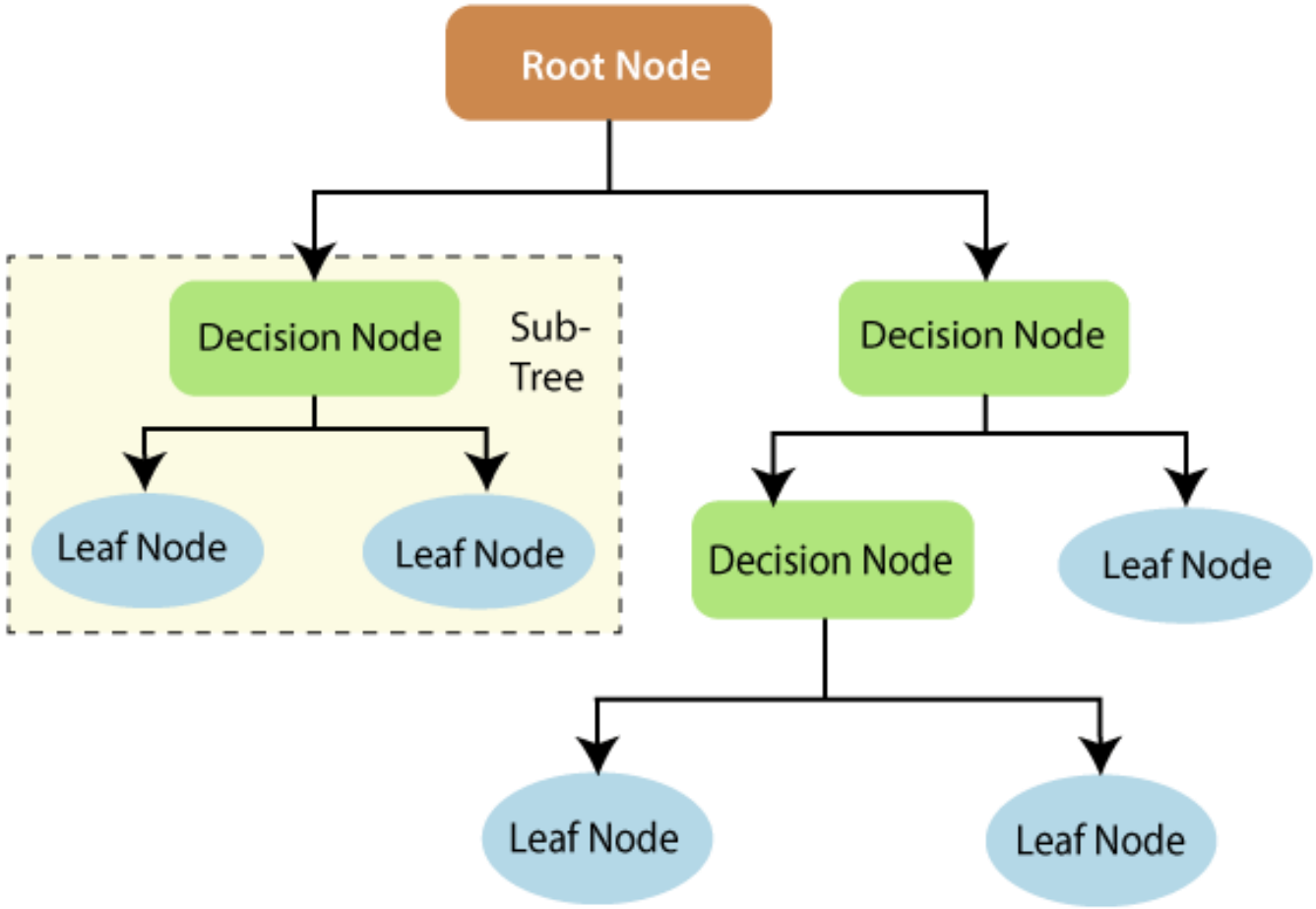
https://github.com/MaksimSkoblyuk/Obfuscation

NAIVE BAYES

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

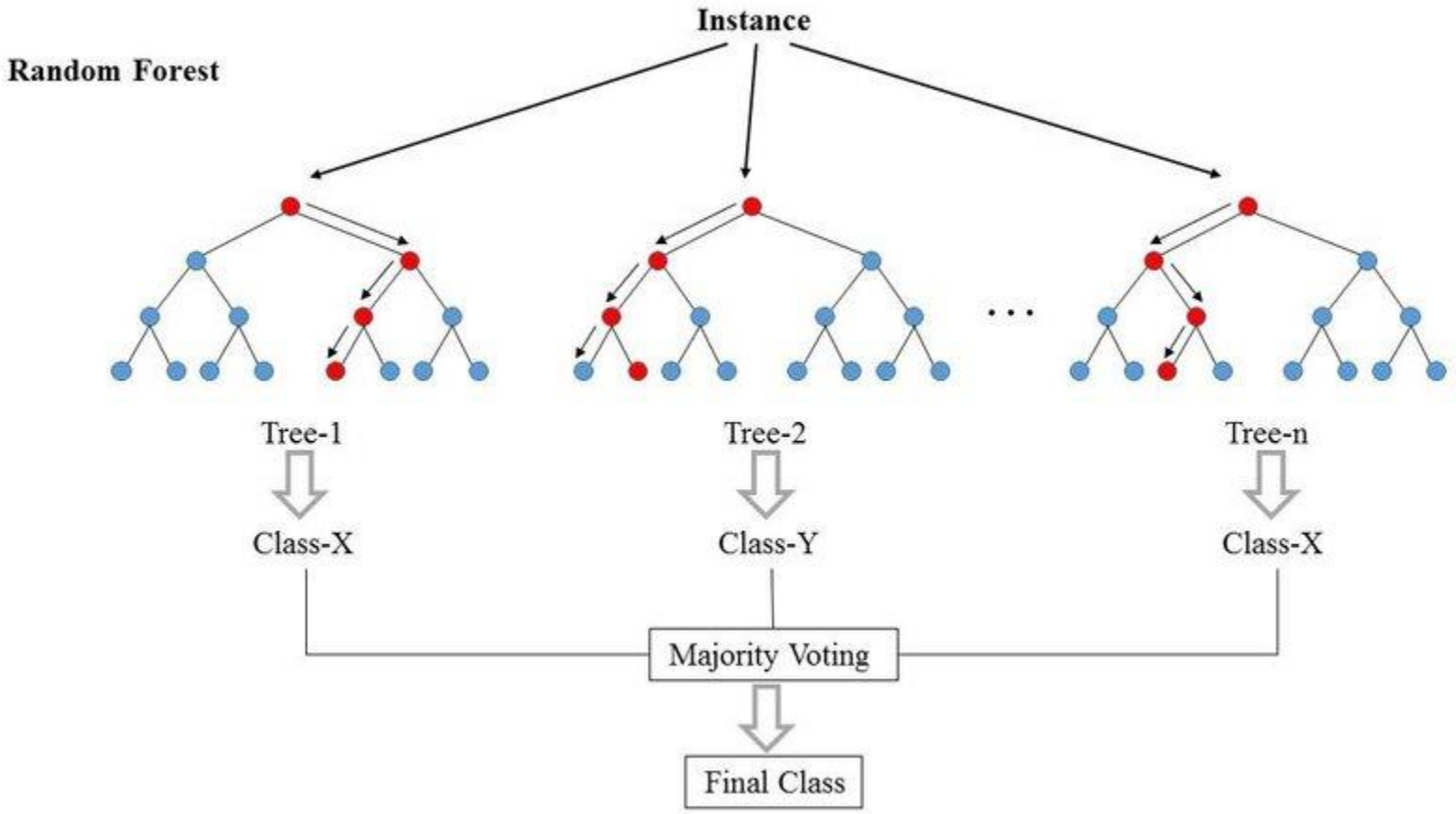
https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

DECISION TREE



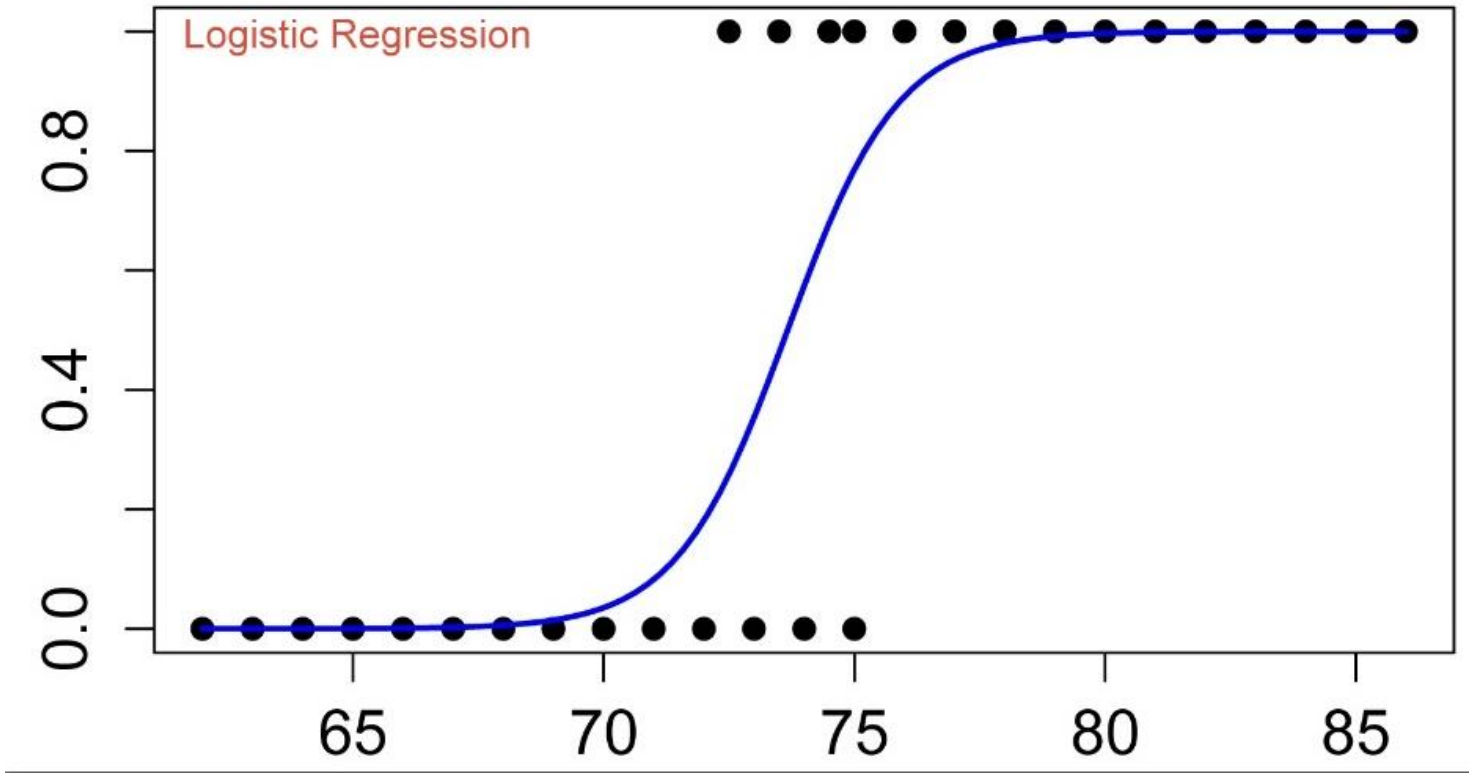
https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

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