

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет (национальный исследовательский университет)»
Вышая 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 метрике MITRE
5. Провести тестирование разработанного приложения

MITRE ATT&CK

MITRE ATT&CK													
Matrices ▾ Tactics ▾ Techniques ▾ Data Sources Mitigations ▾ Groups Software Campaigns Resources ▾ Blog 🔗 Contribute <input type="text" value="Search Q"/>													
ATT&CK Matrix for Enterprise													
layout: side ▾ show sub-techniques hide sub-techniques													
Reconnaissance	Resource Development	Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Command and Control	Exfiltration	Impact
10 techniques	8 techniques	9 techniques	14 techniques	19 techniques	13 techniques	42 techniques	17 techniques	31 techniques	9 techniques	17 techniques	16 techniques	9 techniques	13 techniques
Active Scanning (3)	Acquire Access	Drive-by Compromise	Cloud Administration Command	Account Manipulation (5)	Abuse Elevation Control Mechanism (4)	Abuse Elevation Control Mechanism (4)	Adversary-in-the-Middle (3)	Account Discovery (4)	Exploitation of Remote Services	Adversary-in-the-Middle (3)	Application Layer Protocol (4)	Automated Exfiltration (1)	Account Access Removal
Gather Victim Host Information (4)	Acquire Infrastructure (8)	Exploit Public-Facing Application	Command and Scripting Interpreter (9)	BITS Jobs	Access Token Manipulation (5)	Access Token Manipulation (5)	Brute Force (4)	Application Window Discovery	Internal Spearphishing	Archive Collected Data (3)	Communication Through Removable Media	Data Transfer Size Limits	Data Destruction
Gather Victim Identity Information (3)	Compromise Accounts (3)	External Remote Services	Container Administration Command	Boot or Logon Autostart Execution (14)	Access Token Manipulation (5)	Access Token Manipulation (5)	Credentials from Password Stores (5)	Browser Information Discovery	Lateral Tool Transfer	Audio Capture	Data Encoding (2)	Exfiltration Over Alternative Protocol (3)	Data Encrypted for Impact
Gather Victim Network Information (6)	Compromise Infrastructure (7)	Hardware Additions	Deploy Container	Boot or Logon Autostart Scripts (5)	Boot or Logon Autostart Execution (14)	Boot or Logon Autostart Execution (14)	Exploitation for Credential Access	Cloud Infrastructure Discovery	Remote Service Session Hijacking (2)	Automated Collection	Data Obfuscation (3)	Exfiltration Over C2 Channel	Data Manipulation (3)
Gather Victim Org Information (4)	Develop Capabilities (4)	Phishing (3)	Exploitation for Client Execution	Browser Extensions	Boot or Logon Initialization Scripts (5)	Boot or Logon Initialization Scripts (5)	Forced Authentication	Cloud Service Dashboard	Remote Services (7)	Browser Session Hijacking	Dynamic Resolution (3)	Exfiltration Over Other Network Medium (1)	Defacement (2)
Phishing for Information (3)	Establish Accounts (3)	Replication Through Removable Media	Inter-Process Communication (3)	Compromise Client Software Binary	Create or Modify System Process (4)	Create or Modify System Process (4)	Forge Web Credentials (2)	Cloud Service Discovery	Replication Through Removable Media	Clipboard Data	Encrypted Channel (2)	Exfiltration Over Physical Medium (1)	Disk Wipe (2)
Search Closed Sources (2)	Obtain Capabilities (6)	Supply Chain Compromise (3)	Native API	Create Account (3)	Domain Policy Modification (2)	Domain Policy Modification (2)	Input Capture (4)	Cloud Storage Object Discovery	Software Deployment Tools	Data from Cloud Storage	Fallback Channels	Exfiltration Over Web Service (3)	Endpoint Denial of Service (4)
Search Open Technical Databases (5)	Stage Capabilities (6)	Trusted Relationship	Scheduled Task/Job (5)	Create or Modify System Process (4)	Escape to Host	Event Triggered Execution (16)	Modify Authentication Process (8)	Container and Resource Discovery	Taint Shared Content	Data from Configuration Repository (2)	Ingress Tool Transfer	Scheduled Transfer	Firmware Corruption
Search Open Websites/Domains (3)	Valid Accounts (4)	Serverless Execution	Shared Modules	Event Triggered Execution (16)	Exploitation for Privilege Escalation	Exploitation for Defense Evasion	Multi-Factor Authentication Interception	Debugger Evasion	Use Alternate Authentication Material (4)	Data from Information Repositories (3)	Multi-Stage Channels	Exfiltration Over Web Service (3)	Inhibit System Recovery
Search Victim-Owned Websites		Valid Accounts (4)	Software Deployment Tools	Event Triggered Execution (16)	External Remote Services	File and Directory Permissions Modification (2)	Multi-Factor Authentication Request Generation	Device Driver Discovery		Data from Local System	Non-Application Layer Protocol	Exfiltration Over Web Service (3)	Network Denial of Service (2)
			System Services (2)	Event Triggered Execution (16)	Hijack Execution Flow (12)	Hide Artifacts (10)	Network Sniffing	Domain Trust Discovery		Data from Network Shared Drive	Non-Standard Port	Scheduled Transfer	Resource Hijacking
			User Execution (3)	Event Triggered Execution (16)	Hijack Execution Flow (12)	Impair Defenses (10)	OS Credential Dumping (8)	File and Directory Discovery		Data from Removable Media	Protocol Tunneling	Transfer Data to Cloud Account	Service Stop
			Windows Management Instrumentation	Implant Internal Image	Process Injection (12)	Indicator Removal (9)	Steal Application Access Token	Group Policy Discovery		Data Staged (2)	Proxy (4)		System Shutdown/Reboot
				Modify Authentication Process (8)	Scheduled Task/Job (5)	Indirect Command Execution	Steal or Forge Authentication Certificates	Network Service Discovery		Email Collection (3)	Remote Access Software		
				Office Application Startup (8)	Valid Accounts (4)	Masquerading (8)	Modify Authentication Process (8)	Password Policy Discovery		Input Capture (4)	Traffic Signaling (2)		
				Pre-OS Boot (5)		Modify Cloud Compute		Peripheral Device Discovery		Screen Capture	Web Service (3)		

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

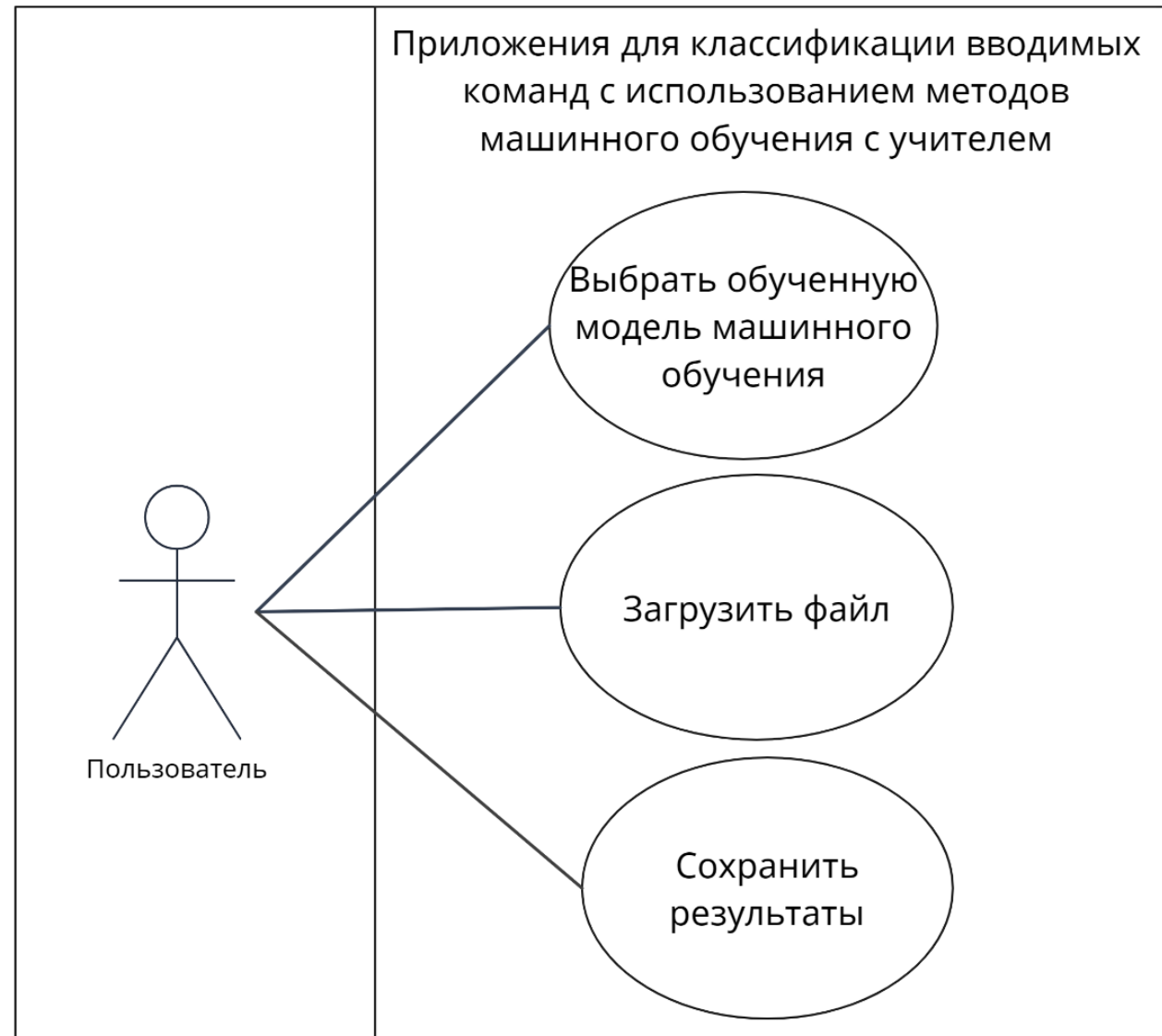


ДИАГРАММА ПОСЛЕДОВАТЕЛЬНОСТИ

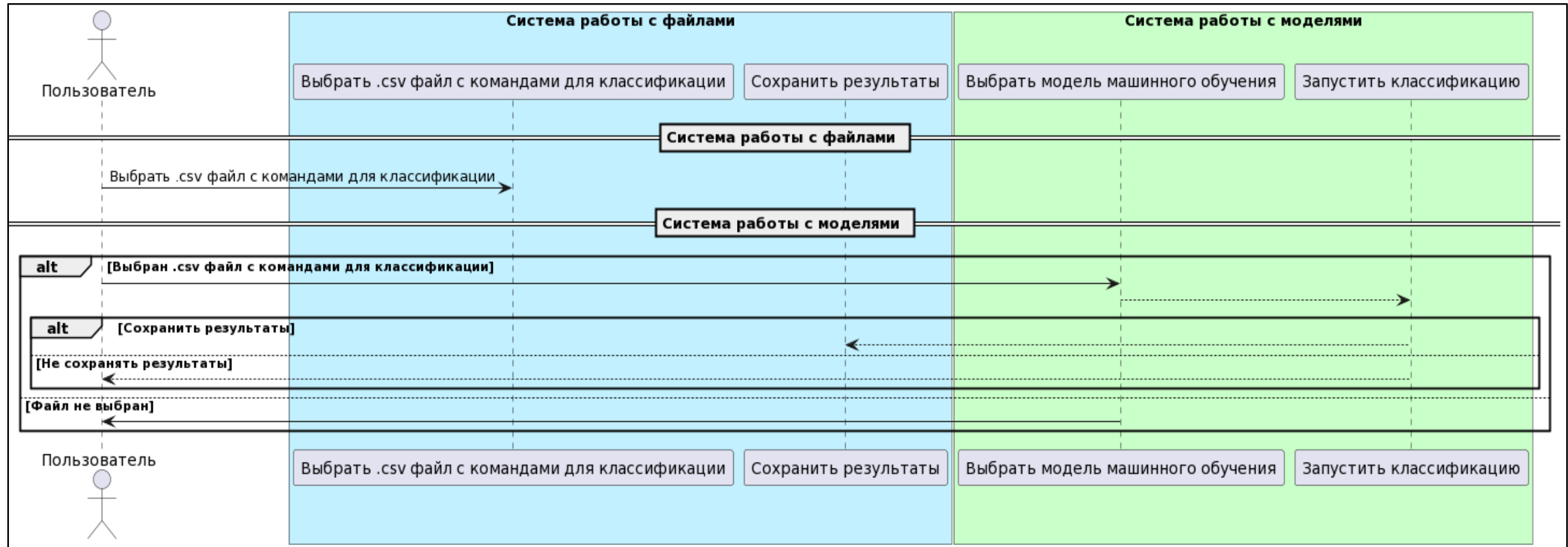
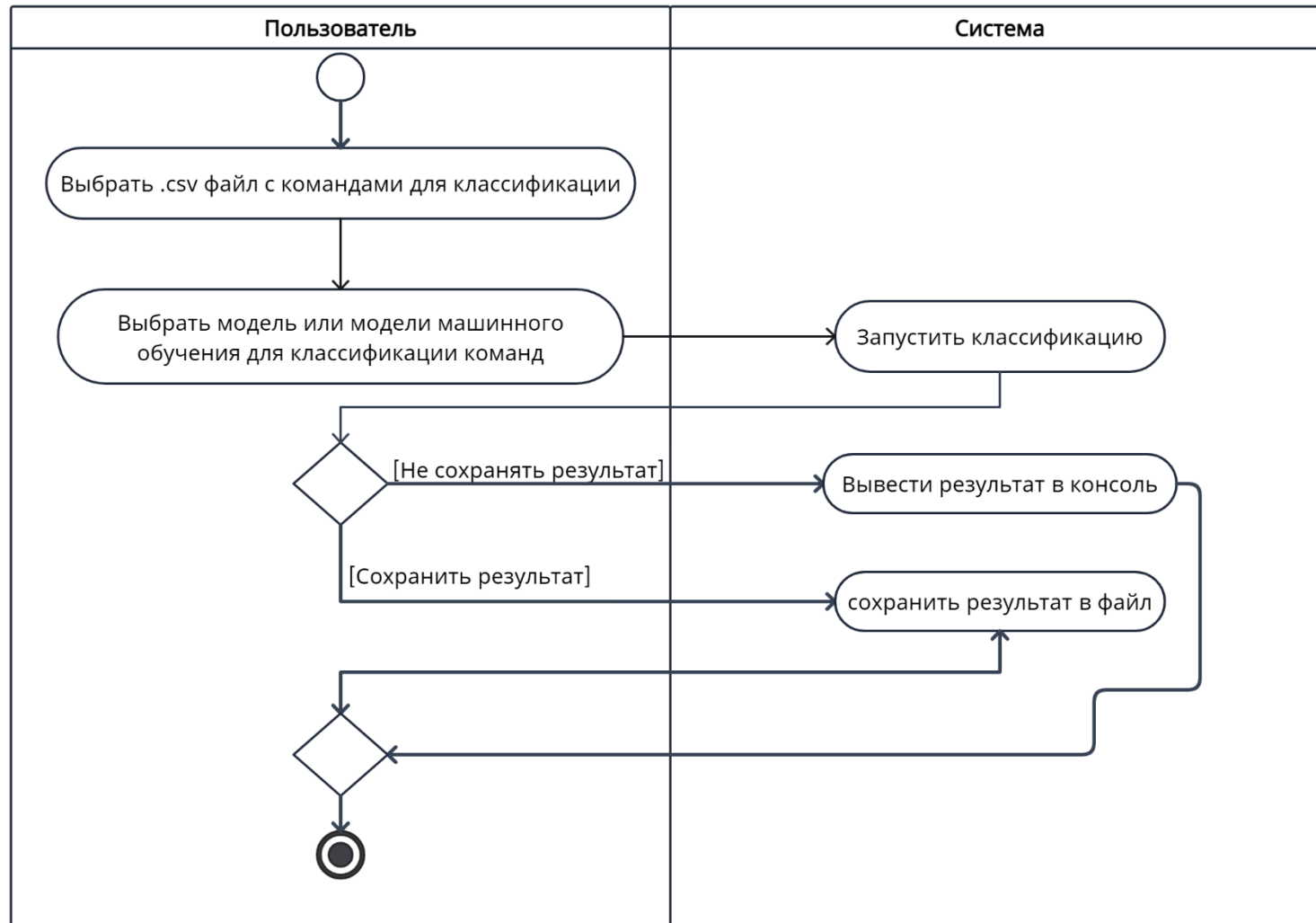


ДИАГРАММА ДЕЯТЕЛЬНОСТИ РАБОТЫ С ПРИЛОЖЕНИЕМ



СРЕДСТВА РАЗРАБОТКИ



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

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)

ПРЕДОБРАБОТКА



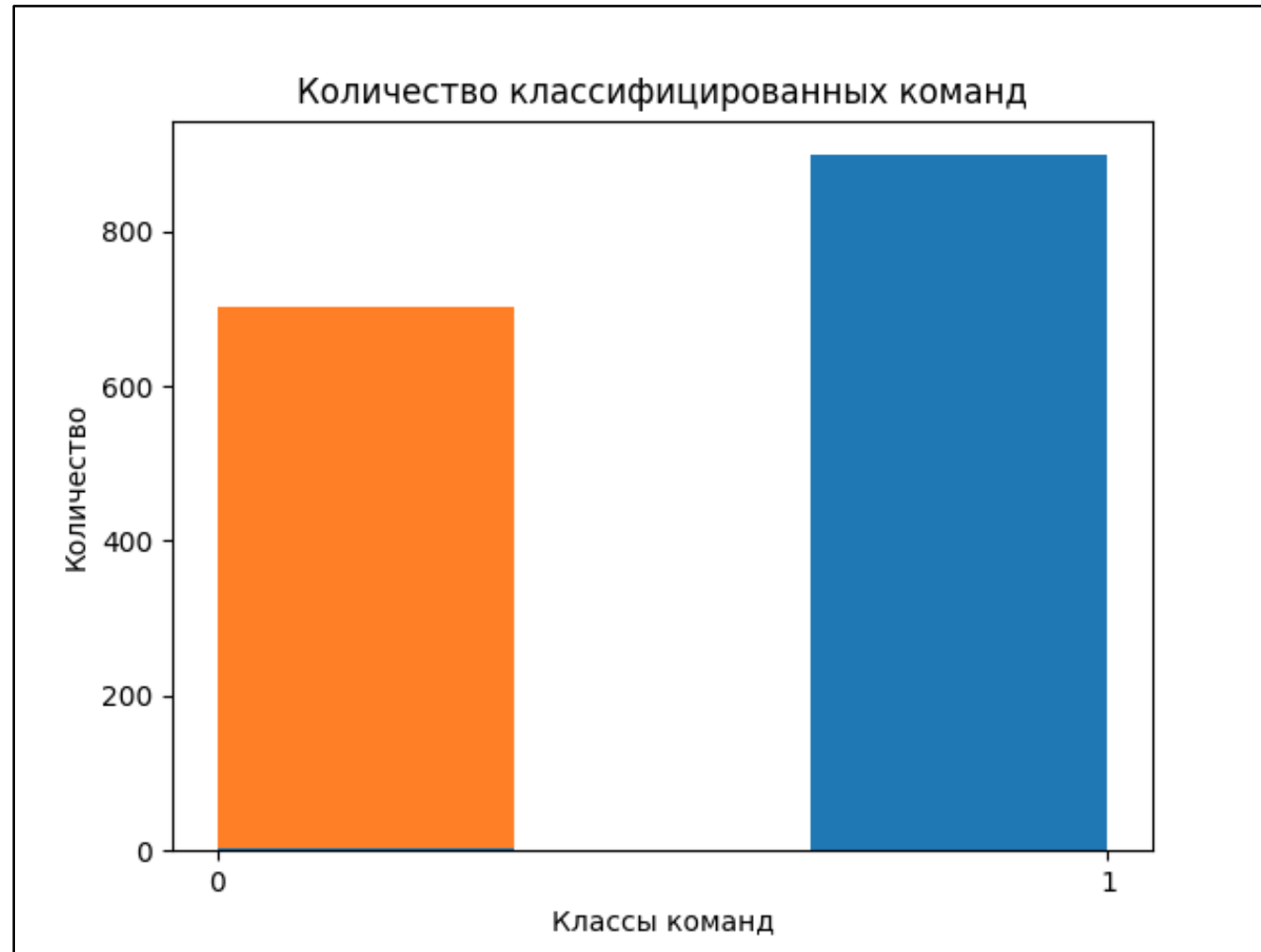
1. Проверка на наличие пустых значений и оценка типов признаков

2. Трансформация признаков

Итог:

- Количество записей: 1 600

РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ ПОСЛЕ ПРЕДОБРАБОТКИ



АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ



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

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)

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

РАБОТА ПРИЛОЖЕНИЯ (1)



```
o 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
```

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

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
```

РАБОТА ПРИЛОЖЕНИЯ (4)



Activities Files 45.9°C 61.3°C июн 6 21:20:46 en 98 %

Home / PycharmProjects / DataAnalytics / CourseWork / dist / mitre_main

Name	Size	Modified
answers.csv	30,9 kB	21:19 ☆
base_library.zip	1,1 MB	18:45 ☆
bayes_model.sav	636,7 kB	21:19 ☆
_cffi_backend.cpython-310-x86_64-linux-gnu.so	983,8 kB	12 окт 2022 ☆
decision_tree_model.sav	8,8 kB	21:19 ☆
libBLT.2.5.so.8.6	1,5 MB	23 мар 2022 ☆
libbrotlicommon.so.1	137,6 kB	23 мар 2022 ☆
libbrotlidec.so.1	51,5 kB	23 мар 2022 ☆
libbsd.so.0	89,1 kB	25 янв 2022 ☆
libbz2.so.1.0	74,8 kB	23 мар 2022 ☆
libcrypto.so.3	4,4 MB	24 мая ☆
libexpat.so.1	194,9 kB	18 ноя 2022 ☆
libffi.so.8	47,7 kB	17 янв 2022 ☆
libfontconfig.so.1	298,1 kB	23 мар 2022 ☆
libfreetype.so.6	813,1 kB	2 мая ☆
libfreetype-9b226f05.so.6.18.3	1,4 MB	13 окт 2022 ☆
libgcc_s.so.1	125,5 kB	13 мая 2022 ☆
libgfortran-040039e1.so.5.0.0	2,7 MB	13 окт 2022 ☆
libgomp-a34b3233.so.1.0.0	168,2 kB	8 ноя 2022 ☆
libharfbuzz-2a5452dc.so.0.40401.0	3,0 MB	13 окт 2022 ☆
libjpeg-a4c3d5e9.so.62.3.0	687,3 kB	13 окт 2022 ☆

ФУНКЦИОНАЛЬНОЕ ТЕСТИРОВАНИЕ



№	Название теста	Шаги	Ожидаемый результат	Тест пройден?
1	Выбор файла	<ol style="list-style-type: none">В меню ввести цифру 1.Ввести абсолютный путь до файла.	Программа должна вывести первые несколько строчек файла.	Да
2	Выбор одной модели	<ol style="list-style-type: none">В меню ввести цифру 2.Ввести цифру 1.С помощью цифр 1-4 выбрать алгоритм.	Программа должна вывести результаты в виде таблицы	Да
3	Выбор двух и более моделей	<ol style="list-style-type: none">В меню ввести цифру 2.Ввести цифру от 2 до 4.С помощью цифр 1-4 поочередно выбрать алгоритмы.	Программа должна вывести результаты в виде таблицы	Да
4	Ввод неверных параметров	<ol style="list-style-type: none">В меню выбора интерфейса ввести значение не из списка доступных интерфейсов.	Программа должна вывести сообщение об ошибке «Выберете заново».	Да

А/В 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.

30.05.2023 г.



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



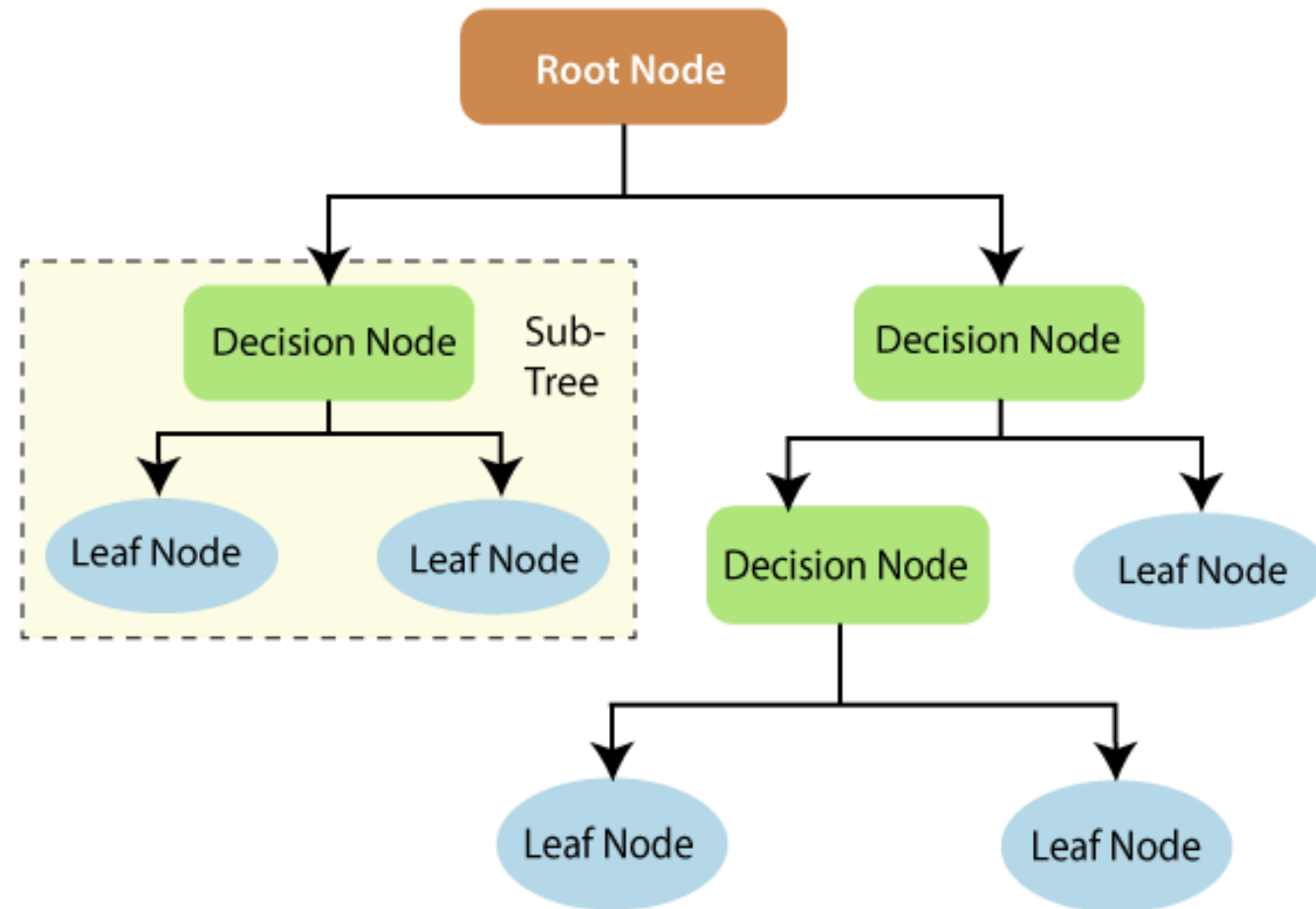
Вид команды	Команды MITRE
Execution	<code>\$assembly = [Ref].Assembly.GetType('{0}{1}i{2}' -f \$a,\$b,\$u))</code>
Command and Control	<code>\$field.SetValue(\$null,\$true)</code>
<u>Discovery</u>	<code>\$ping = New-Object System.Net.Networkinformation.Ping</code>
Privilege Escalation	<code>\$computer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

NAIVE BAYES

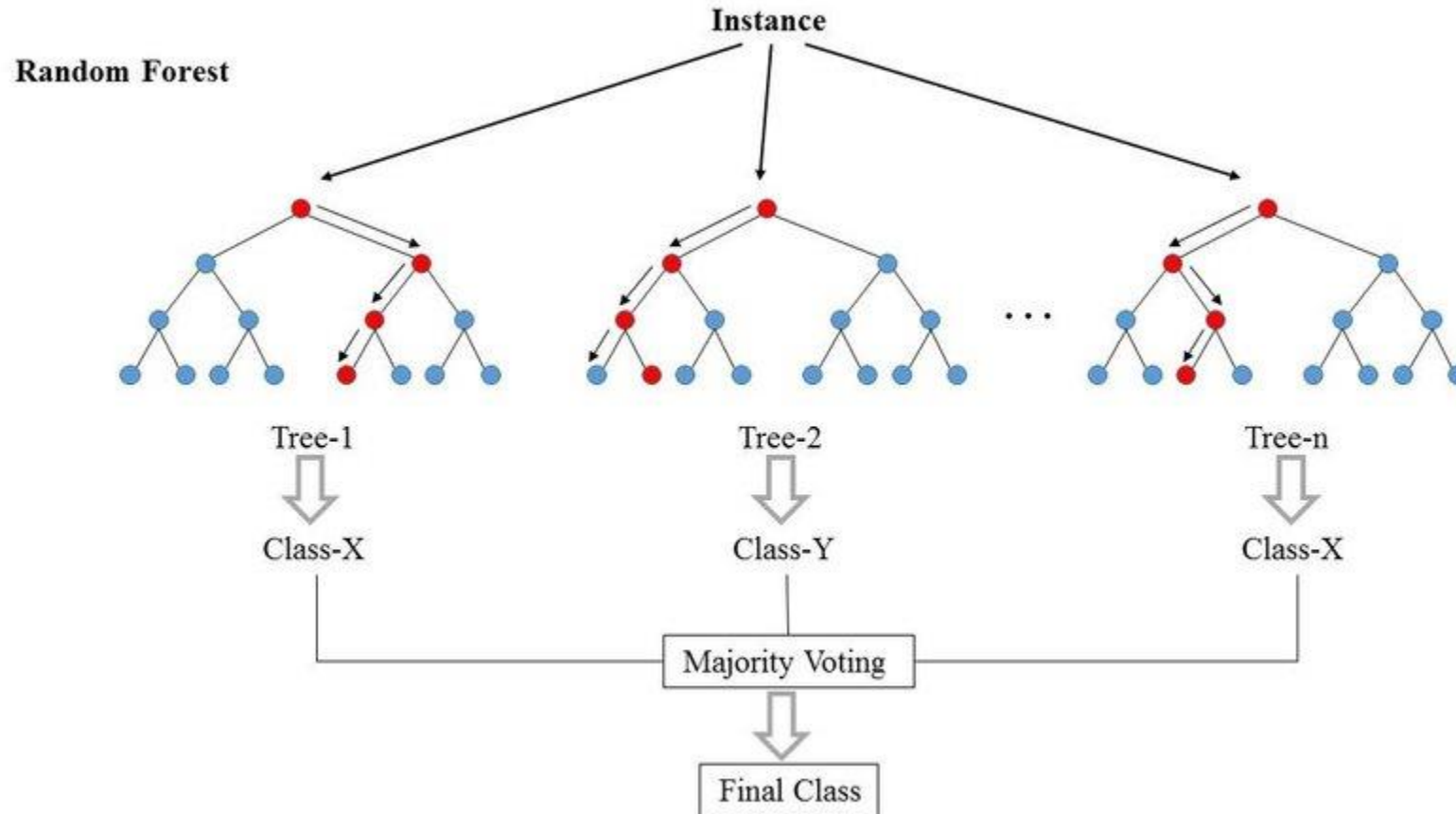


$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

DECISION TREE



RANDOM FOREST



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

LOGISTIC REGRESSION

