

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ ФИШИНГОВЫХ САЙТОВ

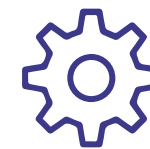
Зверков Сергей
Группа 743-1
Email: sergey_zverkov@vk.com

Цели и задачи

Целью данной курсовой работы является разработка и исследование системы обнаружения фишинговых сайтов на основе современных методов глубокого обучения. Особое внимание уделяется комплексному анализу как текстового содержимого веб-страниц, так и структурных характеристик URL-адресов.

Задачи:

- Исследовать тематику;
- Собрать набор данных;
- Разработать модель нейронной сети;
- Обучить нейронную сеть;
- Проанализировать результаты.

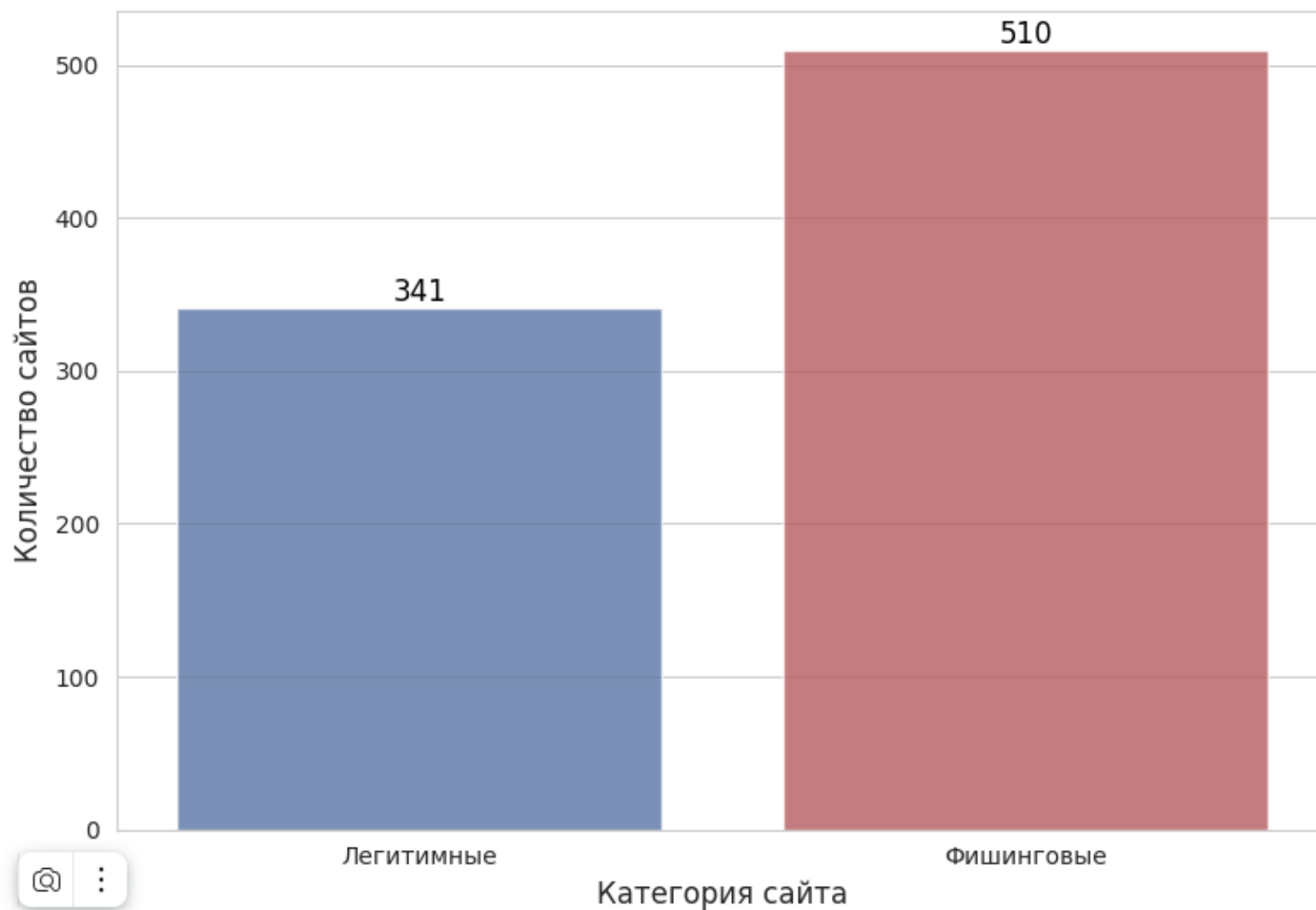


Набор данных

URL	Text Content	Label	url_length	num_subdomains	has_hyphen	has_https	domain_length	phishing_keywords_in_url	phishing_keywords_in_text	text_length
http://eb177.cc/	eb177.cc Не удается получить доступ к сайту Пр...	1	16	0	False	False	8	False	False	937
odaddysites.com/	Metamask Login usametameskloggin usametamesk...	1	42	1	False	False	34	False	True	658
http://pyfrcz.top/	WhatsApp Group Invite WhatsApp Group Invite Ho...	1	18	0	False	False	10	False	False	79
-theta.vercel.app/	Amazon Sign In Account & Lists Returns & Order...	1	44	1	True	False	36	False	True	7840
odaddysites.com/	404 Not Found File not found (404 error) If yo...	1	42	1	True	False	34	False	False	124
iplieradvisor.com/	Вход в учетную запись Войти Нет учетной записи...	1	39	1	True	False	31	False	False	187
iy.fuckgwf.eu.org/	hy.fuckgwf.eu.org Не удается получить доступ к...	1	25	2	False	False	17	False	False	964
m/ltv/owacrt.html	Outlook Email address: Password: sign in The u...	1	40	1	False	True	16	False	False	156
es-vless.bavi-far...	海外频道_央视网(cctv.com) 新闻 熊猫 熊猫一刻 熊猫星团 大熊猫故事 大熊猫国 家...	1	62	2	True	False	54	False	False	3130
https://pub-70802fa2aaeb1.r...	Outlook Web App To use Outlook Web App, browse...	1	74	1	True	True	43	False	False	575
ef365.3656aak.cc/	bet365 -No.1 体育投注, 世界锦标赛斯诺克赔率, 百家乐娱乐场, 扑克牌, 电 子游戏 系...	1	24	1	False	False	16	False	False	7206
odaddysites.com/	404 Not Found File not found (404 error) If yo...	1	40	1	False	False	32	False	False	124
blond.vercel.app/	Instagram c Home Search Explore Reels Message ...	1	45	1	True	False	37	False	False	1031
-clone.vercel.app/	Facebook – log in or sign up Facebook helps yo...	1	33	1	True	False	25	True	True	157
odaddysites.com/	Phantom Extension – How to connect it with Led...	1	41	1	False	False	33	False	False	942

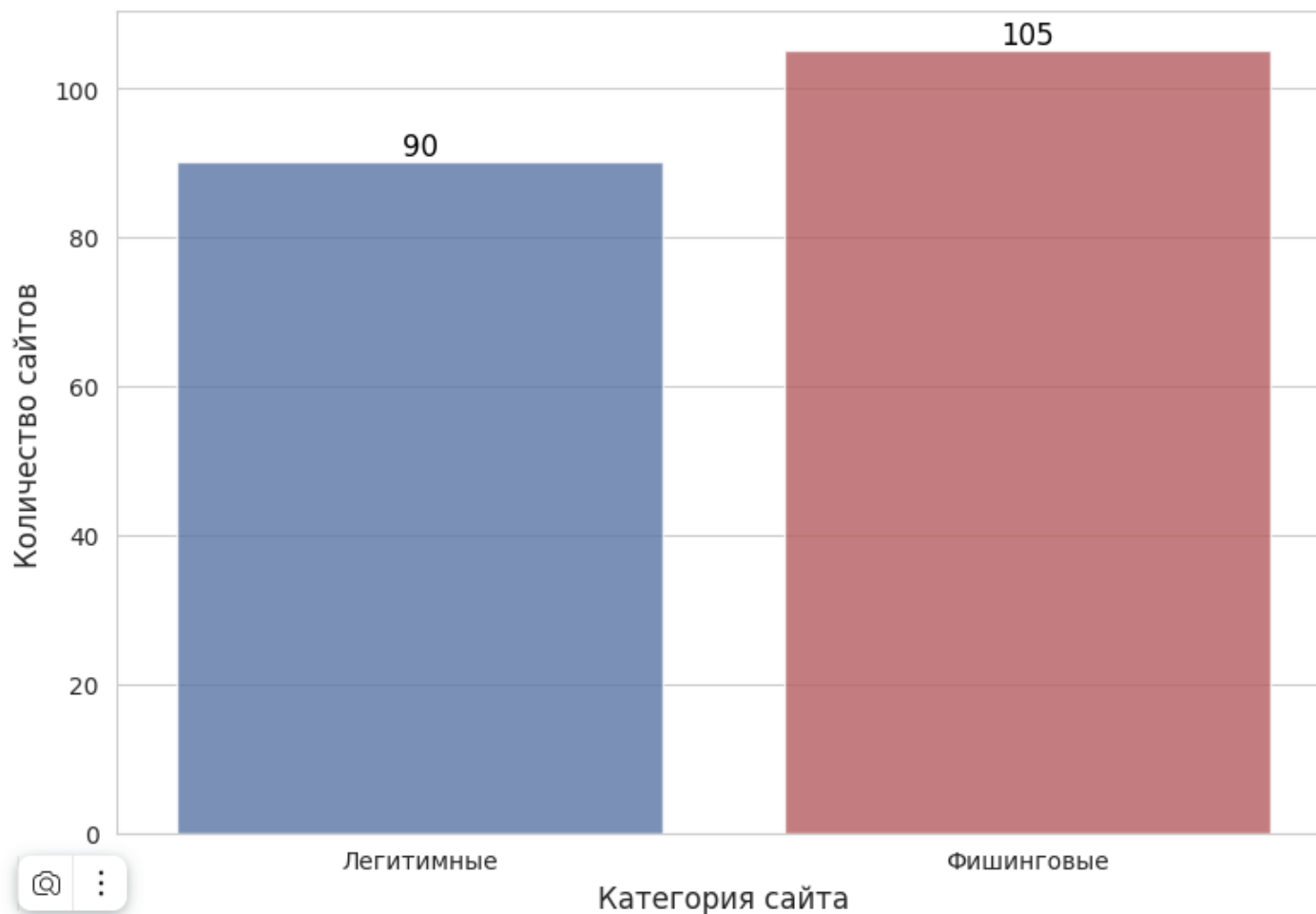
Размер набора данных

Распределение фишинговых и легитимных сайтов



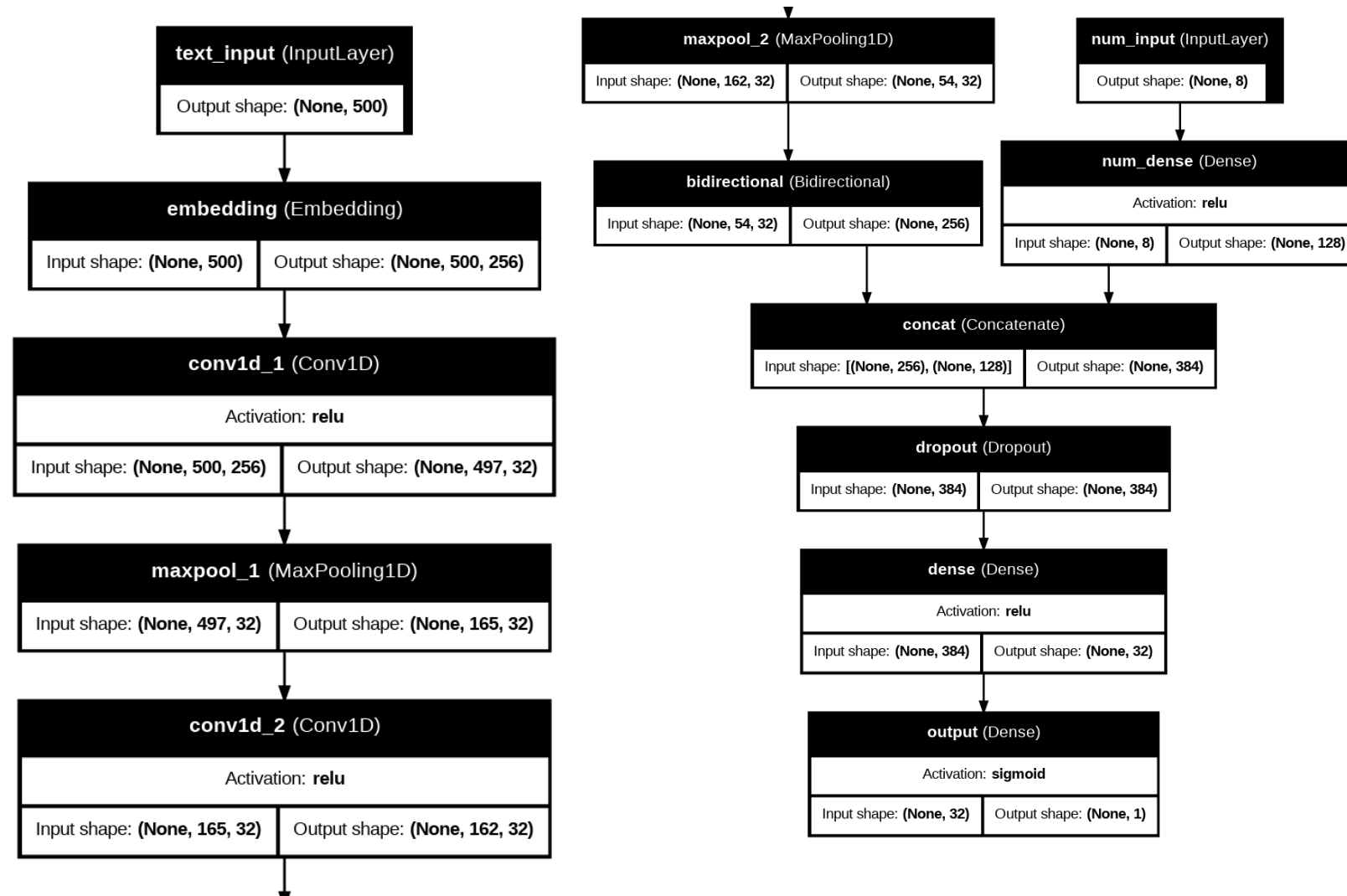
Набор данных для тестирования модели

Распределение фишинговых и легитимных сайтов в тестовом наборе



Архитектура модели

Архитектура представляет собой гибридную модель, которая объединяет обработку текстовых данных (через CNN + Bidirectional LSTM) и числовых признаков (через полносвязные слои).



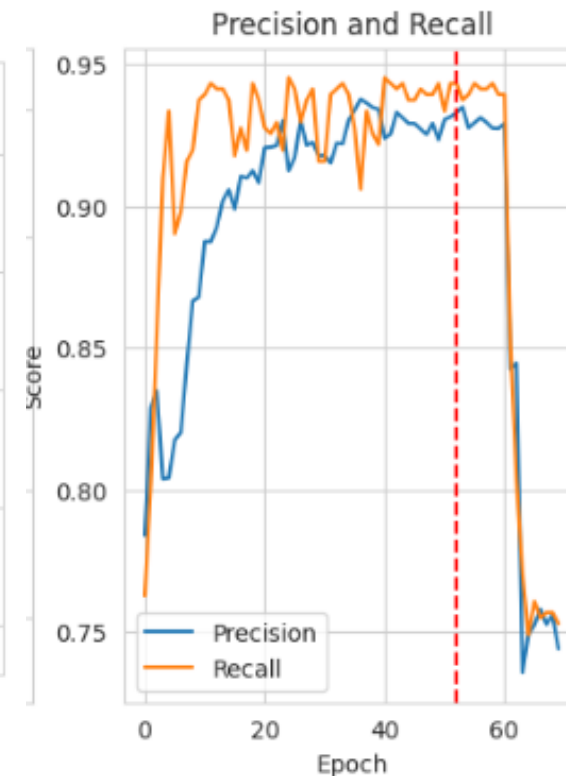
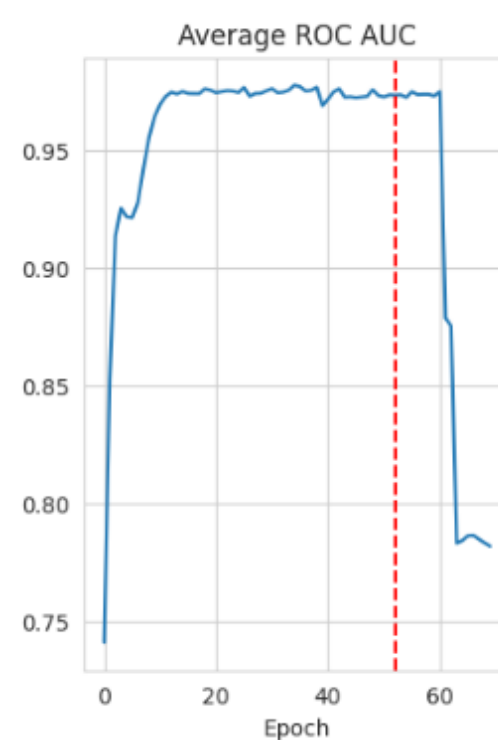
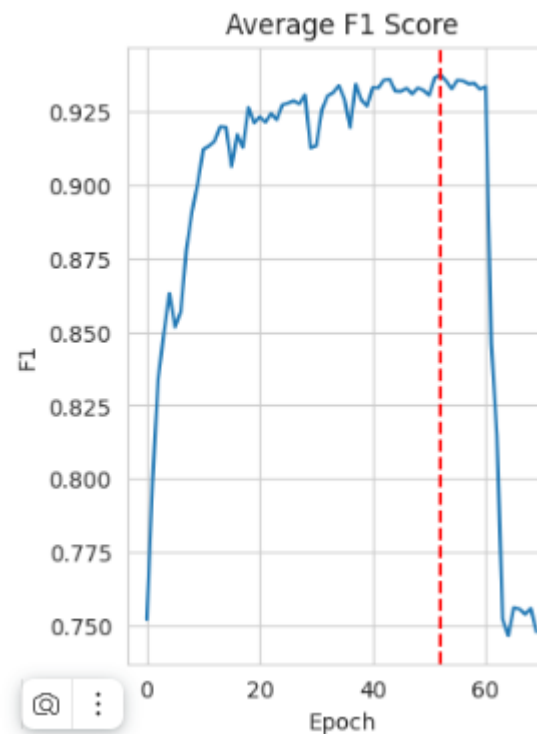
Оценка качества модели

F1-Score составило 0.925

ROC AUC составило 0.95

Precision ≈ 0.95 — означает, что 95% сайтов, классифицированных как фишинговые, действительно являются таковыми.

Recall ≈ 0.90 — показывает, что модель обнаруживает 90% всех реальных фишинговых сайтов.



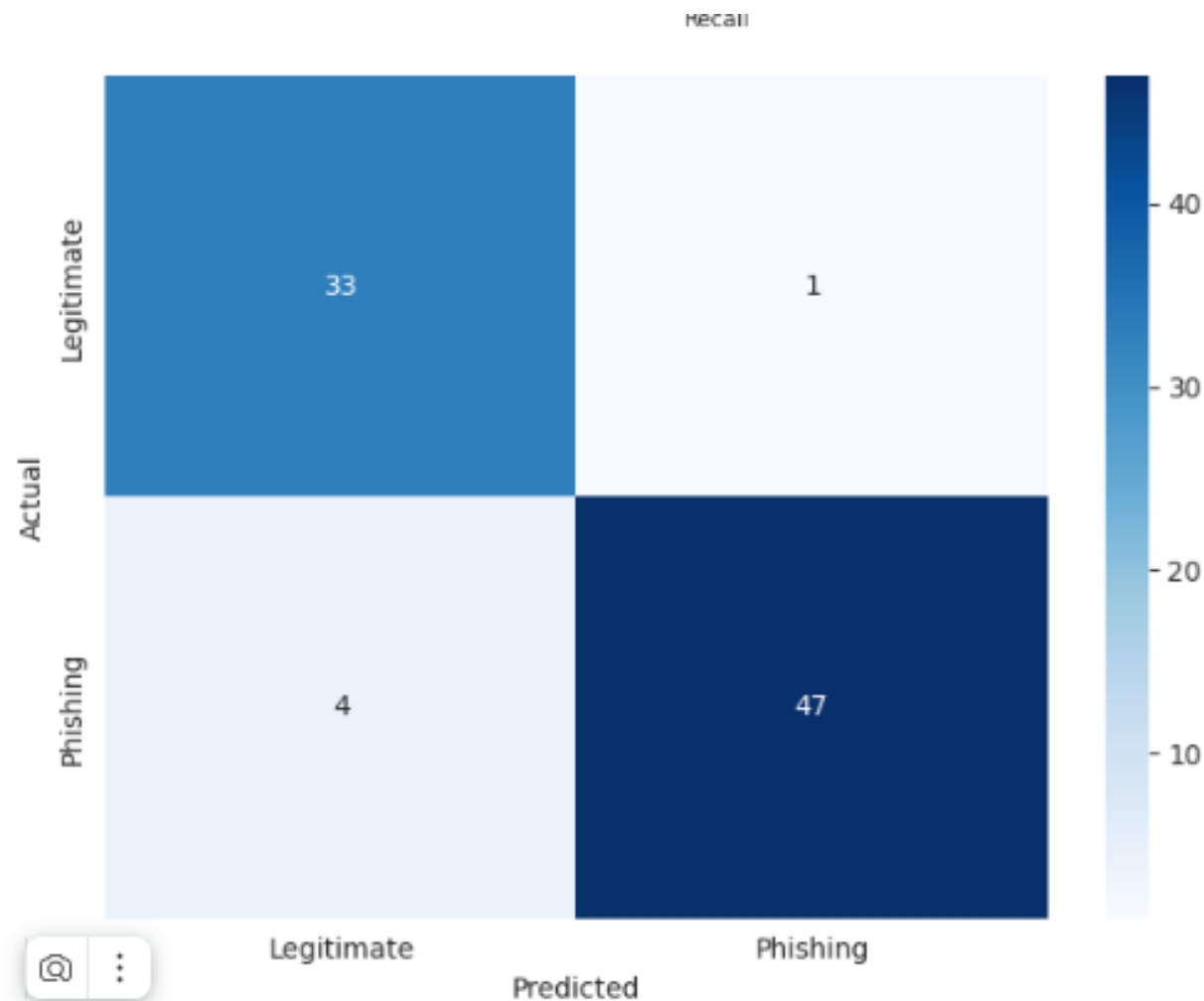
Матрица ошибок

Легитимные сайты (Legitimate):

- Правильно классифицировано: 33 (True Negative);
- Ошибочно помечены как фишинг: 1 (False Positive).

Фишинговые сайты (Phishing):

- Правильно классифицировано: 47 (True Positive);
- Ошибочно помечены как легитимные: 4 (False Negative).



Отчет по метрикам для каждого класса

	precision	recall	f1-score	support
Legitimate	1.00	0.97	0.99	34
Phishing	0.98	1.00	0.99	51
accuracy			0.99	85
macro avg	0.99	0.99	0.99	85
weighted avg	0.99	0.99	0.99	85

Тестирование модель на новом наборе данных

	precision	recall	f1-score	support
Легитимный	0.81	0.73	0.77	90
Фишинг	0.79	0.86	0.82	105
accuracy			0.80	195
macro avg	0.80	0.80	0.80	195
weighted avg	0.80	0.80	0.80	195

ROC AUC Score: 0.8232

F1 Score: 0.8219

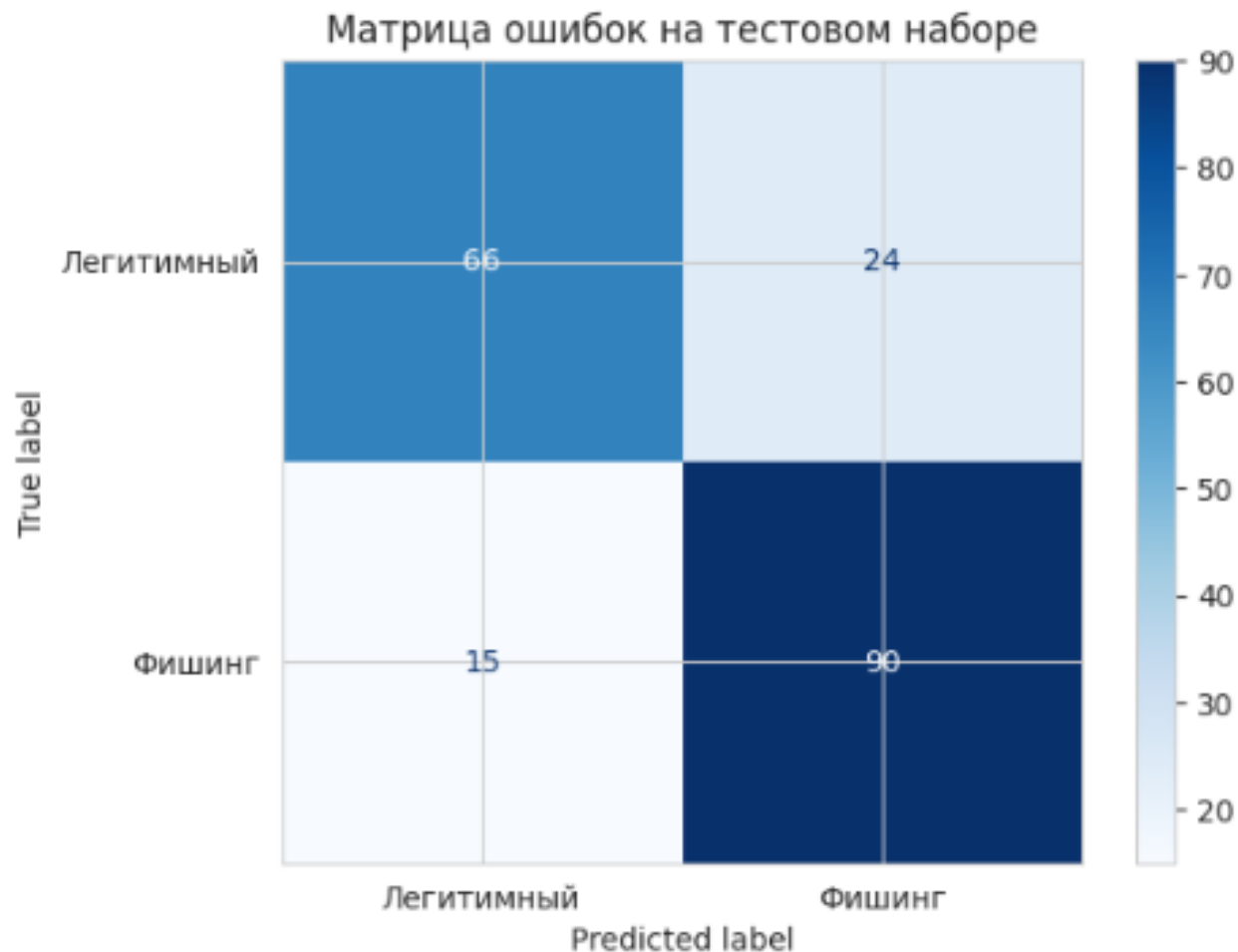
Тестирование модель на новом наборе данных

Легитимные сайты (Legitimate):

- Правильно классифицировано: 66 (True Negative);
- Ошибочно помечены как фишинг: 24 (False Positive).

Фишинговые сайты (Phishing):













- Правильно классифицировано: 90 (True Positive);
- Ошибочно помечены как легитимные: 15 (False Negative).



Классификация ссылок

На слайде представлено 6 случайных ссылок из датасета (3 легитимных и 3 фишинговых) и проверяет, правильно ли модель их классифицировала.

URL	Предсказание	Верно?	Истинный класс

1/1  0s 326ms/step http://www.ug.ru/ Легитимный (0.0132) Да			Легитимный
1/1  0s 32ms/step http://www.vaco.ru/ Легитимный (0.0130) Да			Легитимный
1/1  0s 32ms/step http://www.detlit.ru/ Легитимный (0.0129) Да			Легитимный
1/1  0s 32ms/step http://help-fb-recovery-center.github.io/notification Фишинговый (0.5279) Да			Фишинговый
1/1  0s 33ms/step http://maha-git55.github.io/instagrame Фишинговый (0.8523) Да			Фишинговый
1/1  0s 32ms/step https://tinfrnnjaw.com/ Фишинговый (0.8438) Да			Фишинговый
1/1  0s 32ms/step			
1/1  0s 31ms/step			
1/1  0s 33ms/step			
1/1  0s 33ms/step			
1/1  0s 32ms/step			
1/1  0s 33ms/step			
Итоговая точность на выбранных примерах: 6/6 (100.00%)			

Спасибо за внимание!



Зверков Сергей
группа 743-1
Email: sergey_zverkov@vk.com