



מציגים: איל רון וגל ארז
מנחה: ד"ר אריק פארן
המרכז האקדמי רופין



מבוא

פשינג הוא איום מרכזי על אבטחת מידע בעידן הדיגיטלי. בשנת 2023 אובחנו מעל 1.2 מיליארד ניסיונות פשינג, עם נזק כלכלי עולמי של 20 מיליארד דולר בשנה. 90% ממתקפות הסייבר מתחילות בפשינג, ו-97% מהמשתמשים מתקשים לזהות הודעות מתוחכמות – דבר המדגיש את הצורך בפתרונות מתקדמים להתמודדות עם האיום.



המוטיבציה לפרויקט

פרויקט PhishGuard AI שואף לספק פתרון אוטומטי ומהיר לזיהוי ומניעת פשינג, ולהגן על מידע רגיש. המוטיבציה לפרויקט נובעת מהתחכום הגובר של מתקפות הפשינג, חוסר היעילות של פתרונות ידניים, והצורך בהגנה בזמן אמת. הפרויקט משלב טכנולוגיות מתקדמות כמו NLP ולמידת מכונה לשיפור דיוק הזיהוי.



האתגר

האתגר המרכזי נובע מהתחכום הגובר של איומי הפישינג, עם שיטות הונאה המתפתחות במהירות והקושי בזיהוי ידני של הודעות חשודות. מתקפות אלו גורמות לנזק כלכלי משמעותי ומסכנות מידע רגיש של משתמשים וארגונים.





האתגר

מאגר המחקר שלנו כלל 18,588 הודעות דוא"ל, מתוכן 8,000 הודעות פישניג מאומתות ו-10,588 הודעות לגיטימיות. בארגון ממוצע מתקבלות אלפי הודעות דוא"ל ביום, וזמן הבדיקה הידנית הממוצע לכל הודעה הוא 2-3 דקות. בשל היקף העבודה הנדרש, סינון ידני אינו אפשרי באופן יעיל.



למה נדרש פתרון אוטומטי?

- ריבוי הודעות דורש עיבוד מהיר
- צורך בזיהוי בזמן אמת
- חיסכון משמעותי בזמן ומשאבים
- דיוק גבוה יותר מבדיקה אנושית
- מניעת טעויות הנובעות מעייפות ושחיקה
- הגנה רציפה 24/7



הפתרון - PhishGuard AI

PhishGuard AI מציע זיהוי חכם של הודעות פשינג באמצעות NLP עם דיוק של 97.62%. המערכת כוללת ממשק ידידותי למשתמש, התראות מיידיות ויכולת עיבוד של מספר הודעות במקביל. בעתיד, המערכת תשפר את יכולות הזיהוי שלה באמצעות למידה מתמדת.



מטרות הפרויקט (מדו"ח איפיון)

- **זיהוי אלמנטים מרמזים:** איתור דפוסים בשפה ובאסטרטגיות פסיכולוגיות המשמשות בהודעות פשינג, כדי לזהות טקסטים וקישורים חשודים.
- **דיוק בזיהוי:** המערכת שואפת להשיג דיוק גבוה בזיהוי פשינג, תוך שימוש באלגוריתמים שמזהים תבניות התקפה ומסווגים אותן כניסיונות פשינג במטרה למנוע חשיפה למידע רגיש.
- **מהירות תגובה:** פיתוח מערכת לזיהוי מהיר של פשינג כדי לספק הגנה מיידית.



שלב הפיתוח



איסוף וארגון הנתונים

בשלב הראשוני בפרויקט PhishGuard AI התמקדנו באיסוף נתונים לאימון המודלים לזיהוי פישנינג. לאחר חיפוש במאגרים שונים, בחרנו במאגר של פלטפורמת Hugging Face בשל נגישותו, קלות העיבוד, והרלוונטיות של הנתונים. המאגר כלל 18,000 הודעות דוא"ל, מהן כ-8,000 הודעות פישנינג והשאר בטוחות.




איסוף וארגון הנתונים

הנתונים שנאספו דרשו ניקוי וארגון לפני השימוש במודלי למידת מכונה. התהליך בוצע ב-Jupyter Notebook עם ספריית Pandas וכלל הסרת ערכים חסרים ורווחים מיותרים, והמרת סיווג ההודעות למספרים ("פשינג" ל-1 ו"בטוח" ל-0). לאחר הניקוי, הנתונים נשמרו כקובץ CSV לשימוש באימון המודלים, כדי להבטיח איכות ואמינות ביכולת ההבחנה בין הודעות פשינג לבטוחות.



D	C	B	A	
	Safe Ema ecology or	920	934	
	Phishing E please rea	929	933	
	Safe Ema article : ' b	930	934	
	Phishing Ere : q&w&g	931	935	
	Phishing E urgent sta	932	936	
	Phishing E	933	937	
	Safe Ema final sched	934	938	
	Phishing E Here It	935	939	
	Safe Ema gtcs , form	936	940	
	Phishing E buy med a	937	941	
	Safe Ema year end 2	938	942	
	Safe Ema negative c	939	943	
	Safe Ema corpuyr lir	940	944	
	Safe Ema new nat ga	941	945	
	Phishing E Me and	942	946	
	Safe Ema Not	943	947	
	Safe Ema linguistics	944	948	
	Phishing E how is you	945	949	
	Safe Ema associate	946	950	
	Phishing E stern trillio	947	951	
	Safe Ema re : boat i	948	952	
	> A	949	953	
		its stated	954	
		common r	955	
	and Our f	Our Allies opportuniti	956	
		with Wear	957	
		The grave	958	
		technology	959	
		along with weapons	960	
		even weak	961	
		and strike grea	962	
		have been	963	
	or to harm	or to harm capability	964	
		we will op	965	
		New York West Poir	966	
	2002The	Jun-01	967	
		allies and	968	
		producing force	969	



B	A	
Email Typ	Email Text	1
	1	2
	1 software at incredibl	3
	1 entourage , stockmo	4
	1 we owe you lots of m	5
	1 make her beg you to	6
	1 formal invite for chas	7
	1	8
	1	9
	1	10
	1 lowers blood pressu	11
	1 <!--	12
	1 premium adult conte	13
	1 25 mg trick how to s	14
	1 Help wanted. We	15
	1 important message	16
	1 BUY 2 ADULT	17
	1 gino , who do u want	18
	1 unbelievable new ho	19
	1 discover the new wit	20
	1 empty	21
	1 free portable dvd pla	22
	1 urgent mr . johnson	23
	1 refinancing has neve	24
	1 your in - home sourc	25
	1 Digital Publishing	26
	1 hi again are story pe	27
	1 re : hi . . . y - 0 - u - r	28
	1	29
	1 "Now	30
	1 confidence attn : ma	31
	1 current anaiysis on f	32
	1 re [11] bands leona	33

איסוף וארגון הנתונים



תיאור המערכת לחילוץ נתונים

בשלב הבא פותחה מערכת לחילוץ נתונים מתוך הודעות דוא"ל, שמעשירה את תהליך עיבוד הנתונים עבור אימון המודלים לזיהוי פשינג. המערכת מפיקה מאפיינים ונתונים סטטיסטיים מהטקסט, תוך שימוש בספריות כמו NLTK לניתוח תחבירי, זיהוי ישויות ותבניות לשוניות, וניתוח סנטימנט. תהליך זה מאפשר הבנה מעמיקה יותר של מבנה ההודעות ותומך בזיהוי תבניות המאפיינות הודעות פשינג לעומת הודעות לגיטימיות.



תיוג ועיבוד הנתונים

במהלך תהליך תיוג ועיבוד הנתונים, הופקו מאפיינים (Features) ייחודיים מהודעות הדוא"ל כדי לשפר את יכולת זיהוי הפישינג של המודלים. הפרמטרים שנבחרו:

- כמות המילים
- כמות כתובות URL
- כמות כתובות דואר אלקטרוני
- כמות המשפטים
- אורך ממוצע של משפטים
- גיוון לשוני
- תדירות בי-גרמים
- ניתוח תחבירי
- זיהוי ישויות
- ניתוח סנטימנט
- ציון קריאות (readability)
- מספר תווים מספריים
- זיהוי כתובות IP
- נוכחות תווים מיוחדים, כגון '@' ו-'

Text Analysis Tool

Text Analysis Tool

Input Text:

angry!

URL/Email address:

eyal@gal.com

Analyze

Save to CSV

Load File

Process CSV

Analysis Results

Word Count: 2 (Total number of words)

URL Count: 0 (Total number of URLs found)

Email Count: 1 (Total number of email addresses found)

Sentence Count: 1 (Total number of sentences)

Average Sentence Length: 2.00 (Average number of words per sentence)

Lexical Diversity: 1.00 (Proportion of unique words to total words)

Sentiment Score: -0.5562 (Score indicating the emotional tone; closer to 1 is more positive)

Readability Score: 77.91 (Flesch Reading Ease score; higher is easier to read)

NumDots: 0 (Number of periods in the text)

UrlLength: 12 (Length of any URLs in characters)

NumDash: 0 (Number of dashes in the text)

AtSymbol: Not present (Presence of @ symbol)

NumNumericChars: 0 (Number of numeric characters in the text)

IpAddress: Not present (Presence of an IP address)

Most Common Bigrams: I angry: 1; angry !: 1 (The five most common two-word combinations and their frequencies)

תיוג ועיבוד הנתונים

AA		Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A	h																
		()	&	%	\$	#	^	!	Readabil	Sentiment	Named E	POS	Tag	Lexical	Di	Ip	Address	Random	S	Num	Num	At	Symbol	Num	Dash	Most	Com	Url	Length	Num	Dots	Average	S	Sentence	Email	Col	URL	Cour	Word	Cou			
0	1	1	0	0	0	0	0	0	0	1	74.59	0.964	Helio	NP	Helio	NNP	0.822917	FALSE	N/A	29	FALSE	14	I am	4	for	N/A	36	9.090909	11	0	0	0	0	96	2									
0	1	1	0	0	0	1	0	0	0	0	74.96	0.0534	software	-	software	0.714286	FALSE	N/A	2	FALSE	0	software	4	N/A	13	6	13	0	0	0	0	0	91	3										
17	16	9	0	4	8	0	0	0	0	30	5	41.09	0.939	entourage	-	entourage	0.434249	FALSE	N/A	130	FALSE	11	9	gen	N/A	49	23	1429	56	0	0	0	0	1384	4									
0	0	0	0	0	1	2	0	0	0	0	63.7	0.2124	we	PRP	-	0.605634	FALSE	N/A	12	FALSE	24	-	22	for	N/A	8	16	75	8	0	0	0	0	142	5									
0	0	0	5	0	2	0	0	0	0	6	67.28	0.9961	make	VB	make	VB	0.559603	FALSE	N/A	11	FALSE	23	-	no	4	rx	N/A	5	26	45455	11	0	0	0	0	302	6							
0	0	0	0	0	0	0	0	1	2	2	74.15	0.6229	formal	JJ	formal	JJ	0.826923	FALSE	N/A	5	FALSE	0	here	2	1	N/A	5	6	428571	7	0	0	0	0	52	7								
0	0	0	0	0	0	0	0	0	0	0	60.01	0.9469	GPE	Que	Question	0.716814	FALSE	N/A	13	FALSE	0	Do	4	D	N/A	8	7	14286	14	0	0	0	0	113	8									
0	0	0	1	0	0	0	0	0	0	7	72.97	0.5108	PROMOT	-	PROMOT	0.687831	FALSE	N/A	13	FALSE	14	4	4	4	N/A	3	19	4	10	0	0	0	0	189	9									
0	0	0	0	0	0	0	0	0	0	0	206.84	0	0	FALSE	N/A	0	FALSE	N/A	0	FALSE	0	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	2	0	0	0	0	0	0	0	-76.38	-0.8469	lowers	NN	lowers	NN	0.714286	FALSE	N/A	0	FALSE	1	's	2	to	yc	N/A	0	0	0	0	0	0	0	0	161	11							
29	29	0	16	0	0	0	0	1	4	9	-34.32	0	-NN	V	-	0.795181	FALSE	N/A	21	FALSE	5		8	f	(N/A	45	2	509091	55	0	0	0	0	83	12								
0	1	1	3	0	0	0	0	0	0	0	61.29	0.8402	premium	-	premium	0.677083	FALSE	N/A	7	FALSE	2	hardcore	N/A	2	31	3	0	0	0	0	0	0	96	13										
0	0	0	0	0	0	0	0	0	0	1	62.54	0.9041	25	CD	mx	25	CD	mx	0.813187	FALSE	N/A	10	FALSE	1	25	mg	1	N/A	8	9	111111	9	0	0	0	0	91	14						
0	0	0	0	0	0	0	0	0	0	0	59.9	0.9477	PERSON	-	Help	NNP	0.728155	FALSE	N/A	27	FALSE	2	We	3	a	N/A	13	8	538462	13	0	0	0	0	103	15								
0	0	0	2	0	0	0	0	0	0	2	50.06	0.9697	important	-	important	0.652174	FALSE	N/A	11	FALSE	3	and	the	3	N/A	2	25	28571	7	0	0	0	0	184	16									
0	0	0	0	0	0	0	0	0	0	2	76.22	0	BUY	NNP	-	0.972222	FALSE	N/A	8	FALSE	0	BUY	2	1	N/A	2	13	3	0	0	0	0	36	17										
0	0	0	0	0	0	0	0	0	0	0	82.61	0.8588	gino	NN	-	0.857143	FALSE	N/A	4	FALSE	0	here	2	(N/A	3	7	4	5	0	0	0	0	42	18									
0	0	0	1	0	0	0	1	0	0	1	72.56	0.9523	unbeliev	-	unbeliev	0.793103	FALSE	N/A	10	FALSE	1	unbeliev	N/A	3	20	75	4	0	0	0	0	87	19											
0	0	0	1	0	0	0	0	0	0	1	70.19	-0.126	discover	v	discover	v	0.803571	FALSE	N/A	0	FALSE	0	a	2	a	pe	N/A	4	10	2	5	0	0	0	0	56	20							
0	0	0	0	0	0	0	0	0	0	0	36.62	-0.2023	empty	JJ	empty	JJ	-	FALSE	N/A	0	FALSE	0	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	0	0	0	0	0	0	0	0	65.63	0.9571	free	JJ	potter	JJ	po	0.702703	FALSE	N/A	26	FALSE	0	of	the	2	N/A	9	15	44444	9	0	0	0	0	148	22							
3	3	1	0	2	2	0	0	0	0	0	55.74	-0.5632	urgent	JJ	urgent	JJ	0.4576	FALSE	N/A	31	FALSE	71	-	70	mr	N/A	32	17	02857	35	0	0	0	0	625	23								
2	2	7	0	0	0	0	0	1	2	14	78.48	0.9957	refinancin	-	refinancin	0.332797	FALSE	N/A	15	TRUE	14	-	186	N/A	30	12	23404	47	0	0	0	0	622	24										
0	0	0	0	0	0	0	0	0	0	0	67.76	0.8253	your	PRP	your	PRP	0.851852	FALSE	N/A	0	FALSE	1	nothing	to	N/A	4	12	5	0	0	0	0	0	54	25									
0	0	0	0	0	0	0	0	0	0	0	54.32	0.9625	CRICANZ	-	Digital	NN	0.808819	FALSE	N/A	14	FALSE	5	Digital	Put	N/A	5	17	8	0	0	0	0	176	26										
0	0	0	0	0	0	0	0	0	0	0	99.53	0.4215	hi	NN	aga	hi	NN	aga	0.740741	FALSE	N/A	20	TRUE	6	hi	again	1	N/A	10	7	1	10	0	0	0	0	81	27						
7	0	0	0	0	0	0	0	0	0	0	95.17	0.5423	re	NN	-	0.57377	FALSE	N/A	4	TRUE	4)	4	;	3	N/A	7	7	714286	7	0	0	0	0	61	28								
1	1	1	0	0	0	1	0	0	0	7	76.82	0.9717	Are	NNP	-	0.753333	FALSE	N/A	4	FALSE	4	1	We	3	A	N/A	2	11	92308	13	0	0	0	0	150	29								
0	0	0	0	0	9	0	0	0	0	6	78.45	0.9851	'	JJ	Now	'	JJ	Now	0.737327	FALSE	N/A	55	FALSE	3	credit	car	N/A	38	6	419667	36	0	0	0	0	217	30							
1	1	1	0	0	1	1	0	0	0	0	64.36	0.8662	confidence	-	confidence	0.431655	FALSE	N/A	31	FALSE	0	;	7	now	to	N/A	14	26	47868	16	0	0	0	0	556	31								



הניסויים שבוצעו



אלגוריתמים שנבחנו

רגרסיה לוגיסטית – מתאים לסיווג בינארי ומספק משקלים פרשניים לכל תכונה.

KNN – סיווג מבוסס שכנים קרובים, מתאים ללא הנחות על התפלגות הנתונים אך דורש שמירת כל הנתונים.

SVM – בונה היפר-מישורים להפרדת מחלקות, יעיל במיוחד בנתונים לא ליניאריים.

Random Forest – מבוסס על אנסמבל של עצי החלטה, מספק יציבות ומונע התאמת יתר.

Gradient Boosting – מזהה דפוסים מורכבים ומתקן טעויות, אך דורש זמן חישוב רב.

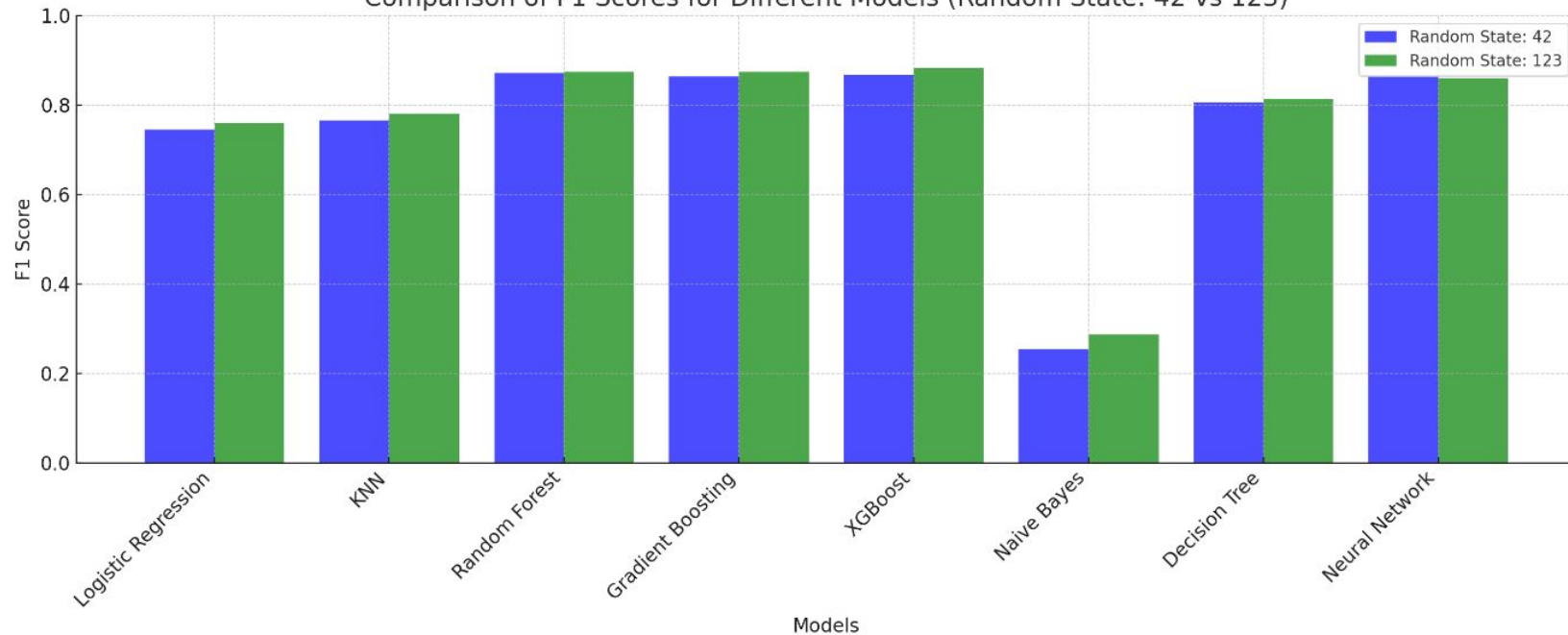
XGBoost – גרסה משופרת ומהירה של Gradient Boosting, מתאים לנתונים גדולים.

נאיב בייס – מהיר ויעיל בטקסטים, מתבסס על הנחת עצמאות בין תכונות.

עץ החלטה – פשוט להבנה ופרשנות, אך עלול להטות יתר בעצים עמוקים.

רשת נוירונים – מתאים לדפוסים מורכבים, מדמה את פעולת המוח האנושי לזיהוי דפוסים בטקסטים.

Comparison of F1 Scores for Different Models (Random State: 42 vs 123)





תהליך אימון המודל ושיפור התוצאות

תהליך אימון המודלים כלל שלבים לשיפור דיוק ועמידות בזיהוי פשינג. בשלב הראשוני השתמשנו במגוון אלגוריתמים וביצענו חלוקת נתונים לאימון ובדיקה. לשם אופטימיזציה של הפרמטרים, נעשה שימוש ב-Grid Search כדי למצוא את ההגדרות המיטביות לכל מודל.



תהליך אימון המודל ושיפור התוצאות

נבחנו פרמטרים עיקריים לשיפור דיוק המודלים:

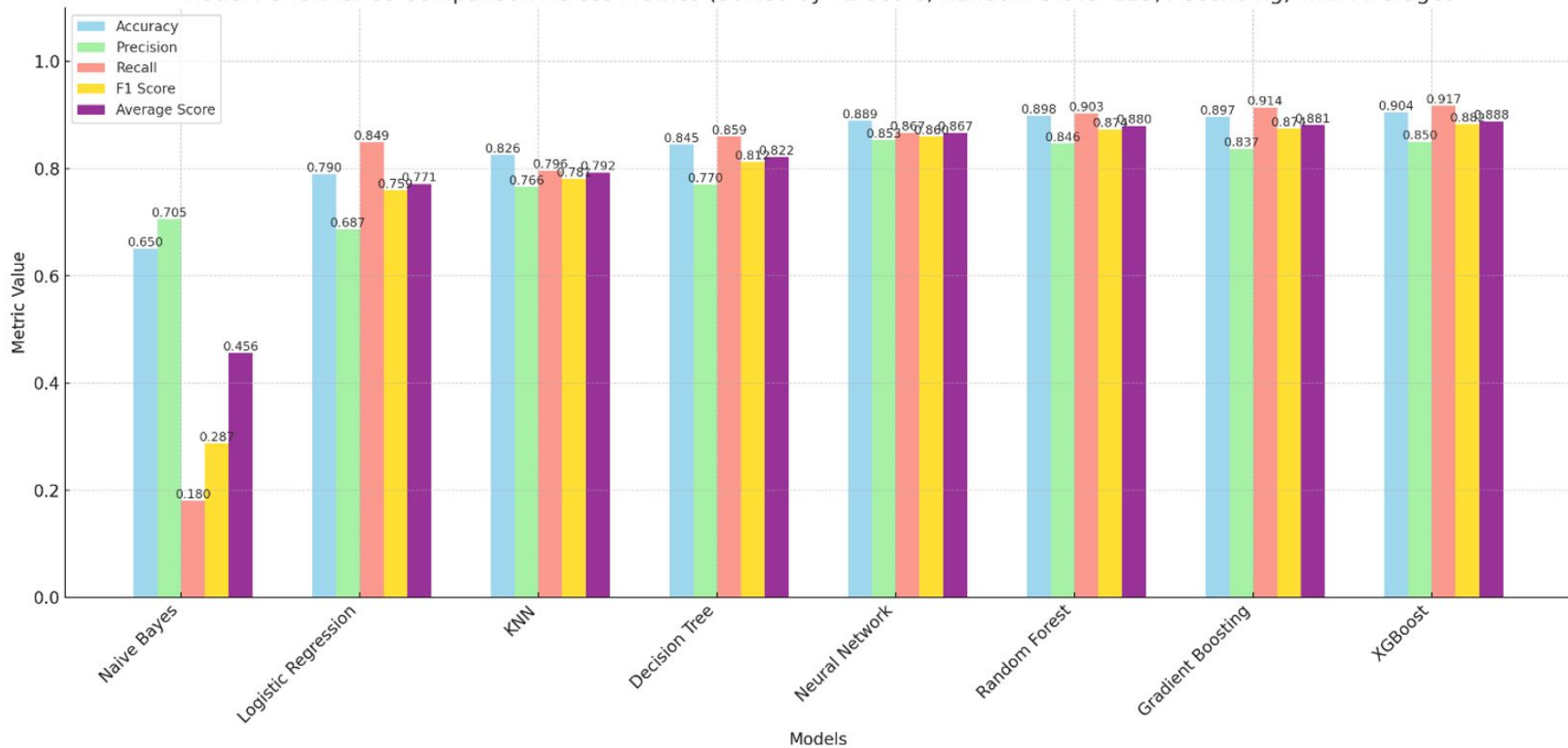
- **מספר העצים (`n_estimators`)** במודלים כמו Random Forest ו-XGBoost, עם איזון בין דיוק למשאבים.
- **עומק מקסימלי (`max_depth`)** לאיזון בין פשטנות לדיוק יתר.
- **קצב למידה (`learning_rate`)** במודלים כמו Gradient Boosting.
- **פרמטר C** ב-SVM לקביעת דיוק הסיווג.
- **מבנה הרשת (`hidden_layer_sizes`)** ברשתות נוירונים לזיהוי דפוסים מורכבים. בנוסף, בדקנו יציבות המודלים עם ערכי Random State שונים כדי להבטיח אמינות בתוצאות.



תוצאות ושיפור

לאחר אופטימיזציה באמצעות Grid Search, מדדנו את ביצועי המודלים בעזרת Accuracy, Precision, Recall, ו-F1 Score. המודל XGBoost נמצא כמתאים ביותר לפרויקט מבין כל האלגוריתמים של למידת מכונה, עם ביצועים גבוהים בכל המדדים.

Model Performance Comparison Across Metrics (Sorted by F1 Score, Random State: 123, Ascending) with Averages





שימוש ב-NLP

בפרויקט נבדקה גישת עיבוד שפה טבעית (NLP) לצד למידת מכונה. בעוד שלמידת מכונה מתמקדת בזיהוי תבניות מתוך מאפיינים, גישת ה-NLP מתמקדת בהבנה מעמיקה של השפה וההקשר, כדי לזהות דקויות שעשויות להעיד על פשיג.

בפרויקט השתמשנו ב-DistilBERT, גרסה קלה ויעילה של מודל BERT, שמאפשר הבנה מעמיקה של הקשרים בטקסט לזיהוי פשיג המתחזה לשפה רגילה, באופן מהיר וללא פגיעה משמעותית בביצועים.



שלבי הטמעת המודל

שלבי הטמעת המודל:

- **טוקניזציה:** שימוש בטוקניזר של DistilBERT לחילוק הטקסט לטוקנים להבנת ההקשר.
- **בניית Dataset:** הכנת מערך נתונים מסודר להכנת המודל.
- **אימון המודל:** אימון על GPU תוך התאמת קצב הלמידה.
- **הערכת ביצועים:** הערכה על בסיס Precision, Recall ו-F1.
- **שמירת המודל:** שמירה לשימוש עתידי, כולל קובץ תצורה.



יתרונות השימוש ב-NLP ומודלים טרנספורמטיביים

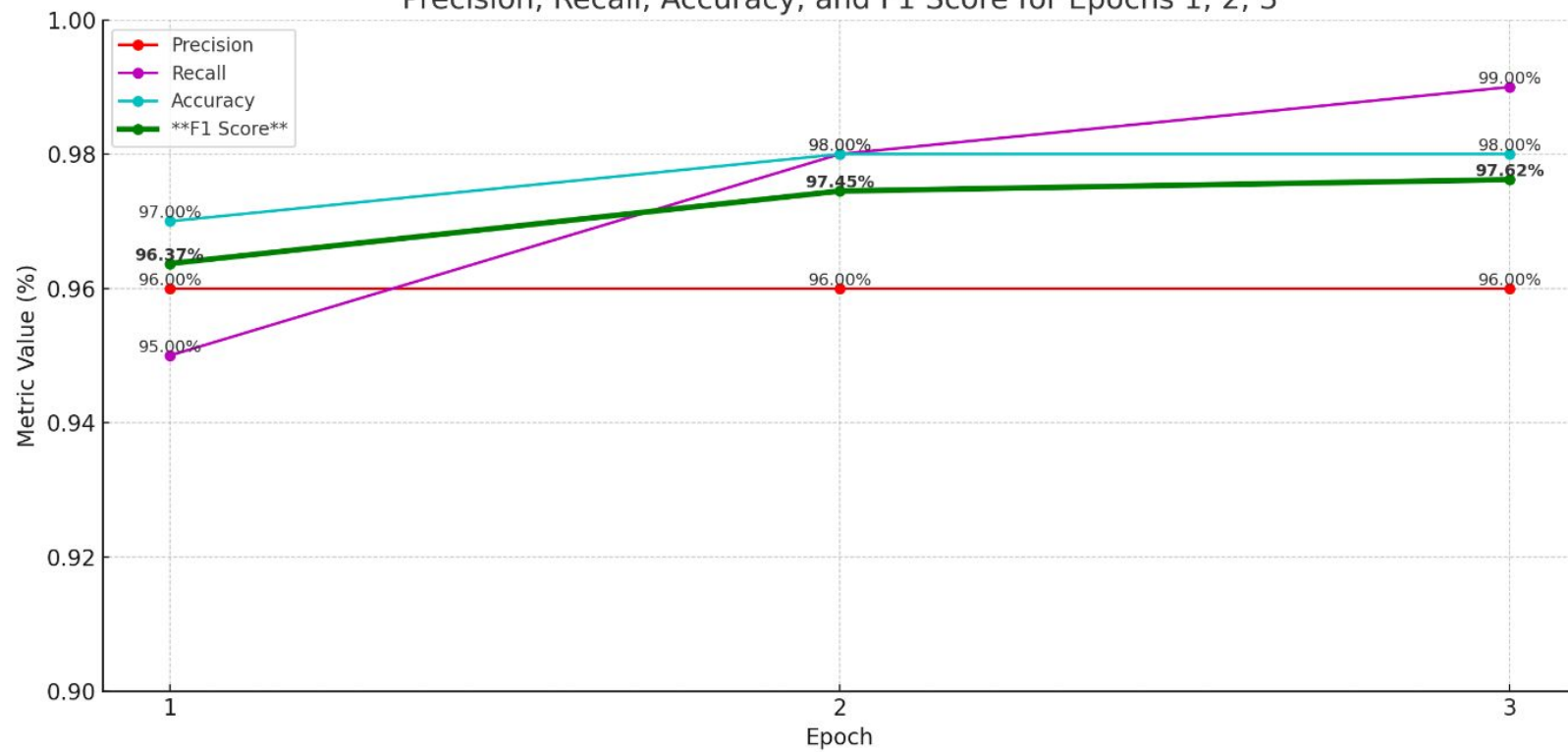
שימוש ב-NLP ומודלים טרנספורמטיביים כמו DistilBERT מאפשר הבנת הקשרים בטקסט והבחנה בין טקסט רגיל למתחזה, יתרון חשוב בזיהוי פישיוג. מודלים אלו מזהים מילים וסימנים חריגים ומנתחים כוונות בטקסט, ובכך תורמים לזיהוי יעיל ומדויק יותר של ניסיונות הונאה, תוך הפחתת שגיאות ושיפור הבטיחות למשתמשים.



תהליך אימון המודל ושיפור התוצאות

אימון מודל DistilBERT כלל חלוקת נתונים לאימון ובדיקה כדי להבטיח הכללה טובה. האימון בוצע לאורך 3 מחזורי Epoch, לשיפור דיוק המודל בזיהוי פשינג מבלי לגרום ל-Overfitting. ביצועי המודל נבדקו באמצעות מדדי Accuracy, Precision, Recall ו-F1, שהראו שיפור עם כל מחזור אימון. בסיום האימון נשמר המודל לשימוש עתידי, כולל קובצי תצורה, כך שהוא מוכן לסיווג אוטומטי של הודעות דוא"ל לשיפור אבטחת המשתמשים.

Precision, Recall, Accuracy, and F1 Score for Epochs 1, 2, 3





תהליך אימון המודל ושיפור התוצאות

אימון נוסף של מודל DistilBERT על הודעות אותנטיות העלה את F1-score ב-1.11%, מה שמצביע על שיפור אמיתי, ולא על Overfitting. השימוש בנתונים עכשוויים משפר את התאמת המודל לשיטות פישניג משתנות, ומבטיח רלוונטיות לאורך זמן באמצעות אימון מתמשך.



תוצאות והערכה

לצורך בחירת המודל הטוב ביותר לזיהוי פשינג, השתמשנו במדדי ביצוע מרכזיים:

1. **Accuracy** – אחוז הסיווגים הנכונים, אך עלול להטעות במאגרי נתונים לא מאוזנים.
2. **Precision** – מדויק לסיווג פשינג, מקטין סיווג שגוי של הודעות לגיטימיות.
3. **Recall** – בודק יכולת זיהוי פשינג מלא, כולל במחיר של סיווגים שגויים.
4. **F1-Score** – משלב בין Precision ו-Recall למדד מאוזן לביצועים.

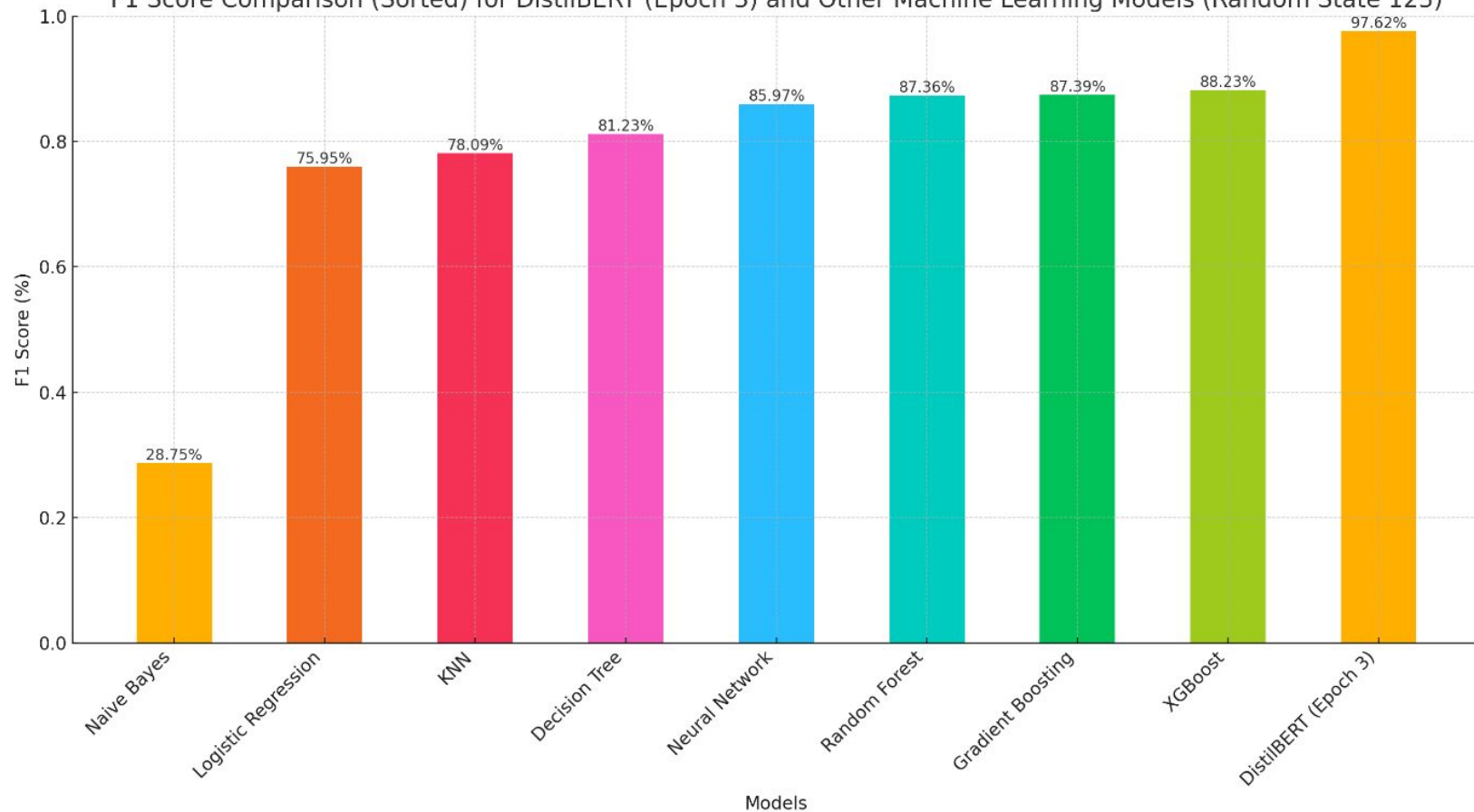
מדדים אלו עוזרים לבחור מודל שמאזן בין זיהוי מדויק של פשינג לבין הימנעות מסיווג שגוי של הודעות לגיטימיות



תוצאות הניסויים וההבדלים בין הגישות

XGBoost השיג F1-score של 0.8823 ודיוק גבוה בזיהוי פשינג בנתונים גדולים. DistilBERT בגישת NLP הציג F1-score של 0.9762 עם הבנה לשונית מעמיקה, אך דורש יותר משאבי חישוב. מודלים פשוטים יותר הציגו דיוק נמוך יותר עקב המורכבות בטקסטי הפשינג.

F1 Score Comparison (Sorted) for DistilBERT (Epoch 3) and Other Machine Learning Models (Random State 123)





Model	Random State	Accuracy	Precision	Recall	F1 Score
Naive Bayes	42	63.80%	65.08%	15.75%	25.36%
Naive Bayes	123	65.05%	70.50%	18.05%	28.75%
Logistic Regression	42	77.95%	67.93%	82.46%	74.49%
Logistic Regression	123	79.00%	68.70%	84.89%	75.95%
KNN	42	81.45%	75.56%	77.59%	76.56%
KNN	123	82.55%	76.60%	79.64%	78.09%
Decision Tree	42	83.80%	75.53%	86.56%	80.67%
Decision Tree	123	84.50%	77.04%	85.92%	81.23%

Gradient Boosting	42	88.90%	83.16%	89.76%	86.33%
Neural Network	42	88.90%	83.16%	89.76%	86.33%
Neural Network	123	88.95%	85.26%	86.68%	85.97%
XGBoost	42	89.10%	83.00%	90.65%	86.66%
Random Forest	42	89.55%	84.05%	90.40%	87.11%
Gradient Boosting	123	89.70%	83.70%	91.42%	87.39%
Random Forest	123	89.80%	84.63%	90.27%	87.36%
XGBoost	123	90.45%	85.04%	91.68%	88.23%
DistilBERT (NLP)	Epoch 3	98.00%	96.00%	99.00%	97.62%



מסקנות

המודל הטוב ביותר היה מודל DistilBERT במסגרת השימוש בגישת ה-NLP, שהצליח להשיג ביצועים גבוהים במיוחד בזיהוי הודעות פישיוג עם F1-score של 0.9762. מודל XGBoost, במסגרת המודלים של למידת מכונה, גם הוא השיג תוצאות טובות עם F1-score של 0.8823, והוכיח שהוא מתאים כאשר נדרש איזון בין דיוק ובין זמן ריצה ותשתית חישובית.



דיון בתוצאות

מודלים מבוססי NLP כמו DistilBERT מציעים יתרון משמעותי בזיהוי פשינג עם דיוק ורגישות גבוהים, הודות להבנת ההקשרים הלשוניים. לעומתם, המודלים הקלאסיים מהירים ופשוטים אך מתקשים בזיהוי שפה מתוחכמת. למרות העלות החישובית, NLP מהווה פתרון יעיל יותר להגנה מפני הונאות פשינג.



ממשק GUI למשתמש קצה

פיתחנו ממשק משתמש (UI) ב-Tkinter ב-Python, שמאפשר למשתמשים לנתח הודעות דוא"ל ולזהות ניסיונות פשינג, עם דגש על נוחות ואינטואיטיביות.



ממשק GUI למשתמש קצה

- **טעינת ושמירת הודעות:** אפשרות לטעינת קבצי טקסט ולשמירתם לאחר ניתוח.
- **ניתוח הודעה:** הכנסת טקסט ההודעה ולחיצה על "Analyse Email" לחיזוי באמצעות DistilBERT.
- **חישוב הסתברויות ותצוגת תוצאות:** הצגת הסתברויות לרמת סיכון, כולל אינדיקציות צבע והמלצות לפעולה.
- **ניהול היסטוריה:** שמירת היסטוריית ניתוחים בקובץ JSON, כולל פרטי הניתוח.
- **עזרה למשתמש:** חלון "עזרה" עם הנחיות שימוש ומידע על המערכת, והסברת רמות הסיכון.



תיאור פעולת הממשק

המערכת משתמשת ב-DistilBERT לחיזוי הסתברות שהודעה היא פשינג, מציגה המלצות לפעולה ושומרת את הניתוחים בהיסטוריה לגישה עתידית. הממשק מעוצב לנוחות ואינטואיטיביות, מתאים למשתמשים ללא ידע טכני, עם עיצוב נקי ואלמנטים ויזואליים להדגשת רמות סיכון. כל הפעולות המרכזיות נגישות בלחיצה, ומאפשרות קבלת תובנות מהירות וברורות כדי לסייע למשתמשים לקבל החלטות בטוחות.



פונקציונליות מרכזית

- סריקה וניתוח אוטומטי של הודעות
- זיהוי תבניות פשינג בזמן אמת
- חישוב הסתברויות מדויקות
- מתן המלצות פעולה מיידיות



זרימת העבודה

1. קליטת הודעה (טקסט או קובץ)
2. ניתוח מידי באמצעות מנוע DistilBERT
3. הצגת רמת סיכון והסתברויות
4. מתן המלצות פעולה מותאמות
5. תיעוד אוטומטי בהיסטוריית המערכת

תיאור פעולת הממשק



תיאור פעולת הממשק





פונקציונליות נוספת למשתמש הקצה

מערכת PhishGuard AI מאפשרת ניתוח מקבילי של מספר הודעות דוא"ל, כולל קבצי MSG, eml, txt להצגת דירוגי סיכון והמלצות לכל הודעה. המערכת שומרת דוחות וסטטיסטיקות, מספקת ממשק עזרה אינטראקטיבי, ומתחזקת היסטוריית ניתוחים מלאה למעקב אחרי דפוסים חוזרים – כל זאת לשיפור יעילות העבודה ונוחות המשתמשים.



סיכום והישגי הפרויקט



הישגי הפרויקט

- פיתוח מערכת אוטומטית לזיהוי פישיוג
- השגת דיוק של 97.62% באמצעות מודל DistilBERT
- יצירת ממשק משתמש אינטואיטיבי ונוח
- מערכת המאפשרת זיהוי בזמן אמת



כיווני פיתוח עתידיים

הרחבת יכולות:

- שיפור הלמידה מדאטה חדש
- הוספת תמיכה בשפות נוספות
- פיתוח יכולות זיהוי נוספות

הרחבת יכולות:

- אינטגרציה עם מערכות דוא"ל קיימות
- אופטימיזציה של זמני תגובה
- הרחבת יכולות הניתוח האוטומטי



מבט קדימה - PhishGuard AI

מציעה פתרון מתקדם ויעיל לאתגרי אבטחת המידע המודרניים, עם יכולת להתמודד עם PhishGuard AI האיומים המשתנים בעולם הפישינג. המערכת מבוססת על טכנולוגיות מתקדמות המאפשרות לא רק הגנה בזמן אמת, אלא גם למידה מתמדת המבטיחה דיוק הולך ומשתפר. עם בסיס טכנולוגי חזק מספקת מענה אמין, גמיש ודינמי לצרכים של היום ושל PhishGuard AI, ופוטנציאל להתפתחות עתידית. המחר.




קשיים ואתגרים

שילוב לימודים, עבודה והמצב הביטחוני היוו אתגר משמעותי שדרש איזון בין מחויבויות. פתרנו זאת באמצעות תעדוף משימות ותקשורת שוטפת.

אתגר טכני עיקרי היה הכרת האלגוריתמים, שנפתר באמצעות למידה עצמית, ייעוץ ושימוש בקוד פתוח.

התמודדנו עם כמות דאטה גדולה בעזרת כלים מתאימים (כמו pandas) ומחשב בעל יכולות עיבוד גבוהות.

גאנט



•מאי - (השוואה ולמידה של אלגוריתמים): איסוף נתונים, תיוג ועיבוד הדאטה, השוואה בין סוגים שונים של אלגוריתמים, מדידת הביצועים שלהם ובחירת האלגוריתמים בהם נשתמש.

•יוני – (התחלת פיתוח המערכת): פיתוח סביבה ראשונית ותחילת סיווג המידע בעזרת האלגוריתמים הנבחרים.

•יולי – (פיתוח UI ובדיקות): פיתוח ויצירת ממשק משתמש UI והתחלת בדיקות המערכת.

•אוגוסט - (סיום פיתוח המערכת): סיום פיתוח מערכת עם שילוב תכונות מלא של האלגוריתמים שנבחרו. עריכת בדיקות מקיפות מול דרישות, חידוד ושיפור הפרויקט על סמך משוב ודיוק התוצאות הרצויות עד להגשה סופית.

•ינואר - (דוח איפיון): הערכת ותיזמון היקף הפרויקט, היעדים והדרישות המפורטות.

•פברואר - (לימוד החומרים החדשים הנדרשים לפרויקט): רכישת ידע תיאורטי ומיומנויות טכניות הכרחיות בעולם למידת המכונה בינה NLP

•מרץ - (תכנון המערכת): פיתוח ארכיטקטורת המערכת, החלטה כיצד אנו רוצים שתעבוד ובאיזה אופן לאחר הבנת הרקע והתיאוריה.

•אפריל - (הפקת נתונים ותיוגם): התחלת איסוף נתונים, ייצור נתונים בעזרת מערכות שונות, תיוג ועיבוד ראשוני לפיתוח מערכת.

•אמצע אפריל עד תחילת מאי - חופשת פסח