



الجمهورية العربية السورية

جامعة الشام الخاصة
كلية الهندسة المعلوماتية

تقرير النموذج الأولي
العام الدراسي 2023-2024

Base Line Report



الخلاصة

في السنوات الأخيرة ومع الاعتماد الكبير على الخدمات الالكترونية وتطور تقنيات الذكاء الصناعي زاد الاهتمام بقضية المراجعات الكاذبة على الخدمات الالكترونية، وتم طرح الكثير من الحلول في هذا الخصوص، ومن الحلول المطروحة استخدام نموذج BERT وفوقه طبقة Drop-out وفوقها طبقة عصبية تسقط الدخل الى الخرج او استخدام النموذج مع طبقتين عصبيتين، وتم استخدام مجموعة بيانات AFRD المطروحة في بحث اخر وكذلك اعتمدنا على نفس البحث في تدريب النموذج على بيانات مجال واحد او عدة مجالات، تم الاعتماد على عدة معايير في التقييم وهي Precision, Recall, F1, Accuracy, confusion matrix، وعلى نتيجة تم الوصول في مجال محدد هي 0.90، وعلى في مجالات متعددة كانت 0.83.

1. المقدمة

مع الأبحاث العديدة المطروحة في قضية اكتشاف المراجعات الزائفة ومع تنوع الطرق والأساليب المستخدمة لتحقيق أفضل النتائج، وبفضل التطور الكبير الذي حصل في السنوات الأخيرة أصبح هنالك مجموعة جديدة من الأدوات التي يمكن استخدامها للوصول الى حلول أفضل من سابقتها سواء على مستوى الدقة او الكلفة او وقت التدريب او....، وكان للغة العربية حصة أيضا في هذا التطوير الحاصل وذلك على الرغم من التأخر الذي كان يشهده من قبل، فأصبح هنالك مقابل عربي لنموذج BERT وهو AraBERT وكذلك تم انشاء مجموعات بيانات ذات جودة عالية من قبل الخبراء مثل AFRD.

2. الأهداف

نهدف في هذا البحث الى استخدام نموذج ال AraBERT او بصورة ادق AraBERT V2 مع طبقة Drop-out وطبقة لأسقاط الدخل الى الخرج وكذلك سنستخدم النموذج مع إضافة طبقتين اليه وذلك لتكرار طريقة حل مطروحة في بحث سابق مع بعض التغييرات، والبيانات المستخدمة هي من بحث سابق أيضا وتسمى مجموعة البيانات ب AFRD، حيث الهدف هو الوصول الى نفس النتائج المقدمة في هذه الأبحاث.

3. الطريقة

لقد قمنا بأشاء نموذجين، الأول يتكون من نموذج AraBERT لعملية ال embedding مع طبقة Drop-out بنسبة 0.1 وطبقة أخيرة تتكون من 768 دخل وخرجين ونرمز له ب m01، والثاني يتكون من نموذج AraBERT لعملية ال embedding وطبقة تتكون من 768 دخل و256 خرج والطبقة الثانية من 256 دخل وخرجين ونرمز لها ب m02، والبيانات المستخدمة هي AFRD التي تحوي بيانات عن ثلاث مجالات الفنادق والمطاعم والاغراض ونرمز اليها بالترتيب H,R and P حيث قمنا بتدريب النماذج على مستوى مجال واحدة و ونختبره على هذا المجال، وكذلك نقوم بتدريب النموذج على مجال ما ونعيد تدريبه على المجال



الذي سيستخدم فيه، وقمنا بتدريب النموذج على بيانات لمجالات متعددة ، فمثلا النموذج ذو الهيكلية m01 والمدرّب على بيانات الفنادق ثم المطاعم ثم الأغراض نرّمز له بـ HRPmodel_m01، وتم تحديد معدل Drop-Out كما ذكر في بحث سابق ليكون 0.1، ومعدل التعلم بـ $(2.5e-5)$ ، و Batch Size بـ 16، وطول الـ embedding بـ 265، عدد العقود 3-4، وقمنا باستخدام مجموعة معايير للاختبار وهي Precision, Recall, F1, Accuracy, confusion matrix.

4. النتائج

نقوم في هذا القسم بعرض النتائج التي توصلنا اليها من اجل كل نموذج وبمختلف الطرق المستخدمة للتعامل مع البيانات:

1. Model01(m01)

• Single Domain

➤ تدريب النموذج على بيانات المجال المراد استخدامه واختباره عليها.

	precision	recall	f1-score	support
Truthful	0.78	0.83	0.81	83
Fake	0.85	0.80	0.82	96
accuracy			0.82	179
macro avg	0.82	0.82	0.82	179
weighted avg	0.82	0.82	0.82	179

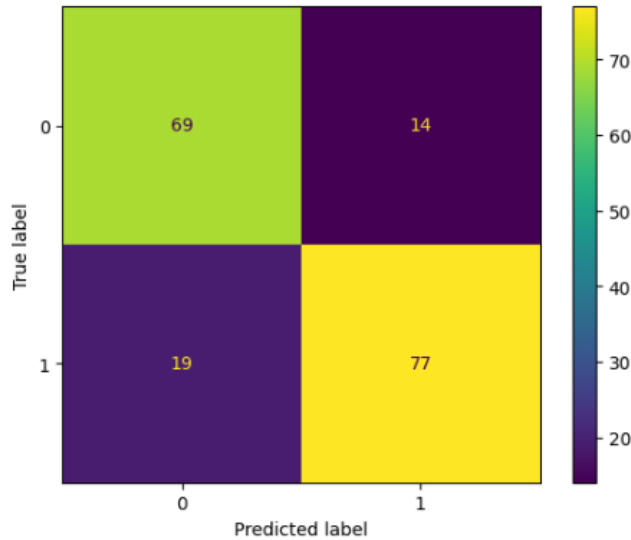


Figure 2 Restaurant Data Set

	precision	recall	f1-score	support
Truthful	0.94	0.73	0.82	44
Fake	0.73	0.94	0.82	34
accuracy			0.82	78
macro avg	0.83	0.83	0.82	78
weighted avg	0.85	0.82	0.82	78

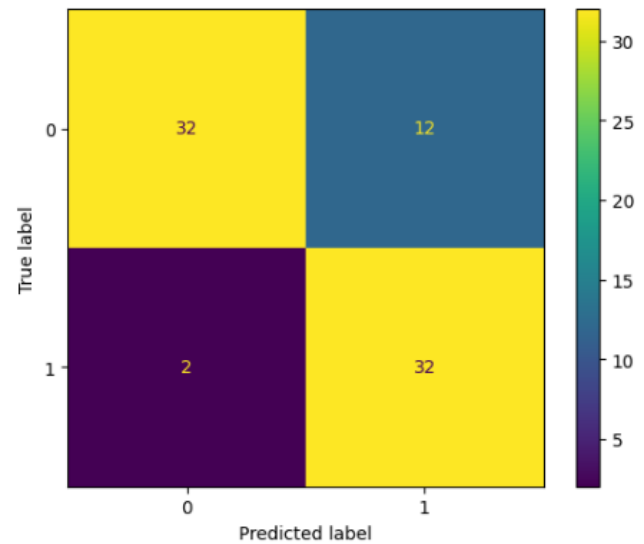


Figure 1 Hotel Data Set



الجمهورية العربية السورية

جامعة الشام الخاصة
كلية الهندسة المعلوماتية

تقرير النموذج الاول
العام الدراسي 2023-2024

	precision	recall	f1-score	support
Truthful	0.78	0.82	0.80	84
Fake	0.83	0.79	0.81	92
accuracy			0.81	176
macro avg	0.81	0.81	0.81	176
weighted avg	0.81	0.81	0.81	176

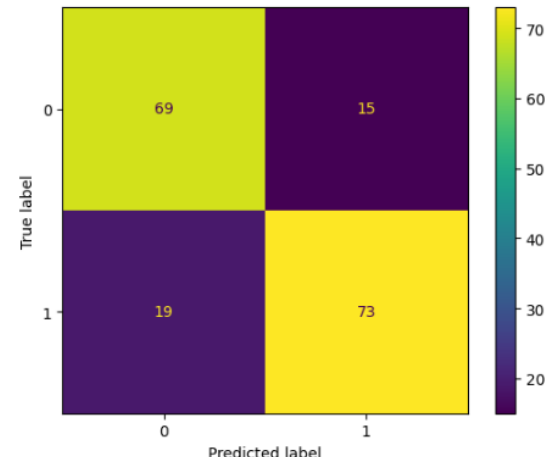


Figure 3 Product Data Set

➤ تدريب النموذج على بيانات مجال ما من ثم تدريب النموذج واختباره على بيانات المجال المراد استخدامه فيه

	precision	recall	f1-score	support
Truthful	0.95	0.67	0.79	83
Fake	0.78	0.97	0.86	96
accuracy			0.83	179
macro avg	0.86	0.82	0.82	179
weighted avg	0.86	0.83	0.83	179

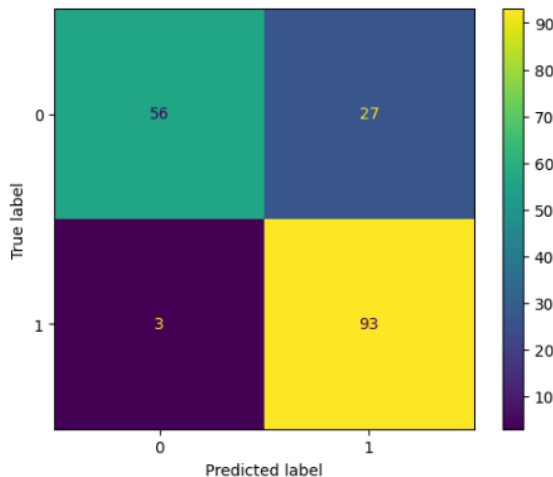


Figure 4 Product data than Restaurant data

	precision	recall	f1-score	support
Truthful	0.95	0.86	0.90	44
Fake	0.84	0.94	0.89	34
accuracy			0.90	78
macro avg	0.90	0.90	0.90	78
weighted avg	0.90	0.90	0.90	78

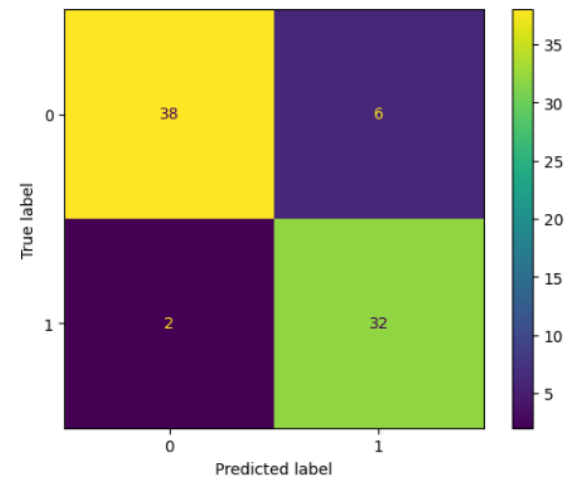


Figure 5 Product data than Hotel data



الجمهورية العربية السورية

جامعة الشام الخاصة
كلية الهندسة المعلوماتية

تقرير النموذج الاول
العام الدراسي 2023-2024

	precision	recall	f1-score	support
Truthful	0.73	0.87	0.79	84
Fake	0.86	0.71	0.77	92
accuracy			0.78	176
macro avg	0.79	0.79	0.78	176
weighted avg	0.80	0.78	0.78	176

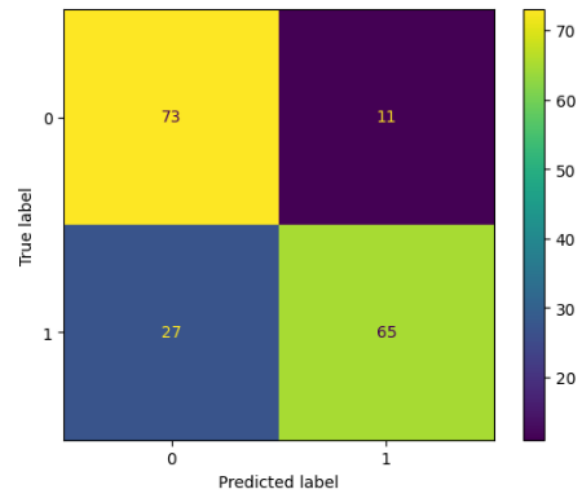


Figure 8 Hotel data than Product data

➤ تدريب النموذج على بيانات مجالين من ثم تدريب النموذج واختباره على بيانات المجال المراد استخدامه فيه.

	precision	recall	f1-score	support
Truthful	0.95	0.72	0.82	83
Fake	0.80	0.97	0.88	96
accuracy			0.85	179
macro avg	0.88	0.85	0.85	179
weighted avg	0.87	0.85	0.85	179

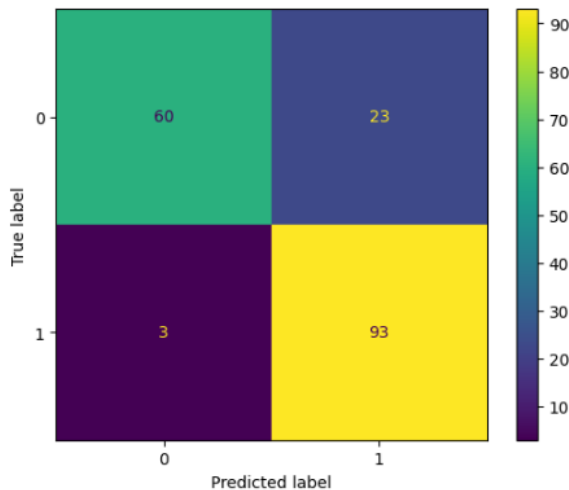


Figure 6 Hotel & Product data than Restaurant data

	precision	recall	f1-score	support
Truthful	0.91	0.89	0.90	44
Fake	0.86	0.88	0.87	34
accuracy			0.88	78
macro avg	0.88	0.88	0.88	78
weighted avg	0.89	0.88	0.88	78

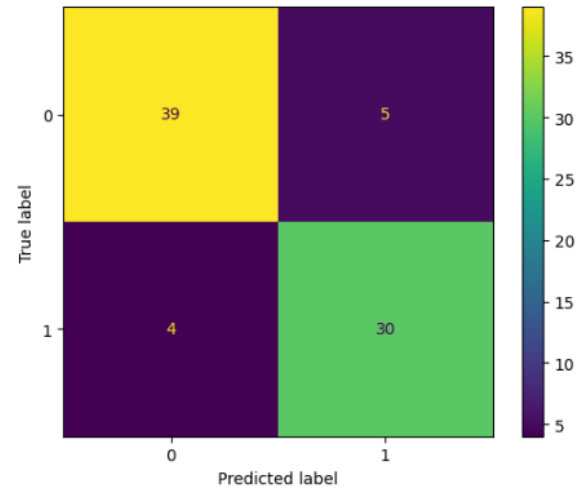


Figure 7 Product & Restaurant data than Hotel data



	precision	recall	f1-score	support
Truthful	0.77	0.79	0.78	84
Fake	0.80	0.78	0.79	92
accuracy			0.78	176
macro avg	0.78	0.78	0.78	176
weighted avg	0.78	0.78	0.78	176

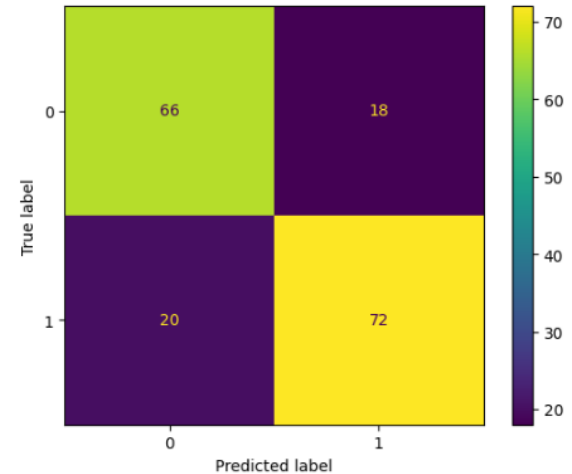


Figure 9 Hotel & Restaurant data than Product data

• Multi Domain: ندرّب النموذج على جميع المجالات بترتيب معين ونقوم باختباره على جميع البيانات.

• Mixed Data: ندرّب النموذج على جميع البيانات معا.

	precision	recall	f1-score	support
Truthful	0.84	0.83	0.84	224
Fake	0.82	0.83	0.82	208
accuracy			0.83	432
macro avg	0.83	0.83	0.83	432
weighted avg	0.83	0.83	0.83	432

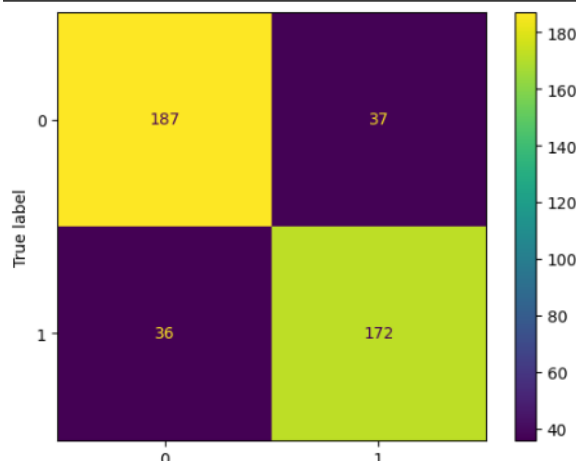


Figure 10 Mixed Data

	precision	recall	f1-score	support
Truthful	0.80	0.88	0.84	211
Fake	0.87	0.79	0.83	222
accuracy			0.83	433
macro avg	0.83	0.83	0.83	433
weighted avg	0.83	0.83	0.83	433

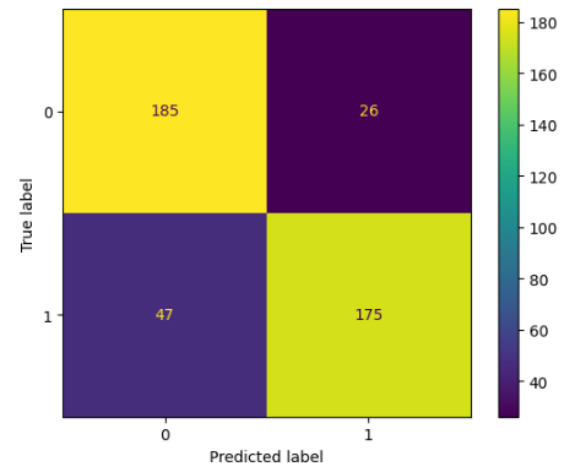


Figure 11 Multi Domain (PRH)



Model02(m02) .II

Single Domain •

➤ تدريب النموذج على بيانات المجال المراد استخدامه واختباره عليها.

	precision	recall	f1-score	support
Truthful	0.93	0.77	0.84	83
Fake	0.83	0.95	0.88	96
accuracy			0.87	179
macro avg	0.88	0.86	0.86	179
weighted avg	0.87	0.87	0.86	179

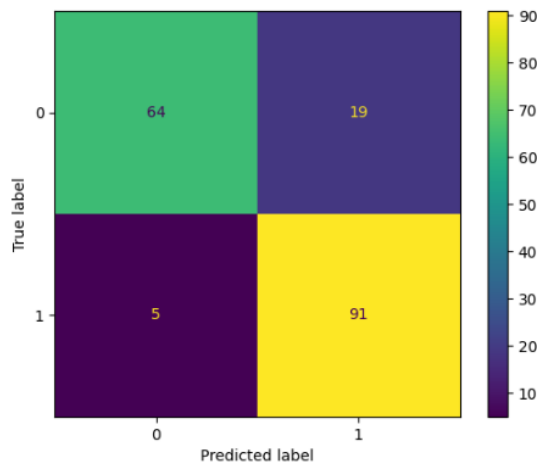


Figure 14 Restaurant Data

	precision	recall	f1-score	support
Truthful	0.89	0.73	0.80	44
Fake	0.71	0.88	0.79	34
accuracy			0.79	78
macro avg	0.80	0.80	0.79	78
weighted avg	0.81	0.79	0.80	78

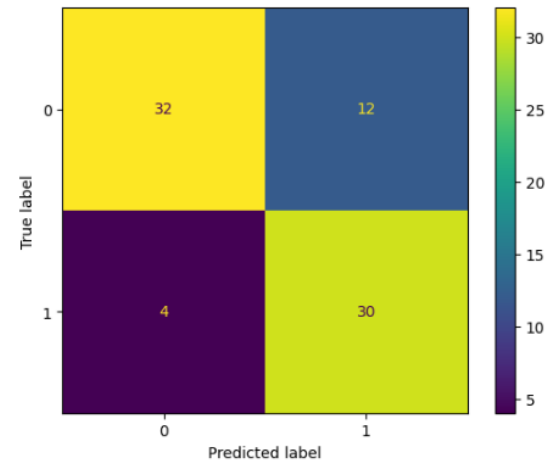


Figure 13 Hotel Data

	precision	recall	f1-score	support
Truthful	0.89	0.73	0.80	44
Fake	0.71	0.88	0.79	34
accuracy			0.79	78
macro avg	0.80	0.80	0.79	78
weighted avg	0.81	0.79	0.80	78

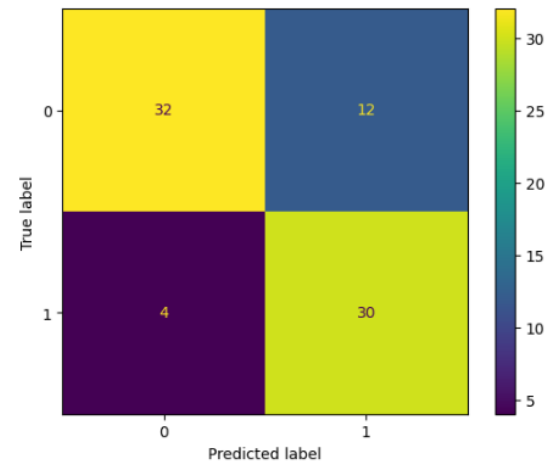


Figure 12 Product Data



➤ تدريب النموذج على بيانات مجال ما من ثم تدريب النموذج واختباره على بيانات المجال المراد استخدامه فيه.

	precision	recall	f1-score	support
Truthful	0.92	0.73	0.82	83
Fake	0.81	0.95	0.87	96
accuracy			0.85	179
macro avg	0.86	0.84	0.84	179
weighted avg	0.86	0.85	0.85	179

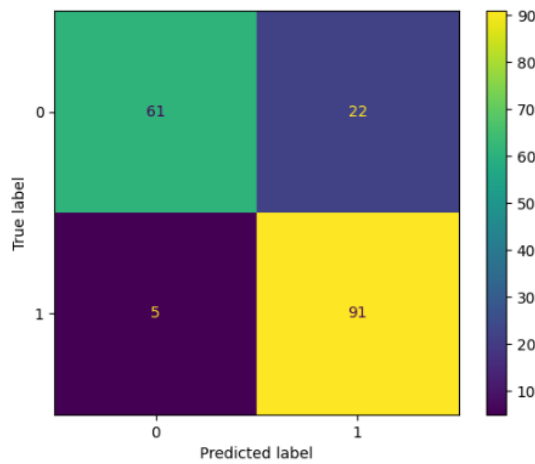


Figure 16 Hotel data than Restaurant data

	precision	recall	f1-score	support
Truthful	0.89	0.89	0.89	44
Fake	0.85	0.85	0.85	34
accuracy			0.87	78
macro avg	0.87	0.87	0.87	78
weighted avg	0.87	0.87	0.87	78

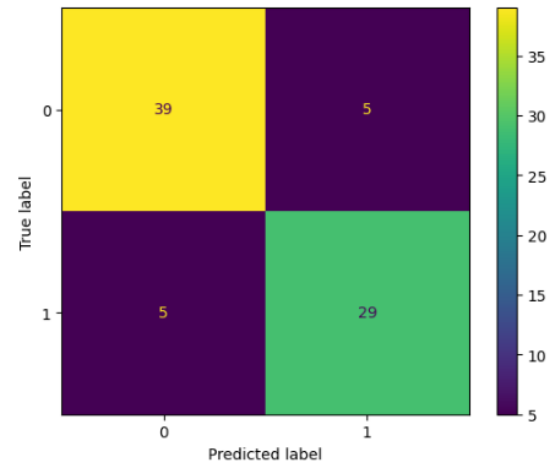


Figure 17 Restaurant data than Hotel data

	precision	recall	f1-score	support
Truthful	0.77	0.81	0.79	84
Fake	0.82	0.78	0.80	92
accuracy			0.80	176
macro avg	0.80	0.80	0.80	176
weighted avg	0.80	0.80	0.80	176

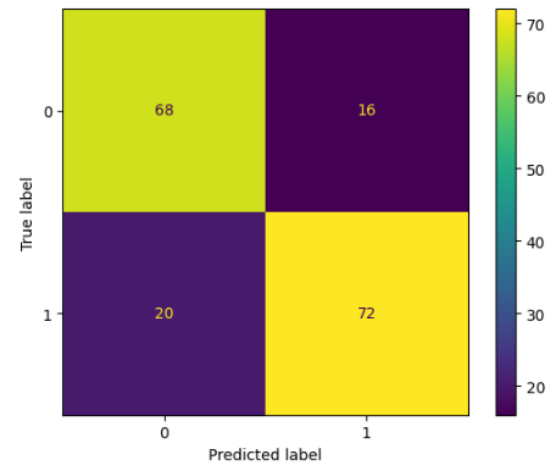


Figure 15 Restaurant data than Product data



➤ تدريب النموذج على بيانات مجالين من ثم تدريب النموذج واختباره على بيانات المجال المراد استخدامه فيه.

	precision	recall	f1-score	support
Truthful	0.85	0.84	0.85	83
Fake	0.87	0.88	0.87	96
accuracy			0.86	179
macro avg	0.86	0.86	0.86	179
weighted avg	0.86	0.86	0.86	179

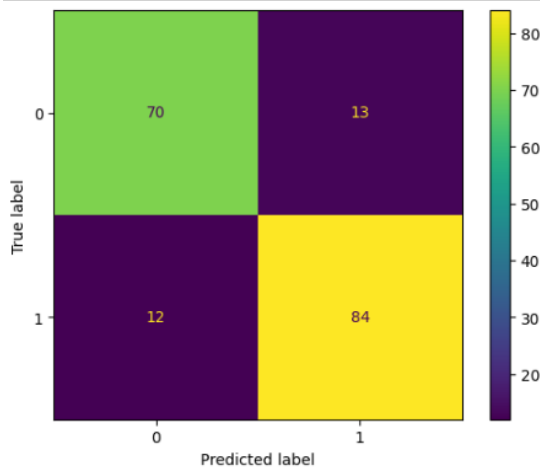


Figure 20 Product & Hotel data than Restaurant data

	precision	recall	f1-score	support
Truthful	0.93	0.89	0.91	44
Fake	0.86	0.91	0.89	34
accuracy			0.90	78
macro avg	0.89	0.90	0.90	78
weighted avg	0.90	0.90	0.90	78

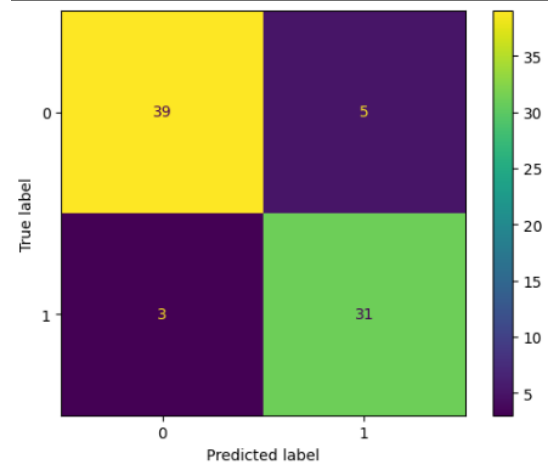


Figure 19 Product & Restaurant data than Hotel data

	precision	recall	f1-score	support
Truthful	0.76	0.86	0.80	84
Fake	0.85	0.75	0.80	92
accuracy			0.80	176
macro avg	0.80	0.80	0.80	176
weighted avg	0.81	0.80	0.80	176

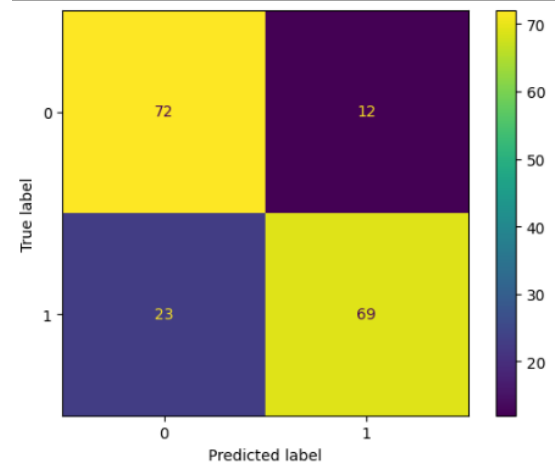


Figure 18 Hotel & Restaurant data than Product data



- Multi Domain: ندرّب النموذج على جميع المجالات بترتيب معين ونقوم باختباره على جميع البيانات.
- Mixed Data: ندرّب النموذج على جميع البيانات معاً.

	precision	recall	f1-score	support
Truthful	0.78	0.90	0.84	224
Fake	0.87	0.73	0.79	208
accuracy			0.82	432
macro avg	0.83	0.81	0.81	432
weighted avg	0.82	0.82	0.82	432

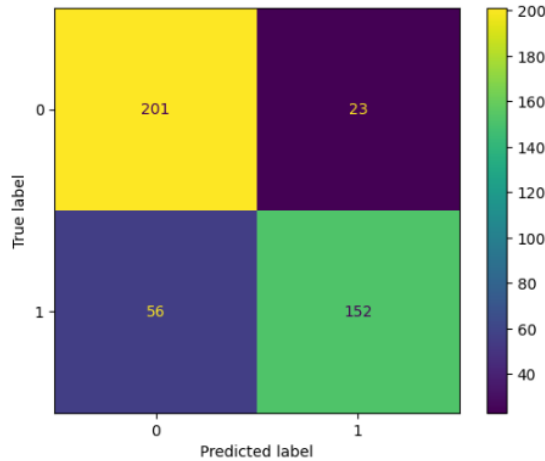


Figure 21 Mixed Data

	precision	recall	f1-score	support
Truthful	0.80	0.85	0.82	211
Fake	0.85	0.79	0.82	222
accuracy			0.82	433
macro avg	0.82	0.82	0.82	433
weighted avg	0.82	0.82	0.82	433

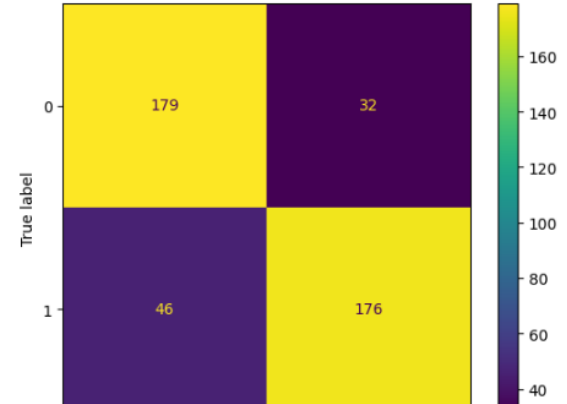


Figure 22 Multi Domain (PHR)

Table 1 All Models

F1	Recall	Precision	Accuracy	Model
0.82	0.94	0.73	0.88	Hmodel_m01
0.82	0.80	0.85	0.82	Rmodel_m01
0.81	0.79	0.83	0.81	Pmodel_m01
0.89	0.94	0.84	0.90	PHmodel_m01
0.86	0.97	0.78	0.83	PRmodel_m01
0.77	0.71	0.86	0.78	HPmodel_m01
0.87	0.88	0.86	0.88	PRHmodel_m01
0.88	0.97	0.80	0.85	HPRmodel_m01
0.79	0.78	0.80	0.78	HRPmodel_m01
0.83	0.79	0.87	0.83	MultiData_m01 (PRH)
0.82	0.83	0.82	0.83	MixedData_m01
0.79	0.88	0.71	0.79	Hmodel_m02
0.88	0.95	0.83	0.87	Rmodel_m02
0.79	0.88	0.71	0.79	Pmodel_m02
0.85	0.85	0.85	0.87	RHmodel_m02



0.87	0.95	0.81	0.85	HRmodel_m02
0.80	0.78	0.82	0.80	RPmodel_m02
0.89	0.91	0.86	0.90	PRHmodel_m02
0.87	0.88	0.87	0.86	PHRmodel_m02
0.80	0.75	0.85	0.80	HRPmodel_m02
0.82	0.79	0.85	0.82	MultiData_m02 (PHR)
0.79	0.73	0.87	0.82	MixedData_m02

5. الاستنتاجات

بعد الاختبارات ومراقبة عملية التدريب نلاحظ الاتي:

- النتائج توفق البحث من ناحية تحدد معدل ال Drop-Out ب 0.1 لكون الزيادة تؤثر سلبا.
- عند زيادة طول ال embedding تتحسن النتائج ولكن يتم استهلاك الموارد بالضعف، لذلك قررنا تحديد الطول ليكون 256.
- عند التدريب على مجال واحد تكون النتائج أفضل بعدد عقود يكون 4 وعند استخدام بيانات عدة مجالات تكون أفضل عند 3 او 2.
- تكون النتائج أفضل إذا كان المجال المراد واحدا وتقل بشكل ضعيف عند التعامل مع مجالات متعددة.
- عند تدريب النموذج على بيانات من مجالات اخر ثم تدريبه على المجال المراد نحصل على نتائج أفضل، وبشكل خاص عندما تكون البيانات الأولى هي الخاصة بالمنتجات.
- ترتيب ادخال البيانات يؤثر على النتائج.
- عند تدريب النموذج على بيانات مختلفة تصبح الدقة لديه أفضل في اكتشاف التعليقات المزيفة ولكنه تقل الدقة في تحديد التعليقات الحقيقية.

6. المراجع

- [Detecting Fake Online Reviews using Fine-tuned BERT](#)
- [Multiscale cascaded domain-based approach for Arabic fake reviews detection in e-commerce platforms](#)