## -1- تقييم النتائج

قمنا باختبار الخوارزمية بعد تدريبها وذلك من ناحية قدرتها على تقديم توصيات جيدة بالنسبة لمستخدم أعطى تقييمات سابقاً ضمن النظام، وبالنسبة لمستخدم جديد (مشكلة البداية الباردة)، وذلك على مجموعة التدريب نفسها، ومجموعة عشوائية أخرى من مستخدمين لا ينتمون لمجموعة التدريب. كانت المجموعة المتوقعة Predicted Label هي مجموعة التوصيات المقترحة من الخوارزمية، بينما المجموعة الحقيقة Truth Label هي مجموعة الأفلام التي قام المستخدم بإعطائها تقييم أكبر أو يساوي 3 (بسبب اختيارنا العتبة 3 عند تحويل مصفوفة التقييمات لمصفوفة ثنائية عند إدخالها لخوارزمية العنقدة). قمنا بحساب عدة معايير تتضمن: True Positives، True Negatives، False Positives، False Negatives ومن ثم حساب قيم Precision، Recall، F1-Measure، Accuracy بالاعتماد عليها.

يوضّح الجدول 2 النتائج التي حصلنا عليها:

جدول 2: نتائج تقييم خوارزمية Q-Learning بعد تدريبها

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Average Accuracy | Average F1-Score | Average Recall | Average Precision | Average FN | Average FP | Average TN | Average TP |  |
| 55% | 56% | 44% | 83% | 39 | 7 | 25 | 31 | المستخدمين الموجودين في مجموعة التدريب (الاعتماد على تقييماتهم المسبقة) |
| 50% | 52% | 42% | 74% | 40 | 11 | 22 | 29 | المستخدمين الموجودين في مجموعة التدريب (الاعتماد على معلوماتهم الديموغرافية – مشكلة البداية الباردة) |
| 65% | 34% | 55% | 33% | 12 | 24 | 54 | 12 | مجموعة عشوائية من المستخدمين (الاعتماد على تقييماتهم المسبقة) |
| 60% | 29% | 52% | 26% | 12 | 29 | 51 | 10 | مجموعة عشوائية من المستخدمين (الاعتماد على معلوماتهم الديموغرافية – مشكلة البداية الباردة) |

بالنظر للجدول، يمكننا استنتاج مجموعة ملاحظات:

* عدم انخفاض الدقة Accuracy عند الاختبار على مجموعة مغايرة لمجموعة التدريب، وهذا يدل على عدم وجود حالة overfitting على معطيات التدريب.
* نلاحظ انخفاض قيمة الضبط Precision بشكل ملحوظ عند الاختبار على مجموعة مغايرة لمجموعة التدريب، في حين لا نلاحظ هذا الانخفاض في قيمة الاسترجاع Recall. أي بمعنى آخر، نلاحظ أن الخوارزمية بعد التدريب أصبحت قادرة على التقاط نسبة كبيرة من الأفلام التي ستعجب المستخدم ولكن مع وجود ضجيج (أفلام ليست ذات صلة). نرى بأنه يجب العمل على تحسين هذه النتائج وخصوصاً أن الهدف الأساسي هو اقتراح أفلام ستعجب المستخدم، وليس عدم تفويت أي فيلم سيعجبه عليه.
* الأداء بشكل عام يقل بشكل طفيف عند استخدام المعلومات الديموغرافية عوضاً عن التقييمات (بمراقبة قيمة كل من F1-Score والدقة Accuracy)، وهذا يدل على قدرة هذه المعلومات على إعطاء تمثيل بديل جيد للمستخدم في حال عدم وجود تقييمات له (قدرة المنهجية المقدمة على إعطاء حل جيد لمشكلة البداية الباردة).