جامعة حلب



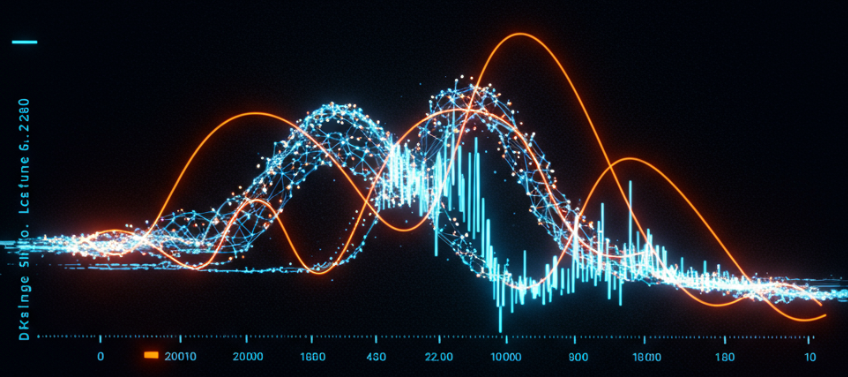
كلية الهندسة المعلوماتية

للعام الدراسي 2024/2025

Regression of Netflix stocks   
using Elman network

بإشراف: المهندسة هديل عبد الرحمن

الاسم: محمود بنّان



* **ملخص المشروع:**

يهدف هذا المشروع إلى استكشاف إمكانية التنبؤ بأسعار أسهم Netflix باستخدام الشبكات العصبية المتكررة (RNN)، وتحديداً شبكة Elman. تم تدريب النموذج على مجموعة بيانات "netflix-stocks" والتي تحتوي على بيانات تاريخية مفصلة لأسعار أسهم Netflix (مثل أسعار الافتتاح والإغلاق وأعلى وأدنى سعر وحجم التداول) على مدار فترة زمنية محددة.

قمت بتصميم وتدريب شبكة عصبونية اصطناعية بهدف التنبؤ **بسعر السهم الاغلاق المعدل للشركة**.   
أظهر النموذج أداءً ممتازاً على بيانات التدريب، حيث كانت **نسبة خطأ منخفضة جداً حيث بلغت 0.000008%.**   
عند تقييم النموذج على بيانات اختبار جديدة، كانت **نسبة خطأ قد بلغت %0.004.**

* **حول المشروع:**

يُعد **التنبؤ بأسعار الأسهم** أحد أكثر المجالات تحديًا وإثارة في التمويل وعلوم البيانات. تتأثر أسعار الأسهم بعدد لا يحصى من العوامل الاقتصادية، السياسية، الاجتماعية، وحتى النفسية، مما يجعل حركتها متقلبة وغير خطية بطبيعتها. ومع ذلك، فإن القدرة على التنبؤ ولو بجزء بسيط من اتجاهات الأسهم يمكن أن تحمل **قيمة هائلة** لمختلف الأطراف.

* **حول البيانات:**

تُعد مجموعة البيانات المستخدمة في هذا المشروع حول **أسهم Netflix (المُشار إليها بالرمز NFLX)**، وهي شركة أمريكية رائدة في مجال خدمات البث المباشر وإنتاج المحتوى، تأسست في 29 أغسطس 1997. تشتهر Netflix بمكتبتها الضخمة من الأفلام والمسلسلات التلفزيونية، بالإضافة إلى إنتاجاتها الأصلية المعروفة باسم "Netflix Originals".

تكتسب بيانات أسهم Netflix أهمية خاصة حاليًا نظرًا **للتحديات المالية** التي واجهتها الشركة في الآونة الأخيرة. هذا التقلب يجعلها موضوعًا مثيرًا للاهتمام للدراسة والتحليل، حيث يمكن استخدام البيانات التاريخية لإنشاء نماذج قادرة على محاولة التنبؤ بتحركاتها المستقبلية.

حيث ان هذه البيانات تحتوي على معلومات أسهم الشركة لمدة 5036 يوم.

**معلومات عن السمات:**

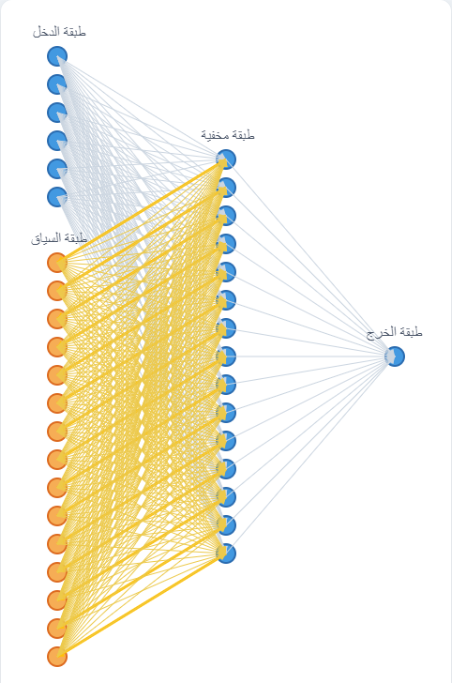
1. **(سعر الافتتاح):** سعر السهم عند بداية يوم التداول.
2. **(أعلى سعر):** أعلى سعر وصل إليه السهم خلال يوم التداول.
3. **(أدنى سعر):** أدنى سعر وصل إليه السهم خلال يوم التداول.
4. **(سعر الإغلاق):** سعر السهم عند نهاية يوم التداول.
5. **(سعر الإغلاق المعدل):** سعر الإغلاق بعد تعديله لأي عمليات تقسيم أسهم (Stock Splits) أو توزيعات أرباح. يُفضل استخدام هذا السعر في التحليل لأنه يعكس القيمة الحقيقية للعائد.
6. **(حجم التداول):** عدد الأسهم المتداولة في ذلك اليوم، وهو مؤشر على سيولة السهم واهتمام السوق به.

* في مرحلة التحضير، تم إجراء عملية معالجة أولية لمجموعة البيانات بهدف ضمان جودتها، حيث لوحظ وجود بعض القيم غير الصحيحة التي استدعت التصحيح. بعد التأكد من سلامة البيانات، تم تقسيمها إلى مجموعتين رئيسيتين:   
  خُصصت نسبة 80% من البيانات لعملية تدريب الشبكة العصبونية،   
  بينما تم حجز نسبة الـ 20% المتبقية كمجموعة بيانات اختبار لتقييم أداء النموذج على بيانات لم يرها مسبقًا.
* **بنية الشبكة:**

تم بناء النموذج باستخدام شبكة عصبونية اصطناعية تعتمد على نموذج شبكة Elman

، وتتألف بنيتها من الطبقات التالية:

* **طبقة الإدخال: (Input Layer)** مكونة من 6 عصبونات، حيث يستقبل كل عصبون إحدى الميزات (features) الـ 6 من مجموعة البيانات.
* **الطبقة المخفية: (Hidden Layer)** طبقة واحدة تحتوي على 15 عصبوناً، مسؤولة عن استخلاص الأنماط من بيانات الإدخال.
* **طبقة الإخراج: (Output Layer)** تحتوي على عصبون واحد فقط، وظيفته تخمين سعر السهم المعدل عند الاغلاق
* **طبقة السياق (context layer): تحتوي على 15 عصبون من اجل تذكر الخرج السابق**

وقد تم استخدام دالة التفعيل **Sigmoid** لكل العصبونات الموجودة في الطبقة المخفية وطبقة السياق طبقة الإخراج .

* **منهجية العمل والنتائج:**

اتبعت منهجية عمل منظمة بدأت بمرحلة **معالجة البيانات الأولية**، حيث تم تنظيف مجموعة البيانات لضمان جودتها، وتضمنت هذه المرحلة خطوتين أساسيتين: أولا، تصحيح الأخطاء في تنسيق الأرقام، مثل تحويل الفاصلة العشرية من "," إلى "." (على سبيل المثال: 15,5 أصبحت 15.5) ثانيا، تم حذف الصفوف (السجلات) التي تحتوي على حقول فارغة. بعد ذلك، تم **تقسيم البيانات**، حيث خُصصت نسبة **80%** من السجلات لمجموعة **التدريب** ونسبة **20%** المتبقية لمجموعة **الاختبار**.

في شبكة Elmanوبما ان الطبقة المخفية لديها نسخة من تنشيطاتها السابقة (من خطوة زمنية سابقة) يتم تغذيتها إلى "طبقة السياق" (Context Layer). هذه الطبقة السياقية تعمل كنوع من الذاكرة قصيرة المدى للشبكة.

تم اعتماد طول الخطوة الزمنية 5 من اجل الوصول لحل امثل.

للوصول إلى البنية المثلى للشبكة، تم إجراء سلسلة من التجارب المنهجية مع تغيير عدد العصبونات في الطبقة المخفية ومعدل التعلم وكانت أبرز النتائج كالتالي:

* **عند استخدام 10 عصبونات:**

معدل التعلم 0.1 و(عدد الخطوات الزمنة 5) , بعد 10000 دورة تدريب تم الوصول الى معدل خطأ 0.00001% من اجل البيانات التي تدرب عليها ومعدل خطأ 0.004% من اجل البيانات الجديدة.

* **عند استخدام 10 عصبونات:**

معدل التعلم 0.1 و(عدد الخطوات الزمنة 1) , بعد 10000 دورة تدريب تم الوصول الى معدل خطأ 0.002% من اجل البيانات التي تدرب عليها ومعدل خطأ 0.25% من اجل البيانات الجديدة.

* **عند استخدام 15 عصبوناً:**

معدل التعلم 0.1 و(عدد الخطوات 5) , بعد 10000 دورة تدريب تم الوصول الى معدل خطأ 0.000008% من اجل البيانات التي تدرب عليها ومعدل خطأ 0.004% من اجل البيانات الجديدة.

تُظهر هذه المقارنات أثر تغيير هذه البارامترات بشكل مباشر على أداء النموذج، وقد ساعدت هذه العملية في تحديد التوليفة الأفضل من البارامترات التي قادت في النهاية إلى تحقيق أفضل نتيجة ممكنة على بيانات الاختبار، والتي هي:

1. 15 عصبونا في الطبقة المخفية (hidden layer) وعصبون واحدة لطبقة الخرج
2. معدل تعلم0.1
3. تابع تفعيل لكلا من الطبقتين hidden, output , context, هو تابع Sigmoid
4. 0,0001دورة تدريب لكامل البيانات (entire dataset)

رسم بياني للخطأ اثناء التدريب:

