تقرير مشروع الذكاء الاصطناعي النهائي

تحليل البيانات والتعلم الآلي (Regression & Clustering)

1. تحليل التوقع (Linear Regression)

تم إجراء تجربة لمقارنة أداء النموذج بين البيانات النظيفة والبيانات المشوشة (Synthetic) لقياس مدى تأثير جودة البيانات على دقة التوقع.

نوع البيانات	قيمة الخطأ (MSE)	النتيجة
بيانات نظيفة (Clean Data)	**11.3675**	دقة عالية جداً
بيانات مشوشة (Synthetic Data)	**74.0740**	خطأ مرتفع بسبب الضوضاء

الاستنتاج: جودة البيانات هي العامل الأساسي في تقليل نسبة الخطأ؛ حيث ارتفع الخطأ بنسبة كبيرة عند وجود قيم مفقودة ومشوشة.

2. تقسيم العملاء (Customer Segmentation)

باستخدام خوارزمية K-Means، تم تحليل سلوك العملاء بناءً على معدل الإنفاق السنوي وتقييم العميل.

طريقة الكوع (Elbow Method): أظهر التحليل أن أفضل عدد للمجموعات هو **3 مجموعات**.(Clusters)

المجموعة (Cluster)	الوصف المقترح	الخصائص
اللون الأحمر	العملاء المستقرون	إنفاق متوسط وتقييم متوسط
اللون الأزرق	عملاء واعدون	إنفاق منخفض مع تقييم مرتفع
اللون الأخضر	عملاء VIP (الحيتان)	إنفاق عالٍ جداً وولاء مرتفع

3. الملخص الفني

- تمت معالجة القيم المفقودة (NaN) باستخدام المتوسط الحسابي.
- تم تحويل جميع القيم الرقمية السالبة إلى موجبة لضمان منطقية البيانات.
- تم استخدام StandardScaler لتوحيد مقاييس البيانات قبل عملية التقسيم.

تم إعداد هذا التقرير آلياً بناءً على مخرجات الأكواد البرمجية المنفذة.