تقريرعن Dataset (Breast Cancer Prediction)

تقرير تحليل بيانات سرطان الثدي في ويسكونسن الأصلي، والتي تم التبرع بها في 14 يوليو 1992، وهي متاحة في مستودع UCI للتعلم الآلي تستخدم هذه المجموعة عن نطاق واسع في مهام التصنيف لتمييز الأورام الحميدة Benign عن الأورام الخبيثة Malignant

1 مصدر البيانات (Data Source)

المصدر:

- University of Wisconsin Hospitals, Madison
- تم جمع البيانات بواسطة Dr. William H. Wolberg

2 هدف البيانات (Objective)

التنبؤ إذا كان الورم:

- M = Malignant سرطاني خبيث
- B = Benign غير سرطاني

باستخدام خصائص يتم استخراجها من صور أو فحوصات الخلايا.

3 شكل البيانات (Dataset Structure)

- 683 عينة (samples)
- 9 أعمدة (features)
- عمود الهدف (Diagnosis)

📊 جدول الأعمدة ووصفها:

اسم العمود	القيم	النوع	الوصف
Sample Code Number		ID	رقم تعريف فريد لكل عينة
Clump Thickness	(1 - 10)	Feature (int)	سماكة تكتل الخلايا (من 1 إلى 10) — القيم الأعلى قد تشير لورم غير طبيعي
Uniformity of Cell Size	(1 - 10)	Feature (int)	مدى تجانس حجم الخلايا — عدم التجانس قد يدل على سرطان
Uniformity of Cell Shape	(1 - 10)	Feature (int)	مدى تجانس شكل الخلايا — اختلاف الشكل قد يشير إلى خلايا سرطانية
Marginal Adhesion	(1 - 10)	Feature (int)	درجة التصاق الخلايا ببعضها — انخفاض الالتصاق قد يكون علامة على الورم
Single Epithelial Cell Size	(1 - 10)	Feature (int)	حجم الخلايا الظهارية المفردة — القيم الكبيرة قد ترتبط بالخباثة
Bare Nuclei	(1 - 10)	Feature (int)	عدد الأنوية العارية (بدون سيتوبلازم) — مؤشر مهم على الخباثة
Bland Chromatin	(1 - 10)	Feature (int)	نسيج الكروماتين داخل النواة — عدم الانتظام قد يدل على سرطان
Normal Nucleoli	(1 - 10)	Feature (int)	عدد وشكل النوية الطبيعية داخل الخلايا
Mitoses	(1 - 10)	Feature (int)	معدل الانقسام الخلوي — ارتفاعه غالبًا يدل على خلايا سرطانية نشطة
Class	(2 أو 4)	Target	التصنيف النهائي: 2 = حميد(Benign) ، 4 = خبيث (Malignant)