

القسم الثاني (1)

لبناء أي نموذج، يجب أن تخضع لعمليات عدة لا بد منها قبل البدء في عملية إنشاء النموذج أولاً يجب تحديد الهدف من النموذج الذي تعمل عليه هو خطوة أساسية في عملية تطوير النماذج فهم الهدف يساعدك:

- على تحديد كيفية صياغة المشكلة.

- اختيار الخوارزميات المناسبة.

- تحديد مقياس الأداء لتقييم النموذج.

ومن ثم يجب طرحها على صاحب المشروع او القائد هو كيف تبدو الحلول الحالية (إن وجدت) الوضع الحالي غالباً ما يوفر مرجعاً للأداء، بالإضافة إلى رؤى حول كيفية حل المشكلة.

ثانياً نحدد نوع الإشراف التدريبي الذي سيحتاجه النموذج هل هو تعلم تحت إشراف أو بدون إشراف أو شبه تحت إشراف، أو تعلم ذاتي أو تعلم تعزيزي وهل هو مهمة تصنيف أو مهمة انحدار أو شيء آخر، هل يجب استخدام تقنيات التعلم الدفعي أو التعلم الأونلاين.

هذه العملية تساعد في تحديد الأولويات والتركيز على الجوانب الأكثر أهمية للمشروع، وتجنب العمل على مهام قد لا تضيف قيمة حقيقية للنتائج النهائية.

ثالثاً نقوم باستيراد البيانات واقرءها في ملف البايتون ثم ابداء بالتعامل مع هذه البيانات وافهم سلوك البيانات مثلاً الإحصاءات الوصفية :

- **العدد (Count):** يشير إلى عدد القيم غير الفارغة على سبيل المثال.

- **المتوسط (Mean):** متوسط قيم البيانات.

- **الأدنى (Min):** أصغر قيمة في البيانات.

- **الأعلى (Max):** أكبر قيمة في البيانات.

- **الانحراف المعياري (std):** يقيس مدى تشتت القيم (انحراف معياري منخفض يعني أن القيم تميل للقرب من المتوسط، بينما انحراف معياري مرتفع يعني أن القيم موزعة على نطاق أوسع).
- **النسب المئوية الـ 100:** أكبر قيمة في العاود.

- **النسبة المئوية الـ 25 (الربع الأول):** تحت هذه القيمة، يقع 25% من البيانات.

- **النسبة المئوية الـ 50 (الوسيط):** نصف البيانات تحت هذه القيمة.

- **النسبة المئوية الـ 75 (الربع الثالث):** تحت هذه القيمة، يقع 75% من البيانات.

رابعاً أقوم برسمه توزع هذه البيانات وكيفية سلوكها من اجل ان تعامل بها بأريحية واقوم بتصحيح الأخطاء فيها والاستفسار عن التوزعات الغير مفهومة ومن ثم أقوم بتقسيم البيانات الي بيانات تدريب وبيانات اختبار التي تشكل 20% من البيانات وهذا التقسيم يجب ان يكون ثابت بحيث ليس في كل مرة أقوم بتشغيل الكود يحدث اختلاف مع الانتباه الى عدم حدوث تجيز في البيانات في عملية التقسيم أي تكون الفئات مقسمة في التدريب والاختبار بشكل صحيح ومن ثم نقوم بدراسة معامل الارتباط بين الاعمدة في البيانات ونرسمه لتحقيق تصور لهذه العلاقات .