**تكليف مقدمة الى علم البيانات**

**عمل الطالبة/ تحت اشراف/**

**أزهار أحمد القديمي. د/ حسين الخضر.**

**(نظري)**

1**:- ماهي اهمية تحليل البيانات وعلم البيانات بشكل عام بالنسبة للشركات التجارية والمؤسسات الحكومية والقطاعات المختلفة في بلادنا؟ دلل بالأمثلة الواقعية والملموسة بدون الاستعانة بالذكاء الاصطناعي؟**

* **أهميته تكمن في إتاحة الفرصة لاتخاذ قرارات أفضل، وزيادة الكفاءة، وتحسين الخدمات المقدمة.**
* **الأمثلة الواقعية والملموسة:**

1. تحليل البيانات في الشركات التجارية.

مثل شركات الاتصالات مثل "يمن موبايل" و"سبأفون".

المحلات التجارية الكبرى (مثل محلات الجملة في صنعاء وعدن).

1. تحليل البيانات في المؤسسات الحكومية.

مثل وزارة الصحة العامة والسكان

حملات التوعية واللقاحات إلى المحافظات الأشد تأثرًا.

وزارة التربية والتعليم

تحليل نتائج امتحانات الثانوية العامة لاكتشاف ضعف التحصيل في مناطق معينة.

بناء خطط لتطوير التعليم في المديريات المتأخرة.

2:- مع تطور مصادر البيانات بشكل كبير وتزايد يومياً , كم هو معدل زيادة البيانات بشكل سنوي(ابحث عن ذلك من مصادر علمية موثوقة)؟

المعدل السنوي لزيادة البيانات العالمية يتراوح تقريباً بين 24% و 22% أي ان حجم البيانات يتضاعف تقريباً كل 3-4 سنوات.

**3:- اذكر اهم خمس تقنيات ومنهجيات وبرمجيات رائدة في مجال علم البيانات وتحليل البيانات؟ مع ارفاق صور للنافذة الرئيسية لكل تقنية؟**

1. Python
2. R
3. Tableau
4. Apache Spark
5. Power BI

**(عملي)**

1**:- اكتب ملخص عن اهم عشرة من الاوامر التي شدتك للانتباه من مكتبة Numpy من 40 التمرين بشر ط ان لا تكون من الاوامر والامثلة التي ذكرت في المحاضرة؟**

**1\_ توزيع Zipf باستخدام دالة الزيتا.**

plt.plot(k, n\*(k\*\*-a)/zeta(a), 'k.-', alpha=0.5, label='expected count')

الوصف: يمثل توزيع Zipf الاحتمالي، حيث التكرار يقل عكسيًا مع الرتبة.

التقنية: zeta) (من SciPy) تطبّق لتطبيع توزيع Zipf.

الاستخدام: واسع في تحليل النصوص (NLP)، التوزيع السكاني، علوم الشبكات.

الميزة: يعرض القانون الأسّي بطريقة دقيقة باستخدام تطبيع رياضي.

**2\_ اختبار T اليدوي (T-test).**

t = (np.mean(intake) 7725)/(intake.std(ddof=1)/np.sqrt(len(intake)))

الوصف: حساب يدوي لاختبار T لعينة واحدة.

الاستخدام: مقارنة متوسط عينة بقيمة معيارية.

الوظيفة: يُستخدم لتقييم الفرضيات الإحصائية في التجارب الطبية والبحثية.

الميزة: لا يعتمد على مكتبة خارجية مثل `scipy.stats`.

**3\_ توزيع Logseries وتحليل الكثافة.**

s = rng.logseries(a, 10000)

plt.plot(centres, logseries(centres, a) \* s.size, 'r')

الوصف: توزيع احتمالي للأحداث النادرة.

الوظيفة: يستخدم لتمثيل البيانات النادرة (مثل الكلمات القليلة التكرار).

التميّز: قليل الاستخدام نسبيًا، لكنه مهم لنمذجة البيانات الثقيلة الذيل.

**4\_ توزيع Pareto وتحليل دالة الكثافة (PDF.**

pdf = a / (x+1)\*\*(a+1)

plt.plot(x, pdf, linewidth=2, color='r')

الوصف: حساب يدوي لكثافة توزيع Pareto.

الاستخدام: تحليل توزيع الثروات، الأعطال، أو المخاطر الطبيعية.

الميزة: يوضح قانون 80/20 بطريقة واضحة.

**5\_ توزيع Cauchy مع تقليم الأطراف.**

s = s[(s > -25) & (s < 25)]

الوصف: توزيع ثقيل الذيل، لا يحتوي على متوسط أو تباين معرف.

الوظيفة: تقليم البيانات لتحسين التمثيل البياني.

التميّز: نادر الاستخدام، لكنه مهم في الحالات المتطرفة.

**6\_ تحليل توزيعي F و Noncentral-F.**

plt.plot(F[1][1:], F[0])

plt.plot(NF[1][1:], NF[0])

الوصف: مقارنة توزيع F القياسي مع نظيره غير المركزي.

الاستخدام: مفيد في ANOVA وتحليل النماذج الإحصائية المعقدة.

الميزة: يوضح تأثير الانحراف المركزي على الشكل الإحصائي للتوزيع.

**7\_ استخدام `np.polyfit()` لتقريب متعدد الحدود.**

z = np.polyfit(x, y, 3)

الوصف: حساب معاملات دالة تقريب من الدرجة الثالثة.

الاستخدام: تحليل الانحدار، وتفسير العلاقات غير الخطية.

الميزة: دالة قوية ومرنة لبناء نماذج بيانية بسهولة.

**8\_ استخدام `np.ndenumerate()` لاستعراض العناصر بمؤشراتها.**

for index, x in np.ndenumerate(a):

print(index, x)

الوصف: استعراض كل عنصر مع موقعه داخل مصفوفة متعددة الأبعاد.

الاستخدام: التصحيح اليدوي، تحليل مصفوفات الصور أو البيانات المجمعة.

الميزة: أداة قوية لفهم وتفقد البيانات بدقة.

**9- تحليل الفرق بين توزيعين: `noncentral\_chisquare` و `chisquare**

plt.plot(values[1][0:-1], values[0] - values2[0], 'ob')

الوصف: توضيح الفرق بين توزيعات كاي-تربيع المركزية وغير المركزية.

الاستخدام: مقارنة الأثر الإحصائي للمعلمات المختلفة.

الميزة: تمثيل مرئي دقيق للفروقات التراكمية بين توزيعين.

**10\_ تقاطع بيانات مقنعة باستخدام `np.ma.intersect1d()`.**

np.ma.intersect1d(x, y)

الوصف: تقاطع بين مصفوفتين مع دعم للبيانات المقنعة (masked arrays).

الاستخدام: التعامل مع بيانات تحتوي على قيم مفقودة أو محجوبة.

الميزة: يجنب الأخطاء عند التعامل مع بيانات غير مكتملة، وهو مفيد جدًا في تنظيف البيانات.