بسم الله الرحمن الرحيم

جامعة النيلين

ماجستير تقنية المعلومات

الدفعة السادسة عشر

**Project clustering(3)**

المشكلة:

شركة تبيع منتج الدقيق وتريد ان تعرف مدى جودة اداء البيع للدقيق .

يمكننا تقسيم العملاء بناءً على سلوكهم الشرائي في السوق.

خلفية المشكلة:

السودان منذ عام 2018 زلت الشعب السوداني يعاني من أزمة الخبز وباتت مشهد الصفوف اليومي أمام المخابز مألوف اكثر من ذي قبل ،رغم مخاطر فيروس كورونا الذي تنتشر رفعة عدواه في الزحام.

التقنيات:

Clustering هو تقنية تعليمية غير خاضعة للإشراف تأخذ بيانات الإدخال وتنظمها في مجموعات بحيث ينتهي الأمر بالكائنات المتشابهة في نفس المجموعة. يمكن استخدام هذا .

K-mean Clustering خوارزمية

لأداء مهمة التجميع. يقوم بتعيين جميع نقاط البيانات في مساحة السودان ، ثم يضع بشكل عشوائي مراكز مجموعة k في الفضاء (الأمر متروك للمبرمج ليقرر العدد ). كل مركز كتلة هو مجرد نقطة في الفضاء.

بعد ذلك يتم تعين جميع الأقرب إلي مركز كل مجموعه عن اي مركز آخر .

بعد تكرار العملية نزل كل نقطة في نفس المجموعه التي كانت عليها من قبل .

نكون قد وصلنا الي توازن ونهاية الخوارزمية .

مستخدمين لغة بايثون في البرمجة .

**التحليل:**

جمع البيانات

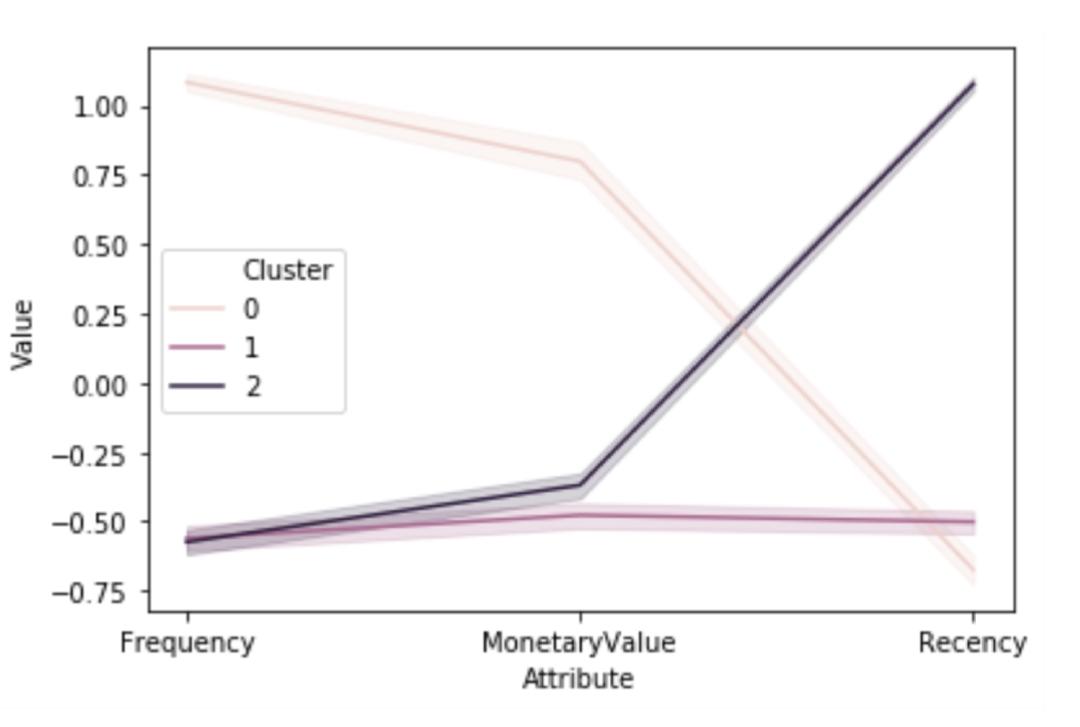
في هذه الخطوة ، سنجمع البيانات أولاً. في هذه الحالة ، سنأخذ البيانات من UCI Machine Learning التي تسمى مجموعة بيانات البيع بالتجزئة عبر الإنترنت.

يمثل كل صف المعاملة التي تحدث. يتضمن اسم الدقيق والكمية والسعر والأعمدة الأخرى التي تمثل المعرف.

**النتيجة:**

**ها هي النتيجة**

**~**

****

نستنتج أن المجموعة 0 متكررة ، وتنفق أكثر ، ويشترون المنتج مؤخرًا. لذلك ، يمكن أن تكون مجموعة العملاء المخلصين .

بعد ذلك ، تكون المجموعة 1 أقل تكرارًا وأقل إنفاقًا ، لكنهم يشترون المنتج مؤخرًا. لذلك ، يمكن أن تكون مجموعة العملاء الجدد .

أخيرًا ، تكون المجموعة 2 أقل تكرارًا ، وأقل إنفاقًا ، ويشترون المنتج في الوقت القديم. لذلك ، يمكن أن يكون مجموعة من العملاء المتخبطين

References

[1] Daqing C., Sai L.S, and Kun G. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining (2012), Journal of Database Marketing and Customer Strategy Management.

[2] Millman K. J, Aivazis M. Python for Scientists and Engineers (2011), Computing in Science & Engineering.

[3] Radečić D. Top 3 Methods for Handling Skewed Data (2020), Towards Data Science.

[4] Elbow Method for optimal value of k in KMeans, Geeks For Geeks.

مجموعة (9)

احمد محمد حسين رمضان

سوزان بشير عبدالله عثمان

مزمل محمود عمر محمود