**Data Mining Project**

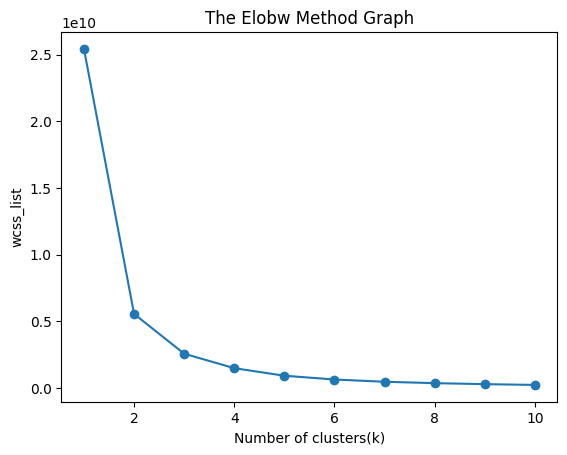
**في القسم الأول** تم الحديث عنه في doc\_classifier

**في القسم الثاني** تم تطبيق kmeans مع the elbow method لإيجاد العدد المثالي للعناقيد

تم استدعاء kmeans من sklearn  
تم المرور على عدد cluster من 1 إلى 10 وتطبيق kmeans في كل مرة ثم إضافة قيم intertia إلى مصفوفة wcss\_list

تم رسم الخط البياني بين عدد cluster و مصفوفة wcss\_list

نلاحظ من الشكل أن العدد المثالي للعناقيد هو k=2 بسبب وجود نقطة ذروة واضحة عند هذا العدد



تم تطبيق kmeans مع عدد cluster يساوي 2

تم حساب عدد instance لكل cluster وطباعة النتائج

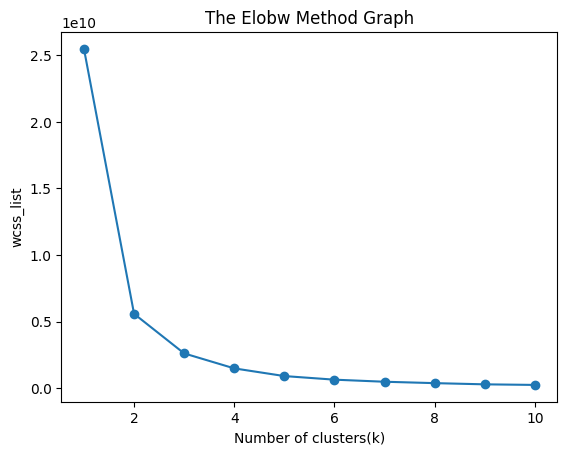
**في القسم الثالث** تم تطبيق minibatchkmeans مع the elbow method لإيجاد العدد المثالي للعناقيد

تم استدعاء minibatchkmeans من sklearn  
تم المرور على عدد cluster من 1 إلى 10 وتطبيق minibatchkmeans مع batch\_size=100 حيث في minibatchkmeans يتم اخذ الداتا على batchs وليس دفعة واحد كما يحصل في kmeans

إضافة قيم intertia إلى مصفوفة wcss\_list

تم رسم الخط البياني بين عدد cluster و مصفوفة wcss\_list

نلاحظ من الشكل أن العدد المثالي للعناقيد هو k=2 بسبب وجود نقطة ذروة واضحة عند هذا العدد



تم تطبيق minibatchkmeans مع عدد cluster يساوي 2

تم حساب عدد instance لكل cluster وطباعة النتائج

**إعداد:**

**علي ديب زيد ناصر**

**عمار معلا**