**ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC**

**1. Tiêu đề:**

Thuật toán K – Means phân cụm dữ liệu và ứng dụng

**2. Đối tượng nghiên cứu:**

Đối tượng nghiên cứu của thuật toán K – Means phân cụm dữ liệu , nghĩa là tìm cách phân cụm dữ liệu vào những nhóm khác nhau sao cho những dữ liệu trong cùng một nhóm có đặc điểm tương tự nhau và khác với những dữ liệu ở nhóm khác .

**3. Mục tiêu:**

- Hiểu được bài toán phân cụm dữ liệu .

- Nghiên cứu về các độ đo khoảng cách .

- Hiểu được thuật toán K – Means và lập trình trên ngôn ngữ Python .

- Ứng dụng được cho bài toán thực tế cụ thể là bài toán tách vật thể ảnh .

**4. Ý nghĩa khoa học:**

Thuật toán K – Means là một công cụ quan trọng trong lĩnh vực khoa học dữ liệu vì nó giúp chúng ta phân tích và khám phá các cấu trúc và các tiềm ẩn trong tập dữ liệu lớn . Cụ thể , K – Means phân cụm dữ liệu bằng cách tìm các phân loại các điểm dữ liệu vào các cụm dữ liệu khác nhau sao cho các điểm dữ liệu trong cùng một cụm có tính chất giống nhau và khách nhau với các điểm dữ liệu trong các cụm khác . Qúa trình này giúp giảm thiểu số lượng dữ liệu cần xử lý , giúp chúng ta dễ dàng phân tích và hiểu được các mẫu dữ liệu . Điều này có ý nghĩa rất lớn trong ngành khoa học dữ liệu .

**5. Ý nghĩa thực tiễn:**

K – Means có ý nghĩa rất lớn trong nhiều lĩnh vực khác nhau , bao gồm kinh doanh , tài chính , y tế , khoa học xã hội , giáo dục , thị trường và nhiều lĩnh vực khác . Trong kinh doanh thuật toán được áp dụng nhiều để phân tích dữ liệu khách hàng giúp cải thiện chiến lược kinh doanh , giảm chi phí và tăng doanh số . Tài chính nó còn có thể được sử dụng để phân tích các dữ liệu tài chính và giúp các công ty tìm ra các mẫu tiềm ẩn trong tình hình tài chính của họ , từ đó đưa ra các quyết định kinh doanh thông minh , v.v .

**6. Tính cấp thiết:**

Thuật toán K – Means có tính cấp thiết cao trong thời đại số hóa và dữ liệu ngày nay . Với sự phát triển của công nghệ thông tin , các dữ liệu được sản sinh ra với tốc độ ngày càng nhanh , đồng thời cũng tăng cường tính phức tạp và đa dạng của dữ liệu . K – Means là một trong những phương pháp phâm cụm phổ biển nhất và hiệu quả cáo , giúp cho việc khai thác , xử lý , và hiểu được dữ liệu trở nên dễ dàng hơn .

**7. Tính khả thi:**

Thuật toán K – Means có ưu điểm đơn giản dễ hểu , tốc độ xử lí nhanh , hiệu quả độ chính xác cao , dễ dàng mở rộng có ứng dụng đa dạng . Vậy nên tính khả tinh là rất cao .

**8. Nội dung chi tiết:**

- Tìm hiểu về phân cụm dữ liệu : phân cụm dữu liệu là gì ? Các phương pháp phân cụm , tìm hiểu về một số độ đo khoảng cách được dùng trong phân cụm .

- Tìm hiểu về thuật toán K- Means và lập trình trên ngôn ngữ Python .

- Ứng dụng vào bài toán tách vật thể ảnh .

**9. Kế hoạch thực hiện:**

