

อัลกอริทึมเค-มีนส์ เป็นวิธีที่นิยมใช้ในการจัดกลุ่มข้อมูลในปัจจุบัน อย่างไรก็ตามเมื่อข้อมูลเข้ามีขนาดใหญ่อัลกอริทึมจะทำงานช้ามาก เพราะอัลกอริทึมเค-มีนส์มีความซับซ้อนของเวลาเป็น  $O(krn)$  เมื่อ  $k$  เป็นจำนวนของกลุ่มที่ต้องการ  $r$  เป็นจำนวนของการวนทำงาน และ  $n$  เป็นจำนวนของจุด เป็นที่รู้กันว่า เค-มีนส์มีลักษณะเป็นเชิงเส้น (linear rate) เมื่อเข้าใกล้การเบรเข้าหากัน (convergence) ถ้า  $r$  เป็นสัดส่วนกับ  $n$  เวลาที่ใช้ในการทำงานของ เค-มีนส์จะเท่ากับ  $O(n^2)$  ซึ่ง  $n$  ที่มีขนาดใหญ่มาก จะไม่สามารถทำงานได้จริงในขอบเขตของเวลาที่ต้องการ ในผลการค้นคว้านี้ได้นำเสนออัลกอริทึมการจัดกลุ่มข้อมูลด้วยวิธีการใช้ต้นไม้แบบทอดข้ามค่าสุด มีความซับซ้อนของเวลาเป็น  $O(n \lg n)$  ซึ่งมีเวลาทำงานเร็วกว่าอัลกอริทึมเค-มีนส์ นอกจากนี้อัลกอริทึมการจัดกลุ่มข้อมูลด้วยวิธีการใช้ต้นไม้แบบทอดข้ามค่าสุดยังมีผลลัพธ์ที่ได้ดีกว่าอัลกอริทึมเค-มีนส์

K-means algorithm is one of the most popular clustering algorithms in use today. However, when the input size is large, the algorithm runs very slow. This is because it has time complexity of  $O(krn)$  where  $k$  is the number of desired clusters,  $r$  is the number of iterations, and  $n$  is the number of points. It is known that K-means approaches a linear rate when it is close to convergence. If  $r$  is proportional to  $n$ , the running time of K-means is  $O(n^2)$  which is impractical for a large  $n$ . In this papers, we offer the minimum spanning tree-based clustering (MST-based clustering). It has time complexity of  $O(n \lg n)$  and its running time is faster than that of K-means algorithm. Besides, the clustering result of our algorithm is better than that of K-means algorithm.