



خوشه بندی دیتاست Iris Flowers با استفاده از دو الگوریتم خوشه بندی k-means و الگوریتم DBSCAN در نرم افزار داده کاوی با پایتون (python)

محمدعلی جوادزاده^{۱*}، سعید خدادادیان^۲

۱- استادیار دانشگاه جامع امام حسین علیه السلام

۲- دانشجوی کارشناسی ارشد دانشگاه امام حسین (ع)

چکیده

خوشه بندی^۱، کار بخش بندی داده ها در گروه هایی است که به آن ها خوشه^۲ گفته می شود. هدف از خوشه بندی، تقسیم کردن داده ها به صورتی است که در یک خوشه حداکثر مشابهت میان نقاط داده و بین داده ها در خوشه های گوناگون، کمترین مشابهت وجود داشته باشد. به عبارت دیگر، هدف از خوشه بندی قرار دادن نقاط داده در خوشه هایی با حداکثر مشابهت درون خوشه ای و حداقل مشابهت میان خوشه ای است. ابتدا الگوریتم خوشه بندی K-Means روی یک مجموعه داده اعمال و سپس الگوریتم DBSCAN روی آن اعمال می شود. سپس، خروجی های هر دو الگوریتم مورد بررسی قرار می گیرند و با یکدیگر مقایسه می شوند. لازم به ذکر است که دو الگوریتم نامبرده از الگوریتم های بدون نظارت هستند که در این مقاله بر روی دیتاست Iris Flowers اعمال شده است.

واژگان کلیدی: خوشه بندی، DBSCAN، K-Means، Iris Flowers

¹ Clustering

² Cluster