

بهبود ایجاد شده در کد استفاده از خوشه‌بندی تجمعی به عنوان جایگزینی برای k-means و همچنین استفاده از پارامترهای بهتری برای تابع "TfidfVectorizer" است. در جداول زیر نتایج الگوریتم اولیه و بهبود داده شده مقایسه شده‌اند.

جدول ۱. تعداد نمونه‌ها در هر خوشه در الگوریتم اولیه و بهبود داده شده

تعداد نمونه‌ها در هر خوشه	
{0: 782, 1:615, 2:510, 3:1480}	الگوریتم اولیه
{0:1476, 1:719, 2:119, 3:1073}	الگوریتم بهبود داده شده

با توجه با جدول ۱ الگوریتم بهبود داده شده توانسته است، تعداد نمونه‌ها را در خوشه‌ها به صورت متوازن تری نسبت به الگوریتم اولیه خوشه بندی کند.

جدول ۲. برچسب خوشه‌ها در الگوریتم اولیه و بهبود داده شده

برچسب‌های خوشه‌ی ۰	برچسب‌های خوشه‌ی ۱	برچسب‌های خوشه‌ی ۲	برچسب‌های خوشه‌ی ۳	
comp.graphics: 733 sci.space: 44 alt.atheism: 4 talk.religion.misc: 1	alt.atheism: 365 talk.religion.misc: 247 comp.graphics: 2 sci.space: 1	sci.space: 508 alt.atheism: 1 comp.graphics: 1	sci.space: 434 alt.atheism: 429 talk.religion.misc: 380 comp.graphics: 237	الگوریتم اولیه
alt.atheism: 664 talk.religion.misc: 578 sci.space: 150 comp.graphics: 84	sci.space: 687 comp.graphics: 15 alt.atheism: 9 talk.religion.misc: 8	alt.atheism: 112 talk.religion.misc: 7	comp.graphics: 874 sci.space: 150 talk.religion.misc: 35 alt.atheism: 14	الگوریتم بهبود داده شده

۱. تفسیر نتایج:

با توجه با جدول ۱ الگوریتم بهبود داده شده توانسته است، تعداد نمونه‌ها را در خوشه‌ها به صورت متوازن تری نسبت به الگوریتم اولیه خوشه بندی کند. در جدول ۲ تعداد برچسب‌ها در هر خوشه نشان داده شده که مشخص می‌کند آیا خوشه بندی درست انجام شده است؟ الگوریتم اولیه خوشه‌ی ۰ را تقریباً به دسته‌ی گرافیک اختصاص داده، خوشه‌ی ۱ را به دسته‌ی مذهبی و کفر اختصاص داده است که با توجه به همپوشانی این دو دسته طبیعی است، خوشه‌ی ۲ را به دسته‌ی فضا اختصاص دهد و مدل خوشه‌ی ۳ را نتوانسته است به درستی تشخیص دهد. از طرف دیگر الگوریتم بهبود داده شده خوشه‌ی ۰ را تقریباً به دسته‌ی مذهبی و کفر اختصاص داده، خوشه‌ی ۱ را به دسته‌ی فضا، خوشه‌ی ۲ را به دسته‌ی کفر اختصاص داده و خوشه‌ی ۳ را به دسته‌ی گرافیک اختصاص داده است که نسبت به الگوریتم اولیه بهتر خوشه‌بندی کرده است.