

## استفاده از روش ترکیبی الگوریتم ژنتیک و کی-مینز در گروه‌بندی تصور افراد از بدن خود

معصومه هوشمند<sup>۱</sup>؛ فاطمه معدنی<sup>۲</sup>؛ پرستو نقدی<sup>۳</sup>؛ پدram پیوندی<sup>۴</sup>

### چکیده

امروزه با رشد روزافزون فناوری، حجم وسیعی از داده‌ها با سرعت بالایی جمع‌آوری و انبارش می‌شوند. لذا اهمیت دست‌یابی و پردازش اطلاعات در راستای تصمیم‌گیری سازمان با دو پارامتر حداقل زمان و عدم دخالت انسان، احساس می‌شود که نمی‌توان به صورت یکپارچه از این حجم داده‌ها استفاده نمود، داده‌کاوی یکی از مهمترین روش‌هایی است که به مسئله استخراج اطلاعات از پایگاه داده‌ها می‌پردازد هدف از این مقاله استفاده از روش‌های داده‌کاوی به عنوان یک تکنیک به منظور گروه‌بندی تصور افراد از بدن خود با بررسی پارامترهایی اجتماعی و فردی راحتی مثل معیارهای اجتماعی، محل زندگی، علاقه و ارزش فرد می‌باشد. بدین منظور از طریق مطالعه میدانی ۴۰۱ پرسش‌نامه مؤثر در بین دانشجویان خوابگاه دختران دانشگاه یزد و گرفتن سایز (۱۸ اندازه از بدن افراد) و همچنین با بکارگیری روش داده‌کاوی ترکیبی کی-مینز و الگوریتم ژنتیک روابط بین پارامترهای اجتماعی و فردی راحتی و تصور افراد از بدن خود بررسی شد و با استفاده از الگوریتم ژنتیک یک مدلی برای توصیف داده‌ها نسبت به روش K-Means بکار گرفته شده است.

### کلمات کلیدی :

گروه‌بندی، کی-مینز، الگوریتم ژنتیک، راحتی روانشناختی

## Using GA and k- means hybrid method for grouping people Image of their body

Masoume Hushmand; Fateme madani; Soraya Naghdi; Pedram Payvandy

### Abstract:

Today the high volume of data are collecting are storing very fast. The important of obtaining and processing information are needed but without handing of human and time. Data-mining is one of the important ways that deals of exploiting information from the database and is defined with expression that discovery of hidden knowledge. In this paper, data mining as a technique for grouping people Image of their bodies is using and personal comfort with social parameters such as social norms, location, interest and Personal values are investigated so using of field study (spread questionnaire) from the number of students in girls dormitory of Yazd University and taking their sizes (size of persons shape) and also using of K- Means hybrid method and GA, relations between parameters such as social and individual comfort and image peoples of their body was examined. The results show that real shape of persons with their image shape that they are Significantly wrong about their size.

### Keywords

Grouping , k-means , genetic algorithm , Psychological comfort

۱. کارشناسی ارشد، مهندسی تکنولوژی نساجی، دانشگاه یزد. hooshmand.yazd@yahoo.com
۲. دانشجوی کارشناسی ارشد مهندسی تکنولوژی نساجی، دانشگاه یزد madani\_jut@yahoo.com
۳. کارشناسی مهندسی نساجی، دانشگاه یزد
۴. استادیار دانشکده مهندسی نساجی دانشگاه یزد . دانشگاه یزد. peivandi@yazd.ac.ir

## ۱. مقدمه

افراد تمایل دارند لباسی بپوشند که متناسب با بدن آنها باشد و فرم بدن هرکس متفاوت و منحصر به فرد است. توجه به تصور افراد از بدن خودشان از نظر روانشناسی و جامعه‌شناسی از کار Paul schailder در سال ۱۹۲۰ سرچشمه می‌گیرد. وی تصور از بدن را به عنوان تصویری از بدن که در ذهن شکل گرفته است یا به عبارت دیگر شیوه‌ای که بدن به نظر می‌آید تعریف می‌کند [۵]. Grogan از چندین مطالعه دریافت که تصور از بدن از فاکتورهای زیادی مثل خانواده، دوستان، معلم، همسر و جامعه تاثیر می‌گیرد [۶]. در سال ۱۹۹۷ Cash, Ansis, Strachan از رابطه بین تصور از بدن و اعتماد بنفس دریافتند که تصور از بدن متفی به کاهش اعتماد به نفس منجر می‌شود [۳]. Tompson Vandenberg در سال ۲۰۰۲ تصور از بدن با بررسی چهار جزء: خشنودی ذهنی کلی، احساسات، شناخت‌ها را ارزیابی کرده است. اندازه‌گیری‌های روانشناختی از تصور بدن یک یا بیشتر این جزء‌ها را ارزیابی می‌کند. زنان بیشتر از مردان تمایل دارند تا در مورد وزنشان قضاوت کنند [۹]. در سال ۲۰۰۲ Thomas Cash یک مدل رفتارشناختی از تصور از بدن را توسعه داد و آزمایش کرد. در ارزیابی از بدن بیشتر از نظر اجتماعی- فرهنگی، مشخصه‌های میان فردی، مشخصه‌های فیزیکی و ویژگی‌های شخصیت تاکید شده است [۲].

در هریک از پروژهای بیان شده و دیگر پایگاه‌ها، با حجم بالایی از داده‌های ذخیره شده مواجه هستیم که نیاز به ابزاری برای پردازش داده‌های ذخیره شده می‌باشد. داده کاوی یکی از مهمترین این روشها است که به وسیله آن الگوهای مفید در داده‌ها با حداقل دخالت کاربران شناخته می‌شوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می‌دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمانها اتخاذ شوند. داده کاوی در سال‌های اخیر، تاثیرات شگرفی در محیط‌های آکادمیک و صنعتی ایجاد کرده و کاربردهای فراوانی در زمینه‌های مختلف یافته است. به عنوان نمونه می‌توان به کاربردهای تجاری، مدیریت و کشف فریب، پزشکی، ورزشی، متن کاوی و وب کاوی اشاره کرد [۷]. در زمینه صنعت نساجی نیز تحقیقاتی در مورد تکنیک‌ها و روش‌های داده کاوی انجام شده است. آقای شهرابی و همکارش در سال ۲۰۱۰ با استفاده از روش داده کاوی یک سیستم جدید ساینزبندی توسعه داد [۸]. در سال ۲۰۱۰ با استفاده از تکنیک داده کاوی یک سیستم ساینزبندی برای پسران ایرانی دارای سنین بین ۱۶ و ۲۲، بر اساس نوع بدن مشخص گردید [۱]. نتایج نشان داد که سه نوع بدن و سیستم ساینزبندی دارای عملکرد مناسبی بودند. در سال ۲۰۰۸ پژوهش‌هایی تحت عنوان پیش بینی فروش پوشاک و منسوجات با استفاده از شبکه عصبی و خوشه‌بندی نیز انجام گرفت [۴]. در سال ۲۰۱۲ با استفاده از روش خوشه‌بندی-K-Means جهت گروه‌بندی فرم بدن با جامعه آماری ۱۰۲۰ دختران یزدی در محدوده سنی ۱۵ الی ۱۷ سال استفاده شد که تمام دختران یزدی دارای فرم بدن مشابه "ساعت شنی" هستند بنابراین با افزودن سه پارامتر دیگر، تقسیم بندی فرم بدن در سه گروه انجام شد [۱۰]. در تحقیقی دیگر در سال ۲۰۱۲ با مطالعه موردی دانشجویان دختر و پسر، تصور از بدن و مشکلات فرم قسمت‌های بدن انجام پذیرفت [۱۱]. هدف از انجام این مقاله ارائه یک روش مؤثر با استفاده از الگوریتم ژنتیک برای روش خوشه بندی K-Means می‌باشد. در این میان خوشه بندی خود را به عنوان یک قالب مفهومی و الگوریتمی غنی جهت تحلیل و تفسیر داده‌ها مطرح ساخته است. خوشه بندی به دنبال کشف ساختار در داده‌های جمع‌آوری شده می‌باشد [۱]. هدف از انجام این مقاله ارائه یک روش مؤثر ترکیبی کی- مینز و الگوریتم ژنتیک برای روش خوشه بندی داده‌ها می‌باشد. از الگوریتم پیشنهادی مدلی برای تصور افراد از بدن خود با بررسی روابط بین پارامترهای اجتماعی و فردی راحتی داده شد. نتایج نشان داد اکثریت افراد در مورد اندام واقعی اشتباه تصور می‌کنند بطوری که بیشتر اندام‌ها مثلث است ولی اکثر افراد سعی می‌کنند خودشان را به اندام ساعت شنی نزدیک کنند.

## ۲. جمع‌آوری داده

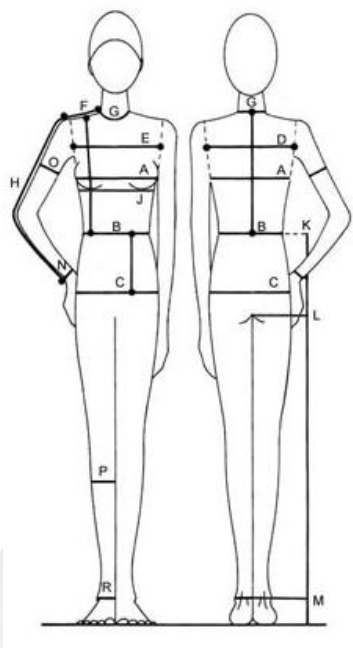
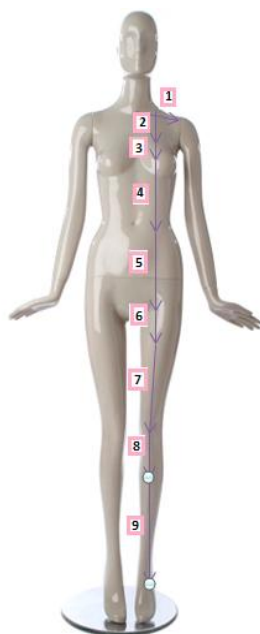
پرسشنامه شامل دسته‌ای از پرسش‌هاست که برطبق اصول خاصی تدوین گردیده است و به صورت کتبی به افراد ارائه می‌شود و پاسخگو، بر اساس تشخیص خود جواب‌ها را در آن می‌نویسد. هدف از ارائه پرسشنامه در این مقاله، کسب اطلاعات لازم از تصور افراد از بدن خود است. بزرگ بودن جامعه مورد مطالعه یکی از دلایلی بود که این روش را برای جمع‌آوری داده اهمیت بخشید. اطلاعات به دست آمده در این پروژه از طریق مطالعه میدانی (پخش پرسشنامه) در مجموعه خوابگاه دختران دانشگاه یزد به دست آمده است انجام کار بدین صورت بود که به هر فرد پرسشنامه‌ای داده می‌شد و اندازه‌هایی (۱۸ اندازه) نیز از بدن وی گرفته می‌شد شماره روی پرسشنامه هر فرد با شماره اندازه‌ها بررسی می‌شد که مثل هم باشند و اندازه همان فردی که پرسشنامه را پر کرده است داشته باشیم در پرسشنامه علاوه بر سوالات ۱۲ تصویر به هر فرد نشان داده می‌شد که به چند سوال پرسشنامه شامل تصویری که از اندام خود دارند اندام مورد علاقه‌شان و اندامی که فکر می‌کنند مورد علاقه یک مرد است پاسخ دهند و هر فرد باید یکی از شماره‌های (۱-۱۲) را به عنوان جواب سوال می‌نوشت.

### جدول (۱) سؤالات پرسشنامه

عامل مورد بررسی	نمونه سوالات
سن	سن -
وضعیت تاهل	- الف) مجرد □ ب) متاهل □
شهر	- شهر
وضعیت مذهبی	- شما جز کدامیک از گروههای زیر هستید؟ الف) مذهبی معمولی ب) مذهبی مقید ج) غیر مذهبی
افراد مؤثر در تصمیم گیری	- در انتخاب پوشش بیشتر تحت تأثیر کدامیک از موقعیتهای زیر هستید؟ الف) پدر و مادر ب) بازیگران ج) ورزشکاران د) روحانیون ه) کارکنان دانشگاه و) دوستان ی) هیچکدام
معیارهای اجتماعی مؤثر بر راحتی	- در شهری که شما زندگی میکنید در مورد پوشش کدامیک از موارد زیر بیشتر تسلط دارد؟ الف) قوانین حکومتی ب) سنتهای اجتماعی (عرفی) ج) احکام و دستورات دینی
عکس العمل افراد در مقابل خلاف هنجار بودن	- در جامعه ای که پوشش شما برخلاف معیارهای آن جامعه است چه عکس العملی نشان میدهید؟ الف) خودم را با معیارهای آنها تطبیق میدهم. ب) سعی میکنم با معیارها بجنگم و آنها را متقاعد کنم. ج) اگر مخالفت با جامعه، برایم مضر باشد، پوشش مرا مطابق معیارهای جامعه میکنم.
تأثیر مکان جغرافیایی در راحتی افراد	- اگر شما به مکان جغرافیایی دیگر (خارج از ایران) بروید چگونه لباسی می پوشید؟ الف) لباسهای با تری میپوشم چون از نظر فرهنگی تحت فشار نیستم. ب) هر جور مردم آن منطقه باشند مثل آنها لباس می پوشم. ج) در همه شرایط یک جور لباس می پوشم.
تأثیر علاقه در راحتی	- اگر شما پوششتان را خودتان انتخاب کنید چگونه لباسی میپوشید؟ الف) لباسی میپوشم که به پوشش معیار جهانی نزدیک باشم. ب) لباس محلی میپوشم. ج) پوششی را انتخاب میکنم که جامعه معرفی و بر آن تأکید دارد. د) لباسی میپوشم که مورد پذیرش فرهنگ کشورم باشد. (با تغییرات در طرح و رنگ)
تأثیر ارزش های فردی بر راحتی	- چه نوع پوششی برای شما جذاب و ایده آل است؟ الف) کاملاً پوشیده ب) متنوع و پوشیده ج) راحت د) کاملاً باز
تأثیر تصور از بدن در راحتی	- اگر شما به مهمانی بروید چگونه لباسی می پوشید؟ (افراد مهمانی خانم هستند) الف) به گونه ای لباس می پوشم که اندامم به چشم بیاید. ب) لباسی می پوشم که نقصه های اندامم را بپوشاند. ج) به گونه ای لباس می پوشم که اندامم اصلاً دیده نشود. د) از شرکت در مهمانی و جاهاییکه اندامم در معرض دید قرار گیرد دوری میکنم
الگوی دانشجویان در مورد پوشش و اندام	کدامیک از موارد زیر در پوشش و اندام، الگوی شما هستند؟ - الف) مادر یا خواهر ب) بازیگران و مدلها ج) دوستان و آشنایان د) الگویی ندارم. (تابع نظر خودم هستم.)
تصور افراد از شکل اندام خود	- به نظر شما کدامیک از تصاویر، به اندام شما نزدیک است؟
اندام ایده آل فرد	- کدامیک از تصاویر برای شما اندام ایده آل است؟
اندام ایده آل از نظر یک مرد	- به نظر شما کدامیک از تصاویر، از نظر یک مرد ایده آل است.
	<p>1 2 3 4 اندام مثلث معکوس</p> <p>5 6 7 8 اندام ساعت شنی</p> <p>9 10 11 12 اندام گلابی</p>

هدف از اندازه گیری سایز فرد پاسخگو، بدین دلیل بود که اندازه واقعی فرد را داشته باشیم تا بتوان با تصویری که از بدن خود دارد (تصاویر موجود در پرسشنامه) مقایسه گردد که در صورت اختلاف، عوامل مؤثر از این تفکر را با بررسی عواملی چون اجتماع، مکان زندگی، علایق فرد و تصور افراد از بدن خود را تجزیه و تحلیل کنیم. با توجه به توضیحات فوق، نمونه سؤالات پرسشنامه در جدول (۱)، سایزهای مورد نیاز هر فرد برای اندازه گیری در شکل (۱) و جدول (۲) ارائه شده است.

جدول (۲) راهنمای تصویر



شماره	سایز	شماره	سایز
G	دور گردن	۱	سرشانه
E	دور بغل	۲	سرشانه تا بغل
A	دور سینه	۳	سرشانه تا سینه
B	دور کمر	۴	سرشانه تا کمر
C	دور باسن	۵	سرشانه تا باسن
P	دور زانو	۶	سرشانه تا ران
O	دور ساق	۷	سرشانه تا زانو
P	دور مچ	۸	سرشانه تا ساق
L	دور ران	۹	سرشانه تا مچ

شکل (۱) اندازه های عمودی و افقی بدن

برای انجام مدل سازی، تعداد ۴۰۱ داده که شامل ۳۳ نوع متغیر می باشد، از پرسشنامه های مؤثر استخراج شد که متغیرها در جدول ۳ نشان داده شده است.

جدول (۳) پارامترهای مورد استفاده در فرآیند داده کاوی

شماره	نام متغیر	شماره	نام متغیر	شماره	نام متغیر
۱	دور گردن	۱۴	سرشانه تا باسن	۲۷	تأثیر علاقه در راحتی
۲	دور بغل	۱۵	سرشانه تا ران	۲۸	تأثیر ارزش های فردی بر راحتی
۳	دور سینه	۱۶	سرشانه تا زانو	۲۹	تأثیر تصور از بدن در راحتی
۴	دور کمر	۱۷	سرشانه تا ساق	۳۰	الگوی دانشجویان در مورد پوشش و اندام
۵	دور باسن	۱۸	سرشانه تا مچ	۳۱	تصور افراد از شکل اندام خود
۶	دور ران	۱۹	سن	۳۲	اندام ایده آل فرد
۷	دور زانو	۲۰	وضعیت تاهل	۳۳	اندام ایده آل از نظر یک مرد
۸	دور ساق	۲۱	شهر		
۹	دور مچ	۲۲	وضعیت مذهبی		
۱۰	سرشانه	۲۳	افراد مؤثر در تصمیم گیری		
۱۱	سرشانه تا بغل	۲۴	معیارهای اجتماعی مؤثر بر راحتی		
۱۲	سرشانه تا سینه	۲۵	عکس العمل افراد در مقابل خلاف هنجار بودن		
۱۳	سرشانه تا کمر	۲۶	تأثیر مکان جغرافیایی در راحتی افراد		

### ۳. روش کار

#### ۱.۳. کارهای انجام شده

داده های جمع آوری شده در برنامه اکسل مرتب گردید. سپس بر روی داده های خام متغیرها پردازش صورت گرفته و به صورت داده هایی که قابلیت تحلیلی بیشتری داشته باشند در آمدند. چون بعضی از سؤالات پرسش نامه بصورت ناقص تکمیل شده بود ، به همین علت مقادیر مربوط به تعدادی از سؤالات خالی می ماند، برای رفع این مشکل تنها از پرسش نامه های کامل استفاده گردید و عمل پاک سازی و تمیز کردن داده ها انجام شد. هنگامی که داده های مورد کاوش مشخص گردیدند، معمولاً" به تبدیل های خاصی روی داده ها نیاز است. تبدیل های ساده همچون تبدیل داده ها به داده های نرمال تعریف شدند. برای نرمال سازی داده ها از رابطه ۱ استفاده شد.

(۱)

که در رابطه ۱:

$P_{ij}^*$ : مقدار متغیر بعد از نرمال سازی

$P_{ij}$ : مقدار متغیر قبل از نرمال سازی

$\bar{P}_j$ : مقدار میانگین

$S_j$ : واریانس  $j$ ام متغیر از همه نمونه ها می باشد.

از متغیرهای نرمال شده به عنوان ورودی الگوریتم ژنتیک ترکیبی با الگوریتم **k-means** استفاده شد. اصولاً" وارد کردن داده ها به صورت خام باعث کاهش سرعت و دقت الگوریتم انتخابی می شود.

#### ۲.۳. الگوریتم ژنتیک ترکیبی با الگوریتم k-means

الگوریتم های ترکیبی زیادی براساس تئوری تکامل جهت بهبود خوشه بندی با روش **k-means** تاکنون ارائه شده اند. یکی از این روشها که همگرایی آن در بهینه کلی به صورت تئوری ثابت شده است، الگوریتم **GKA** می باشد که توسط **K.Kirishna** در سال ۱۹۹۹ معرفی گردیده است. عملکرد این الگوریتم براساس بهینه نمودن پارامتر مجموع واریانس درون خوشه ها (**TWCV**(total within cluster variation) می باشد که به آن مجموع مربعات خطای کلی (**SE**) نیز گفته می شود. که روش محاسبه آن به شرح زیر است.

(۲)

در معادله (۲):

$n$ : تعداد داده ها (patterns)

$d$ : تعداد پارامترهای هر داده (dimension)

$k$ : تعداد خوشه ها

$\{x_i, i=1, 2, 3, \dots, n\}$ : مجموعه داده ها می باشد.

در این صورت می توان ماتریس  $W$  را به صورت زیر تعریف نمود:

$$V \quad (3)$$

ماتریس  $W = [w_{ij}]$  دارای خصوصیات زیر می باشد:

$$w \quad (4)$$

اگر  $C_k = (C_{k1}, C_{k2}, C_{k3}, \dots, C_{kd})$  مراکز پارامترهای مرکز خوشه  $C_k$  باشد در این صورت برای محاسبه مراکز خوشه ها خواهیم داشت:

$$C \quad (5)$$

با استفاده از معادله (4) میزان واریانس درون هر خوشه برابر است با:

$$S \quad (6)$$

و در نهایت مجموع واریانس درون خوشه ها  $TWCV$  برابر خواهد بود با:

$$S \quad (7)$$

که هدف پیدا کردن  $W^*$  می باشد که در آن  $S(W^*)$  کمترین شود یعنی:

$$S \quad (8)$$

یکی از روش های معمول برای پیدا کردن خوشه با کمترین مربع خطا (SE)، استفاده از روش  $K\_maens$  می باشد (با محاسبه فاصله اقلیدسی). اگر نقاط اولیه در الگوریتم  $K\_maens$  خوب انتخاب نشده باشند الگوریتم در یک بهینه محلی خواهد افتاد و  $SW$  کوچکترین  $SW$  ممکن نخواهد بود.

- $GKA$  شامل مجموعه ای از جواب های کد شده می باشد که جمعیت اولیه بر اساس انتخاب متغیرها به صورت تصادفی مشخص شده است و در نسل های بعدی بر اساس عملگرهای الگوریتم ژنتیک صورت می پذیرد و تولید نسل ها تا زمان رسیدن به شرایط خواسته شده ادامه پیدا می کند. عملگرهای ژنتیکی که در  $GKA$  استفاده شده اند عبارت اند از:

عملگر انتخاب

عملگر جهش بر اساس فاصله

عملگر  $K\_maens$



در این بخش به نحوه کد کردن و عملگر عملگرهای ژنتیکی پرداخته می شود:

### ۱,۲,۳. کد گذاری

هدف، کدگذاری ماتریس  $W$  می باشد. برای این کار رشته ای به طول  $n$  (تعداد داده ها) در نظر گرفته می شود که اعداد داخل رشته نمایانگر خوشه ای می باشند که داده به آن تعلق دارد. که مقادیر داخل کروموزم  $\{1,2,3,\dots,K\}$  می باشند. به این روش کدگذاری string-of-group-numbers encoding گفته می شود.

### ۲,۲,۳. جمعیت اولیه

جمعیت اولیه  $P(0)$  به صورت تصادفی مشخص می گردد. رشته کروموزم از اعداد بین ۱ تا  $K$  به صورت تصادف با توزیع نرمال پر می شوند. برای اینکه احتمال دارد در فرآیند انتخاب تصادفی یک خوشه اصلاً انتخاب نشود برای جلوگیری از این امر یک عدد صحیح کوچکتر از  $n/K$  انتخاب شده و به این تعداد ژن یک گروه اختصاص داده می شود. این عمل برای  $K$  گروه تکرار شده و مابقی ژن ها به صورت دلخواه پر می شوند.

### ۳,۲,۳. انتخاب

عملگر انتخاب به صورت تصادفی و بر اساس توزیع زیر از جمعیت اقدام به انتخاب کروموزم می نماید.

$$P \quad (9)$$

که در آن  $F(S_i)$  معرف مقدار تابع برازندگی کروموزم  $S_i$  می باشد. جهت ارزیابی میزان خوب بودن کروموزم های موجود، نیاز به تابع برازندگی می باشد که در اینجا مجموع واریانس درون خوشه ها به عنوان معیار خوب یا بد بودن کروموزم ها در نظر گرفته شده است. هر چه  $S(W)$  کمتر باشد به این معنی است که کروموزم دارای برازندگی بیشتری می باشد. جهت تعریف تابع برازندگی بر اساس  $S(W)$  روش های زیادی موجود می باشد که در این جا از روش زیر استفاده شده است:

$$f \quad (10)$$

و

$$g \quad (11)$$

که در آن  $\bar{f}$  و  $\sigma$  معرف میانگین و انحراف معیار  $f(S_W)$  در نسل فعلی می باشد و ثابت  $C$  عددی مابین ۱ و ۳ می باشد. در نهایت تابع برازندگی به شرح زیر خواهد بود:

$$F \quad (12)$$

## کنفرانس داده کاوی ایران

### ۴,۲,۳. جهش

عملگر جهش بر پایه فاصله معرفی شده است. در جهش، مقدار یک ژن با توجه به فاصله نقطه انتخابی از مراکز خوشه ها تغییر پیدا می کند. در واقع هر ژن معرف یک داده و مقدار آن معرف خوشه ای است که آن داده به آن اختصاص دارد، می باشد. عملگر جهش به گونه ای تعریف می گردد که تمایل دارد مقدار ژن را به گونه ای تغییر دهد که داده به خوشه ای که به مرکز آن نزدیک تر است تعلق گیرد.

جهت انجام این عمل بر روی آن  $s_w(i)$  که معرف ژن انتخابی در کروموزم  $s_w$  می باشد و با توجه به این که ژن  $i$  ام در ارتباط با داده  $x_i$  می باشد می توان فاصله ژن از مرکز خوشه  $J$  به صورت زیر تعریف نمود:

$$d \quad (13)$$

و سپس با احتمال زیر خوشه داده با همان مقدار ژن جایگزین می گردد.

که در آن  $c_m$  عددی بزرگتر از یک و  $d_{\max} = \max_j\{d_j\}$  است. باید توجه داشت احتمال این که خوشه خالی در این عملگر به وجود آید، وجود دارد.

مثلاً اگر چهار خوشه داشته باشیم توسط عملگر جهش فقط خوشه های ۱ و ۲ و ۴ انتخاب شده باشند و خوشه ۳ اصلاً انتخاب نشده باشد و با توجه به این که هر چه تعداد خوشه کمتر باشد SE طبیعتاً بیشتر خواهد بود. باید از خالی شدن خوشه ها جلوگیری به عمل آید. یک روش سریع برای بررسی نمودن اینکه احتمال تشکیل خوشه خالی وجود دارد این هست که آیا جامعه داده  $x_i$  از خوشه خودش  $C_{sw}(i)$  بزرگتر از صفر هست یا نه. در صورتی که  $d_{sw}(i) = 0$  بدین منظور است که داده و مرکز خوشه یکی می باشد. بنابراین فقط در مواردی که فاصله بزرگتر از صفر است عملگر جهش می تواند بر روی داده اعمال گردد. کروموزمی که دارای K خوشه غیر خالی باشد کروموزم معتبر شناخته می شود و در غیر اینصورت کروموزم معتبر نیست.

### ۵.۲.۳. عملگر دورگه شدن K\_maens

در الگوریتم ترکیبی به جای عمل دو رگه شدن، عملگر K\_maens قرار داده شده است جهت افزایش سرعت همگرایی از الگوریتم K\_maens استفاده شد که اگر sw کروموزم انتخابی باشد مراحل زیر توسط عملگر K\_maens بر آن اعمال می گردد.

- ۱ - مراکز خوشه با توجه با فرمول (۵) برای ماتریس W محاسبه می گردد.
  - ۲ - هر داده به نزدیکترین خوشه نسبت داده می شود و ماتریس  $\tilde{W}$  تشکیل می گردد.
- احتمال دارد که حاصل این عملگر منجر به کروموزم نامعتبر گردد. در این حالت کروموزم نامعتبر به روش زیر به کروموزم معتبر تبدیل می گردد:
- ۱ - خوشه C که دارای بیشترین واریانس درون خوشه ای است انتخاب می گردد.
  - ۲ - به تعداد مورد نظر از دورترین داده ها به مرکز خوشه C جهت پر کردن خوشه خالی استفاده می گردد.

### ۳. نتایج و بحث

هدف از این تحقیق استفاده از روش خوشه بندی ترکیبی کی - مینز و الگوریتم ژنتیک برای گروه بندی تصور افراد از بدن خود است. این روش یکی از تکنیک های کشف الگوها از پایگاه داده های مشخص در داده کاوی است. ایده اصلی در این الگوریتم بهره مندی از مزایای هر دو الگوریتم k-means و ژنتیک می باشد و در واقع این ترکیب، امکان رسیدن به خوشه بندی بهینه را فراهم می آورد. در ادامه روش کار ، متغیر ها خوشه بندی شدند و متغیرها بر اساس شباهت خود در خوشه های مختلفی قرار گرفتند.

پارامترهای تنظیم شده برای الگوریتم ژنتیک ترکیبی با الگوریتم k-means در جدول ۴ ارائه شده است.

جدول (۴) پارامترهای الگوریتم ژنتیک ترکیبی با الگوریتم k-means

تعداد نسل	تعداد جمعیت	احتمال جهش	ثابت C	ثابت $C_m$
۱۰۰	۵۰	۰.۰۵	۲	۱

### ۱.۳. شاخص DB

برای بررسی و تایید روش انتخاب شده از شاخص دیویس - بولدین (DB) استفاده شد شاخص های اعتبارسنجی برای سنجش میزان صحت نتایج خوشه بندی به منظور مقایسه بین روشهای خوشه بندی مختلف یا مقایسه نتایج حاصل از یک روش با پارامترهای مختلف مورد استفاده قرار می گیرند که یکی از این شاخص ها ، شاخص دیویس بولدین می باشد که این معیار از شباهت بین دو خوشه بر اساس پراکندگی یک خوشه  $R_{ij}$  و  $S_i$  و عدم شباهت بین دو خوشه  $d_{ij}$  تعریف می شود استفاده می کند. به بیان دیگر این شاخص یک تابع از نسبت مجموع "پراکندگی داخل خوشه" به " دوری بین خوشه ها" می باشد.



برای تعیین بهترین دسته بندی، تعداد خوشه های مختلف مورد بررسی قرار گرفت. در جدول ۵ تعداد داده ها در هر خوشه و مقادیر و DB برای هر خوشه مشخص شده است.

مطابق با جدول ۵ بهترین حالت برای داده ها خوشه بندی با  $K=8$  انتخاب شد. زیرا دارای کمترین مقدار DB است. با در نظر گرفتن  $K=8$  به عنوان بهترین تعداد خوشه، داده های متعلق به هر گروه مشخص شد.

جدول (۵) تعداد داده ها در هر خوشه و مقادیر DB برای هر خوشه

تعداد خوشه (K)	تعداد داده ها در هر خوشه	DB
۳	۱۴۰ ۱۷۵ ۸۶	۲,۲۸۶
۴	۱۰۵ ۱۲۸ ۸۳ ۸۵	۲,۲۵۹
۵	۵۹ ۵۸ ۱۰۳ ۱۰۴ ۷۷	۲,۱۶۴
۶	۸۳ ۸۰ ۵۸ ۶۶ ۵۹ ۵۵	۲,۱۳۷
۷	۵۸ ۳۷ ۵۵ ۷۳ ۷۵ ۵۹ ۴۴	۲,۰۹۷
۸	۷۱ ۴۹ ۳۲ ۵۲ ۵۸ ۵۷ ۳۹ ۴۳	۲,۰۶۶

# کنفرانس داده کاوی ایران

جدول (۶) مراکز گروه ها برای بهترین خوشه (K=۸)

شماره گروه								شماره متغیر
۸	۷	۶	۵	۴	۳	۲	۱	
۳۰.۵۹۰	۳۲,۸۲۸	۳۱,۳۷۷	۳۱,۷۱۸	۳۲,۱۰۲	۳۱,۱۲۵	۳۳,۵۴۷	۳۲,۵۷۷	۱
۷۹,۹۰۹	۸۶,۸۸۵	۸۱,۰۹۴	۸۲,۲۱۸	۸۳,۸۹۷	۸۱,۱۲۵	۹۱,۳۰۱	۸۶,۶۷۶	۲
۸۱,۹۳۱	۹۱,۸۸۵	۸۳,۸۱۱	۸۵,۳۹	۸۷,۳۴۶	۸۳,۸۱۲	۹۶,۷۱۶	۹۱,۲۱۱	۳
۶۵,۹۳۱	۷۴,۵۷۱	۶۷,۷۳۵	۶۹,۶۷۱	۷۰,۹۵۹	۶۷,۵	۷۹,۱۱۳	۷۴,۰۷	۴
۸۹,۷۲۷	۹۹,۵۷۱	۹۱,۷۵۴	۹۳,۰۱۵	۹۳,۴۰۸	۹۱,۵۳۱	۱۰۱,۲۸۸	۹۶,۶۹	۵
۴۹,۴۵۴	۵۸,۲۸۵	۵۱,۳۹۶	۵۲,۸۷۵	۵۳,۰۶۱	۵۱,۰۹۳	۵۸,۷۵۴	۵۵,۴۵	۶
۳۴,۲۹۵	۳۸,۵۷۱	۳۵,۸۳	۳۶,۲۱۸	۳۷,۰۲	۳۵,۵	۴۰,۳۷۷	۳۸,۵۲۱	۷
۳۲,۵۶۸	۳۷,۲۵۷	۳۳,۹۲۴	۳۴,۵۳۱	۳۵,۳۴۶	۳۲,۸۱۲	۳۸,۰۱۸	۳۶,۳۹۴	۸
۲۱,۲۲۷	۲۲,۸۸۵	۲۱,۹۶۲	۲۱,۸۹	۲۲,۴۲۸	۲۱,۴۶۸	۲۳,۳۲	۲۲,۷۰۴	۹
۱۵,۴۵۴	۱۶,۶	۱۵,۸۶۷	۱۶,۲۰۳	۱۶,۵۳	۱۵,۴۳۷	۱۷,۳۷۷	۱۶,۶۴۷	۱۰
۹,۸۱۸	۱۰,۵۱۴	۹,۹۰۵	۹,۳۷۵	۹,۷۱۴	۱۰,۵۳	۹,۵۸۴	۹,۸۴۵	۱۱
۱۹,۵۴۵	۲۱,۱۷۱	۲۰,۸۶۷	۱۹,۹۳۷	۲۰,۳۰۶	۱۹,۹۳۷	۲۱,۱۶۹	۲۰,۶۳۳	۱۲
۳۱,۷۲۷	۳۴,۶۲۸	۳۲,۸۸۶	۳۲,۲۵	۳۳,۰۶۱	۳۲,۶۲۵	۳۴,۸۴۹	۳۳,۳۸	۱۳
۵۴,۷۲۷	۵۸,۶۵۷	۵۶,۴۵۲	۵۵,۰۷۸	۵۶,۰۸۱	۵۳,۷۸۱	۵۷,۹۶۲	۵۸,۱۵۴	۱۴
۶۵,۸۸۶	۶۹,۸	۶۷,۱۵	۶۷,۴۸۴	۶۸,۳۴۶	۶۴,۸۱۲	۶۹,۸۱۱	۷۰,۲۸۱	۱۵
۹۱,۲۷۲	۹۶,۳۴۲	۹۲,۷۵۴	۹۳,۳۴۳	۹۴,۱۴۲	۹۲,۹۶۸	۹۵	۹۵,۳۸	۱۶
۱۰۲,۳۴	۱۰۷,۹۴۲	۱۰۴,۷۵۴	۱۰۵,۵۶۲	۱۰۶,۲۴۴	۱۰۵,۲۱۸	۱۰۷,۶۹۸	۱۰۷,۴۰۸	۱۷
۱۲۶,۹۷۷	۱۳۲,۷۴۲	۱۲۹,۱۶۹	۱۲۹,۷۸	۱۳۱,۰۶۱	۱۲۸,۶۲۵	۱۳۱,۶۰۳	۱۳۲,۲۳۹	۱۸
۲۰,۰۶۸	۲۱,۱۴۲	۲۰,۳۳۹	۲۰,۳۷۵	۲۰,۳۸۸	۲۰,۸۴۳	۲۰,۰۳۷	۲۰,۲۸۱	۱۹
۱,۰۶۸	۱,۰۲۸	۱,۰۱۸	۱,۱۷۱	۱,۱۲۲	۱,۰۹۳	۱,۲۴۵	۱,۱۲۶	۲۰
۳,۲۰۴	۲,۰۲۸	۳,۳۰۱	۲,۸۱۲	۳,۱۴۲	۳,۶۵۶	۲,۶۷۹	۲,۴۶۴	۲۱
۱,۲۰۴	۱,۱۱۴	۱,۰۱۸	۱,۲۸۱	۱,۲۲۴	۱,۰۹۳	۱,۱۸۸	۱,۱۸۳	۲۲
۶,۳۴	۳,۶۵۷	۱,۱۱۳	۶,۴۰۶	۱,۲۰۴	۱,۲۱۸	۱,۳۹۶	۶,۴۹۲	۲۳
۲,۰۶۸	۲,۰۲۸	۲,۱۸۸	۱,۸۹	۲,۲۰۴	۲	۲,۳۲	۲,۱۸۳	۲۴
۱,۸۱۸	۲,۴	۲,۱۸۸	۲,۴۸۴	۱,۹۳۸	۲,۵۹۳	۲,۳۰۱	۲,۲۶۷	۲۵
۲,۸۸۶	۱,۹۴۲	۲,۹۰۵	۱,۰۴۶	۲,۷۷۵	۱,۰۶۲	۲,۷۷۳	۲,۸۱۶	۲۶
۳,۶۵۹	۱,۰۵۷	۳,۵۲۸	۲,۵۷۸	۳,۳۲۶	۳,۷۵	۳,۶۲۲	۳,۸۳	۲۷
۲	۲,۵۷۱	۲,۰۷۵	۲,۳۹	۱,۸۹۷	۲,۳۴۳	۲,۰۹۴	۲,۱۲۶	۲۸
۱,۵۹	۱,۵۷۱	۱,۸۶۷	۱,۳۷۵	۲,۳۰۶	۱,۳۴۳	۲,۱۱۳	۲,۰۱۴	۲۹
۳,۶۸۱	۳,۴۸۵	۳,۸۸۶	۳,۵۹۳	۱,۰۴	۳,۱۵۶	۳,۹۰۵	۳,۶۰۵	۳۰
۴,۰۶۸	۶,۷۴۲	۳,۶۴۱	۳,۶۰۹	۴,۵۳	۴,۱۲۵	۶,۱۸۸	۵,۰۹۸	۳۱
۴,۳۴	۴,۶۸۵	۳,۳۳۹	۳,۵۱۵	۴,۰۸۱	۳,۶۸۷	۴,۴۳۳	۳,۵۶۳	۳۲
۵,۸۴	۵,۱۱۴	۴,۳۳۹	۴,۵۷۸	۵,۴۸۹	۴,۲۵	۴,۵۸۴	۴,۳۸	۳۳

با توجه به نتایج خروجی از الگوریتم ترکیبی مورد استفاده طبق جدول ۶، مبنی بر این که اگر اختلاف بین داده ها زیاد باشد آن پارامتر مؤثر در تصور افراد از بدن می باشد. مشخص گردید پارامترهایی چون افراد مؤثر در تصمیم گیری، معیارهای اجتماعی و مکان جغرافیایی در پارامتر خروجی در تصور افراد از بدن خود مؤثرند همچنین اکثریت افراد در مورد اندام خود با تفاوت در سایزهای واقعی افراد و سایز ایده آل مانند دور سینه، دور ساق و فاصله سرشانه تا زانو اشتباه تصور می کنند بطوری که بیشتر اندام ها مثلث است ولی اکثر افراد سعی می کنند خودشان را به اندام ساعت شنی نزدیک کنند.

#### ۴. نتیجه گیری

در این تحقیق با توجه به ویژگی های الگوریتم ژنتیک، از یک الگوریتم ژنتیک ترکیبی با الگوریتم k-means جهت دسته بندی داده ها استفاده شد. جهت بررسی روش پیشنهادی پس از انتخاب مجموعه داده ها با استفاده از برنامه متلب، الگوریتم مورد نظر نوشته شد. داده های مورد پردازش مربوط به مطالعه میدانی ۴۰۱ پرسش نامه مؤثر در بین دانشجویان خوابگاه دختران دانشگاه یزد و گرفتن سایز ۱۸ اندازه از بدن افراد بود. دقت الگوریتم از پارامترهای مهم تحلیل روش پیشنهادی بود. نتایج نشان داد که خوشه بندی با استفاده از روش ترکیبی کی - مینز و الگوریتم ژنتیک به عنوان یک تکنیک داده کاوی، یک ابزار مؤثر برای راحتی روانشناختی تصور افراد از بدنشان است.

#### ۵. منابع :

[۱] Bagherzadeh, R., Latifi, M. and Faramarzi A.R., "Employing a Three-Stage Data Mining Procedure to Develop Sizing System" World Applied Sciences Journal, Vol. ۸, pp. ۹۲۳-۹۲۹, ۲۰۱۰.

[۲] Cash 'Cognitive-behavioral perspectives on body image,' in T.F. Cash and T. Pruzinsky (eds) Body image: A handbook of theory, research, and clinical practice, New York: Guilford. (۳۸-۴۶), (۲۰۰۲).

[۳] Cash, T., Ansis, J. and Strachan, M. 'Gender attitudes, feminist identity, and body images among college women,' Sex Roles, ۳۶: ۴۳۳-۴۷, (۱۹۹۷).

[۴] Ghazala Khan., Naila Khan., "Susceptibility to Informational Social Influence on Purchase Decisions of Designer Lable Apparel: The Mediating role of Gender" The Business Review, Cambridge; Vol ۴, No ۱; Summer ۲۰۰۵, P ۳۲-۳۶.

[۵] Schilder, P. the image and appearance of the human body, New York: International Universities Press (۱۹۵۰).

[۶] Grogan Sarah, Body Image Understanding body dissatisfaction in men, women, and children by Routledge Taylor & Francis e-Library, (۲۰۰۷)

[۷] Larose, Daniel T., "Discovering knowledge in data; an introduction to datamining". John Wiley & Sonc, Inc., ۲۰۰۵

[۸] SHAHRABI, J. and Salehi, M., "DEVELOPMENT OF A NEW SIZING SYSTEM BASED ON DATA MINING APPROACHES" ۷th International Conference – TEXSCI, ۲۰۱۰.

[۹] Thompson, J.K. and Van Den Berg, P. 'Measuring body image attitudes among adolescents and adults,' in T.F. Cash and T. Pruzinsky (eds) Body image. A handbook of theory, research, and clinical practice, New York: Guilford Press. ۱۴۲-۲۵۴, (۲۰۰۲).

[۱۰] Zare ,A., Dehghan , N., Payvandy ,P., Hadzade ,M ,. "K-mean clustering method for grouping the body form", ۸th National Conference on Textile Engineering, ۲۰۱۲.

[۱۱] Zarenejad , Z., mohammadi ,T., Naghdi ,S., Payvandy ,P., "Study on Body Image and Dissatisfaction of Body Parts (Case Study: Yazd University Students) ", ۸th National Conference on Textile Engineering, ۲۰۱۲.