

سری سوال: یک ۱

زمان آزمون (دقیقه): تستی: ۴۰ تشریحی: ۳۵

تعداد سوالات: تستی: ۲۵ تشریحی: ۷

عنوان درس: (داده کاوی (آموزش محور

رشته تحصیلی/کد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

استفاده از ماشین حساب ساده مجاز است

۱- داده ای که موارد زایدش حذف شده و به حداقل ممکن برای تصمیم گیری تقلیل یافته، چه نام دارد؟

۱. الگو ۲. اطلاعات ۳. دانش ۴. مدل

۲- رگرسیون نمونه ای از کدام روش داده کاوی است؟

۱. پیش بینی ۲. تقطیع ۳. مدل سازی وابستگی ۴. مصور سازی

۳- کدامیک از روشهای داده کاوی زیر توصیفی هستند؟

۱. خوشه بندی ۲. پیش بینی ۳. دسته بندی ۴. سریهای زمانی

۴- معیار یک روش خوشه بندی خوب کدام است؟

۱. تعداد مناسب خوشه ها ۲. شباهت کم بین نقاط خوشه های مختلف
۳. شباهت کم نقاط داخلی هر خوشه ۴. شباهت زیاد بین نقاط خوشه های مختلف

۵- اگر $d = 0/33$ (مریم و محمد) d

و $d = 0/67$ (علی و محمد) d

و $d = 0/75$ (علی و مریم) d باشد. آنگاه:

۱. شباهت مریم و محمد کمتر از شباهت علی و محمد است.
۲. شباهت علی و محمد کمتر از شباهت مریم و علی است.
۳. شباهت مریم و علی بیشتر از شباهت مریم و محمد است.
۴. شباهت مریم و محمد بیشتر از شباهت مریم و علی است.

۶- متغیرهای گسسته ای که با توجه به ارزش حالتهایشان مرتب شده اند، چه نام دارند؟

۱. متغیر اسمی ۲. متغیر رتبه ای ۳. متغیر عددی ۴. متغیر متقارن

۷- کدام گزینه صحیح است؟

۱. در الگوریتم K-means هر خوشه با یکی از اشیاء که در نزدیکی مرکز خوشه جای گرفته است، نمایش داده می شود.
۲. در الگوریتم K-medoids هر خوشه با میانگین اشیاء آن خوشه یا مرکز خوشه نمایش داده می شود.
۳. الگوریتمهای هیورستیک در پایگاه داده های کوچک تا متوسط به خوبی کار می کند.
۴. الگوریتمهای هیورستیک برای یافتن خوشه هایی با اشکال پیچیده مناسب هستند.

۸- الگوریتم AGNES (تجمیعی) از کدام روش برای خوشه بندی استفاده می کند؟

۱. افزازی ۲. سلسله مراتبی
۳. مبتنی بر چگالی ۴. مبتنی بر مشبک کردن فضا

سری سوال: ۱ یک

زمان آزمون (دقیقه): تستی: ۴۰ تشریحی: ۳۵

تعداد سوالات: تستی: ۲۵ تشریحی: ۷

عنوان درس: داده‌کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره‌وری ۱۳۱۴۰۴۰

۹- در روش پیوند تکی، فاصله بین خوشه‌ها چگونه محاسبه می‌شود؟

۱. بر حسب حداقل فاصله ممکنه بین عناصر آنها
۲. بر حسب حداکثر فاصله ممکنه بین عناصر آنها
۳. بر حسب میانگین فاصله ممکنه بین عناصر آنها
۴. بر حسب فاصله بین مراکز خوشه‌ها

۱۰- اگر p در شعاع فرضی e از q قرار گرفته باشد و در شعاع e از q حداقل نقاط مورد نظر ما نیز وجود داشته باشد، آنگاه نقطه p از q چه حالتی دارد؟

۱. مستقیماً قابل دسترس چگال است.
۲. قابل دسترس چگال است.
۳. متصل چگال است.
۴. پرت و مغشوش است.

۱۱- در قواعد تلازمی کدام گزینه صحیح است؟

۱. پشتیبان به معنی میزان وابستگی یک کالای خاص به دیگری است.
۲. اگر $x \Rightarrow y$ و $x \Rightarrow z$ آنگاه $x \Rightarrow z$
۳. اگر $x \Rightarrow A$ آنگاه $x + y \Rightarrow A$
۴. قواعد تلازمی ماهیتاً قواعد احتمالی هستند.

۱۲- در قواعد تلازمی اگر $(x) = \frac{5}{8}$ پشتیبان و $(x \cup y) = \frac{2}{8}$ پشتیبان باشد. شاخص درجه وابستگی بین دو مجموعه X و Y

(شاخص اطمینان) چقدر است؟

۱. ۰/۲۵
۲. ۰/۶۲
۳. ۰/۴۰
۴. ۰/۱۵

۱۳- کدام گزینه صحیح است؟

۱. درختهای تصمیم نمونه‌ای از دسته بندی‌های کاهل هستند.
۲. دسته بندی‌های کاهل در زمان دسته بندی بسیار سریع عمل می‌کنند.
۳. دسته بندی‌های مشتاق زمان زیادی را در مرحله آموزش، صرف ساخت مدل می‌کنند.
۴. روش نزدیکترین همسایگی نمونه‌ای از دسته بندی‌های مشتاق است.

۱۴- دسته بندی‌های مشتاق زمان زیادی را در مرحله آموزش، صرف ساخت مدل می‌کنند.

۱. تابع اقلیدسی
۲. تابع مانهاتان
۳. تابع ایها
۴. تابع ترکیب

۱۵- کدام گزینه در یک شبکه عصبی مصنوعی، معادل "سیناپسها" در شبکه عصبی زنده است؟

۱. ورودی
۲. خروجی
۳. تابع تحریک‌گره
۴. اتصالات

سری سوال: ۱ یک

زمان آزمون (دقیقه): تستی: ۴۰ تشریحی: ۳۵

تعداد سوالات: تستی: ۲۵ تشریحی: ۷

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴

۱۶- برای اندازه گیری اثربخشی یک درخت تصمیم پس از ایجاد، از چه معیاری استفاده می کنند؟

۱. درصد داده هایی که درست دسته بندی می شوند.
۲. کیفیت شاخه های ایجاد شده
۳. تعداد شاخه های ایجاد شده
۴. الف و ب هر دو

۱۷- کدام عامل بر تعداد داده های مورد نیاز برای استفاده از الگوریتمهای درخت تصمیم مؤثر نمی باشد؟

۱. تعداد ویژگیها
۲. تابع فعالسازی
۳. تعداد دسته ها
۴. پیچیدگی مدل

۱۸- کدام گزینه صحیح است؟

۱. درخت تصمیم نیاز به محاسبات پیچیده برای دسته بندی داده ها دارد.
۲. درخت تصمیم برای داده های طبقه ای قابل استفاده نمی باشد.
۳. روش درخت تصمیم نسبت به سایر روشهای دسته بندی دقت پائین تری دارد.
۴. درخت تصمیم برای هر مسیری از ریشه تا یک برگ، یک قاعده ایجاد می کند.

۱۹- در چه مواقعی از رگرسیون لجستیک استفاده می شود؟

۱. هنگامیکه تعداد متغیرهای مستقل در معادله بیش از یکی باشد.
۲. هنگامیکه داده ها دارای وابستگی خطی نباشند.
۳. هنگامیکه متغیر وابسته تنها دو مقدار ۰ و ۱ را بپذیرد.
۴. هنگامیکه از داده های پرت و مغشوش استفاده می شود.

۲۰- شاخصهای ارزیابی روشهای مختلف دسته بندی کدامند؟

۱. صحت، دقت و سرعت
۲. سرعت، پایداری و تفسیرپذیری
۳. صحت، دقت و پایداری
۴. دقت، پایداری و جمع و جور بودن

۲۱- مهم بودن عامل زمان در ساختار انبار داده ها بیانگر کدام ویژگی یک انبار داده است؟

۱. موضوع محوری
۲. جامعیت
۳. پویا پذیری
۴. پایائی

۲۲- کدام گزینه صحیح است؟

۱. پردازشی که بر روی انبار داده ها انجام میگیرد OLTP نام دارد.
۲. کاربران OLTP اپراتورها هستند در حالیکه کاربران OLAP کارشناسان خبره.
۳. نحوه دسترسی در OLAP به شکل خواندن و نوشتن است در صورتیکه در OLTP به شکل کاوش و کشف.
۴. OLTP در تصمیم گیری کاربرد دارد ولی OLAP در کارهای روزمره.

سری سوال: ۱: یک

زمان آزمون (دقیقه): تستی: ۴۰ تشریحی: ۳۵

تعداد سوالات: تستی: ۲۵ تشریحی: ۷

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/کد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

۲۳- اگر ابعاد داده ها بیش از سه بعد باشد، با چه مدلی می توان آن را نمایش داد؟

۱. ستاره ای ۲. برف دانه ۳. فلکی ۴. شبکه ستاره ای

۲۴- در معماری انبار داده ها، جزئی که توضیح می دهد چه چیزی و کجا وجود داشته باشد، چه نام دارد؟

۱. سیستمهای منبع ۲. مخزن مرکزی ۳. مخزن فراداده ۴. بازارچه داده

۲۵- مقدار داده ها و چگونگی توزیع آنها جزء کدام دسته از اطلاعات است؟

۱. اطلاعات ساختاری ۲. اطلاعات متریک ۳. اطلاعات تجاری ۴. اطلاعات انتقالی

سوالات تشریحی

- ۱- مراحل متعدد کشف دانش را نام ببرید. ۱،۰۰۰ نمره
- ۲- خوشه بندی را تعریف کنید و ۵ کاربرد آن را نام ببرید. ۱،۰۰۰ نمره
- ۳- چهار روش از روشهای مختلف دسته بندی را نام برده و یکی از آنها را توضیح دهید. ۱،۰۰۰ نمره
- ۴- لایه های یک شبکه عصبی ساده را با شکل نشان دهید. ۱،۰۰۰ نمره
- ۵- چهار ویژگی مهم یک انبار داده را نام ببرید. ۱،۰۰۰ نمره
- ۶- دو روش عمده جهت طراحی انبار داده ها از دیدگاه مهندسی نرم افزار نام برده و تعریف کنید. ۱،۰۰۰ نمره
- ۷- متغیر پیش بینی و متغیر برچسب دسته را در درخت تصمیم همراه با مثال تعریف کنید. ۱،۰۰۰ نمره



سوال ۱: یک

زمان آزمون (دقیقه): ۶۰ : نسی : ۶۰ : تشریحی : ۶۰

تعداد سوالات : نسی : ۲۰ : تشریحی : ۴

موضوع درس : داده کاوی (آموزش محور)

رشته تحصیلی / گد درس : مهندسی صنایع - مدیریت سیستم و بهره وری ۱۳۱۴۰۴

سوالات تشریحی

- ۱- شش خودرو با مشخصات زیر موجود است. با استفاده از روش k-means این شش خودرو را در دو گروه خوشه بندی کنید.
فرض کنید مراکز اولیه خوشه ها دو خودروی B و E بوده و خوشه بندی را تا دو مرحله انجام دهید. ضمناً برای محاسبه فواصل از فاصله اقلیدسی استفاده کنید.

نوع خودرو	قیمت (۱۰۰۰)	کیفیت (۱۰۰۰)
A	۹	۸
B	۵	۵
C	۴	۳
D	۶	۷
E	۸	۷
F	۶	۵

- ۲- جدول مقایسه دو شخص با شش ویژگی دودویی نامتقارن مطابق زیر است. میزان تمایز این دو شخص چقدر می باشد؟

ویژگی ها	شخص ۱	شخص ۲
ویژگی اول	۱	۱
ویژگی دوم	۰	۱
ویژگی سوم	۰	۰
ویژگی چهارم	۱	۰
ویژگی پنجم	۰	۱
ویژگی ششم	۱	۱

- ۳- مزایا و معایب استفاده از شبکه های عصبی در تکنیک های داده کاوی را نام برده و نمونه هایی از کاربرد این روش ها در داده کاوی را نام ببرید.

- ۴- کاربران نهایی ابزار داده ها چه کسانی هستند؟ نیازها و ویژگی های آنها را تشریح نمایید.

تعداد سوالات: تستی: ۲۰؛ تشریحی: ۴

زمان آزمون (دقیقه): تستی: ۶۰؛ تشریحی: ۶۰

عنوان درس: داده‌کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره‌وری ۱۳۱۴۰۴۰

دفتر آزمون و سنجش
مرکز آزمون و سنجش
دانشگاه پیام نور

۱۸- جدول داده‌های زیر مفروض می‌باشد:

سن	نوع ماشین	ریسک
۲۳	خانوادگی	زیاد
۱۷	اسپورت	زیاد
۴۳	اسپورت	زیاد
۶۸	خانوادگی	کم
۳۲	باری	کم
۲۰	خانوادگی	زیاد

به منظور دسته‌بندی با روش درخت تصمیم، مقدار شاخص $I(S)$ برای حالت $43 < \text{سن}$ چقدر می‌باشد؟

۱. ۰.۴۱۶ ۲. ۰.۳۲۲ ۳. ۰.۴۴۴ ۴. ۰.۲۶۶

۱۹- اهمیت عامل زمان و در نظر گرفتن افق زمانی در انبار داده‌ها، بیانگر کدام یک از ویژگی‌های انبار داده‌ها می‌باشد؟

۱. جامعیت ۲. پویابذیری ۳. پایایی ۴. جاری بودن

۲۰- کدام مدل (نوع) از انبار داده‌ها است که کلیه اطلاعات درباره موضوعات معین داخل سازمان را گردآوری می‌کند و نیازمند مدل‌سازی کسب و کار گسترده بوده و طراحی و ساخت آن ممکن است سال‌های زیادی به طول انجامد؟

۱. انبار مجازی ۲. بازارچه داده ۳. انبار جامع سازمانی ۴. انبار بنگاه اقتصادی



تقریباً ۱۰ سوال تشریحی

هرگز شهیرانات

معاونت آزمون

تعداد سوالات: تئوری: ۲۰، تشریحی: ۴

زمان آزمون (دقیقه): تئوری: ۶۰، تشریحی: ۶۰

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره‌وری ۱۳۱۴-۱۳۱۵

۱۳- کدام جمله درباره روش های دسته بندی بر مبنای نزدیک ترین همسایگی و درخت تصمیم صحیح می باشد؟

۱. روش نزدیکترین همسایگی و روش درخت تصمیم، هر دو نمونه ای از دسته بندی های مشتاق هستند.
۲. روش نزدیکترین همسایگی نمونه ای از دسته بندی های مشتاق و روش درخت تصمیم نمونه ای از دسته بندی های کاهل می باشد.
۳. روش نزدیکترین همسایگی و روش درخت تصمیم، هر دو نمونه ای از دسته بندی های کاهل هستند.
۴. روش نزدیکترین همسایگی نمونه ای از دسته بندی های کاهل و روش درخت تصمیم نمونه ای از دسته بندی های مشتاق می باشد.

۱۴- می خواهیم داده های زیر را با روش k -medoids خوشه بندی کنیم. اگر داده های ۱۱، ۳۵، ۴۶ و ۲ به عنوان medoid های اولیه انتخاب شوند کدام یک از دو داده در یک خوشه قرار خواهند گرفت؟

{۱، ۲، ۴، ۵، ۷، ۸، ۱۱، ۱۲، ۱۵، ۱۹، ۲۲، ۲۴، ۳۱، ۳۵، ۴۰، ۴۲، ۴۵، ۴۶، ۴۹، ۵۳، ۵۸، ۶۱، ۶۲}

۵، ۷، ۱ ۱۹، ۲۲، ۲ ۴۰، ۴۲، ۲ ۲۲، ۲۴، ۴

۱۵- در دسته بندی بر مبنای نزدیکترین همسایگی، تاثیر مقدار پارامتر k بر داده های مغشوش چیست؟

۱. مقدار k اثری بر نقش داده های مغشوش ندارد.
۲. با افزایش مقدار k ، اثر داده های مغشوش کم می شود.
۳. با کاهش مقدار k ، اثر داده های مغشوش کم می شود.
۴. مقدار k خیلی کم و خیلی زیاد باعث افزایش اثر داده های مغشوش می شود.

۱۶- کدام جمله درباره روش دسته بندی بر مبنای نزدیکترین همسایگی صحیح است؟

۱. روش نزدیکترین همسایگی، پیچیدگی محاسباتی زیادی دارد اما از سرعت بالایی برخوردار است.
۲. روش نزدیکترین همسایگی، پیچیدگی محاسباتی زیادی داشته و سرعت کمی دارد.
۳. روش نزدیکترین همسایگی، روشی ساده و موثر است و از سرعت بالایی برخوردار است.
۴. روش نزدیکترین همسایگی، روشی ساده و موثر است ولی سرعت کمی دارد.

۱۷- در دسته بندی به روش درخت تصمیم، منظور از گره چیست؟

۱. متغیر وابسته یا همان برجسب دسته می باشد.
۲. مقیاسی است که خروجی بر اساس آن تعیین می شود.
۳. متغیر مستقلی است که آزمون روی آن صورت می گیرد.
۴. داده ای است که دسته مربوط به آن تعیین شده است.



فرماندهان محترم
مرکز سنجش و آزمون
معاونت آموزشی

زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰

تعداد سوالات: تستی: ۲۰ تشریحی: ۴

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴

۸- کدام یک از روش های خوشه بندی، پیچیدگی محاسباتی را کاهش و سرعت را افزایش می دهد؟

۱. روش های مبتنی بر شبک کردن فضا
۲. روش های مبتنی بر چگالی
۳. نقشه های خود سازمان ده
۴. روش های سلسله مراتبی

۹- در یک تحلیل سبد بازار، اطلاعات زیر در دست می باشد:

۰.۳ (X) پشتیبان	۰.۶ (Y) پشتیبان	۰.۲ (XUY) پشتیبان
میزان وابستگی کالای X به کالای Y چقدر می باشد؟		
۱. ۰.۳۳	۲. ۰.۵	۳. ۰.۶۶
		۴. ۰.۷۵

۱۰- در قواعد تلازمی، به چه عناصری مکرر گفته می شود؟

۱. عناصری که وابستگی زیادی به یکدیگر دارند
۲. عناصری که در همه سبدهای خرید قرار دارند
۳. عناصری که همواره همراه با عناصر دیگر می باشند
۴. عناصری که حداقل پشتیبانی لازم را داشته باشند

۱۱- در فرآیند استخراج قواعد تلازمی، اولین مرحله کدام است؟

۱. محاسبه میزان وابستگی اقلام به یکدیگر
۲. تعیین همه زیرمجموعه های اقلام مکرر
۳. محاسبه پشتیبان هر یک از اجزا و تعیین اقلام مکرر
۴. استخراج قواعد قوی

۱۲- کدام یک از جملات زیر درباره دسته بندی و پیش بینی صحیح می باشد؟

۱. مدل های دسته بندی در تحلیل داده های گسسته و طبقه ای استفاده می شود اما مدل های پیش بینی بیشتر بر روی داده های پیوسته کار می کنند.
۲. مدل های دسته بندی و مدل های پیش بینی هر دو بر روی داده های گسسته و پیوسته کار می کنند.
۳. مدل های پیش بینی در تحلیل داده های گسسته و طبقه ای استفاده می شود اما مدل های دسته بندی بیشتر بر روی داده های پیوسته کار می کنند.
۴. مدل های دسته بندی در تحلیل داده های گسسته، طبقه ای و پیوسته استفاده می شود اما مدل های پیش بینی فقط بر روی داده های گسسته کار می کنند.



۱۳۴۵
۹۳-۲۱۷

تعداد سوالات: نستی: ۲۰؛ تشریحی: ۴
 عنوان درس: (داده کاوی (آموزش محور
 رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴
 زمان آزمون (دقیقه): نستی: ۶۰؛ تشریحی: ۶۰
 تعداد سوالات: نستی: ۲۰؛ تشریحی: ۴
 عنوان درس: (داده کاوی (آموزش محور
 رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴

استفاده از ماشین حساب ساده مجاز است

- ۱- کدام یک از روش های زیر جزء روش های داده کاوی توصیفی محسوب می شود؟
 ۱. دسته بندی ۲. قوانین تلازمی ۳. رگرسیون ۴. سری های زمانی
- ۲- کدام یک از جملات زیر درباره داده کاوی و کشف دانش (KDD) صحیح است؟
 ۱. استفاده از پایگاه داده با تعداد فیلدها و جداول کم اما تعداد رکوردهای خیلی زیاد معمول است.
 ۲. استفاده از پایگاه داده با تعداد فیلدها، جداول و رکوردهای کم معمول است.
 ۳. استفاده از پایگاه داده با تعداد فیلدها، جداول و رکوردهای خیلی زیاد معمول است.
 ۴. استفاده از پایگاه داده با تعداد فیلدها و جداول زیاد اما تعداد رکوردهای خیلی کم معمول است.
- ۳- تقسیم داده ها به گروه های مشابه بر اساس اصل حداکثر کردن شباهت داخل گروه ها و حداقل کردن شباهت بین گروه ها، چه نام دارد؟
 ۱. خلاصه سازی ۲. خوشه بندی ۳. دسته بندی ۴. یکپارچه سازی
- ۴- اگر دو شیء A و B از هشت ویژگی اسمی دارای دو ویژگی یکسان باشند آنگاه میزان تمایز (عدم شباهت) آنها چقدر می باشد؟
 ۱. ۰.۷۵ ۲. ۰.۲۵ ۳. ۰.۶۰ ۴. ۰.۵۰
- ۵- در یک مرحله از خوشه بندی به روش k-means، دو خوشه زیر به دست آمده است. در خوشه بندی مرحله بعد کدام دو عدد در یک خوشه قرار نمی گیرند؟
 $K1 = \{2, 3, 4\}$ ، $K2 = \{10, 11, 12, 20, 25, 30\}$
 ۱. ۳ و ۴ ۲. ۴ و ۱۰ ۳. ۱۰ و ۱۱ ۴. ۱۱ و ۱۲
- ۶- کدام یک از جملات زیر در مورد خوشه بندی صحیح است؟
 ۱. در روش های خوشه بندی سلسله مراتبی، تعداد خوشه ها از ابتدا مشخص است.
 ۲. در روش های خوشه بندی سلسله مراتبی، تخصیص انجام شده در یک مرحله قابل تغییر در مراحل بعد نیست.
 ۳. روش های خوشه بندی سلسله مراتبی معمولا سریعتر از غیرسلسله مراتبی عمل می کنند.
 ۴. روش k-means برای خوشه بندی داده هایی با ویژگی های طبقه ای بسیار کاراست.
- ۷- طبق معیار پیوند تکی، فاصله دو خوشه زیر چقدر است؟
 $K1 = \{9, 15, 17\}$ ، $K2 = \{40, 41, 53, 60, 65, 69\}$
 ۱. ۲۳ ۲. ۳۱ ۳. ۵۲ ۴. ۶۰



تعداد سوالات: تستی: ۲۰ تشریحی: ۴

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰

تعداد سوالات: تستی: ۲۰ تشریحی: ۴

سوالات تشریحی

۱- شش خودرو با مشخصات زیر موجود است. با استفاده از روش k -means این شش خودرو را در دو گروه خوشه بندی کنید.

فرض کنید مراکز اولیه خوشه ها دو خودروی B و E بوده و خوشه بندی را تا دو مرحله انجام دهید. ضمناً برای محاسبه فواصل از فاصله اقلیدسی استفاده کنید.

نوع خودرو	قیمت (۱۰-۰)	کیفیت (۱۰-۰)
A	۹	۸
B	۵	۵
C	۴	۳
D	۶	۷
E	۸	۷
F	۶	۵

۲- جدول مقایسه دو شخص با شش ویژگی دودویی نامتقارن مطابق زیر است. میزان تمایز این دو شخص چقدر می باشد؟

ویژگی ها	شخص ۱	شخص ۲
ویژگی اول	۱	۱
ویژگی دوم	۰	۱
ویژگی سوم	۰	۰
ویژگی چهارم	۱	۰
ویژگی پنجم	۰	۱
ویژگی ششم	۱	۱

۳- مزایا و معایب استفاده از شبکه های عصبی در تکنیک های داده کاوی را نام برده و نمونه هایی از کاربرد این روش ها در داده کاوی را نام ببرید.

۴- کاربران نهایی انبار داده ها چه کسانی هستند؟ نیازها و ویژگی های آنها را تشریح نمایید.



تعداد سوالات: تستی: ۲۰ تشریحی: ۴

زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

۱۸- جدول داده های زیر مفروض می باشد:

سن	نوع ماشین	ریسک
۲۳	خانوادگی	زیاد
۱۷	اسپورت	زیاد
۴۳	اسپورت	زیاد
۶۸	خانوادگی	کم
۳۲	باری	کم
۲۰	خانوادگی	زیاد

به منظور دسته بندی با روش درخت تصمیم ، مقدار شاخص $I(S)$ برای حالت $43 = < \text{سن چقدر می باشد؟}$

۱. ۰.۴۱۶ ۲. ۰.۲۲۲ ۳. ۰.۴۴۴ ۴. ۰.۲۶۶

۱۹- اهمیت عامل زمان و در نظر گرفتن افق زمانی در انبار داده ها، بیانگر کدام یک از ویژگی های انبار داده ها می باشد؟

۱. جامعیت ۲. پویاپذیری ۳. پایایی ۴. جاری بودن

۲۰- کدام مدل (نوع) از انبار داده است که کلیه اطلاعات درباره موضوعات معین داخل سازمان را گردآوری می کند و نیازمند مدل سازی کسب و کار گسترده بوده و طراحی و ساخت آن ممکن است سال های زیادی به طول انجامد؟

۱. انبار مجازی ۲. بازارچه داده ۳. انبار جامع سازمانی ۴. انبار بنگاه اقتصادی



تعداد سوالات: تستی: ۲۰ تشریحی: ۴

زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/کد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

۱۳- کدام جمله درباره روش های دسته بندی بر مبنای نزدیک ترین همسایگی و درخت تصمیم صحیح می باشد؟

۱. روش نزدیکترین همسایگی و روش درخت تصمیم، هر دو نمونه ای از دسته بندی های مشتاق هستند.
۲. روش نزدیکترین همسایگی نمونه ای از دسته بندی های مشتاق و روش درخت تصمیم نمونه ای از دسته بندی های کاهل می باشد.
۳. روش نزدیکترین همسایگی و روش درخت تصمیم، هر دو نمونه ای از دسته بندی های کاهل هستند.
۴. روش نزدیکترین همسایگی نمونه ای از دسته بندی های کاهل و روش درخت تصمیم نمونه ای از دسته بندی های مشتاق می باشد.

۱۴- می خواهیم داده های زیر را با روش k -medoids خوشه بندی کنیم. اگر داده های ۴۶، ۳۵، ۱۱ و ۲ به عنوان medoid های اولیه انتخاب شوند آنگاه کدام یک از دو داده در یک خوشه قرار خواهند گرفت؟

{۱، ۲، ۴، ۵، ۷، ۸، ۱۱، ۱۲، ۱۵، ۱۹، ۲۲، ۲۴، ۳۱، ۳۵، ۴۰، ۴۲، ۴۵، ۴۶، ۴۹، ۵۳، ۵۸، ۶۱، ۶۲}

۲۲، ۲۴، ۴

۴۰، ۴۲، ۳

۱۹، ۲۲، ۲

۵، ۷، ۱

۱۵- در دسته بندی بر مبنای نزدیکترین همسایگی، تاثیر مقدار پارامتر k بر داده های مغشوش چیست؟

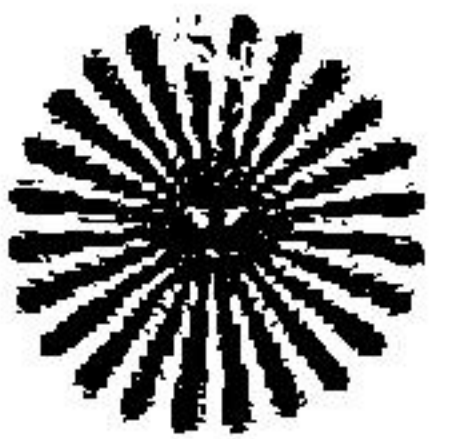
۱. مقدار k اثری بر نقش داده های مغشوش ندارد.
۲. با افزایش مقدار k ، اثر داده های مغشوش کم می شود.
۳. با کاهش مقدار k ، اثر داده های مغشوش کم می شود.
۴. مقادیر خیلی کم و خیلی زیاد k باعث افزایش اثر داده های مغشوش می شود.

۱۶- کدام جمله درباره روش دسته بندی بر مبنای نزدیکترین همسایگی صحیح است؟

۱. روش نزدیکترین همسایگی، پیچیدگی محاسباتی زیادی دارد اما از سرعت بالایی برخوردار است.
۲. روش نزدیکترین همسایگی، پیچیدگی محاسباتی زیادی داشته و سرعت کمی دارد.
۳. روش نزدیکترین همسایگی، روشی ساده و موثر است و از سرعت بالایی برخوردار است.
۴. روش نزدیکترین همسایگی، روشی ساده و موثر است ولی سرعت کمی دارد.

۱۷- در دسته بندی به روش درخت تصمیم، منظور از گره چیست؟

۱. متغیر وابسته یا همان برچسب دسته می باشد.
۲. مقیاسی است که خروجی بر اساس آن تعیین می شود.
۳. متغیر مستقلی است که آزمون روی آن صورت می گیرد.
۴. داده ای است که دسته مربوط به آن تعیین شده است.



تعداد سوالات: تستی: ۲۰ تشریحی: ۴

عنوان درس: داده کاوی (آموزش محور)

رشته تحصیلی/گد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰

سری سوال: ۱ یک

زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰

۸- کدام یک از روش های خوشه بندی، پیچیدگی محاسباتی را کاهش و سرعت را افزایش می دهد؟

۱. روش های مبتنی بر مشبک کردن فضا
۲. روش های مبتنی بر چگالی
۳. نقشه های خود سازمان ده
۴. روش های سلسله مراتبی

۹- در یک تحلیل سبد بازار، اطلاعات زیر در دست می باشد:

- پشتیبان $(X) = 0.3$ ، پشتیبان $(Y) = 0.6$ ، پشتیبان $(X \cup Y) = 0.2$
- میزان وابستگی کالای X به کالای Y چقدر می باشد؟
۱. ۰.۳۳
 ۲. ۰.۵
 ۳. ۰.۶۶
 ۴. ۰.۷۵

۱۰- در قواعد تلازمی، به چه عناصری مکرر گفته می شود؟

۱. عناصری که وابستگی زیادی به یکدیگر دارند
۲. عناصری که در همه سبدهای خرید قرار دارند
۳. عناصری که همواره همراه با عناصر دیگر می باشند
۴. عناصری که حداقل پشتیبانی لازم را داشته باشند

۱۱- در فرآیند استخراج قواعد تلازمی، اولین مرحله کدام است؟

۱. محاسبه میزان وابستگی اقلام به یکدیگر
۲. محاسبه پشتیبان هر یک از اجزا و تعیین اقلام مکرر
۳. تعیین همه زیرمجموعه های اقلام مکرر
۴. استخراج قواعد قوی

۱۲- کدام یک از جملات زیر درباره دسته بندی و پیش بینی صحیح می باشد؟

۱. مدل های دسته بندی در تحلیل داده های گسسته و طبقه ای استفاده می شود اما مدل های پیش بینی بیشتر بر روی داده های پیوسته کار می کنند.
۲. مدل های دسته بندی و مدل های پیش بینی هر دو بر روی داده های گسسته و پیوسته کار می کنند.
۳. مدل های پیش بینی در تحلیل داده های گسسته و طبقه ای استفاده می شود اما مدل های دسته بندی بیشتر بر روی داده های پیوسته کار می کنند.
۴. مدل های دسته بندی در تحلیل داده های گسسته، طبقه ای و پیوسته استفاده می شود اما مدل های پیش بینی فقط بر روی داده های گسسته کار می کنند.



تعداد سوالات: تستی: ۲۰ تشریحی: ۴
عنوان درس: (داده کاوی) آموزش محور
رشته تحصیلی/کد درس: مهندسی صنایع-مدیریت سیستم و بهره وری ۱۳۱۴۰۴۰
زمان آزمون (دقیقه): تستی: ۶۰ تشریحی: ۶۰
سری سوال: یک ۱

استفاده از ماشین حساب ساده مجاز است

۱- کدام یک از روش های زیر جزء روش های داده کاوی توصیفی محسوب می شود؟

۱. دسته بندی ۲. قوانین تلازمی ۳. رگرسیون ۴. سری های زمانی

۲- کدام یک از جملات زیر درباره داده کاوی و کشف دانش (KDD) صحیح است؟

۱. استفاده از پایگاه داده با تعداد فیلدها و جداول کم اما تعداد رکوردهای خیلی زیاد معمول است.
۲. استفاده از پایگاه داده با تعداد فیلدها، جداول و رکوردهای کم معمول است.
۳. استفاده از پایگاه داده با تعداد فیلدها، جداول و رکوردهای خیلی زیاد معمول است.
۴. استفاده از پایگاه داده با تعداد فیلدها و جداول زیاد اما تعداد رکوردهای خیلی کم معمول است.

۳- تقسیم داده ها به گروه های مشابه بر اساس اصل حداکثر کردن شباهت داخل گروه ها و حداقل کردن شباهت بین گروه ها، چه نام دارد؟

۱. خلاصه سازی ۲. خوشه بندی ۳. دسته بندی ۴. یکپارچه سازی

۴- اگر دو شیء A و B از هشت ویژگی اسمی دارای دو ویژگی یکسان باشند آنگاه میزان تمایز (عدم تشابه) آنها چقدر می باشد؟

۱. ۰.۷۵ ۲. ۰.۲۵ ۳. ۰.۶۰ ۴. ۰.۵۰

۵- در یک مرحله از خوشه بندی به روش k-means، دو خوشه زیر به دست آمده است. در خوشه بندی مرحله بعد کدام دو عدد در یک خوشه قرار نمی گیرند؟

$$K1 = \{2, 3, 4\} \quad , \quad K2 = \{10, 11, 12, 20, 25, 30\}$$

۱. ۳ و ۴ ۲. ۴ و ۱۰ ۳. ۱۰ و ۱۱ ۴. ۱۱ و ۱۲

۶- کدام یک از جملات زیر در مورد خوشه بندی صحیح است؟

۱. در روش های خوشه بندی سلسله مراتبی، تعداد خوشه ها از ابتدا مشخص است.
۲. در روش های خوشه بندی سلسله مراتبی، تخصیص انجام شده در یک مرحله، قابل تغییر در مراحل بعد نیست.
۳. روش های خوشه بندی سلسله مراتبی معمولاً سریعتر از غیرسلسله مراتبی عمل می کنند.
۴. روش k-means برای خوشه بندی داده هایی با ویژگی های طبقه ای بسیار کاراست.

۷- طبق معیار پیوند تکی، فاصله دو خوشه زیر چقدر است؟

$$K1 = \{9, 15, 17\} \quad , \quad K2 = \{40, 41, 53, 60, 65, 69\}$$

۱. ۲۳ ۲. ۳۱ ۳. ۵۲ ۴. ۶۰