

الله يحيى

الله يحيى

الله يحيى

نام و نام حنا و حنا

واحد حرس

رسن

احمد

پیش‌نیاز ۴ Data PreProcessing

A: حیرا Data cleaning (علم از داده اهمیت) (از زیر چای سازی داده در علم > ۱۰۰ از اهمیت بالایی برخورده است. افزار اسی کیفیت تحلیل هاو نول ها: داده های نیز و فقط منجر بر تحلیل های قابل اعتماد تر و نول های دقیق تر شوند. داده های نادرست یا ناقص می‌توانند بعثت ایجاد الگوهای خاکریست و نیز به کثیری های استیاه اسوزن کاهش خطاها، داده های پاکسازی سهولت‌شامل خطاها کوئی هستزای خطاها را توانند خاصی از خود را داده نادرست است. اینها در جمیع آورده‌ها یا مسکلات انتقال داده‌ها که افزار اسی اعتبار داده‌ها: داده های نیز به افزار اسی اعمال کاربران و خیان) بر داده‌ها کمک می‌کنند این امر در همکاری و تبادل داده با مایل سازمانها و افراد اهمیت در این می‌باشد. داده کی مرحله خود را هر فرآیند علم داده است که به اطمینان از صحیحیت در قات و قابلیت داده کی مرحله خود را همکاری و تبادل داده با مایل سازمانها و تجربه نتایج بهتر و دفعی کثیری های موثر ترمیمه شود.

B: Missing Values (جگونه مدیریت مفقودی می‌بینیم به مقادیر کمترین نزد داده‌ها که از حالات های اصلی بر تحلیل داده است. روشن‌قای مختلف برای برخورد با این مقادیر وجود دارد که انتخاب بهترین وسیله نسبت به نوع داده های گیمسیز و هدف تحمل شی دارد. ۱. حذف مشاهدات: خذف کامل سطحی؛ اگر سطحی شامل مقادیر کمترین باشد می‌توان حذف می‌شود این روشن‌قای سازه است. اما اگر ترانزیت منجر به از دست رفتن الایات زیادی شود ۲- حذف کامل ستون. اگر ستونی شامل مقادیر کمترین زیادی باشد می‌توان حذف این روشن‌قای ترانزیت منجر به از دست رفتن الایات زیادی شود و فقط در صورتی که متوجه می‌شوند برای تحلیل ضروری برای نیاز دارند از این روشن‌قای ترانزیت منجر به از دست رفتن الایات زیادی شود این روش استفاده از المتریک لعای پیش‌فرسته؛ در بخش موادر از المتریک دو می‌بینیم می‌بینیم ترانزیت المتریک های یاد کثیری مایسین برای پیش‌بینی مقداری کمترین / استفاده می‌شود

1104/1114

Data PreProcessing مقدمة

نکات ایجاد کننده خوبی و عیوبی در توانایی آن ها را سنجیده و مقابله می‌کند. این نکات ایجاد کننده خوبی و عیوبی در توانایی آن ها را سنجیده و مقابله می‌کند. این نکات ایجاد کننده خوبی و عیوبی در توانایی آن ها را سنجیده و مقابله می‌کند. این نکات ایجاد کننده خوبی و عیوبی در توانایی آن ها را سنجیده و مقابله می‌کند.

Data Preprocessing

Label Encoding), Encoding Techniques (one-hot : E

چه تفاوتی مارکر گهرد و روش برای تبدیل داده های مسنهای Encoding
طبقه ای به راههای امدادی هسته اما روش های انجام برای انجام این کار هسته های
املاک آنها بعبارتند از رمزگذاری بر حسب درسته کی عدد بفرد اختلاف
اما مسند برای مثال اگریک و پریک بنام زنگ داشته باشد که این روش معمولاً استفاده مسورد
که ورژن های درای ارتباها ترتیب باشند مثل وظایف های از جمله های از مهایی این روش
این است که میتوان ابتدا مدلی برای ترتیب احتمال معنای خاصی در این حیثی اگر این مدل را نمایند.
رمزگذاری که داشت در این روش به صورت مسند است که سهون جزوی از قوهای برای مثال اگر و پریک زنگ با
سیزه رمز $\langle \text{سیزه} \rangle$ و $\langle \text{آبی} \rangle$ داشته باشند ماتریس باشند که برای مسورد است
که را کمتر کنند. عزم رمز سیزه آبی این روش در صورتی که داده های موارد خاصی ندارند.

Model-building \rightarrow Feature Selection

در اینجا مسقیم کمالدرد مدل: هدف سهایی انتخاب ویرگی بهمراه کمالدرد کی مدل بیارگی
ماشین است با ساختن واژه‌ای کی مدل باز نیز مجموعه‌های مختلف ویرگی هایی تواند
کلیه رستقیم تا شیر ویرگی های انتخاب شده را بر کمالدرد مدل سنجیده این ارزیابی به مانند
کند تا بهترین ریز مجموعه ویرگی ها را برای دستیابی به بالاترین دقت کنند خواهد بیا
سایر محیا رهای کمالدرد مفرود قدر سناسایی کنیم بهمراه قابلیت تفسیر؛ بدل هایی
ساده‌تر با تعداد ویرگی های کمتر محبوب لازمات است، تر قابل تفسیر هستند این امر به مانند
کند تا در گنجینه از روایا بین ویرگی ها و متغیر هدف داشت باشیم و
در اینجا جدیدی از حداه ها استخراج کنیم.

Data PreProcessing

G: Duplicate Data

حذف داده های تکراری در روابط ادله های وظایفی هم رایج است و بسته به سمع مدیریت پایه ای داده کارهای استفاده و ساختار داده ها روشن نهادی مختلفی برای انجام آن وجود دارد ا- سیاست سازی رایم های تکراری با اینکه ابتدا با دید مسخن کنند که سیاست رکوردهای تکراری هستند این کار محدود با استفاده از دستور ROLLUP BY و تابع COUNT(*) در QL کار انجام می شود ۲- انتحاب کی رکوردهای تکراری را زین رکوردهای تکراری باشد تعمیم بپذیرید که سیاست رکوردهای تکراری خود را حذف کنند و ممکن است بخواهید قدیمی ترین جزوی داده را حذف کنید چنانچه با استفاده از این روش می توانیم جزوی داده را حذف کنیم.

H: Machine Learning

ایجاد مدل های مرتبط با داده های نامربوطه ممکن است متوجه رایج بینی های ماسنین شوند ایجاد مکث شوکا های دقت داده های نامربوطه ای تواند این مجموعه های داده بپذیرد این اسرائیلی بینی های ثانی است که می شود افراد ماسنین پیچیده مدل: داده های نامربوطه ای توانند باعث شوندن مدل های پیچیده تر شوند این پیچیدگی ای توانند نتیجه ایجاد ماسنین زمان آن موزیک و کالسین قابلیت هستش مدل شود که می تواند سرعت آن موزیک پردازش و تحلیل داده های نامربوطه زمان و منابع پیچیده را پیچیده کرده باشد که می تواند سرعت آن موزیک مدل را کاهش دهد.

Data Preprocessing بخش ۴

۱: Missing Values Data Imputation
 کے لئے برائی پر کردن
 کاربرد خارجی استفادہ از مدل سازی کاربرائی پر کردن مقادیر کم مسحہ در دادہا اہمیت
 ریاضی خارجی و جملی اصلی برائی این کا رویا رات ایسے از اس حفظ حجم و اطلاعات دادہا
 جیسا ری از الگوریتم های یادگیری ماسن و تحلیل دادہا یا از جا فرکہ دادہا کامل باشند
 حذف نیوٹھا دارای مقادیر کم مسحہ ممکن اسے منجر ہے از دست رفت) حجم زیادی
 از دادہا و کامیابی دقت امداد شود۔ جلوگیری از خطا ای تحلیل و تحلیل نادرست
 دادہا کی ناقص میں اسے باعث نتایج خا درست یا درست ہی نادرست در تحلیل
 کی آماری سنجو نوہا ای باعث افراسنی صحت و احتمال تحلیل ہاو مرل کا
 میں سوچو ۳۔ سیمیل مدل پردازش و تحلیل کی بعدی یہی ریاضی از الگوریتم ها مانند کریستین
 درخت تعمیم سبک کی جمعیت و نیاز جائز کار دادہا کامل باشند پر کردن مقادر کم مسحہ
 خراں پیو در سامنہ ترویج ترجیح کنزوں

۲: چیتو نرم تو اندر لا Normality را خردا دهای کددی ببری کیزیم
 برائی ببری نرمال بعدن دادہا کی ددی چڈیں روشن و جو در دار کری تو اندر ازان ہا
 استفادہ کیزی این روشن ھائی ای تر صون کا ای آماری روشن کی ٹرافیک و برسی بعری
 دادہا ہستند روشن کا ای گر افکر، ہستیوگرام توزیع دادہا را نشان سے دھد اگر دادہا
 نرمال باشند ہستیوگرام تقریباً سلسلہ نتیولہ خواهد بود و حل میانیں متفاہن
 اسے نمودار جعبہ ای؛ نمودار جعبہ ای میں تو اندر برائی سنسا سایر دادہا کی پریت و
 برائی متفاہن توزیع دادہا مور داستانی قرار گیرد در توزیع نرمال جمیں نمودار
 باید حقیر بیام و مسحہ قدر گیرد و حلول سیلہا باید حقیر بیا برابر باشند.
 محسیدی: کسیدی میزان نیزی یا بھن بعدن توزیع دادہا را نشان سے دھد توزیع
 نرمال داری کیتیہ کم ۳ نٹا ن دھنند توزیع نیزت و مقادر کسٹر از ۳ نٹا ن دھنند
 توزیع بھن تراست.