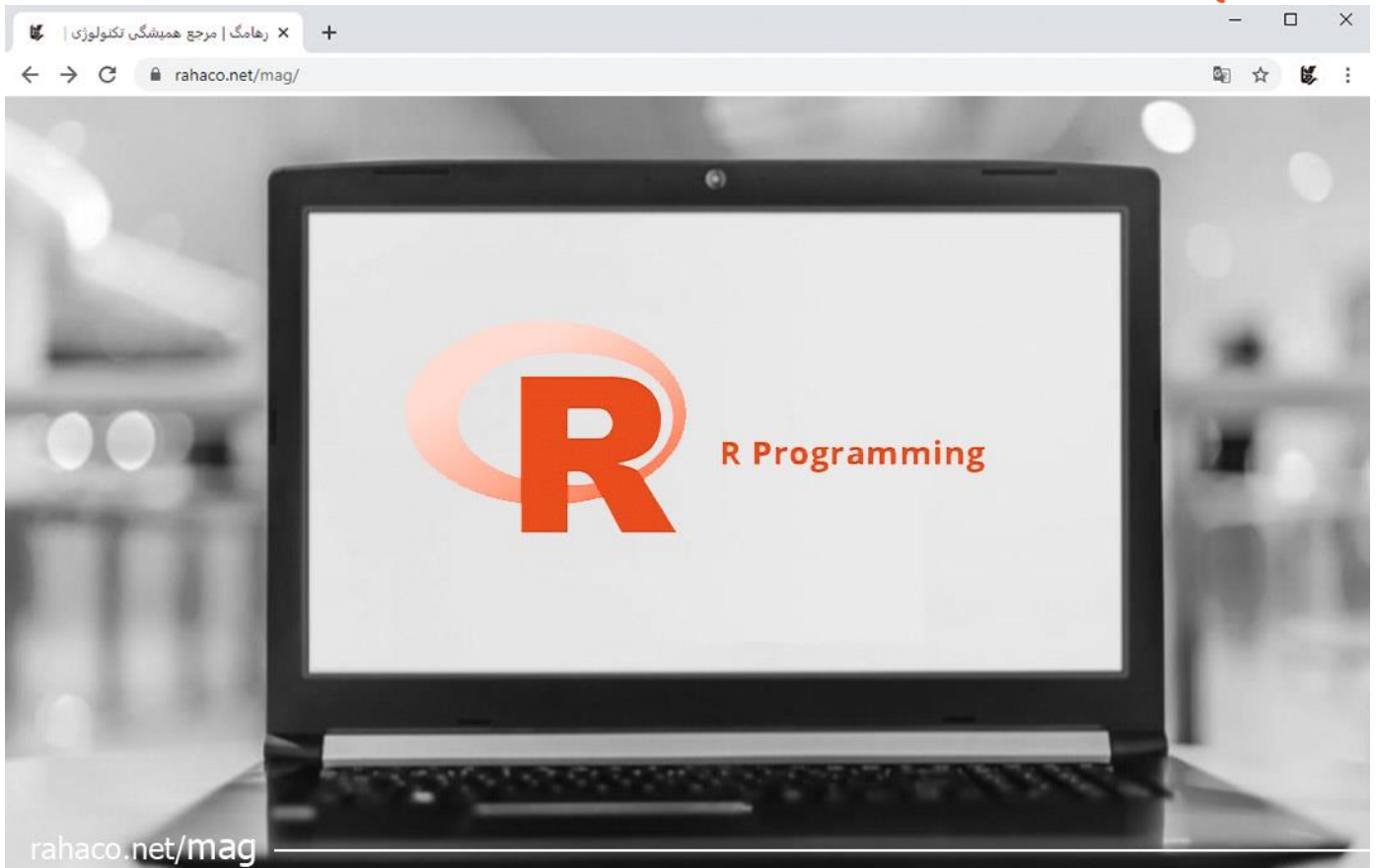




مجموعه شرکت های مهندسی دانش بنیان رها

زبان برنامه نویسی R برای تحلیلگران داده

شرکت رهاکو



فهرست

- 3 زبان برنامه نویسی R چیست؟
- 3 محیط زبان برنامه نویسی R
- 4 ویژگی های زبان برنامه نویسی R
- 4 چرا زبان برنامه نویسی R؟
- 4 کاربرد زبان برنامه نویسی R در دنیای واقعی
- 5 روش پردازش داده ها در زبان R
- 5 نتیجه گیری



زبان برنامه نویسی R برای محاسبات آماری و گرافیکی است که توسط تیم R Core و بنیاد R برای محاسبات آماری پشتیبانی می شود. R توسط راس ایهاکا و رابرت جنتلمن در دانشگاه اوکلند نیوزلند ایجاد شد. این زبان در میان داده کاویان، بیوانفورماتیکان و آماردانان برای تجزیه و تحلیل داده ها و توسعه نرم افزارهای آماری استفاده می شود. R به صورت رایگان تحت مجوز عمومی گنو در دسترس است و نسخه های باینری از پیش کامپایل شده برای سیستم عامل های مختلف مانند: لینوکس، ویندوز و مک ارائه شده است. زبان برنامه نویسی R از حرف اول نام دو توسعه دهنده آن، راس ایهاکا و رابرت جنتلمن گرفته شده است.

زبان برنامه نویسی R چیست؟

R یک زبان برنامه نویسی منبع باز برای تجزیه و تحلیل پیش بینی داده ها است. نسخه اولیه R در سال 1955 منتشر شد تا افرادی که در رشته آمار تخصص دارند بتوانند با مهارت های برنامه نویسی پیشرفته داده های پیچیده را تجزیه و تحلیل کنند. همچنین می توانند این نتایج را در نمودارهای بصری نمایش دهند.

R شامل توابعی است که از مدل سازی خطی، مدل سازی غیر خطی، آمار کلاسیک، طبقه بندی و خوشه بندی پشتیبانی می کند. این برنامه بر روی پلتفرم های یونیکس و سایر سیستم ها از جمله لینوکس، ویندوز و MacOS کامپایل و اجرا می شود. R programming language برای برنامه نویسان نرم افزار، آماردانان و داده کاویانی که مشتاقانه منتظر توسعه نرم افزارهای آماری با استفاده از برنامه نویسی R هستند، طراحی شده است.

محیط زبان برنامه نویسی R

محیط برنامه نویسی زبان R حول یک رابط خط فرمان استاندارد تعریف شده است. کاربران از این محیط برای خواندن داده ها و بارگذاری آن ها در فضای کاری، تعیین دستورات و دریافت نتایج استفاده می کنند. دستورات می توانند هر چیزی باشند از عملگرهای ساده ریاضی مانند: +، -، * و / گرفته تا توابع پیچیده تر که رگرسیون خطی (رگرسیون یعنی پیش بینی تغییرات یک متغیر بر اساس اطلاعات متغیر دیگر) و سایر محاسبات پیشرفته را انجام می دهند.

محیط زبان R به کاربران اجازه می دهد تا عملیات های مستقل مانند: بیرون کشیدن یک متغیر ثابت، اجرای رگرسیون را در مجموعه داده های حاصل در یک تابع واحد که می توان بارها و بارها استفاده کرد، ترکیب کنند. توابع حلقه در محیط برنامه نویسی R بسیار محبوب هستند. این توابع به کاربران امکان می دهند که اقداماتی را به طور مکرر انجام دهند، مانند: بیرون کشیدن نمونه ها از یک مجموعه داده بزرگ تر.



ویژگی های زبان برنامه نویسی R

- **آمار پایه:** رایج ترین اصطلاحات آماری پایه عبارتند: از میانگین، حالت و میانه. همه ی این ها به عنوان معیارهای گرایش مرکزی شناخته می شوند بنابراین با استفاده از R programming language می توانیم گرایش مرکزی را به راحتی اندازه گیری کنیم.
- **تجزیه و تحلیل داده ها:** مجموعه ای منسجم و یکپارچه از ابزارها را برای تجزیه و تحلیل داده ها فراهم می کند.
- **بسته های R:** یکی از ویژگی های اصلی R این است که کتابخانه ها در دسترس کاربران است R. دارای CRAN (شبکه آرشیو جامع (R می باشد: مخزنی که بیش از 10000 بسته را در خود جای داده است.
- **محاسبات توزیع شده:** محاسبات توزیع شده مدلی است که در آن اجزای یک سیستم نرم افزاری بین چندین کامپیوتر به اشتراک گذاشته می شود تا کارایی و عملکرد را بهبود بخشد.

چرا زبان برنامه نویسی R؟

زبان R به عنوان یک ابزار پیشرو برای یادگیری ماشین، آمار و تجزیه و تحلیل داده ها استفاده می شود. در محیط برنامه نویسی R اشیا، توابع و بسته ها به راحتی ایجاد می شوند R. یک زبان برنامه نویسی مستقل است؛ یعنی می توان آن را برای تمام سیستم عامل ها اعمال کرد R programming language. یک زبان رایگان منبع باز است که هر کسی می تواند آن را در هر سازمانی بدون خرید مجوز نصب کند. زبان برنامه نویسی R به ما اجازه می دهد تا با زبان های دیگر (C، ++C متصل شویم. بنابراین، شما به راحتی می توانید با بسیاری از منابع داده و بسته های آماری تعامل داشته باشید.

جامعه وسیعی از کاربران زبان برنامه نویسی دارد که روز به روز در حال افزایش است R. در حال حاضر یکی از پرتقاضاترین زبان های برنامه نویسی در بازار کار علم داده است که آن را به داغ ترین گرایش امروزی تبدیل کرده است. زبان برنامه نویسی R ترکیب درستی از سادگی و قدرت است و شرکت ها در سراسر جهان از آن استفاده می کنند.

کاربرد زبان برنامه نویسی R در دنیای واقعی

علم داده: هاروارد بیزینس ریویو، دانشمند داده را جذاب ترین شغل قرن بیست و یکم معرفی کرد. با ظهور دستگاه های اینترنت اشیا می توان از ترابایت داده ای ایجاد می کنند برای تصمیم گیری بهتر استفاده کرد، علم داده حوزه ای است که راهی جز پیشرفت و توسعه ندارد.



زبان های برنامه نویسی مانند زبان برنامه نویسی R به دانشمند داده قدرت فوق العاده ای می دهد R. به آن ها این قابلیت را می دهد که داده ها را در یک بازه زمانی مشخص جمع آوری کنند، تجزیه و تحلیل انجام دهند، تصویر سازی ایجاد کنند و نتایج عملی را منتقل نمایند. بیشتر دوره های علوم داده R programming language را در برنامه درسی خود گنجانده اند.

یادگیری ماشین R: در یادگیری ماشین کاربرد زیادی پیدا کرده است و بسته های مختلفی برای آن دارند، مانند: رگرسیون خطی و غیر خطی و بسیاری موارد دیگر. تقریباً تمام افراد از علاقه مندان به یادگیری ماشین گرفته تا محققان از R برای پیاده سازی الگوریتم های یادگیری ماشین در زمینه هایی مانند: امور مالی، ژنتیک، بازاریابی و مراقبت های بهداشتی استفاده می کنند.

محاسبات آماری R: محبوب ترین زبان برنامه نویسی در میان آمردانان است. این زبان به محققان، حتی آن هایی که دارای پیشینه علوم کامپیوتری نیستند اجازه می دهد تا به سرعت داده ها را از منابع مختلف وارد و تجزیه و تحلیل کنند R. همچنین دارای قابلیت ترسیم نمودار است. به این معنی که می توانید داده های خود را رسم کنید و تصاویر جالبی را از هر مجموعه داده ایجاد نمایید.

روش پردازش داده ها در زبان R

ساختارهای R programming language شامل بردارها، آرایه ها و فهرست ها می شوند. بردارها مجموعه ای از مقادیر هستند و می توانند در مجموعه هایی چند بعدی به ترتیب ستون اصلی قرار بگیرند. یعنی با توجه به مجموعه ای از ابعاد مرتب شده، ابتدا مقادیر را در بعد اول پر می کنند، سپس بعد دوم پر می شود و همین ترتیب ادامه دارد R. programming language مانند زبان هایی مثل APL و MATLAB است.

R و کتابخانه های آن تکنیک های آماری مختلفی از جمله مدل سازی خطی، آزمون های آماری کلاسیک، تحلیل مداوم، طبقه بندی و غیره را پیاده سازی می کنند. برای کارهای محاسباتی فشرده، کدهای C، ++C و Fortran را می توان در زمان اجرا فراخوانی کرد. یکی دیگر از نقاط قوت R گرافیک استاتیک است که می تواند نمودارهایی با کیفیت تولید کند.

نتیجه گیری

زبان برنامه نویسی R یک زبان متن باز است که به طور گسترده به عنوان یک نرم افزار آماری و ابزار تجزیه و تحلیل داده استفاده می شود R. به طور کلی با رابط خط فرمان ارائه شده و در پلتفرم های پرکاربرد مانند: ویندوز، لینوکس و مک در دسترس است. براساس نظر سنجی های کاربران و مطالعات پایگاه های اطلاعاتی، R یکی از رایج ترین زبان های برنامه نویسی برای استفاده در رشته داده کاوی است.



مجموعه شرکت های مهندسی دانش بنیان رها