



Namatek
True Education

R
R

**Programming
Language**

www.namatek.com

زبان برنامه نویسی R

فهرست مطالب

۱. ۲. ۳. ۴.

زبان برنامه نویسی R یکی از زبان های برنامه نویسی رایانه ای برای تجزیه و تحلیل اطلاعات است. زبان های برنامه نویسی مهم ترین گام به سوی مدیریت و توسعه تکنولوژی هستند. R نیز در همین راستا به عنوان یک ابزار پیشرو در علوم داده ارائه شده است. اما این زبان چگونه به محبوب ترین زبان برنامه نویسی در دنیای علوم داده تبدیل شده است؟

اگر شما نیز به این زبان علاقمند هستید و دوست دارید در مورد آن اطلاعات بیشتری کسب کنید تا انتهای مقاله با ما همراه باشید.

#۱ معرفی زبان برنامه نویسی R



R یک زبان برنامه نویسی منبع باز (Open Source) و رایگان است. این زبان توسط مبتکران زبان برنامه نویسی روبرت جنتلمن (Robert Gentleman) و روس ایهاکا (Ross Ihaka) در سال ۱۹۹۳ ایجاد شد.

زبان R در واقع به عنوان یک نرم افزار محاسبات آماری و ابزاری برای تجزیه و تحلیل شناخته می شود که می توانید از آن برای تجزیه و تحلیل و نمودار داده های خود استفاده کنید. این زبان دارای فهرست گسترده ای از روش های آماری و گرافیکی است که برخی از آن ها شامل موارد زیر هستند:

- الگوریتم های یادگیری ماشین (machine learning algorithms)
- تکنیک آماری رگرسیون خطی (linear regression)
- مدل های سری های زمانی (time series)
- تحلیل داده توسط استنباط آماری (statistical inference)

زبان برنامه نویسی R به طور گسترده توسط محققان در رشته های مختلف برای برآورد و نمایش نتایج، همچنین توسط معلمان آمار و روش های تحقیق استفاده می شود. این زبان در سیستم عامل های پرکاربردی نظیر Windows، Linux و macOS نیز مورد استفاده قرار گرفته است. بیشتر کتابخانه های زبان برنامه نویسی R با خود این زبان نوشته شده اند؛ اما برای انجام کارهای محاسباتی سنگین در آن، ترجیحا از کدهای C، C++ و Fortran استفاده می شود.

R نه تنها مورد اعتماد دانشگاهیان است؛ بلکه بسیاری از شرکت های بزرگ نیز از این زبان برنامه نویسی استفاده می کنند که برخی از آن ها عبارتند از:

- Uber •
- Google •
- Airbnb •
- Facebook •
- Twitter •
- Bing •
- Accenture •
- Wipro •

#۲ ویژگی های زبان برنامه نویسی R

ویژگی های زبان برنامه نویسی R به طور کلی به دو دسته ویژگی های آماری و ویژگی های برنامه نویسی در این زبان طبقه بندی می شوند که در ادامه به بررسی این ویژگی ها و زیرمجموعه های آن ها خواهیم پرداخت.



#۱-۲ ویژگی های آماری زبان R

۱. آمار پایه (Basic Statistics)

رایج ترین اصطلاحات آمار پایه عبارتند از:

- میانگین (Mean)
- حالت یا نما (Mode)
- میانه (Median)

این موارد در آمار پایه به عنوان سنجش های گرایش به مرکز یا Measures of Central Tendency شناخته می شوند که در توصیف و نمایش داده ها کاربرد دارند. یکی از ویژگی های زبان برنامه نویسی R این

است که با استفاده از آن می توان گرایش مرکزی را به راحتی اندازه گیری کرد.

۲. تکنیک گرافیک آماری (Statistical Graphical Techniques):

تکنیک گرافیک آماری، گرافیکی است که در زمینه آمار برای تجسم داده ها استفاده می شود. زبان R دارای امکانات زیادی برای ایجاد و توسعه گرافیک های آماری جالب است.

برخی از این نمودارهای پرکاربرد عبارتند از:

- نقشه های گرافیکی (graphic maps) هستند.
- نقشه موزاییک (Marimekko یا Mosaic plot) که یک روش گرافیکی برای تجسم داده ها از دو یا چند متغیر کیفی است.
- نمودار دوقطبی (Biplot) که یک نوع نمودار اکتشافی است که در آمار استفاده می شود.

نمودار دوقطبی اجازه می دهد اطلاعات هر دو نمونه و متغیرهای یک ماتریس داده به صورت گرافیکی نمایش داده شوند.

۳. توزیع احتمال (Probability Distributions):

توزیع احتمال نقش مهمی در آمار بازی می کند. با استفاده از زبان R ما می توانیم به راحتی انواع توزیع احتمال مانند توزیع دو جمله ای (Binomial Distribution)، توزیع عادی (Normal Distribution)،

توزیع کای دو (Chi-squared Distribution) و بسیاری دیگر را مدیریت کنیم.

۴. تجزیه و تحلیل داده ها (Data Analysis):

زبان R مجموعه وسیع، منسجم و یکپارچه ای از ابزارها برای تجزیه و تحلیل داده ها را فراهم می کند.



#۲-۲ ویژگی های برنامه نویسی زبان R

۱. پکیج های زبان برنامه نویسی آر (R Packages):

یکی از ویژگی های اصلی زبان R این است که کتابخانه های زیادی برای آن در دسترس است. زبان R دارای یک شبکه جامع بایگانی با نام CRAN است که این مخزن بیش از ۱۰۰۰۰۰ بسته را در خود جای داده است.

۲. محاسبه توزیع شده (Distributed Computing):

محاسبه توزیع شده مدلی است که در آن اجزای یک سیستم نرم افزاری برای بهبود کارایی و عملکرد بین چندین رایانه به اشتراک گذاشته می شود. دو پکیج جدید ddR و multidplyr که برای برنامه نویسی توزیع شده در R استفاده می شوند، در نوامبر ۲۰۱۵ منتشر شدند.

#۳ دلیل محبوبیت زبان R



از R برای علم داده استفاده می کنیم. این زبان انواعی از کتابخانه های مربوط به آمار و محیطی برای محاسبه و طراحی آماری دارد.

تحلیلگران کمی (Quantitative Analysts)، R را به عنوان ابزار برنامه نویسی استفاده می کنند. آن ها مدل های ریاضی پیچیده ای را برای غربالگری داده ها پیاده سازی می کنند تا شرکت ها از آن ها برای تصمیم گیری در مورد مدیریت ریسک استفاده کنند؛ بنابراین، زبان R به غربالگری داده ها کمک می کند. زبان برنامه نویسی R رایج است و تحلیلگران داده و برنامه نویسان تحقیقاتی (Research Programmer) از آن استفاده می کنند. آن ها از این ابزارها برای تجزیه و تحلیل، جمع آوری داده ها، تبدیل داده ها جهت یافتن بهترین راه حل و ابزاری اساسی در داده های مالی استفاده می کنند.

زبان R ابزاری پیشرو برای یادگیری ماشین، آمار و تجزیه و تحلیل داده ها است. اشیا، توابع و پکیج ها را می توان به راحتی توسط آن ایجاد کرد. R زبانی مستقل از پلتفرم است؛ یعنی در تمام سیستم عامل ها اجرا می شود.

زبان R رایگان و منبع باز است؛ یعنی همه می توانند آن را در هر سازمانی بدون خرید مجوز نصب کنند. R یک بسته آماری است که به ما امکان ادغام با زبان های دیگر (C، C++) را هم می دهد؛ بنابراین می توانید با بسته های آماری و منابع داده ای بررسی شده توسط سایر زبان ها تعامل داشته باشید. زبان برنامه نویسی R کاربران زیادی دارد که در حال افزایش هستند.

R امروزه از متداول ترین زبان های برنامه نویسی در علم داده است که

داغ ترین روند برای جمع آوری، تجزیه و تحلیل انواع داده ها را دارد. در جامعه برنامه نویسان زبان R، همه از ارائه پکیج های جدید، رفع اشکال و بهبود کد استقبال می کنند.

#۴ معایب زبان برنامه نویسی R



در زبان R، برخی از پکیج ها کاملا استاندارد نیستند. اگرچه دستورات R فشار کمی به مدیریت حافظه می رسانند؛ اما ممکن است این زبان برنامه نویسی تمام حافظه موجود را مصرف کند. زبان برنامه نویسی R بسیار کندتر از سایر زبان های برنامه نویسی مانند Python و MATLAB است.