کاربرد برنامه نویسی پایتون در تجزیه‌های آماری

مرتبط با تکنولوژی بذر و ارزیابی ارقام‌گیاهی

عنایت رضوانی خورشیدی[1] | سعید سهیلی وند[2]

۱- عضو هیأت علمی مؤسسه تحقیقات ثبت و گواهی بذر و نهال
۲- عضو هیأت علمی پژوهشگاه بیوتکنولوژی کشاورزی ایران

آمار حوزه‌ای از ریاضیات کاربردی است که به طور کلی به روش جمع‌آوری داده‌ها، جدول‌بندی و تفسیر داده‌ها می‌پردازد. بخش کشاورزی برای توسعه برنامه‌ها و سیستم‌های تقویت شیوه‌های کشاورزی به انواع زبان‌ها و فناوری‌های برنامه‌نویسی متکی است. انتخاب زبان و تکنولوژی اغلب به نیازهای خاص عملیات کشاورزی بستگی دارد. با ادامه رشد جمعیت جهان، نقش برنامه‌ریزی در کشاورزی در تضمین امنیت غذایی و پایداری زیست‌محیطی حیاتی‌تر خواهد شد. پایتون (Python) یکی از زبان‌های برنامه‌نویسی است که در استفاده از آن مزایای متعددی وجود دارد و علی‌رغم اینکه یک نرم‌افزار داده باز است، اجازه کار حرفه‌ای و پیچیده در مدیریت داده‌ها و محاسبات آماری را می‌دهد. در علوم و تکنولوژی بذر نیز همانند بسیاری از علوم، نیاز به تجزیه و تحلیل و استنتاج آماری است. در این مقاله سعی شده است ضمن مقایسه پایتون با نرم‌افزارهای مهم آماری مبتنی بر برنامه‌نویسی مثل SAS و R، کاربرد و موارد برتری پایتون نسبت به بقیه نرم‌افزارها بررسی و با مطالعه آن، متخصصان کشاورزی تشویق شوند که از کتابخانه‌های پایتون، هم به عنوان توسعه دهنده نرم‌افزارهای آماری و هم به عنوان محیط برنامه‌نویسی برای رفع نیازهای تحقیقاتی و اجرایی استفاده کنند. در این مطالعه سعی شده است تا با استفاده از منابع بروز و بررسی کتابخانه‌های مهم پایتون، به کاربردهای اصلی پایتون در تجزیه و تحلیل و رسم گراف‌های آماری در علوم و تکنولوژی بذر و توانایی‌های آن با تحلیل خروجی‌های برنامه، پرداخته شود

کلیاتی در مورد آمار

در اصطلاح عامیانه، آمار توصیفی به طور کلی به معنای توصیف داده‌ها با کمک برخی روش‌ها مانند نمودارها و گراف‌ها در برنامه های شناخته شده‌ای مانند میکروسافت اکسل است. داده‌ها به گونه‌ای توصیف می‌شوند که بتوانند اطلاعات معناداری را بیان کنند که می‌توان از آنها برای یافتن اطلاعات خلاصه شده‌تر و قابل فهم‌تر استفاده کرد. بطوری که حتی با یک نگاه، بتوان نتیجه‌گیری کلی از کل تجزیه داده‌ها را به سرعت دریافت کرد. بطور مثال، روندهای توصیف و خلاصه کردن یک متغیر را تحلیل تک متغیره می‌نامند. توصیف یک رابطه آماری بین دو متغیر تحلیل دو متغیری گفته

می‌شود و توصیف رابطه آماری بین چند متغیر را تحلیل چند متغیره می‌نامند. بطور خلاصه، آمار توصیفی بر مبنای دو معیار مهم یعنی گرایش مرکزی (مانند میانگین، میانه، مد و ...) و اندازه‌گیری تغییرپذیری (مانند دامنه، واریانس و انحراف معیار) تحلیل می‌شود.

دو رویکرد اصلی برای نمایش اطلاعات وجود دارد. رویکرد بصری که برای نشان دادن داده‌ها با هیستوگرام، جداول، نمودارها استفاده می‌شود و رویکرد کمی، که توصیفات و خلاصه‌ای از داده‌ها را به صورت عددی ارائه می‌کند. در حال حاضر، نمودارهای آماری پرکاربردترین روش برای نمایش تحلیل داده‌ها هستند. نمایش بصری چنین داده‌هایی اجازه می‌دهد تا اطلاعات بیشتری به شیوه‌ای واضح ارائه شود. برنامه‌ها و ابزارهای زیادی برای تجزیه و تحلیل آماری استفاده می‌شود که از روال‌ها و زیربرنامه‌های ریاضی به بسته های نرم‌افزاری و ابزارهای حرفه‌ای سطح بالا مانند Microsoft SAS تبدیل شده‌اند

در آمار، جامعه آماری، مجموعه‌ای از تمام اقلام یا عناصری است که با یک یا چند صفت مشخص شناخته می‌شوند. جوامع معمولاً بزرگ یا بی‌نهایت هستند به همین خاطر، مطالعه کل جامعه با جمع‌آوری داده‌ها و تجزیه و تحلیل آنها، تقریبا امری غیرممکن است. بنابراین، نمونه‌گیری از جامعه به عنوان یک زیرمجموعه همگن از آن و بررسی پارامترهای نمونه برای برآورد آماره های یک جامعه، بهترین روش ممکن می باشد. با این اوصاف، زیرمجموعه جامعه، نمونه نامیده می‌شود که در هر حالتی، باید ویژگی‌های اصلی جامعه آماری را در خود داشته باشد. اصطلاح پرکاربرد دیگر واحد آزمایشی می باشد در واقع، فرم سازماندهی انجام آزمایش است که در جمع‌آوری داده‌ها استفاده می‌شود

فرآوری داده‌ها، پس از پردازش و سازماندهی آنها انجام می‌شود. این کار شامل جستجوی موارد تکراری، خطاها و ناهماهنگی‌ها در داده‌ها و متعاقباً حذف داده‌های نامناسب یا اصطلاحا پرت می‌باشد. در بیانی دقیق‌تر، فرآوری داده‌ها شامل فعالیت‌هایی مانند مرتب‌سازی داده‌ها، تطبیق سوابق، شناسایی نقاط پرت، شناسایی داده‌های نادرست، حفظ کیفیت داده‌ها و بررسی املای داده‌های متنی است. این کار در نهایت به جلوگیری از نتایج غیرمنتظره در ارائه داده‌هایی با کیفیت بالا کمک می‌کند که برای یک نتیجه‌گیری قوی و قابل قبول بسیار با اهمیت است

بنابراین فرآوری مجموعه داده‌ها و عاری از خطا کردن آنها، اولین قدم بوده و پس از این مرحله، تجزیه و تحلیل می‌تواند انجام شود. سپس روش‌های مختلفی از تکنیک‌هایی مانند تجزیه و تحلیل اولیه داده‌ها با درک پیام‌های موجود در داده‌های به‌دست‌آمده و نیز آمار توصیفی، یافتن مد، میانه، میانگین و سایر روش‌ها برای مصورسازی داده‌ها استفاده می شود. مصورسازی داده‌ها، یکی از تکنیک‌های مورد استفاده است که در آن داده‌ها به صورت گرافیکی نمایش داده می‌شوند تا برای به دست آوردن مشاهدات اضافی در مورد اطلاعات موجود در داده‌ها، مورد استفاده قرار گیرند. در نهایت در مواردی مانند استفاده از تجزیه علیت و همبستگی، مدل‌ها یا فرمول‌های ریاضی (به نام الگوریتم) به داده‌ها اضافه می‌شوند تا روابط درون متغیرها را کشف کنند. کیفیت داده‌ها با کیفیت ابزار (نرم‌افزارهای مورد استفاده) و تخصص تحلیلگر مطابقت دارد

کلیاتی در مورد زبان‌های برنامه نویسی

زبان های برنامه نویسی به دو دسته اصلی سطح بالا و سطح پایین تقسیم می‌شوند. زبان‌های سطح بالا برای خوانایی انسان و سهولت درک آن‌ها طراحی شده‌اند. این زبان‌ها معمولا می‌توانند روی هر پلتفرمی اجرا شوند و به طور گسترده مورد استفاده قرار می‌گیرند. اما زبان‌های سطح پایین به کد ماشین نزدیک‌تر بوده و برای کارهای خاص کارآمدتر هستند. با این‌حال، آنها وابسته به ماشین هستند و برای برنامه‌نویسی‌های دقیق‌تر و پیشرفته در سطح مدیریت میکروپروسسورهای خاص در صنعت استفاده می‌شوند. زبان‌های برنامه‌نویسی از تنوع بالایی برخوردارند. اینکه بهترین زبان برنامه‌نویسی برای یادگیری چیست، پاسخ قطعی برای این سوال وجود ندارد. اما آنچه که در مجوع قابل توصیه است اینست که زبان‌های برنامه‌نویسی ساختارهای مشترکی دارند، جریان‌های کاری، الگوها و موارد مشابهی دارند. هنگامی که تجربه بیشتری کسب شود، دیدن الگوهای مشترک در بین زبان‌ها ممکن خواهد بود و داشتن انگیزه، کاربر را به یادگیری و عمیق‌تر شدن در آن سوق می‌دهد.

نرم افزارهای آماری

نرم‌افزارهای متعدد آماری برای تجزیه و تحلیل داده‌ها استفاده

می‌شوند. با نگاهی به تاریخچه کاربرد نرم افزارهای آماری در علوم کشاورزی، مشخص می شود که قبل از طراحی سیستم عامل ویندوز، نرم‌افزارهای تحت DOS مثل MSTAT-C و Statgraph در علوم زارعی و باغی متداول بودند. در نسل بعدتر، نرم‌افزارهایی مثل MATLAB که محیط برنامه‌نویسی مدونی داشتند به خصوص در علوم مرتبط با تغذیه دام و طیور و محاسبات ریاضی مرتبط با فنی و مهندسی آب به عرصه آمدند. تولید نرم افزارهای آماری بعد از ورود سیستم عامل ویندوز، شتاب بیشتری گرفته و در نسخه‌های تحت ویندوز، این نرم‌افزارها با امکانات بیشتر و کاربرپسندتر (User-friendly) ارائه شدند. در همین دوره، نرم‌افزار قدرتمند SPSS معرفی شد که عمده نیازها در تجزیه آماری در علوم کشاورزی را بصورت کاربر پسندتر بر عهده گرفت. سپس نرم‌افزار قوی SAS به عنوان پرقابلیت‌ترین نرم‌افزار آماری، معرفی شد که یک محیط برنامه‌نویسی برای کدنویسی ارائه داد. البته لازم به ذکر است که این نرم افزار رایگان نبود. آخرین نرم‌افزاری عمومی آماری که به خصوص در بخش کشاورزی بسیار متداول شد، نرم افزار R بود که بر خلاف SAS بصورت داده باز و رایگان در دسترس عموم محققان بود. در حال حاضر کتابخانه‌های متعدد و پیشرفته‌ای از نرم‌افزار R وجود دارد که امکان تجزیه و تحلیل تخصصی در حوزه‌های مختلف کشاورزی را فراهم می‌کند.

زبان‌ها و فناوری‌های کلیدی برنامه نویسی در کشاورزی

کشاورزی مدرن و توسعه آن، نیاز غیرقابل تفکیکی به برنامه‌ها و برنامه نویسان دارد. انتخاب نوع زبان برنامه‌نویسی و تکنولوژی مرتبط با آن، اغلب به نیازهای خاص و عملیات مورد نیاز در کشاورزی مدرن، بستگی دارد. از بین زبان‌های برنامه‌نویسی که بطور روزمره در حوزه کشاورزی استفاده می شود می توان به موارد و حوزه های کاری زیر اشاره کرد

پایتون : پایتون یک زبان همه کاره است که اغلب در کشاورزی برای تجزیه و تحلیل داده‌ها، محاسبات علمی و توسعه نرم‌افزار و مدیریت مزرعه استفاده می‌شود. سادگی و طیف گسترده‌ای از کتابخانه‌ها، برای توسعه برنامه‌های کاربردی برای نظارت بر محصولات کشاورزی، دام و ماشین‌آلات و بیوتکنولوژی و بیوانفورماتیک، آن را یک کاندید ایده‌آل می‌سازد

R : زبان برنامه‌نویسی R برای تجزیه و تحلیل آماری و تجسم داده‌ها در کشاورزی استفاده می‌شود. این نرم‌افزار بویژه برای تجزیه و تحلیل عملکرد محصول، داده‌های آب و هوا و مدل‌سازی بیماری، ارزشمند است

جاوا (Java) و سی‌پلاس‌پلاس (C++): جاوا و سی‌پلاس‌پلاس بیشتر در توسعه سیستم‌های تعبیه شده برای ماشین آلات‌کشاورزی و حسگرها استفاده می‌شوند. این زبان‌ها عملکرد و کنترل لازم را برای برنامه‌های کاربردی همزمان ارائه می‌دهند

GIS یا سیستم‌های اطلاعات جغرافیایی: نرم‌افزارهای GIS (مانند ArcGIS و QGIS) برای نقشه‌برداری و تحلیل فضایی در کشاورزی حیاتی است. این سیستم به کشاورزان اجازه می‌دهد تا سلامت خاک را ارزیابی کرده و استفاده از زمین و آبیاری را بهینه و مدیریت کنند

برنامه‌نویسی در حوزه اینترنت اشیا: دستگاه‌ها و حسگرهای اینترنت اشیا در کشاورزی برای نظارت بر رطوبت خاک، شرایط آب و هوایی و سلامت دام استفاده می‌شوند. برنامه‌نویسی برای اتصال این دستگاه‌ها و جمع‌آوری داده‌ها و کنترل سیستم‌ها، ضروری است

برنامه‌نویسی در زمینه اتوماسیون و رباتیک: برنامه‌نویسی برای توسعه سیستم‌های کنترلی برای تراکتورهای خودران، پهپادها و دروگرهای رباتیک استفاده می‌شود. این سیستم‌ها عملیات دقیق و کارآمد را در کشاورزی امکان‌پذیر می‌کنند

نرم افزار پیش‌بینی مدل‌های آب و هوا: توسعه مدل‌های پیش‌بینی آب و هوا و شبیه‌سازی آن بر برنامه‌نویسی متکی است. این مدل‌ها اطلاعات ضروری را برای تصمیم‌گیری‌های مهم مانند زمان سم‌پاشی یا مدیریت رژیم آبیاری، در اختیار کشاورزان قرار می‌دهد

نرم‌افزارهای سنجش از راه دور و تصاویر ماهواره‌ای: برنامه‌نویسی برای پردازش و تجزیه و تحلیل تصاویر ماهواره‌ای و داده‌های سنجش از راه دور، استفاده می‌شود که درک دقیقتری از سلامت محصول، شیوع بیماری و شاخص‌های پوشش گیاهی ارائه می‌دهد

زبان برنامه‌نویسی پایتون

زبان برنامه‌نویسی پایتون اواخر دهه ۱۹۸۰ میلادی توسط «خودو فان روسوم» در موسسه ملی تحقیقات ریاضی و رایانه

هلنـد، ایجـاد شـد. از سـال ۱۹۹۱ زبـان برنامه‌نویـسی پایتون بـه عنـوان یـک زبـان برنامه‌نویـسی آسـان، جهـت نوشـتن برنامه‌هـای سـاده بـرای حـل مسـاله‌های بـزرگ، بـه کار گرفتـه شـد. در چنـد سـال اخیـر، زبان برنامه‌نویـسی پایتـون، در توسـعه نرم‌افزارهـای پیشـرفته، مدیریـت زیرسـاختها و نیـز تجزیـه و تحلیـل داده‌ها، مـورد اسـتفاده زیـاد قـرار گرفتـه اسـت. همچنیـن ایـن زبـان بـه دلیـل سـادگی و قـدرت آن، در محیط‌هـای دانشـگاهی و آکادمیـک، در حـال جایگزیـن شـدن بـا زبـان برنامه‌نویـسی Mathlab می‌باشـد. در بـازار کار نیـز، زبـان برنامه‌نویـسی پایتـون، بـه علـت اینکـه تولیـد و توسـعه نرم‌افـزار را سـرعت می‌بخشـد، مـورد اسـتقبال قـرار گرفتـه اسـت. بسـیاری از عملیـات و آنالیـز و تحلیل‌هـا کـه در سـایر زبان‌هـای برنامه‌نویـسی کدهـای زیـادی را نیـاز دارد، در زبـان پایتـون بـه دلیـل وجـود سـاختارهای داده متعـدد، ماژول‌هـا و ابزارهـای آمـاده بخصـوص کتابخانـه‌هـای متنـوع، بـه سـرعت و بـا نوشـتن کمتریـن تعـداد خـط ، قابـل انجـام می باشـد. امـروزه ایجـاد برنامه‌هـای کاربـردی تحـت وب و تحـت سیسـتم عامـل و نیـز کاربـرد گسـترده در تجزیـه و تحلیـل داده‌هـای بـزرگ و هـوش ماشیـن، بـه لطـف اسـتفاده از زبـان برنامه‌نویـسی پایتـون، بسـیار هموارتـر شـده اسـت. ایـن زبـان، کاملا متـن‌بـاز می‌باشـد و برنامه‌هـای نوشـته شـده بـا آن، بـه سـادگی بـر روی سیسـتم عامـل‌هـای مختلـف، قابـل اجـرا اسـت. تسـلط و یادگیـری پایتـون آسـان اسـت. اکثـر مـردم می‌تواننـد آن را یـاد بگیرنـد، حـتی کسـانی کـه دانـش برنامه‌نویـسی کمتـری دارنـد. بـا اسـتفاده از یـک زبـان برنامـه نویـسی رایـج، امکانـات زیـادی بـرای یافتـن راه‌حـلی بـرای مشـکلاتی کـه ممکـن اسـت پیـش بیایـد، وجـود دارد. نوشـتن کـد در پایتـون آسـان اسـت. همچنیـن پایتـون بـا طـراحی قابـل دسـترسی اسـت و آن را بـه یـکی از سـریع‌ترین زبان‌هـا از نظـر سـرعت توسـعه تبدیـل می‌کنـد. خوانـدن کـد پایتـون بصـری اسـت و حفـظ آن را آسـان می‌کنـد. طـرح زبـان تقریبـاً نزدیـک بـه انگلیـسی و خوانـا اسـت و تفسیـر آن را آسـان می‌کنـد (واچاراسـاپات و همـکاران، ۲۰۲۲)

مزایای کلیدی پایتون

موفقیـت پایتـون حـول چندیـن مزیـتی اسـت کـه در سـطوح مختلـف دانـش کاربـران از تـازه‌کاران تـا متخصصـان حرفـه‌ای در دسـترس اسـت. پایتـون زبـانی اسـت کـه بـه آسـانی می‌تـوان آن را آموخـت و اسـتفاده کـرد. دسـتور زبـان یـا گرامر زبـان برنامه‌نویـسی پایتون بـا رویکـرد خوانـا بـودن و سـاده بـودن طـراحی شـده اسـت. ایـن سـادگی، پایتـون را زبـانی ایـده‌آل بـرای یادگیـری قـرار می‌دهـد و بـه تـازه‌کاران اجـازه می‌دهـد تـا آن را سـریع بیاموزنـد. در نتیجـه، توسـعه‌دهندگان، زمـان بیشـتری را بـرای فکـر کـردن روی مشـکلی کـه می‌خواهنـد حـل نماینـد صـرف می‌کننـد و زمـان کمتـری در مورد پیچیـدگی زبـان یـا رمزگشـایی کدهـای دیگـران صـرف می‌نماینـد. اگـر شـخصی دررشـته‌ای غیـر از مهنـدسی نـرم افـزار متخصـص می باشـد و قصـد داشـته باشـد تـا وارد دنیـای برنامه‌نویـسی شـود، بهتریـن گزینـه ممکن، زبـان برنامه‌نویـسی پایتون اسـت

پایتـون امـکان بهینه‌سـازی و افـزودن امکانـات جدیـد را بـه متخصصـان و برنامه‌نویسـان می‌دهـد. زبـان برنامـه نویـسی پایتـون بـه صـورت گسـترده‌ای پشـتیبانی می‌شـود. بـر اسـاس رتبه‌هـای بـالا در نظرسـنجی‌ها (نظیـر شـاخص Tiobe و تعـداد زیـاد پروژه‌هـای پایتـون درGitHub) زبـان پایتـون هـم شـناخته شـده‌تر اسـت و هـم بـه صـورت گسـترده اسـتفاده می‌شـود (شـکلهای ۱ و ۲). پایتـون روی هـر سیسـتم عامـل و بـه ویـژه بیشـتر سیسـتم عامل‌هـای کوچک‌تـر اجـرا می‌شـود. بسـیاری از کتابخانه‌هـای مهـم و خدمـات مبتـنی بـر API بـه پایتـون ایـن امـکان را می‌دهـد تـا آزادانـه بـه وسیلـه آن، بـا سـرویس‌ها تعامـل داشـته باشـد یـا مسـتقیما از آن کتابخانه‌هـا اسـتفاده کند. زبـان برنامه‌نویـسی پایتون ممکـن اسـت کـه سـریع‌ترین زبـان برنامـه نویـسی نباشـد در عـوض، تطبیق‌پذیـری آن نسـبت بـه زبانهـای دیگـر بالاتـر می‌باشـد

برنامه‌نویـسی برنامه‌هـای کاربـردی عمـومی بـا پایتـون

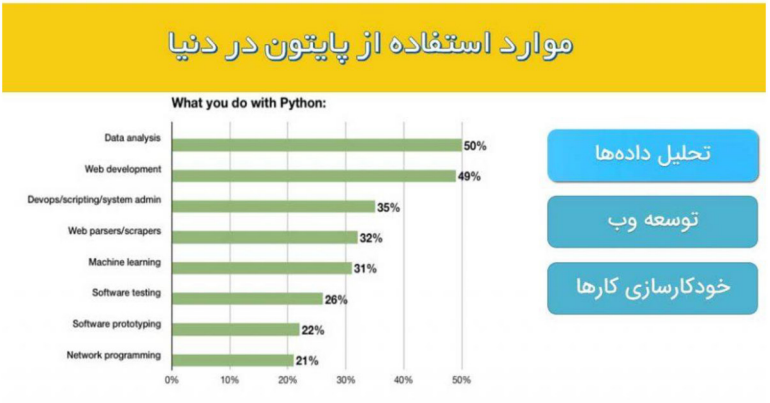
یـک از توانـایی هـای مهـم پایتـون در تولیـد برنامه‌هـای کاربـردی بـا رابـط گرافیـکی بـدون وابسـته بـه سیسـتم عامـل اسـت. علاوه بـر آن، تجزیـه و تحلیـل داده‌هـای پیچیـده بـه یـکی از سـریع‌ترین بخش‌هـای فنـاوری اطلاعـات و مـوارد اصـلی اسـتفاده از پایتـون تبدیـل شـده اسـت. اکثریـت قریـب بـه اتفـاق کتابخانه‌هـای مـورد اسـتفاده در علـم داده و یادگیـری ماشیـن، از رابط‌هـای پایتـون اسـتفاده می‌کننـد بنابرایـن ایـن زبـان را بـه محبوب‌تریـن زبـان برنامه‌نویـسی رابـط سـطح بـالا بـرای کتابخانه‌هـای یادگیـری ماشیـن و سـایر الگوریتم‌هـای عـددی تبدیـل کـرده اسـت

معایب پایتون

پایتـون زبـان سـطح بـالایی اسـت لـذا بـرای برنامه‌نویـسی در سـطح سـخت‌افزار یـا برنامه‌نویـسی سـطح سیسـتم یـا هسـته

شکل ۱- آمار پروژه های ارائه شده در وبگاه GitHub برگرفته از وبگاه https://gisplus.ir

شکل ۲- موارد استفاده از پایتون در دنیا برگرفته از وبگاه https://gisplus.ir



سیستم عامل، مناسب نیست. زمانی که سرعت، اولویت مطلق در برنامه‌های محاسباتی باشد، پایتون بهترین انتخاب نبوده و بهتر است از زبان‌هایی نظیر C/C++ یا دیگر زبان‌های مناسب استفاده شود

ارتباط و اتصال پایتون به سایر نرم افزارهای آماری

با وجود اینکه نرم‌افزار SPSS و زبان برنامه‌نویسی پایتون دو برنامه مجزا و مستقل هستند ، از نسخه ۱۵ نرم‌افزار SPSS، افزونه‌ای (Add ins) طراحی شد که ارتباط بین SPSS و پایتون را برقرار می‌کند. بنابراین به راحتی می توان در محیط برنامه‌نویسی SPSS از توابع و دستورات پایتون استفاده کرد. به این ترتیب متغیرها و ساختار آن‌ها از SPSS به پایتون ارسال شده و دستورالعمل‌ها به زبان پایتون یک به یک اجرا شده و نتیجه محاسبات به محیط SPSS ارجاع داده می شوند. از طرف دیگر، امکان فراخوانی مولفه‌هایی که درون پنجره خروجی‌های SPSS قرار دارند، توسط پایتون وجود دارد، در نتیجه می‌توان مقدارهای موجود در جدول‌ها و نمودارهای SPSS را به پایتون انتقال و شیوه نمایش آن‌ها را تغییر داد

پایتون یکی از پرکاربردترین زبان‌های برنامه‌نویسی برای توسعه نرم‌افزار است. ساختار و ماهیت این زبان و همچنین رویکرد شی‌گرایی آن به برنامه‌نویسان کمک می‌کند تا کدهای واضح‌تر و کوتاهتر برای پروژه‌های خود بنویسند. از این نظر، پایتون زبانی است که می تواند برای توسعه نرم افزارهای کاربردی علمی، آمار، شبکه، برنامه های دسکتاپ، بازی ها، برنامه های کاربردی وب و غیره مورد استفاده قرار گیرد

کتابخانه‌های پایتون

چندین کتابخانه در پایتون وجود دارد که بر موضوع آمار متمرکز بوده و می‌توانند در مواردی که مجموعه داده‌ها خیلی بزرگ نیستند یا اگر به ورود اطلاعات از کتابخانه‌های دیگر متکی نیستید، استفاده شوند. برخی از پرکاربردترین و محبوب‌ترین کتابخانه‌ها عبارتند از NumPy، Pandas، Seaborn و Matplotlib که در ادامه توضیحاتی در خصوص آنها ارائه می‌شود

NumPy : به کتابخانه‌ای اشاره دارد که برای کار با ماتریس‌ها طراحی شده است و برای استفاده با آرایه‌های تک‌بعدی و چندبعدی استوار است. این کتابخانه شامل چندین روال است که برای تجزیه و تحلیل آماری و همچنین برای ذخیره داده‌ها در آرایه‌ها استفاده می‌شود.

Matplotlib : کتابخانه‌ای برای مصورسازی داده‌ها است. ترکیب کتابخانه‌های SciPy، NumPy و Panda با Matplotlib برای گرفتن عملکرد و نتیجه بسیار مناسب کمک می‌کند

Pandas: انواع کتابخانه‌هایی هستند که برای محاسبات عددی براساس NumPy استفاده می‌شوند. Pandas در مدیریت داده‌های نامگذاری شده از یک نوع با مجموعه‌ای از اشیاء به خوبی کار می‌کند. Pandas از طریق تبدیل داده‌ها به شکل جدولی، اجازه می‌دهد داده‌ها به راحتی خوانده شوند و ساختار بهتری داشته باشند.

Seaborn: یک بسته مصورسازی داده پایتون مبتنی بر mat-plotlib است که عمدتاً به ساختارهای داده Pandas متصل است. مصورسازی جزء اصلی Seaborn است که به درک و کاوش داده‌ها کمک می‌کند

Sklearn/Scikit-learn: مهمترین کتابخانه‌ها برای یادگیری ماشین در پایتون هستند که شامل بسیاری از ابزارهای حیاتی برای خوشه (کلاستر)بندی، گروه‌بندی، کاهش ابعاد و رگرسیون است

SciPy : کتابخانه‌ای است که برای محاسبات علمی استفاده می‌شود که مبتنی بر NumPy است. SciPy یک قابلیت اضافه‌تر از NumPy ارائه می‌کند که شامل آمار SciPy برای تجزیه و تحلیل آماری است.

زیرکتابخانه‌های پایتون

هر کتابخانه در پایتون ممکن است از زیرکتابخانه‌های متعدد و تخصصی تشکیل شده باشد. به عنوان مثال در آنالیز واریانس (ANOVA) که از روش‌های آماری پرکاربرد در تکنولوژی بذر بویژه ارزیابی ارقام گیاهی است وقتی صحبت از

اجـرا (تحلیـل واریانـس) در پایتـون می‌شـود، چندیـن زیـر کتابخانـه وجـود دارد کـه می‌تـوان از آنهـا اسـتفاده شـود. بطـور مثـال، کتابخانـه scipy.stats تابـعی را بـرای تجزیـه واریانـس یـک طرفـه ارائـه می‌دهـد. بدیـن صـورت کـه گروه‌هـا را بـه عنـوان ورودی می‌گیـرد و مقادیـر F و سـطح احتمـال (P-value) را برمی‌گردانـد. همچنیـن زیـر کتابخانـه statsmodels ابزارهـای جامع‌تـری را بـرای ANOVA و آزمون‌هـای پـس از آن فراهـم می‌کنـد. ایـن کتابخانـه بـه شـما ایـن امـکان را می‌دهـد تـا تجزیـه و تحلیل‌هـای عمیق‌تـری، از جملـه بـررسی فرضیـات و انجـام آزمون‌هـای بعـدی را انجـام دهیـد.

مثال عملی از کاربرد پایتون در علوم و تکنولوژی بذر

در پایـان بـه عنـوان مثـال کاربـردی، کدنویـسی تجزیـه بـه مولفه‌هـای اصـلی کـه بـه عنـوان یـکی از روش‌هـای پیشـرفته تجزیـه پایـداری عملکـرد ارقـام گیـاهی اسـت، در سـه برنامـه نرم‌افـزاری مختلـف، فـارغ از نحـوه آماده‌سـازی و خوانـش داده‌هـای آن و نیـز انتخاب‌هـایی کـه بـرای اطلاعـات تکمیلـی قابـل دسـترس اسـت، بصـورت زیـر قابـل ارائـه و مقایسـه اسـت. در مقـالات بعـدی سـعی خواهـد شـد تمـام برنامه‌هـا و کدهـای پایتـون کـه بـرای تجزیـه و تحلیـل مقدمـاتی و پیشـرفته علـوم و تکنولـوژی بـذر و ارزیـابی ارقـام گیـاهی مـورد نیـاز اسـت، معـرفی شـده و نیـز خـروجی برنامه‌هـا مـورد بـررسی و تحلیـل قـرار گیـرد

الف) در نرم افزار SAS

```
Proc princomp out= prin N=... std;
Var x1-x5;
Proc plot;
Plot prin2*prin1;
Run; quit;
```

ب) در نرم افزار R

```
pca.com=prcomp(comp)
summary (pca.com)
biplot (pca.com, main="biplot', scale= … , cex= ….)
```

ج) در برنامه پایتون

۱- با استفاده از کتابخانه statsmodels

```
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.multivariate.pca import PCA


((plt.rc("figure", figsize=(16, 8
data = sm.datasets.fertility.load_pandas().data
data.head()
```

2- ۲-با استفاده از کتابخانه sklearn

```
import matplotlib.pyplot as plt
        From sklearn.decomposition import PCA

pca= PCA(n_component=2)
principalComponent=pca.fit_transform(x)

principalIDf=pd.DataFrame(data= principalComponent,
column=[' principalComponent 1', ' principalComponent
2'])
```

توصیه ها

در دوره معاصـر، تأثیـر برنامه‌نویـسی در کشـاورزی دگرگـون کننـده بـوده اسـت و دقـت، کارایی و پایـداری را بـه شیوه‌هـای کشـاورزی مـدرن اضافـه کـرده اسـت. زبان‌هـا و فناوری‌هـای برنامه‌نویـسی، تصمیم‌گیـری مبتـنی بـر داده، اتوماسیـون و بهینه‌سـازی منابـع در کشـاورزی را ممکـن سـاخته‌اند. بـا ادامـه رشـد جمعیـت جهـان، نقـش برنامه‌ریـزی در کشـاورزی در تضمیـن امنیـت غـذایی و پایـداری زیسـت‌محیطی حیاتی‌تـر خواهـد بـود. زبان‌هـای برنامه‌نویـسی کلیـدی ماننـد پایتـون، R و جـاوا بـه همـراه فناوری‌هـای مرتبـط بـا نرم‌افزارهـای جغرافیـایی ماننـد GIS، اینترنـت اشیـا و اتوماسیـون، در توسـعه برنامه‌هـا و سیسـتم‌های کشـاورزی حیـاتی هسـتند

در آنالیـز داده‌هـا در پایتـون در مقایسـه بـا زبان‌هـای دیگـر ماننـد جـاوا یـا C ، از خطـوط کـد کمتـری اسـتفاده می‌شـود. بدیـن صـورت کـه وقـتی کدهـا، خـط کمتـری بـرای مـرور نیـاز داشـته باشـند و خوانـدن آنهـا، بیشـتر شـبیه انگلیـسی باشـد، بازبینـی کدهـا، بسیـار سـریع‌تر و آسـان‌تر خواهـد بـود. بنابرایـن هـر بـار کـه کـد

تغییر می‌کند، به روزرسانی کمتری احتیاج دارد، تا آنجا که به هر تابع مربوط می‌شود، محاسبه به سرعت انجام می‌شود.

با نگاهی به امکانات دو نرم‌افزار قدیمی‌تر SAS و R در مقایسه با پایتون، جهت تجزیه‌های آماری، شاید استفاده از نرم‌افزار R در آمار تخصصی به عنوان یک محیط برنامه‌نویسی داده باز رایگان، دارای برتری نسبت به پایتون (بویژه در انعطاف‌پذیری در نمایش نمودارهای آماری) باشد اما با یادگیری پایتون که به مراتب آسانتر است، علاوه بر تامین نیازها در تجزیه و تحلیل آمار تخصصی در علوم و تکنولوژی بذر و ارزیابی ارقام، می‌توان از کاربردهای عمومی آن نیز در توسعه اپلیکیشن‌های کاربردی استفاده کرد و برنامه‌نویسی را به عنوان یک تخصص به دامنه دانش و مهارت یک متخصص تکنولوژی بذر اضافه کرد

منابع

-Bødker, M. S., Wilkinson, C. J., Mauro, J. C., & Smedskjaer, M. M. (2022). StatMechGlass: Python based software for composition-structure prediction in oxide glasses using statistical mechanics. SoftwareX, 17, 100-913. https://doi.org/10.1016/j.softx.2021.100913.

-Giroux, B. (2021). ttcrpy: A Python package for travel-time computation and raytracing. SoftwareX, 16, 100-134. https://doi.org/10.1016/j.softx.2021.100834.

-Gunaratne, C., & Garibay, I. (2021). NL4Py: Agent-based modeling in Python with parallelizable NetLogo workspaces. SoftwareX,16 , https://doi.org/10.1016/j.softx.2021.100801.

-Hudson Pérez, M. C., Förster Marín, C. E., Rojas-Barahona, C. A., Valenzuela Hasenohr, M. F., Valdés, P. R., & Ferré, A. R. (2013). Comparison of the effectiveness of two strategies teaching methodology in the development of the reading comprehension in the first school year. Perfiles Educativos, 35 (140), 100-118. https://doi.org/10.1016/s0185-2698(13)71824-5.

-Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). Applied linear statistical models (5th). New York, NY: McGraw-Hill Irwin.

Ott, R. L., and Longnecker, M. (2010). An introduction to statistical methods and data analysis. Belmon, CA: Brooks/Cole.

org/10.1016/j.cpc.2021.108245.

-Takefuji, Y. (2021). SCORECOVID: A Python Package Index for scoring the individual policies against COVID-19. Healthcare Analytics,1 (August), 100-105. https://doi.org/10.1016/j.health.2021.100005

-Watcharasupat, K. N., Lee, J., & Lerch, A. (2022). Latte: Cross-framework Python package for evaluation of latent-based generative models. Software Impacts, 100-222. https://doi.org/10.1016/j.simpa.2022.100222.

-Zolotov, O., Romanovskaya, Y., & Knyazeva, M. (2021). pyFIRI - A free and open source Python software package of the non-auroral Earth's lower ionosphere. SoftwareX, 16, 100-105. https://doi.org/10.1016/j.softx.2021.100885

-Zurheide, F. T., Hermann, E., & Lampesberger, H. (2021). PyBNBowTie: Python library for Bow-Tie Analysis based on Bayesian Networks. Procedia Computer Science, 180 (2019), 344-351. https://doi.org/10.1016/j.procs.2021.01.172.

-Matsuda, F., Maeda, K., Taniguchi, T., Kondo, Y., Yatabe, F., Okahashi, N., & Shimizu, H. (2021). mfapy: An open-source Python package for 13C-based metabolic flux analysis. Metabolic Engineering Communications, 13(July), 177-188. https://doi.org/10.1016/j.mec.2021.e00177.

-Meng, S., Fu, Z., Qin, J., Ma, X., Li, Y., Hao, L., Liu, Y., Sun, K., & Chen, D. (2021). magcoilcalc: A Python package for modeling and optimization of axisymmetric magnet coils generating uniform magnetic field for noble gas spin-polarizers. SoftwareX, 16, 100-105. https://doi.org/10.1016/j.softx.2021.100805.

-Naranjo, S, I, L,, Brito, N, A, E., Núñez, V, A, V& Ordóñez, E, M, R., (2022), Analysis of the use of the Python programming language for statistical calculations. Spirals multidisciplinary journal of scientific research, 6 (2): 1-13

-Pérez-García, F., Sparks, R., & Ourselin, S. (2021). TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine,208 , 200-230. https://doi.org/10.1016/j.cmpb.2021.106236.

-Rokhlin, V., Szlam, A and Tygert, M.2009. A randomized algorithm for principal component analysis.

-Schwarz, S., Uerlich, S. A., & Monti, A. (2021). pycity_scheduling-A Python framework for the development and assessment of optimisation-based power scheduling algorithms for multi-energy systems in city districts. SoftwareX, 16, 100-104. https://doi.org/10.1016/j.softx.2021.100839.

-Silvestri, L. G., Stanek, L. J., Dharuman, G., Choi, Y., & Murillo, M. S. (2022). Sarkas: A fast pure-python molecular dynamics suite for plasma physics. Computer Physics Communications, 272, 1100-1140. https://doi.