



عنوان مقاله:

آشنایی با نقش‌های مؤثر در اجرای گام‌های فرآیند داده

چکیده مقاله:

در فرآیند ETL، داده‌ها می‌بایست از منابع معتبر اخذ شوند و حاوی ارزش‌های صحیح باشد. کیفیت داده‌ها از طریق بررسی نویز و نقاط دور افتاده **Noise and outlier**، داده از دست رفته **Missing values** و داده تکراری **Duplication** مورد ارزیابی قرار می‌گیرد. نقاط دورافتاده یا نویز در اثر تغییرات ناخواسته در داده‌های اصلی ایجاد شده‌اند و ویژگی آن‌ها به صورت قابل ملاحظه‌ای با دیگر نقاط داده متفاوت است. در تعیین نقاط دورافتاده، می‌توان از پروفایل رفتارهای نرمال استفاده کرد.

از مجموع مقالات آکادمی نیک درس

[لینک دسترسی به این مقاله در وب سایت آکادمی نیک درس](#)



دانشمند داده (Data Scientist)

وظیفه اصلی تحلیل‌های پیچیده برعهده دانشمند داده می‌باشد. معمولاً انتظار می‌رود که یک دانشمند داده بتواند سؤال‌های پیچیده کسب‌وکار را با روش‌های مبتنی بر آمار و یادگیری ماشین پاسخ دهد. خیلی وقت‌ها لازم است با استفاده از داده‌ها اتفاقی را پیش‌بینی نماید و خیلی وقت‌های دیگر هم لازم است اطلاعات بیشتری را از داده‌های موجود کسب نماید. طبیعتاً لازم است که به آن کسب‌وکار اشراف نسبی پیدا کند.

دانشمند داده در خیلی از موارد برای انجام اموراتش مجبور است تصمیم‌گیری نماید که کدام بخش از داده‌ها بهتر است که در تحلیل‌ها باشند و خیلی وقت‌ها هم مجبور می‌شوند که نواقص داده‌ای را جبران نمایند. درنهایت لازم است که بتوانند نتایج را به صورت یک داستان علمی دقیق همراه با ارائه یکسری گرافیک از مراحل را که طی شده به صورت گزارش علمی دقیق ارائه کنند. خیلی وقت‌ها در روند بررسی سؤال‌های کسب‌وکار، دانشمند داده به صورت تجزیه و تحلیل اولیه‌ای که بانام تجزیه و تحلیل اکتشافی داده شناخته می‌شود، به بینش جدید و کامل‌تری از داده‌های موجود دست پیدا می‌کند که این خودش می‌تواند مجدد یافته ارزشمندی برای آن سازمان یا کسب‌وکار باشد.

وظایف دانشمند داده:

- شناسایی منابع داده‌ای با ارزش و اتوماتیک سازی فرآیندهای جمع‌آوری داده‌ها
- انجام پردازش روی داده‌های ساختار یافته و بدون ساختار
- تحلیل مقدار زیادی از اطلاعات و کشف الگوها و روندها

- ساخت مدل‌های پیش‌بینی‌کننده و الگوریتم‌های یادگیری ماشین
- ارائه اطلاعات با استفاده از تکنیک‌های مصورسازی
- پیشنهاد راه‌حل‌ها و استراتژی‌هایی برای چالش‌های کسب‌وکار

مهارت‌های لازم برای این تخصص عبارت است از:

- آمار
- برنامه‌نویسی
- قصه‌گویی
- بصری‌سازی
- تجزیه و تحلیل کسب‌وکار

برخی ابزارهای قابل بررسی برای این تخصص عبارت است از:

- SQL
- Python
- R
- Apache spark
- Rapidminer
- Knime



نیک درس



مهندسی داده (Data Engineer)

مهندس داده، معمار استخوان بندی مسیر جمع آوری، یکپارچه سازی و پاک سازی اولیه داده را برعهده دارد. در واقع مهندس داده شخصی است که داده ها را از طیف وسیعی از منابع ساختار یافته و بدون ساختار جمع آوری کرده و آن ها را یکپارچه نموده و در نهایت، برای تحلیل های تحلیل گر یا دانشمند داده آماده سازی می نماید. معمولاً اغلب مهندسان داده تجربه کار قبلی به عنوان تحلیل گر یا بعضاً دانشمند داده را داشته اند.

وظایف مهندس داده عبارت است از:

- طراحی، توسعه، ساخت، آزمایش و نگهداری معماری داده ها
- هم راستاسازی معماری داده ها با نیازهای کسب و کار
- طراحی و ایجاد زیرساخت مورد نیاز برای جمع آوری داده ها بر اساس فرآیندهای کاری
- طراحی و ایجاد زیرساخت مورد نیاز برای استخراج بهینه، تبدیل، تغییر، ترمیم، پاک سازی و بارگذاری داده ها از طیف گسترده ای از منابع اطلاعاتی با استفاده از فناوری SQL و Big data
- ایجاد راهکارهایی جهت ارتقای کیفیت و قابلیت اطمینان داده ها
- یافتن الگوهای پنهان با استفاده از داده ها
- تبدیل و تغییر داده ها به منظور استفاده توسط تحلیل گر یا دانشمند داده و ارائه مدل های پیش بینی و مدل سازی تجویزی

مهارت های لازم برای این تخصص عبارت است از:

- برنامه‌نویسی
- پایگاه داده
- طراحی الگوریتم

برخی ابزارهای قابل بررسی برای این تخصص عبارت است از:

- SQL
- Python
- R
- Apache spark
- Hadoop
- SCALA



تحلیل‌گر داده (Data Analyst)

تحلیل‌گر داده، داده‌های خام را دریافت نموده و با پردازش آن، روابط پنهان داده‌ها را کشف کرده و از انبوه داده‌های بدون ساختار، پاسخ سؤال‌های کسب‌وکار را می‌یابد. این افراد با داده‌های بسیار زیادی همچون اعداد و ارقام، حقایق و داده‌های

خام و مشکلاتی که شرکت‌های مختلف با آن دست‌وپنجه نرم می‌کنند، سروکار دارند. بنابراین، یک تحلیل‌گر داده حرفه‌ای باید توانایی تحلیل و شناخت مشکلات، مهارت ریاضی و دقت در جزئیات را داشته و با استفاده از داده‌های موجود، بهترین راه‌حل‌ها را شناسایی و ارائه نماید. همچنین باید با استفاده از تفکر انتقادی، بتواند روابط پنهان بین اعداد و حقایق را کشف نماید.

وظایف تحلیل‌گر داده:

- برای تعیین اهداف اولیه با بخش‌های مربوط به فناوری اطلاعات، مدیران و کارکنان در تعامل باشد.
- داده‌ها را از منابع اولیه و حتی ثانویه جمع‌آوری کند.
- داده‌ها را پاک‌سازی کند به این معنی که داده‌های مشکل‌دار را ترمیم نموده و داده‌های تکراری را حذف کند، داده‌هایی که دارای مقادیر غلط هستند یا خالی هستند را با مقادیر مناسب نظیر میانگین یا اعداد تصادفی جایگزین کند و ...
- تجزیه و تحلیل داده‌ها را استفاده از ابزارها و روش‌های استاندارد آماری انجام دهد.
- روندها، همبستگی‌ها، الگوها را شناسایی و کشف کند (با استفاده از تفکر نقادانه و ابزارهای کمکی).
- فرصت‌های جدید را برای هر چه بهتر کردن روندهای موجود شناسایی کند.
- گزارش‌های مختصر و مفید و در عین حال کامل و جامع برای کمک به تصمیم‌گیری مدیران تهیه کند.
- پایگاه داده‌هایی طراحی و ایجاد کند تا دانش استخراج شده را ذخیره‌سازی و در زمان موردنیاز از آن استفاده کند.

مهارت‌های لازم برای این تخصص عبارت است از:

- قصه‌گویی
- بصری سازی
- تجزیه و تحلیل کسب و کار
- تهیه شاخص‌های کلیدی عملکرد

برخی ابزارهای قابل بررسی برای این تخصص عبارت است از:

- Power BI
- Tableau
- Qlikview
- QlikSense

سر فصل‌های کلی پیشنهادی برای تحلیل پیشرفته کسب و کار و علم داده

سرفصل‌های کلی پیشنهادی برای تحلیل پیشرفته کسب‌وکار و علم داده عبارت است از:

- Data warehouse concepts, design, and data integration
 - Introduction to data, information and knowledge
 - What is data model (dimension vs 3NF)
 - What is data warehouse
 - Business side of data warehousing
 - Why build a data warehouse?
 - The value of a data warehouse
 - Key concepts (measurement, facts, dimensions, ...)
 - Data warehouse architectures (Inmon, Kimball and Stand-alone data marts)
 - Kimball's dimensional modeling (dimension design, fact table design)
 - DW performance
 - ETL processes and tools

• Data management fundamentals

- مدیریت داده به‌عنوان یک دارایی سامانی
- مدیریت داده: ضرورت همسویی با اهداف کسب‌وکار
- مدیریت داده: اصول، چالش‌ها، موانع و فرصت‌ها
- مدیریت داده: چارچوب‌ها، روش‌ها و استانداردها
- مدیریت داده: چارچوب DMBOK

• The data management body of knowledge

- مروری بر DMBOK2
- حاکمیت داده
- متادیتا
- کیفیت داده
- معماری و مدل‌سازی داده
- داده‌های مرجع و Reference

- Statistics for business analytics
 - Key concepts of statistics
 - Data preparation
 - Data quality and consistency assessment
 - Linear methods
 - Regression samples and application
 - Correlation vs cause/effects
 - Probability vs statistics
 - Statistical significant
 - KPI definitions

- DW/BI project management
 - Project management concepts
 - DW/BI project management methodologies
 - Agile DW/BI project management
 - BI projects and BI program in organization
 - Which BI tool is the right choice?
 - BI roadmap
 - Real cases of applied projects
 - Why do BI projects fail?

- Data driven decision making for managers
 - سیستم‌های تفکر دو گانه و تله‌های متداول تصمیم‌گیری
 - خطاهای شناختی متداول در برآوردهای کمی
 - انواع تصمیم‌گیری و اقتضائات آن در شرایط متفاوت محیطی کسب‌وکار
 - ابزارهای تحلیلی (Analytics) و نقش آن در تبدیل اطلاعات خام به بینش لازم برای تصمیم‌گیری
 - گام‌ها و فرآیندهای حل مسئله
 - تکنیک‌های ساده و مؤثر برای جمع‌آوری و استفاده از اطلاعات در جهت حل مسئله
 - روش‌های کمی پیشرفته‌تر برای تصمیم‌گیری و حل مسئله

• Data mining concepts, tools, and applications

- مقدمه‌ای بر مفاهیم داده‌کاوی و کاربردهای آن
- آشنایی با متن‌کاوی، وب‌کاوی، فرآیندکاوی، تحلیل احساسات و یادگیری عمیق
- آشنایی با برنامه‌نویسی در MATLAB و داده‌کاوی سریع در نرم‌افزار Rapidminer
- آشنایی با منطق فازی و طراحی، پیاده‌سازی و تحلیل سیستم استنتاج فازی
- پیاده‌سازی الگوریتم‌های رگرسیون (داده‌های پیوسته) و طبقه‌بندی (داده‌های گسسته)
- پیاده‌سازی الگوریتم‌های خوشه‌بندی داده‌ها
- آشنایی با شبکه‌های عصبی و پیاده‌سازی شبکه عصبی چندلایه
- آشنایی با سیستم‌های استنتاج فازی عصبی انطباقی و پیاده‌سازی آن

• Information virtualization and BI dashboards

- Introduction to information virtualization
- Data abstraction
- Fundamental graphs and data transformation
- Graphical components and mapping strategies
- Dashboard and storytelling with data

• Python programming

- مقدمه‌ای بر زبان برنامه‌نویسی پایتون و محیط برنامه‌نویسی آن
- آشنایی با کتابخانه‌های مهم در پایتون
- استفاده از انواع متغیرها، عملگرها، ساختارهای داده و مروری بر کاربرد آن‌ها
- پیاده‌سازی دنباله‌ها، لیست‌ها، تاپل‌ها، دیکشنری‌ها و مجموعه‌ها به همراه متدهای مرتبط
- انواع شرط‌ها، حلقه‌ها، دستورات کنترلی، توابع و کاربرد آن‌ها
- آشنایی با برنامه‌نویسی شیء‌گرا، کلاس‌ها و وراثت در پایتون
- کار با انواع فایل‌ها، مصورسازی داده‌ها، ترسیم و تحلیل نمودارها
- آشنایی با کتابخانه‌های مرتبط با ایجاد ساختارهای پیشرفته داده در علم داده و یادگیری ماشین

• Data science and machine learning in python

- آشنایی با مفاهیم علم داده، تحلیل‌های پیشرفته، یادگیری ماشین و روند تکامل آن‌ها

- مروری بر ارتباط بین علم داده و رایانش ابری، داده‌های عظیم، اینترنت اشیا و بلاک‌چین و آینده آن‌ها
- بررسی آرایه‌ها، ماتریس‌ها، سری‌ها، چارچوب‌ها و پیش‌پردازش داده در پایتون
- مروری بر تحلیل فرضیه‌ها، آزمون‌های آماری، تحلیل واریانس، کواریانس و همبستگی متغیرها
- پیاده‌سازی انواع روش‌های یادگیری ماشینی نظارت شده و اعتبارسنجی آن‌ها
- پیاده‌سازی انواع روش‌های یادگیری ماشینی بدون نظارت و اعتبارسنجی آن‌ها
- مروری بر روش‌های فرا ابتکاری و الگوریتم‌های تکاملی و کاربرد آن‌ها
- بررسی کاربردهای متن‌کاوی، تحلیل احساسات و یادگیری عمیق در تحلیل پیشرفته کسب‌وکار

- R programming

- Overview of the R language, basic concepts (vectors, matrices and data frames)
- Reading and writing data in R
- If, loops and functions
- Efficient computation in R
- Data virtualization: basic and advanced methods (ggplot2)
- Exploring, cleaning, and preparing data

- Data science for business with R

- Data virtualization
- Statistical tests
- Multiple linear regression
- Logistic regression
- Classification using a nearest neighbor analysis
- Cluster analysis
- Market basket analysis
- Churn analysis

- Introduction to big data and distributed data processing

- Big data: why and where
- Characteristics of big data and dimensions of scalability

- Big data standards
 - Data lake architecture and its best practice
 - Big data ingestion tools and solutions
 - Big data storage tools and solutions
 - Big data processing tools and solutions
 - Big data governance
- Big data analytics and stream processing
 - Implement stream processing using Apache Spark streaming
 - Consume events from source, apply logics and send it to a data sink
 - Understand message deliveries in stream data processing
 - Create a job to analyze data in real-time using the Apache Spark streaming API
 - Single event processing and micro-batch approach to processing events
 - Real-time event processing
 - Real-time dashboard
 - Advanced topics in business analytics
 - Data warehouse in the age of AI maturity
 - Real-time data warehousing
 - Real-time data analytics
 - The path from reports to AI
 - The path to predictive analytics and machine learning
 - Business science problem framework

عنوان مقاله:

آشنایی با نقش‌های مؤثر در اجرای گام‌های فرآیند داده

نتیجه‌گیری:

داده‌های هر سیستم کسب‌وکار در خلال مجموعه‌ای از فرآیندها ایجاد می‌شوند. بررسی فرآیندهای کسب‌وکار و دستیابی به بینش حاصل از لاگ‌های سیستم می‌تواند گامی مؤثر در تحلیل و دستیابی هوشمندی باشد.

از مجموع مقالات آکادمی نیک درس

[لینک دسترسی به این مقاله در وب سایت آکادمی نیک درس](#)

آکادمی نیک درس، در انتهای تابستان 1400 فعالیت خود را باهدف انتقال تجربه مدرسین در حوزه‌های مختلف از جمله دروس مدرسه و دانشگاه (تمامی رشته‌ها و تمامی مقاطع)، دروس حوزه علمیه، مهارت‌های بازار کار، هنر، صنعت، آشپزی، نرم‌افزارهای کاربردی و ... باکیفیت مطلوب جهت سربلندی کشور اسلامی عزیزمان ایران، آغاز نموده است.

تلاش شبانه‌روزی همکاران ما در آکادمی نیک درس جهت ارائه خدمات آموزش باکیفیت و مطلوب، باقیمتی پایین جهت پیش برد اهداف از قبل تعیین شده از جمله تحقق عدالت آموزشی، دسترسی آسان و باکیفیت به آموزش‌های متنوع در زمینه‌های گوناگون برای اقشار مختلف جامعه در استان‌ها، شهرها و روستاها و حتی در مناطق کمتر برخوردار می‌باشد.

امید است که با معرفی آکادمی نیک درس به دوستان و آشنایان خود، ما را جهت پیشبرد این اهداف یاری نمایید.

در صورت تمایل به تدریس و یا تمایل به نویسندگی مقالات در آکادمی نیک درس، مهارت‌های خود را در [فرم همکاری با ما](#) در آکادمی نیک درس، ثبت نمایید.



www.NikDars.com