

فهرست مطالب

مقدمه ناشر ۵

فصل اول

چگونه یک تیم داده تشکیل دهیم ۷

- ۷ تیم داده
- ۸ چه زمانی به علم داده نیاز است؟
- ۱۰ مهندس داده: صلاحیت‌ها و مهارت‌ها
- ۱۲ دانشمند داده: صلاحیت‌ها و مهارت‌ها
- ۱۵ مدیر تیم علم داده: صلاحیت‌ها و مهارت‌ها
- ۱۷ مصاحبه برای علم داده
- ۱۹ مدیریت تیم علم داده
- ۲۱ ارزیابی موفقیت تیم
- ۲۲ توانمندسازی دیگران برای استفاده از داده‌ها
- ۲۴ مشکلات مشترک متقابل
- ۲۵ مشکلات داخلی مشترک
- ۲۸ چه کاری لازم است انجام دهید تا یک مدیر عالی شوید؟
- ۳۱ چه موقع باید متخصص شد
- ۳۲ آیا شما آماده MACHINE LEARNING هستید؟
- ۳۴ نحوه استخدام دانشمند داده

فصل دوم

استخدام تیم داده ۳۸

- ۱- درک روشنی از اینکه چرا می‌خواهید یک دانشمند داده استخدام کنید، داشته باشید ۳۸
- ۲- پروفایل‌های علم داده را بفهمید و بدانید چه شخصیتی را می‌خواهید استخدام کنید ۳۹
- ۳- کاری را که می‌خواهید انجام دهند مشخص کنید ۴۰
- ۴- یک فرهنگ مبتنی بر داده ایجاد کنید که به دانشمندان داده کمک می‌کند ۴۰
- ۵- تصمیم بگیرید که در سازمان شما کجا و چگونه قرار می‌گیرند ۴۱
- ۶- یک پشتیبان یا مدیر محصول اختصاصی داشته باشید ۴۱
- موارد استفاده سازمانی از علم داده ۴۸

۴۹	چرخه عمر پروژه علم داده
۴۹	پارامترهای شغلی علم داده
۵۰	مهارت‌هایی برای ارزیابی در یک دانشمند داده
۵۳	همه کاره و هیچ کاره؟
۵۳	بنابراین، یک کسب‌وکار چه کاری می‌تواند انجام دهد؟
۵۶	استخدام دانشمند داده: خوب، بد و زشت

فصل سوم

فرآیند علم داده تیمی چیست؟

۶۰	اجزای اصلی TDSP
۶۱	چرخه حیات علم داده (DATA SCIENCE LIFECYCLE)
۶۴	ساختار پروژه استاندارد شده
۶۶	زیرساخت‌ها و منابع برای پروژه‌های علوم داده
۶۷	ابزارها و برنامه‌های کاربردی برای اجرای پروژه
۶۸	نقش و وظایف فرآیند علم داده تیم
۶۸	تعریف و وظایف برای چهار نقش TDSP
۶۹	وظایفی که باید توسط چهار نقش انجام شود
۷۱	گردش کار اجرای پروژه علم داده
۷۳	توسعه سریع پروژه‌های علم داده
۷۴	وظایف آماده‌سازی داده‌ها برای یادگیری پیشرفته ماشین

فصل چهارم

قوانین یادگیری ماشین: بهترین روش‌ها برای مهندسی ML

۸۲	واژه‌شناسی
۸۳	بررسی اجمالی
۸۴	قبل از یادگیری ماشین

مقدمه ناشر

تقدیم به

آنان که دوستشان دارم

پدر و مادرم عزیزم، همسر فداکار و مهربانم و فرزندانم

دکتر امین نظارات

فصل اول

چگونه یک تیم داده تشکیل دهیم

تیم داده

اولین چیزی که باید هنگام ساخت یک تیم علم داده بدانید، این است که علم داده یک ورزش گروهی است. حل پروژه‌های واقعی و عملی در موضوع داده‌ها، نیاز به همکاری گروه بزرگی از افراد دارد. این افراد شامل دانشمندان داده، به‌علاوه مدیران آنها و همچنین مهندسان داده که زیرساخت‌ها را انجام و توسعه می‌دهند می‌شود. در ادامه به تشریح وظایف و جزئیات هر یک از این نقش‌ها خواهیم پرداخت.

مهندس داده: شخصی است که ممکن است یک پایگاه داده بسازد، یا ممکن است داده‌ها را از آن پایگاه داده بیرون بکشد تا افراد تجزیه و تحلیل کنند یا ممکن است الگوریتم‌های یادگیری ماشینی در سطح تولید را بسازد و آنها را بر روی سرورها پیاده‌سازی کند، به طوری که بتوانند در یک پایگاه داده عظیم از مشاهدات داده‌ها اعمال شوند و حتی ممکن است به صورت برخط اجرا شوند؛ بنابراین، آنها هم از نظر سخت‌افزاری و هم از نظر نرم‌افزاری مهارت‌هایی در توسعه زیرساخت‌ها دارند و در واقع کمتر تحلیل داده‌های روزانه را انجام می‌دهند.

تحلیل داده‌ها به شخصی که **دانشمند داده** خوانده می‌شود واگذار می‌شود؛ بنابراین، یک دانشمند داده ممکن است شخصی باشد که داده‌ها را از یک پایگاه داده بیرون بکشد، سپس ممکن است داده‌ها را تجزیه و تحلیل کرده، آزمایش‌هایی روی آن داده‌ها انجام دهد، بصری‌سازی را انجام داده و نتایج را به مدیر علوم داده و سایر افراد سازمان منتقل نماید تا کارها را رو به جلو ادامه دهند. غالباً، یک دانشمند داده هر الگوریتم یادگیری ماشینی یا الگوریتم پیش‌بینی را توسعه داده، آنگاه با محوریت یک مهندس داده اقدام به پیاده‌سازی مقیاس‌پذیر آن می‌نماید.

سومین شخص کلیدی **مدیر علوم داده** است. مدیر علوم داده اطمینان حاصل می‌کند که همه اعضای تیم داده با یکدیگر ارتباط برقرار می‌کنند. آنها همچنین تیم متخصصین و مهندسين داده را جذب و آماده به کار

می‌سازند. یک مدیر داده با مدیران بالادست در سازمان و همینطور با همکارانی که در سطح سازمان در یک سطح هستند ارتباط دارد. به‌عنوان یک مدیر داده اطمینان حاصل کنید که همه اطلاعات در اختیار شما قرار می‌گیرد. یکی از وظایف مدیر علم داده معرفی و ایجاد آشنایی سایر گروه‌ها و واحدهای سازمان از توانایی‌ها و قابلیت‌های تیم داده است. تشریح اینکه توانایی‌های آنها چیست، چه کاری می‌توانند انجام دهند و چگونه می‌توانند به شما کمک کنند. آنها به‌عنوان یک واحد با هم کار می‌کنند و غالباً هر یک از این افراد در حال کار بر روی پروژه‌های منفرد یا زیرمجموعه‌های فردی موارد کاربرد علم داده هستند. به‌طور مرسوم اعضای تیم داده گرد هم می‌آیند و جلسات گروهی مشترک و ارائه‌های مشترک دارند و در آنجا ایده‌های خود را مورد بحث قرار می‌دهند. آنها همچنین با افراد خارج از سازمان ارتباط برقرار می‌کنند. ممکن است از طریق جلسات فردی با افراد خارج از تیم علم داده ارتباط مستقیم داشته باشند، یا امکان دارد از طریق مدیر تیم داده با یکدیگر تعامل داشته باشند؛ بنابراین، به هر صورت آنها باید بتوانند با زیرساخت علوم داده و آنچه به تیم علم داده مرتبط است، ارتباط برقرار کنند.

چه زمانی به علم داده نیاز است؟

نحوه ارتباط با تیم علم داده کمی به نوع سازمانی بستگی دارد و تا حدودی بستگی به اندازه سازمانی دارد که قرار است تیم داده در آن تشکیل شود؛ بنابراین، هنگامی که شما تازه کار خود را شروع می‌کنید، زمانی که یک شرکت در مراحل اولیه است، یا فقط یک نفر یا تیم بسیار کوچک هستید، شاید نیاز نباشد سازمان خود را درگیر فرایندهای پیچیده تحلیل داده کنید و همچنین لازم نباشد که در مورد چگونگی انجام ارزیابی، نحوه یادگیری ماشینی، چگونگی انجام پیش‌بینی و پردازش پایین‌دستی نگران باشید. اولین قدم این است که فقط مطمئن شوید که داده‌های سازمان مرتب است و روش انجام این کار اطمینان از تمرکز بر زیرساخت‌ها است؛ بنابراین، اولین کاری که شما باید انجام دهید ایجاد زیرساخت برای ذخیره داده‌ها، پایگاه داده‌ها و موارد دیگر است. قطعاً در ادامه به نرم‌افزارهایی برای نگهداری و مدیریت داده‌ها، بصری‌سازی، ساخت گزارشات و ... نیاز خواهید داشت؛ بنابراین، همه این موارد در ابتدا نیاز به ایجاد زیرساخت مناسب دارند. اغلب اولین افرادی که می‌خواهید در یک تیم علم داده استخدام شوند، افرادی نیستند که شما لزوماً به آنها دانشمند داده می‌گویید، آنها داده‌ها را تجزیه و تحلیل نمی‌کنند، یادگیری ماشینی را انجام نمی‌دهند. ممکن است آنها مقدار کمی از این کار را انجام دهند، اما بیشتر

فقط در جهت اطمینان از کارکرد سامانه‌ها و اطمینان از جمع‌آوری اطلاعات، ایمن‌بودن، ذخیره‌شدن و کاربردی بودن آن دخیل هستند؛ بنابراین، هنگامی که شما یک سازمان متوسط هستید، احتمالاً زیرساخت‌های اساسی را در اختیار دارید و می‌توانید به فکر ساخت تیم علم داده واقعی باشید. برای انجام این کار می‌توانید افرادی را که در واقع دانشمندان داده نامیده می‌شوند، استخدام کنید. این افراد کسانی هستند که واقعاً از داده‌ها استفاده می‌کنند؛ بنابراین، ممکن است داده‌ها را از پایگاه داده استخراج کنند و برخی آزمایشات و تحلیل‌ها را انجام دهند. دانشمندان داده ممکن است الگوریتم‌های یادگیری ماشینی را بسازند و داده‌ها را تجزیه و تحلیل کنند تا ببینند آیا می‌توانند الگوها یا روندهای رفتاری را که برای شما مهم است شناسایی کنند یا خیر. در آن مرحله، شما در حال ساختن تیم علم داده و همچنین در حال پیاده‌سازی این ایده‌ها و محصولات علم داده هستید؛ بنابراین، دانشمند داده ممکن است چیزی مانند الگوریتم یادگیری ماشینی بسازد که مثلاً رفتار مصرف‌کننده را پیش‌بینی نماید. هنگامی که آن الگوریتم را ساختید، ممکن است لازم باشد دوباره آن را در سیستم خود پیاده‌سازی کنید و ممکن است لازم باشد آن را مقیاس‌پذیر نمایید تا بتواند روی کل مجموعه داده (احتمالاً بیگ دیتا) اجرا شود. ممکن است بخواهید به ساخت نوعی از خروجی‌های قابل فهم بپردازید که افرادی که لزوماً دانشمندان داده نیستند قابلیت درک و فهم آنها را داشته باشند. این امر می‌تواند توسط گروه دیگری از افراد تیم با سمت مهندسی داده انجام شود؛ بنابراین هنوز نگرانی‌های زیرساختی وجود دارد، زیرا شما مجموعه بزرگی از داده‌ها را دارید که باید امیدوار بود در این مرحله جمع‌آوری کرده باشید. باید در این خصوص مطمئن عمل کنید، باید یک پایگاه داده داشته باشید و بتوانید آن را مقیاس‌پذیر نمایید، اما اکنون شما به‌نوعی فارغ‌التحصیل شده و دیدگاه کاملی از علم داده به‌دست آورده‌اید. برای یک سازمان بزرگ نیز همین اجزا مورد نیاز است. اکنون یک زیرساخت داده دارید، ممکن است یک تیم علم داده داشته باشید که ارزیابی‌ها را انجام می‌دهد و امکان دارد از این ارزیابی‌ها برای تصمیم‌گیری استفاده کنید.

اکنون شما یک مؤلفه اضافی دارید که آن مدیریت تیم و هماهنگ‌کردن همه افراد است. در این مرحله نقش مدیر علوم داده کمی بیشتر درگیر می‌شود، به این معنا که ممکن است شما چندین تیم از دانشمندان داده را که روی پروژه‌های مختلف کار می‌کنند، هماهنگ کنید. ممکن است تیمی داشته باشید که منحصراً در ساخت پروژه‌های یادگیری ماشینی کار کند. همچنین امکان دارد تیم دیگری داشته باشید که منحصراً بر روی

اجرای آزمایشات و استنباط آنچه می‌توانید از آن آزمایشات به‌دست آورید کار کنند. کسی باید مسئول هماهنگی آن فعالیت‌ها باشد تا مطمئن شود که آنها به افراد مناسب سازمان شما متصل هستند. چه این تیم بازاریابی باشد، چه گروه تجاری یا هر کس دیگری که با آنها همکاری می‌نمایید. شما باید بتوانید آن افراد را بهم وصل کنید و بنابراین در این مقیاس، زیرساخت کامل علم داده موجود است. در ادامه به تشریح و توضیح بیشتر هریک از این نقش‌ها پرداخته خواهد شد.

مهندس داده: صلاحیت‌ها و مهارت‌ها

یکی از اعضای اصلی تیم علم داده، مهندس داده است. در اینجا در مورد شرایط و مهارت‌هایی که ممکن است در یک مهندس داده نیاز باشد صحبت خواهد شد. یک مهندس داده چه کاری انجام می‌دهد؟ آنها ممکن است کارهایی مانند ایجاد زیرساخت را انجام دهند؛ بنابراین، آنها می‌توانند پایگاه داده شما و سخت‌افزار آن را ایجاد کنند. خواه خرید یا سازماندهی و سفارش آن باشد. همچنین مهندسان داده بر فرایند تهیه، توسعه و اجرای نرم‌افزارهای بالادستی اعم از آنهایی که بر روی دیتابیس کار می‌کنند یا آنهایی که در لایه بالای سرور قرار می‌گیرند نظارت داشته و یا خود اجرا را برعهده می‌گیرند. علاوه بر این آنها ذخیره‌سازی و استفاده از داده‌ها را مدیریت می‌نمایند. مهندسين داده ممکن است نظارت کنند که چه کسی، چه زمانی و از چه داده‌هایی استفاده می‌کند. آنها ممکن است داده‌ها را جدا کرده و به افراد دیگری بدهند و یا اینکه ابزار تولید را پیاده‌سازی نمایند. همان‌طور که می‌بینید هریک از این امور متفاوت می‌باشد و ممکن است شما بیش از یک نفر داشته باشید که همه این کارها را انجام دهد.

مطمئناً در ابتدا، هنگامی که شما یک سازمان بسیار کوچک هستید، امکان دارد به یک شخص نیاز داشته باشید که به‌نوعی بتواند کل زیرساخت داده را اداره کند و با داشتن یک سازمان بزرگتر، ممکن است افراد متخصص بیشتری داشته باشید. حال سؤال این است که آنها به چه مهارت‌هایی نیاز دارند؟

ممکن است به برخی از دانش‌ها در مورد سخت‌افزار مناسب برای جستجو، چه از نظر ذخیره‌سازی و چه از نظر محاسبات، نیاز باشد. ممکن است لازم باشد که در مورد نرم‌افزار پایگاه داده اطلاعاتی داشته باشند. از این نظر، آنها باید بتوانند با دانشمندان داده ارتباط برقرار کنند. مهندسين داده همچنین باید با مفاهیم و اصول پردازش

داده‌ها در مقیاس بزرگ آشنا بوده و یا تجربه کاری داشته باشند. در سازمان‌های بزرگ این موضوع به دلیل اینکه اکثر مسائل داده محور با حجم بزرگی از داده‌ها مواجه هستند ضروری است. مهندس داده باید بتواند آن مقیاس‌های پردازشی و الگوریتم‌های پیش‌بینی داده‌ای را که توسط دانشمندان داده و تیم‌های تحلیل توسعه داده شده‌اند در مقیاس بزرگ اجرا کند. پیشینه مهندسی داده اغلب علوم کامپیوتر و مهندسی کامپیوتر است، اما آنها می‌توانند از تخصص‌های دیگر نیز باشند.

نکته اصلی که مدیران تیم داده به دنبال آن هستند این است که آیا مهندسی داده می‌تواند آن مشاغل را که سازمان شما برای انجام آن مسائل به آنها احتیاج دارد انجام دهند؟ آیا آنها قادر به ایجاد زیرساخت هستند؟ آیا می‌توانند زیرساخت‌هایی را که سازمان به آنها نیاز دارد تا بتوانند به خوبی سرویس دهد، حفظ کنند؟ آیا مهندسی داده می‌تواند علم داده یا الگوریتم‌های یادگیری ماشین یا بررسی‌های آماری را در مقیاس لازم به روشی که سازمان می‌خواهد پیاده‌سازی کنند؟ بنابراین مهندسی داده باید بتواند مشکلاتی را که اکنون به پاسخ آنها نیاز دارید حل کنند. با وجود یک مهندس داده، تعادل راه‌حل‌ها در برابر نرم‌افزار ممکن است کمی متفاوت از دانشمند داده باشد. از آنجاکه این فرد شخصی است که به نوعی داده‌ها را برای شما حفظ می‌کند، شما به آنها نیاز دارید تا بتوانند این کار را به روشی سازگار با نحوه انجام کار سازمان شما انجام دهند؛ لذا آنها باید بتوانند ایده‌ای در مورد نیازهای نرم‌افزاری و سخت‌افزاری خاص سازمان داشته باشند.

چند نکته اساسی وجود دارد که این افراد باید بدانند، ممکن است لازم باشد بدانند چگونه برخی از پایگاه داده‌ها را ایجاد و مدیریت کنند، مواردی مانند SQL، مواردی مانند MongoDB همچنین ممکن است لازم باشد بدانند که چگونه کارهایی مانند پیاده‌سازی یا اجرای کدهای SQL را انجام دهند، ابزارهایی مانند Hadoop و Spark که یک زیرساخت پردازش موازی هستند از لوازمات ضروری این متخصصین است که باید بدانند. البته لزوماً نیازی نیست که آنها هریک از این واژه‌های پر سروصدا را بدانند و به آنها مسلط باشند، اما لازم است ترکیبی از مهارت‌ها را که به ایشان امکان می‌دهد تا زیرساخت داده‌ای ایجاد کنند که از قابلیت نصب و راه‌اندازی این ابزارها برخوردار باشد و بتوان آن را نگهداری کرد، داشته باشند؛ بنابراین در هنگام جستجوی یک مهندس داده، دو ویژگی اصلی را جستجو می‌کنید. ابتدا، آنها باید تمایل به یافتن پاسخ مسائلی که با آنها مواجه می‌شوند داشته باشند بنابراین، اغلب باید بتوانند خودشان به این سوالات پاسخ دهند. آنها باید بتوانند جستجو کرده و بفهمند که

سخت‌افزار مناسب چیست، اقدامات امنیتی مناسبی که باید انجام شود و موارد دیگر. غالباً آنها مجبورند این کار را کمی به‌تنهایی انجام دهند، به این معنا که تیم مهندسی داده اغلب در یک زمینه بسیار خاص متخصص هستند. باید بدانند که داده چگونه استفاده، چگونه واکنشی و چگونه تحلیل می‌شود. در حالت ایده‌آل، آنها از نزدیک با دانشمندان داده همکاری خواهند کرد، البته شاید قبلاً این کار را کرده باشند. همچنین باید بتوانند تحت فشار، خوب کار کنند. زیرساخت داده برای یک سازمان اغلب بسیار مهم است و اگر از حالت عملیاتی خارج شود، سازمان قادر به انجام هیچ‌گونه تجزیه و تحلیل نیست، یا سازمان به‌نوعی متوقف می‌شود؛ لذا داشتن یک مهندس داده که بتواند تحت فشار کاری زیاد به‌خوبی کار کند، قادر به ادامه کار و حفظ امور است ضروری خواهد بود. آنها باید بتوانند با سایر افراد سازمان تعامل داشته و قادر باشند با اصطلاحات ساده توضیح دهند که چرا لازم است تصمیمات خاصی بگیرند؛ بنابراین اغلب تصمیمات، به‌ویژه این روزها، کاملاً فنی و مرتبط با سخت‌افزار و نرم‌افزار است؛ بنابراین اگر آنها بتوانند با زبانی ساده توضیح دهند چه اتفاقی افتاده است، مشکل چیست و چگونه برطرف می‌شود بسیار مفید است. این موضوع می‌تواند یک مزیت بزرگ برای سازمان شما باشد.

دانشمند داده: صلاحیت‌ها و مهارت‌ها

وقتی تیم علم داده خود را می‌سازید، تعجب‌آور نیست که یکی از مهمترین نقش‌هایی که برای استخدام در نظر گرفته می‌شود، یک دانشمند داده باشد؛ بنابراین در این بخش درباره این نقش و صلاحیت‌ها و مهارت‌های آن بیشتر صحبت خواهیم کرد.

ابتدا باید بدانید یک دانشمند داده چه کاری انجام می‌دهد؟ همان‌طور که به‌طور معمول تعریف می‌شود، یک دانشمند داده، آزمایش‌هایی را انجام می‌دهد، داده‌ها را تغییر شکل داده و تمیز می‌کند، آنها را تجزیه و تحلیل کرده و سپس نتایج را به ذینفعان منتقل می‌نماید. آنها باید مجموعه‌ای از مهارت‌ها را داشته باشند که به آنها امکان انجام همه این فعالیت‌ها را بدهد. بسته به تعریف شما از علم داده، ممکن است نیاز به کمی جابجایی در مهارت‌ها یا افراد وجود داشته باشد. این تعریف بسته به بلوغ سازمان می‌تواند کمی متفاوت بوده و یا مهارت‌های متفاوتی را ترسیم نماید. اگر تیم شما در مرحله استخدام دانشمندان داده است، این بخش در تدوین شرح وظایف آنان به شما کمک خواهد کرد. اگر در مرحله خیلی ابتدایی هستید و اولین اعضای تیم علم داده خود را استخدام می‌کنید، شاید

نیاز باشد که این افراد کم‌وبیش هم وظایف مهندسی داده و هم دانشمندان داده را انجام دهند؛ بنابراین در اینجا در مورد نوعی سازمان بالغ صحبت می‌کنیم که دانشمندان داده را استخدام می‌کند. دانشمندان داده لازم است که دانش پایه‌ای خوبی در زمینه آمار داشته باشند، خواه دارای مدرک تحصیلی دانشگاهی بوده یا تعداد زیادی کلاس‌های مرتبط را گذرانده باشند. به‌طور معمول دانشمندان داده دارای دو مهارت متفاوت استنتاج و پیش‌بینی هستند. این موضوع مهم است که بدانید برخی از افراد در یکی بهتر هستند و برخی از افراد در دیگری. انتخاب بین این افراد یا انتخاب هر دو گروه به‌نوعی به آنچه سازمان انجام می‌دهد بستگی دارد. اگر آنها بیشتر در حوزه پیش‌بینی تخصص دارند، لازم است در حوزه یادگیری ماشینی تخصص داشته باشند و یا کسب کنند، اما اگر می‌خواهید آزمایش‌هایی انجام دهید و فرضیه‌های جدیدی ارائه نمایید، ممکن است لازم باشد در استنتاج کمی بهتر باشند. همچنین باید بتوانند تجزیه و تحلیل داده‌ها را انجام دهند. آنها باید بتوانند مجموعه داده‌ها را از یک پایگاه داده واکنشی کرده، تجزیه و تحلیل نموده و تمیز کنند. در نهایت استنباط یا پیش‌بینی آماری یا ماشینی را که به دنبال آن بوده‌اند را جمع‌بندی، مصورسازی و ارائه نمایند؛ بنابراین با کمک این افراد شما می‌توانید مطلع شوید به هنگام تصمیم‌گیری داده‌های مورد نظر تا چه اندازه از تصمیمی که می‌خواهید بگیرید، پشتیبانی می‌کنند. همان‌طور که گفته شد، دانشمندان داده معمولاً چند مهارت زیر را دارند.

- ☑ آنها معمولاً می‌دانند که چگونه از R یا Python استفاده نمایند. این زبان‌ها دانش عمومی در علم داده محسوب می‌شوند که افراد می‌توانند از آنها برای تجزیه و تحلیل داده‌ها استفاده کنند.
- ☑ این افراد می‌دانند چگونه نوعی بصری‌سازی را انجام دهند. گاهی اوقات تجسم تعاملی با چیزی مانند D3.js. همچنین به‌طور قطع باید بدانند که چگونه داده‌ها را از یک پایگاه داده بیرون بکشند؛ بنابراین، آنها حداقل تجربه کار با یک پایگاه داده SQL، MongoDB یا چیزی شبیه به آن را خواهند داشت.
- ☑ این افراد لازم است کمی مهندسی نرم‌افزار را آموخته باشند. آنها کمی در مورد ابزارها و روش‌های کنترل نسخه و کمی در مورد برنامه‌نویسی به زبان‌های دیگر می‌دانند. این موارد به آنها امکان می‌دهد تا با سایر افراد تیم که بخش مهندسی نرم‌افزار را برعهده دارند ارتباط داشته باشند، اما آنها در درجه اول در تحلیل و آمار متمرکز هستند.

دسته دیگری که در بین دانشمندان داده بسیار رایج است این است که آنها درواقع مهندسين نرم‌افزاری هستند که به‌صورت رسمی در دانشگاه در همین رشته تحصیل کرده‌اند. آنها درواقع یک مهندس نرم‌افزار آموزش‌دیده یا یک مهندس نرم‌افزار در حال کارند و فقط به دانش آمار آشنا هستند. انتخاب بین این گروه‌ها به ترکیبی از آنچه سازمان شما به‌دنبال آن است بستگی دارد. آیا سازمان به دنبال شخصی است که با کمی دانش آماری آمیخته، توسعه زیرساخت‌های بیشتری را انجام دهد؟ پس ممکن است بهتر باشد به سراغ شخصی بروید که دارای تخصص مهندسی نرم‌افزار به‌علاوه سابقه آماری باشد. اگر از قبل تیم داده دارید، یک نکته را باید به خاطر بسپارید که یکی از مؤلفه‌های اصلی دانشمندان داده خوب بودن آشنایی با انواع جدید ایده‌ها یا نرم‌افزارها نیست. این بدان معناست که در برخی موارد واقعاً مهم است که به‌دنبال شخصی باشید که دارای مهارت‌های خاص مهندسی نرم‌افزار یا زبان برنامه‌نویسی و مهارت‌های خاصی باشد که سازمان شما به آنها نیاز دارد، اما اغلب با این کار نامزدهای بسیار قدرتمند را از دست خواهید داد؛ بنابراین، یک ایده خوب این است که به‌طور کلی ارزیابی کنید، آیا شخص کاندید یک فرد با خصوصیات مورد نظر شما هست؟ آیا وی قادر به یادگیری زمینه‌های جدید می‌باشد؟ آیا قادر به انجام این نوع وظایف علمی-کاربردی روی داده‌ها مطابق با نیاز سازمان هست؟ و سپس سعی کنید بفهمید که آیا آنها قادر به شروع یک کار جدید هستند؟ آیا آنها با زبان‌های برنامه‌نویسی که شما برای استفاده در اجرای موردهای کاربرد خاص خود به آنها نیاز دارید آشنا هستند یا خیر.

ویژگی‌های کلیدی دانشمندان داده که ممکن است به‌دنبال آنها باشید چیست؟

یکی اینکه آنها مایل به یافتن پاسخ به‌تنهایی هستند؛ بنابراین، اغلب یک دانشمند داده باید بتواند مجموعه اطلاعات مورد نیاز خود برای شناسایی مسائلی که برای حل آنها نیازمند تحلیل داده هستیم را از جامعه مخاطب در قالب مصاحبه یا روش‌های دیگر دریافت کند، آنها باید بتوانند با ذینفعان همکاری کنند تا بفهمند مشکل اصلی برای حل چیست. یک سؤال بسیار هوشمندانه هنگام مصاحبه با یک دانشمند داده، مطرح کردن یک مسئله روی داده‌ای که قبلاً هرگز ندیده‌اند، خواه از سازمان شما باشد یا حتی از داده‌هایی که بسیار دور از ذهن باشد و ببینید که وی چگونه در برابر آن واکنش نشان می‌دهد. اگر آنها به روشی مثبت واکنش نشان دهند، ایده‌هایی داشته باشند، به روش‌هایی فکر کنند که می‌توانند با این مشکل مقابله کنند، بسیار عالی است. اگر آنها با این نوع داده‌ها مرعوب شوند و سعی کنند که دوباره به مجموعه داده‌هایی که می‌شناسند برگردند، به‌احتمال فراوان در آینده با

تغییر داده‌ها در سازمان یا تغییر صورت‌مسئله‌ها با مشکل مواجه خواهید شد. باید ببینید که آیا آنها راحت می‌گویند، من نمی‌دانم. برای شما بسیار معمول است که یک آزمایش را انجام دهید و نتیجه‌گیری این است که ما بیش از آنچه قبل از شروع کار می‌دانستیم نمی‌دانیم. البته این موضوع ناامیدکننده است، و همچنین یکی از قسمت‌های سخت در مورد علم داده همین واقعیت است، اما این موضوع مهم است که بدانید مردم در صورت مواجهه با آن راحت خواهند گفت "من نمی‌دانم"، بنابراین شما می‌دانید که آنها حقیقت را به شما خواهند گفت. آنچه در پاراگراف قبل ذکر شد به این معنا است که علم داده نیاز به صرف زمان فراوان دارد تا تمام جزئیات را به‌درستی درک کرده و تا اطمینان حاصل شود که به‌عنوان یک فرد تحلیلگر با ذینفعان به‌گونه‌ای تعامل دارد که کارها را به جلو سوق می‌دهد. داشتن مهارت‌های نرم رفتاری و کاری همراه با دانشمند داده بودن در یک تیم علم داده که در آن قرار است با افراد دیگر کار کنند، بسیار مهم است. به‌طور معمول، مخاطبین برای مؤلفه‌های فنی بیش‌ازحد ارزش قائل می‌شوند مثل اینکه آیا آنها زبان برنامه‌نویسی X را می‌دانند؟ و اگر سازمان و تیم داده برخی از باورهای عمیق آنها را در مورد داده‌ها زیر سال ببرد، ممکن است باعث ایجاد مقاومت از سمت آنها شود. این نکته بسیار مهم است که اطمینان حاصل شود که دانشمندان داده توانایی ایجاد تعامل با تمامی ذینفعان را داشته و می‌توانند به اهداف مورد نظر پروژه دست یابند.

مدیر تیم علم داده: صلاحیت‌ها و مهارت‌ها

آخرین عضو تیم علم داده مدیر تیم داده است. یک مدیر علوم داده چه کاری انجام می‌دهد؟ اولین چیزی که این فرد ممکن است مسئول آن باشد ساخت تیم علم داده است. آنها مسئول شناسایی و استخدام مهندسان داده و دانشمندان داده در تیم هستند. هدایت نیروهای استخدام‌شده، تعیین اهداف و اولویت‌ها، شناسایی آن دسته از مشکلات موجود در یک سازمان که باید توسط علم داده حل شوند و به‌نوعی قراردادن افراد مناسب بر روی مورد کاربرد مناسب از جمله وظایف مهم این افراد است. آنها فرآیند تولید دانش از داده را مدیریت می‌کنند. فرایند اجرای یک پروژه علم داده یک فرایند تک مرحله‌ای نیست و نیازمند تکرارهای زیاد است. این‌گونه پروژه‌ها دارای یک فرایند تکراری هستند و باید با دقت مدیریت شوند. یکی از مسئولیت‌های مدیران تیم داده اطمینان از تعامل افراد در تیم داده با یکدیگر است، اما همچنین با افراد دیگر در سازمان خارجی و نیز با گروه‌های دیگر تعامل دارد. آنها ممکن است نیاز باشد که به مدیریت بالاتر گزارش دهند. همچنین ممکن است با افراد در سطوح

و واحدهای مختلف سازمانی تعامل داشته باشند یا با آنها همکاری کنند و بنابراین باید مهارت‌های ارتباطی خوبی از این نظر داشته باشند.

آنها به چه نوع مهارت‌هایی نیاز دارند؟ در حالت ایده‌آل، آنها از نرم‌افزار و سخت افزار مورد استفاده آگاهی دارند. اگر به‌نوعی سابقه کار در علم داده یا مهندسی داده داشته باشند، بسیار عالی است، زیرا در این صورت در واقع نوعی از زیرساخت‌های مربوطه را می‌دانند و اگر مشکلی پیش بیاید، ممکن است پیشنهاد خوبی در مورد چگونگی اصلاح زیرساخت علم داده یا نحوه اصلاح الگوریتم یادگیری ماشینی ارائه کنند. آنها باید نقش‌ها را به‌خوبی بشناسند، باید بدانند که یک دانشمند داده چه کاری انجام می‌دهد، یک مهندس داده چه کار می‌کند، تیم‌های دیگر قرار است چه کار کنند. آنها می‌توانند به‌نوعی موارد کاربردی را که لزوماً برای زمان فعلی مناسب نیستند یا به شکل کلی باید از دستور کار خارج شوند، فیلتر کرده و بر روی موارد کاربرد مهمتر تمرکز کنند. اغلب مشکلات و سؤالاتی در کسب‌وکار سازمان وجود دارد که فقط با علم داده حل نمی‌شوند. این موضوع می‌تواند به دلیل در دسترس نبودن داده‌ها باشد یا می‌تواند به این دلیل باشد که الگوریتم‌ها آنقدر خوب نیستند که بتوانند آن نوع پیش‌بینی را در آن زمان انجام دهند. البته این موضوع می‌تواند به این دلیل باشد که ما به‌طور قطع می‌توانیم مشکل را حل کنیم، اما نمی‌توانیم آن را در مقیاس لازم مقیاس بندی کرده که توان پاسخ‌دهی با سؤالات مطرح‌شده را داشته باشد؛ بنابراین یک مدیر باید به‌نوعی تصور کند که طبق تجربیات برتر جهانی آنچه با علم داده انجام می‌شود چیست و چه کاری نمی‌تواند انجام شود و پس از آن برخورداری از مهارت‌های ارتباطی قوی بدیهی است؛ چراکه آنها در حال مدیریت تیمی هستند که هم در درون خود نیاز به فعالیت نزدیک و تیمی است و هم نیازمند داشتن ارتباطات قوی با سایر اجزای سازمان یا تیم‌های کارفرمایی است.

◀ سابقه مدیران داده چیست؟

مدیران داده به‌طور ایده‌آل از نوعی پیشینه علم داده برمی‌آیند. خواه تجزیه و تحلیل داده‌ها باشد، خواه دارای سابقه کار در زیرساخت‌های علم داده که در هر دو حالت لازم است برخی آموزش‌های مدیریتی را دیده یا در حال گذراندن آن باشد. ویژگی‌های کلیدی این افراد دانش کافی در علم داده است. از آنجاکه آنها وظیفه مهم پشتیبانی از عملکرد افراد تیم را برعهده دارند؛ بنابراین مثبت و حمایتی و انگیزشی بودن هنگام شناسایی یک مدیر علم داده بسیار مهم است. اغلب به عهده مدیر علم داده است تا نتایج ناامیدکننده الگوریتم پیش‌بینی یا آزمایشی که

به نتیجه نرسیده است را گزارش دهد، آنها معمولاً باید این موضوع را به مدیریت یا افرادی که ممکن است از این موضوع خیلی راضی نباشند گزارش دهند. مدیران تیم داده باید مایل باشند این کار را به روشی مثبت و دوستانه انجام دهند. این همان چیزی است که شما در یک مدیر علوم داده جستجو می‌کنید. شخصی که می‌تواند به‌نوعی کل فرآیند را سازماندهی کند، ادامه دهد و ذینفعان را به سمت موردهای کاربرد مناسب سوق دهد و مشکلات واقعی سازمان شما را حل کند.

مصاحبه برای علم داده

معمولاً سه مؤلفه اصلی در مصاحبه علوم داده وجود دارد. در ابتدا برگزاری جلسات فردی یا جلسات گروهی است. این موضوع به‌اندازه سازمان شما بستگی دارد. مصاحبه‌شوندگان هم با افراد داخل تیم علوم داده و هم با افراد خارج از تیم علوم داده ملاقات خواهند کرد، بنابراین می‌توانید در مورد اینکه آیا آنها هم با افراد داخلی و هم با افراد خارجی ارتباط خوبی برقرار کرده‌اند بازخورد بگیرید. نکته اصلی در اینجا تمرکز بر روی پروژه‌هایی است که آنها روی آنها کار کرده‌اند. تا جایی که می‌توانید اطلاعات مربوط به نحوه کار آنها در پروژه‌ها را دریافت کنید. همچنین بسیار مهم است که شما در این مرحله ارزیابی مناسبی از آنها داشته باشید. رزومه در مورد مهارت‌های فنی آنها چیزهای زیادی به شما می‌گوید، اما در مورد مهارت‌های بین فردی آنها بسیار کم صحبت خواهد کرد؛ بنابراین این نوعی مرحله حساس است که در آن شما باید کالبدشکافی و شناسایی کنید که آیا آنها شخصی هستند که به‌راحتی با افراد دیگر سازمان شما کار می‌کنند. در صورتی که می‌خواهید ارزیابی جامعی از مهارت‌های فنی آنها داشته باشید می‌توانید از سؤالات فنی استفاده کنید؛ اگرچه، بسیاری از سازمان‌های بزرگ، از جمله سازمان‌های علوم داده واقعاً مشهور، در مصاحبه‌ها، به‌نوعی از نوع معمایی سؤالات بهره می‌برند یا سؤالات پازل از نوع ریاضی را طراحی می‌کنند، اما این موضوع نمی‌تواند توانایی آنها را در مواجهه با مسائل دنیای واقعی محک زند پس بهتر است به‌نوعی سؤالات عمومی بپردازید. چگونه می‌توانید با این نوع مشکلات مقابله کنید؟ از چه راهکاری برای حل این مسئله استفاده می‌کنید؟ در زمان مواجهه با یک مشکل بسیار بزرگ چه روشی را پیش خواهید گرفت؟ در پاسخ به این سؤالات ممکن است مصاحبه‌شونده با مباحث فنی بپردازند اما شما قطعاً می‌خواهید روی درک مفهومی آنها تمرکز کنید. آیا آنها می‌دانند چگونه با مشکلات کنار بیایند؟ آیا آنها نوع صحیح مذاکره و مصاحبه را می‌دانند؟ آیا به‌نظر می‌رسد که آنها تجربه‌ای در حل انواع مشکلاتی دارند که اگر وارد

سازمان شما شوند می‌توانید از آنها کمک بگیرید؟ جزء کلیدی بعدی این است که توانایی آنها را در ارائه موضوعات و نتایج بسنجید. این موضوع به دو دلیل خوب است. اول، این یک توانایی است که آنها بتوانند خلاقیت خود را نشان دهند و سهم منحصربه‌فرد خود را که می‌توانند در سازمان شما داشته باشند نشان دهند؛ بنابراین اگر آنها در مورد تحلیل‌هایی که انجام داده‌اند، یا مدل و نرم‌افزاری که ساخته‌اند، ارائه‌ای داشته باشند می‌توان در خصوص توانایی آنها هم در حوزه قابلیت‌های فردی و هم فنی اطلاعات خوبی به دست آورد. این موضوع همچنین از زاویه بررسی مهارت‌های ارتباطی دارای اهمیت است. تقریباً همیشه دانشمند داده، مهندس داده، مدیر داده، مجبورند نتایج خود را به ذینفعان ارائه دهند. خوب است که درست از ابتدا ارزیابی کنید، آیا آنها قادر به انجام این کار هستند، آیا توانایی پاسخگویی به سؤالات را دارند. در اینجا دو نوع مکتب فکری وجود دارد. یکی بسیار فنی است، شاید حتی مانند مشکلاتی مثل یک معما باشد و دیگر این که در واقع آنها را وادار به تلاش برای حل یک مسئله بسیار کوچک اما واقعی کنند. به نظر می‌رسد که مورد اخیر معمولاً بسیار مؤثرتر است؛ بنابراین برای یک دانشمند داده، ممکن است یک مجموعه داده کوچک به آنها بدهید و ببینید آیا آنها می‌توانند آن را تحلیل کنند و گزارش دهند؟ حل این مسئله ممکن است برای آنها تنها یک ساعت طول بکشد، چون آن را بسیار کوچک و یا چیزی شبیه به آن می‌دانند اما شما بعد از آن با آنها مصاحبه خواهید کرد شما از آنچه انجام داده‌اند سؤال کرده و مطلع می‌شوید و خواهید فهمید که روند تفکر آنها واقعاً چگونه است. این مسئله ممکن است یک نمونه شبیه‌سازی شده یا یک مثال غیرواقعی باشد؛ بنابراین آنها به هیچ داده واقعی دسترسی ندارند، اما به شما ایده می‌دهد. چگونه آنها می‌توانند به صورت مستمر در سازمان فعالیت کنند؟ بنابراین برای یک انتخاب خوب و دقیق سه مؤلفه مورد توجه است، ابتدا جلسات فردی یا گروهی برای ارزیابی شخصیت و توانایی‌های کلی آنها برگزار می‌شود. سپس پروژه بسیار کوچکی را انجام می‌دهند و در نهایت مهارت‌های فنی آنها را ارزیابی می‌کنید. سپس تمام اطلاعات را از هر کسی که با آنها مصاحبه کرده است جمع‌آوری کرده و در مورد اینکه آیا مناسب سازمان شما هستند، قضاوت می‌نمایید.

مدیریت تیم علم داده

هنگامی که تیم شما در حال فعالیت است، فرایند مدیریت یا حفظ تیم با پیشرفت و پاسخگویی به مشکلات علم داده نقش مهمی برای یک مدیر علوم داده است؛ بنابراین اولین چیزی که به طور ایده‌آل اغلب با هریک از اعضای تیم خود می‌خواهید، جلسات فردی است؛ بنابراین این اتفاق می‌افتد به‌ویژه اگر در اولین مرحله مدیریت باشید که فقط چند نفر را مدیریت می‌کنید. هنگامی که سازمان شما کمی بزرگتر می‌شود، یک ساختار سلسله‌مراتبی وجود خواهد داشت و شما ممکن است لزوماً با هر عضو تیم هر هفته مثلاً دیدار داشته باشید، اما اگر شما تنها مدیریت یک تعداد کمی از افراد را برعهده دارید، می‌توان با یک جلسه ایستاده به‌طور منظم و روزانه که در آن آنها می‌توانند هر چیزی را که در حال کار بر روی آن هستند ارائه دهند. باید توجه داشت که جلسات فردی وقت زیادی را می‌گیرد ولی اگر به آن تن دهید می‌تواند یک مزیت بزرگ باشد به این معنا که افراد تیم اغلب در علم داده در مورد چیزی گیر می‌کنند که فقط با نگاه فردی دیگری به‌راحتی قابل حل است. بسیاری از اوقات، این اتفاق برای من افتاده است، جایی که من در حال تجزیه و تحلیل برخی از داده‌ها هستم و فقط در برخی از جزئیات سطح پایین گیر می‌کنم و نمی‌توانم کاملاً آن را کشف کنم و سپس به محض اینکه آن را به شخص دیگری نشان داده‌ام بلافاصله متوجه مشکل من شده است. دومین چیزی که داشتن آن بسیار منطقی است، جلسات تیم علم داده است. در این جلسات تمامی اعضا در جلسه حاضر شده و موارد خود را مطرح کرده و از سایر اعضای تیم بازخورد می‌گیرند؛ بنابراین دوباره آنها می‌توانند هر به‌روزرسانی از هر پیشرفتی را که به‌دست آورده‌اند با دیگران به اشتراک بگذارند. آنها می‌توانند هر مشکلی را که روی آنها کار می‌کنند را مطرح نمایند. به‌ویژه مشکلات خاصی که شاید تیم‌های مختلف آن را لمس می‌کنند؛ بنابراین تیم علم داده لزوماً نمی‌تواند مجموعه داده‌ای را که می‌خواهند ایجاد کنند، زیرا زیرساخت‌ها فراهم نیست، و آن دقیقاً زمان مناسب برای برقراری ارتباط با تیم مهندسی داده است. در این جلسات مراحلی که آنها می‌خواهند بردارند مورد بحث قرار خواهد گرفت؛ این همچنین فرصتی برای نوعی بررسی همگانی از پیشرفت کارها است. نکته کلیدی در اینجا این است که شما باید آن را به‌گونه‌ای تنظیم کنید که افراد در صورت لزوم توانمند شوند یا بتوانند مشکلاتشان را مطرح کنند. آنها می‌توانند نظرات انتقادی درباره عملکرد دیگران را بدون وقاحت و بی‌ادبی بیان کنند. نوعی خط بسیار خوب برای تعادل وجود دارد و بسیاری از افراد در انتقاد خوب مشکل دارند. در این جلسات می‌توان در خصوص سؤالات زیر به

بحث و تبادل نظر پرداخت. مانند: چرا ما داریم این کار را می‌کنیم؟ چرا ما در حال حل همه این مشکلات فنی واقعاً سخت هستیم؟ چه هدفی را تعیین می‌کنیم و چگونه می‌توانیم به آنجا برسیم؟ و این فرصتی است که شما برای ایجاد انگیزه در افراد ایجاد می‌کنید، همه را هیجان‌زده می‌کنید، همه را در مورد موردهای کاربردی که در هر زمان مشخص روی آنها کار می‌کنید تازه و شاداب نگه می‌دارید، اما در جلسات فردی، در آن هر فرد می‌تواند در مورد مسائل شخصی خود که برای مقابله با مشکلات و انجام وظایف با آنها مواجه است صحبت نماید. شما همچنین مسئول نظارت بر تعاملات هستید. دانستن چگونگی تعدیل این نوع درخواست‌ها و گرفتن مقداری انرژی برای علم داده یا مهندس داده‌ای که این درخواست‌ها را دریافت می‌کند، چیزی است که به عهده مدیر داده است. بهترین راه برای انجام این کار طبق تجربه نویسنده سیاست درب باز است؛ بنابراین به‌طور کلی این خوب است که افراد بتوانند در هر زمان و نه فقط در جلسه هفتگی آنها یا جلسه فردی یا جلسه گروهی خود، به شما مراجعه کنند. روش دیگر داشتن سیاستی برای پاسخگویی سریع به سؤالات از طریق ایمیل یا چت است، خصوصاً اگر درخواست‌ها مواردی هستند که خیلی سریع حل می‌شوند. شما همچنین می‌خواهید بتوانید جلسات رسمی تیم را به حداقل برسانید. علوم داده و مهندسی داده‌ها به تلاش بی‌وقفه و متمرکز زیادی نیاز دارند؛ بنابراین معرفی جلسات بیش از حد در مسیر این افراد به معنای آن است که آنها واقعاً کار خود را کند می‌کنند؛ بنابراین، یکی از راه‌های انجام این کار اجازه‌دادن به ارتباط سریع است که لازم نیست در یک جلسه رسمی حل شود. همچنین مدیریت رشد سازمان مهم است و این رشد از چند طریق مختلف است. اطمینان حاصل کنید که فرصت‌هایی برای یادگیری ابزارهای جدید برای آنها وجود دارد. خواه این امر از طریق دسترسی به دوره‌های آنلاین یا سایر موارد باشد. خواه از طریق ارائه در جلسه گروه، جایی که افراد ایده‌های جدیدی ارائه می‌دهند یا ابزارهای جدیدی که در مورد آنها آموخته‌اند تا افراد دیگر بتوانند از آنها استفاده کنند. خواه معرفی آن برای تعاملات جدید باشد؛ بنابراین گاهی اوقات در جلساتی خواهید بود که لزوماً جلساتی نیستند که تیم داده شما در آن شرکت می‌کند، اما ممکن است یک گروه خارجی یا شخصی را شناسایی کنید که واقعاً ارتباط خوبی با تیم علوم داده شما برقرار می‌کند. وظیفه شما به‌عنوان مدیر علوم داده است که این تعامل را تسهیل کنید و سپس، یک مؤلفه بسیار مهم برای مدیر داده بودن و ایجاد یک تیم علمی خوب برای داده‌ها، شناسایی فرصت‌های پیشرفت برای افراد شما است.

ارزیابی موفقیت تیم

در یک سطح، شما نیاز به بحث در مورد موفقیت گروه دارید، موفقیت را می‌توان از چند طریق مختلف تعریف کرد. یکی حل مشکلات سازمانی است؛ بنابراین، مثلاً بگویید ما می‌خواهیم رشد کاربر را با درصد معینی با زمان مشخص افزایش دهیم؛ بنابراین، برای انجام این کار، به چند مورد مختلف نیاز است. این نیاز بستگی به استفاده دقیق از داده‌ها دارد. شما می‌توانید این مشکلات سازمانی را به‌عنوان یک هدف، برای دستیابی به آن شناسایی کنید. در تعیین اهداف خاص مانند افزایش درصد رشد، دقت کنید؛ چراکه همیشه دستیابی به آن سخت است؛ بنابراین شما باید بتوانید کالیبراسیون داخلی انجام دهید. چه زمانی است که ما واقعاً یک مسیر را کاوش کرده‌ایم، چیزها را بهبود بخشیده‌ایم، اما شاید به هدف مورد نظر نرسیده‌ایم، هدفی که در ابتدا فرض کردیم ممکن است بتوانیم انجام دهیم. مورد دیگر حل مشکلات داخلی است. بگویید امنیت داده در مکانی که می‌خواهید انجام شود کاملاً مناسب نیست. شاید این نوعی بدهی فنی باشد. به‌عنوان مثال یکی از تیم‌های شما می‌گوید که کار به پایان رسیده است اما هیچ‌کس نمی‌تواند آن را درک کند و هیچ مستندی درباره آن کد وجود ندارد؛ بنابراین، حل یک مشکل داخلی ممکن است پاک‌کردن کد، تمیزکردن مستندات و آماده‌سازی آن برای استفاده افراد دیگر باشد. این نوع اهداف معمولاً اهدافی هستند که مشخص‌تر و قابل تعریف بوده و دستیابی به آنها آسان است؛ بنابراین، تمایل به شناسایی مشکلات درونی وجود دارد که شما می‌خواهید آنها را حل کنید، زیرا آنها چیزهایی هستند که به‌راحتی در یک بازه زمانی مشخص قابل حل هستند. درحالی‌که اهداف سازمانی مبهم‌تر هستند و دستیابی به آنها کمی دشوارتر است، زیرا لزوماً داده‌ها همیشه دقیقاً همان کاری را که شما می‌خواهید انجام نمی‌دهند؛ بنابراین، معیارهای موفقیت می‌توانند بسیار مشخص یا مبهم باشند. آیا ما در مسیری پیش می‌رویم که می‌خواهیم تیم علم داده برای سازمان ایجاد نماییم؛ بنابراین، توانایی اولویت‌بندی آن معیارها، معیارهای مبهم در مقابل معیارهای بسیار خاص، یک وظیفه اساسی مدیرعامل یا مدیر تیم داده است. مورد دیگر مستندسازی و مدیریت افراد است. یکی از راه‌های انجام این کار شناسایی پروژه‌هایی است که افراد می‌توانند روی آنها کار کنند و هنگامی که آنها آن پروژه را به پایان رسانند، می‌توانید آن را به‌عنوان یک موفقیت در نظر بگیرید، اما اغلب با علم داده این کار آسان نیست زیرا معمولاً برای حل یک مشکل بزرگ علم داده برای یک سازمان، شما نیاز به ایجاد زیرساخت‌های مهندسی داده دارید، باید برخی داده‌ها را تجزیه و تحلیل کرده و

آزمایش‌هایی انجام دهید. هنگام این آزمایش‌ها شما نیاز به کشف کردن دارید و اینکه چگونه پیاده‌سازی آن را انجام دهید نیز مهم است. این موضوع شامل چندین نفر در چندین زمان مختلف در چندین سطح مختلف است. برای این کار لازم است به سؤالات مختلفی پاسخ داده شود. آیا به نظر می‌رسد که افراد دیگر آنچه را از تیم مهندسی داده شما نیاز دارند دریافت می‌نمایند؟ آیا دانشمند داده قادر به جمع‌آوری داده‌های مورد نیاز و... هست؟ آیا آنها تکنیک‌های جدیدی را یاد می‌گیرند؟ همان‌طور که سازمان رشد می‌کند و شما از نیاز به یک پایگاه داده به چندین پایگاه داده می‌رسید، آیا مهندس شما یاد می‌گیرد که چگونه آنها را مدیریت کند، چگونه زیرساخت‌ها را به روشی درست راه‌اندازی و نگهداری کند؟ و آیا راهی وجود دارد که آنها شخصاً بتوانند اهداف خود را بهبود بخشیده و به اهداف خود برسند؟ مورد بعدی این است که شما باید شکست‌ها را بررسی کنید. بدیهی است که این بخش سخت‌ترین قسمت است. یک مطلب این است که در صورت عدم موفقیت، مسئولیت آن را به‌عنوان مدیر تیم داده بپذیرید. به‌عنوان مثال، ایجاد یک فرضیه، شاید توسط برخی از مدیران سازمان بسیار معمول است. آنها پایین می‌آیند و از تیم علم داده می‌خواهند آزمایش‌هایی را انجام دهند. آنها این آزمایشات را انجام می‌دهند، شاید چیزی پیدا نکنند. مورد بعدی شناسایی مشکلات است. گاهی اوقات کمبود ارتباطی بین افراد خارجی که با تیم علم داده صحبت می‌کنند وجود دارد یا گاهی درون تیم علم داده بین مهندسی داده و دانشمندان داده ارتباط مؤثری وجود ندارد. بعضی اوقات اعضای تیم روی مشکلی که باید حل شود کار نمی‌کنند.

توانمندسازی دیگران برای استفاده از داده‌ها

فقط افراد تیم علم داده شما نیستند که می‌خواهند از داده‌ها استفاده و با آنها ارتباط برقرار کنند. بخشی از نقش شما به‌عنوان مدیر تیم داده ارائه فرصت آموزش نحوه استفاده از داده‌ها یا توانمندسازی افراد برای استفاده از داده‌ها در داخل سازمان است. یکی از راه‌های انجام این کار آموزش مبانی علم داده است. چندین روش مختلف وجود دارد که می‌توانید این کار را انجام دهید. یکی اشاره به منابع، جمع‌آوری و گردآوری منابع که به‌صورت آنلاین در دسترس هستند. اینکه آیا کلاس‌های آنلاین و یا مواد آنلاین مربوط به انواع چیزهای که آنها ممکن است نیاز داشته باشند می‌تواند در آموزش آنها مؤثر باشد یا نه نیز باید مورد بررسی قرار گیرد. یک راه واقعاً مفید دیگر داشتن گفتگوهای داخلی است که در دسترس همه است و اگر بخواهند می‌توانند به آنها مراجعه کنند. در

این محیط گفتگوی عمومی، افراد سؤالات مختلفی از یکدیگر می‌پرسند: چگونه داده‌ها را از پایگاه داده بیرون می‌کشید؟ چگونه داده‌ها را به روشی جذاب‌تر تجسم می‌کنید؟ یادگیری ماشینی را چگونه انجام می‌دهید؟ یا یادگیری ماشین چیست؟ نکته‌ای که در مورد تیم‌های داده بسیار دشوار است این است که هر تیم توانایی‌های مختلفی دارد. می‌توانید به‌نوعی جلسات آموزشی یا نوعی جلسات مشاوره درباره نحوه کار تیم داشته باشید. روش دیگر ساخت اسناد یا ساخت ابزارها یا ساخت محصولاتی است که می‌توانند با افراد دیگر به اشتراک گذاشته شوند. کسانی که لزوماً آموزش علم داده یا مهندسی داده ندیده‌اند، اما مایلند با داده‌ها تعامل داشته باشند. یک مثال متداول، نوعی داشبورد و مجموعه‌ای از تجسم‌ها یا مجموعه از طرح‌هاست. آنها به‌گونه‌ای ساخته شده‌اند که فرد می‌تواند بر روی برخی از دکمه‌ها کلیک کرده و با آن تجسم ارتباط برقرار کند. آنها می‌توانند سؤالات خود را بپرسند، فرضیه خود را داشته و با داده‌ها ارتباط برقرار کنند. حتی اگر فقط زیرمجموعه‌ای از داده‌ها باشد. راه دیگر این است که فقط یک تجسم تعاملی می‌تواند واقعاً جذاب باشد، می‌تواند راهی برای تعامل سایرین با داده‌ها باشد و به‌نوعی ببینید که چگونه انتخاب آنها کار می‌کند. همچنین می‌توانید داشبوردی را تنظیم کنید که از پایگاه داده شما به‌صورت خودکار داده‌هایی را واکنشی کند به‌گونه‌ای که همیشه جدیدترین داده‌ها را در اختیار افراد قرار دهد. اگر بتوانید آن را از این طریق خودکار کنید، رایج‌ترین سؤالات افراد در سایر بخش‌های سازمان پاسخ داده می‌شود و نیاز به تعریف سرفصل‌های جدید برای هر درخواست نمی‌باشد. این کار آنها را به سمت نوآوری و تمرکز بر ایده‌های جدید و کارهای جدید سوق می‌دهد که می‌توانند به‌خوبی انجام دهند. مورد دیگری که می‌تواند به‌عنوان روشی برای تعامل افراد با تیم داده به‌واقع مفید باشد، ارزیابی ایده است. غالباً افراد ایده‌هایی در مورد روش‌هایی که فکر می‌کنند ممکن است آنها را قادر سازد تا بتوانند از داده‌هایی استفاده نمایند که مسئله خاصی را حل کنند، خواهند داشت. اما معلوم می‌شود که بعضی اوقات این موارد غیرممکن است. ما داده‌های درست یا زیرساخت‌های پر قدرتی نداریم، مقیاس مناسبی نداریم که بتوانیم آن مشکل را حل کنیم، اما اگر شما بتوانید یک راه یا یک فرمت فراهم نمایید که در آن سایرین قادر باشند ایده‌هایی پیشنهاد کنند، می‌تواند بازخورد مناسبی از اقداماتی که انجام داده‌اید یا در حال انجام آنها هستید باشد. شما نمی‌خواهید فوراً هر ایده‌ای را که کسی مطرح می‌کند رد کنید که در این‌صورت دیگر هرگز ایده‌ای دریافت نخواهید کرد، اما ایده‌ای خوب است که به تنظیم انتظارات کمک کند.

مشکلات مشترک متقابل

هرچقدر هم که تیم علم داده را به خوبی جمع کرده باشید، ممکن است مشکلاتی وجود داشته باشد. اولین مورد عدم تعامل است. این مشکل خصوصاً اگر یک تیم علم داده داشته باشید که به تنهایی قصد انجام کارها را داشته باشد اتفاق می افتد و لزوماً یک تعهد منظم وجود ندارد که بتوان آن را در خود جای داد یا با یک واحد خارجی جلسه‌ای برگزار کرد. در صورت عدم وجود تعامل، باید مشکل را تشخیص دهید. مشکل می‌تواند از دانشمندان داده باشد. آنها نمی‌دانند چگونه با افراد ارتباط بگیرند. آنها نمی‌دانند که چه سؤالاتی را برای شناسایی جوانب پروژه بپرسند. گاهی اوقات این عدم ارتباط به این دلیل است که کاری برای انجام دادن وجود ندارد. این بدان دلیل است که در واقع مشکلی برای اینکه آنها حل کنند وجود ندارد. در این صورت خوب است که زمان دانشمند داده را برای یک پروژه متفاوت دوباره استفاده کنید. آنها با شخصی که با او کار می‌کنند ایمیل و ارتباط برقرار نمی‌کنند؛ بنابراین دوباره، این کمی به مقیاس پروژه بستگی خواهد داشت. اگر شما در حال ایجاد یک الگوریتم یادگیری ماشین سریع هستید قطعاً باید تعداد زیادی تعامل سریع برقرار شود. اگر در حال ایجاد زیرساخت بزرگ مقیاس‌پذیر برای مدیریت داده‌ها هستید، ممکن است کمی بیشتر طول بکشد تا در این باره با تیم رهبری ارتباط برقرار شود؛ بنابراین، اگر این عدم ارتباط وجود داشته باشد، دو گزینه وجود دارد. می‌توانید نقش فعال‌تری داشته باشید. شما می‌توانید شروع به مکالمه کنید، می‌توانید ایمیل‌هایی ارسال کنید که افراد را به هم متصل کند. این اتفاق اغلب اگر دانشمند داده واقعاً در یک تیم خارجی قرار داشته باشد می‌افتد. اگر این اتفاق بیفتد، غالباً دانشمند داده کمی احساس انزوا می‌کند و ممکن است برای تصمیم‌گیری قدرت نداشته باشد؛ بنابراین، واحد خارجی مسئول است. آنها باید در مورد بازاریابی تصمیماتی بگیرند؛ بنابراین اولین چیزی که شما باید انجام دهید این است که با هر دو طرف ارتباط بگیرید و سعی کنید مشکل ارتباط بین طرفین را حل نمایید. در این صورت، این شما هستید که باید شخص دیگر را متقاعد کنید و نوعی ابتکار عمل در توضیح اینکه چرا این مهم است، چرا استفاده از داده‌ها به شما کمک می‌کند داشته باشید. بعضی اوقات ساخت یک الگوریتم یادگیری ماشین واقعاً خوب، زمان زیادی می‌برد؛ بنابراین، بخشی از نقش شما نوعی واسطه بودن و توضیح‌دهنده بودن است. چرا این قدر طول کشید؟ و سپس پشتیبانی داخلی از تیم علم داده خود را ارائه دهید. اگر یک راه‌حل کاملاً خوب پیدا کنید، ممکن است برای انجام همه کارها کافی باشد؛ بنابراین، دانشمندان داده یا مهندسين داده، چون به آنچه

انجام می‌دهند افتخار می‌کنند، سعی می‌کنند الگوریتم یادگیری ماشین را برای ۳٪ آخرین دقت بهینه کنند. آخرین مشکل می‌تواند عدم درک باشد. این عدم درک می‌تواند به این دلیل باشد که تیم علم داده مشکلی را که روی آن کار می‌کنند به درستی درک نکرده‌اند. این اتفاق معمولاً هنگامی رخ می‌دهد که دانشمند داده در یک جدول مشاوره‌ای عمل می‌کند، هنگامی که آنها فقط یکبار مراجعه می‌کنند، افراد سازمان شما را می‌بینند یا یک سؤال مطرح می‌کنند، این آزمون t چگونه کار می‌کند یا این نیروی تصادفی چگونه کار می‌کند آنها ممکن است در واقع در درک مشکل و غیره خوب نباشند. موضوع دیگری که می‌تواند مطرح باشد این است که تیم بازاریابی نمی‌فهمد که علم داده چه کاری می‌تواند انجام دهد، یا تیم محصول نمی‌فهمد که علم داده چه کاری می‌تواند انجام دهد، یا هر همکار نمی‌فهمد که علم داده چه کاری می‌تواند انجام دهد؛ بنابراین نوع کمک به شما به این امر بستگی دارد که مطمئن شوید همه می‌دانند چه چیزی ممکن است و چه چیزی غیرممکن. سپس، شما می‌توانید این کار را انجام دهید، می‌توانید این فرایند را به چند روش مختلف مدیریت کنید. اطمینان حاصل نمایید که اگر شما یک تیم علم داده اختصاصی جداگانه دارید، آنها به‌طور منظم با سایر گروه‌ها که برای مدت‌زمان طولانی تقسیم کار شده‌اند تعامل برقرار می‌کنند. همچنین، اینکه به‌طور منظم هم در مورد قابلیت‌ها، هم در مورد تکنیک‌های جدید و هم در مورد آنچه اتفاق می‌افتد، سخنرانی‌هایی در سازمان انجام می‌شود. سپس داده‌ها را برای تیم درون سازمان معرفی نمایید. اطمینان حاصل کنید که همه می‌دانند به دنبال چه هستیم. این مشکلی است که تیم علم داده می‌توانند از عهده آن برآیند یا در واقع ما هیچ داده‌ای در این زمینه نداریم، اما جمع‌آوری آنها بسیار جالب است. آیا این چیزی است که ما واقعاً برای جمع‌آوری داده‌ها به آن نیاز داریم؟ این نوع از نقش‌هایی که می‌تواند کمک به ساخت یک درک درون سازمان از علم داده نماید بسیار مهم هستند؛ بنابراین اینها دو روش است که می‌توانید در به حداقل رساندن درگیری‌های خارجی در تیم علم داده کمک کنید.

مشکلات داخلی مشترک

برخی از مشکلات درون تیمی مربوط به شخصیت و تعاملات بین افراد است. برخی از آنها مربوط به نحوه کار دانشمندان داده و مهندسان داده است؛ بنابراین، این شما هستید که باید به‌نوعی محیطی را تنظیم نمایید که این نوع مشکلات به حداقل برسد و امکان این وجود داشته باشد که روند را با سرعت و به همان نوع دوستانه ادامه دهید؛ بنابراین اولین اتفاقی که اغلب می‌تواند رخ دهد این است که درگیری بین فردی به وجود آید. برای مثال،

برای بررسی کد، اگر یک نفر در حال بررسی کدهای فرد دیگری است، شما ممکن است برای ارائه فیدبک از روش‌های خاصی استفاده کنید. آیا آنها قصد دارند پیشنهادهایی را ارائه دهند؟ آیا آنها می‌خواهند بلافاصله تغییرات یا اصلاحاتی را در کد افراد ارائه دهند؟ آیا می‌خواهند این کار را به روشی انجام دهند که با کسی که در ابتدا کد را نوشته است ارتباط برقرار کنند یا خیر؟ وقتی افراد در حال مرور کد یکدیگر هستند و در جلسه گروهی، چگونه باید ارتباط برقرار کنند؟ آیا آنها قرار است انتقادی باشند؟ چگونه قرار است آنها بی‌طرف باشند؟ چگونه قرار است آنها با یکدیگر تعامل داشته باشند؟ اکنون به نظر می‌رسد که تنظیم این نوع موارد از قبل، این نوع سیاست‌ها، کمی عرفانی است، اما در پایان در واقع می‌تواند شما را از مشکلات زیادی نجات دهد؛ بنابراین، داشتن یک کد تماس برای جلوگیری از احساس صدمه شخصی یکی از مؤلفه‌های مهم این امر است. اگر فقط موفقیت را با استفاده از الگوریتم یادگیری ماشین تعریف کنید یا این نرم‌افزار هرگز اشکالی در آن وجود نداشته باشد یا سخت‌افزار هرگز خراب نشود، شکست زیادی وجود خواهد داشت، اما اگر شما موفقیت را شناسایی مشکلات، حل آنها و نه همیشه دریافت پاسخی که همه انتظار دارند، تعریف کنید، نوعی از فشار بر روی هرگونه تجزیه و تحلیل فردی را کاهش می‌دهید. این کار به افراد اجازه می‌دهد تا بی‌طرفانه عمل کنند و واقعاً به داده‌ها اجازه می‌دهد خود صحبت کنند. علاوه بر این سیاست شما نیاز به برقراری ارتباط آزاد دارید. به افراد اجازه دهید در مورد اینکه آیا این سیاست اعمال می‌شود مستقیماً با شما و با یکدیگر ارتباط برقرار نمایند و اطمینان حاصل کنید که آنها برای داشتن آن مکالمه با یکدیگر احساس قدرت می‌کنند. این واقعاً هم به معنای کلی و هم به طور خاص برای علم داده بسیار مهم است. این موضوع باید سیاستی باشد که انتقاد آشکار نسبت به تجزیه و تحلیل داده‌های دیگران یا زیرساخت داده‌های دیگران پذیرفته شود و مورد استقبال قرار گیرد، اما در عین حال، این انتقادات برای ایجاد احساس وحشتناک در سازمان ایجاد نمی‌شود یا برای آزار افراد مورد استفاده قرار نمی‌گیرد، بلکه فقط برای بهبود کل روند استفاده می‌شود. به طوری که اگر آن را به این ترتیب سازماندهی کنید، مردم در معرض انتقاد قرار می‌گیرند و اگر احساس نکنند که به کار و چشم‌انداز شغلی آنها آسیب می‌رساند، آنها مایل به ایجاد نوعی تغییر هستند. نکته دیگر این است که، در شرایط سخت که افراد نرمال رفتار نمی‌کنند، به هر دو دلیل عمومی یا به دلایل خاص علم داده، توضیح اینکه چه مرحله‌ای انجام می‌شود بسیار مفید است. بهتر است این حرف را در ابتدا بگویید، به جای اینکه در هنگام کار با یک تیم علم داده، سیاست‌هایی را در حین کار ابداع کنید. به‌ویژه هنگامی که با افراد متنوعی روبرو هستید که دارای تخصص‌های مختلف هستند و شاید انتظارات متفاوتی

در مورد آنچه اتفاق می‌افتد داشته باشند. مورد دیگری که باید با آن کنار بیایید یا مشکل دیگری که اغلب به‌صورت داخلی بروز می‌کند، کندی بین جلسات است؛ بنابراین به‌خصوص اگر جلسات منظم فردی، جلسات منظم گروهی ایجاد کرده باشید، امیدوارید که در هر یک از جلسات پیشرفت داشته باشید. شما امیدوارید که مشکلات حل شده یا حداقل برطرف شوند، اما گاهی اوقات این سرعت کاهش می‌یابد و همان‌طور که اشاره کردم، این اغلب می‌تواند به دلایل بسیار پیش‌پاافتاده‌ای باشد. یک اشکال کوچک در کد وجود دارد و شما فقط نمی‌توانید آن را ببینید؛ بنابراین، شما می‌خواهید شخص دیگری نگاهی به کد شما بیندازد و به شما کمک کند آن را برطرف نماید. کلید اصلی در اینجا شناسایی مسائل اصلی است. مشکل واقعی چیست؟ آیا این باعث می‌شود که فرد انگیزه شخصی نداشته باشد؟ آیا اینکه تعامل بین مهندسان داده‌ها و دانشمندان داده و یا دانشمندان داده و مردم خارجی، اتفاق نمی‌افتد نیست؟ این نوعی تعادل ظریف است، زیرا همان‌طور که اشاره کردم، دانشمندان داده و مهندسان داده برای حل واقعی این مشکلات چالش‌برانگیز فنی به دوره‌های طولانی و بدون وقفه نیاز دارند. همچنین نوعی بررسی انگیزه افراد نیز مفید است؛ بنابراین ممکن است شما یک دانشمند داده داشته باشید که مثلاً با تیم بازار کار کند، اما آنها واقعاً بازاریابی را دوست ندارند. آنها ترجیح می‌دهند با تیم دیگری درگیر شوند، ترجیح می‌دهند با محصول درگیر شوند و آنها دوست دارند در حال ساخت محصولات علمی داده باشند که می‌توانند در خارج از تیم استفاده شوند. با توجه به ماهیت ناامیدکننده علم داده و مهندسی داده برای افرادی که این کار را انجام می‌دهند، انگیزه بسیار حیاتی است؛ بنابراین شناسایی و رفع مشکلات مرتبط با انگیزش یکی از مشاغل کلیدی مدیر علوم داده است. ناامیدی چیزی است که همیشه با تیم‌های علم داده همراه است.

بسیاری از مدیران تیم‌های علوم داده به این دلیل مدیر می‌شوند که همکاران بسیار خوبی بودند و نه لزوماً به این دلیل که آنها مهارت یا آموزش برای رهبری یک تیم را دارند؛ اما مدیریت به خودی خود یک مهارت است و اتکا به تجربه شما به‌عنوان یک همکار موفق برای اینکه اطمینان حاصل کنید که قادر هستید ضمن حفظ آموخته‌ها، محصولات و نتایج ارزشمند به سازمان، توانایی حفظ و رشد استعدادهای عالی را داشته باشید، کافی نیست. دانشمندان عالی داده گزینه‌های شغلی متنوعی دارند و از مدیران بد برای مدت طولانی پیروی نمی‌کنند. اگر می‌خواهید دانشمندان بزرگ داده را حفظ کنید، بهتر است خود را متعهد کنید که یک مدیر عالی باشید.

چه کاری لازم است انجام دهید تا یک مدیر عالی شوید؟

البته در این باره مقالاتی نوشته شده است، از جمله از HBR^۱. اما از نظر من، چند زمینه برای کسانی که تیم‌های علوم داده را هدایت می‌کنند از اهمیت ویژه‌ای برخوردار است. مدیریت عالی به معنای مراقبت از اعضای تیم خود، ارتباط دادن کار آنها به کسب‌وکار و طراحی تیم‌های متنوع، انعطاف‌پذیر و با عملکرد بالا است.

اعتماد ایجاد کنید و صریح باشید

اعتماد، اصالت و وفاداری برای مدیریت خوب ضروری است. این امر به‌ویژه در علم داده صدق می‌کند که سردرگمی در نظم و انضباط و نقش آن در سازمان به این معناست که مدیر تیم مسئول محافظت کردن از اعضای تیم در برابر درخواست‌های غیرمعقول و توضیح نقش تیم برای بقیه اعضای سازمان است.

پشتیبانی از کارمندان به معنای دفاع کورکورانه از آنها به هر قیمتی نیست. این بدان معناست که مطمئن شوید آنها می‌دانند که شما برای مشارکت آنها ارزش قائل هستید. بهترین راه برای انجام این کار این است که اطمینان حاصل کنید اعضای تیم شما پروژه‌های جالبی برای کار دارند و آنها به‌وسیله پروژه‌های با الزامات مبهم یا جدول زمانی غیرواقعی بیش‌ازحد تحت فشار قرار نمی‌گیرند (که با توجه به تقاضای زیاد دانشمندان داده بسیار هم معمول است).

برای ایجاد اعتماد در طول زمان، باید روی صراحت در بیان و رفتار سرمایه‌گذاری کنید. دانشمندان داده افراد باهوشی هستند که در مورد نحوه دستیابی و پردازش اطلاعات آموزش دیده‌اند؛ بنابراین، شیوه اکتشافی این است که حدود ۲۰٪ بیشتر از آنچه فکر می‌کنید باید مستقیم و صریح باشید. با خوب و بد در طی کل مراحل استخدام، ایجاد آمادگی برای شروع کار، فعالیت‌های روزانه، تا بررسی عملکرد و بحث در مورد استراتژی تیم و سازمان شفاف باشید. این موضوع شاید کمی سخت باشد، اما برای موفقیت حیاتی است. لحظه‌ای که شما شروع به "خوب بودن" می‌کنید تا از یک مکالمه سخت و جدی جلوگیری کنید، شما و تیمتان شروع به شکست خوردن کرده‌اید.

1. <https://hbr.org/product/the-harvard-business-review-managers-handbook-the-17-skills-leaders-need-to-stand-out/10004H-HBK-ENG>

سرانجام، بازخورد باید ثابت و دوطرفه باشد و دانشمندان داده بزرگ از فاصله دور هم به سخنان نسنجیده و بیهوده پی خواهند برد. اگر می‌گویید که به‌صراحت اعتقاد دارید، اما حالت دفاعی می‌گیرید یا (بدتر!) درواقع بازخورد خوبی نمی‌دهید، در این‌صورت در بهترین حالت نیروهای شما به سمت ترک کارشان خواهند رفت.

◀ کار افراد را به مفاهیم کسب‌وکار وصل کنید

برای استفاده بیشتر از وقت دانشمند داده، آنها باید درک روشنی از اهداف تجاری پشت هر پروژه داشته باشند. اولویت‌بندی کار تیم شما در چارچوب استراتژی سازمانی گسترده‌تر از جمله مهمترین مشاغل است که یک مدیر علم داده دارد. متأسفانه، انجام آن همیشه آسان نیست. پروژه‌های علوم داده اغلب با سؤال از شخصی خارج از تیم شروع می‌شوند، اما اغلب سؤالی که فرد می‌پرسد دقیقاً همان چیزی نیست که درواقع می‌خواهند بدانند. مقدار زیادی از دانش مدیریت داده شامل بحث و تنظیم دقیق سؤالات مناسب از ذینفعان به‌منظور درک بهتر اطلاعاتی که آنها واقعاً می‌خواهند و نحوه استفاده از آن است. اجازه ندهید سؤالات یا درخواست‌ها برای تیمتان تبدیل به پروژه شود تا زمانی که دقیقاً بدانید ذینفعان چه می‌خواهند و چگونه از آن استفاده خواهند کرد. داشتن اهداف کاملاً واضح برای سؤالات مربوط به داده که پیش روی شما قرار می‌گیرد، یکی از مهمترین چیزهایی است که می‌توانید برای تیم خود فراهم کنید.

البته ذینفعان همیشه نمی‌توانند به‌تنهایی به این سؤالات پاسخ دهند. آنها ممکن است ایده روشنی از شکل ظاهری یک محصول نهایی علم داده (یا نحوه استفاده از آن) نداشته باشند. برای پرکردن این خلاء، اطمینان حاصل کنید که اعضای تیم علم داده به‌طور منظم به جلسات محصول و استراتژی دعوت می‌شوند. از این طریق آنها می‌توانند به‌جای اینکه صرفاً پاسخگوی سؤالات باشند، وارد روند خلاقیت شوند.

◀ تیم‌های عالی طراحی کنید

بسیاری از متخصصان در تلاشند تا به "جذاب‌ترین حرفه قرن بیست و یکم" برسند و بنابراین، به‌عنوان یک مدیر علوم داده، پیشنهادهای زیادی دریافت خواهید کرد و باید انتخاب کنید. از آن بهره ببرید تا راه‌های درست را برگزینید. به روند استخدام خود اهمیت بدهید.

یکی از بزرگترین زمینه‌های شکست افراد به‌عنوان مدیر، سبک سنگین کردن بین کوتاه‌مدت و بلندمدت است. به‌عنوان مثال، آسان است که فکر کنیم که وقت استخدام ندارید. این یک اشتباه بزرگ است. اگر وقت کافی برای یافتن اعضای عالی تیم و بررسی دقیق مصاحبه و مراحل پردازش خود ندارید تا اطمینان حاصل نمایید که افراد خوبی را در محل کار خود قرار داده‌اید، پس دیگر فرصتی برای مدیریت گزارش مستقیم جدید ندارید. ایجاد یک فرآیند عالی استخدام در بلندمدت نتیجه می‌دهد.

روند استخدام عالی چگونه است؟ یک چیز، فقط مهارت‌های فنی نیست. مهارت‌های اجتماعی مانند همدلی و ارتباط در علوم داده و رشته‌هایی که دانشمندان داده معمولاً از آنها پدید می‌آیند، ارزش کمی دارند، اما برای یک تیم بسیار مهم هستند. این کار را بخشی از استخدام خود در نظر بگیرید. به‌جای تمرکز بر اینکه آیا می‌توانید با یک داوطلب کنار بیایید، از خود بپرسید که آیا ذره‌بینی وجود دارد که این فرد جهان را ببیند و مرزهای حوزه دانش تیم را گسترش دهد و به همان اندازه که برای سایر ویژگی‌های فنی همچون توانایی و تخصص ارزش قائل هستید، به آن ارزش‌ها نیز پایبند باشید. به همین دلیل این نکته مهم را مدنظر داشته باشید که تنوع^۲ را در دستور کار خود قرار دهید که این امر شامل تنوع نظم دانشگاهی و تجربه شغلی و همچنین جنبه فکری و تجربه زندگی می‌باشد.

چند حوزه به‌ویژه برای علم داده مهم هستند. **اول**، فقط افراد ارشد را استخدام نکنید. نه تنها تقاضای آنها زیاد و گران است، بلکه کارمندان با تجربه کمتر "پذیرش ناآگاهی" دارند و می‌توانند سؤالات "گنگ" را بپرسند. البته این سؤالات درحقیقت گنگ نیستند، اما با فرض‌های معمول که متخصصان باتجربه‌تر از آن آگاهی دارند و کم‌تجربه‌ترها مطلع نیستند سازگار نمی‌باشد. شیفته یک روش خاص شدن برای انجام کارها و فراموش کردن طرح این سؤال که آیا رویکرد مورد علاقه هنوز بهترین راه‌حل برای یک کار جدید است یا خیر، می‌تواند مضر باشد.

دوم اینکه، دانشمندان داده از زمینه‌های تحصیلی گوناگونی می‌آیند: علوم کامپیوتر، فیزیک، آمار و بسیاری موارد دیگر. آنچه بیش از همه اهمیت دارد، داشتن ذهن خلاق همراه با مهارت‌های تفکر انتقادی درجه یک است. من یک عضو تیم دارم که زیست‌شناسی دریایی را مطالعه کرده و دارای اثرگذاری بالایی در تیم است، این تنوع تخصص بسیار ارزشمند است و کارایی آن به اثبات رسیده است.

2. <https://hbr.org/2016/11/why-diverse-teams-are-smarter>

سوم اینکه، مهم است به جای ساختن تیمی که همه در یک حوزه سرآمد باشند، افرادی را استخدام کنیم که نقاط قوت آنها مکمل یکدیگر است. یک فرد "مصورساز"، کسی که می‌تواند داستان‌ها را با داده بیان کند و یک متخصص حرفه‌ای مصورسازی که با هم کار می‌کنند می‌توانند با هم همکاری کنند تا چیزهایی تولید نمایند که هیچ‌کس به‌طور مستقل قادر به انجام آن نیست. برای به‌دست‌آوردن بیشترین استفاده از این مهارت‌های مکمل، اطمینان حاصل نمایید که تیم در واقع به‌عنوان یک گروه عمل و همکاری می‌کند. شما می‌خواهید که تیمتان با یکدیگر کار کنند و نه فقط در کنار هم. به‌طور مرتب اعضا را ملزم به خواندن کد و گزارش‌های یکدیگر و تقویت فعالیت‌های تیم با محوریت بحث‌های فنی کنید تا از این نوع تنوع بیشترین بهره را ببرید.

سرانجام، ساختن تیمی که نشان‌دهندهٔ افرادی باشد که داده‌های آنها را تجزیه و تحلیل می‌کنید نیز مهم است. این تنها راه برای اطمینان از این است که شما تیم انعطاف‌پذیری دارید که سؤالات بهتری می‌پرسد و از روزنه دید وسیع‌تری از جنبه فکری برخوردار است که می‌توانید از آنها این سؤالات را بپرسید. به این ترتیب، نقاط کور هر فرد توسط تجربیات گذشته و مجموعه مهارت‌های دیگران پوشش داده می‌شود.

چه موقع باید متخصص شد

زمانی که یک تیم علوم داده تازه شروع به کار می‌کند، همه افراد حاضر در آن "در همه کارها حضور دارند" و انواع مختلف زیادی از پروژه‌های علم داده را انجام می‌دهند. این درست مثل وقتی است که یک نفر به یک استارت‌آپ بپیوندد، اما همان‌طور که تیم شما بالغ می‌شود و ارزش خود را ثابت می‌کند، تشخیص می‌دهید که نقش‌های بیشتر تعریف می‌شوند و برخی فعالیت‌ها به تیم‌های دیگر (زیرساخت، عملیات و غیره) منتقل می‌شوند. تخصص فقط زمانی کار می‌کند که الزامات کاملاً مشخص و واضحی برای جبران تأخیرهای هماهنگی و هزینه‌های مرتبط با همکاری چندین تیم وجود داشته باشد. یافتن دانشمندان داده "Full stack" بسیار دشوار است، اما می‌توان دانشمندان داده "partial stack" هوشمند را پیدا کرد که بتوانند با کمی مربیگری اختصاصی یاد بگیرند که چگونه یک مسئله را به‌طور مناسب چارچوب‌بندی کنند، یک پروژه کوچک را مدیریت نموده، توسعه دهند و یک مدل را آموزش دهند، با API ها ادغام کرده و به سمت تولید سوق دهند.

اگر وظیفه خود را به عنوان مدیر به درستی انجام داده باشید، این روند پیشرفت نسبتاً آرام پیش می‌رود. شما در استخدام خود سختگیر بوده و یک تیم عالی با مهارت‌های متعادل ایجاد کرده‌اید. کارمندان شما به شما اعتماد خواهند کرد و آنها درک خواهند کرد که چگونه تغییرات از سازمان و اهداف آن حمایت می‌کند.

آیا شما آماده Machine Learning هستید؟

مسئله‌ای را بیان کنید که به حل آن نیاز دارید

زمانی که یک شرح مشکل روشن وجود داشته باشد یادگیری ماشینی بهترین نتیجه را می‌دهد. تعریف مسئله باید شامل اقداماتی باشد که امکان‌پذیر است و/یا نتایج قابل اندازه‌گیری است. حتی بهتر؟ اگر مسئله به‌وضوح به چالش‌های عملیاتی یا اهداف استراتژیک عالی مرتبط باشد. از آنجاکه یادگیری ماشینی به زمان و داده نیاز دارد، ارزیابی حیاتی از اینکه آیا مدل‌های تحلیلی موجود / رویکردها یا راه‌حل‌های جایگزین ممکن است اعمال شود نیز مناسب است. این تضمین می‌کند که مقدار بالقوه متناسب با ورودی است.

توجه داشته باشید نقاط تصمیم‌گیری معمول که حجم بالایی دارند، به پاسخ بی‌درنگ یا سریع نیاز دارند و یا به ورودی‌های بسیار متغیری وابسته هستند، این نقاط کاندیدهای خوبی برای یادگیری ماشینی هستند.

یک ذهنیت تجربی ایجاد کنید

یادگیری ماشینی یک فرایند تکراری و آزمایشی است؛ اگرچه الگوریتم‌های اصلی به‌طور فزاینده‌ای پرکاربرد می‌شوند، اما هر پروژه باید براساس زمینه و داده‌های تجاری سفارشی شود. مانند هر آزمایش خوب، نادرست بودن برخی فرضیه‌ها ثابت خواهد شد. ممکن است نیاز به تهیه یا ایجاد داده‌های جدید باشد. یا بیان مسئله براساس آنچه یافت شده دوباره اصلاح شده است. در نتیجه، تصمیم‌گیرندگان و اعضای تیم به‌طور یکسان باید یک ذهنیت آزمون و یادگیری را برای موفقیت در یادگیری ماشینی اتخاذ نمایند. برای ارزیابی سریع پیشرفت، از یک فرایند باز شونده، تکرارشونده استفاده کنید که قابلیت و چابکی را فراهم نماید تا تعیین کنید آیا روش جایگزین تضمین شده است یا چه زمانی کافی خواهد بود.

◀ یک تیم علم داده همکار را به خدمت بگیرید

تخصص ML یک الزام است. به همان اندازه مهم است که یک مدل تیمی پویا که متخصصان متنوعی را در زمینه تجارت، داده و تخصص فنی درگیر می‌کند. این شامل کارشناسان داده است که می‌توانند دارایی‌های داده‌های مورد نیاز را ارزیابی کرده و در اختیار داشته باشند، متخصصان بازرگانی برای تهیه دانش زمینه، ارزیابی پیامدهای (تجاری، اجتماعی، اخلاقی) اقدامات پیشنهادی یا خدمات جدید یا خدمات و پرسنل فناوری اطلاعاتی که اکوسیستم‌های فنی را مستقر و نگهداری می‌کنند.

◀ یک استراتژی و اکوسیستم داده‌های قوی ایجاد کنید

Machine learning بر روی داده‌های بسیار زیاد اجرا می‌شود. ایجاد یک فرآیند داده برای شناسایی مؤثر، دستیابی (یا ایجاد)، تهیه و دسترسی به داده‌های با کیفیت بالا و دارایی‌های اطلاعاتی بسیار مهم است. برای این منظور، سیاست‌های حاکمیتی و اکوسیستم داده باید از محیط‌های اکتشافی (که اغلب به آنها سندباکس می‌گویند) و همچنین از محیط‌های تولیدی پشتیبانی کنند. این امر مستلزم رویکردی چندلایه برای توازن در دسترسی و چابکی بدون تضعیف امنیت، حریم خصوصی یا کیفیت است. معرفی منابع داده‌ای غیرعادی (بزرگ) شامل متن، ساختار صوتی، تصاویر و غیره بدون ساختار نیز ممکن است به قابلیت‌های جدید مدیریت داده نیاز داشته باشد.

تحمل ریسک سازمان را از توافق در مورد معیارهای "به‌اندازه کافی خوب" گرفته تا درک چگونگی اعتبارسنجی و توسعه مدل‌ها ارزیابی نمایید. یادگیری ماشین اغلب رویکردهای سنتی تضمین کیفیت و مدیریت ریسک را به چالش می‌کشد. چرا؟ در برخی موارد، چرخ‌های تمرین یا در این حالت، داده‌های آموزش باید از طریق آنها انجام شود. اعتبار واقعی از آزمایش عملکرد الگوریتم‌های یادگیری ماشینی در برابر داده‌های جدید حاصل می‌شود.

اغلب، این موضوع مستلزم قراردادن سیستم در تمرین عملیاتی است. ممکن است از انجام "تست A / B" در تولید برای تأیید مدل ایجاد رفتارهای مطلوب مشتری گرفته تا راندن اتومبیل در جاده با ناظر انسانی آماده برای گرفتن چرخ در یک بحران باشد.

متعهد شوید که با فرایندهای تجاری مستقر تطبیق دارد

چه اتوماتیک کردن یک نقطه تصمیم‌گیری فعلی و چه امکان ایجاد یک محصول یا خدمات جدید خالص، ML می‌تواند مغل باشد. ارزیابی پیامدهای بالقوه در فرآیندها، عملکردها و نقش‌های موجود تجاری کلیدی است. این بدان معنا نیست که قبل از شروع، کل تغییر را معماری کنید، اما یک چک کردن سریع می‌تواند پتانسیل تمرینات پرهزینه را کاهش دهد. به‌منظور برنامه‌ریزی برای این موضوع، با طرح این سؤال شروع کنید: «اگر به این سؤال پاسخ دهیم یا این فرضیه را ارائه دهیم، چه کاری می‌توانیم با این اطلاعات انجام دهیم؟ این امر چگونه ممکن است بر فرآیندهای موجود تأثیر بگذارد؟ آیا ما مایل و قادر به ایجاد تغییرات لازم هستیم؟»

متعهد به پذیرش شیوه‌های جدید فناوری اطلاعات باشید

به‌محض استقرار، مدل‌سازی تکراری و تنظیم مدل یادگیری ماشینی باید ادامه یابد. وزن مورد نیاز برای به‌روزرسانی‌ها قابل پیش‌بینی نیست و با پنجره‌های استقرار برنامه‌ریزی شده سنتی مطابقت ندارد. در نتیجه، استفاده از یادگیری ماشینی نسبت به روش‌های سنتی توسعه فناوری اطلاعات، به مدل‌های مختلف QA و استقرار، مجموعه مهارت‌ها و سطح خدمات نیاز دارد.

نحوه استخدام دانشمند داده

آیا رقبای شما محصولات مربوط به یادگیری ماشینی را می‌سازند و مدیرعامل شما احساس می‌کند که باید همین کار را انجام دهید؟ حالا وقت آن رسیده که اولین متخصص علوم داده خود را استخدام کنید و به بازی برگردید. با این حال، به‌زودی متوجه خواهید شد که استخدام دانشمند داده در مقایسه با یک توسعه‌دهنده نرم‌افزار بی‌نهایت دشوارتر است، شاید به سه دلیل زیر باشد:

(۱) نوشتن شرح وظیفه برای نقش دانشمند داده دشوار است،

(۲) تعداد زیادی از دانشمندان داده مایل به درخواست‌دادن هستند، اما تعداد کمی از آنها تجربه مورد نیاز را دارند،

(۳) تعداد کمی از استانداردها و معیارهای صنعت در دسترس هستند.

◀ ابتدا از آنها بخواهید یک نسخه اولیه بسازند

علم داده و موضوعات نزدیک آن مانند یادگیری ماشینی واقعاً نوید تغییر نتیجه کار شما را می‌دهند. با این حال، مهم است به یاد داشته باشید که اولین محصول داده شما احتمالاً باید چیزی بسیار ساده‌تر باشد، به عنوان مثال، یک داشبورد هوش تجاری که می‌تواند با استفاده از KPI های کلیدی، سلامت کلی کسب‌وکار را کنترل نماید. ساختن چیزی ساده و در عین حال مؤثر به شما کمک می‌کند تا مدیریت بالاتر را به سرمایه‌گذاری بیشتر در تیم علوم داده خود تشویق کرده و به پروژه‌های داده چالش‌برانگیز بپردازید.

همچنین، مهم است که به یاد داشته باشید پیش‌بینی نتایج برای کسب‌وکار با استفاده از ابزارهای پیشرفته مانند یادگیری ماشینی فقط پس از تکرارهای متعدد قابل اعتماد می‌شود و ساخت آن اغلب ماه‌ها طول می‌کشد، اگر نه شاید سال‌ها به طول بینجامد.

◀ از آنها بخواهید داده‌های شما را بازبینی کنند

داده‌های تولیدشده یا تجمیع شده توسط تجارت شما می‌تواند از متون گرفته تا فایل‌های صوتی، تصاویر و حتی فیلم‌ها باشد. اگر سازمان شما سوابق پزشکی یا مالی را کنترل می‌کند، اغلب اقدامات امنیتی و استانداردهای صنعتی دیگری برای ذخیره و بازبازی این مجموعه داده‌ها وجود دارد. اطمینان حاصل کنید که اولین استخدام شما تجربه قبلی در مدیریت این نوع داده‌ها را دارد.

◀ از آنها بخواهید Data-pipeline داده بسازند

داده‌های تولیدشده توسط مشاغل اصلی شما اغلب به‌اندازه کافی نیستند. به عنوان مثال، بخش‌های بزرگی از داده‌ها را می‌توان در یک فرمت غیردیجیتالی ذخیره کرد و یا می‌تواند زیر قفل و کلید خصوصی نگهداری شود. همچنین، داده‌ها با گذشت زمان ممکن است خراب شوند و امکان دارد داده‌های جدید نوع داده متفاوتی داشته باشند. اطمینان حاصل کنید که اولین استخدام شما در جمع‌آوری داده‌ها از چندین منبع داده دارای تجربه است و در ساخت خطوط لوله داده نیز تخصص دارد. این به شما کمک می‌کند تا یک اسکلت عملیاتی برای کسب‌وکار خود ایجاد نمایید که می‌تواند برای تولید یک یا چند محصول اطلاعاتی به کار رود. یک خط لوله داده ساخته شده

کارآمد همچنین به وسیله اطمینان دادن از آنکه شما می‌توانید دسترسی محدود به کارمندان جدید بدهید، امنیت داده‌ها را بهبود می‌بخشد.

GETTY

◀ به تخصص‌ها در استخدام توجه کنید

علم داده اغلب در صنعت به عنوان یک چتر برای کل مجموعه مهارت‌های دارای همپوشانی استفاده می‌شود که از تهیه داده‌ها گرفته تا هوش مصنوعی و تجسم داده‌ها را شامل می‌شود. از این رو ابزارهای نرم‌افزاری مورد استفاده توسط انواع مختلف دانشمندان به مجموعه مهارت آنها بستگی دارد. به عنوان مثال، یک برنامه‌نویس با سابقه کامپیوتر چندین زبان برنامه‌نویسی مانند Python، R، ++C و برخی از چارچوب‌های داده‌های بزرگ مانند Hadoop، Spark، NoSQL را می‌داند. در حالی که یک دانشمند می‌تواند تعداد کمی از زبان‌های برنامه‌نویسی را بداند، اما ممکن است تکنیک‌های پیشرفته مانند mapreduce و یادگیری ماشینی را بداند.

بنابراین، عاقلانه است که انتظار داشته باشید اولین استخدام شما مهندس "کارشناسی ارشد" باشد، نه "یک متخصص تجاری". همچنین دانستن خلاء موجود در تخصص‌های استخدام اول به شما کمک می‌کند تا یک نقشه راه استخدام برای تیم علوم داده خود برنامه‌ریزی کنید. معمولاً استخدام یک مهندس داده ابتدا برای ساخت خط لوله داده‌ها و سپس استخدام یک متخصص هوش تجاری یا آماری برای بهینه‌سازی تولید، عاقلانه است.

◀ اطمینان حاصل کنید که آنها نمونه کارهای اصلی دارند

در حالی که برای یک موقعیت شغلی مصاحبه می‌کنید، بسیاری از نامزدها نمونه کارهای خارق‌العاده‌ای را که از قبل ساخته‌اند به شما نشان می‌دهند. این نمونه کارها ممکن است شامل چیزهایی مانند برنامه‌ای باشد که قادر است به صورت خودکار کلمات یا ربات گپ را که می‌تواند به عنوان دستیار مجازی کار کند، شناسایی نماید. لازم به یادآوری است که بسیاری از این پروژه‌ها در مجامع علم داده منبع باز موجود هستند. در حین مصاحبه برای یک موقعیت شغلی، اطمینان حاصل کنید که داوطلب در پروژه سهیم بوده است.

هنگام استخدام اولین دانشمند داده خود، شعار شما باید "افراد با استعداد را استخدام کن، به آنها آموزش مهارت‌های فنی بده" باشد. اغلب بهترین استخدام، استفاده از یک برنامه‌نویس نرم‌افزار باتجربه است که

علاقه‌مند به ایجاد حرکت شغلی در علم داده‌ها است. از آنجاکه علم داده یک زمینه نوظهور است، بهتر است آنها را تشویق کنید تا به‌طور مداوم از انجمن‌ها، MOOCها یا دانشگاه‌هایی که دوره‌های مربوطه را ارائه می‌دهند، یاد بگیرند.

◀ اطمینان حاصل کنید که صرفاً دانشگاهی نیستند

علم داده شاخه‌ای از آمار است. در نتیجه، بسیاری از محققان دانشگاهی از رشته‌هایی مانند اقتصاد، فیزیک، ریاضیات، علوم کامپیوتر و مهندسی بر روی علوم داده جهش کرده‌اند تا بتوانند به سمت صنعت سوق پیدا کنند؛ اگرچه برخی از بهترین دانشمندان داده از دانشگاه‌ها و آزمایشگاه‌های تحقیقاتی وابسته آمده‌اند، اما لازم به یادآوری است که این متخصصان باید تنها برای بهینه‌سازی و نه ساخت محصول داده شما استخدام شوند.

علم داده ترکیبی از تجارت، مهندسی نرم‌افزار و آمار است. جاش ویلز یک دانشمند داده را چنین توصیف می‌کند: "شخصی که در آمار بهتر از هر مهندس نرم‌افزار و در مهندسی نرم‌افزار از هر آمارشناسی بهتر است." بنابراین، اولین استخدام شما باید فردی با چندین سال تجربه مهندسی نرم‌افزار و همچنین برخی از تجربه‌های مدیریت باشد. این کار تضمین می‌کند که او می‌تواند کارها را اولویت‌بندی کرده و بر روی کسب‌وکار شما تأثیر بگذارد.

فصل دوم

استخدام تیم داده

قبل از استخدام دانشمند داده چه کاری باید انجام دهید؟ هوش مصنوعی نوید رشد نمایی و ارتقاء تجارت شما را می‌دهد. جای تعجب نیست که در مورد استفاده از هوش مصنوعی (AI) هیجان زیادی وجود دارد. بسیاری از شرکت‌ها برای استخدام اولین دانشمند داده یا ساخت یک تیم علم داده درست در همان لحظه، عجله دارند. اشتیاق آنها قابل درک است، زیرا آنها می‌خواهند با داده‌ها نوآوری کنند و در بازار آزمایش و خطا نمایند. در اینجا چند توصیه وجود دارد که شما می‌توانید قبل از سرمایه‌گذاری در تیم داده در شرکت خود انجام دهید:

۱- درک روشنی از اینکه چرا می‌خواهید یک دانشمند داده استخدام کنید، داشته باشید

می‌توانید با شناسایی مشکلات و فرصت‌های تجاری که می‌خواهید آنها را برطرف کنید، شروع نمایید. شما لزوماً نیازی به داشتن مقادیر زیادی داده ندارید، اما قطعاً به برخی از داده‌ها که مربوط به مشکلات تجاری شناسایی شده هستند نیاز دارید. به‌عنوان مثال، یک صفحه گسترده با چند صد هزار ردیف با مشخصه‌های داده مناسب که اکثریت توزیع جمعیت را پوشش می‌دهد، ممکن است کافی باشد.

اگر برای شروع یک "مشکل تجاری" بسیار خطرناک به نظر می‌رسد، می‌توانید با لیست کوتاه و اولویت‌بندی **use case** های ساده‌تر شروع کنید مانند:

تشخیص و کاهش تقلب مبتنی بر صدا در مرکز مراقبت از مشتری.

توصیه‌های محصول برای یک سایت تجارت الکترونیکی.

پیش‌بینی ریزش برای یک شرکت B2B SaaS.

اجازه دهید از دانشمند داده بخواهیم داده‌های موجود را بررسی کند و سپس یک ارزش تجاری تولید کند، این یک استراتژی بد است.

در صورت تمایل، ابتدا می‌توانید یک مشاور مستقل استخدام کنید یا با یک ارائه‌دهنده خدمات همکاری نمایید. قبل از اینکه بخواهید خودتان یک تیم بسازید، باید ابتدا فرضیه‌ها را آزمایش کنید.

۲- پروفایل‌های علم داده را بفهمید و بدانید چه شخصیتی را می‌خواهید استخدام کنید

دانشمندان داده از زمینه‌های مختلف مانند بازاریاب، طراح، مدیر محصول یا هر حرفه دیگر می‌آیند. برخی از آنها در ساخت مدل‌های یادگیری ماشینی مهارت دارند؛ در حالی که دیگران در تجزیه و تحلیل و بصری‌سازی قوی هستند. برخی فقط در بینایی ماشین و برخی دیگر فقط در پردازش زبان طبیعی (NLP) کار کرده‌اند. برخی کلی‌گرا هستند در حالی که برخی دیگر متخصص‌اند.

دانشمندان داده مرتبط با موضوعات خود را استخدام نمایید

نیازی نیست که یک PhD را استخدام کنید و یا از Kaggle به‌عنوان ابزار اولیه خود به‌منظور ارزیابی فردی استفاده نمایید. دانشمندان داده خوبی با مدرک دکتری که هرگز در رقابت‌های Kaggle شرکت نکرده‌اند، وجود دارند.

هنگامی که شما دکترای بدون تجربه تجاری را استخدام می‌کنید، تعصب در رویکرد سبک دانشگاهی وجود دارد تا تفکر تجاری. داشتن مدرک دکتری در یک رشته مرتبط و مشارکت در ابتکارات اجتماعی مانند Kaggle چیز خوبی است، اما نباید معیارهای اجباری استخدام باشند. به‌استثنای تعداد کمی از مؤسسات دانشگاهی که با شرکت‌ها همکاری می‌کنند، دانشگاهیان زمینه حل مشکلات تجاری را ندارند. یک فرد آکادمیک و بدون تجربه تجاری ممکن است در زمانی که تنها چیزی که نیاز دارید این است که او یک نمونه اولیه سریع برای رفتن به مرحله بعد را آماده کند، اقدام به صرف وقت در حل ناهنجاری‌های داده کرده و انورمالی‌ها را شناسایی نماید.

اولین استخدام شما باید دانش کافی در مورد رشته Data Science به‌طورکلی و برخی دانش‌های تجاری داشته باشد. ممکن است ابتدا به یک "مهندس داده" نیاز داشته باشید که در مهندسی نرم‌افزار، ذخیره‌سازی داده‌ها، استخراج و مدیریت به‌جای ریاضیات و آمار مانند دانشمند داده، قدرتمند باشد.

۳- کاری را که می‌خواهید انجام دهند مشخص کنید

دانشمندان داده وظیفه اصلی حل مشکلات پیچیده به‌وسیله استفاده از تجربه و تخصص خود را، در حالی که از تحقیقات دانشگاهی استفاده کرده و به آنها کمک می‌کند برعهده دارند. از مواردی که دانش زیادی ارائه نمی‌دهند یا اطلاعات عمومی که توجه زیادی را به خود جلب نمی‌کنند دوری کنید. به‌عنوان مثال به‌جای صحبت کردن در مورد شرکت بیمه میلیارد تومانی که ممکن است MBA ها را جذب کند، شما باید در مورد استفاده از مواردی مانند بازرسی خودکار آسیب خودرو، افزایش دقت و صحت ادعاهای جعلی و غیره صحبت کنید.

به‌دنبال کسی باشید که همه چیز را بداند، علم داده معادل جستجوی یک اسب تک‌شاخ، در تجارت، تکنولوژی، ریاضی، مدل‌سازی، برنامه‌نویسی و آمار است. شما باید مهارت‌های کلی را حساب کنید، نه فقط مهارت‌های سخت فنی را. آنها باید به‌عنوان بخشی از تیم شما کار کنند و خلاء مهارت‌هایی را که در حال حاضر در تیم شما از دست رفته است، پر نمایند.

۴- یک فرهنگ مبتنی بر داده ایجاد کنید که به دانشمندان داده کمک می‌کند

این نکته بسیار مهم است که به داخل نگاه کنید و ببینید که آیا فرهنگ شرکت شما برای دانشمندان داده مفید است یا اینکه علیه آنهاست. برخی از این موارد شامل نداشتن هیچ شخص فنی یا تجاری با درک اساسی از Data Science یا نداشتن هیچ‌گونه اطلاعات و دانش فنی می‌باشد.

اگر شرکت تجربه کار با داده‌ها را نداشته باشد و تصمیمات مبتنی بر داده را اتخاذ نکند، دانشمندان داده برای تبدیل کار خود به تجارت و ارزش مشتری تلاش خواهند کرد.

همچنین باید بودجه کلی برای پشتیبانی از یک دانشمند داده و یا یک تیم را تهیه کنید. ممکن است استارت‌آپ‌های هوش مصنوعی با صندوق سرمایه‌گذاری با سرمایه VC نیازی به نگرانی در مورد این موضوع نداشته باشند. برای شرکت‌ها، با توجه به مقیاس داده‌ها و موردهای کاربردی که می‌خواهید برطرف کنید، این هزینه‌ها ممکن است به صدها مورد برسد که می‌تواند تلاش‌های شما را به بن‌بست برساند. نرم‌افزار منبع آزاد و ابزار Data Science برخی از این هزینه‌ها را کاهش می‌دهند.

شرکت‌هایی که در تلاش‌های خود برای هوش مصنوعی شکست می‌خورند، سرانجام دانشمندان داده را به دلیل مهارت‌های ضعیف مقصر می‌دانند، در حالی که در حقیقت، ممکن است این آنها باشند که در ارائه حمایت مناسب، محیط، بودجه و تیم مناسب موفق نبودند.

۵- تصمیم بگیرید که در سازمان شما کجا و چگونه قرار می‌گیرند

برای اداره، پشتیبانی و هدایت ابتکارات علم داده، شما به یک رهبر داخلی از محصول، تکنولوژی یا عملکرد تجاری نیاز دارید. تیم علم داده باید برای ایجاد ردیابی، پاسخگویی و بازخورد مداوم از بقیه مشاغل، آمادگی داشته باشد.

پرسش اجتناب‌ناپذیر گروه متمرکز در برابر گروه غیرمتمرکز دیر یا زود مطرح می‌شود. عواملی مانند فرآیندهای فعلی شما، مدل سازمان، بلوغ داده و اندازه تیم وجود دارد که تصمیم را به نفع یک رویکرد نسبت به یک رویکرد دیگر تحت تأثیر قرار می‌دهند. برای یک استارت‌آپ AI، ممکن است دانشمندان داده بسیاری را در یک تیم پیدا کنید. برای غول‌های فن‌آوری مانند گوگل و فیسبوک، ممکن است شما برای محصولات سنگین داده در هر تیم محصول یک دانشمند داده پیدا کنید. برای سازمان‌های چابک، امکان دارد یک گروه علم داده با تمرکز بر محصول AI/ML پیدا کنید.

در هنگام ایجاد عملکرد علم داده، با هر مدلی که تصمیم به استفاده از آن دارید، انعطاف‌پذیری را برای آزمایش و تکامل تیم حفظ کنید.

۶- یک پشتیبان یا مدیر محصول اختصاصی داشته باشید

برخی از رهبران تجاری بر این باورند که یک دانشمند داده اکثر کار خود را در انزوا انجام می‌دهد (تصویر یک دانشمند در اتاقی که روی تخته‌سیاه کار می‌کند). این امر نمی‌تواند دور از حقیقت باشد. دانشمندان داده برای تحویل پروژه‌ها به تیمی از مهندسان، سهامداران تجاری، مدیران پروژه و سایر اعضای تیم احتیاج به ارتباط دارند. شما باید یک مدیر اختصاصی یا پشتیبانی‌کننده محصول را به مورد کاربردی که دانشمند داده در حال کار بر روی آن است اختصاص دهید. مدیران محصول می‌توانند از کارهای بی‌شماری مانند جمع‌آوری نیازها، بینش مشتری،

تجزیه و تحلیل داده‌ها، مدل عملیاتی، دستورالعمل‌های قانونی، ارائه یک محصول، جمع کردن منابع و تولید کالاها مراقبت کنند. دانشمندان داده باید در تعدادی از این فعالیت‌ها مشارکت داشته باشد، اما اگر پشتیبانی نشود، ارزش زیادی از تلاش‌های او کسب نخواهد کرد.

تعداد کمی از دانشمندان داده با سابقه از کار با مدیران محصولات و Scrum Master ها استقبال می‌کنند و اظهار می‌دارند:

مدیران محصول می‌خواهند اعتبار زیادی را بدون افزودن هیچ‌گونه ارزشی مطالبه کنند.

این جمله می‌تواند برای برخی درست باشد، اما نه برای همه. اکثریت مدیران محصول هرگز قبلاً پروژه‌های علم داده را تحویل نداده‌اند. آنها در حال یادگیری چیزهایی هستند که زمینه و هنر مدیریت محصول داده محور در آنها تأثیرگذار است.

مدیران محصول می‌توانند هماهنگی و انتظارات را در سراسر کسب و کار تنظیم کرده و برای از بین بردن موانع موجود بسیار ارزشمند باشند. نتیجه‌گیری این است که باید یک تنظیم نقش مشخص بین مدیر محصول و یک دانشمند داده و آنچه هر یک از اعضاء در حال ارائه به میز کار این همکاری هستند، وجود داشته باشد.

علم داده یک زمینه در حال ظهور است و در حال حاضر نقش‌ها و همچنین صلاحیت‌ها مشخص نیستند. با توجه به پیچیدگی در این زمینه و کمبود بالقوه تخصص تجزیه و تحلیل در شرکت‌هایی که می‌خواهند دانشمند داده یا تیمی از دانشمندان داده را استخدام کنند، کار ایجاد یک تیم تجزیه و تحلیل یا استخدام اولین دانشمند داده شرکت می‌تواند سخت باشد. با این حال، با مروری اجمالی بر انواع دانشمندان داده و سؤالات نمونه برای ارزیابی هر نوع، مدیران استخدام می‌توانند مشخصات مناسب‌تری برای استخدام‌کنندگان فراهم نموده و داوطلبان را در مورد مهارت‌های احتمالی لازم برای پرکردن نقش ارزیابی کنند.

دانشمندان داده به‌طور معمول در ۳ زمینه اصلی مهارت دارند: ریاضیات / آمار / یادگیری ماشین، برنامه‌نویسی / مهندسی نرم‌افزار و تخصص در صنعتی که در آن به دنبال استخدام هستند.