

دانشگاه صنعتی امیرکبیر  
( پلی تکنیک تهران )

دانشکده مهندسی کامپیوتر

گزارش درس روش پژوهش و ارائه

# آشنایی با هوش تجاری و عملکرد آن

نگارش:

محمد جواد رضوانیان

استاد راهنما:

دکتر رضا صفابخش

اردیبهشت ۱۴۰۱



## سپاسگزاری

در ابتدا از خداوند منان سپاسگزارم که همواره و در هر لحظه پشتیبان بندگان خود بوده و هر کجا که آنها در بن بست قرار گرفته‌اند، آنها را راهنمایی کرده. از خدایی سپاسگزارم که توفیق شاگردی استاد گرانقدر و بزرگوار، جناب آقای دکتر رضا صفابخش را به من عطا کرد تا با راهنمایی‌ها و دلسوزی‌های این استاد بزرگوار امکان تهیه این گزارش میسر شود.

محمد جواد رضوانیان

اردیبهشت ۰۱

## چکیده

هوش تجاری<sup>۱</sup> یکی از پرکاربردترین ابزارهای امروز دنیای کسب و کار است. ابزاری که اثبات کرد که می‌توان از کوچک‌ترین رخدادهای حوزه کاری خود نیز، اطلاعات مفید استخراج کرد؛ اطلاعاتی مثل تعداد کلیک‌های کاربران، مقدار زمان صرف شده در هر صفحه، رنگ‌های محبوب کاربران و مثال‌هایی از این قبیل. ابزاری بسیار مفید که با پیوند زدن مفاهیم علوم کامپیوتری به صنعت، فصل جدیدی از پیوند فناوری‌های کامپیوتری و دنیای کسب و کار را رقم زد.

مسئله اصلی استفاده از فناوری‌های هوش تجاری، هزینه بسیار بالای استفاده از آن است؛ به طور معمول کسب و کارهای نوپا<sup>۲</sup> هنگامی که شروع به فعالیت می‌کنند، بدلیل نداشتن سرمایه کافی امکان استفاده گسترده از این فناوری را ندارند. این فناوری معمولاً هنگامی استفاده می‌شود که شرکت‌ها و کسب و کارها به یک حد کافی از درآمد رسیده باشند و برای افزایش مزیت رقابتی خود در بازار، شروع به استفاده از داده‌هایی می‌کنند که رقیبان خود از وجود آنها غافل بوده و یا آنها را کم اهمیت می‌پندارند.

بدون شک، استفاده‌ی دقیق از این فناوری و شناسایی داده‌های مورد غفلت واقع شده از دید رقیبان، باعث ایجاد یک نقطه مثبت نسبت به سایرین خواهد شد اما سوالی که به وجود می‌آید این است که «آیا این امکان وجود ندارد که از این تکنولوژی در همان ابتدای شروع به کار یک بنگاه اقتصادی استفاده شود؟ آیا این امکان وجود دارد که هزینه استفاده از این فناوری به اندازه‌ای کاهش پیدا کند که کسب و کارهای کوچک هم بتوانند از قابلیت‌های این ابزار استفاده کنند؟»

برای پاسخ به این سوال، ابتدا باید مفاهیم اولیه هوش تجاری شناخته شده و تعاریف مربوط به آن فهمیده شود. سپس بعد از فهم اولیه نسبت به موضوع چيستی هوش تجاری، باید با اجزای مختلف آن آشنا شد و پس از فهم نحوه‌ی قرارگیری آنها کنار یکدیگر اقدام به طراحی کرد. در مرحله آخر نیز باید روش‌های استفاده از این ابزار و علوم مرتبط با آن را به خوبی درک کرد تا بتوان بیشترین بازدهی را از آن گرفت.

**واژه‌های کلیدی:** هوش تجاری، مدیریت فناوری اطلاعات، پردازش‌های تحلیلی آنلاین، ساختارهای دسترسی به داده، سیستم‌های پشتیبانی از تصمیم.

<sup>۱</sup> Business Intelligence

<sup>۲</sup> Startup

## فهرست مطالب

صفحه	عنوان
۱.....	<b>فصل اول: مقدمه</b>
۲.....	۱-۱. مقدمه
۴.....	<b>فصل دوم: معرفی هوشمندی کسب و کار</b>
۵.....	۱-۲. تعریف هوشمندی کسب و کار
۵.....	۲-۲. تاریخچه‌ای کوتاه
۶.....	۳-۲. کاربردها
۷.....	نتیجه‌گیری
۸.....	<b>فصل سوم: معماری در هوشمندی کسب و کار</b>
۹.....	۱-۳. معماری هوش تجاری
۹.....	۲-۳. داده‌های ورودی و انتقال جریان داده
۱۰.....	۳-۳. پایگاه داده تحلیلی
۱۱.....	۴-۳. سرورهای میانی
۱۲.....	۵-۳. ارائه نمایش خروجی به کاربر
۱۲.....	نتیجه‌گیری
۱۳.....	<b>فصل چهارم: ذخیره و بازیابی اطلاعات</b>
۱۴.....	۱-۴. ساختارهای دسترسی به داده
۱۵.....	۲-۴. فشرده‌سازی اطلاعات
۱۶.....	۱-۲-۴. سرکوب پوچ‌ها
۱۶.....	۲-۲-۴. فشرده‌سازی فرهنگ لغات
۱۶.....	۳-۲-۴. رمزگذاری طول اجرا
۱۷.....	۳-۴. پردازش دستورات

۱۸.....	۴-۴. سرورهای رابطه‌ای
۱۸.....	۴-۴-۱ بهینه‌سازی پرس‌وجو
۱۹.....	۴-۴-۲ پردازش موازی پرس‌وجو
۲۰.....	نتیجه‌گیری
۲۱.....	<b>فصل پنجم: نتیجه‌گیری و پیشنهادها</b>
۲۲.....	نتیجه‌گیری
۲۲.....	پیشنهادها
۲۳.....	<b>مراجع</b>

## فهرست اشکال

صفحه	عنوان
۲.....	شکل ۱-۱ میزان درآمد یک کارشناس حوزه هوش تجاری در شرکت‌های مختلف در سال ۲۰۲۱.....
۷.....	شکل ۲-۱ دید ابتدایی نسبت به هوش تجاری.....
۹.....	شکل ۱-۳ معماری اولیه هوش تجاری.....
۱۷.....	شکل ۱-۴ مثالی از الگوریتم RLE.....
۱۸.....	شکل ۲-۴ یک نمایش از داده‌های چند بعدی.....

## فصل اول

### ۱. مقدمه



## ۱-۱ مقدمه

با گذشت زمان و ایجاد جوامع و تمدن‌های بشری، نیازهای انسان‌ها هم افزایش پیدا کرد. با افزایش تقاضا، افرادی درصدد رفع نیاز برآمده و مشاغل مختلف را در طول تاریخ ایجاد کردند. صاحبان صنایع، راز کیفیت کار خود را استفاده از مواد خام بهتر برای تولید با کیفیت بالاتر می‌دانستند. با گذشت زمان و رشد علم، بشر این توانایی را پیدا کرد که کیفیت محصولات را نه تنها با استفاده از افزایش کیفیت مواد خام، بلکه با استفاده از بهبود فرایند تولید بالا ببرد. در نتیجه اولین افزایش کیفیت محصول با استفاده از بهبود فرایند در تاریخ، برای کسب جایگاه برتر در بازار رقم خورد [۱].

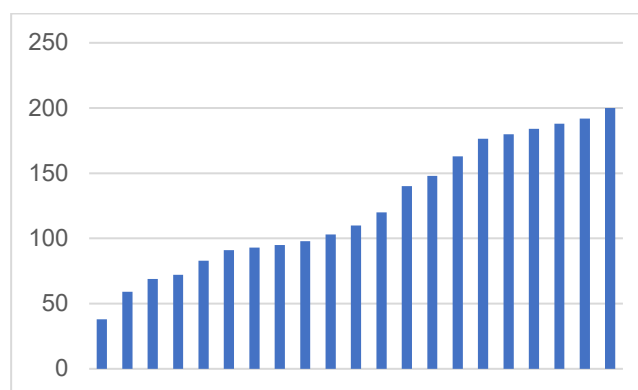
با گذشت زمان این رقابت بین صاحبان صنایع در جریان بود و هر کسی سعی داشت تا با پیدا کردن داده‌ی مناسب، فرایند تولید محصول خود را افزایش دهد. داده‌هایی که می‌توانستند نقاط قوت و ضعف محصول و یا خدمت را به ذی‌نفعان نشان داده و ایشان را به سوی توسعه آنها راهنمایی کنند. اما مهمترین چالش هنوز باقی است؛ این داده‌ها از کجا باید جمع‌آوری شود؟ بعد از پاسخ به این سوال، سوال جدید مطرح می‌شود؛ چگونه باید این داده‌ها را ذخیره و سپس در زمان مناسب بازیابی و استخراج کرد؟

با پیشرفت دانش در حوزه علوم کامپیوتری، متخصصان این حوزه به این فکر افتادند که آیا این امکان وجود دارد که بتوان از فناوری‌های این حوزه در عرصه‌ی اقتصاد و کسب و کار هم استفاده کرد؟ و در اینجا ایده‌ی اولین سامانه‌ی هوشمندی کسب و کار به وجود آمد. سامانه‌ای که هدف آن آسان‌سازی فرایند جمع‌آوری، دسته‌بندی و تحلیل داده‌های مربوط به یک حوزه کاری است.

به ظاهر انجام این کار امکان پذیر بود تا زمانی که یک مسئله مشترک در بین فعالیتهای اقتصادی مختلف به وجود آمد؛ «برای شناسایی داده‌های کاربردی هر حوزه‌ی کاری، به یک تجربه و دانش عمیق نسبت به آن حوزه نیاز است.» همین موضوع باعث می‌شود تا هزینه استفاده از این فناوری بسیار بالا برود؛ زیرا تعلیم تمامی نکات قابل توجه در یک حوزه کاری به یک نیروی متخصص، مسئله بسیار هزینه‌سازی است. غالب هزینه شرکت‌ها و کسب و کارها در این حوزه صرف جمع‌آوری داده‌ها از منابع مختلف و تحلیل آنها می‌باشد.

همانطور که در شکل ۱-۱ مشاهده می‌کنید، حقوق یک متخصص حوزه هوش تجاری از مبلغ ۳۸ هزار دلار تا ۲۰۰ هزار دلار متغیر است. این مبلغ برای یک کسب و کار کوچک هزینه کمی نیست. پس نیاز به راهکاری است که بتوان به وسیله آن، هزینه استفاده از این فناوری را کاهش داد و در نتیجه امکان استفاده از داده‌های ایجاد شده در کسب و کارهای مختلف را برای حوزه‌های اقتصادی کوچک نیز فراهم کرد [۲].

شکل ۱-۱ میزان درآمد یک کارشناس حوزه هوش تجاری در شرکت‌های مختلف در سال ۲۰۲۱



علت محبوبیت این فناوری در میان کسب و کارهای نوین، توجه و دقت ویژه به داده‌ها و رخدادهایی است که از طرف کاربران مورد توجه واقع نمی‌شود ولی می‌توان از آن‌ها اطلاعات مفیدی را استخراج کرد. از مهمترین وظایف هوش تجاری، تبدیل داده‌های خام به اطلاعات مفید و قابل ارزیابی برای ذی‌نفعان است.

هدف ما از این پژوهش دستیابی به مدلی ساده و قابل استفاده برای عموم است. این که بتواند در انتهای این پژوهش یک مدلی برای ساخت یک محصول با کاربری عمومی برای اکثر کسب و کارهای امروزی در کشور ارائه شود. محصولی که برای استفاده از آن نیازی به یادگیری الگوریتم‌ها و مدل‌های پیچیده علوم کامپیوتری نباشد.

در همین راستا در این گزارش، ابتدا بعد از تعریف مفاهیم اولیه هوش تجاری و کاربردهای مختلف این فناوری، نگاهی اجمالی به معماری استفاده شده در آن می‌اندازیم. بعد از آن به بررسی اجزای این سامانه پرداخته و در نهایت، به چگونگی نحوه‌ی عملکرد آن در ذخیره‌سازی اطلاعات و بازیابی آنها خواهیم پرداخت.

## فصل دوم

### ۲. معرفی هوشمندی کسب و کار

## ۲-۱ تعریف هوشمندی کسب و کار

هوشمندی کسب و کار یا هوش تجاری<sup>۳</sup> - به صورت اختصاری (BI) - مجموعه‌ای از فناوری‌های پشتیبانی از تصمیم است که هدف آن سرمایه‌گذاری برای توانمندسازی کارکنانی مثل مدیران اجرایی، مدیران راهبردی و نظارتی و تحلیلگران شرکت‌ها است تا فرایند تصمیم‌گیری بهتر و سریع‌تر انجام شود. در واقع هوشمندی کسب و کار روش‌ها، نظریات، معماری‌ها و فناوری‌هایی است که برای تبدیل داده‌های خام<sup>۴</sup> به اطلاعات<sup>۵</sup> مفید و معنادار استفاده می‌شود. اطلاعات موجود در یک کسب و کار، شامل طیف‌های گسترده و مقادیر بسیار بزرگی از اطلاعات و داده‌های متنوع می‌باشد. هوشمندی کسب و کار تلاش می‌کند تا با استفاده از نظریات و فرایندهای اثربخش خود، این اطلاعات را برای شناسایی و توسعه فرصت‌های جدید بکار بگیرد [۱].

استفاده کردن از فرصت‌های ایجاد شده و بکار گرفتن یک راهبرد مناسب می‌تواند مزیت بازار رقابتی و ماندگاری بلندمدت محصول (و یا خدمت) را برای کسب و کارها به همراه داشته باشد.

در رابطه با کلمه‌ی «هوش تجاری» می‌توان با توجه به نوع استفاده، دو معنا برای آن در نظر گرفت؛ معنای اول که به ندرت مورد استفاده قرار می‌گیرد، بکارگیری ظرفیت‌های «هوش انسانی» در امور و یا فعالیت‌های تجاری است. هوشمندی کسب و کار یک حوزه‌ی جدیدی است که شامل بررسی قوای شناختی انسان و فناوری‌های هوش مصنوعی و یادگیری ماشین<sup>۶</sup> در مدیریت و پشتیبانی از تصمیم‌گیری در مسائل مختلف تجاری یک کسب و کار است.

در واقع می‌توان گفت که با استفاده از این تعریف، هوش تجاری یعنی استفاده‌ی خلاقانه مدیران اجرایی و مسئولان تحلیل و راهبرد کسب کارها از داده‌ها و اطلاعات خام موجود، بدون در نظر گرفتن نظریه‌ها و شاخص‌ها و صرفاً با نظر گرفتن هوش انسانی و استفاده از ابزارهای هوش مصنوعی و یادگیری ماشین برای ارائه‌ی داده‌ها و دیداری کردن<sup>۷</sup> آنها برای مسئولین مربوطه.

تعریف دوم، مربوط می‌شود به تعریف «هوش تجاری» به عنوان اطلاعاتی که برای هزینه‌ها و ارتباط بین آن‌ها ارزش‌گذاری می‌شوند. این اطلاعات، دانش و فناوری‌های تخصصی کارآمد در مدیریت کسب و کارهای سازمانی و فردی است. بنابر این، از این نظر هوش تجاری دسته وسیعی از برنامه‌ها و فناوری‌ها برای جمع‌آوری، ارائه دسترسی<sup>۸</sup> و تجزیه و تحلیل داده‌ها به منظور کمک به کاربران سازمانی در تصمیم‌گیری بهتر تجاری است. این اصطلاح به معنای داشتن یک دانش جامع از تمام عوامل مؤثر بر روی تجارت است. با توجه به این موضوع، الزامی است که شرکت‌ها و کسب و کارها دانشی عمیقی درباره‌ی عواملی مانند مشتریان، رقیبان، شرکای تجاری، محیط‌های اقتصادی و عملیات‌های داخلی کسب کنند تا تصمیمات تجاری مؤثر و با کیفیت خوبی اتخاذ کنند. هوش تجاری شرکت‌ها را قادر می‌سازد تا از این قبیل تصمیمات بگیرند [۳].

## ۲-۲ تاریخچه‌ای کوتاه

اولین استفاده از اصطلاح هوش تجاری در کتابی از آقای «ریچارد میلار دونس»<sup>۹</sup> در کتاب «دایرة المعارف حکایت‌های بازرگانی و تجاری» در سال ۱۸۶۵ بوده است. ایشان در این کتاب از اصطلاح هوش تجاری برای توصیف روش کسب سود توسط شخصیت

<sup>۳</sup> Business Intelligence (BI)

<sup>۴</sup> Data

<sup>۵</sup> Information

<sup>۶</sup> Machine Learning (ML)

<sup>۷</sup> Visualization

<sup>۸</sup> Providing access

<sup>۹</sup> Richard Millar Devens

داستانی خود، آقای فرنسی<sup>۱۰</sup>، استفاده می‌کند؛ به این صورت که وی در سراسر هلند، فلاند، فرانسه و آلمان یک مجموعه کامل را به عنوان هوش تجاری ثبت کرده بود. بدین ترتیب، خبر جنگ‌های زیادی که در آن دوران اتفاق می‌افتاد را او زودتر از همه واری می‌کرد و با جمع‌آوری اطلاعاتی کامل از شرایط کشورها توانست سود زیادی کسب کند. مفهومی که آقای دونس به آن اشاره می‌کند این است که «توانایی جمع‌آوری اطلاعات و واکنش مناسب بر اساس اطلاعات بدست آمده، در هوش تجاری حایز اهمیت است.»

در دو دهه‌ی گذشته، یک رشد انفجاری، هم در تعداد محصولات و خدمات ارائه شده و هم در پذیرش و اتخاذ این فناوری توسط صنعت دیده شده است. این رشد به دلیل کاهش هزینه‌های بدست آوردن و ذخیره‌سازی مقادیر بسیار زیادی از داده‌هایی است که ناشی از منابع مختلفی مثل تراکنش‌های مشتریان در بانکداری، خرده‌فروشی و همچنین در کسب و کارهای الکترونیکی برای ردیابی ایمیل‌ها، موجودی حساب و گزارش‌های جست و جو برای صفحه‌های اینترنتی<sup>۱۱</sup>، وبلاگ‌های بررسی محصولات، هستند. امروزه شرکت‌ها داده‌ها را با جزئیات دقیق‌تری جمع‌آوری می‌کنند که با توجه به همین موضوع حجم داده‌های نگهداری شده توسط شرکت‌ها بسیار زیاد خواهد شد [۶].

## ۲-۳ کاربردها

هوش تجاری بسیار کاربرد گسترده‌ای در صنعت دارد. هوش تجاری بدلیل «تحلیل‌گر بودن نرم‌افزارهای مرتبط» با آن و فراهم کردن امکان گزارش‌دهی دقیق و تحلیل‌های جامع، راه ورود خود را به اکثر صنایع باز کرده است. همانطور که در بخش معرفی گفته شد، هوش تجاری شامل یکسری نظریات، ابزارها، معماری‌ها و فناوری‌هایی است که قابلیت تحلیل و ارزیابی را برای مسئولین کسب و کارها فراهم می‌کند. به طور مثال یک شرکت بیمه را متصور شوید؛ امروزه راهکارهای محصول‌گرایی شرکت‌های بیمه، دیگر کاربردی ندارند و مشتریان انتظارات متفاوت و متنوعی را از شرکت‌های بیمه دارند. در ضمن طی سال‌های اخیر و با ورود شرکت‌های بیمه خصوصی و موسسات ارائه دهنده خدمات مالی، بازار از شرایط انحصاری خود خارج شده و رقابت تنگاتنگی بین شرکت‌های بیمه، برای جذب مشتری و ارائه خدمات متنوع ایجاد شده است.

شرکت‌های بیمه، داده‌های حجیم و متنوعی را از جمله داده‌های بیمه گزاران، بیمه نامه‌ها و خسارات را در بانک‌های اطلاعاتی خود که شامل بیمه‌های عمر، اتومبیل، درمان، مسولیت، معلم و ... ذخیره و نگهداری می‌کنند. در این بین مشکل بزرگی در شکل یک کابوس، برای تیم فناوری اطلاعات ایجاد می‌شود که بدون استفاده از هوش تجاری، فرآیند بازیابی و جمع‌آوری داده‌ها و انتقال آنها به نرم‌افزارهای اطلاعاتی، بسیار وقت گیر و کند پیش می‌رود. از این رو پیاده‌سازی هوش تجاری نقش بسیار مهمی را در همه شاخه‌های صنعت بیمه ایفا می‌کند تا مدیران این صنعت بتوانند با تحلیل داده‌های حجیم ذینفعان بهترین مدل بیمه را ارائه دهند.

پس این شرکت‌ها می‌توانند با استفاده از روش‌های هوش تجاری عملکرد شرکت خود را ارتقا دهند؛ به طور مثال:

- هوش تجاری در شناسایی مشتریان بالقوه برای فعالیت‌های بازاریابی بیمه‌ای بسیار مفید است.
- تحلیل سودآوری و کارایی محصولات با دقت بالا انجام خواهد شد.

<sup>۱۰</sup> Furnese

<sup>۱۱</sup> Web sites

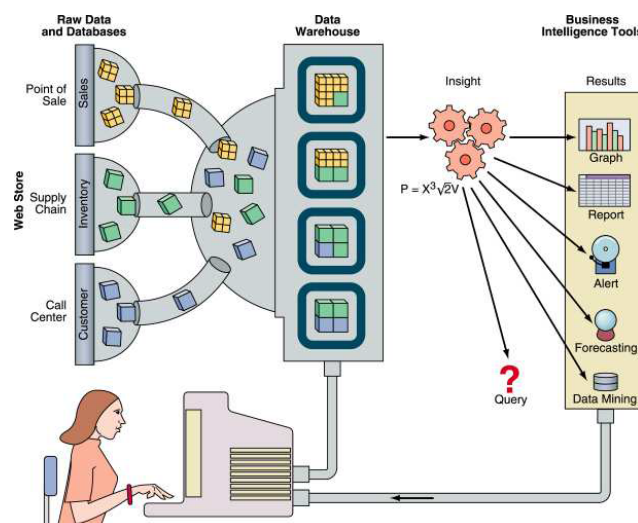
- تحلیل و بررسی دقیق طرح‌های مختلف بیمه و کنترل سود و زیان هر کدام، با استفاده از هوش تجاری انجام خواهد شد.
- کنترل و مدیریت ریسک‌ها با دقت بالایی انجام خواهد شد و فرصت‌های عالی موجود در بخش‌های ویژه بازار را تعیین می‌نماید. در نتیجه، بخش‌های مختلف بازار را که می‌توانند به یکدیگر مربوط باشند را پیدا خواهد کرد.
- کنترل و مدیریت ادعاهای دروغین خسارت و شناسایی فریب کاری‌ها، به بهترین شکل ممکن انجام خواهد شد.
- با استفاده از سیستم‌های هوش تجاری، محدودیت‌های ذاتی انسان در عدم توانایی پاکسازی حجم زیادی از داده‌ها از بین خواهد رفت و در نتیجه احتمال تصمیم‌های مخرب و با ریسک بالا کاهش خواهد یافت.
- با استفاده از تکنیک‌های هوش تجاری، مدیریت شعب و نمایندگی‌ها به بهترین نحو انجام می‌شود.

مثال‌های بسیار زیادی می‌توان برای کاربردهای هوشمندی کسب و کار در صنعت اشاره کرد از قبیل مدیریت ناوگان‌های دریایی و هوایی، در حوزه‌ی مخابرات برای شناسایی دلایل ریزش مشتریان، در تأسیسات صنعتی برای تجزیه و تحلیل مصرف برق، برای تجزیه و تحلیل نتایج پزشکی در حوزه‌ی بهداشت و سلامت و ... [۷].

## نتیجه‌گیری

هوشمندی کسب و کار یا همان هوش تجاری، یکی از پرکاربردترین فناوری‌هایی است که امروزه جایگاه خود را در بین تمامی صنایع و جوامع دارد پیدا می‌کند. فناوری که ارتباط تنگاتنگی با دانش عمیق در حوزه‌های مختلف صنعتی دارد و اهمیت استفاده از دانش‌های روز دنیا نظیر هوش مصنوعی و یادگیری ماشین را نشان می‌دهد. هوشمندی کسب و کار به نشان داد که صنایع مختلف برای حفظ بازار رقابتی خود به ناچار مجبور هستند که به سمت استفاده از نظریات، روش‌ها و فرایندهای هوشمندی کسب و کار بروند تا بتوانند در بازار کار آینده جایگاه خود را حفظ کنند.

هوش تجاری مجموعه‌ای از فناوری‌های پشتیبانی از تصمیم است که فرایند تصمیم‌گیری را برای مدیران در عرصه‌ها و جایگاه‌های مختلف آسان‌تر و سریع‌تر می‌کند.



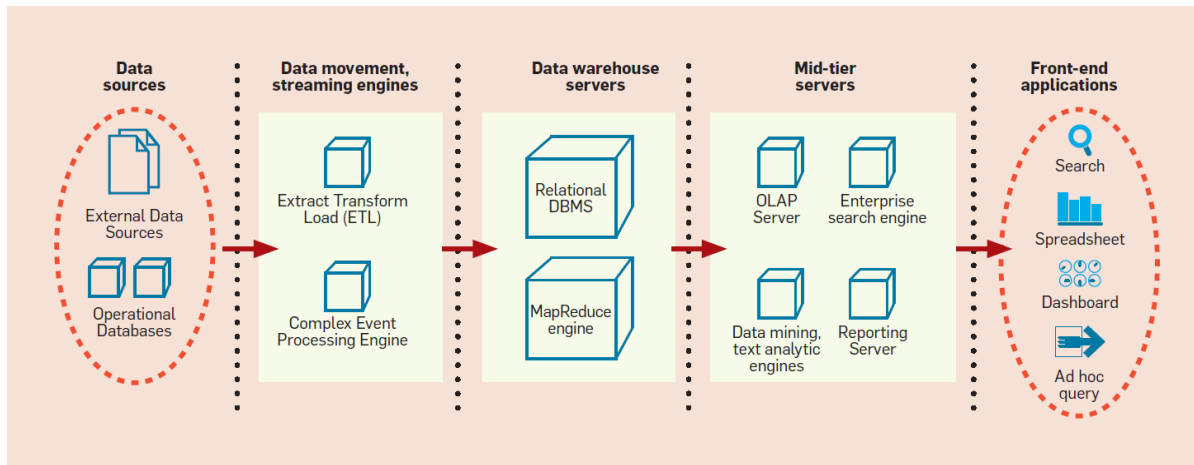
شکل ۲-۱ دید/پنداری نسبت به هوش تجاری

## فصل سوم

### ۳. معماری در هوشمندی کسب و کار

### ۳-۱ معماری هوش تجاری

یکی از مهم‌ترین موضوعات در هر فناوری، نحوه‌ی قرارگیری اجزای<sup>۱۲</sup> آن در کنار یکدیگر و رابط‌های بین آن‌ها است. هوشمندی کسب و کار هم این قضیه مستثنی نیست. بنابراین داشتن دانش حداقلی درباره اجزا و نحوه‌ی قرارگیری آنها در کنار یکدیگر بسیار لازم و ضروری است. به طور کلی می‌توان گفت که امروزه سامانه‌های مبتنی بر هوش تجاری، دارای ۴ بخش اصلی هستند؛ داده‌های ورودی و جریان داده، پایگاه‌داده‌های تحلیلی، سرورهای میانی و رابط کاربری و ارائه‌ی اطلاعات به کاربران و ذی‌نفعان مربوطه [۷].



شکل ۳-۱ معماری اولیه هوش تجاری

بررسی هر کدام از این قسمت‌ها بسیار مهم و حیاتی است و برای ایجاد هر کدام از آن‌ها نیاز به دانش تخصصی درباره علوم کامپیوتری و دانش عمیقی در حوزه‌ی خاص صنعتی است که قصد استفاده از فناوری‌های هوش تجاری را دارد؛ به همین دلیل است که هزینه‌ای که شرکت‌ها و کسب و کارها برای این حوزه می‌کنند، هزینه‌ی بسیار زیادی است. هر کسب و کاری که در حوزه‌ی علوم کامپیوتری فعالیت نمی‌کند باید هزینه‌ای را صرف متخصص حوزه‌ی فناوری اطلاعات<sup>۱۳</sup> کند تا با کسب و کار آن‌ها و انواع متغیرهای موجود در این حوزه آشنا شود. این همان هزینه‌ای است که ما قصد کاهش آن را با انجام این پژوهش داریم.

### ۳-۲ داده‌های ورودی و انتقال جریان داده

داده‌ها در یک شرکت ممکن است که از منابع مختلفی تولید شوند؛ ممکن است که از پایگاه‌داده‌های مختلفی که در یک شرکت و یا سازمان وجود دارد تولید شوند و یا اینکه از یک منبع خارج از سازمان و یا شرکت وارد شوند. تنوع داده‌های ورودی باعث می‌شود که شرکت و یا سازمان با داده‌هایی با کیفیت‌های مختلفی مواجه شود. داده‌هایی که قالب‌ها و کدگذاری‌های متفاوتی دارند که برای مورد استفاده قرارگرفتن، باید با یکدیگر تطبیق داده شوند. اولین چالشی که گروه‌های متخصص هوش تجاری با آن مواجه هستند، یکپارچه سازی<sup>۱۴</sup>، پاکسازی<sup>۱۵</sup> و استاندارد کردن<sup>۱۶</sup> داده‌ها به جهت آماده کردن آن‌ها برای فرایندهای هوش تجاری است. بارگذاری کارآمد داده‌ها برای هوش تجاری بسیار لازم و ضروری است.

<sup>۱۲</sup> Components

<sup>۱۳</sup> IT

<sup>۱۴</sup> Integrating

<sup>۱۵</sup> Cleansing

<sup>۱۶</sup> Standardizing



علاوه بر این، معمولاً با رسیدن داده‌های جدید، وظایف هوش تجاری به صورت تدریجی انجام خواهد شد. ورودی تدریجی داده‌ها، بارگذاری داده‌های کارآمد و مقیاس پذیر را ساده‌تر کرده و قابلیت تازه‌سازی را برای سازمان ممکن می‌سازد. این فناوری‌ها پشتیبانی برای آماده‌سازی داده‌ها برای هوش تجاری هستند که در مجموع به آن‌ها ابزارهای استخراج-انتقال-بارگیری<sup>۱۷</sup> می‌گویند.

امروزه به طور فزاینده‌ای نیاز به پشتیبانی از وظایف هوش تجاری به صورت در لحظه<sup>۱۸</sup> وجود دارد؛ یعنی تصمیمات تجاری بر اساس خود داده‌های عملیاتی، اتخاذ می‌شوند. برای پشتیبانی از چنین روندی، موتورهای تخصصی به نام موتورهای پردازش رویداد پیچیده<sup>۱۹</sup> پدید آمده‌اند [۱].

عملیات استخراج-انتقال-بارگیری و موتورهای پردازش رویداد پیچیده از مهم‌ترین قسمت‌های روند انتقال داده‌ها به سمت پایگاه‌داده‌های تحلیلی هستند. دو عنصری که در پاکسازی و آماده‌سازی داده‌ها نقش کلیدی ایفا می‌کنند.

### ۳-۳ پایگاه داده تحلیلی

برنامه‌های هوش تجاری، از داده‌های جمع‌آوری شده در پایگاه‌های داده (به اصطلاح به آن‌ها انبار داده گفته می‌شود) با یک ساختار داده استفاده می‌کنند. در یک انبار داده، نسخه‌ای از داده‌هایی که قرار است عملیات تحلیل روی آن‌ها انجام شود و تصمیم‌گیری‌های تجاری توسط آن‌ها انجام شود، نگهداری می‌شود. داده‌هایی که وظایف هوش تجاری بر روی آن‌ها انجام می‌شود، توسط یک یا چند سرور انبار داده مدیریت می‌شود. یکی از محبوب‌ترین انتخاب‌ها از بین موتورهای جست‌وجویی که قابلیت پرسیدن<sup>۲۰</sup> و پاسخ‌دهی دارند، سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای<sup>۲۱</sup> است.

در طول دو دهه گذشته، چندین ساختار داده، بهینه‌سازی و تکنیک‌های پردازش پرس‌وجو عمدتاً برای اجرای دستورات پیچیده بر روی حجم زیادی از داده‌ها توسعه داده شده‌اند که نیاز کلیدی برای هوش تجاری است. برای فهم بهتر موضوع، به مثال زیر که یک نمونه از دستور تک‌کاره‌ی پیچیده است توجه کنید: «یافتن مشتریانی که در سه ماهه گذشته سفارشی را ثبت کرده‌اند که مبلغ آن‌ها حداقل ۵۰ درصد از میانگین مقدار سفارش بیشتر است.» برای انجام چنین دستوری در انبارهای داده بزرگ معمولاً موتورهای RDBMS را به صورت موازی را مستقر می‌کنند تا پرس‌وجوها را بتوان روی حجم زیادی از داده با تأخیر کم اجرا کرد. از آنجایی که داده‌های بیشتری به صورت دیجیتالی متولد می‌شوند، تمایل فزاینده‌ای برای معماری‌های داده کم هزینه وجود دارد که می‌توانند حجم داده‌های بسیار بیشتری نسبت به آنچه که به طور سنتی توسط RDBMS مدیریت می‌شود، پشتیبانی کنند.

این اغلب به عنوان چالش «داده بزرگ»<sup>۲۲</sup> توصیف می‌شود. با هدایت این هدف، موتورهای مبتنی بر الگوی کاهش‌نگاشت<sup>۲۳</sup> - که در ابتدا برای تجزیه و تحلیل اسناد وب و گزارش‌های جستجوی وب ساخته شده بودند - اکنون برای تجزیه و تحلیل سازمانی هدف قرار می‌گیرند. چنین موتورهایی در حال حاضر برای پشتیبانی از پرس‌وجوهای پیچیده که برای سناریوهای انبار

<sup>۱۷</sup> Extract-Transform-Load (ETL)

<sup>۱۸</sup> Real Time

<sup>۱۹</sup> Complex Event Processing (CEP)

<sup>۲۰</sup> Querying

<sup>۲۱</sup> Relational Database Management Systems (RDBMS)

<sup>۲۲</sup> Big Data

<sup>۲۳</sup> MapReduce

داده سنتی سازمانی ضروری است، گسترش می‌یابند. سرورهای انبار داده با مجموعه‌ای از سرورهای سطح میانی تکمیل می‌شوند که عملکردهای تخصصی را برای سناریوهای مختلف هوش تجاری ارائه می‌کنند [۱].

### ۴-۳ سرورهای میانی

سرورهای پردازش تحلیلی آنلاین<sup>۲۴</sup> به طور موثر نمای چندبعدی داده‌ها را در معرض دید برنامه‌ها یا کاربران قرار می‌دهند و عملیات رایج هوش تجاری مانند فیلتر کردن، تجمیع، تمرین کردن، و چرخش را فعال می‌کنند. انبار داده و پردازش تحلیلی آنلاین عناصر ضروری پشتیبانی تصمیم هستند که به طور فزاینده‌ای به کانون توجه صنعت پایگاه داده تبدیل شده است. بسیاری از محصولات و خدمات تجاری در حال حاضر در دسترس هستند و همه فروشندگان اصلی سیستم مدیریت پایگاه داده اکنون در این زمینه‌ها پیشنهادهایی دارند. پشتیبانی تصمیم‌گیری الزامات متفاوتی را در فناوری پایگاه داده در مقایسه با برنامه‌های سنتی پردازش تراکنش آنلاین ایجاد می‌کند.

برای تسهیل تحلیل‌ها و تجسم پیچیده، داده‌ها در انبار معمولاً به صورت چند بعدی مدل‌سازی می‌شوند. به عنوان مثال، در یک انبار داده فروش، زمان فروش، منطقه فروش، فروشنده و محصول ممکن است برخی از ابعاد مورد علاقه باشد. اغلب، این ابعاد سلسله مراتبی هستند. زمان فروش ممکن است به صورت سلسله مراتب روز-ماه-سه ماهه سال، محصول به عنوان سلسله مراتب محصول-دسته-صنعت سازماندهی شود.

عملیات پردازش تحلیلی آنلاین معمولاً شامل افزایش سطح تجمیع<sup>۲۵</sup>، کاهش سطح تجمیع یا افزایش جزئیات<sup>۲۶</sup> در امتداد یک یا چند سلسله مراتب بعدی خود، انتخاب و طرح ریزی<sup>۲۷</sup>، محور<sup>۲۸</sup> و جهت یابی مجدد نمای چند بعدی از داده‌ها است [۳].

علاوه بر سرورهای سنتی پردازش تحلیلی آنلاین، موتورهای جدیدتری به نام «رویکرد حافظه در هوش تجاری»<sup>۲۹</sup> ظاهر می‌شوند که از اندازه‌های بزرگ حافظه اصلی امروزی برای بهبود چشمگیر عملکرد پرس‌وجوهای چند بعدی استفاده می‌کنند.

سرورهای گزارش‌دهی<sup>۳۰</sup>، تعریف، اجرای کارآمد و ارائه گزارش‌ها را امکان‌پذیر می‌کنند - برای مثال، کل فروش را بر اساس منطقه برای سال جاری گزارش می‌کند و با فروش سال گذشته مقایسه می‌کند - افزایش در دسترس بودن و اهمیت داده‌های متنی مانند بررسی محصول، ایمیل و رونوشت‌های مرکز تماس برای هوش تجاری چالش‌های جدیدی را به همراه دارد. برای حل این موضوع، موتورهای جستجوی سازمانی<sup>۳۱</sup> ایجاد شدند که از الگوی جستجوی کلمه کلیدی بر روی متن و داده‌های ساختاریافته در انبار داده پشتیبانی می‌کنند (به عنوان مثال، یافتن پیام‌های ایمیل، اسناد، تاریخچه خریدها و تماس‌های پشتیبانی مربوط به یک مشتری خاص) و در دهه‌ی گذشته به ابزاری ارزشمند برای هوش تجاری تبدیل شده‌اند.

<sup>۲۴</sup> Online analytic processing (OLAP)

<sup>۲۵</sup> rollup

<sup>۲۶</sup> drill-down

<sup>۲۷</sup> Slice-and-dice

<sup>۲۸</sup> pivot

<sup>۲۹</sup> In-Memory BI

<sup>۳۰</sup> Reporting servers

<sup>۳۱</sup> Enterprise search engines

موتورهای داده کاوی<sup>۳۲</sup> تجزیه و تحلیل عمیق داده‌ها<sup>۳۳</sup> را امکان‌پذیر می‌سازند که فراتر از آنچه توسط پردازش تحلیلی آنلاین یا سرورهای گزارش‌دهی ارائه می‌شود، توانایی ساخت مدل‌های پیش‌بینی‌کننده برای کمک به پاسخ‌گویی به سؤالاتی مانند: «مشتریان فعلی احتمالاً به پست‌های کاتالوگ آتی من پاسخ خواهند داد؟» باشند.

موتورهای تجزیه و تحلیل متن می‌توانند مقادیر زیادی از داده‌های متنی را تجزیه و تحلیل کنند (به عنوان مثال، پاسخ‌های نظرسنجی یا نظرات مشتریان) و اطلاعات ارزشمندی را استخراج کنند که در غیر این صورت به تلاش دستی قابل توجهی نیاز دارد، به عنوان مثال، محصولاتی که در پاسخ‌های نظرسنجی ذکر شده‌اند و موضوعاتی که بسیار پرتکرار هستند.

چندین برنامه کاربردی محبوب وجود دارد که کاربران از طریق آنها وظایف هوش تجاری را انجام می‌دهند؛ مثل صفحه‌های گسترده<sup>۳۴</sup>، پورتال‌های سازمانی برای جستجو<sup>۳۵</sup>، برنامه‌های کاربردی مدیریت عملکرد که تصمیم‌گیرندگان را قادر می‌سازد تا شاخص‌های عملکرد کلیدی کسب‌وکار را با استفاده از داشبوردهای بصری ردیابی<sup>۳۶</sup> کنند، ابزارهایی که به کاربران امکان می‌دهد پرسوچوهای موردی را مطرح کنند و بیندگانی مدل‌های داده‌کاوی را دنبال می‌کنند و غیره.

تجسم سریع و موقت داده‌ها می‌تواند کاوش پویا از الگوها و موارد غیرمرتبط را ممکن کند و به کشف حقایق مرتبط برای هوش تجاری، کمک کند [۱].

### ۳-۵ ارائه نمایش خروجی به کاربر

یکی از مهم‌ترین بخش‌های هر فرایند و پروژه‌ای، گرفتن خروجی و ارائه‌ی کار به کارفرما و ذی‌نفعان است. هوش تجاری هم از این قضیه مستثنی نیست. ایجاد داشبوردهای دسته‌بندی شده و قابل استفاده برای مدیران، یکی از مهم‌ترین مراحل در هوش تجاری است.

مدیران باید بتوانند که به سادگی به تجزیه و تحلیل داده‌های کسب و کار خود دسترسی داشته و بتوانند به وسیله‌ی آنها تصمیمات تجاری متناسب را اتخاذ کنند و جایگاه خود را در بازار حفظ کنند. همانطور که در قسمت ۲-۲ گفته شد، هوش تجاری علم جمع‌آوری اطلاعات و نحوه‌ی انتخاب یک راهبرد مناسب برای پیمودن ادامه مسیر در کسب و کار و فعالیت‌های تولیدی/خدماتی است [۶].

### نتیجه‌گیری

با توجه به مطالب گفته شده در قسمت‌های قبلی، هوش تجاری دارای معماری پیچیده و تخصصی است. معماری که برای کنار هم قرار دادن اجزای آن، نیاز به داشتن دانش در علوم مختلف کامپیوتری و استفاده‌ی تخصصی از ابزارهای این حوزه است. بسیار لازم است که علاقه‌مندان به این حوزه، علاوه بر دانش استفاده از ابزارهای کامپیوتری، نسبت به طراحی و معماری‌های مربوطه دانش عمیقی کسب کنند.

<sup>۳۲</sup> Data mining engines

<sup>۳۳</sup> in-depth analysis

<sup>۳۴</sup> Spreadsheets

<sup>۳۵</sup> enterprise portals for searching

<sup>۳۶</sup> track

## فصل چهارم

### ۴. ذخیره و بازیابی اطلاعات

## ۴-۱ ساختارهای دسترسی به داده

دستورات پشتیبانی از تصمیم، نیازمند یکسری از عملیاتی‌هایی مثل گزینش کردن<sup>۳۷</sup>، پیوستن<sup>۳۸</sup> و تجمیع کردن<sup>۳۹</sup> هستند. برای پشتیبانی کارآمد از این قبیل عملیات، ساختار داده‌های ویژه‌ای توسعه داده شده‌اند. یکی از این ساختار داده‌ها، «ساختارهای شاخص<sup>۴۰</sup>» است. این ساختار شاخص، دسترسی را به صورت انجمنی<sup>۴۱</sup> بر اساس مقادیر یک ستون خاص امکان‌پذیر می‌کند؛ هنگامی که یک پرس‌وجو دارای چند نوع شرایط گزینش باشد، گزینش‌پذیری‌های<sup>۴۲</sup> این شرایط را می‌توان (از طریق فهرستی که از شاخص‌های از پیش ساخته شده استفاده کرد) با همانندسازی<sup>۴۳</sup> کردن فهرست، مورد استفاده قرار داد. (به عنوان مثال، یک فهرست در ستون StoreId می‌تواند به بازیابی همه فروش‌ها برای StoreId = ۲۳ کمک کند) در کنار این می‌توان هنگامی که از چند شرط متعامد<sup>۴۴</sup> استفاده می‌شود، از قابلیت تقاطع<sup>۴۵</sup> بهره برد.

این عملیات می‌تواند به طور قابل توجهی نیاز به دسترسی به جداول پایه را کاهش می‌دهد. در برخی موارد به طور کلی این نیاز را از بین می‌برد. به طور مثال زمانی که ما شاخص ما شامل تمام ستون‌های مورد نیاز باشد، دیگر نیازی به پردازش بر روی جداول پایه نیست. این کار سرعت پردازش را بسیار افزایش می‌دهد [۵].

پرس‌وجوهای گزارش‌دهی<sup>۴۶</sup> اغلب به داده‌های خلاصه<sup>۴۷</sup> نیاز دارند، به عنوان مثال، فروش کل در سه ماهه اخیر و سال مالی جاری. برای ساده‌سازی این کار از مفهومی به عنوان «نمای واقعی‌سازی<sup>۴۸</sup>» استفاده می‌شود. نمای واقعی‌سازی شده با استفاده از فرایند جداسازی قسمتی از جداول پایه، می‌تواند تسریع بسیار چشمگیری به درخواست‌های پشتیبانی تصمیم داشته باشد. از این رو نیاز است که پیش‌محاسباتی<sup>۴۹</sup> و واقعی‌سازی داده‌های خلاصه انجام شود تا به سرعت پاسخ‌دهی به شدت افزایش یابد.

مهم‌ترین نقطه قوت نمای واقعی‌سازی شده، توانایی آن در هدف قرار دادن پرس‌وجوهای خاص و ذخیره مؤثر نتایج آنها در حافظه نهان<sup>۵۰</sup> است. با این حال، همین نقطه قوت می‌تواند یک نقطه ضعف جدی نیز برای این ساختار داده نیز هست و می‌تواند عملکرد آن را محدود کند؛ یعنی برای یک پرس‌وجوهای و کمی متفاوت‌تر از شاخص ساخته شده، دیگر امکان استفاده از شاخص ساخته شده نیست. در صورتی که پرس‌وجوی ما مقداری متفاوت باشد، نمای واقعی‌سازی شده قادر به پاسخگویی به پرس‌وجوی مطرح شده نیست.

<sup>۳۷</sup> Filtering

<sup>۳۸</sup> Join

<sup>۳۹</sup> Aggregation

<sup>۴۰</sup> Index structures

<sup>۴۱</sup> Associative

<sup>۴۲</sup> Selectivity

<sup>۴۳</sup> Scan

<sup>۴۴</sup> Multiple

<sup>۴۵</sup> Intersection

<sup>۴۶</sup> Reporting queries

<sup>۴۷</sup> Summary

<sup>۴۸</sup> Materialized views

<sup>۴۹</sup> Precomputing

<sup>۵۰</sup> Cache

نمای واقعی‌سازی شده ممکن است که در تضاد با مفهوم شاخص باشد؛ نمای واقعی‌سازی شده یک مفهوم بسیار کلی‌تری است اما تأثیر آن بر عملکرد پرس‌وجو ممکن است به اندازه‌ی یک نمای تحقق‌یافته چشمگیر نباشد. به طور معمول یک طراحی فیزیکی خوب شامل ترکیبی دقیق از شاخص‌ها و نماها است.

یکی دیگر از راهکارهای دسترسی به داده‌ها، استفاده از قسمت‌بندی<sup>۵۱</sup> است. قسمت‌بندی کردن داده‌ها را می‌توان برای بهبود عملکرد و مدیریت‌پذیری آنها استفاده کرد. قسمت‌بندی این قابلیت را به جداول و نمایه‌ها می‌دهد تا به واحدهای کوچک‌تر و قابل مدیریت‌تر تقسیم شوند. عملیات نگهداری پایگاه‌داده‌ها مانند بارگذاری و پشتیبان‌گیری را می‌توان بر روی قسمت‌بندی به جای کل جدول یا فهرست‌ها انجام داد. انواع رایج قسمت‌بندی که امروزه پشتیبانی می‌شوند عبارتند از ادغام کردن<sup>۵۲</sup> و تعیین محدوده<sup>۵۳</sup>. طرح‌های ترکیبی که اولین قسمت‌بندی بر اساس محدوده و به دنبال آن قسمت‌بندی کردن ادغام‌های انجام شده و در هر قسمت محدوده نیز رایج است.

پایگاه‌داده‌های سنتی، داده‌ها بر اساس سطرها در جدول دسته‌بندی می‌شوند؛ اما رویکرد دیگری وجود دارد که داده‌ها به جای اینکه بر اساس سطرها داده‌ها دسته‌بندی شوند، با رویکرد ستون‌گرا<sup>۵۴</sup> دسته‌بندی خواهند شد. در این روش، بدلیل اینکه مقدار داده‌های تکراری ما بسیار بیشتر از روش سطرگرا<sup>۵۵</sup> می‌باشد، عملیات فشرده‌سازی به صورت بهینه‌تری انجام پذیر است. دومین فایده و بهره این روش این است که تنها نیاز است که ستونی که مورد پرس‌وجو قرار گرفته است اسکن شود. با توجه به فوایدی که در قسمت نمای واقعی‌سازی شده اشاره شد، این کار می‌تواند در بهبود سرعت عمل سامانه نقش بسیار زیادی داشته باشد؛ در رویکرد سطرگرا، بدلیل اینکه نمی‌توانیم به راحتی از ستون‌هایی که نیاز به دسترسی نداریم، عبور کنیم، سرعت سامانه را به سرعت کاهش می‌دهد.

## ۴-۲ فشرده‌سازی اطلاعات

عملیات فشرده‌سازی اطلاعات، چند مزیت و کاربرد برای سامانه دارد:

۱. فشرده‌سازی نقش بسیار زیادی در کاهش هزینه‌های ورودی و خروجی پرس‌وجوها دارد.
۲. فشرده‌سازی بدلیل اینکه فضای ذخیره‌سازی را کاهش می‌دهد، می‌تواند هزینه‌های ذخیره‌سازی و پشتیبان‌گیری را نیز کاهش دهد.
۳. فشرده‌سازی به طور موثر میزان داده‌های ذخیره شده در حافظه را کاهش می‌دهد؛ زیرا صفحه‌ها را می‌توان به صورت فشرده نگه داشت و فقط در صورت نیاز از حالت فشرده خارج کرد.
۴. برخی از عملیات رایج پرس‌وجو (به عنوان مثال، شرایط برابری و حذف تکراری) اغلب می‌توانند بر روی خود داده‌های فشرده بدون نیاز به اینکه از حالت فشرده خارج شوند، انجام شود.

در نهایت، فشرده‌سازی داده‌های منتقل شده از طریق شبکه به طور موثر پهنای باند شبکه موجود را افزایش می‌دهد. این برای سامانه‌های مدیریت پایگاه‌داده‌های موازی که در آن داده‌ها باید بین گره‌ها منتقل شوند، مهم است. این عملیات فشرده‌سازی

<sup>۵۱</sup> partitioning

<sup>۵۲</sup> Hash

<sup>۵۳</sup> Range

<sup>۵۴</sup> column-oriented

<sup>۵۵</sup> row-oriented

نه تنها در این سامانه‌ها، بلکه در موتورهای تخصصی دیگر مثل موتورهای تحلیل پردازش‌های برخط نیز قابل مشاهده است. در ادامه به چند مهارت<sup>۵۶</sup> برای فشردسازی داده‌ها برای سامانه‌های مدیریت پایگاه‌داده‌های رابطه‌ای، اشاره خواهیم کرد.

#### ۴-۲-۱ سرکوب پوچ‌ها<sup>۵۷</sup>

این را می‌دانیم که چندین نوع داده رایج در سیستم‌های مدیریت پایگاه‌داده، دارای طول ثابت هستند؛ به عنوان مثال مقدار متغیر برای اعداد صحیح<sup>۵۸</sup> و تاریخ زمان<sup>۵۹</sup>. حال اگر همین نوع داده‌ها به عنوان طول متغیر برای اهداف ذخیره‌سازی در نظر گرفته شوند، عملیات فشردسازی را به طور چشمگیری امکان‌پذیر می‌کنند. در این روش، فقط قسمت غیر پوچ<sup>۶۰</sup> به همراه طول واقعی مقادیر را ذخیره می‌کند [۱].

مشابه این فرایند در ذخیره‌سازی شاخص‌ها در موتورهای جست‌وجو انجام می‌شود؛ در آنجا آدرس شاخص ساخته شده را با استفاده از مقادیر ثابتی ذخیره می‌کنند و این باعث می‌شود که حجم بالایی از داده‌ها را با فضای کمتری ذخیره کنیم [۵].

#### ۴-۲-۲ فشردسازی فرهنگ لغات<sup>۶۱</sup>

در این مهارت با شناسایی مقادیر تکراری در داده‌ها و ساخت فرهنگ لغاتی<sup>۶۲</sup> که چنین مقادیری را به صورت فشردتر در خود ذخیره کرده است، به نوعی عملیات فشردسازی را انجام داده است. به عنوان مثال، ستونی که حالت حمل و نقل را برای یک سفارش ذخیره می‌کند ممکن است حاوی مقادیر رشته‌ای مانند «کامیون»، «قایق» و «هواپیما» باشد. هر مقدار را می‌توان با استفاده از دو بیت با نگاشت آنها به مقادیر ۰، ۱ و ۲ نمایش داد.

#### ۴-۲-۳ رمزگذاری طول اجرا<sup>۶۳</sup>

برخلاف طرح‌های فشردسازی در ذخیره کردن جداول ردیف‌گرا که در آن هر نمونه از یک مقدار نیاز به ورودی دارد، در ذخیره‌سازی جداول ستون‌گرا، سایر تکنیک‌های فشردسازی مانند رمزگذاری طول اجرا (RLE) می‌توانند مؤثرتر شوند. در فشردسازی RLE، دنباله‌ای از  $k$  نمونه با مقدار  $v$  توسط جفت  $(v, k)$  کدگذاری می‌شود. RLE به ویژه زمانی جذاب است که دوره‌های طولانی با همان ارزش اتفاق بیفتد. این می‌تواند برای ستون‌هایی با مقادیر نسبتاً کمی متمایز یا زمانی که مقادیر ستون مرتب شده‌اند اتفاق بیافتد [۱].

<sup>۵۶</sup> Technique

<sup>۵۷</sup> Null suppression

<sup>۵۸</sup> Int

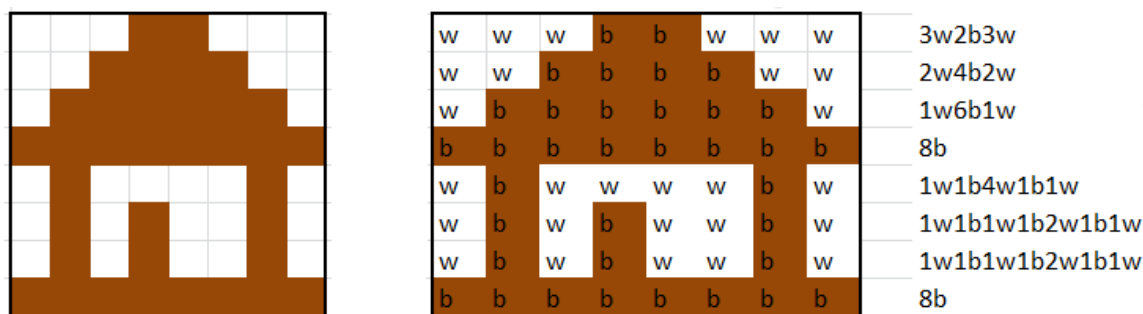
<sup>۵۹</sup> Datetime

<sup>۶۰</sup> Null

<sup>۶۱</sup> Dictionary compression

<sup>۶۲</sup> Dictionary

<sup>۶۳</sup> Run-length encoding (RLE)



شکل ۴-۱ مثالی از الگوریتم RLE

اگر ما دارای  $n$  مقدار متفاوت باشیم، آنگاه رابطه‌ی RLE برای رشته‌ی  $L$  به صورت زیر در می‌آید که در آن  $v_i$  ها عبارت‌های افراز شده‌ی رشته‌ی  $L$  و  $k_i$  تعداد تکرار افراز  $i$  ام است.

$$RLE(L) = \sum_i^n \{k_i, v_i\} \rightarrow k_i = \text{number of } v_i \text{ in string } L$$

برای درک بهتر رابطه‌ی بالا، به مثال زیر توجه کنید. اگر ما رشته‌ای به نام  $L$  داشته باشیم، آنگاه  $RLE(L)$  به صورت زیر در می‌آید:

$$\begin{aligned} L &= \{v_i, v_i, v_{i+1}, v_{i+1}, v_{i+1}, v_{i+1}, v_{i+1}, v_{i+1}, v_{i+2}, v_{i+2}\} \\ &\rightarrow RLE(L) = \{\{k_i, v_i\}, \{k_{i+1}, v_{i+1}\}, \{k_{i+2}, v_{i+2}\}\} \\ L &= \{5, 5, 8, 9, 9, 9, 9, 9, 30, 30, 30, 22, 22, 22, 22, 22, 22, 12, 12\} \\ &\rightarrow RLE(L) = \{\{2, 5\}, \{1, 8\}, \{6, 9\}, \{3, 30\}, \{6, 22\}, \{2, 12\}\} \end{aligned}$$

در این روش، در ابتدا باید یک دنباله به طول  $n$  را بعد از مرتب‌سازی، با توجه مقادیر مشابه افراز می‌کنیم. سپس تعداد افرازهای یکسان را برابر با مقدار  $k$  قرار می‌دهیم و افراز مورد نظر را برابر با  $v$ . آنگاه می‌توانیم یک نگاشت  $RLE(v, k)$  داشته باشیم.

با وجود چنین مهارت‌هایی، باز هم چندین چالش فنی جالب در فشرده‌سازی داده‌ها وجود دارد:

- اول، تکنیک‌های فشرده‌سازی جدید مناسب برای انبارهای داده بزرگ بوده و یک مبادله<sup>۶۴</sup> قابل قبول بین هزینه‌های رفع فشرده‌سازی و به‌روزرسانی‌های مهم ایجاد می‌کنند.
- دوم، حتی برای تکنیک‌های فشرده‌سازی شناخته‌شده، مشکلات باز مهم باقی می‌مانند؛ برای مثال، برای RLE، انتخاب ترتیب مرتب‌سازی جدول می‌تواند به طور قابل توجهی بر میزان فشرده‌سازی تأثیر بگذارد.

تعیین بهترین ترتیب مرتب‌سازی برای استفاده، یک مسئله بهینه‌سازی غیر پیش پا افتاده است. در نهایت، تصمیم‌گیری در مورد فشرده‌سازی ساختارهای دسترسی به حجم کاری بستگی دارد. بنابراین، نیاز به ابزارهای طراحی فیزیکی خودکار وجود دارد که بر اساس اطلاعات بار کاری، آیا ساختارهای دسترسی باید فشرده شوند و چگونه این فشرده‌سازی انجام شود [۳]؟

## ۴-۳ پردازش دستورات

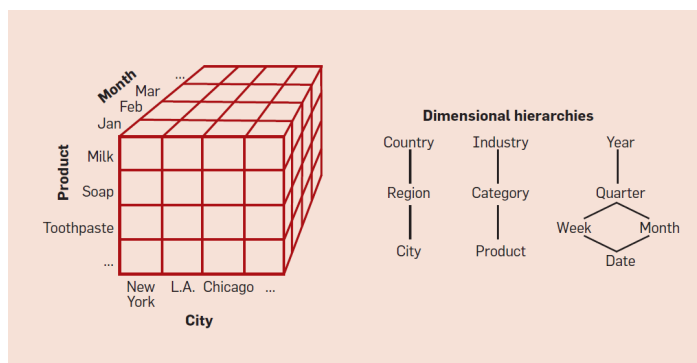
یک مدل مفهومی محبوب که برای وظایف هوش تجاری استفاده می‌شود، استفاده از نمای چند بعدی داده‌ها است، همانطور که در شکل ۴-۲ نشان داده شده است. در یک مدل داده چند بعدی، مجموعه‌ای از معیارهای عددی وجود دارد که موضوع تجزیه

<sup>۶۴</sup> Trade-off



و تحلیل هستند. نمونه‌هایی از این اقدامات عبارتند از فروش، بودجه، درآمد و موجودی. هر یک از معیارهای عددی با مجموعه‌ای از ابعاد مرتبط است که زمینه را برای اندازه‌گیری فراهم می‌کند. عنوان مثال، ابعاد مرتبط با مبلغ فروش می‌تواند محصول، شهر و تاریخ انجام فروش باشد.

بنابراین، یک اندازه‌گیری را می‌توان به عنوان یک مقدار در فضای چند بعدی ابعاد مشاهده کرد. هر بعد با مجموعه‌ای از ویژگی‌ها توصیف می‌شود، برای مثال، بعد محصول ممکن است از ویژگی‌های زیر تشکیل شود: دسته، صنعت، شماره مدل، و سال معرفی آن. [۱]



شکل ۴-۲ یک نمایش از داده‌های چند بعدی

ویژگی‌های یک بعد ممکن است از طریق سلسله مراتبی از روابط مرتبط باشند؛ به عنوان مثال، یک محصول از طریق یک رابطه سلسله مراتبی به دسته خود و ویژگی‌های صنعت مربوط می‌شود (شکل ۴-۲). یکی دیگر از ویژگی‌های متمایز مدل مفهومی، تاکید آن بر تجمیع معیارها توسط یک یا چند بعد است. به عنوان مثال، محاسبه و رتبه بندی کل فروش بر اساس هر شهرستان برای هر سال در این روش امکان‌پذیر است.

## ۴-۴ سرورهای رابطه‌ای

سرورهای پایگاه داده رابطه‌ای به طور سنتی به عنوان پشتیبان انبارهای داده بزرگ عمل می‌کنند. چنین انبارهای داده‌ای باید بتوانند پرس‌وجوهای پیچیده SQL را به بهترین نحو ممکن در برابر پایگاه‌های داده بسیار بزرگ اجرا کنند. اولین فناوری کلیدی مورد نیاز برای رسیدن به این هدف، بهینه‌سازی پرس‌وجو است که یک پرس‌وجو پیچیده را می‌گیرد و آن پرس‌وجو را در یک برنامه اجرا می‌کند.

برای اطمینان از اینکه طرح اجرا می‌تواند به خوبی در پایگاه‌های داده بزرگ مقیاس شود، قسمت‌بندی داده‌ها و پردازش پرس‌وجوی موازی به طور گسترده‌ای مورد استفاده قرار می‌گیرند. بنابراین ما دو بخش از فناوری کلیدی را مورد بحث قرار می‌دهیم؛ بهینه‌سازی پرس‌وجوها و پردازش موازی پرس‌وجوها.

## ۴-۴-۱ بهینه‌سازی پرس‌وجو

بهینه‌سازی پرس‌وجوها فناوری است که کلید دستیابی به هوش تجاری است. بهینه‌ساز پرس‌وجو مسئول انتخاب یک برنامه اجرایی<sup>۶۵</sup> برای پاسخ دادن به یک درخواست است. این مهارت وظیفه دارد تا یک طرح ترکیبی از عملگرهای مختلف مانند مرتب‌سازی، اسکن کردن و غیره را هنگام ارزیابی درخواست اعمال کرده و نتایج پرس‌وجو را تولید کند. [۱].

<sup>۶۵</sup> execution plan

یکی دیگر از وظایف آن تغییر ترتیب عملیات‌های منطقی<sup>۶۶</sup> روی درخواست کاربران است. به عنوان مثال برای عملیات‌هایی که شامل عملگر اشتراک<sup>۶۷</sup> می‌باشند، ترتیب عملیات را به گونه‌ای رقم می‌زند تا بتواند فرایند جستجو را تا حد مطلوبی کاهش دهد. با یک تغییر مناسب در ترتیب انجام عملیات‌های منطقی می‌توان تا حد خوبی از اتلاف زمان در سامانه جلوگیری کرد.

از دیگر وظایف بهینه ساز پرس و جو انجام عملیات‌های ریشه‌گیری<sup>۶۸</sup> و عادی‌سازی<sup>۶۹</sup> جستجو است؛ بدین صورت که کاربران آنچه را که نیاز دارند در سامانه وارد می‌کنند. سپس سامانه با انجام عملیات‌های فوق، پرس‌وجوی آنها را پردازش کرده و نتایج مشابهی را برای ایشان باز می‌گرداند. در فرایند ریشه‌گیری و عادی‌سازی، مقداری از کلماتی که کاربران وارد کردند توسط سامانه صرف نظر می‌شود و همین باعث می‌شود که نتایج مشابه بیشتری را تولید کند [۵].

این فرایند در موتورهای جستجو بسیار کاربردی است؛ اما باید دید که در باقی موتورها چه نتیجه‌ای را خواهد داشت. بررسی این موضوع بسیار با جنس داده‌هایی که در کسب و کارهای مختلف تولید می‌شود، ارتباط مستقیم دارد. در برخی از کسب و کارها ما نیاز داریم که نتایج کاملاً با پرسشی وارد شده تطابق داشته باشد، بدیهی است که در چنین سامانه‌هایی استفاده از فناوری بالا اصلاً بهینه نیست و نتایج مورد رضایت واقع نخواهد شد.

#### ۴-۴-۲ پردازش موازی پرس‌وجو

موازی سازی نقش مهمی در پردازش پرس و جوها در پایگاه‌داده‌های عظیم ایفا می‌کند. عملگرهای رابطه‌ای مانند انتخاب، طرح ریزی، پیوستن و تجمیع، فرصت‌های زیادی برای موازی‌سازی ارائه می‌دهند. ایده اصلی، موازی سازی داده‌ها است، یعنی عملگرهای رابطه‌ای به صورت موازی بر روی زیرمجموعه‌های متمایز از داده‌ها که در قسمت ۴-۱ صبحت شد، اعمال شده و سپس نتایج را با یکدیگر ترکیب کنیم. چندین سال است که تمامی فروشندگان عمده سیستم‌های مدیریت پایگاه داده، فناوری قسمت‌بندی داده‌ها و پردازش پرس‌وجوی موازی را ارائه کرده‌اند.

دو معماری اساسی برای موازی سازی وجود دارد:

- حافظه مشترک<sup>۷۰</sup>: در آن هر پردازنده<sup>۷۱</sup> یک حافظه اصلی<sup>۷۲</sup> خصوصی دارد اما حافظه‌های جانبی<sup>۷۳</sup> را با تمام پردازنده‌های دیگر به اشتراک می‌گذارد.
- هیچ چیز مشترکی<sup>۷۴</sup>: جایی که هر پردازنده دارای حافظه اصلی و حافظه جانبی خصوصی است و معمولاً یک ماشین ارزان قیمت است.

جالب توجه است، در حالی که قدمت این معماری‌ها به حدود دو دهه می‌رسد، هیچ کدام هنوز به عنوان یک برنده آشکار در صنعت ظاهر نشده‌اند و امروزه اجرای موفق هر دو وجود دارد [۱].

<sup>۶۶</sup> Logical operations

<sup>۶۷</sup> Intersection

<sup>۶۸</sup> Stemming

<sup>۶۹</sup> Normalization

<sup>۷۰</sup> Shared disk

<sup>۷۱</sup> CPU

<sup>۷۲</sup> RAM

<sup>۷۳</sup> Secondary storage

<sup>۷۴</sup> Shared nothing

## نتیجه‌گیری

از کلیدی‌ترین واحدهای هوش تجاری، واحد ذخیره و بازیابی اطلاعات است. وظیفه این واحد طراحی و پیاده‌سازی روش‌هایی است که بتوان به وسیله آنها حجم استفاده شده از حافظه و زمان برای پاسخ به پرس‌وجوی کاربران را کاهش داد. با توجه به نکات گفته شده، این فناوری نیازمند دانش عمیقی در حوزه‌های مختلف مثل مدیریت پایگاه‌های داده، طراحی موتورهای جستجو و طراحی و پیاده‌سازی الگوریتم‌های فشرده‌سازی می‌باشد.

قدم بعدی برای این توسعه این فناوری، طراحی یک ماشین با رویکرد هوش مصنوعی است که عملیات خودکارسازی<sup>۷۵</sup> فرایندهای بالا را با استفاده از داده‌ها و دانش‌های قبلی و پرس‌وجوهای جدید را انجام دهد. رویکرد در طراحی این ماشین به این صورت خواهد بود که بتواند با استفاده از درخواست‌های کاربران، مدل‌های بهینه‌تری را برای حوزه‌های فوق ارائه کرده و سپس اقدام به طراحی و پیاده‌سازی آنها نماید. به عنوان مثال، هدف طراحی یک نرم افزار مدیریت پایگاه‌های داده‌ای است که خود بتواند با توجه به نوع داده‌های کسب و کار، مدل بهینه را تشخیص داده و سامانه را مطابق با آن تغییر دهد.

---

<sup>۷۵</sup> Automation

## فصل پنجم

### نتیجه گیری و پیشنهادها

## نتیجه‌گیری

در این گزارش به بررسی اجمالی هوش تجاری و معرفی برخی از اجزای آن پرداخته شد؛ هوش تجاری یک فناوری پشتیبان از تصمیم است که به وسیله‌ی آن، انجام تصمیم‌گیری‌های کلان برای شرکت‌ها و کسب و کارها ساده‌تر و هوشمندانه‌تر خواهد شد. هوش تجاری به ما کمک می‌کند تا بتوانیم تصویر واضح‌تر از آینده کسب و کار خود و وضعیت بازار رقابتی داشته باشیم. اما اولین چالش در استفاده از این فناوری، بالا بودن هزینه‌های استفاده از آن به دلیل نیازمند بودن به دانش‌های تخصصی مختلف می‌باشد. هدف این گزارش بررسی اجمالی فناوری فوق به جهت پیدا کردن راهکاری برای ساده‌سازی و خودکارسازی این فرایندها است تا هزینه استفاده از آن کاهش پیدا کرده و کسب و کارهای کوچک‌تر نیز بتوانند از هوش تجاری استفاده کنند [۶].

همانطور که گفته شد هوش تجاری در تحقیق و صنعت امروز بسیار محبوب و دارای فراز و نشیب‌های فراوان است. با استفاده از آن جمع‌آوری داده‌ها آسان‌تر شده و انبارهای داده بزرگی با حجم ۱۰ تا ۱۰۰ ترابایت یا حتی بیشتر برای داده‌های رابطه‌ای استفاده می‌شوند. داده‌های متنی نیز به عنوان یک منبع ارزشمند هوش تجاری مورد استفاده قرار می‌گیرند. تغییرات در فناوری سخت‌افزاری مانند کاهش هزینه حافظه اصلی بر نحوه معماری انبارهای داده بزرگ تأثیر می‌گذارد. علاوه بر این، با ورود سرویس‌های داده ابری، تغییرات بیشتری در معماری هوش تجاری انتظار می‌رود [۷].

در نهایت، تقاضای فزاینده‌ای برای ارائه تجربیات تعاملی هوش تجاری در دستگاه‌های تلفن همراه وجود دارد. فرصت‌های زیادی برای فعال کردن برنامه‌های جدید، غنی و تعاملی هوش تجاری در نسل بعدی دستگاه‌های تلفن همراه وجود دارد. بنابراین، نرم‌افزار هوش تجاری چالش‌ها و فرصت‌های فنی بسیار هیجان‌انگیزی دارد که همچنان به تغییر شکل چشم‌انداز و عملکرد خود ادامه خواهند داد [۱].

## پیشنهادها

اگر چه تحقیقات در حوزه‌های مالی به صورت گسترده‌ای در حال انجام است، اما یکی از موضوعاتی که بسیار می‌تواند نقش تحول‌آفرینی در حوزه اقتصاد کشور ایفا کند، پیوند علوم کامپیوتری و هوش مصنوعی با این حوزه است. موضوعی که اخیراً بسیار مورد توجه سیاستمداران و دانشمندان این عرصه قرار گرفته است. استفاده از الگوریتم‌های یادگیری ماشین برای طراحی مدل‌های بهینه‌تر، ایجاد رابط کاربری ساده‌تر و قابل فهم‌تر، نمونه‌هایی از پیوند دانش‌های علوم کامپیوتری با صنعت کشور است. پیوندی که می‌توان به وسیله آن به رشد کسب و کارهای کوچک کشور کمک کرد و در ایجاد شغل و رونق اقتصادی کشور گامی بزرگ برداشت.

با تبدیل نقاط قوت کسب و کارها از اشخاص به ابزارها، می‌توان در ایجاد مشاغل خرد و کسب و کارهای کوچک‌تر برای افراد دارای بضاعت مالی کمتر، کمک بسیار بزرگی کرد. وجود ابزارهای قدرتمندی که توان انجام کارهای پیچیده و دارای دانش‌های بسیار تخصصی و عمیق را داشته باشد، می‌تواند در این مسیر بسیار کمک‌کننده باشد.

اما ترس احتمالی از کنار رفتن دانشمندان حوزه هوش تجاری در شرکت‌ها و کسب و کارهای بزرگ، اصلاً درست نیست؛ زیرا تمامی شرکت‌ها و کسب و کارهای بزرگ از الگوریتم‌ها و فرایندهای هوش تجاری مخصوص به خودشان استفاده می‌کنند. همان‌طور که در گزارش اشاره شد، الگوریتم‌های هوش تجاری بسیار وابسته به نوع داده‌های موجود در یک کسب و کار است و

---

همین موضوع فرایندهای استفاده از این داده را در کسب و کارهای مختلف سخت‌تر می‌کند. لذا در کسب و کارهای بزرگ نیاز به دانشمندان این حوزه بسیار ملموس خواهد بود.

هوش تجاری می‌تواند در آینده به صورت یک سکو<sup>۷۶</sup> برای شرکت‌ها استفاده بشود به گونه‌ای که شرکت‌ها و کسب و کارهای کوچک می‌توانند از سکوه‌ای آماده و شرکت‌های بزرگ از سکوه‌ای خصوصی‌سازی شده و شخصی‌سازی شده خودشان استفاده کنند و سکوه‌ای هوش تجاری، می‌تواند یک محصول تجاری برای متخصصان این عرصه باشد.

با توجه به نکات گفته شده امکان برکناری متخصصان هوش تجاری پس از دستیابی به مدل‌های هوشمند شده این فناوری وجود ندارد؛ زیرا این افراد مسئولین توسعه این پروژه خواهند بود. این افراد هستند که می‌توانند مسیر اقتصادی جوامع و کشورها را بگونه‌ای که می‌خواهند تغییر دهند و عملاً چرخ‌های اقتصاد فردای جامعه را دانشمندان حوزه هوش تجاری و امور مالی به دست خواهند گرفت.

---

<sup>۷۶</sup> Platform

## مراجع

- [1]. Chaudhuri, S., et al. (2011). "An overview of business intelligence technology." *Communications of the ACM* **54**
- [2]. Built In, 2021. Online Community for Startups and Tech Companies. [online] Available at: <https://builtin.com/salaries/data-analytics/business-intelligence-analyst> [Accessed 19 May 2022]
- [3]. Ranjan, J. (2009). "Business intelligence: Concepts, components, techniques and benefits." *Journal of theoretical and applied information technology* **9**(1): 60-70.
- [4]. Chaudhuri, S. and U. Dayal (1997). "An overview of data warehousing and OLAP technology." *ACM Sigmod record* **26**(1): 65-74.
- [5]. Manning, C.D., Raghavan, P., Schütze, H., 2008. *Introduction to Information Retrieval*. London: Cambridge University Press.
- [6]. Turban, E., et al. (2008). *Business intelligence: A managerial approach*, Pearson Prentice Hall Corydon eIndiana Indiana.
- [7]. Watson, H. J. (2009). "Tutorial: business intelligence—past, present, and future." *Communications of the Association for Information systems* **25**(1): 39