

# Review and Prioritization of the Risks Affecting the Use of Organizational Big Data

**Sadra Ahmadi\***

Phd in Information Technology; Assistant Professor; CyberSpace Research Institute; Shahid Beheshti University; Tehran, Iran;  
Email: sa\_ahmadi@sbu.ac.ir

**Sanaz Ghorbanloo**

MSc Student in Information Technology Management; Faculty of Management and Accounting, Department of Industrial Management and Information Technology; Shahid Beheshti University; Tehran, Iran Email: s.ghorbanloo@mail.sbu.ac.ir

Received: 14, Jul. 2021 Accepted: 10, Nov. 2021

**Abstract:** Organizations produce massive volumes of variant and valuable data on a daily basis. Identifying the risks of using and processing organizational big data and finding solutions to manage and reduce these risks, paves the way for organizations to succeed and gain a competitive advantage. On the other hand, in organizations, due to the dynamics of the business environment, the factors affecting the reduction of risks affect each other, and a change in the status of one factor will directly or indirectly affect other factors. For this purpose, by reviewing the thematic literature and the opinions of experts, a framework consisting of 4 influential areas and 44 effective factors was formed. According to the qualitative nature of the influencing factors, fuzzy concepts have been used for modeling and analysis. Furthermore, to model the relationships between factors and determine the most effective and influential factor in organizations, the FUZZY DEMATEL technique has been used. In this article, the modelling is based on the opinion of experts and subsequently, the results of modeling and calculations have been evaluated with the opinion of experts. By using the achievements and modelling methods mentioned in this research and according to the needs of the organization, they can model and plan to minimize the risks of using their organizational big data as much as possible.

**Keywords:** Big Data, Big Data Lifecycle, Dematel Fuzzy Technique, Big Data Risk, Capability

Iranian Journal of  
**Information  
Processing and  
Management**

Iranian Research Institute  
for Information Science and Technology  
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 37 | No. 4 | pp. 1183-1216

Summer 2022

<https://doi.org/10.35050/JIPM010.2022.007>



\* Corresponding Author

# بررسی و اولویت‌بندی مخاطره‌های تأثیرگذار در استفاده از کلان داده‌های سازمانی

صدرا احمدی

دکتری فناوری اطلاعات، استادیار؛ پژوهشکده فضای  
مجازی؛ دانشگاه شهید بهشتی؛ تهران، ایران؛  
پدیداور رابط sa\_ahmadi@sbu.ac.ir

ساناز قربانلو

دانشجوی کارشناسی ارشد مدیریت فناوری اطلاعات؛  
پژوهشکده مدیریت و حسابداری؛ گروه مدیریت  
صنعتی و فناوری اطلاعات؛ دانشگاه شهید بهشتی؛  
تهران، ایران s.ghorbanloo@mail.sbu.ac.ir



دریافت: ۱۴۰۰/۰۴/۲۳ پذیرش: ۱۴۰۰/۰۸/۱۹ مقاله برای اصلاح به مدت ۳۵ روز نزد پدیداوران بوده است.

نشریه علمی | رتبه بین‌المللی  
پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایرانداک)

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۸۳۳۱-۲۲۵۱

نما به در .LISTA، .ISC، و SCOPUS

jipm.irandoc.ac.ir

دوره ۳۷ | شماره ۴ | صص ۱۱۸۳-۱۲۱۶

تابستان ۱۴۰۱

<https://doi.org/10.35050/JIPM010.2022.007>



**چکیده:** سازمان‌ها روزانه حجم زیادی از داده‌های متفاوت و ارزشمند تولید می‌کنند. شناسایی مخاطره‌های استفاده و پردازش کلان داده‌های سازمانی و یافتن راهکارهایی جهت مدیریت و کاهش این مخاطرات، راه را برای موفقیت سازمان‌ها و دست یافتن به مزیت رقابتی باز می‌کند. از سوی دیگر، به دلیل پویایی محیط کسب و کار، عوامل اثرگذار بر کاهش مخاطرات بر روی یکدیگر در سازمان‌ها تأثیرگذار هستند و تغییر در وضعیت یک عامل به صورت مستقیم یا غیرمستقیم بر عوامل دیگر تأثیرگذار خواهد بود. هدف از تحقیق پیش رو، افزون‌بر شناسایی عوامل مؤثر بر کاهش مخاطره‌های ناشی از استفاده کلان داده‌های سازمانی، تحلیل ارتباط بین عوامل و در نهایت، اولویت‌بندی عوامل با در نظر گرفتن رابطه بین آنهاست. به همین منظور، با بررسی مرور ادبیات موضوعی و نظرات خبرگان، چارچوبی متشکل از ۴ حوزه تأثیرگذار و ۴۴ عامل مؤثر تشکیل گردید. با توجه به ماهیت کیفی عوامل تأثیرگذار از مفاهیم فازی جهت مدل‌سازی و تحلیل استفاده شده و به منظور مدل‌سازی روابط میان عوامل و تعیین تأثیرگذارترین و تأثیرپذیرین عامل در سازمان‌ها از تکنیک «دیمتل فازی» استفاده شده است. مدل‌سازی‌های انجام‌شده در این مقاله بر اساس نظر خبرگان است و به دنبال آن، نتایج مدل‌سازی و محاسبات با نظرات خبرگان ارزیابی گردیده است. سازمان‌ها با استفاده از دستاوردهای این پژوهش و روش مدل‌سازی اشاره‌شده در این تحقیق می‌توانند مدل‌سازی

متناسب با شرایط خود را انجام داده و نسبت به برنامه‌ریزی جهت کاهش هر چه بیشتر مخاطرات استفاده از کلان‌داده‌های سازمانی خود اقدام نمایند.

**کلیدواژه‌ها:** کلان‌داده، چرخه حیات کلان‌داده، تکنیک فازی دیمتل، مخاطره کلان‌داده، قابلیت

## ۱. مقدمه

امروزه، داده‌ها به‌عنوان یکی از دارایی‌های مهم و حیاتی در فرایند تصمیم‌گیری سازمان‌ها محسوب می‌شوند، به‌طوری‌که استفاده از داده‌ها در سازمان‌ها امری اجتناب‌ناپذیر است. بدون دسترسی به داده‌هایی با کیفیت بالا فراهم کردن اطلاعات مناسب در زمان مناسب، طراحی، نظارت، و ارزیابی خط‌مشی‌های مؤثر تقریباً غیرممکن است (Wang, Zeshui & Witold 2017). از سوی دیگر، با گسترش روزافزون استفاده از فناوری اطلاعات، اینترنت، دستگاه‌های دیجیتال و نرخ بالای تولید اطلاعات، حجم عظیمی از داده‌ها در حال تولید است. بر اساس گزارش منتشرشده از سوی «گروه تحقیقاتی بین‌المللی داده‌ها»<sup>۱</sup>، میزان کل داده‌های تولید، ذخیره، و مصرف‌شده در جهان به‌سرعت در حال افزایش است، به‌طوری‌که در سال ۲۰۲۰، به بالاترین میزان خود، در حدود ۶۴/۲ زتابایت رسیده است. با توجه به این‌که حجم داده‌ها با سرعت غیرقابل‌تصور در حال افزایش است، پیش‌بینی می‌شود که طی پنج سال آینده (تا سال ۲۰۲۵) حجم داده‌های تولیدشده در جهان به بیش از ۱۸۰ زتابایت افزایش یابد (Holst 2021). این حجم بسیار زیاد داده، ما را با مفهوم و واژه جدیدی تحت عنوان «کلان‌داده»<sup>۲</sup> روبه‌رو ساخته است؛ داده‌های بزرگی که چند سالی است در زمینه فناوری اطلاعات حضور پیدا کرده و با اشاره به حجم بزرگی از داده‌ها که توسط سازمان‌های بزرگ، مانند «گوگل» و «ناسا» ذخیره و تحلیل می‌شوند، مورد استفاده قرار می‌گیرند (Sivarajah et al. 2017). واژه «کلان» در کلان‌داده تنها به حجم مربوط نیست؛ در حالی‌که کلان‌داده به‌طور قطع، دربرگیرنده داده‌های زیادی است. اما اصطلاح کلان‌داده تنها به حجم اشاره ندارد؛ یعنی در صورتی‌که مسئله‌ای کلان‌داده باشد، تنها بحث تحلیل حجم انبوهی از داده‌ها مطرح نیست، بلکه داده‌ها با سرعت تولید می‌شوند و در قالب‌های پیچیده‌ای از منابع داده گوناگون قرار می‌گیرند (Sonkar 2020). در گذشته، سازمان‌ها دارای سنت طولانی در زمینه ثبت داده‌های تراکشی بودند، در حالی‌که امروزه،

1. International Data Corporation (IDC)

2. big data

سازمان‌ها داده‌های افزوده‌ای را نیز از محیط عملیاتی دریافت و با سرعتی در حال افزایش ثبت می‌کنند که از آن جمله می‌توان به داده‌های وب، داده‌های متنی، داده‌های زمانی و مکانی، شبکه‌های هوشمند و داده‌های حسگرها، داده‌های شبکه‌های اجتماعی و غیره اشاره نمود (Borkar, Carey and Li 2012).

از دیدگاه سازمانی، همان‌طور که سازمان‌ها رشد پیدا می‌کنند، داده‌های مرتبط با آن‌ها نیز به‌صورت نمادین افزایش یافته و امروزه، بسیاری از پیچیدگی‌های مربوط به داده‌های آن‌ها (از تولید تا تحلیل داده‌ها) به یک چالش بزرگ تبدیل شده است (Oussous et al. 2018). طبق مطالعاتی که در پژوهش (McAfee et al. 2012) انجام شده، با توجه به تمایز کلان‌داده‌ها با داده‌هایی که در سازمان‌ها ایجاد می‌شود، سازمان‌هایی که از کلان‌داده‌ها برخوردار هستند، دارای سه ویژگی کلیدی هستند: ۱) در این سازمان‌ها به‌صورت روزانه حجم<sup>۱</sup> انبوهی از داده ایجاد می‌شود، ۲) شتاب تولید<sup>۲</sup> داده در این سازمان‌ها بسیار بالاست و تحلیل آنی داده این امکان را برای سازمان فراهم می‌آورد که بسیار چابک‌تر از رقبای خود باشند. برای مثال، در شرکتی مانند «المارت»<sup>۳</sup> در هر ساعت بیش از ۲/۵ پتابایت داده از تراکنش‌های مشتریان ایجاد می‌شود که این موضوع نشان‌دهنده حجم و سرعت بالای تولید اطلاعات در این شرکت است، و ۳) در این سازمان‌ها تنوع<sup>۴</sup> زیادی از داده وجود دارد که بایستی تحلیل شوند. این داده‌ها از منابع مختلفی همچون پیام‌ها، مطالب و تصاویر ارسال‌شده در شبکه‌های اجتماعی، اطلاعات دریافت‌شده از سنسورها و حتی سیگنال‌های GPS دریافت‌شده از تلفن‌های همراه مشتریان که نشان‌دهنده موقعیت فیزیکی آن‌هاست، دریافت می‌شود (Anshari et al. 2019). ساختار داده تولیدشده در هر کدام از این منابع با دیگران متفاوت است و نیاز است برای تحلیل این داده‌ها از ابزارهای مختلفی استفاده نمود. همچنین، با توجه به تنوع داده‌ها، پایگاه‌های داده ساخت‌یافته‌ای که تا پیش از این اطلاعات سازمان‌ها را ذخیره می‌کردند، برای ذخیره و پردازش کلان‌داده‌ها با چنین تنوع و حجم مناسب نیستند (Al-Yadumi et al. 2021).

مطالعات نشان می‌دهد که در طیف وسیعی از صنایع، استفاده از تجزیه و تحلیل کلان‌داده‌ها محرک اصلی موفقیت سازمانی است (McAfee et al. 2012)؛ به‌ویژه در سازمان‌هایی که کسب‌وکار اصلی آن‌ها مبتنی بر داده و فناوری اطلاعات است یا فناوری

اطلاعات و سیستم‌های جامع سازمانی به‌عنوان بخش جدایی‌ناپذیر سازمان ایفای نقش می‌کنند، نگهداری و تجزیه و تحلیل کلان‌داده بسیار حائز اهمیت است (Shabbir and Gardezi 2020). برای مثال، شرکت‌های خرده‌فروش آنلاین مانند «آمازون»، «ای‌بی»<sup>۱</sup> و یا Kogan.com که کسب‌وکار اصلی آن‌ها مبتنی بر فناوری اطلاعات و یا اینترنت است، روزانه حجم عظیمی از داده توسط فروشندگان و مشتریان در قالب محصولات که برای خرید یا فروش در نظر می‌گیرند و یا بازخوردهایی که از طریق نظرهای ثبت‌شده در سایت ایجاد می‌شود. همچنین، سازمان‌هایی که مدیریت شبکه‌های اجتماعی را بر عهده دارند، مانند «فیس‌بوک» و «توییتر»، روزانه بایستی حجم عظیمی از پست‌های افراد، تبلیغات و واکنش کاربران به تبلیغات را تحلیل نمایند. همچنین، در سازمان‌هایی مانند بیمارستان‌ها یا بانک‌ها روزانه حجم عظیمی از داده‌های ساخت‌یافته و غیرساخت‌یافته توسط بیماران یا مشتریان بانک ایجاد می‌شود (Palanisamy and Thirunavukarasu 2019). اغلب منابع داده سنتی ساخت‌یافته هستند. برای مثال، می‌توان به رسید خرید محصولات توسط مشتری از خواروبارفروشی، داده‌های پرداخت حقوق کارکنان یک سازمان، اطلاعات حسابداری ثبت‌شده در صفحه گسترده‌ها و هر آنچه به‌خوبی در پایگاه داده‌های رابطه‌ای قرار گرفته، اشاره نمود. در واقع، هر بخش از اطلاعات پیش از زمان تعیین‌شده شناسایی و در قالب خاصی ارائه می‌شود و همچنین، به‌ترتیبی خاص به وقوع می‌پیوندد. در مقابل، منابع داده غیرساخت‌یافته آن‌هایی هستند که کاربر کنترلی روی قالب آن‌ها ندارد. داده‌های متنی، داده‌های ویدیویی و صوتی همه در این دسته قرار می‌گیرند. کار با داده‌های غیرساخت‌یافته پیچیده است، زیرا معانی آن‌ها از پیش تعریف نشده است (Hu et al. 2014). در سازمان‌هایی مانند خطوط هوایی، هتل‌ها و یا شرکت‌های رزرو بلیط یا اتاق، روزانه حجم زیادی از اطلاعات توسط خریداران ایجاد می‌شود که بایستی تحلیل شوند و نتیجه تحلیل در اختیار فروشندگان سرویس قرار گیرد. اگرچه کلان‌داده‌ها ارزش بالایی دارند، ولی با توجه به این که منابع داده‌ای مختلفی در مکان‌های توزیع‌شده قرار دارند و آن‌ها تعیین‌کننده آسیب‌پذیری‌های امنیتی هستند، از این رو، سازمان‌ها باید پیش از استفاده از کلان‌داده‌ها، مسائل و چالش‌های مربوط به حریم خصوصی<sup>۲</sup> و امنیتی آن را نیز در نظر بگیرند (Raguseo 2018).

1. Ebay

2. privacy

در حال حاضر، بسیاری از داده‌های در دسترس بسیاری از سازمان‌های بزرگ با توجه به منابع تولیدشان از ساختارهای متنوعی برخوردارند و نمی‌توانند در پایگاه‌های داده ساخت یافته ذخیره و تحلیل شوند (Dash et al. 2019). از سوی دیگر، با توجه به این که سازمان‌ها با انبوهی از داده‌ها سروکار دارند و این حجم از داده‌ها با رشدی فزاینده همراه است، این سازمان‌ها نیازمند شناسایی مخاطره‌ها و به طبع کاهش آن‌ها هستند. در واقع، مخاطرات یا ریسک‌ها از جمله اجزایی هستند که در زمان تصمیم‌گیری سازمان‌ها برای سرمایه‌گذاری در راه‌حل‌های کلان‌داده‌ها باید مد نظر قرار گیرند (Lu et al. 2020) تا از این طریق، راه برای استفادهٔ بهینه از کلان‌داده‌ها در سازمان‌ها فراهم شود. این است که سازمان‌ها به دنبال راهکاری هستند که تا حد ممکن ریسک استفاده از کلان‌داده‌های خود را کاهش دهند.

با در نظر گرفتن این مهم که امروزه، کاهش چالش‌ها و مخاطره‌های کلان‌داده‌ها از دغدغه‌های اصلی در هر سازمان است، هدف اصلی این پژوهش، شناسایی و اولویت‌بندی عامل‌های تأثیرگذار در کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌های سازمانی در سطحی است که با چالش‌های قابل توجهی در ارتباط با حجم بسیار زیاد، سرعت تولید، تنوع داده‌ها و تغییرپذیری در ویژگی داده‌ها مواجه هستند و این چالش‌ها نیاز به مدیریت، فناوری‌ها و روش‌های حمل داده متمایزی را برای سازمان‌ها به منظور تحلیل و استفاده از آن‌ها مطرح می‌سازند. با توجه به هدف اصلی، اهداف فرعی مورد نظر در این پژوهش عبارت‌اند از: شناسایی مخاطره‌های کلان‌داده‌ها در سازمان‌ها و طبقه‌بندی آن‌ها در سطوح سازمانی. در راستای دستیابی به هدف پژوهش، سؤالاتی که در این باره مطرح می‌شوند، به شرح زیر است:

- ◇ عوامل مؤثر در کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌ها در سازمان‌ها کدام‌اند؟
- ◇ چه ارتباطی بین عوامل مؤثر شناسایی شده از لحاظ علی و معلولی برقرار است؟
- ◇ رتبه‌بندی عامل‌های مؤثر در کاهش مخاطره‌ها به چه صورت است؟

در این پژوهش، پس از معرفی چرخهٔ حیات کلان‌داده‌ها و ابعاد چهارگانهٔ توانایی‌ها و قابلیت‌های سازمانی، عامل‌هایی که در زیر هر یک از این ابعاد قرار خواهند گرفت، از طریق مصاحبه با خبرگان و متخصصان فعال حوزهٔ کلان‌داده‌ها در سازمان شناسایی شده‌اند، معرفی می‌شوند. در فرایند تحلیل مخاطره‌ها، ابتدا عامل‌های تحلیل مخاطره‌های

کلان داده‌ها معرفی می‌شوند. سپس، روابط و ضرایب علی و معلولی میان مخاطره‌های کلان داده‌ها به‌عنوان یکی از شاخص‌های تحلیل با به‌کارگیری تکنیک دیمتل فازی محاسبه و اولویت‌بندی می‌شوند. سرانجام، ضمن طبقه‌بندی آن‌ها، تأثیرگذارترین و تأثیرپذیرترین عامل شناسایی می‌شود.

## ۲. مبانی نظری و پیشینه پژوهش

### ۲-۱. مفهوم کلان داده و کلان داده سازمانی

امروزه، داده‌ها به‌عنوان عنصری روبه‌رشد در زندگی سازمان‌ها شناخته می‌شوند. همزمان با افزایش تولید داده‌ها، کاربرد آن‌ها فراگیرتر شده و منجر به شکل‌گیری مفهومی جدید به‌نام کلان داده‌ها گردیده است؛ مفهومی که نه تنها نحوه ارائه خدمات و محصولات سازمان‌ها و کسب و کارها، بلکه فرایند شناسایی، ذخیره‌سازی و تحلیل داده‌ها را دستخوش تغییر خواهد نمود. کلان داده‌ها یکی از امیدوارکننده‌ترین روندهای فناوری امروزی هستند. هرچند محور کسب و کار سازمان‌های بزرگی مانند «فیس‌بوک»، «گوگل» و «نتفلیکس»<sup>۱</sup> کلان داده‌هاست، اما بسیاری از بازارهای کوچک مانند خرده‌فروشی، بیمه و مراقبت‌های بهداشتی در حال وارد شدن به صحنه هستند (Baltzan 2019). تعاریف متنوعی از دیدگاه‌های مختلف در ارتباط با مفهوم کلان داده‌ها ارائه شده است که در قالب جدول ۱، به آن خواهیم پرداخت.

### جدول ۱. کلان داده و تعاریف آن

مؤلف/مراجع	تعریف	حوزه‌های تمرکز
Favaretto et al. (2020)	اصطلاح کلان داده به‌طور معمول، برای توصیف طیف وسیعی از مفاهیم مختلف مورد استفاده قرار می‌گیرد. این اصطلاح از جمع‌آوری و تجمیع حجم وسیعی از داده‌ها رفتاری گرفته تا انبوهی از تکنیک‌های دیجیتال پیشرفته را که برای نشان دادن الگوهای رفتاری طراحی شده‌اند، شامل است.	جمع‌آوری داده، تجمیع داده، تکنیک دیجیتال، الگوی رفتاری

1. Netflix

مؤلف/مراجع	تعریف	حوزه‌های تمرکز
Nguyen et al. (2019)	کلان‌داده‌ها را با چند کلمه که در انگلیسی با حرف (V) حجم، سرعت، تنوع، آغاز می‌شوند، تعریف می‌کنند. از جمله ویژگی حجم، صحت، تغییرپذیری، ارزش، سرعت و تنوع و همچنین صحت، تغییرپذیری، مصورسازی، مصورسازی ارزش و غیره.	
Raguseo (2018)	یک دارایی اطلاعاتی با حجم بالا، سرعت بالا و تنوع به‌موقع بودن، تنوع، بالاست که نیازمند انواع مقرون‌به‌صرفه و نوآورانه پردازش مقرون‌به‌صرفه، تحلیل اطلاعات برای دستیابی به اطلاعات بیشتر و تصمیم‌گیری داده، دارایی اطلاعاتی، بهتر است.	
Huang and Chaovalitwongse (2015)	مانند چتری هستند که مجموعه‌ای از داده‌های کلان و پیچیده مجموعه داده کلان و پیچیده، را در بر می‌گیرند؛ به‌گونه‌ای که ذخیره‌سازی، پردازش و ذخیره‌سازی، پردازش، تجزیه و تحلیل آن‌ها با استفاده از روش‌های سنتی و مرسوم تحلیل، روش سنتی امکان‌پذیر نیست.	
Khan et al. (2014)	اغلب برای توصیف داده‌های جریان بزرگ، پیچیده و در زمان واقعی، مدیریت پیچیده، زمان واقعی استفاده می‌شود که نیاز به مدیریت پیچیده، تحلیل، جریان کلان‌داده، تحلیلی و تکنیک‌های پردازش برای فهمیدن دارند. تکنیک پردازش	
Che, Mejd and Zhiyong (2013)	پدیده‌ای است که هدف آن ارائه یک جایگزین برای پایگاه‌داده، تحلیل داده، راه‌حل‌های سنتی بر اساس پایگاه‌داده و تجزیه و تحلیل شناسایی ارزش، راه حل داده‌هاست و صرفاً در مورد ذخیره‌سازی یا دسترسی به جایگزین، ذخیره‌سازی، داده‌ها نیستند، بلکه راه‌حل‌های آن به‌منظور تجزیه و تحلیل روش سنتی، دسترسی داده‌ها به‌منظور شناسایی آن‌ها و بهره‌برداری از ارزش آن‌هاست.	

با توجه به تعاریف ارائه شده از سوی پژوهشگران و مراجع علمی مختلف، در یک تعریف جامع‌تر از کلان‌داده‌ها می‌توان گفت: کلان‌داده‌ها، مجموعه‌ای بزرگ و پیچیده از داده‌ها شامل داده‌های ساخت‌یافته و غیرساخت‌یافته است که نمی‌تواند توسط پایگاه داده‌ها و روش‌های سنتی تحلیل شود. مدیران بیشتر درگیر تصمیم‌گیری‌های کسب‌وکارند؛ مانند این که چقدر از هر محصول تولید شود، چه مواد اولیه‌ای خریداری شود، یا چه تعداد کارمند با توجه به درک و بینش آن‌ها به خدمت گرفته شوند. در عصر اطلاعات مدیران موفق روزانه با حجم انبوهی از داده‌هایی که به آن‌ها جهت تصمیم‌گیری صحیح کمک می‌کند، روبه‌رو بوده و آن‌ها را تحلیل و پردازش می‌کنند.

از سوی دیگر، باید به این مهم اشاره نمود که اگرچه مزایای کلان‌داده‌ها واقعی و قابل توجه است، اما هنوز چالش‌های فراوانی وجود دارد که برای رسیدن به پتانسیل مورد نظر کلان‌داده‌ها باید به آن‌ها توجه کرد. برخی از این چالش‌ها به ویژگی‌های



کلان‌داده‌ها، برخی مربوط به مدل‌ها و روش‌های تجزیه و تحلیل و برخی نیز مربوط به محدودیت سیستم‌های پردازش جاری مربوط می‌شوند (Sivarajah et al. 2017). اگر نتوانیم این چالش‌ها را برطرف کنیم، کلان‌داده‌ها به یک سنگ معدن طلا تبدیل خواهند شد که ما قادر به کشف آن نیستیم، به‌ویژه هنگامی که اطلاعات بیش از توانایی ما برای مهار کردن باشد (Fikri et al. 2019; Mikalef et al. 2020).

## ۲-۲. چالش‌های مطرح درباره کلان‌داده‌ها

با مطالعه مقالات مختلف می‌بینیم که چالش‌های کلان‌داده‌ها به‌صورت پراکنده مورد بررسی قرار گرفته و دسته‌بندی‌های مختلفی برای آن در نظر گرفته شده است. برای دستیابی به هدف پژوهش، تحقیقی جامع در مقالات مختلف انجام شد که شامل مقالات مجلات علمی و کنفرانس‌ها در پایگاه‌های علمی معتبر است.

«حریری، فردریکس و باورز» در پژوهش خود تحت عنوان «عدم قطعیت در تجزیه و تحلیل کلان‌داده‌ها: بررسی فرصت‌ها و چالش‌ها» به این موضوع می‌پردازند که تجزیه و تحلیل کلان‌داده‌ها با استفاده از هوش مصنوعی طیف وسیعی از چالش‌ها را به‌خصوص زمانی که در معرض عدم قطعیت قرار می‌گیرند - با خود به همراه دارند. به‌طوری که این عدم قطعیت می‌تواند در کل فرایند تجزیه و تحلیل (برای مثال، جمع‌آوری، سازماندهی و تحلیل کلان‌داده‌ها) نهفته باشد (Hariri, Fredericks and Bowers 2019). بر اساس مطالعات صورت گرفته توسط (Darwish 2020) چالش‌های کلان‌داده را می‌توان در رشد، زیرساخت، صحت، سرعت، تنوع، مصورسازی، انطباق و حاکمیت داده‌ها دسته‌بندی نمود. در پژوهشی دیگر، «سیواراجا» و همکارانش چالش‌های کلان‌داده‌ها را در نیاز شدید به تحلیلگرانی با درک دقیق و بالا از نحوه استفاده از کلان‌داده‌ها تعریف می‌کنند (Sivarajah et al. 2017). همچنین، «میکالف، وترینگ و کروگزیت» سرمایه‌گذاری با هدف آموزش پرسنل کلیدی جهت استفاده از کلان‌داده‌ها در سازمان را به‌عنوان یکی از چالش‌های مطرح در این حوزه تعریف می‌کنند (Mikalef, Wetering and Krogsite 2021). بر اساس مطالعات صورت گرفته توسط «اوسوس» و همکارانش، چالش‌های کلان‌داده‌ها را می‌توان در دو بُعد چالش‌های مدیریتی و فناورانه در نظر گرفت. به عقیده «اوسوس» و همکاران او چالش‌های مدیریتی نسبت به چالش‌های فناورانه از اهمیت بیشتری برخوردار است. از این رو، مدیران باید در نحوه مدیریت خود تغییراتی ایجاد کنند (Oussous et al. 2018). امروزه، چالش‌های زیادی

در بهره‌برداری از پتانسیل کلان‌داده‌ها از طراحی سیستم‌های پردازش در لایه پایین‌تر گرفته تا روش‌های تحلیل در لایه بالاتر، و همچنین، مجموعه‌ای از مشکلات باز در تحقیقات علمی وجود دارد (Willems et al. 2019; Janev 2020). بر اساس چرخه حیات داده‌ها، چالش‌های کلان‌داده‌ها را می‌توان در قالب سه دسته اصلی الف) چالش‌های داده‌ای، ب) چالش‌های فرایندی، و ج) چالش‌های مدیریتی طبقه‌بندی نمود (Hussein 2020). دسته اول، چالش‌های مربوط به ویژگی‌های خود داده‌ها را نشان می‌دهد؛ در حالی که چالش‌های دسته دوم، مربوط به مجموعه‌ای از فنون است؛ از جمله نحوه ذخیره داده‌ها، تغییر داده‌ها، ادغام داده‌ها، انتخاب مدل مناسب جهت تجزیه و تحلیل داده‌ها و در نهایت، نحوه ارائه نتایج. در دسته‌بندی سوم، چالش‌های مربوط به حریم خصوصی، امنیت، حاکمیت و جنبه‌های اخلاقی قرار می‌گیرند. بر اساس پژوهش‌های صورت گرفته در کنفرانس بین‌المللی «انجمن مهندسان برق و الکترونیک»<sup>۱</sup> در سال ۲۰۱۸، چالش‌های کلان‌داده‌ها را در دو دسته چالش‌های مربوط به الف) مدیریت داده‌ها، و ب) تجزیه و تحلیل داده‌ها تقسیم نموده‌اند (Wang et al. 2018). همان‌طور که در پژوهش «وانگ» و همکارانش به آن اشاره شده، از یک سو به دلیل گستردگی توزیع منابع داده‌ای، دسترسی به داده‌های سازمانی، یکپارچه‌سازی و به اشتراک‌گذاری آن‌ها با مشکل مواجه شده، و از سوی دیگر، با توجه به این که داده‌ها برای ادغام و مدیریت به شکل مناسبی آماده نشده‌اند و زیرساخت‌های فنی فاقد خدمات مناسب برای پشتیبانی از فرایند تجزیه و تحلیل کلان‌داده‌ها هستند، جمع‌آوری این داده‌ها به صورت تجمیع شده به عنوان یک چالش کلیدی شناخته می‌شوند (همان).

### ۲-۳. عامل‌های تأثیرگذار بر کلان‌داده‌های سازمانی

شناسایی چالش‌های کلان‌داده‌ها امری بسیار ضروری در راستای ایجاد بستری مناسب جهت استفاده سازمان‌ها از این داده‌هاست و همواره راهکارهای کاهش مخاطره‌های کلان‌داده‌ها، چالش و دغدغه بزرگی برای مدیران و کارکنان کسب‌وکارها و سازمان‌ها محسوب می‌شود. از این رو، در این تحقیق مجموعه‌ای از عواملی که بر روند استفاده از کلان‌داده‌ها در سازمان‌ها تأثیر داشت، مورد بررسی قرار گرفت. بر اساس مرور ادبیات پژوهش و مطالعات کتابخانه‌ای، عامل‌های کلیدی و تأثیرگذار در ارتباط با کاهش مخاطره‌های کلان‌داده‌ها در سازمان‌ها شناسایی شدند که از جمله آن‌ها می‌توان به موارد

1. Institute of Electrical and Electronics Engineers (IEEE)

زیر اشاره نمود: ۱) پشتیبانی مدیریت ارشد از پروژه کلان داده‌ها به منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده (Pal, Meher and Skowron 2015)، ۲) درک صحیح سازمان از مزایای استفاده از گزارشات تحلیلی مستخرج از پروژه کلان داده‌ها (Bharathi 2017)، ۳) تعامل مناسب میان بخش‌های مختلف سازمان به منظور ایجاد دسترسی به پایگاه داده آن‌ها و جمع‌آوری داده‌ها در یک انبار (Wang, Kung and Byrd 2018)، ۴) شناخت دقیق نیازمندی‌های سازمان جهت تعیین خروجی‌های مورد انتظار در استقرار کلان داده‌ها (Saltz, Shamsurin and Connors 2017)، ۵) آموزش کارکنان سازمان جهت کار با سامانه‌های سازمانی به منظور تولید داده صحیح (Sivarajah et al. 2017)، ۶) آموزش نیروهای متخصص که توانایی لازم جهت شناسایی داده‌های مورد نیاز در پایگاه‌های داده مختلف در سطح سازمان را داشته باشند (همان)، ۷) وجود نیروهای متخصص به منظور انجام مراحل استخراج داده از پایگاه‌های داده موجود در بخش‌های مختلف سازمان (همان)، ۸) وجود افراد با مهارت‌های تحلیلی برای تفسیر داده‌ها (Hariri, Fredericks and Bowers 2019)، ۹) استفاده از فناوری مناسب و جدید جهت تولید داده در عملیات‌های مختلف سازمانی (Saheb & Saheb 2020; Maritz, Eybers & Hattingh 2020)، ۱۰) استفاده از فرم‌های جدید پایگاه‌های داده مانند Not Only SQL NoSQL برای ذخیره‌سازی داده‌ها (Gandomi and Haider 2015)، ۱۱) اطمینان از صحت منابع داده‌ای (Baltzan 2019)، ۱۲) وجود سخت‌افزار مناسب و یکپارچه که جوابگوی نیاز استخراج و ذخیره‌سازی کلان داده‌ها باشد (Brahamza et al. 2017)، ۱۳) وجود سیستم عامل و سیستم مرکز داده که جوابگوی نگهداری فرمت‌های مختلف کلان داده‌ها باشند (Gandomi and Haider 2015)، ۱۴) وجود زیرساخت پیچیده مانند صحت، محرمانگی، دسترس پذیری و پاسخگویی که امنیت داده‌ها را تضمین می‌کند، (Sivarajah et al. 2017; Wang, Kung and Byrd 2018)، ۱۵) قابلیت پیش‌بینی و پیشگیری مشکلات شبکه و واحد پردازش مرکزی در زمان استخراج داده (Zhou et al. 2018; Nguyen et al. 2019)، ۱۶) برقراری ارتباط بین پایگاه‌های داده‌ای متفاوت و انتخاب روشی مناسب جهت استخراج داده‌ها (Sivarajah et al. 2017)، ۱۷) وجود برنامه‌های مناسب جهت تجزیه و تحلیل داده‌ها (Mikalef et al. 2019)، ۱۸) استفاده از ابزار تجسم کلان داده‌ها (برای مثال، Tableau) (Sivarajah et al. 2017)، ۱۹) استفاده از روش‌های محاسباتی موازی برای مثال، Hadoop برای پردازش کلان داده‌ها (Oussous et al. 2018)، ۲۰) استفاده از سرویس‌های مبتنی بر ابر برای پردازش و تجزیه و تحلیل داده (Li et al.).

(2020)، (۲۱) تدوین استاندارد مربوط به نحوه و مدت‌زمان ذخیره‌سازی داده (Obilikwu and Ogbuju 2020)، (۲۲) تعیین استاندارد برای مشخص نمودن اعتبار داده‌ها و حذف داده‌های منقضی‌شده (Pérez-Martín, Pérez-Torregrosa and Vaca 2018)، (۲۳) توانایی مدیریت مجموعه کلان‌داده‌های غیرمرتبط یا غیرساختاری (Wordsworth et al. 2018)، (۲۴) توانایی گرفتن داده‌ها در قالب‌های متنوع و از منابع گوناگون (Sivarajah et al. 2019; Baltzan 2017)، (۲۵) توانایی مدیریت داده‌های ناهمگن، همه‌جا، و ماهیت پویای منابع و دستگاه‌های مختلف تولید اطلاعات (Tian et al. 2015; Sivarajah et al. 2017)، (۲۶) طبقه‌بندی، مدل‌سازی و نقشه‌برداری از داده‌ها به همان صورتی که گرفته شده‌اند ذخیره می‌شوند، به‌طور عمده به دلیل ماهیت ساختار پیچیده آن‌ها (Sivarajah et al. 2017;)، (۲۷) ایجاد قوانین و مقررات در خصوص جمع‌آوری و نحوه ذخیره داده (Fikri et al. 2019)، (۲۸) استفاده از فیلترهای هوشمند و قوی برای ضبط اطلاعات مفید و دور انداختن اطلاعات بی‌فایده که حاوی اشتباهات و یا ناسازگاری است (Sivarajah et al. 2017)، (۲۹) وجود قوانین و مقررات برای حفظ حریم خصوصی استاندارد در انبار داده (Sivarajah et al. 2017; Raguseo 2018)، (۳۰) وجود کنترل‌های امنیتی مناسب برای اطمینان از عدم تغییر در اطلاعات و داده (Sivarajah et al. 2017)، (۳۱) استفاده از ابزار و تجزیه و تحلیل برای مدیریت و استخراج داده‌های غیرقابل اعتماد (Mehmood et al. 2020; Caesarius and Hohenthal 2018; Lei and Kong 2020)، (۳۲) به کارگیری الگوریتم‌های تحلیلی کارآمد برای درک رسیدن به داده‌ها و پردازش کلان‌داده‌ها و کاهش داده‌ها قبل از ذخیره‌سازی (Mikalef et al. 2019; Lei and Kong 2020)، (۳۳) استفاده از تجزیه و تحلیل بلادرنگ و برنامه‌ریزی مبتنی بر شواهد برای مدیریت داده‌های تولیدشده توسط دستگاه‌های دیجیتال که در همه‌جا قرار گرفته‌اند و داده‌های زیادی تولید می‌کنند (Lei and Kong 2020)، (۳۴) استفاده از روش‌های جدید تجزیه و تحلیل و مدل‌سازی برای مدیریت کلان‌داده‌ها برای حداکثر تأثیر و ارزش تجاری (Mikalef et al. 2019; Lei and Kong 2020)، (۳۵) پشتیبانی مدیران فناوری اطلاعات از حضور یک استراتژی رسمی کلان‌داده (Del Vecchio et al. 2018)، (۳۶) تأمین مالی پروژه‌های تجزیه و تحلیل کلان‌داده‌ها (Sivarajah et al. 2017)، (۳۷) در اختیار قرار دادن زمان کافی به پروژه‌های تحلیلی کلان‌داده‌ها برای رسیدن به اهداف خود (Stergiou and Psannis 2017)، (۳۸) درک مدیران تجزیه و تحلیل کلان‌داده‌ها از نیازهای کسب‌وکار، مدیران عملکرد، تأمین‌کنندگان و مشتریان

(Mikalef et al. 2019)، (۳۹) تقویت فرهنگ سازمان در خصوص به اشتراک گذاشتن داده (Díaz, Rowshankish and Saleh 2018; Eastwood 2021)، (۴۰) ایجاد ساختار سازمانی مناسب جهت استفاده از استقرار پروژه کلان‌داده‌ها در سازمان (Günther et al. 2017)، (۴۱) تدوین قوانین و مقررات و روال‌های سازمانی جهت استقرار و پیاده‌سازی پروژه‌های کلان‌داده‌ها در سازمان (Sivarajah et al. 2017)، (۴۲) زمینه‌سازی ایجاد فرهنگ سازمانی که پذیرای استفاده از کلان‌داده‌ها در سازمان باشد (Díaz, Rowshankish and Saleh 2018; Eastwood 2021)، (۴۳) وجود پرسنل ماهر یا کارکنان با مهارت‌های تجزیه و تحلیل برای درک کلان‌داده‌ها (Sivarajah et al. 2017)، (۴۴) توسعه و اجرای زیرساخت‌های مدیریت و پردازش کلان‌داده‌ها (Song et al. 2018)، و (۴۵) آموزش کارکنان برای تجزیه و تحلیل داده‌ها (Wordsworth et al. 2018).

با توجه به این‌که در موضوع اولویت‌بندی مخاطره‌های تأثیرگذار در استفاده از کلان‌داده‌های سازمانی چندین مطالعه انجام شده و به‌رغم توجه فزاینده به این حوزه تحقیقاتی، تاکنون تأثیرگذاری و تأثیرپذیری عوامل مؤثر در کاهش مخاطره‌ها به‌طور نظام‌مند بررسی و مورد سنجش قرار نگرفته است، از این رو، در این پژوهش برآنیم تا با شناسایی و اولویت‌بندی عامل‌های مؤثر به سازمان‌ها کمک کنیم تا نسبت به برنامه‌ریزی جهت کاهش هرچه بیشتر مخاطرات استفاده از کلان‌داده‌های سازمانی خود اقدام نمایند.

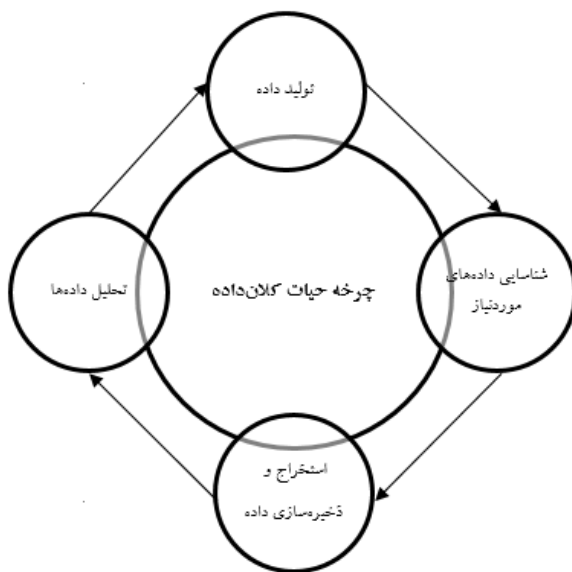
### ۳. چارچوب عوامل تأثیرگذار در استفاده از کلان‌داده‌های سازمانی

هدف از ارائه این مقاله بررسی و شناسایی عوامل کلیدی مؤثر در کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌های سازمانی، طبقه‌بندی و ارائه یک چارچوب از این عوامل است. در این راستا با استفاده از منابع کتابخانه‌ای، عوامل کلیدی مؤثر در کاهش مخاطره‌ها مورد شناسایی قرار گرفت و سپس، با استفاده از مطالعات صورت گرفته پیرامون این موضوع، چارچوبی از عوامل مؤثر بر کاهش مخاطره در قالب دو بعد قابلیت‌های سازمانی و چرخه حیات کلان‌داده‌ها ارائه گردید که در ادامه، به تشریح هر یک از این دو بُعد و ارائه چارچوب پیشنهادی پرداخته خواهد شد.

#### ۳-۱. چرخه حیات کلان‌داده‌ها

چرخه حیات داده مراحل دنباله‌داری است که در آن یک واحد خاص از داده تولید،

جمع‌آوری و ذخیره می‌شود و در پایان عمر مفیدش حذف می‌شود (Wigmore 2017). چرخه حیات داده سطح بالایی از مراحل درگیر در مدیریت موفق و حفظ داده‌ها برای استفاده و استفاده مجدد را فراهم می‌کند (Shah, Peristeras and Magnisalis 2021). همان‌طور که اشاره شد، سازمان‌ها با حجم بالایی از داده‌ها مواجه هستند. در همین راستا نیاز است که داده‌ها به‌خوبی شناسایی شده، مدت‌زمان مناسبی نگهداری و ذخیره شوند و با تحلیل صحیح و به‌موقع، سازمان‌ها را در جهت رسیدن به اهداف و حفظ مزیت رقابتی یاری کنند. با توجه به این که نسخه‌های چندگانه برای چرخه حیات داده وجود دارد، چرخه حیات مورد نظر در این پژوهش در قالب شکل ۱، به تصویر کشیده شده است.



شکل ۱. چرخه حیات کلان‌داده‌ها

مرحله اول، تولید داده: تولید داده با ورود اطلاعات شروع می‌شود. چالش زمانی شروع می‌شود که در ورود اطلاعات نقصی وجود داشته باشد و یا نقصی در ترکیب داده‌های سیستم‌های موروثی باشد (Lei and Kong 2020).

مرحله دوم، شناسایی داده: یکی از ویژگی‌های کلان‌داده‌ها این است که ساختارمند نیستند. به همین دلیل تعیین معیاری برای سنجش آن‌ها و استخراج دسته‌های مختلف داده‌ای که در تصمیم‌گیری‌های مربوط به کسب و کار استفاده می‌شود، مشکل است. به‌عبارتی،

داده‌ها قبل از تجزیه و تحلیل باید به‌خوبی ساخته شوند (Fikri et al. 2019). شناخت داده، آماده‌سازی داده جهت دریافت نتایج بهتر و باکیفیت‌تر از تجزیه و تحلیل، چالشی را به‌وجود می‌آورد که بسیار بااهمیت است. مراحل نظیر فرایند استخراج-پالایش-بارگذاری، تمیز کردن داده‌ها و ادغام امری بسیار مهم در تفسیر و سرانجام، تصمیم‌گیری محسوب می‌شوند (Fikri, et al. 2019).

مرحله سوم، ذخیره‌سازی داده: اولین و مهم‌ترین ویژگی کلان‌داده‌ها حجم آنهاست که ما را با چالش ذخیره‌سازی آن روبه‌رو می‌سازد (Oussous et al. 2018; Mikalef et al. 2019; Fikri et al. 2019). به‌دلیل حجم بالای داده‌ها، پالایش آن‌ها در یک موضوع خاص، زمان‌بر است (Jha, Agi and Ngai 2020). بنابراین، باید این امر سنجیده شود که آیا داده‌ها همان داده‌هایی است که در تصمیم‌گیری مورد استفاده قرار خواهد گرفت یا خیر. و همچنین، باید اطمینان حاصل شود که مقدار داده‌ای آن‌ها هنوز درست است و قابلیت اطمینان برای این مقدار وجود دارد. به‌عبارتی، جمع‌آوری و اجتماع داده از اهمیت بالایی برخوردار است (Lei and Kong 2020).

مرحله چهارم، تحلیل داده: در خصوص بهره‌برداری از پتانسیل کلان‌داده‌ها (Willems et al. 2020; Janev 2019) نیز چالش‌هایی وجود دارد؛ از جمله طراحی سیستم‌های پردازش در لایه پایین‌تر به روش‌های تجزیه و تحلیل در لایه بالاتر (Oussous et al. 2018) و همچنین، مجموعه‌ای از مشکلات باز که به تحقیقات علمی مربوط می‌شود.

### ۳-۲. قابلیت‌های سازمانی مورد نیاز جهت استفاده از کلان‌داده‌ها

هر سازمانی با توجه به توانایی‌های خود می‌تواند بهره‌وری متفاوتی از کلان‌داده‌ها داشته باشد. برای مثال، نیروی انسانی بر اساس امکانات، مهارت‌ها، دانش و نگرش خویش که از آموزش کسب کرده است، در طی مراحل مختلف می‌تواند نقش به‌سزایی در بهبود و افزایش بهره‌وری از کلان‌داده‌ها داشته باشد. به‌طور کلی، این توانایی‌ها و قابلیت‌ها را می‌توان در قالب چهار دسته زیر تقسیم‌بندی کرد (Capstera 2018):

◇ **قابلیت‌های سازمانی:** به زبان ساده یعنی توانمندی‌هایی که آن سازمان دارد. قابلیت‌ها در واقع کارهایی هستند که یک سازمان با استفاده از فرایندها، افراد و فناوری خود





در واقع، جهت کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌ها در سازمان‌ها می‌توان در هر مرحله از چرخه حیات کلان‌داده‌ها (تولید، شناسایی، ذخیره‌سازی، و تحلیل) از انواع قابلیت‌های موجود در سازمان‌ها (مدیریتی، زیرساختی و فنی، نیروی انسانی، و سازمانی) استفاده کرد. از این رو، در این پژوهش با بهره‌گیری از روش اسنادی و مبتنی بر مطالعات کتابخانه‌ای و پژوهش‌های موجود، عوامل تأثیرگذار بر کاهش مخاطره‌ها استخراج گردید. سپس، با بررسی مرور ادبیات موضوعی و نظرات خبرگان، چارچوبی متشکل از ۴ حوزه تأثیرگذار و ۴۴ عامل مؤثر تشکیل گردید که در جداول (۲ تا ۵) فهرستی از عامل‌هایی که جهت کاهش مخاطره در هر مرحله از فرایند چرخه حیات کلان‌داده‌ها می‌توان مد نظر قرار داد، در ۴ حوزه سازمانی، انسانی، فنی و مدیریتی دسته‌بندی و نگاشت شده است.

## جدول ۲. قابلیت‌های سازمانی

علامت اختصاری	عامل	مراحل چرخه حیات داده	قابلیت
C42	زمینه‌سازی ایجاد فرهنگ سازمانی که پذیرای تولید و استفاده از کلان‌داده‌ها در سازمان باشد.	تولید داده	سازمانی
C40	ایجاد ساختار سازمانی مناسب جهت استفاده استقرار پروژه کلان‌داده‌ها در سازمان (تعیین مسئول صحت‌سنجی)		
C4	شناخت دقیق نیازمندی‌های سازمان جهت تعیین خروجی‌های مورد انتظار از استقرار کلان‌داده	شناسایی داده	
C3	تعامل مناسب میان بخش‌های مختلف سازمان به‌منظور ایجاد امکان دسترسی به پایگاه‌های داده آن‌ها و تجمیع داده‌ها در یک انبار	ذخیره‌سازی داده	
C1	پشتیبانی مدیریت ارشد از پروژه کلان‌داده‌ها به‌منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده		
C35	پشتیبانی مدیران فناوری اطلاعات از حضور یک استراتژی رسمی کلان‌داده		
C2	درک صحیح سازمان از مزایای استفاده از گزارشات تحلیلی مستخرج از پروژه کلان‌داده‌ها	تحلیل داده	

### جدول ۳. قابلیت‌های انسانی

علامت اختصاری	عامل	مراحل چرخه حیات داده	قابلیت
C5	آموزش کارکنان سازمان جهت کار با سامانه‌های سازمانی به‌منظور تولید داده صحیح	تولید داده	انسانی
C6	آموزش نیروهای متخصص که توانایی لازم جهت شناسایی داده‌های مورد نیاز در پایگاه‌های داده مختلف در سطح سازمان را داشته باشند	شناسایی داده	
C7	وجود نیروی متخصص به‌منظور انجام مراحل استخراج داده از پایگاه‌های داده موجود در بخش‌های مختلف سازمان	ذخیره‌سازی داده	
C8	وجود افراد با مهارت‌های تحلیلی برای تفسیر داده‌ها	تحلیل داده	
C43	وجود پرسنل ماهر یا کارکنان با مهارت‌های تجزیه و تحلیل برای درک کلان‌داده‌ها		

### جدول ۴. قابلیت‌های فنی

علامت اختصاری	عامل	مراحل چرخه حیات داده	قابلیت
C9	استفاده از فناوری مناسب و جدید جهت تولید داده در عملیات‌های مختلف سازمانی (برنامه‌های نرم‌افزاری مناسب جهت ثبت اطلاعات)	تولید داده	فنی
C10	استفاده از فرم‌های جدید پایگاه‌های داده مانند Not Only SQL (NoSQL) برای ذخیره‌سازی داده‌ها	ذخیره‌سازی داده	
C11	اطمینان از صحت منابع داده‌ای (شناسایی دقیق منابع داده‌ها)	شناسایی داده	
C12	وجود سخت‌افزار مناسب و یکپارچه که جوابگوی نیاز استخراج و ذخیره‌سازی کلان‌داده‌ها باشد.	ذخیره‌سازی داده	
C13	وجود سیستم عامل و سیستم مرکز داده که جوابگوی نگهداری فرمت‌های مختلف files, db's, (voices, picture, etc) کلان‌داده‌ها باشند.		
C14	وجود زیرساخت پیچیده که امنیت داده‌ها را تضمین می‌کند؛ مانند صحت، محرمانگی، دسترس‌پذیری و پاسخگویی		
C15	قابلیت پیش‌بینی و پیشگیری مشکلات شبکه و واحد پردازش مرکزی در زمان استخراج داده		
C16	برقراری ارتباط بین پایگاه‌های داده‌ای متفاوت و انتخاب روشی مناسب (با توجه به وجود سیستم‌های قدیمی) جهت استخراج داده‌ها		
C17	وجود برنامه‌های مناسب جهت تجزیه و تحلیل داده (اطمینان از صحت فرایندهای تجزیه و تحلیل)	تحلیل داده	

علامت اختصاری	مراحل چرخه حیات داده	عامل	قابلیت
C18	استفاده از ابزار تجسم کلان داده‌ها (برای مثال Tableau)		
C19	استفاده از روش‌های محاسباتی موازی (برای مثال، Hadoop) برای پردازش کلان داده‌ها		
C20	استفاده از سرویس‌های مبتنی بر ابر برای پردازش و تجزیه و تحلیل داده		
C44	توسعه و اجرای زیرساخت‌های مدیریت و پردازش کلان داده‌ها		

### جدول ۵. قابلیت‌های مدیریتی

علامت اختصاری	مراحل چرخه حیات داده	عامل	قابلیت
C21	تدوین استاندارد مربوط به نحوه و مدت زمان ذخیره‌سازی داده		مدیریتی تولید داده
C22	تعیین استاندارد برای مشخص نمودن اعتبار داده‌ها و حذف داده‌های منقضی شده		
C23	توانایی مدیریت مجموعه کلان داده‌های غیر مرتبط یا غیر ساختاری		
C24	توانایی گرفتن داده‌ها در قالب‌های متنوع و از منابع گوناگون		
C25	توانایی مدیریت داده‌های ناهمگن، همه‌جا، و ماهیت پویای منابع و دستگاه‌های مختلف تولید اطلاعات		
C26	طبقه‌بندی، مدل‌سازی و نقشه‌برداری از داده‌ها به همان صورتی که گرفته و ذخیره می‌شوند، به‌طور عمده به‌دلیل ماهیت ساختار و پیچیده آن‌ها		شناسایی داده
C27	ایجاد قوانین و مقررات در خصوص جمع‌آوری و نحوه ذخیره داده (متمرکز یا توزیع شده) (در خصوص ترکیب داده‌ها از منابع مختلف)		ذخیره‌سازی داده
C28	استفاده از فیلترهای هوشمند و قوی برای ضبط اطلاعات مفید و دور انداختن اطلاعات بی‌فایده که حاوی اشتباهات و یا ناسازگاری است		
C29	وجود قوانین و مقررات برای حفظ حریم خصوصی استاندارد در انبار داده		
C30	وجود کنترل‌های امنیتی مناسب برای اطمینان از عدم تغییر در اطلاعات و داده		
C31	استفاده از ابزار و تجزیه و تحلیل برای مدیریت و استخراج داده‌های غیر قابل اعتماد (داده‌های نادرست و مبهم)		
C32	به‌کارگیری الگوریتم‌های تحلیلی کارآمد برای درک رسیدن به داده‌ها و پردازش کلان داده‌ها و کاهش داده‌ها قبل از ذخیره‌سازی		
C41	تدوین قوانین و مقررات و روال‌های سازمانی جهت استقرار و پیاده‌سازی پروژه کلان داده‌ها در سازمان		
C39	تقویت فرهنگ سازمان در خصوص به‌اشتراک گذاشتن داده‌ها		

علامت اختصاری	مراحل چرخه حیات داده	قابلیت
C33	استفاده از تجزیه و تحلیل بلادرنگ و برنامه‌ریزی مبتنی بر شواهد برای مدیریت داده‌های تولیدشده توسط دستگاه‌های دیجیتال که در همه‌جا قرار گرفته‌اند و داده‌های زیادی تولید می‌کنند.	تحلیل داده
C34	استفاده از روش‌های جدید تجزیه و تحلیل و مدل‌سازی برای مدیریت کلان‌داده‌ها برای حداکثر تأثیر و ارزش تجاری	
C38	درک مدیران تجزیه و تحلیل داده‌های بزرگ از نیازهای کسب‌وکار، مدیران عملکرد، تأمین‌کنندگان و مشتریان	
C37	در اختیار دادن زمان کافی به پروژه‌های تحلیلی کلان‌داده‌ها برای رسیدن به اهداف خود	
C36	تأمین مالی پروژه‌های تجزیه و تحلیل کلان‌داده‌ها	

#### ۴. روش پژوهش

امروزه، اهمیت و ضرورت داده‌ها بر کسی پوشیده نیست؛ به‌ویژه با توجه به حجم وسیع داده‌های تولیدشده که سازمان‌ها را مجاب می‌کند تا بهترین تصمیم را گرفته و از آن به‌عنوان یک مزیت رقابتی استفاده کنند. از سوی دیگر، استفاده از این داده‌ها همانند سایر فناوری‌ها مخاطره‌های خاص خود را به همراه دارد که سازمان‌ها لازم است آن‌ها را شناسایی کنند تا ضمن کاهش مخاطره‌ها، بهره‌وری خود را بالا ببرند. از این رو، پژوهش حاضر از لحاظ هدف جزء پژوهش‌های کاربردی است، زیرا انتظار می‌رود با شناسایی عوامل کلیدی تأثیرگذار بتوان به سازمان‌ها در کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌ها کمک کرد. همچنین، با توجه به این که فرایند گردآوری داده‌ها در این پژوهش، بر اساس دو شیوه مطالعات کتابخانه‌ای و پرسشنامه‌ای (مصاحبه با خبرگان) صورت گرفته است، پس می‌توان بیان کرد که پژوهش حاضر بر اساس ماهیت و روش گردآوری داده‌ها یک پژوهش توصیفی-پیمایشی محسوب می‌شود. در ابتدا، با مطالعه متون نظری پژوهش، عوامل مؤثر در خصوص کاهش مخاطره‌های کلان‌داده‌ها در سازمان شناسایی شد. سپس، جهت دستیابی به هدف پژوهش و با بهره‌گیری از مصاحبه‌های نیمه‌ساخت یافته با ۱۵ نفر از افراد درگیر در سازمان تأثیر این عوامل مشخص شد و نظرات آن‌ها در قالب پرسشنامه‌ای در ارتباط با مخاطره‌های کلان‌داده‌ها ثبت گردید. جامعه آماری ترکیبی از خبرگان و متخصصان بانکی و اعضای هیئت علمی دانشگاه‌ها بودند. برای نمونه‌گیری از

این جامعه آماری و انتخاب افراد خبره برای برگزاری مصاحبه‌ها ترکیبی از روش‌های غیراحتمالی هدفمند قضاوتی و گلوله‌برفی به کار گرفته شد.

در ابتدای پژوهش، نظر بر این بود که با توجه به زمان و منابع در اختیار، برای انجام تحقیق از خبرگان و متخصصانی استفاده شود که با مقوله کلان‌داده‌ها و همچنین، مخاطره‌های امنیتی در این حوزه آشنایی دارند. بر این اساس، جامعه آماری محدودی با ۱۵ نفر از خبرگان و متخصصان فعال در شبکه‌های بانکی، مدیریت، امنیت و شبکه در سازمان پیش روی محققان (بانک مرکزی) شناسایی شد. به‌رغم تلاش‌های فراوان و با توجه به شرایط فراگیری کرونا و عدم امکان حضور خبرگان، از بین ۱۵ نفر خبره شناسایی شده، تنها ۹ نفر حاضر به انجام مصاحبه - با طول زمانی ۱۵ تا ۳۰ دقیقه - گردیدند. به‌منظور تجزیه و تحلیل داده‌ها از تکنیک «دیمتل فازی»<sup>۱</sup> استفاده شده است. علم مدیریت فازی، ضمن ایجاد انعطاف‌پذیری در مدل، داده‌هایی نظیر دانش، تجربه و قضاوت انسانی را در مدل وارد کرده و پاسخ‌های کاملاً کاربردی را ارائه می‌دهد. امروزه، رویکرد فازی در بیشتر روش‌های تصمیم‌گیری وارد شده و ترکیبی از رویکرد فازی و روش‌های تصمیم‌گیری در حال گسترش است. این تکنیک، ابزاری کارآمد برای حل مسائلی است که در آن‌ها ممکن است دانش، غیرقطعی باشد. در واقع، نظریه فازی ذهنیت رفتار را نشان می‌دهد و به‌عبارتی، به روش کار کردن مغز انسان‌ها نزدیک‌تر است. در ادامه، توضیح مختصری از تکنیک مورد استفاده در این پژوهش در فضای فازی بیان می‌شود.

### تکنیک دیمتل فازی

تکنیک «دیمتل» روشی برای نشان دادن ساختار روابط علی در یک سیستم پیچیده از روابط است که آن را می‌توان برای ایجاد و تحلیل یک مدل ساختاری متشکل از روابط علی میان عوامل مدل به کار گرفت. مراحل تکنیک «دیمتل فازی» به‌صورت الگوریتمیک به‌شرح زیر است (Jeng and Tzeng 2012).

**گام اول) تشکیل ماتریس روابط مستقیم فازی:** ماتریس روابط مستقیم فازی  $\tilde{X}$  یک ماتریس زوجی  $n \times n$  است که هر یک از عناصر آن میزان تأثیر مستقیم عامل  $i$  بر عامل  $j$  را به‌صورت یک عدد فازی مثلثی نشان می‌دهد. میزان تأثیر هر عامل بر سایر عوامل طبق

1. Fuzzy DEMATEL technique

نظر خبرگان و با به کارگیری اعداد فازی در قالب (جدول ۶) تعیین شد. سپس، میانگین نظرات خبرگان محاسبه شده و ماتریس نهایی روابط مستقیم فازی به دست آمد.

جدول ۶. متغیر زبانی و اعداد فازی معادل آن ها

متغیر زبانی	معادل قطعی	معادل فازی
بدون تأثیر	۰	(۰/۲۵، ۰/۰۰، ۰/۰۰)
تأثیر خیلی کم	۱	(۰/۵۰، ۰/۲۵، ۰/۰۰)
تأثیر کم	۲	(۰/۷۵، ۰/۵۰، ۰/۲۵)
تأثیر زیاد	۳	(۱/۰۰، ۰/۷۵، ۰/۵۰)
تأثیر خیلی زیاد	۴	(۱/۰۰، ۱/۰۰، ۰/۷۵)

گام دوم) نرمال کردن ماتریس روابط مستقیم فازی: ماتریس روابط مستقیم فازی نرمال شده  $\tilde{N}$  بر اساس فرمول (۱) محاسبه می شود.

$$k = \frac{1}{\max \sum_{j=1}^x a_{ij}}$$

گام سوم) محاسبه ماتریس روابط کامل: ماتریس روابط کامل بر اساس فرمول (۲) محاسبه می شود. برای محاسبه ماتریس ارتباط کامل، ابتدا یک ماتریس یکه یا همانی تشکیل می شود که قطر اصلی آن ۱ و بقیه درایه های آن صفر است.

$$\tilde{T} = \tilde{N} \times (1 - \tilde{N})^{-1}$$

گام چهارم) تشکیل روابط علی و معلولی: در این گام، مجموع سطرها (D) و مجموع ستون ها (R) در ماتریس مجموعه روابط فازی ( $\tilde{T}$ ) به شکل جداگانه محاسبه می شوند که  $\tilde{D}$  و  $\tilde{R}$  به ترتیب، ماتریس  $(n \times 1)$  و  $(1 \times n)$  هستند.

◇ در ماتریس (D)، جمع عناصر هر سطر برای هر عامل نشانگر میزان تأثیرگذاری آن عامل بر سایر عامل های سازمان است که از فرمول (۳) محاسبه می شود. (میزان تأثیرگذاری متغیرها)؛

$$\tilde{D} = (\tilde{D}_i)_{n \times 1} = [\sum_{j=1}^n \tilde{T}_{ij}]_{n \times 1}$$

◇ در ماتریس (R)، جمع عناصر ستون برای هر عامل نشانگر میزان تأثیرپذیری آن عامل از سایر عامل های سازمان است که از فرمول (۴) محاسبه می شود. (میزان تأثیرپذیری متغیرها)؛

$$\bar{R} = (\bar{R}_i)_{1 \times n} = [\sum_{i=1}^n \bar{T}_{ij}]_{1 \times n}$$

در مرحله بعد، میزان اهمیت شاخص‌ها  $(\bar{D}_i + \bar{R}_i)$  و رابطه بین عامل‌ها  $(\bar{D}_i - \bar{R}_i)$  مشخص می‌گردد. اگر  $\bar{D}_i - \bar{R}_i > 0$  باشد، عامل مربوطه اثرگذار و اگر  $\bar{D}_i - \bar{R}_i < 0$  باشد، عامل مربوطه اثرپذیر است. در دستگاه مختصات روابط علی و معلولی، محور طولی بردار  $(\bar{D}_i + \bar{R}_i)$  میزان تأثیرگذاری و تأثیرپذیری عامل مورد نظر در سازمان است. به عبارت دیگر، هر چه مقدار  $(\bar{D}_i + \bar{R}_i)$  برای یک عامل بیشتر باشد، آن عامل تعامل بیشتری با سایر عوامل سازمان دارد. محور عرضی بردار  $(\bar{D}_i - \bar{R}_i)$  قدرت تأثیرگذاری هر عامل را نشان می‌دهد. به طور کلی، اگر  $(\bar{D}_i - \bar{R}_i)$  مثبت باشد، یک متغیر علی محسوب می‌شود و اگر منفی باشد، یک متغیر معلول محسوب می‌شود. بنابراین، نمودار روابط علی و معلولی از طریق ترسیم مقادیر دیفازی شده  $(\bar{R} + \bar{D}, \bar{R} - \bar{D})$  برای هر عامل حاصل می‌شود. در واقع، موقعیت هر عامل با نقطه‌ای به مختصات  $(\bar{R} + \bar{D}, \bar{R} - \bar{D})$  در دستگاه مختصات دکارتی تعیین و در نهایت، یک نمودار گرافیکی حاصل می‌شود. در مرحله بعد، به منظور تبدیل مقادیر فازی به دست آمده از مرحله قبل به مقادیر قطعی (دیفازی کردن) می‌توان از فرمول (۵) استفاده نمود که در حقیقت B دیفازی شده مقادیر  $\bar{A} = (a_1, a_2, a_3)$  است.

$$B_{ij}^{def} = \frac{a_1 + a_3 + 2 \times a_2}{4}$$

## ۵. تجزیه و تحلیل یافته‌ها

سؤال اول طبقه‌بندی عامل‌های مؤثر در کاهش مخاطره‌های شناسایی شده ناشی از استفاده از کلان داده‌ها در سازمان‌ها

در این مقاله تلاش شده است که با شناسایی و تحلیل عوامل مؤثر بر کاهش مخاطره‌های ناشی از استفاده از کلان داده‌ها در سازمان، چارچوبی مدون در این زمینه ارائه شود. بر این اساس، با مروری بر پیشینه ادبیات مطرح در این حوزه و گردآوری اطلاعات، ۴ حوزه تأثیرگذار و ۴۴ عامل مؤثر شناسایی و در قالب چارچوبی برای شناسایی و تحلیل عوامل مؤثر بر کاهش مخاطره‌های ناشی از استفاده از کلان داده‌ها در سازمان‌ها ارائه گردید. این عوامل شامل ۴ بعد قابلیت‌های موجود در سازمان‌ها (مدیریت داده، زیرساخت و فنی، نیروی انسانی، و سازمانی) و ۴ بعد چرخه حیات کلان داده‌ها (تولید، شناسایی، ذخیره‌سازی، و تحلیل) هستند که در بخش (۳) و در قالب چارچوب عوامل تأثیرگذار طبقه‌بندی و به آن پرداخته شده است.

### سؤال دوم) بررسی و تحلیل ارتباط بین عامل‌ها (میزان تأثیرگذاری و تأثیرپذیری)

با مرور مبانی نظری پژوهش و بررسی مخاطره‌ها و چالش‌های مطرح در حوزه کلان‌داده‌ها در مجموع ۴۴ عامل شناسایی شد. برای محاسبه میزان وابستگی متقابل عامل‌های شناسایی‌شده، بر اساس تکنیک «دیمتل فازی» پرسشنامه‌ای تهیه و در اختیار متخصصان و خبرگان این حوزه قرار گرفت. برای شناسایی روابط (N) عامل، ابتدا یک ماتریس (n×n) تشکیل و از خبرگان خواسته شد با به‌کارگیری اعداد فازی در قالب جدول ۶، میزان تأثیر هر عامل بر سایر عامل‌ها را با عددی از صفر تا چهار مشخص کنند. سپس، میانگین نظرات خبرگان محاسبه و ماتریس نهایی روابط مستقیم فازی حاصل شد. تمامی محاسبات بر اساس جدول میانگین نظر خبرگان (پیوست الف) صورت گرفته است. همان‌طور که در تکنیک «دیمتل فازی» بیان شد، جمع عناصر هر سطر (D) نشانگر میزان تأثیرگذاری آن عامل بر سایر عامل‌ها و مجموع عناصر ستون (R) برای هر عامل نشانگر میزان تأثیرپذیری آن عامل از سایر عامل‌های سیستم است. همچنین، بردار افقی (D+R) میزان تأثیرگذاری و تأثیرپذیری عوامل مورد نظر در سیستم را نشان می‌دهد. هر چه مقدار (D+R) عاملی بیشتر باشد، آن عامل تعامل بیشتری با سایر عوامل سیستم دارد. بردار عمودی (D-R)، قدرت تأثیرگذاری هر عامل را نشان می‌دهد. به‌طور کلی، اگر (D-R) مثبت باشد، متغیر یک متغیر علی (علت) محسوب می‌شود و اگر منفی باشد، معلول محسوب می‌شود. مجموع سطرها، ستون‌ها و نیز بردار اهمیت و بردار روابط در حالت فازی بر اساس فرمول‌های (۳) و (۴) محاسبه و در جدول ۷، نشان داده شده است.

جدول ۷. مجموع سطرها و ستون‌های ماتریس مجموعه روابط و بردار اهمیت و روابط

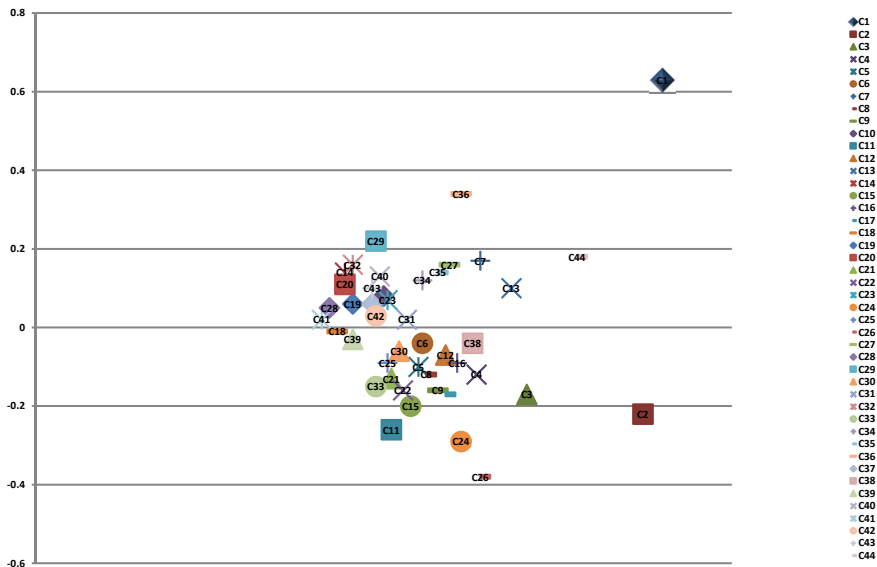
مطلوب	علت	D+R بازدار رابطه فازی	D-R بازدار رابطه فازی	مجموع اثرات عامل	بازدار اهمیت فازی	مطلوب	علت	D+R بازدار رابطه فازی	D-R بازدار رابطه فازی	مجموع اثرات عامل	بازدار اهمیت فازی	
✓	✓	۰/۰۷	۰/۹۱	۰/۴۲	۰/۴۹	C۲۳	✓	۰/۶۳	۱/۶۲	۰/۴۹۵	۱/۱۲۵	C۱
✓	✓	-۰/۲۹	۱/۱	۰/۶۹۵	۰/۴۰۵	C۲۴	✓	-۰/۲۲	۱/۵۷	۰/۸۹۵	۰/۶۷۵	C۲
✓	✓	-۰/۰۹	۰/۹۱	۰/۵	۰/۴۱	C۲۵	✓	-۰/۱۷	۱/۲۷	۰/۷۲	۰/۵۵	C۳
✓	✓	-۰/۳۸	۱/۱۵	۰/۷۶۵	۰/۳۸۵	C۲۶	✓	-۰/۱۲	۱/۱۴	۰/۶۳	۰/۵۱	C۴
✓	✓	۰/۱۶	۱/۰۷	۰/۴۵۵	۰/۶۱۵	C۲۷	✓	-۰/۱	۰/۹۹	۰/۵۴۵	۰/۴۴۵	C۵



عامل	مجموع اثرات عامل پر سایر عامل‌ها D	مجموع اثرات سایر عامل‌ها بر عامل R	پر دار اهمیت فازی D+R	پر دار رابطه فازی D-R	علت	مطلوب	عامل	مجموع اثرات عامل پر سایر عامل‌ها D	مجموع اثرات سایر عامل‌ها بر عامل R	پر دار اهمیت فازی D+R	پر دار رابطه فازی D-R	علت	مطلوب
C6	۰/۴۸	۰/۵۲	۱	-۰/۰۴		✓	C28	۰/۴۰۵	۰/۳۵۵	۰/۷۶	۰/۰۵	✓	
C7	۰/۶۶	۰/۴۹	۱/۱۵	۰/۱۷	✓		C29	۰/۵۵	۰/۳۳	۰/۸۸	۰/۲۲	✓	
C8	۰/۴۴۵	۰/۵۶۵	۱/۰۱	-۰/۱۲		✓	C30	۰/۴۴	۰/۵	۰/۹۴	-۰/۰۶		✓
C9	۰/۴۴	۰/۶	۱/۰۴	-۰/۱۶		✓	C31	۰/۴۹	۰/۴۷	۰/۹۶	۰/۰۲	✓	
C10	۰/۴۹	۰/۴۱	۰/۹	۰/۰۸	✓		C32	۰/۴۹	۰/۳۳	۰/۸۲	۰/۱۶	✓	
C11	۰/۳۳	۰/۵۹	۰/۹۲	-۰/۲۶		✓	C33	۰/۳۶۵	۰/۵۱۵	۰/۸۸	-۰/۱۵		✓
C12	۰/۴۹۵	۰/۵۶۵	۱/۰۶	-۰/۰۷		✓	C34	۰/۵۶	۰/۴۴	۱	۰/۱۲	✓	
C13	۰/۶۶۵	۰/۵۶۵	۱/۲۳	۰/۱	✓		C35	۰/۵۹	۰/۴۵	۱/۰۴	۰/۱۴	✓	
C14	۰/۴۷	۰/۳۳	۰/۸	۰/۱۴	✓		C36	۰/۷۲	۰/۳۸	۱/۱	۰/۳۴	✓	
C15	۰/۳۸۵	۰/۵۸۵	۰/۹۷	-۰/۲		✓	C37	۰/۴۶۵	۰/۴۰۵	۰/۸۷	۰/۰۶	✓	
C16	۰/۵	۰/۵۹	۱/۰۹	-۰/۰۹		✓	C38	۰/۵۴۵	۰/۵۸۵	۱/۱۳	-۰/۰۴		✓
C17	۰/۴۴۵	۰/۶۱۵	۱/۰۶	-۰/۱۷		✓	C39	۰/۳۹۵	۰/۴۲۵	۰/۸۲	-۰/۰۳		✓
C18	۰/۳۸۵	۰/۳۹۵	۰/۷۸	-۰/۰۱		✓	C40	۰/۵۱	۰/۳۸	۰/۸۹	۰/۱۳	✓	
C19	۰/۴۴	۰/۳۸	۰/۸۲	۰/۰۶	✓		C41	۰/۳۸	۰/۳۶	۰/۷۴	۰/۰۲	✓	
C20	۰/۴۵۵	۰/۳۴۵	۰/۸	۰/۱۱	✓		C42	۰/۴۵۵	۰/۴۲۵	۰/۸۸	۰/۰۳	✓	
C21	۰/۳۹۵	۰/۲۵۲	۰/۹۲	-۰/۱۳		✓	C43	۰/۴۸۵	۰/۳۸۵	۰/۸۷	۰/۱	✓	
C22	۰/۳۹۵	۰/۵۵۵	۰/۹۵	-۰/۱۶		✓	C44	۰/۷۹	۰/۶۱	۱/۴	۰/۱۸	✓	

نمودار روابط علی که بر حسب مقادیر دو بردار اهمیت و رابطه ترسیم می‌شود، در شکل ۲، نشان داده شده است. در این نمودار، جمع عناصر هر سطر (D) نشانگر میزان تأثیرگذاری آن عامل بر سایر عامل‌های شبکه است. بر این اساس، عامل «پشتیبانی مدیریت ارشد از پروژه کلان‌داده‌ها به‌منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده» از بیشترین تأثیرگذاری، و عامل «اطمینان از صحت منابع داده‌ای (شناسایی دقیق منابع داده‌ها)» از کمترین تأثیرگذاری برخوردار است. جمع عناصر ستون (R) برای هر عامل نشان‌دهنده میزان تأثیرپذیری آن عامل از سایر عامل‌های شبکه است. بر این اساس، عامل «درک صحیح سازمان از مزایای استفاده از گزارشات تحلیلی مستخرج از

پروژه کلان‌داده‌ها» از میزان تأثیرپذیری بسیار زیاد، و عامل «به‌کارگیری الگوریتم‌های تحلیلی کارآمد برای درک رسیدن به داده‌ها و پردازش کلان‌داده‌ها و کاهش داده‌ها قبل از ذخیره‌سازی» از تأثیرپذیری بسیار کمی برخوردار هستند. بردار افقی (D+R)، میزان تأثیرگذاری و تأثیرپذیری عامل مورد نظر در شبکه است. به‌عبارت دیگر، هرچه مقدار (D+R) یک عامل بیشتر باشد، آن عامل تعامل بیشتری با سایر عامل‌های شبکه دارد. بر این اساس، عامل «پشتیبانی مدیریت ارشد از پروژه کلان‌داده‌ها به‌منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده» بیشترین تعامل را با سایر عامل‌های مورد مطالعه دارد. بردار عمودی (D-R)، قدرت تأثیرگذاری هر عامل را نشان می‌دهد. به‌طور کلی، اگر (D-R) مثبت باشد، متغیر یک متغیر علت محسوب می‌شود و اگر منفی باشد، معلول محسوب می‌شود. هرچه (D-R) عدد مثبت بزرگ‌تری باشد، به معنای تأثیر علی بیشتر عامل مورد نظر بوده و به‌عبارت دیگر، بیانگر اولویت بالاتر آن عامل است. عامل «پشتیبانی مدیریت ارشد از پروژه کلان‌داده‌ها به‌منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده» بزرگ‌ترین عدد مثبت مقدار (D-R) است و عامل «طبقه‌بندی، مدل‌سازی و نقشه‌برداری از نحوه ذخیره داده‌ها به همان صورتی که گرفته شده‌اند»، به‌طور عمده به‌دلیل ماهیت ساختار پیچیده آن‌ها کمترین مقدار (D-R) را دارد.



شکل ۲. نمودار روابط علی و معلولی میان عامل‌های شناسایی شده

**سؤال سوم) رتبه‌بندی عامل‌های مؤثر در کاهش مخاطره‌های شناسایی شده**

باتوجه به شکل ۲، عامل‌های مؤثر در کاهش مخاطره‌های شناسایی شده ناشی از استفاده از کلان‌داده‌ها در سازمان‌ها را می‌توان در قالب چهار ناحیه جدول ۸، دسته‌بندی و اولویت‌بندی نمود.

**جدول ۸. میزان اهمیت عامل‌ها با در نظر گرفتن چهار ناحیه**

D+J (بالا)	توجه خاص	توجه بالا
	توجه کم	توجه همیشگی
	D+J (بالا)	

عامل‌هایی که با سایر عامل‌ها از میزان تقابل بالایی برخوردار هستند و میزان تأثیرگذاری بالایی دارند، در ناحیه «توجه بالا» قرار داده می‌شوند.

عامل‌هایی که با سایر عامل‌ها از میزان تقابل بالایی برخوردار هستند و میزان تأثیرگذاری کمتری دارند، در ناحیه «توجه خاص» قرار داده می‌شوند.

عامل‌هایی که با سایر عامل‌ها از میزان تقابل کمتری برخوردار هستند و میزان تأثیرگذاری بالایی دارند، در ناحیه «توجه همیشگی» قرار داده می‌شوند.

عامل‌هایی که با سایر عامل‌ها از میزان تقابل کمتری برخوردار هستند و میزان تأثیرگذاری کمتری دارند، در ناحیه «توجه کم» قرار داده می‌شوند.

باتوجه به جدول ۸، می‌توان عامل‌های شناسایی شده را بر اساس میزان اهمیت (بسیار بالا، بالا، کم، خیلی کم) در قالب چهار ناحیه (توجه بالا، توجه خاص، توجه همیشگی، توجه کم) در جدول ۹، اولویت‌بندی نمود.

**جدول ۹. اولویت‌بندی عامل‌های شناسایی شده بر اساس میزان اهمیت**

	تقابل کم	تقابل زیاد
تأثیرگذاری زیاد	C15; C20; C29; C32; C40	C1; C7; C17; C34; C35; C34; C44
تأثیرگذاری کم	C10; C12; C29; C18; C19; C20; C21; C22; C23; C25; C28; C27; C41; C42; C33	C2; C3; C5; C6; C8; C9; C12; C13; C14; C16; C17; C24; C26; C30; C31; C38

با بررسی پاسخ‌های ارائه‌شده در قالب پرسشنامه، مصاحبه‌ها و جمع‌آوری اطلاعات از ادبیات پژوهش و با در نظر گرفتن قابلیت‌های موجود در سازمان از بُعد سازمانی، انسانی، فنی، مدیریتی و بررسی چرخه حیات کلان‌داده‌ها در سازمان جهت مواجهه و کاهش مخاطره‌های این حوزه استفاده شده است. بر اساس نمودار روابط علی و معلولی میان عامل‌های شناسایی‌شده در قالب شکل ۲ و اولویت‌بندی عامل‌های شناسایی‌شده بر اساس میزان اهمیت که در جدول ۹، به آن‌ها پرداخته شده، نتایج حاصل از این پژوهش بیانگر این مهم است که عامل (C1) تحت عنوان «پشتیبانی مدیریت ارشد از پروژه کلان‌داده‌ها به منظور ایجاد هماهنگی میان بخش‌های مختلف سازمان جهت استخراج داده‌ها»، بالاترین میزان تأثیرگذاری را به عنوان یک قابلیت سازمانی داراست. عامل (C44)، «توسعه و اجرای زیرساخت‌های مدیریت و پردازش» در اولویت دوم تأثیرگذاری بر سایر عوامل قرار دارد. به عبارتی، هرچه یک سازمان زیرساخت‌های خود را در خصوص پردازش کلان‌داده‌ها توسعه دهد، امکان کاهش مخاطره‌ها و چالش‌های کلان‌داده‌ها بیشتر است. با توجه به این که هزینه اجرا و پیاده‌سازی پروژه‌های کلان‌داده‌ها بسیار بالاست و از سوی دیگر، سازمان باید قادر به پرداخت آن باشد، عامل (C36)، «تأمین مالی پروژه‌های تجزیه و تحلیل کلان‌داده‌ها» در اولویت سوم تأثیرگذاری قرار دارد. برای این منظور، عامل (C2)، «درک صحیح سازمان از مزایای استفاده از گزارشات تحلیلی مستخرج از پروژه کلان‌داده‌ها» که در اولویت چهارم تأثیرگذاری قرار دارد، سازمان را در برداشتن گام‌های مؤثر جهت استفاده از کلان‌داده‌ها در سازمان یاری می‌کند. همچنین، برای نگهداری کلان‌داده‌ها به یک سیستم عامل و سیستم مرکز داده که جوایگوی نگهداری فرمت‌های مختلف داده باشد، نیاز است. از این رو، عامل (C13)، «وجود سیستم عامل و سیستم مرکز داده مناسب» از قابلیت‌های زیرساختی و فنی سازمان در اولویت پنجم تأثیرگذاری قرار دارد. همچنین، عامل (C5)، «نیروی انسانی متخصص» که بتواند داده‌های مورد نیاز را استخراج کرده و تجزیه و تحلیل مناسبی روی آن‌ها انجام دهد، در اولویت ششم تأثیرگذاری قرار دارد. همان‌طور که می‌دانید، اطلاعات باید طبق یک قانون و مقررات خاصی گردآوری و ذخیره شوند. از این جهت، عامل (C27)، «ایجاد قوانین و مقررات» در اولویت هفتم تأثیرگذاری قرار دارد. عامل (C35)، «پشتیبانی مدیران فناوری اطلاعات از حضور یک استراتژی رسمی» و عامل (C34)، «استفاده از روش‌های تجزیه و تحلیل و مدل‌سازی برای مدیریت کلان‌داده‌ها برای داشتن حداکثر تأثیر و ارزش تجاری» به ترتیب، در اولویت هشتم

و نهم قرار گرفتند. سرانجام، به منظور ایجاد امکان دسترسی به پایگاه‌های داده و تجمیع داده‌ها در یک انبار داده عامل (C۳)، «تعامل مناسب میان بخش‌های مختلف سازمان» در اولویت دهم از تأثیرگذاری قرار گرفت. همچنین، عامل (C۲)، «درک صحیح سازمان از مزایای استفاده از گزارشات تحلیلی مستخرج از پروژه کلان‌داده‌ها» و عامل (C۲۶)، «طبقه‌بندی، مدل‌سازی و نقشه‌برداری از نحوه ذخیره‌سازی داده‌ها» از بیشترین میزان تأثیرپذیری برخوردار هستند.

## ۶. نتیجه‌گیری و پیشنهادات

شناسایی چالش‌های کلان‌داده‌ها در سازمان امری بسیار ضروری در راستای ایجاد بستری مناسب جهت استفاده و پردازش این داده‌هاست. از این رو، در این تحقیق مجموعه‌ای از عواملی که بر روند استفاده از کلان‌داده‌ها در سازمان‌ها تأثیر دارند، در قالب یک چارچوب پیشنهادی مورد بررسی قرار گرفته و با استفاده از نظرات خبرگان و تکنیک «دیمتل فازی» تحلیل و اولویت‌بندی شدند. تحقیق حاضر دارای سه نوآوری در حوزه روش شناختی است: ۱) روابط علی را بین عوامل مرتبط با کاهش مخاطره‌های ناشی از استفاده از کلان‌داده‌ها در سازمان مدل‌سازی می‌نماید، ۲) می‌تواند ماهیت مبهم فرایند ارزیابی تأثیرگذاری عوامل بر روی یکدیگر را با استفاده از منطق فازی به‌طور کامل منعکس سازد، و ۳) با استفاده از تحلیل تأثیرگذاری عوامل بر روی یکدیگر، عوامل را اولویت‌بندی نموده و تأثیرگذارترین و تأثیرپذیرترین عوامل را مشخص نماید. در نظر گرفتن روابط علی بین عوامل، تصمیم‌گیرندگان سازمان را قادر می‌سازد تا تأثیر مستقیم و غیرمستقیم هرگونه پیشرفت در یک عامل بر عوامل دیگر را بررسی کنند. در کاربردهای عملی، این رویکرد می‌تواند با توجه به ویژگی‌ها و شرایط سازمان‌هایی که قصد مدل‌سازی دارند، خود را تطبیق داده و مدل‌سازی متناسب با شرایط سازمان انجام شود. این امر مستلزم تشکیل یک تیم شامل مدیران ارشد، مدیران بخش و کاربران کلیدی است که در فرایند مدل‌سازی و تحلیل شرکت خواهند کرد.

همچنین، نشان داده شد که چگونه این رویکرد می‌تواند به سازمان‌ها در شناسایی مناطق و عواملی که دارای اولویت بالاتری هستند و بایستی مورد توجه مدیریت قرار گیرند تا در صورت نیاز نسبت به تخصیص منابع برای بهبود وضعیت آن‌ها اقدام گردد، کمک کند. این موضوع نوآوری‌های عملیاتی این رویکرد است و آن را برای سازمان مفید و

ارزشمند می‌سازد تا بتواند عوامل تأثیرگذار و تأثیرپذیر را به‌طور مؤثر مدیریت نماید. با توجه به این که داده‌های سازمان‌ها به‌طور مداوم در حال تغییر و افزایش هستند و در ضمن، دید مدیران و کارشناسان نسبت به اهمیت کلان‌داده‌ها در سازمان تغییر می‌کند، می‌توان تحقیقات جدیدی در هر یک از قابلیت‌ها و در هر یک از مراحل چرخه حیات به‌صورت مجزا مورد بررسی قرار داد.

## References

- Al-Yadumi, Sohaib, Tan Ee Xion, Sharon Goh Wei Wei, and Patrice Boursier. 2021. Review on Integrating Geospatial Big Datasets and Open Research Issues. *IEEE Access* 9. IEEE. 10604-10620.
- Anshari, Muhammad, Mohammad Nabil Almunawar, Syamimi Ariff Lim, and Abdullah Al-Mudimigh. 2019. Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics* 15 (2): 94-101.
- Baltzan, Paige. 2019. *M: Information Systems*. 5th edition. United States: McGraw-Hill Education.
- Bharathi, S.Vijayakumar. 2017. Prioritizing and ranking the big data information security risk spectrum." *Global Journal of Flexible Systems Management* 18 (3): 183-201.
- Borkar, Vinayak R, Mishael J Carey, and Chen Li. 2012. "Big data platforms: what's next? *XRDS: Crossroads, The ACM Magazine for Students* 19 (1): 44-49.
- Brahanza, Ashley, Laurence Brooks, Daniel Nepelski, Maged Ali, and Russ Moro. 2017. Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research* 70: 328-337.
- Caesarius, Leon Michael, and Jukka Hohenthal. 2018. Searching for big data: How incumbents explore a possible adoption of big data technologies. *Scandinavian Journal of Management* 34 (2): 129-140.
- Capstera. 2018. List-of-common-business-capabilities. <https://www.capstera.com/>. <https://www.capstera.com/list-of-common-business-capabilities/>. (accessed Jan. 05, 2022)
- Che, Dunren, Safran Mejdil, and Peng Zhiyong. 2013. "From big data to big data mining: challenges, issues, and opportunities." *In International conference on database systems for advanced applications*. Berlin, Heidelberg: Springer. 1-15.
- Darwish, Dina. 2020. Developing and Implementing Big Data Analytics in Marketing. *International Journal of Data Science and Analysis* 6 (6): 183-203.
- Dash, Sabyasachi, Suchil Kumar Shakyawar, Mohit Sharna, and Sandeep Kaushik. 2019. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data* 6 (1): 1-25.
- Del Vecchio, Pasquale, Gioconda Mele, Valentina Ndou, and Giustina Secundo. 2018. Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management* 54 (5): 847-860.
- Díaz, Alejandro, Kayvaun Rowshankish, and Tamim Saleh. 2018. Why data culture matters. *McKinsey Quarterly* 3 (1): 36-53.
- Eastwood, Brian. 2021. The case for building a data-sharing culture in your company. *MIT Mangement Sloan School*. Sep 9. <https://mitsloan.mit.edu/ideas-made-to-matter/case-building-a-data-sharing-culture-your-company>. (accessed Sept. 9, 2021)
- Favaretto, M, E De Clercq, C.O Schneble, and B.S Elger. 2020. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS one* 15 (2): 0228987.

- Fikri, Noussair, Mohamed Rida, Nouredine Abghour, Khalid Moussaid, and Amina El Omri. 2019. An adaptive and real-time based architecture for financial data integration. *Journal of Big Data* 6 (1): 1-25.
- Gandomi, Amir, and Murtaza Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management* 35 (2): 137-144.
- Günther, Wendy Arianne, Mohammad H. Rezazade Mehrizi, Marleen Huysman, and Frans Feldberg. 2017. Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems* 26 (3): 191-209.
- Hariri, Reihaneh H, Erik M Fredericks, and Kate M Bowers. 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data* 6 (1): 1-16.
- Hee, Ong Choon, and Nanthinee Shanmugam. 2019. A Review of Human Resource Change Management Strategies in the Digital Era. *International Journal of Academic Research in Business and Social Sciences* 9 (3): 521-531.
- Holst, Arne. 2021. Statista." *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025*. Jun 7. <https://www.statista.com/statistics/871513/worldwide-data-created/>. (accessed June 7, 2021)
- Hu, Han, Yonggang Wen, Tat-Seng Chua, and Xuelong Li. 2014. Toward scalable systems for big data analytics: A technology tutorial. *IEEE access* 2. 652-687.
- Huang, Shuai, and W.Art Chaovallitwongse. 2015. "Computational optimization and statistical methods for big data analytics: Applications in neuroimaging." *In The Operations Research Revolution (INFORMS)* 71-88.
- Hussein, Abou\_el\_ela Abdou. 2020. Fifty-Six Big Data V's Characteristics and Proposed Strategies to Overcome Security and Privacy Challenges (BD2). *Journal of Information Security* 11 (04): 304-328.
- Inan, G. Gurkan, and Umit S Bititci. 2015. Understanding organizational capabilities and dynamic capabilities in the context of micro enterprises: a research agenda. *Procedia-Social and Behavioral Sciences* 210: 310-319.
- Janev, Valentina. 2020. "Chapter 7 Challenges for Exploiting the Potential of Big Data." *In Ecosystem of Big Data (In Knowledge Graphs and Big Data Processing)*, 3-19. Cham: Springer.
- Jeng, Don Jyh-Fu, and Gwo-Hshiung Tzeng. 2012. Social influence on the use of clinical decision support systems: revisiting the unified theory of acceptance and use of technology by the fuzzy DEMATEL technique. *Computers & Industrial Engineering* 62 (3): 819-828.
- Jha, Ashish Kumar, Maher AN Agi, and Eric WT Ngai. 2020. A note on big data analytics capability development in supply chain. *Decision Support Systems* 138: 113-382.
- Khan, Nawsher, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. 2014. Big data: survey, technologies, opportunities, and challenges. *The scientific World Journal* doi:<https://doi.org/10.1155/2014/712826>
- Lei, Jiale, and Linghe Kong. 2020. "Fundamentals of big data in radio astronomy." *In Big Data in Astronomy* (Elsevier) 29-58.
- Li, Yun, Manzhu Yu, Mengchao Xu, Jingchao Yang, Dexuan Sha, Qian Liu, and Chaowei Yang. 2020. "Big data and cloud computing." *In Manual of Digital Earth* (Springer) 325-355.
- Lu, Jie, Anjin Liu, Yiliao Song, and Guangquan Zhang. 2020. Data-driven decision support under concept drift in streamed big data. *Complex & Intelligent Systems* 6 (1): 157-163.

- Maritz, Juane, Sunet Eybers, and Marie Hattingh. 2020. Implementation Considerations for Big Data Analytics (BDA): A Benefit Dependency Network Approach. *Responsible Design, Implementation and Use of Information and Communication Technology* 481-492. 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2020, (pp. 481-492). Skukuza, South Africa. doi:[https://doi.org/10.1007/978-3-030-44999-5\\_40](https://doi.org/10.1007/978-3-030-44999-5_40)
- McAfee, Andrew, Erik Brynjolfsson, Thomas H Davenport, D.J Patil, and Dominic Barton. 2012. Big data: the management revolution. *Harvard business review* 90 (10): 60-68.
- Mehmood, Rashid, Royston Meriton, Gary Graham, Patrick Hennelly, and Mukesh Kumar. 2017. Exploring the influence of big data on city transport operations: a Markovian approach. *International Journal of Operations & Production Management*.
- Mesbahi, MohammadReza, Amir Masoud Rahmani, and Mehdi Hosseinzadeh. 2018. Reliability and high availability in cloud computing environments: a reference roadmap. *Human-centric Computing and Information Sciences* 8 (1): 1-31.
- Mikalef, Patrick (b), Rogier van de Wetering, and John Krogsite. 2021. Building dynamic capabilities by leveraging big data analytics: The role of organizational inertia. *Information & Management* 53 (6): 103-412.
- Mikalef, Patrick, Maria Boura, George Lakakos, and John Krogsite. 2020. The role of information governance in big data analytics driven innovation. *Information & Management* 57 (7): 103361.
- Mikalef, Patrick, Maria Boura, George Lekakos, and John Krogsite. 2019. Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research* 98: 261-276.
- Nguyen, Giang, Stefan Dlugolinsky, Martin Bobák, Viet Tran, Álvaro López García, Ignacio Heredia, Peter Malík, and Ladislav Hluchý. 2019. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review* 52 (1): 77-124.
- Obilikwu, Patrick, and Emeka Ogbuju. 2020. A data model for enhanced data comparability across multiple organizations. *Journal of Big Data* 7 (1): 1-25.
- Oussous, Ahmed, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, and Samir Belfkih. 2018. Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences* 30 (4): 431-448.
- Pal, Snakar K, Saroj K Meher, and Andrzej Skowron. 2015. Data science, big data and granular mining. *Pattern Recognit. Lett* 67 (2): 109-112.
- Palanisamy, Venketesh, and Ramkumar Thirunavukarasu. 2019. Implications of big data analytics in developing healthcare frameworks—A review. *Journal of King Saud University-Computer and Information Sciences* 31 (4): 415-425.
- Pérez-Martín, A, Agustín Pérez-Torregrosa, and M Vaca. 2018. Big Data techniques to measure credit banking risk in home equity loans. *Journal of Business Research* 89: 448-454.
- Raguseo, Elisabetta. 2018. Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management* 38 (1): 187-195.
- Rao, P.Ram Mohan, S.Murali Krishna, and AP Siva Kumar. 2018. Privacy preservation techniques in big data analytics: a survey. *Journal of Big Data* 5 (1): 1-12.
- Roush, Joe. 2020. "IT Infrastructure & Components: An Introduction." *bmc blogs*. May 13. <https://www.bmc.com/blogs/what-is-it-infrastructure-and-what-are-its-components/>. (accessed May 13, 2020)
- Saheb, Tahereh, and Tayebah Saheb. 2020. Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *Journal of Big Data* 7 (1): 1-26.
- Saltz, Jeffrey, Ivan Shamshurin, and Colin Connors. 2017. Predicting data science sociotechnical execution challenges by categorizing data science projects. *Journal of the Association for Information Science and Technology* 68 (12): 2720-2728.



- Shabbir, Muhammad Qasim, and Syed Babar Waheed Gardezi. 2020. Application of big data analytics and organizational performance: the mediating role of knowledge management practices. *Journal of Big Data* 7 (1): 1-17.
- Shah, Syed Iftikhar Hussain, Vassilios Peristeras, and Loannis Magnisalis. 2021. DaLiF: a data lifecycle framework for data-driven governments. *Journal of Big Data* 8 (1): 1-44.
- Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research* 70: 263-286.
- Song, Ma-Lin, Ron Fisher, Jian-Lin Wang, and Lian-Biao Cui. 2018. Environmental performance evaluation with big data: Theories and methods. *Annals of Operations Research* 270 (1): 459-472.
- Sonkar, Siddharth . 2020. "What is Big Data? A Quick Introduction for Analytics and Data Engineering Beginners." *Analytics Vidhya*. November 25. <https://www.analyticsvidhya.com/blog/2020/11/what-is-big-data-a-quick-introduction-for-analytics-and-data-engineering-beginners/>. (accessed Nov. 25, 2020)
- Stergiou, Christos, and Kostas E. Psannis. 2017. Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey. *International Journal of Network Management* 27 (3).
- Tian, Xinhui, Rui Han, Lei Wang, Gang Lu, and Jianfeng Zhan. 2015. Latency critical big data computing in finance. *The Journal of Finance and Data Science* 1 (1): 33-41.
- Wang, Hai, Xu Zeshui, and Pedrycz Witold. 2017. An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities. *Knowledge-Based Systems* 118: 15-30.
- Wang, JunPing, Wensheng Zhang, YouKang Shi, and ShiHui Duan. 2018. Industrial big data analytics: challenges, methodologies, and applications. *Transactions on Automation Science and Engineering*. IEEE. doi:<https://doi.org/10.48550/arXiv.1807.01016>
- Wang, Yicjuan, LeeAnn Kung, and Terry Anthony Byrd. 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change* 126: 3-13.
- Wigmore, Ivy. 2017. "data life cycle." *whatis (techtarget)*. July. <https://whatis.techtarget.com/definition/data-life-cycle>. (accessed July 2017)
- Willems, Stefan M, Sanne Abeln, K. Anton Feenstra, Remco de Bree, Egge F van der Poel, Robert J. Baatenburg de Jong, Jaap Heringa, and Michiel WM van den Brekel. 2019. The potential use of big data in oncology. *Oral oncology*: 8-12. doi:<https://doi.org/10.1016/j.oraloncology.2019.09.003>
- Wordsworth, Sarah, Brett Doble, Katherine Payne, James Buchanan, Deborah A. Marshall, Christopher McCabe, and Dean A. Regier. 2018. Using "big data" in the cost-effectiveness analysis of next-generation sequencing technologies: challenges and potential solutions. *Value in Health* 21 (9): 1048-1053.
- Zhou, Donghao, Zheng Yan, Yulong Fu, and Zhen Yao. 2018. A survey on network data collection. *Journal of Network and Computer Applications* 116: 9-23.

#### صدرا احمدی

متولد سال ۱۳۶۱، دارای مدرک تحصیلی دکتری در رشته فناوری اطلاعات، گرایش سیستم‌های اطلاعاتی از دانشگاه موناخ استرالیا است. ایشان هم‌اکنون استادیار گروه ابعاد انسانی فضای مجازی در پژوهشکده فضای مجازی دانشگاه شهید بهشتی است. کلان‌داده‌ها و تحلیل آن‌ها، حاکمیت داده و مدل‌های تصمیم‌گیری فازی، و تحلیل شبکه‌های اجتماعی از جمله علایق پژوهشی وی است.



#### ساناز قربانلو

متولد سال ۱۳۶۶، دانشجوی کارشناسی ارشد رشته مدیریت فناوری اطلاعات دانشگاه شهید بهشتی تهران است. ایشان هم‌اکنون پژوهشگر پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. امنیت اطلاعات، تحلیل و مدیریت مخاطره، اینترنت اشیا، مدیریت اطلاعات و دانش و تحلیلگری داده از جمله علایق پژوهشی وی است.

