



دبکده صله بهره وری



آزمان صله بهره وری ایران

عنوان دوره:

Training Course on Critical Big Data Analytics to Drive Productivity

آموزش مربیان در زمینه تجزیه و تحلیل مهم داده های کلان برای پیشبرد بهره وری

کد دوره:

20-IN-97-GE-TRC-A

محل برگزاری دوره:

تایلند - مجازی

تهیه و تنظیم:

امیرحسن اسدیان

مسئول امور زیرساخت های ارتباطی، شهرداری مشهد

فرشته خلج

وزارت ورزش و جوانان

مجید گلگلی

کارشناس سازمان اداری و استخدامی کشور

تاریخ برگزاری دوره:

26-28 April 2021

۶-۸ اردیبهشت ماه ۱۴۰۰



فهرست مطالب

| | |
|---------|-----------------------|
| ۱..... | معرفی |
| ۳..... | برنامه زمان‌بندی |
| ۵..... | چکیده |
| ۶..... | مقدمه |
| ۷..... | سخنرانی روز اول |
| ۱۴..... | سخنرانی روز دوم |
| ۱۷..... | سخنرانی روز سوم |
| ۲۵..... | جمع‌بندی و نتیجه‌گیری |

معرفی:

- مشخصات سخنرانان:

اسامی و مشخصات سخنرانان به شرح ذیل می باشد:

1. Ms. Poh Choo Tan, Adjunct Faculty, Singapore Management, University Singapore
2. Ms. Su Fen Cheong, Associate Director (Training and Consulting), the New Norma Pte. Ltd. Singapore
3. Mr. Jinsheng Ng, Director (Training and Consulting), the New Norma Pte. Ltd. Singapore

- مشخصات شرکت کنندگان:

مشخصات شرکت کنندگانی از کشورمان که موفق به شرکت در این دوره شدند به شرح ذیل است:

| نام و نام خانوادگی | محل خدمت | شماره همراه | ایمیل |
|--------------------|------------------------------|--------------|-------------------------|
| فرشته خلیج | وزارت ورزش و جوانان | ۰۹۱۲۴۹۶۰۶۸۹ | fekhalaj@gmail.com |
| مجید گلگلی | سازمان اداری و استخدامی کشور | ۰۹۱۲۷۸۷۵۳۷۱ | Majid.golgoli@gmail.com |
| صدیقه یعقوبی فاز | شهرداری مشهد | ۰۹۱۲۳۵۰۷۴۲۸ | syaghoobi99@gmail.com |
| امیرحسین اسدیان | شهرداری مشهد | ۰۹۱۵۵۲۵۳۸۸۵ | assadiyan-a@mashhad.ir |
| محمد رهبری | سازمان ملی بهره‌وری ایران | ۰۹۱۲۴۵۹۹۷۸۵ | mh.rahbari90@gmail.com |
| رویا حسینیان | شرکت سپندار | ۰۹۱۳۷۸۰۳۰۸۲۸ | r.hassanian@gmail.com |
| علی احسانی | سازمان امور مالیاتی | ۰۹۱۲۰۲۰۳۹۸۵ | info@ehsania.ir |

در مجموع، تعداد ۴۸ نفر از کشورهای مختلف در این رویداد به شرح و مشخصات، مطابق با جدول ذیل شرکت داشتند:

List of Participants

| Country of Residence | Title | Full Name | Present Position | Department | Name of Company/Organization |
|-----------------------------|-------|-----------------------|---|--|---|
| 1 Bangladesh | Mr. | Md. Roza Ahamed | Research Officer | | National Productivity Organisation |
| 2 Bangladesh | Mr. | Francis Sharma | Vice President, Mymensingh and Central Executive Committee Member of NASCIB | | National Association of Small and Cottage Industry of Bangladesh (NASCIB) |
| 3 Cambodia | Dr. | Angkara Bong | Deputy Director General | National Institute of Science, Technology and Innovation | Ministry of Industry, Science, Technology and Innovation |
| 4 Cambodia | Ms. | Ngov Sok You | Deputy Chief of Office | Department of Industrial Affairs | Ministry of Industry, Science, Technology and Innovation |
| 5 Cambodia | Mr. | Vutha Cheang | Deputy Director | STI Data Management | Ministry of Industry, Science, Technology and Innovation |
| 6 India | Mr. | Deepak Gupta | Deputy Director | Industrial Engineering | National Productivity Council |
| 7 India | Mr. | Jatin Saini | Assistant Director | Energy Management | National Productivity Council |
| 8 India | Mr. | Onum Samuel | Deputy Director | Environment and Monitoring and Evaluation | National Productivity Council |
| 9 Islamic Republic of Iran | Mr. | Ali Elseini | Group Manager | Data Science Unit | Iranian National Tax Administration |
| 10 Islamic Republic of Iran | Mr. | Amir Hassan Asadiyan | Responsible for Statistics, Evaluation and Communication Infrastructure | Public Relationship | Mashhad Municipality |
| 11 Islamic Republic of Iran | Dr. | Fereshteh Khataj | Head of Productivity and Statistics Group | IT Center | Ministry of Youth and Sport |
| 12 Islamic Republic of Iran | Mr. | Majid Golgoli | Steering Expert in Administrative and Recruitment Affairs of the Country | Business Management Affairs and Payment Systems | Administrative and Recruitment Affairs Organization |
| 13 Islamic Republic of Iran | Mr. | Mohammad Rahbari | Manager | Managing the Organization of National Productivity Network | National Productivity Organization of Islamic Republic of Iran |
| 14 Islamic Republic of Iran | Dr. | Roza Hassaniemadheh | Planning Manager | Planning Department | Spendaar Company |
| 15 Islamic Republic of Iran | Dr. | Sedigheh Yaghoobi Faz | Planning Expert | Deputy of Planning and Development | Mashhad Municipality |
| 16 Malaysia | Mr. | Josfia Bin Ismail | Deputy Director | Human Resource Management and Development | Malaysia Productivity Corporation |

List of Participants_1

| | | | | | | |
|----|-------------------|-----|---|---|---|---|
| 17 | Malaysia | Mr. | Mohamad Norjayadi bin Tamam | Deputy Director | Productivity Growth Department | Malaysia Productivity Corporation |
| 18 | Mongolia | Ms. | Battsetseg Munkhsukh | Business Analyst | Digital Banking Department | Golomt Bank |
| 19 | Pakistan | Mr. | Raja Jawad | Manager | Cloud and Computing | Inbox Business Technology Pvt. Ltd. |
| 20 | Pakistan | Ms. | Tahira Shahid | Manager and Lead | Data Analytics | Modern HealthLine |
| 21 | Philippines | Mr. | Louie Carl Recalde Mandapat | Senior Science Research Specialist | Department of Science and Technology Central Office | Planning and Evaluation Service - Information Technology Division |
| 22 | Philippines | Ms. | Mae Abigail Oberos Miralles | Senior Statistics Specialist | Statistical Methodology Unit | Philippine Statistics Authority |
| 23 | Philippines | Ms. | Mechelle Mendoza Viernes | Supervising Statistical Specialist | Poverty and Human Development Statistics Division, Social Sector Statistics Service | Philippine Statistics Authority |
| 24 | Republic of China | Ms. | Chu-Hui Liu | Engineer | Smart Manufacturing Management Department | China Productivity Center |
| 25 | Sri Lanka | Mr. | Jayathilaka Mudiyansele Saliya Prasanna Bandara Jayathilaka | Director (Development) | | Ministry of Labour |
| 26 | Sri Lanka | Ms. | Thudallage Shyama Nilanthi Fernando | Productivity Development Officer | | National Productivity Secretariat |
| 27 | Thailand | Mr. | Akarint Sittawaswong | Senior Manager | Corporate Strategy Department | Thai Credit Guarantee Corporation (TCG) |
| 28 | Thailand | Mr. | Bovornrach Saengsith | Senior Analyst, Data Analytics Office | Data Management and Analytics | Bank of Thailand |
| 29 | Thailand | Dr. | Chakree Bamrungwong | Civil Engineer, Expert Level Acting Director of Information Technology and Communication Center | Information Technology and Communication Center | Department of Rural Roads |
| 30 | Thailand | Dr. | Chatinai Chusai | Professional Mining Engineer | Office of Primary Industries and Mines Region 1 | Department of Primary Industries and Mines |
| 31 | Thailand | Mr. | Cholathis Cholpraves | Analyst | Safety, Security, Health, and Environment Management Department | PTT Public Company Limited |
| 32 | Thailand | Ms. | Natcha Techachainiran | Senior Engineering Officer | Spectrum Management Bureau | Office of the National Broadcasting and Telecommunications Commission |
| 33 | Thailand | Mr. | Nirat Rujimora | Researcher | Productivity Research Department | Thailand Productivity Institute (FTPI) |
| 34 | Thailand | Mr. | Nitruk Limlamai | Computer Officer | Information Technology Strategy | Metropolitan Waterworks Authority |
| 35 | Thailand | Ms. | Piyanun Wongkul | Computer Technical Officer, Professional Level | Office of Permanent Secretary | Ministry of Justice |

List of Participants_1

| | | | | | | |
|----|----------|-----|-------------------------|--|---|---|
| 36 | Thailand | Mr. | Prakit Roatphunyakit | Analyst | Sustainability Strategy Department | PTT Public Company Limited |
| 37 | Thailand | Ms. | Sukanya Noonsang | Manager | Information Technology and Knowledge Management | National Food Institute, Ministry of Industry, Thailand |
| 38 | Thailand | Ms. | Taweessap Sriwan | Statistician, Professional Level | Government Strategic Information Center | National Statistical Office |
| 39 | Turkey | Ms. | Cavide Pehlivan | SME Expert | Technology, Innovation and Domestication Department | KOSGEB |
| 40 | Turkey | Mr. | Cuma Uykun | Forest Engineer | Foreign Relations, Training and Research Department | General Directorate of Forestry |
| 41 | Turkey | Dr. | Ertug Guney | Head of Department | Planning, Programming and Coordination Unit | Trakya Development Agency |
| 42 | Turkey | Mr. | Hasan Cenk Dedeoğlu | Expert | Program Management Unit | Zafer Development Agency |
| 43 | Turkey | Ms. | Seda Elçi | EU Expert | Economical and Technical Relations | Ministry of Agriculture and Forestry |
| 44 | Vietnam | Mr. | Nguyen Phuong Lam | Manager | VNPI Ho Chi Minh Branch | Vietnam National Productivity Institute |
| 45 | Vietnam | Ms. | Nguyen Thi Phuong Nhung | Deputy Head of Division | Management and Innovation Solutions Division | Vietnam National Productivity Institute |
| 46 | Vietnam | Mr. | Pham Minh Nhan | Verification and Calibration Staff | Volume, Flow and Length Metrology Laboratory | Quality Assurance and Testing Center 2 (QUATEST2) |
| 47 | Vietnam | Mr. | Quach Khoa Duy | Productivity and Quality Management Consultant | DC | Quality Assurance and Testing Center 3 (QUATEST3) |
| 48 | Vietnam | Mr. | Vu Van Thao | Manager | Marketing | ISOCERT |

برنامه زمان بندی:

برنامه زمانی این دوره که در تاریخ ۶-۸ اردیبهشت ماه ۱۴۰۰ مطابق با 26-28 April 2021 برگزار گردید به شرح جدول ذیل است. شایان ذکر است که در جدول زیر، زمان ها به وقت کشور ژاپن ارائه شده است که از نظر زمانی حدود ۵ ساعت و ۳۰ دقیقه از کشور ایران جلوتر است.

| Time (Japan Time) | Agenda | Speaker |
|------------------------------|---|---|
| Day 1: Monday, 26 April 2021 | | |
| 13:30-14:00 | Registration/Zoom Connection | FTPI and APO Secretariat |
| 14:00-14:10 | Opening Session: | NPO Head/Representative Division Head/OIC |
| | Welcome Remarks by NPO | |
| | Opening Remarks by APO Secretariat | |
| 14:10-14:20 | Introduction of Resource Persons and Participants | OIC |
| | Introduction and Course Objectives (Note: APO video can be played here.) | |
| 14:20-15:00 | Session 1: Introduction to Big Data Analytics This session will discuss the big question in data science: asking the right questions for impactful, high-quality data analytics; it will also explore the different types of analytics, including descriptive, predictive, and prescriptive, for various economic sectors. | Ms. Poh Choo Tan Adjunct Faculty Singapore Management University Singapore |
| 15:00-15:10 | Break | |
| 15:10-15:50 | Session 2: 1) Data Management for Various Economic Sectors This session will explore data collection, integration, utilization, and querying for various economic sectors. | Ms. Poh Choo Tan |
| | 2) Sustainable Digital Service Delivery This session will discuss a case in point: The digital divide and inclusion, with approaches to solutions. | |
| | | |
| 15:50-16:00 | Break | |
| 16:00-16:40 | Session 3: Increasing Productivity with Technologies This session will discuss another case in point: Platform development to meet public demand. | Ms. Su Fen Cheong Associate Director (Training and Consulting) The New Norma Pte. Ltd. Singapore |
| 16:40-16:50 | Break | |
| 16:50-17:30 | Session 4: Current Status and Future Trends of Big Data Analytics This session will explain what data analytics mean for businesses and how trends such as augmented | Mr. Jinsheng Ng Director (Training and Consulting) The New Norma Pte. Ltd. Singapore |

| Time (Japan Time) | Agenda | Speaker |
|---------------------------------|--|--------------------------------------|
| | analytics, artificial intelligence, machine learning, and natural language processing will speed up decision-making in the future. | |
| End of Day 1 | | |
| Day 2: Tuesday, 27 April 2021 | | |
| 13:30–14:00 | Registration/Zoom Connection | FTPI and APO Secretariat |
| 14:00–14:40 | Session 5: Harnessing the Power of Big Data Visualization This session will recapitulate the points learned on the previous day and explore the power of data visualization through Excel and Tableau to make informed decisions in business operations. | Ms. Su Fen Cheong |
| 14:40–15:20 | Session 6: Making Sense of Graphs and Charts for Productivity Gains This session will continue the data visualization journey with sense-making explanations of graphs produced for decision-making. | Mr. Jinsheng Ng |
| 15:20–15:30 | Break | |
| 15:30–16:10 | Session 7: Data-driven Methodologies and Step-by-step Processes for Better Decision-making This session will discuss the various methodologies available for data analysts to create a framework for more informed decision-making within enterprises or government agencies. | Ms. Su Fen Cheong |
| 16:10–16:20 | Break | |
| 16:20–17:00 | Session 8: Hands-on Activities This session will give participants hands-on exercises on creating graphs and charts that enhance understanding of how they impact productivity in the workplace. | Mr. Jinsheng Ng Ms. Su Fen Cheong |
| End of Day 2 | | |
| Day 3: Wednesday, 28 April 2021 | | |
| 13:30–14:00 | Registration/Zoom Connection | FTPI and APO Secretariat |
| 14:00–14:40 | Session 9: What Is Machine Learning and How Can I Use It Effectively for Productivity Gains? This session explains the fundamentals of machine learning and how current machine-learning techniques could enhance business operations. | Mr. Jinsheng Ng |
| 14:40–15:20 | Session 10: Automation with Essential Machine-learning Techniques | Ms. Su Fen Cheong |

| Time (Japan Time) | Agenda | Speaker |
|-------------------|--|---|
| | This session will discuss essential machine-learning techniques in greater detail. | |
| 15:20–15:30 | Break | |
| 15:30–16:10 | Session 11: Essential Types of Machine-learning Algorithms This session will introduce the fundamental machine-learning algorithms commonly used in operational management. | Mr. Jinsheng Ng |
| 16:10–16:50 | Session 12: Hands-on Activities The session provides hands-on activities exposing participants to common applications of machine-learning techniques. | Mr. Jinsheng Ng Ms. Su Fen Cheong |
| 16:50–17:00 | Closing Session: Vote of Thanks Closing Remarks by NPO Administrative Announcements by APO Secretariat (Evaluation, Certificates) | Selected Participants NPO Head/Representative OIC |
| End of the Course | | |

پایان این دوره نیز با برگزاری پنل بحث و گفتگو و جمع‌بندی با پاسخ به شرکت‌کنندگان و در نهایت تکمیل فرم‌های ارزشیابی به اتمام رسید.

چکیده:

داده‌کاوی یکی از تخصص‌های بروز و مورد نیاز تقریباً تمام کسب‌وکارهای داده محور است. داده‌کاوی فرآیندی است که طی آن دیتای خام و بی‌مفهوم به اطلاعات مفید و با ارزشی برای تصمیم‌گیری و تعیین اهداف استراتژیک مشخص سازمانی تبدیل می‌شوند. داده‌کاوی یا دیتامینینگ در واقع علمی میان رشته‌ای است که علوم کامپیوتر، هوش مصنوعی، یادگیری ماشین، ریاضی و آمار را دربر می‌گیرد. با ساخت فرآیند داده‌کاوی، می‌توان دیتاهای حجیم و پیچیده سازمان‌ها و مؤسسات تحقیقاتی را کشف و الگوهای پنهان آن را تحلیل نمود. به طور کلی می‌توان داده را برابر با طلا دانست. هر چقدر داده‌های موجود در پایگاه داده یک شرکت افزایش پیدا کند، سرمایه شرکت نیز بیشتر می‌شود. حال، برای تبدیل کردن این داده‌های حجیم و تمیز نشده به اطلاعات مفید و گزارشات مناسب برای تصمیم‌گیری، نیاز است تا از داده‌کاوی استفاده کرد تا با کشف الگوهای پنهان از پردازش میلیون‌ها داده، نتایج شگفت‌انگیزی را بدست آورد و تحلیل کرد. تکنولوژی **big data** در حال بررسی داده‌های حجیم، پیچیده و در حال رشد است که از منابع مستقل و پراکنده تولید می‌گردند. با توسعه‌ی سریع شبکه‌ها، رسانه‌های ذخیره‌سازی و افزایش گنجایش مجموعه‌های داده‌ای، **big data** در تمام حوزه‌های علم و فناوری از جمله علوم فیزیکی، زیستی و پزشکی به سرعت در حال رشد و گسترش است. نظریه‌ی HACE ویژگی‌های تکنولوژی **big data** را مشخص می‌کند و از چشم‌انداز داده‌کاوی یک مدل پردازشی برای **big data** پیشنهاد می‌کند. کار با داده‌های بزرگ شامل جمع‌آوری داده‌ها از منابع اطلاعاتی، کاویدن و تحلیل آن‌ها، مدل کردن آن‌ها بر اساس ویژگی‌های مورد نظر و فراهم کردن تمهیدات امنیتی داده‌ها است.

کلمات کلیدی: داده‌کاوی، داده، حجیم، Big Data

مقدمه:

امروزه با افزایش استفاده از محیط‌های فراگیر، محیطی که نیاز بشر را در هر زمان و هر مکان فراهم می‌سازد فناوری اطلاعاتی و ارتباطاتی در حوزه تعاملی انسانی و ماشینی با سرعتی غیرقابل تصور پیشرفت نموده است. پیش‌بینی و شناخت چالش‌ها، فرصت‌ها و روند توسعه فناوری‌های محیط‌های فراگیر برای تقویت حوزه‌های تحقیقاتی راهبردی، تبدیل چالش به فرصت، فربه‌سازی نیازهای پیاده‌سازی برای کشورها، به یک ضرورت تبدیل شده است. بشر از بدو تولد شروع به تجزیه و تحلیل داده‌ها کرده است. انسان از دوران کودکی در معرض انبوهی از سیگنال‌ها قرار داشته که قادر به دریافت آن‌ها از طریق چشم، گوش، بینی، زبان و اعصاب بوده است. همچنین مغز او به سرعت آموخته که چگونه این سیگنال‌ها را تجزیه و تحلیل کند و نسبت به آن‌ها واکنش نشان دهد. پس از زبان گفتار، زبان نوشتاری شکل گرفت و مخازن وسیعی از داده‌ها را ایجاد کرد که تا به امروز قابل تجزیه و تحلیل می‌باشند. بنابراین بسیاری از مردم اظهار می‌کنند که همین حالا وارد عصر بزرگ داده‌ها شده‌اند، اما باید به این نکته توجه داشت که داده‌ها همیشه در اطراف ما بوده‌اند و فقط مقدار داده‌های ثبت‌شده شروع به شکوفایی کرده و محبوبیت زیادی پیدا کرده‌اند.

داده‌کاوی و یادگیری ماشین، در عین شباهت با هم متفاوت هستند. اغلب اوقات حتی داده‌کاوی را با نام یادگیری ماشین خطاب می‌کنند. هر چند روز به روز این دو به یکدیگر نزدیک‌تر و شبیه‌تر می‌شوند. یادگیری ماشین از تلاش برای ساخت هوش مصنوعی گذر کرد و اکنون هدف اصلی آن ساخت ماشینی است که بتواند یاد بگیرد و خود را با اطلاعات جدید وفق دهد. اولین یادگیری ماشین به سال ۱۹۵۷ برمی‌گردد که مدل *perceptron* اختراع شد. این مدلی است که از نورون‌های مغز انسان ساخته شد. این مدل آغازگر مدل شبکه عصبی بود که در اواخر دهه ۱۹۸۰ ایجاد شد. از دهه ۱۹۸۰ تا دهه ۱۹۹۰ متد درخت تصمیم بسیار محبوب شد. *SVM* در اواسط دهه ۱۹۹۰ اختراع شد و از آن زمان در صنعت به طور گسترده‌ای استفاده می‌شود. رگرسیون منطقی، یک متد قدیمی در آمار است که در یادگیری ماشین از سال ۲۰۰۱، که کتاب یادگیری آمار منتشر شد، رشد چشم‌گیری داشته است. داده‌کاوی و یادگیری ماشین، در عین شباهت با هم متفاوت هستند. اغلب اوقات حتی داده‌کاوی را با نام یادگیری ماشین خطاب می‌کنند. هر چند روز به روز این دو به یکدیگر نزدیک‌تر و شبیه‌تر می‌شوند. در این وبینار ابتدا به تعریف اولیه مفاهیم کاربردی پرداخته شد و سپس یک نمونه کار آزمایشگاهی نیز ارائه گردید که در ادامه به شرح مبسوط محتوای هر ۳ روز می‌پردازیم.

گزارش سخنرانان:

سخنرانی روز اول:

- + Data Management for Various Economic Sectors
- + Sustainable Digital Service Delivery
- + Increasing Productivity with Technologies
- + Current Status and Future Trends of Big Data Analytics

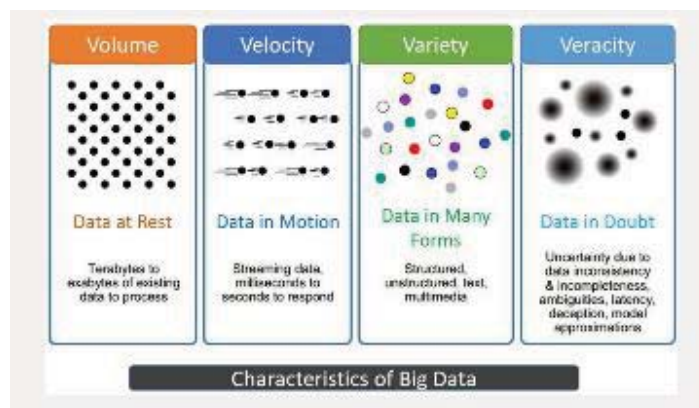


روز نخست با معرفی مسئولین و شرکت کنندگان در وبینار آغاز گردید. سپس یک عکس مجازی از حاضرین در وبینار نیز تهیه گردید. (شکل زیر) محتوای پنل اول با موضوع " تجزیه و تحلیل داده های بزرگ انتقادی برای تولید بهره‌وری " آغاز گردید.

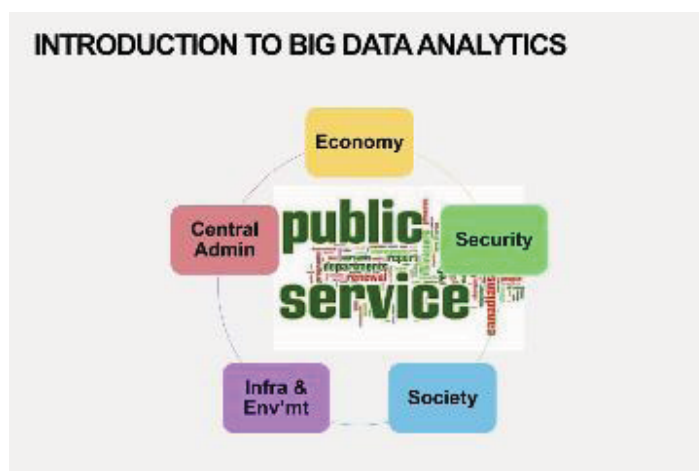


این ارائه که توسط خانم TAN Poh Choo ارائه گردید بیشتر پیرامون چرایی و چگونگی مدیریت داده برای سرویس های عمومی و افزایش بهره‌وری بحث و بررسی صورت گرفت. به عنوان شروع بحث چهار گروه و دسته‌بندی برای شناسایی داده‌های حجیم ارائه گردید:

- داده‌ها در حالت استراحت
- داده‌ها در حالت حرکت
- داده‌ها در اشکال مختلف
- داده‌های مشکوک



همچنین به عنوان چرخه تحلیلی در حوزه داده‌های حجیم شکل زیر معرفی گردید:



در ادامه بحث با هدف مشارکت شرکت‌کنندگان و دریافت بازخورد کلی اعضای حاضر در پنل، یک سوال مطرح گردید مبنی بر اینکه سازمان متبوع هر کدام از حاضرین بیشتر در چه حوزه‌ای از داده‌های شهری درگیر می‌باشد؟ نحوه شرکت در این سوال را در ادامه مشاهده می‌نمایید.

The screenshot shows a survey interface. On the left is a QR code with the number **6911 6612** below it. On the right is the survey content:

From Hindsight to Insight to Foresight

Analytics types: Descriptive, Diagnostic, Predictive, Prescriptive, Proactive.

Which type of analytics does your organization use most commonly?
*You may choose multiple options.

- ☐ Descriptive
- ☐ Diagnostic
- ☐ Predictive
- ☐ Prescriptive
- ☐ Proactive

Submit

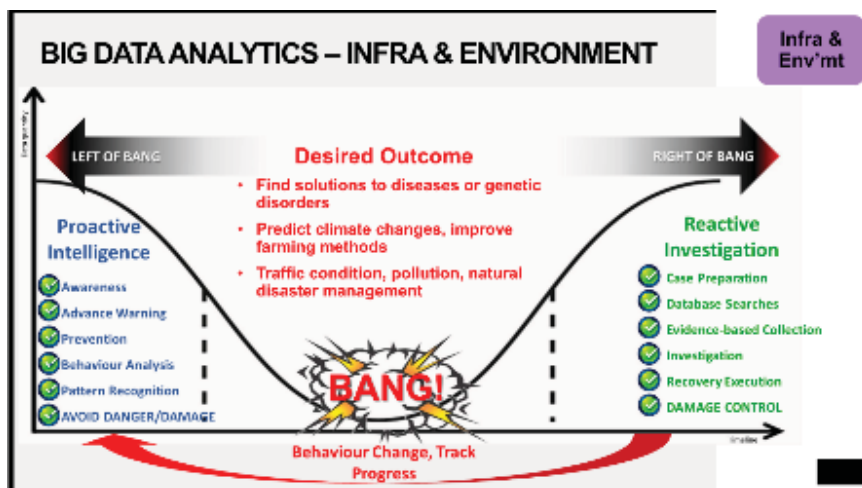
در حوزه زیرساخت اطلاعاتی و تغییرات رفتاری در نحوه مبادله داده و همچنین شناسایی شاخص‌های درگیر در محیط و تأمین زیرساخت‌های لازم برای رسیدن به خروجی مدنظر از یک مجموعه داده، بحث و بررسی صورت گرفت. در این بخش تأثیر عوامل متعددی همچون هوش پیشگیرانه و تحقیق واکنشی به عنوان دو تا از اصلی‌ترین مفاهیم ارائه گردید.

هوش پیشگیرانه به عواملی همچون:

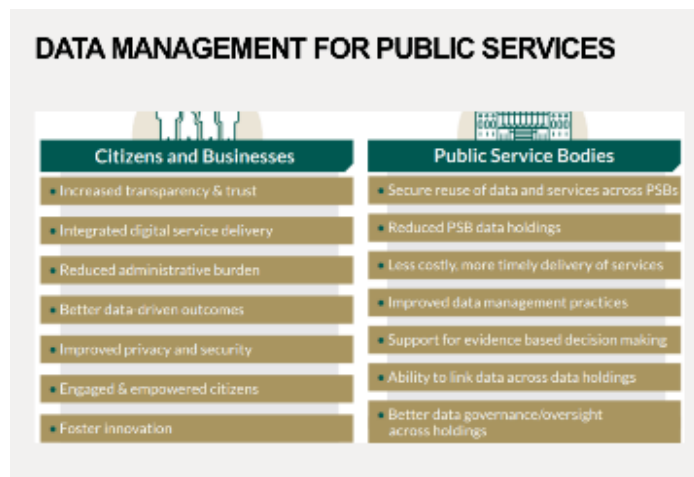
- آگاهی
 - هشدارهای پیشرفته
 - احتیاط
 - تحلیل رفتاری
 - تشخیص الگو
 - پیشگیری از خطر و خرابی
- بستگی دارد.

از سوی دیگر تحقیق واکنشی نیز به‌طور مستقیم با عواملی همچون:

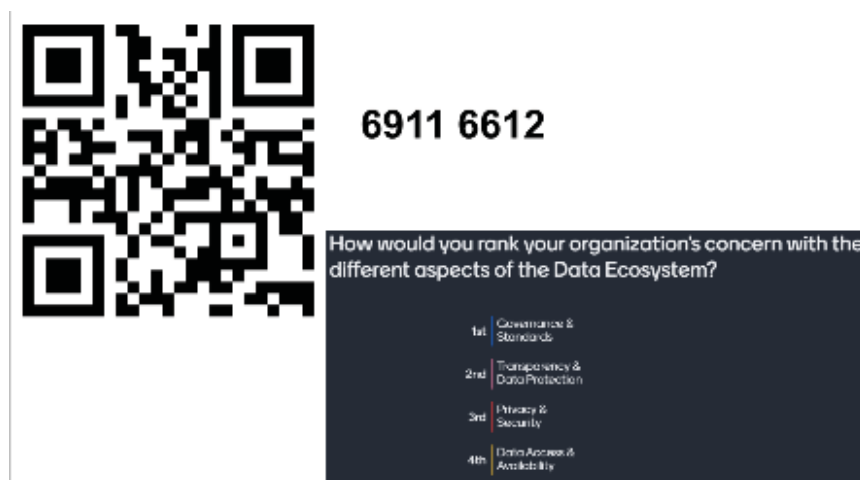
- آماده‌سازی شرایط
 - جستجوهای بانک اطلاعاتی
 - جمع‌آوری مشاهدات
 - سرمایه‌گذاری
 - اجرای فرآیند برگشت‌پذیری
 - و
 - کنترل خرابی
- در ارتباط می‌باشد.



طبق دیدگاه دولت ایرلند، استراتژی داده‌های سرویس عمومی بیان می‌دارد که برای ایجاد یک داده اکوسیستم با هدف بهبود نحوه مدیریت و استفاده مجدد از داده‌ها در امن کردن، کارآمدسازی، و شفاف‌نمودن، بهره‌مندی از شهروندان، مشاغل و سیاست‌گذاران است. برخی از این دیدگاه‌ها را در شکل زیر مشاهده می‌فرمایید:



در خلل برگزاری کنفرانس، پرسش دیگری از حاضرین مطرح گردید که توسط آن میزان ارتباط دستگاه‌های حاکمیتی هر یک از کشورها با جنبه‌های مختلف اکوسیستم داده‌ها بررسی گردید که طی آن استانداردها و حکمرانی، شفافیت و حفاظت از داده، امنیت و حریم خصوصی و دسترس‌پذیری به ترتیب رتبه‌های اول تا چهارم را به دست آوردند.



در ادامه مبحث، موضوع **افزایش بهره‌وری با استفاده از فناوری** مطرح گردید. در این بخش فرآیندهای زیر به عنوان مهم‌ترین اجزاء معرفی گردید:

- به درخواست شهروند، آرزوهایش، افق آینده و نیازهایی که انتظار دارد، گوش کنید
- پیشینه و تاریخچه آن‌ها را بشناسید، تجزیه و تحلیل را در آن اعمال و بهترین اقدام را شناسایی کنید
- ارتباط به موقع، راحت و سازگار در زمان واقعی یا زمان مناسب، برقرار کنید.

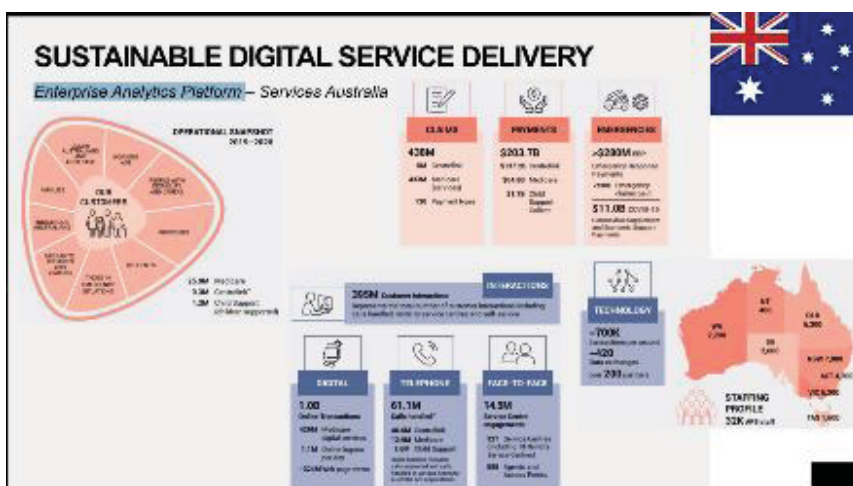
همچنین برخی نمونه‌های اجرایی در همین رابطه معرفی گردید. مانند:

✓ تولید یک برنامه موبایلی در ترکیه با نام "LifeSG App"، با هدف ادغام و به هم پیوستن خدمات در سراسر جهان از نیازهای فردی و سازمانی با هم

✓ تولید برنامه‌های مختلف تلفن همراه برای نظارت و کنترل همه‌گیری (در خصوص ویروس کوید-۱۹)

✓ طراحی و اجرای سریع ارتباطات، حرکت و تحویل بدون تماس بسته‌های کمکی (در خصوص ویروس کوید-۱۹)

بخش چهارم و پایانی صحبت‌های خانم TAN Poh Choo، به موضوع تحویل سرویس دیجیتال پایدار اختصاص داشت. در این بخش به عنوان نمونه یک پلتفرم تجزیه و تحلیل سازمانی از کشور استرالیا مورد نقد و بررسی قرار گرفت و نحوه تأثیر تکنولوژی و کاربرد آن در مدیریت حجم داده‌های کلان، به همراه چالش‌هایی که با آن مواجه بودند معرفی گردید. آمار و محتوای کلی آن را در شکل زیر مشاهده می‌فرمایید:



برخی از چالش‌های این پلتفرم عبارت بودند از:

- رشد طبیعی و تقاضا برای تجزیه و تحلیل بیشتر
- تبدیل شدن به یک آژانس خدمات مشترک
- عدم دسترسی به برنامه‌ها و خدمات کنترل و محدودیت
- وابستگی زیاد به IT برای بارگذاری داده‌ها، ارائه گزارش و برنامه‌ریزی بدون سلف سرویس
- مدت زمان طولانی
- محاسبه حجم کاری که در ساعاتی اوج ۱۰٪ است

جمع‌بندی و نتیجه‌گیری

ایشان در پایان صحبت‌های خود چند نکته کلیدی در مواجهه با شرایط مهم را بیان نمودند:

✓ در ابتدا:

- با توانایی‌ها شروع کنید، نه استراتژی‌ها
- از کوچک شروع کنید، بزرگ فکر کنید
- کاربران را در مرکز خود قرار دهید
- تیم‌های چندرشته‌ای تشکیل دهید

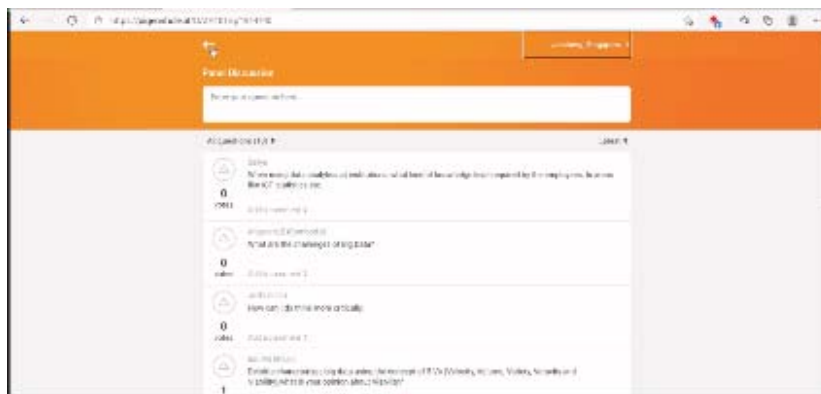
✓ در زمان مقیاس‌بندی:

- رهبری مهم است
- تنظیم KPI ها ، اندازه گیری، بازیابی، تکرار فراموش نشود
- داده ها را به جریان در آورید
- سیستم عامل ها، نه محصولات
- این فقط در مورد فناوری نیست

✓ در زمان پایداری سیستم:

- ابتدا شرایط خود را بررسی کنید
- موفقیت را به نمایش بگذارید، اعتماد را جلب کنید
- بهبود مستمر برای نتایج پایدار داشته باشید.

در پایان نیز به منظور بهره گیری از شرایط یکسان در طرح پرسش و پاسخ های حاضرین در جلسه، یک پلتفرم بسیار ساده و کاربردی معرفی گردید که افراد سوالات خود را در آن مطرح می کردند و سایرین به تناسب به هر سوال امتیاز می دادند. به این طریق هر سوالی که بیشترین امتیاز را داشت مطرح و پاسخ داده می شد.



سخنرانی روز دوم:

- ✚ Harnessing the Power of Big Data Visualization
- ✚ Making Sense of Graphs and Charts for Productivity Gains
- ✚ Data-driven Methodologies and Step-by-step
- ✚ Processes for Better Decision-making



دکتر جینشنگ جلسه دوم را با طرح سؤالاتی نسبت به مطالب ارائه شده در جلسه اول توسط سرکار خانم دکتر پو چون تا شروع کردند. ایشان توضیحاتی در خصوص مطالب مورد نظر برای طرح در سه ساعت جلسه دوم گفتند و در مورد مباحث عنوان شده در بخش معرفی تحلیل داده‌های بزرگ مطالب رو آغاز کردند. داده‌های حجیم به چه نوع داده‌ای اطلاق می‌شود و نقش آن در خدمات عمومی به چه صورتی است؟

ایشان در ادامه انواع آنالیزها و تحلیل‌ها بر داده‌های حجیم را یادآور شدند و دسته‌بندی آن‌ها به صورت توصیفی، پیش‌بینی و آنالیزهای قبل از پیش‌بینی بر اساس داده‌ها و مقایسه این تحلیل‌ها با یکدیگر را بیان کردند. ایشان در ادامه از مهم‌ترین کلمه کلیدی جهانی که در شهر و کشور و اخبارها مطرح است در خصوص (text analysis) سوال کردند که در پاسخ موضوع کوید ۱۹ طرح شد و قدرت تحلیل متن در این زمینه بیان شد. سوال بعدی در خصوص خانم فلورانس نایتینگل طرح شد نایتینگل ۱۸۱۹ متولد شده، تحصیلات او در حوزه آمار بوده، بعنوان پرستار داوطلب در بیمارستان برای کمک به سربازان ارتش انگلیس فعالیت می‌کرد و صد سال پیش درگذشته است. او روزنامه‌نگار نیز بود. او تصمیم گرفت نسبت به جمع‌آوری داده‌های عفونت باکتریایی بین سربازان اقدام کند و از روش‌های آماری، چارت‌های آماری، نمودارهای دایره‌ای و گرافیک آماری و ارائه بصری اطلاعات برای گزارش‌های ارائه شده در داده‌های سلامت استفاده می‌کرد، که راهی بسیار موثر برای بیان داده‌ها بود.

در ادامه بخشی از سخنرانی دکتر هانس روزلینگ، استاد بهداشت بین‌الملل در موسسه کارولینسکا سوئد، میزبان مستند جدیدی در BBC با نام The Joy of Stats است که نگاهی به وسعت و عمق داده‌های موجود امروز برای تحلیلگران و شهروندان خصوصی دارد، پخش شد. در این کلیپ، راسلینگ علاقه اصلی خود به بهداشت جهان را نشان می‌دهد، امید به زندگی و درآمد طی ۲۰۰ سال گذشته را ردیابی می‌کند تا هم پیشرفت قابل توجهی را که ایجاد شده است نشان دهد، بلکه همچنین شکاف عظیم بین افراد سطح بالا (بسیار ثروتمند و افراد سالم) و افراد سطح پایین (افراد بسیار فقیر و بیمار) را بیان می‌کند. روزلینگ با ارائه آماری به روش‌های نوآورانه طرفدار آوازی (و بصری) برای گسترش دانش مردم در جهان بوده است. روزلینگ در این کلیپ تماشایی از "جذابیت آمار" داستان جهان را در ۲۰۰ کشور بیش از ۲۰۰ سال با استفاده از ۱۲۰.۰۰۰ شماره - فقط در چهار دقیقه - بیان می‌کند. با ترسیم امید به زندگی در برابر درآمد برای هر کشور از سال ۱۸۱۰، هانس نشان می‌دهد که چگونه جهانی که ما در آن زندگی می‌کنیم با جهانی که اکثر ما تصور می‌کنیم تفاوت اساسی دارد.

بخش بعدی مباحث طرح شده در خصوص ملموس‌تر کردن گراف‌ها و چارت‌های آماری برای افزایش بهره‌وری طرح شد. در این بخش به داشبوردهای مدیریتی و نرم‌افزارهای مناسب برای طراحی داشبوردها برای تصمیم‌گیری اشاره شد. از جمله به نرم‌افزارهای مناسب Power BI, Tableau, Qlikview برای داده‌های بزرگ با تعداد چند میلیون رکورد اشاره شد. سپس داده‌های اثر کوید ۱۹ بر روی اشتغال در کشور سنگاپور، به عنوان مطالعه موردی مورد بحث و بررسی قرار گرفت. در این مطالعه اشاره شد بعلاوه اثر بیماری همه‌گیر Covid-19 کارگران با دستمزد پایین بسیار آسیب دیده‌اند. ایشان در ادامه داشبورد مرتبط با داده‌های اشتغال و کوید ۱۹ و PMET (professional, manager, engineer, techniques) در سنگاپور توضیحاتی دادند و انواع چارت‌ها و گراف‌های نمایش داده شده برای گزارش این داده‌ها در داشبورد مدیریتی بیان کردند. نمودار دایره‌ای، نمودار خطی و هیستوگرام و نمودار فراوانی از جمله نمودارهایی بودند که در داشبورد مدیریتی برای نمایش داده‌های به صورت تصویری استفاده شده‌اند. در ادامه

تفاوت انواع گراف و دسته‌بندی آن‌ها برای نمایش داده‌ای بیان شد. در این دسته‌بندی به گراف‌هایی که برای تجزیه و تحلیل داده‌ها (Comparative Analysis)، یعنی نمودار دایره‌ای (Pie chart) و نمودار میله‌ای (Bar chart) اشاره شد و به دسته‌بندی گراف‌های مرتبط با تحلیل‌های روندی و نحوه توزیع (Trends and Distributions) که شامل نمودار خطی (chart-Line) و هیستوگرام (Histogram) عنوان شد. سپس بیان شد که نمودار خطی برای تحلیل‌های سری زمانی و پیش‌بینی استفاده می‌شود. سپس در خصوص نحوه انتخاب چارت و گراف در داشبوردهای مدیریتی بر اساس سه‌گام اشاره شد. گام اول، داده (Data)؛ گام دوم، ابزار نمایش داده (Tools)؛ و گام سوم بینشی که این نمودار به ما ارائه می‌دهد (Insights). سپس به صورت عملی بر روی داده‌های سری زمانی مرتبط با جرایم بر اساس سال و بر اساس فایل داده تعریف شده در اکسل به‌طور خلاصه بیان شد:

Data + Tools → Insights

Time Series + Line → Trend analysis

در ادامه نیز در خصوص فراخوانی داده‌ها در نرم‌افزار متن‌باز Tableau توضیح داده شد و با داده‌های سری زمانی دیگری داده‌های ساختاریافته (Structure) و ساختاریافته (Unstructured)، همچنین داده‌های عددی (Numeric) و نحوه دسته‌بندی آن‌ها (Category) توضیحاتی ارائه شد. در ادامه کلاس نیز نصب نرم‌افزار متن‌باز Tableau با توجه به ویندوزهای مختلف به صورت گام به گام بیان شد و فرصتی برای نصب این نرم‌افزار توسط شرکت‌کنندگان داده شد. در این راستا به سوالات مختلف شرکت‌کنندگان در خصوص انواع نمودارها و سایر موارد مرتبط پاسخ داده شد. در ادامه جلسه نیز رسم انواع نمودارهای خطی، روندی و مقایسه‌ای با استفاده از نرم‌افزار داشبوردساز Tableau و تحلیل نمودارها بیان شد.

سخنرانی روز سوم:

- + What Is Machine Learning
- + Automation with Essential Machine-learning Techniques
- + Essential Types of Machine-learning Algorithms



در روز سوم وینار، در خصوص یادگیری ماشین^۱، کاربردها و انواع الگوریتم‌های آن مطالبی ارائه شد. در انتها چند برنامه و نرم‌افزار کاربردی در این زمینه معرفی شد و با تمرین یک مثال از رگرسیون خطی ساده در محیط مایکروسافت اکسل و تحلیل خروجی آن، وینار روز سوم خاتمه یافت.

یادگیری ماشین چیست و چگونه می‌تواند به طور موثر برای دستاوردهای بهره‌وری استفاده شود؟

روش‌های یادگیری ماشین، سیستم‌های مختلف را قادر می‌کنند که یاد بگیرند، بررسی کنند و به ما پیشنهادهای کاربردی ارائه دهند. این سیستم‌ها به مرور زمان که با داده‌ها، شبکه‌ها و افراد تعامل دارند، باهوش‌تر می‌شوند. با استفاده از رویکردهای یادگیری ماشین و هوش مصنوعی، این سیستم‌ها قادر هستند ما را در حل مسائل مهم، کاربردی و روزمره یاری دهند. غالباً این کار با استفاده از داده‌هایی انجام می‌شود که به دلیل حجم زیاد و یا ماهیت نامفهوم، برای ما انسان‌ها چندان قابل استفاده نیست.

تفاوت بین داده‌کاوی و یادگیری ماشین

داده‌کاوی زیر مجموعه‌ای از علم تجزیه و تحلیل تجارت بوده و بر آموزش کامپیوتر متمرکز می‌باشد. همچنین نحوه شناسایی الگوها، روابط یا ناهنجاری‌های ناشناخته در مجموعه داده‌های بزرگ بوده که بعداً انسان‌ها می‌توانند از آن‌ها برای حل یک مشکل تجاری استفاده کنند. اما در حقیقت یادگیری ماشین زیرمجموعه‌ای از هوش مصنوعی بوده و اشاره بر نحوه یادگیری تجزیه و تحلیل داده‌های بزرگ و “یادگیری” الگوهای موجود در آن دارد که می‌تواند به پیش‌بینی داده‌های جدید کمک کند. به بیان دیگر، تمام موارد داده‌کاوی شامل استفاده از یادگیری ماشین می‌باشد، اما تمام یادگیری ماشین شامل داده‌کاوی نیست. به عنوان مثال ممکن است شما از یادگیری ماشین برای داده‌های ترافیک اتومبیل جهت ایجاد الگوهای مربوط به نرخ تصادف استفاده کنید. اما از طرفی اگر بخواهیم در مورد اتومبیل‌هایی با رانندگی خودکار صحبت کنیم، باید به این نکته اشاره داشت که آن‌ها کاملاً مبتنی بر یادگیری ماشین بوده و درگیر هیچ نوع داده‌کاوی نمی‌باشند.

مثال‌هایی از کاربرد یادگیری ماشین در سیستم‌های سلامت و امنیت

- استفاده از هوش مصنوعی در تشخیص عفونت ریه با استفاده از تصاویر اشعه ایکس ریه

در جریان همه‌گیری کوید-۱۹، یک ابزار هوش مصنوعی برای تشخیص سریع عفونت ریه ناشی از این بیماری با استفاده از تصاویر اشعه‌های ایکس ریه در مرکز ملی بیماری‌های عفونی^۲ کشور سنگاپور توسعه یافت. با استفاده از این ابزار پزشکان می‌توانند به سرعت افرادی را که نیاز به اکسیژن و مراقبت‌های بیشتر دارند شناسایی و پیش از وخیم‌شدن حال فرد، رسیدگی‌های لازم را به منظور بستری فرد در بیمارستان انجام دهند. براساس آمار ثبت شده، دو سوم از بستری‌های صورت گرفته در کشور سنگاپور با استفاده از این ابزار صورت گرفته است.

پیش از این برای هر فرد مشکوک به بیماری کوید-۱۹، یک رادیولوژیست می‌بایستی بررسی‌های لازم را به منظور تشخیص بر روی عکس ریه فرد انجام می‌داد که این فرآیند در بازه زمانی یک ساعته صورت می‌گرفت. در حال حاضر این ابزار هوش مصنوعی در

¹ Machine Learning

² NCID

مدت سه ثانیه با دقت ۹۶.۱٪ این فرآیند را انجام می‌دهد. این برنامه که به عنوان نمونه‌ای از یادگیری ماشین می‌باشد، با استفاده از تکنیک یادگیری عمیق^۳ برنامه‌نویسی شده و در این امر از ۵۰۰۰ تصویر شامل ۱۰۰۰ تصویر از ریه دارای عفونت و ۴۰۰۰ تصویر از ریه سالم استفاده شده است.

- استفاده از هوش مصنوعی در تشخیص دیابت چشم

به منظور بررسی علائم اولیه دیابت چشم، مرکز دیابت سنگاپور یک سیستم هوش مصنوعی توسعه داده است که می‌تواند نتایج یک اسکن چشم را در عرض چند دقیقه به جای یک ساعت، آنالیز و بررسی کند. در حال حاضر تمامی غربالگری‌های چشم دیابتی در کشور سنگاپور توسط این برنامه (سلنا^۴) که توانایی تشخیص سه نوع بیماری چشمی را دارد، انجام می‌گیرد.

- ابزارهای هوش مصنوعی در پیشگیری از بیماری

ژن‌های هر فرد می‌توانند میزان خطر بیماری‌هایی که جان انسان را در معرض خطری جدی قرار می‌دهد مانند بیماری‌های قلبی را مشخص نماید. از آنجائیکه اگر فردی آگاه به این موضوع باشد که در معرض خطر جدی می‌باشد بسیاری از رفتارهای خود را به منظور پیشگیری انجام خواهد داد، این موضوع کلید موفقیت در پیشگیری از این دسته از بیماری‌ها است.

- استفاده از یادگیری ماشین به منظور تشخیص الگوهای جرم

تلاش برای تشخیص الگوهای مشخص جرم و رفتار مجرمانه بسیار چالش برانگیز است. تحلیلگران جرم می‌توانند ساعات بی‌شماری را صرف بررسی داده‌ها برای تعیین اینکه آیا یک جرم متناسب با یک الگوی شناخته‌شده و کشف الگوهای جدید است صرف کنند. هنگامی که یک الگو شناسایی شد، اطلاعات می‌تواند برای پیش‌بینی و جلوگیری از جرم مورد استفاده قرار گیرد.

این حرکت به سمت رویکردهای تجربی و مبتنی بر اطلاعات در حال رشد است. با این حال، کار تحلیلگران جرم برای شناسایی الگوهای خاصی از جرم مرتکب شده توسط یک فرد یا گروه تا حد زیادی یک کار دستی باقی مانده است. تحلیلگران برای یافتن الگوهای خاصی باید گزارش‌ها جنایت را به طور دستی بررسی کرده و آن‌ها را با جرایم مربوط به گذشته مقایسه کنند. در حال حاضر یک روش یادگیری ماشین به نام "Finder Series" طراحی شده است که می‌تواند به پلیس برای کشف سری جنایت کمک کند. Finder Series برای تشخیص الگوهای سرقت از منزل آموزش داده شده است و این کار را با استفاده از داده‌های تاریخی انجام می‌دهد. البته این امکان تنها در صورتی وجود دارد که اطلاعات مربوط به دقت در طول سالیان گذشته جمع آوری شده باشد.

اتوماتیک‌سازی فعالیت‌ها با استفاده یادگیری ماشین

به عنوان یک نمونه از خودکارسازی فعالیت‌ها با استفاده از الگوریتم‌های یادگیری ماشین می‌توان به اسکن چشم و صورت در نقاط بازرسی اداره مهاجرت اشاره کرد. در سنگاپور، اسکن چشم و صورت جایگزین اسکن اثر انگشت دیجیتال به عنوان روش اصلی شناسایی مسافران در همه ایست‌های بازرسی مهاجرتی این کشور شده است. شهروندان و دارندگان گذرنامه‌های بلند مدت سنگاپور و سایر مسافرانی که اسکن چشم آن‌ها در اداره مهاجرت این کشور ثبت شده است، می‌توانند از این سیستم جدید در تمامی نقاط

³ Deep Learning

⁴ Selena

بازرسی زمینی، دریایی و هوایی استفاده کنند. کم‌تر از یک ثانیه کافی است تا این سیستم اطلاعات بیومتریک مسافر را با اطلاعات ثبت شده در پایگاه داده اداره مهاجرت تطبیق دهد. این سیستم علاوه بر سرعت بالاتر نسبت به موارد دستی از دقت بالاتری نیز برخوردار است.

انواع الگوریتم‌های یادگیری ماشین

پاسخ این سوال که از کدام الگوریتم یادگیری ماشین استفاده نماییم بستگی به نوع داده و نوع نتیجه مورد نیاز دارد.

- انواع داده‌ها

داده‌ها را می‌توان از نوع ساختار به دو دسته ساختار یافته^۵ و غیرساختاریافته^۶ دسته‌بندی کرد. به طور ساده داده‌های ساختاریافته داده‌هایی هستند که به صورت سطر و ستون قابل ذخیره‌سازی می‌باشند. به طور مثال می‌توان به میزان صادرات کالاهای مختلف در طول سالیان گذشته اشاره کرد. در طرف مقابل، داده‌های غیرساختاریافته قرار دارند که به عنوان مثالی از این نوع داده‌ها می‌توان به تصاویر و فیلم‌های ثبت شده اشاره کرد.

از سوی دیگر می‌توان داده‌ها را به صورت نوع عددی^۷ و اسمی^۸ دسته‌بندی کرد. در نوع عددی داده‌ها قابلیت جمع، تفریق و ... را دارند. به طور مثال، برای داده‌های عددی می‌توان به میزان ریالی صادرات و برای داده‌های اسمی به کشور صادرکننده اشاره کرد.

- انواع مختلف الگوریتم‌های یادگیری ماشین

دو دسته اصلی یادگیری نظارت شده^۹ و یادگیری نظارت نشده^{۱۰} از الگوریتم‌های یادگیری ماشین وجود دارد.

- یادگیری نظارت‌شده

اغلب روش‌های یادگیری ماشین از یادگیری نظارت شده استفاده می‌کنند. در یادگیری ماشین نظارت شده، سیستم تلاش می‌کند تا از نمونه‌های گذشته که در اختیار آن قرار گرفته بیاموزد. به عبارت دیگر، در این نوع یادگیری، سیستم تلاش می‌کند تا الگوها را بر اساس مثال‌های داده شده به آن فرا بگیرد. در استفاده از این دسته از الگوریتم‌ها، هدف بدست آوردن یک داده خروجی مشخص می‌باشد. به عنوان نمونه برای پیش‌بینی میزان رشد تورم در سال آتی با استفاده از داده‌های گذشته از این دسته از الگوریتم‌های یادگیری ماشین استفاده می‌شود.

از الگوریتم‌های این دسته می‌توان به موارد زیر اشاره کرد:

^۵ Structured

^۶ unstructured

^۷ numerical

^۸ categorical

^۹ Supervised machine learning

^{۱۰} Unsupervised machine learning

- دسته‌بندی^{۱۱}: درخت تصمیم‌گیری^{۱۲} برای تشخیص از تکنیک‌های این دسته از الگوریتم‌ها می‌باشد. در این نوع مسائل متغیر خروجی یک دسته یا گروه می‌باشد. برای مثالی از این امر می‌توان به تعلق یک نمونه به دسته‌های «سیاه» یا «سفید» و یک ایمیل به دسته‌های «هرزنامه» یا «غیر هرزنامه» اشاره کرد.
- رگرسیون^{۱۳}: یک مساله هنگامی رگرسیون است که متغیر خروجی یک مقدار حقیقی باشد. در این دسته‌بندی با استفاده از تکنیک‌های آماری سعی در پیش‌بینی مقدار عددی یک متغیر خروجی بر اساس داده‌های پیش‌بینی‌کننده^{۱۴} می‌باشد. به عنوان مثال، برای پیش‌بینی میزان قد یک فرد با استفاده از وزن فرد می‌توان از رگرسیون خطی استفاده کرد. براساس تعداد متغیرهای پیش‌بینی کننده دو نوع رگرسیون خطی ساده^{۱۵} و چندگانه^{۱۶} وجود دارد.
- شبکه‌های عصبی^{۱۷}: سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. ایده اصلی این گونه شبکه‌ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده‌ها و اطلاعات به منظور یادگیری و ایجاد دانش می‌باشد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است.

- یادگیری نظارت‌نشده

در یادگیری نظارت‌نشده، الگوریتم باید خود به تنهایی به دنبال ساختارهای جالب موجود در داده‌ها باشد. به بیان ریاضی، یادگیری نظارت‌نشده مربوط به زمانی است که در مجموعه داده فقط متغیرهای ورودی وجود داشته باشند و هیچ متغیر داده خروجی موجود نباشد. به این نوع یادگیری، نظارت نشده گفته می‌شود زیرا برخلاف یادگیری نظارت شده، هیچ پاسخ صحیح داده شده‌ای وجود ندارد و ماشین خود باید به دنبال پاسخ باشد.

از الگوریتم‌های این دسته می‌توان به موارد زیر اشاره کرد:

- تحلیل خوشه‌بندی^{۱۸}: تحلیل خوشه‌بندی یا به‌طور خلاصه خوشه‌بندی، فرآیندی است که به کمک آن می‌توان مجموعه‌ای از اشیاء را به گروه‌های مجزا افراز کرد. هر افراز یک خوشه نامیده می‌شود. اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه هستند و در عوض میزان شباهت بین خوشه‌ها کمترین مقدار است. از کاربردهای این دسته از الگوریتم‌ها می‌توان به تشخیص ناهنجاری^{۱۹} اشاره کرد.
- الگوریتم‌های اپریوری^{۲۰}: یکی از روش‌های پرکاربرد برای کاوش مجموعه اقلام مکرر و قواعد وابستگی^{۲۱} است. به طور مثال «مردهای جوان آمریکایی که بعد از ظهرهای جمعه پوشک بچه تهیه می‌کنند، مستعد خرید آبنجو نیز

¹¹ classification

¹² Decision Tree

¹³ Regression

¹⁴ Predictors

¹⁵ SLRM Simple Linear Regression Models

¹⁶ SLRM Multiple Linear Regression Models

¹⁷ Neural Networks

¹⁸ Cluster Analysis

¹⁹ Anomaly Detection

²⁰ Association Apriori

²¹ association rule mining

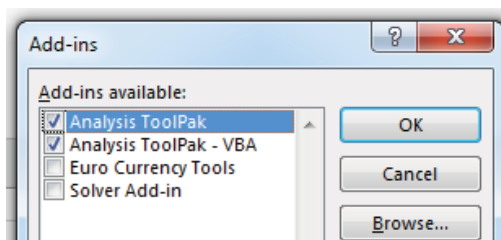
هستند». این عبارت یک مثال مشهور از کاوش قواعد وابستگی از زندگی روزمره افراد است. چنین اطلاعاتی را می‌توان به عنوان پایه‌هایی برای تصمیم‌سازی درباره فعالیت‌های بازار مانند قیمت‌گذاری تبلیغاتی یا تحلیل سبد خرید استفاده کرد.

معرفی برخی برنامه‌های مفید در به کارگیری الگوریتم‌های یادگیری ماشین

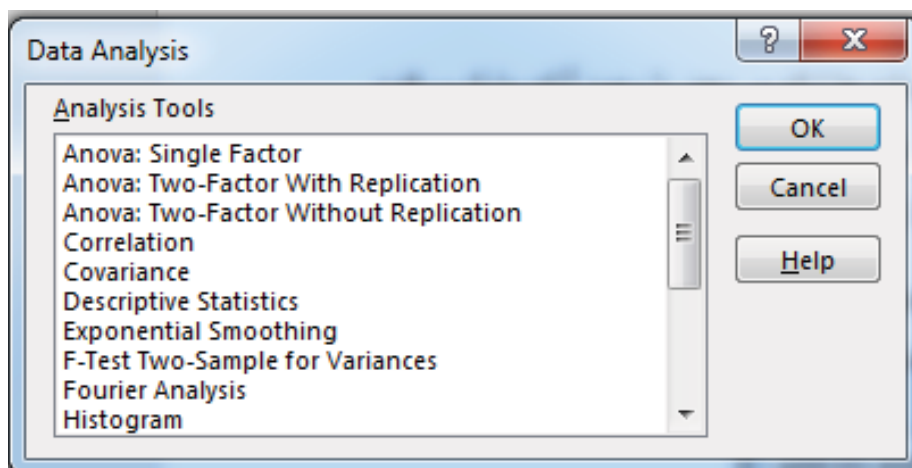
- مایکروسافت اکسل

بسته آنالیز دیتای مایکروسافت، انواع ابزارهای آماری-تحلیلی را برای تحلیل داده‌ها ارائه می‌دهد. با وجود آنکه مایکروسافت اکسل برای پردازش داده‌های با بیشتر از یک میلیون رکورد از عملکرد قابل قبولی برخوردار نمی‌باشد، برای حجم داده‌های کمتر می‌توان آن را مورد استفاده قرار داد. برای استفاده از این بسته ابتدا باید از مسیر زیر این بسته را فعال نمود.

File -> Options -> Add-ins -> Manage: Excel Add-ins Go -> Check Analysis Tool-Pack And VBA



پس از انتخاب کردن داده‌ها، از تب دیتا گزینه دیتا آنالیز را انتخاب می‌کنیم و تحلیل مورد نیاز از جمله رگرسیون، آزمون‌های آماری، بدست آوردن آماره‌های آماری را مشخص می‌نماییم (مطابق شکل زیر).



به طور مثال برای داده‌های قد و وزن افراد و به منظور پیش‌بینی وزن از روی قد، رگرسیون را انتخاب می‌نماییم. متغیر Y را معادل ستون وزن (به عنوان متغیر وابسته) و برای متغیر X ستون نظیر قد را (متغیر مستقل) انتخاب می‌نماییم.

و در انتها خروجی به صورت شکل زیر می باشد:

| | | | | | | | | | |
|-----------------------|--------------|----------------|-------------|-------------|----------------|-------------|--------------|-------------|--|
| SUMMARY OUTPUT | | | | | | | | | |
| | | | | | | | | | |
| Regression Statistics | | | | | | | | | |
| Multiple R | 0.71187772 | | | | | | | | |
| R Square | 0.506769888 | | | | | | | | |
| Adjusted R Square | 0.502659637 | | | | | | | | |
| Standard Error | 9.616764637 | | | | | | | | |
| Observations | 122 | | | | | | | | |
| | | | | | | | | | |
| ANOVA | | | | | | | | | |
| | df | SS | MS | F | Significance F | | | | |
| Regression | 1 | 11402.5094 | 11402.5094 | 123.2941482 | 3.87806E-20 | | | | |
| Residual | 120 | 11097.85945 | 92.48216209 | | | | | | |
| Total | 121 | 22500.36885 | | | | | | | |
| | | | | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% | |
| Intercept | 9.121829433 | 7.198639337 | 1.267160224 | 0.207551672 | -5.130975466 | 23.37463433 | -5.130975466 | 23.37463433 | |
| Height | 0.453296034 | 0.040823542 | 11.10378981 | 3.87806E-20 | 0.372468261 | 0.534123807 | 0.372468261 | 0.534123807 | |

در تحلیل جدول خروجی رگرسیون آگاهی از موارد زیر لازم می باشد:

- مربع - R^2 همواره عددی بین صفر و یک می باشد. هر چه این عدد به ۱ نزدیک تر باشد میزان قابل اعتماد بودن مدل، بالاتر خواهد بود. در این مثال، مربع - R برابر ۰.۵۰۶ می باشد که به معنی آن است که مدل بیشتر از ۵۰٪ قابل اعتماد است.

- در آزمون مربوط به ضرایب مدل، در صورتی که مقدار P-Value مربوط به ضریب کمتر از ۰.۰۵ باشد، به معنی معنی داری ضریب می باشد. در مثال فوق مقدار ضریب قد معنی دار می باشد در حالی که مقدار عرض از مبدا مدل، فاقد معنی داری آماری است.
- مقدار ضریب قد در مدل که عدد 0.4329 می باشد بدین معنی است که به ازای هر ۱ سانتی متر افزایش در قد، وزن به میزان 0.4329 کیلوگرم افزایش می یابد.

- زبان برنامه نویسی R

این زبان برنامه نویسی و نرم افزار رایگان مناسب برای داده های دارای ساختار می باشد. R دارای کتابخانه های گسترده ای از انواع روش های آماری و شیوه های ترسیم نمودارها است. در این زبان برنامه نویسی، «الگوریتم های یادگیری ماشین»، «رگرسیون خطی»، «سری های زمانی» و «استنباط آماری» قابل استفاده هستند. یکی از این کتابخانه ها، Rattle می باشد. برای آموزش نحوه استفاده از این کتابخانه، کتاب داده کاوی با Rattle و R^{۲۳} معرفی می شود.

- از دیگر نرم افزارهای مرتبط، زبان برنامه نویسی پایتون و SAS Enterprise Miner معرفی گردید.

²³ Williams, G.J. (2011), Data Mining with Rattle and R, the art of excavating data for knowledge discovery, Springer

جمع‌بندی و نتیجه‌گیری

داده‌های بزرگ که در زبان انگلیسی به آن Big Data گفته می‌شود، مجموعه‌ای از داده‌ها است که یک سازمان برای بهبود فرآیندهای مختلف خود همچون بازاریابی، فروش، طراحی استراتژی و مواردی مشابه به آن نیاز دارد. در واقع این داده‌ها هستند که بینش لازم را در اختیار افراد قرار می‌دهند تا بتوانند تصمیم‌گیری کنند. بدون داده، تصمیم‌گیری‌ها اغلب تصادفی و غیرقابل اطمینان هستند. در این میان نقش کلان داده در بازاریابی هر روز بیشتر از گذشته می‌شود. کلان داده‌ها می‌توانند بسیاری از سوالات اساسی بازاریاب‌ها را پاسخ دهند. یافتن پاسخ سؤالاتی که می‌تواند برای مدت‌ها برای آن‌ها بی‌جواب مانده باشد با کمک گرفتن و تحلیل صحیح کلان داده‌ها امکان‌پذیر می‌شود. این پژوهش از نوع کاربردی و روش آن مطالعه‌سندی است. نتایج بررسی‌ها نشان می‌دهد که داده‌های بزرگ مزایای مشخص و معقولی دارند و به‌طور قطع تأثیر بسیاری در افزایش فروش کسب و کارها خواهند داشت. اما این روزها بازدهی لازم را ندارند، چرا که سازمان‌ها نمی‌توانند از آن‌ها به شیوه‌ی صحیح بهره‌مند شوند. جمع‌آوری داده‌ها بسیار مهم است، اما آنچه که مهم‌تر است، جمع‌آوری داده‌های صحیح و تحلیل آن‌ها می‌باشد. شرکت‌ها و مدیران تا زمانی که نتوانند تغییر را بپذیرند نمی‌توانند از مزایای بیگ‌دیتا بهره‌مند شوند.

کلان داده دید وسیعی می‌دهد که توسط آن می‌توانید به کشف فرصت‌های کسب و کار بپردازید. به خاطر پدید آمدن منابع ناهمگون تولید دیتا، تنوع زیادی در کلان داده‌ها به وجود آمده است. در گذشته، صفحات گسترده (مثل میکروسافت اکسل و گوگل‌شیت) و دیتابیس‌ها، تقریباً تنها منبع دیتا به حساب می‌آمدند. امروزه، داده‌های بسیار متنوعی به فرم ایمیل، عکس، ویدیو، PDF، صوت و ... نیز در کاربردهای متنوع مورد توجه قرار گرفته‌اند. تنوع زیاد داده‌های بی‌ساختار پیچیدگی‌های فراوانی برای ذخیره‌سازی و آنالیز این داده‌ها ایجاد می‌کنند. توانایی پردازش کلان داده فواید فراوانی دارد؛ برخی از این فایده‌ها عبارتند از:

- کسب و کارها می‌توانند از منابع هوشی قدرتمند در تصمیم‌گیری‌هایشان بهره‌مند شوند.
- پشتیبانی از مشتریان با خودکارسازی فرایندها بسیار راحت‌تر می‌شود.
- می‌توان با استفاده از کلان داده یا همان بیگ دیتا، ریسک‌های احتمالی استارت یک کسب و کار را پیش‌بینی کرد.
- بهینه‌سازی بیش‌تر عملیات‌های مختلف، به سهولت صورت می‌پذیرد.

در نهایت این دوره با تلاش به منظور تفهیم اطلاعات اولیه به شکلی ساده و کاربردی در قالب یک کارگاه عملی چندساعته سعی بر این داشت تا نکات مهم و کلیدی را به مخاطبان ارائه دهد. قطعاً میزان رضایت شرکت‌کنندگان در نظرسنجی صورت گرفته، می‌تواند گویای این مطلب باشد.

با سپاس از شما خواننده گرامی، متولیان محترم سازمان ملی بهره‌وری و تمامی عزیزانی که در راستای گسترش علم همکاری می‌نمایند.

پیوست‌ها

• کاربردها و دستاوردهای دوره برای سازمان مربوطه

- استفاده از داده‌های کلان موجود در سامانه‌های سازمان جهت برنامه‌ریزی و سیاست‌گذاری
- استفاده از داده‌های کلان موجود در سامانه‌های سازمان جهت نظارت بر حسن اجرای قوانین و مقررات مربوط
- جمع‌آوری و الزام دستگاه‌های اجرایی جهت تولید داده‌های عملکردی و استفاده از آن در ارزیابی دستگاه‌های اجرایی

• پیشنهادات شرکت‌کننده در جهت بهبود وضعیت کنونی محل خدمت با توجه به دستاوردهای دوره

- پیاده‌سازی نمونه‌ای از سیستم‌های هوشمند تجاری جهت ارائه بر خط داشبوردهای مدیریتی جهت اتخاذ تصمیمات داده-محور در سازمان
- استفاده از داده‌های کلان سازمان جهت پیش‌بینی تأثیرات مصوبات و بخشنامه‌های پیشنهادی
- استفاده از داده‌های کلان موجود در سامانه‌های سازمان جهت نظارت بر رعایت مقررات ابلاغی، از جمله مقررات در حوزه نظام پرداخت
- برگزاری دوره‌های کاربردی تخصصی با استفاده از نرم‌افزارهای تحلیل بیگ‌دیتا
- آموزش ارزیابی شاخص‌های کلید بهره‌وری بر اساس بیگ‌دیتا و داده‌های حجیم در اختیار

• نقاط قوت

- طرح مبحث بیگ دیتاها در حوزه تخصصی بهره‌وری
- آشنایی اولیه و ضمنی با نرم‌افزارهای گزارش‌ساز و داشبوردهای مدیریتی برای نمایش برخی از شاخص‌های بهره‌وری
- برنامه زمان‌بندی بسیار دقیق و مناسب و استفاده از تجربیات سایر کشورها برای مباحث تخصصی طرح شده در خصوص بهره‌وری بر اساس بیگ دیتاها

• ضعف‌ها و محدودیت‌های دوره آموزشی

- عدم ارائه پاورپوینت مناسب در طول روزهای دوم و سوم
- با توجه به کاربردی بودن دوره و نیاز به استفاده از نرم‌افزار؛ بنظر می‌رسد این دوره باید بصورت **workshop** برگزار می‌شد.
- بخش بیان تاریخچه و مقدمات نسبت به زمان کل دوره و فشردگی مطالب، بصورت طولانی بیان شد.
- محدودیت زمانی برای نصب نرم‌افزار داشبورد مدیریتی آموزش داده شده (Tableau)
- کوتاه‌بودن مدت زمان آموزش کاربردی و عملیاتی با نرم‌افزار (Tableau) برای ارائه گزارشات توصیفی داده‌ها و داشبورد اولیه مدیریتی