Unlocking the Power of Big Data: “The Key Workforce Skills Operator’s Need for the Future of Analytics”

A Technical Paper prepared for SCTE by

Steven R. Harris
VP Global Market Development | Office of the CEO
Society of Cable Telecommunications Engineers, Inc.
140 Philips Rd, Exton, PA 19341
1-610-594-7324
sharris@scte.org

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

1. Introduction

Today's global broadband and telecommunications companies are deep into their digital transformation strategy, compelling operators to be more analytics and data-driven shortly to remain highly competitive in our industry. Liberty Global and other organizations like Comcast, Cox, Rogers, Orange, Plume, and Vodafone all are spearheading digital transformation in broadband and telecommunications.

Figure 1 - Digital Transformation by Liberty Global

Operators who have a handle on vast amounts of data allow them to have a deeper understanding of the business, as well as a competitive edge. For example, with the advancement of data collection and analytical capabilities, all the analytics teams across Cox's technology and operations were centralized [14]. The vast amounts of data paradigm will be referred to herein as "big data". Operators that leverage workforce-related big data, along with data analysis will enhance their operational goals, as well as predict behaviors of their workforce and forecast productivity improvements to drive customer quality of experience (QoE). Moreover, a proficient workforce in big data, analytics, and associated technologies will play a pivotal role in enabling their organizations to succeed in their endeavors toward digital transformation strategies. Contemporary broadband and telecommunications data processing systems need to exhibit greater agility compared to their predecessors as data is ubiquitous, and data ingest is growing exponentially each year. Every day, Facebook produces an immense volume of data, totaling 4 petabytes (PB), equivalent to one million gigabytes. This vast dataset finds its home in a storage system referred to as the "Hive", boasting a staggering storage capacity of approximately 300 PB [9].

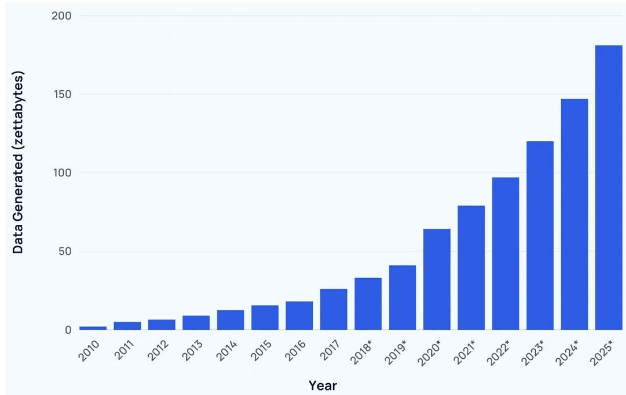

Figure 2 - Global Data Generated Annually, in 2025, hitting 181 zettabytes (Source: Statista)

Furthermore, identifying the significance of ingesting, navigating, storing, transforming, and organizing big data pipelines will lead to workforce efficiencies in the business. The significance of big data pipelines is that they help organizations make sense of the vast quantities of big data they create and stream at unprecedented rates. A workforce that recognizes the key fundamentals of data science and data analysis allows those who work within our industry to be more efficient and streamlined in the areas of big data transformation. For example, correlating data from a data lake, a central repository of structured, semi-structured, and unstructured data, offers the association of big data to provide a deeper insight into workforce and customer key performance indicators (KPI). Using enhanced KPIs allows operators to track trends when a new product is launched, a network outage occurs, or when the school returns to session. In addition, utilizing statistics, artificial intelligence (AI), machine learning (ML), deep learning, and automation will drive big data-based decision-making within the organization.

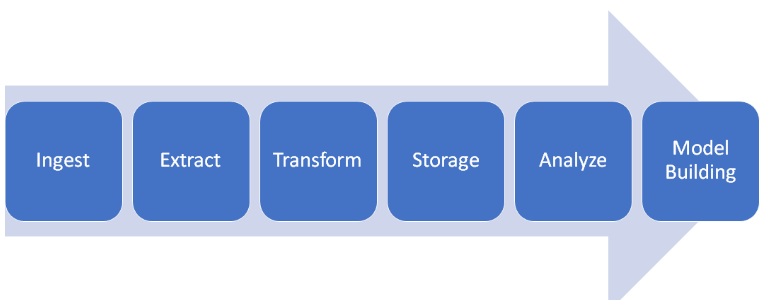

Figure 3 - The Data Science Process

Using structured query language (SQL) and complex pivot tables for today's data visualization is not always a feasible option. In light of ongoing digital transformation efforts within our networks, operators must extend their focus beyond SQL and delve into unstructured big data systems to fully harness the advantages they offer to our industry. This shift will be essential for capitalizing on the potential benefits of these advanced big data management approaches. Summarizing and visualizing broadband and telecommunications big data aids in its accessibility and comprehension for decision-makers, the workforce as a whole, as well as learning and development (L&D) teams. Workforce proficiency in data science (e.g., Kafka, Hadoop), data sources (e.g., Absorb LMS) visualization tools (e.g., Tableau, MS Power BI), data transformation and programming languages (e.g., Python, Scala) are essential for working with big data and data lakes, as well as familiarity with its architecture.

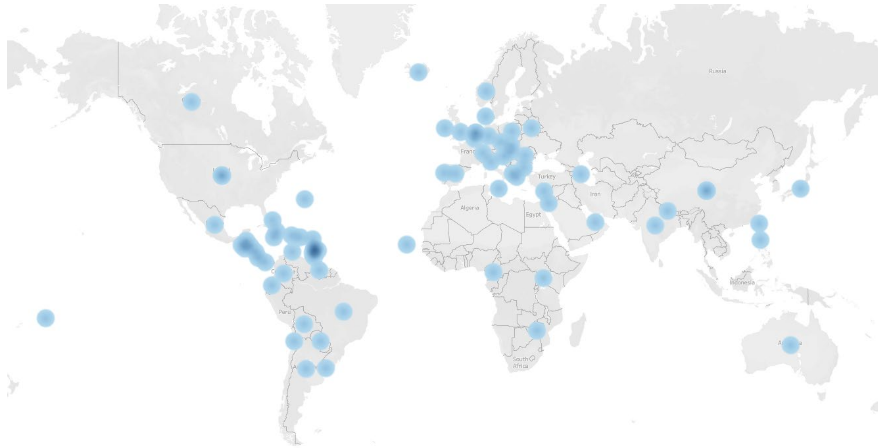

Figure 4 - Tableau, SCTE 2023 Members by Country

1.1. Importance of Big Data in the Broadband and Telecommunication's Industry

According to Gartner [10], by 2024, organizations that fail to build sustainable and scalable big data analytics frameworks will likely experience performance setbacks of at least two years. As shared, big data refers to extremely large and complex sets of data that cannot be effectively processed, managed, analyzed, and visualized using traditional data processing tools and methods such as a relational database (e.g., SQL). The term big data encompasses three primary characteristics, often referred to as the three Vs known as volume, velocity, and variety [1]. Big data involves vast amounts of data generated from various sources, such as social media sites, our broadband CPE, network sensors, mobile device data (images, texts, videos), website-based data, employee key KPIs, eLearning interactions, psychometric data analysis, and more. The sheer volume of data in the broadband and telecommunication industry is typically in the terabytes (TB), PB, or even exabytes (EB). Today with operator digital transformations of their networks, big data is generated at an incredibly high speed and velocity. Often data like network performance or a technician install is produced in real-time as well as near-real-time, and its accumulation can be rapid, requiring immediate pipeline processing and data analysis. Big data comes in various formats and types, including structured data (e.g., SQL-based), semi-structured data (e.g., XML, JSON), and unstructured data (e.g., text, PDF, multimedia, spatial). Data may reside in different vendor-based systems (e.g., DOCSIS, PON) and require aggregation into a data lake for processing. Furthermore, the diversity of data formats adds complexity to the data analysis process. Beyond the three Vs, big data is also characterized by a fourth and fifth "V" known as veracity and value, referring to the trustworthiness, reliability, and worth of the data collected [14]. Big data may include field inconsistencies, errors, and noise due to its diverse sources and unstructured nature, making data quality an important consideration. According to Tractica, companies from the telecommunications industry are expected to invest $36.7 billion per year in AI-related solutions like software, hardware, and services by 2025 [4].

Big data analytics has become increasingly important across various industries besides broadband and telecommunications, including finance, healthcare, retail, and marketing. In the L&D space, data analytics has become crucial and will reign supreme in the future. As budgets tighten, big data analysis becomes even more important to keep a pulse on field and network activities, as well as aligning business stakeholders and showing return on investment, or ROI, for training events. L&D folks refer to the Kirkpatrick New World Model or the Phillips ROI Model that builds on Kirkpatrick to show ROI. Big data and analytics allow operators to be proactive instead of reactive. According to

Knowledge Base Value Research, the global telecommunications analytics market is expected to grow to $8.7 billion by 2025 [4].

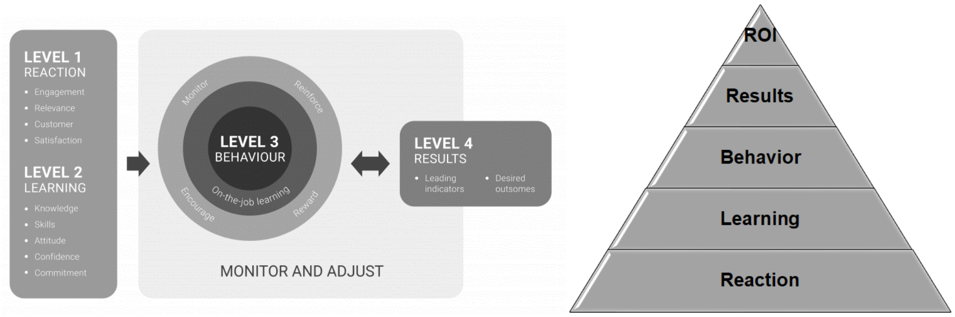

Figure 5 - Kirkpatrick New World Model and Phillips ROI Model

"Data observability has become a necessity, not a luxury, for us. As the business has become more data-driven, nothing is worse than allowing leadership to make a decision based upon data that you don't have trust in. That has tremendous costs and repercussions." - Alex Tverdohleb, VP of Data Services, Fox Networks

By harnessing the potential of big data, operators can gain valuable insights, make data-driven decisions, improve calculated risk, identify patterns, understand trends, develop correlations, and ultimately improve efficiency, productivity, and competitiveness, and drive down cloud costs. Reducing cloud expenses holds significance, given that 81% of IT leaders indicate that their c-suite has directed either a halt in extra expenditures or a decrease in cloud-related costs [3]. Operator L&D organizations and those utilizing SCTE L&D can harness big data throughout their operations to analyze their workforce, including conducting productivity trend analyses.

Figure 6 - Absorb Analyze

Analytics solutions that mine structured and unstructured data are important as they will help operators gain insights not only from their privately acquired data but also from large amounts of data publicly available on the Internet [2]. Such data can be used to improve employee and customer QoE. For example, Vodafone analytics was created to improve business intelligence using big data, using location and mobility insights to help their business customers. Handling broadband and telecommunication big data requires advanced technologies, tools, and techniques that can efficiently process, store, and analyze large volumes of data. Traditional relational databases and data processing methods are often inadequate to cope with our big data challenges. As a result, specialized technologies like distributed computing frameworks (e.g., Hadoop, Azure), AI models, and ML models are used to derive insights and value from our big data sets. Comcast's big data initiative has grown from a small Hadoop cluster to over 10 PB and over 430 nodes in less than two years, allowing them to understand how customers are using their products [5]. The Hadoop framework consists of the Apache Spark (formerly MapReduce) data processing framework and the Hadoop distributed file system (HDFS) serving as the data storage framework.

In the education sector, these technologies allow higher education to analyze learner behavior, assessment data, learner preferences (e.g., modality), learner training time, asset likes, career pathing of learners, and society educational requirements. Data from L&D and higher education will revolutionize the way our workforce learns and instructors present materials at in-person training-led events. The goal is to create a better learner or workforce, meeting them where and how they learn. The facilitator can monitor all learners in real-time and have a deeper conversation on the topics with the learner, providing a stronger understanding of the desired learning behavior [6].

1.2. The Need for Advanced Workforce Skills

As our industry changes at a rapid pace, technology is evolving quickly, and operators must pivot faster than ever. Broadband and telecommunications rely on a knowledgeable and well-skilled workforce with high competency, delivering the latest connectivity options to our customers. Our future workforce must maintain complex systems, implement technological advancements, apply network security, optimize our network, connect a proliferation of devices, make informed decisions with analytics, be well-versed in cloud technologies, collaborate, have strong interdisciplinary skills, and leverage emerging technologies! Advanced workforce skills are essential to meeting the future challenges and opportunities of our industry, furthering our growth and innovation. An employee base that ensures an exceptional customer QoE will result in cost savings for operators through churn reduction and lowered expenses related to acquiring new customers [4]. Leveraging big data additionally empowers our learners to deliver a personalized customer QoE, in the case of Comcast and Vodafone, understanding how customers are using their products and how employees interact with their customers.

It was cited by Pearson Education, that roughly seven (7) in ten (10) employees are currently in job roles where the future is uncertain. It was also mentioned by Pearson Education that occupation redesign coupled with workforce training programs could promote growth in occupational areas that will shrink [7]. Occupation redesign increases productivity, reduces employee churn, increases employee morale, develops new skills (e.g., data science), increases motivation (e.g., less monotone), and encourages innovation. Occupation or job role redesign typically involves a comprehensive analysis of job roles, tasks, skills, behaviors, responsibilities, and interactions within the broader organizational context. Employers are seeking more soft skills such as critical thinking, analysis, complex problem-solving, self-management, and active learning skills states the World Economic Forum (WEF). From a technical perspective, SCTE has collected data on ten trending technical skills: fiber optics, wireless (e.g., 5G, IoT), Wi-Fi, Cybersecurity, cloud computing, DOCSIS, construction practices, network design, network connectivity, and RF theory. All of these areas will require soft skills, as well as technical skills.

1.3. Overview of the Paper's Objective

With broadband and telecommunication data being ubiquitous and its collection growing exponentially, operators and their workforce need to recognize the significance of ingesting, navigating, transforming, storing (e.g., cloud), organizing, and harvesting big data. Understanding the core principles of big data, data science, and data analytics will enable our workforce and networks to operate with enhanced efficiency and seamless coordination.

The paper intends to provide an understanding of how operators can utilize statistics, AI, ML, and automation to drive data-based decision-making as well as call to attention the skills needed by our workforce. Operator data can be used to personalize a customer's experience as well as drive business decisions more effectively. The paper provides definitions and illustrations of practical real-world applications of supervised and unsupervised AI/ML big data, along with how big data applies to broadband and telecommunication networks.

The use of complex data sets or big data requires advanced workforce skills, innovation in operator tools, and updated techniques to process and analyze the information. Managing big data involves the use of data pipelines to ingest, organize, and store large amounts of data from various sources. Using SQL for data is an option today, looking beyond at unstructured big data databases and their benefits to our industry and workforce. Summarizing and visualizing broadband and telecommunication data helps make it more accessible and understandable for decision-makers, as well as our workforce. Proficiency in data science, data transformation with Alteryx, visualization tools like Tableau, and programming languages such as Python and Scala are crucial for the effective handling of big data. Additionally, a working knowledge of architectures like Hadoop or Microsoft Azure is essential for managing unstructured data.

Finally, for immediate action and takeaway, this session will share a view of SCTE's new programs for Python and data analytics, and how attendees can start the occupation redesign process while earning skill-based badges and professional credentials.

2. Understanding Big Data and its Relevance to Operators

Data is created every day in broadband telecommunications, through daily activities servicing our networks or our customer premises. All this data represents big data in the context of broadband and telecommunications. The advancement of data sharing between systems has led to data proliferation across an operator's ecosystem. Here are a few examples of various data sources that are relevant in the context of improving the workforce:

1. Customer care detail records are information about call center calls: such as call duration, time of call, origination of the call, location data, reason for the call, and more.
2. L&D KPIs include courses/journey completion, enrollment, and engagement rate, as well as time spent, assessment performance/patterns, retention rate, dropout rate, level 1 evaluation, skills acquisition, course ratings, and certification achievement.
3. Network device and sensor logs include cable modems, gateways, Wi-Fi extenders, routers, switches, optical nodes, headend equipment, and cell towers providing insights into network performance, traffic patterns, and potential issues caused by craft or failures.
4. Device data encompasses CPE, such as mobile devices, TV displays, and Internet of Things (IoT) devices, along with details about their functionalities and software revisions.
5. Customer data includes all the information about our subscribers, in the case of Comcast, and insights into how customer products are used and connected.
6. Operator social media and online interactions include qualitative data about our customer sentiment and behavior.
7. Quantitative installation data metrics such as discrete service calls on installs, trouble calls, etc.

Operators who effectively gather and correlate big data can make informed decisions about workforce development opportunities. Operators can gather data that has been observed, data that has been volunteered, and data that can be inferred. Adding a layer of visualization, operators can present complex workforce data patterns and trends.

The significance of data collection in the telecommunications industry has been growing rapidly in recent years due to technological advancements, changing consumer behavior, and the increasing demand for better services and experiences. There are examples of Comcast and Vodafone developing their own analytics and big data programs. Big data becomes more significant as our customers are looking for personalized services, personalized customer QoE with our workforce, tailored recommendations, and targeted marketing campaigns. Operators may use big data with data analytics to build an employee profile or persona to align their workforce personnel to particular customer requests. In addition, the network data analysis allows operators to optimize network performance, predict potential issues (e.g., DOCSIS PNM), and ensure seamless connectivity for their customers. Predicted maintenance may be conducted by our workforce from data analysis of various network and customer components, ensuring seamless connectivity. Our workforce can be on top of QoS issues affecting Internet services such as traffic latency, jitter, and delay. A workforce that is rooted in big data provides a significant advantage over its competition. Operators can be more agile in their understanding of market trends, emerging technologies, consumer preferences, and workforce resources.

With the sheer volume and velocity of big data, operators require a place to store the variety of collected data. The cloud computing model plays a crucial role in handling big data due to its scalability, accessibility, cost-effectiveness, and flexibility. Data lake cloud storage addresses many of the operator challenges posed by big data such as scalability, cost-effectiveness, accessibility, flexibility, redundancy, availability, analysis, elastic computing, parallel processing, disaster recovery, diverse data sets, and real-time streaming. In essence, cloud storage empowers operators to handle the challenges of big data. Cloud allows operators to focus on deriving insights from their data without the complexities of managing an on-premises infrastructure like a headend or data center.

3. Fundamentals of Data Science and Engineering for Operators

Several types of data analytics are important to workforce and network efficiency. The first type is known as "descriptive analytics" describing what has happened over a given period. Has the number of truck rolls gone up over the month? Are service calls more this month than last? The second type is known as "diagnostic analytics" which focuses more on why something happened. This involves more diverse data inputs and a bit of hypothesizing. Did the network affect service installs? Did that latest outage impact customer churn? The third type is known as "predictive analytics" and looks at what is likely going to happen in the near term. What happened to service calls the last time we switched Wi-Fi versions at the premises? How many cable modem models predicted network issues? Finally, the "prescriptive analytics" suggests a course of action. If the likelihood of a hot summer week is measured as two days over 100 degrees F, we should add additional technicians to the on-call list as the temperature changes network attenuation [1].

Figure 7 - Types of Data Analytics

As mentioned, statistics, data science, AI, ML, and automation are all key pillars for operators to make data-driven decision making. These pillars collectively enable organizations to analyze big data, extract meaningful insights, and make informed business decisions that drive results in growth and innovation. Let's explore how each of these elements contributes to the process.

Statistics involves the ingestion, analysis, transformation, organization, and presentation of big data. Statistics provides techniques for summarizing and understanding data patterns, relationships, and trends. For data-driven decision-making, statistics help in descriptive statistics, summarizing and describing data through measures like mean, median, and standard deviation. Inferential statistics serve as a valuable tool for deriving conclusions and making predictions about populations, such as the workforce, by analyzing sample data. Hypothesis testing involves evaluating the accuracy of hypotheses and making well-informed data-driven decisions. For example, is a skill like optical power budget measurements in fiber networks missing in the workforce? Is this missing skill increasing service calls on installs?

AI encompasses the development of intelligent systems that simulate human intelligence and decision-making processes. AI technologies contribute to data-driven decision-making by natural language processing (NLP) to extract insights from unstructured text data, enabling sentiment analysis and language understanding. For example, the Vodafone Remedy ticket system employs an AI decision engine powered by NLP. This engine predicts the resolution of raised incidents and initiates existing automation to address and close them, all without requiring human intervention [15]. These AI expert systems automate decision-making processes by encoding domain expertise into software. Another example is DOCSIS PNM, which can automate the decision-making process at the cable modem termination system (CMTS), allowing a CMTS to compensate for poor network conditions by sending data to a cable modem to adapt to conditions. Furthermore, AI expert systems are used to provide real-time workforce information to improve/re-enforce skills or deliver a more personalized customer QoE. A future area of AI algorithms will be cognitive computing or the notion of emulating human thinking and decision-making.

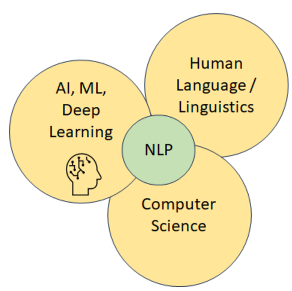

Figure 8 - Natural Language Processing

ML involves the development of algorithms and models that enable computers to learn patterns from data and make predictions or decisions. In the case of DOCSIS PNM, operators can teach the software to recognize network events that may impact the customer QoE. ML techniques enhance operator data-driven decisions by allowing for predictive analytics to forecast future outcomes based on historical data patterns and trends. ML provides classification, categorizing data into predefined classes or categories. Finally, ML can cluster or correlate, identifying groups within data to reveal patterns.

Automation involves the use of technology to perform tasks or processes with minimal human intervention. Operator Internet service self-installs are a good example of where truck rolls and technicians are typically not required. In the data-driven decision-making model, automation contributes by providing data preprocessing or automatically cleaning, transforming, and preparing data for analysis. Automation provides mechanized reporting and visualization to communicate insights. Automation can assist decision-makers by providing relevant data and insights in real time.

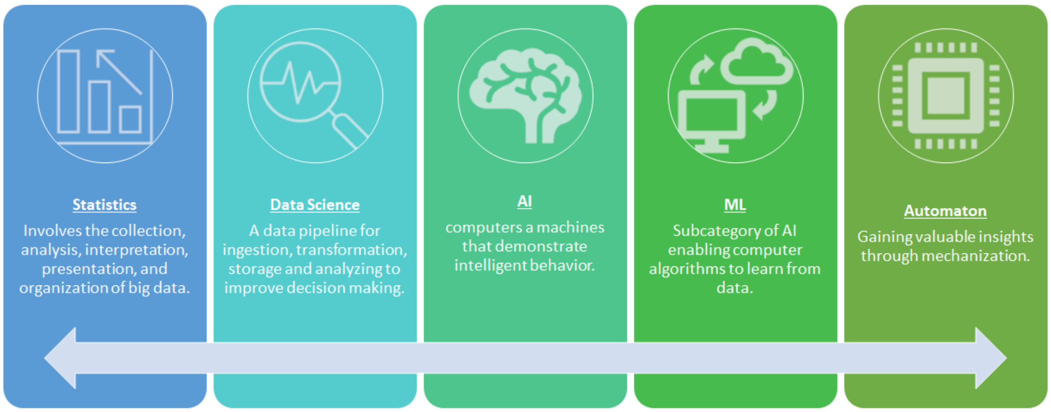

Figure 9 - Differences between Statistics, Data Science, AI, ML, and Automation

There are numerous advantages to utilizing statistics, data science, AI, ML, and automation in the context of data-driven decision-making with big data. Their application will improve an operator's efficiency, and accuracy, reveal insights, expose correlations, be scalable with cloud-based solutions, drive innovation into future solutions, and offer a competitive advantage. In summary, integrating statistics, data science, AI, ML, and automation into data-driven decision-making processes empowers organizations to harness the full potential of their big data, make strategic choices, and achieve their goals in a rapidly evolving business.

3.1. Personalization of the Customer QoE and Business Decisions through Data Utilization

Personalization of customer experiences and business decisions through data utilization is a powerful strategy that leverages big data to create tailored interactions, services, products, and strategies for individual customers. Personalization can also be used with an operator's workforce. For customers, social media is one place that provides valuable customer data that can be harnessed to create a personalized experience. Such an approach enhances customer satisfaction, loyalty, and engagement while also driving more effective and efficient business outcomes [8].

Big data utilization enables personalization in both customer QoE and business decisions in customer profiling by collecting and analyzing customer data. Operators can create detailed personas to understand unique needs and interests that include preferences, behaviors, demographics, and past operator interactions. Furthermore, big data drives segmentation, dividing customers into groups based on shared characteristics. This allows operators to target marketing efforts, product recommendations and offers more effectively. Operators monitor customer behavior, or perform behavior tracking, at various touchpoints (e.g., website, mobile app, set-top box, social media, chatbot) helping in understanding the customer journey, preferences, and pain points. Future operator interactions can then be tailored to meet the customer's specific needs.

Utilizing customer data, operators can deliver personalized marketing messages, recommendations, and promotions that resonate with each individual, leading to higher engagement and conversion rates.

Websites and apps can use data to dynamically display content, such as product recommendations or relevant articles, based on the user's preferences and browsing history. Analyzing big data allows operators to gain insights into market trends, customer preferences, and operational performance. These insights guide strategic decisions, product development, optimized pricing, and resource allocation (e.g., workforce).

By employing predictive models, operators can forecast future trends and customer behaviors. This informs decisions related to inventory management, optimizing supply chain, sales forecasting, enhanced risk assessment, workforce re-skilling, and resource planning. Operators can use big data-driven dashboards and analytics platforms to perform real-time monitoring of KPIs, enabling swift and informed decisions.

The benefits of personalization through data utilization not only enhance customer engagement but allow for improved conversion rates with personalized marketing, increased revenue, cost efficiency with optimized decisions, competitive advantages, and a more innovative organization.

In conclusion, personalization of customer QoE and operator decisions through big data utilization and data analytics is a transformative strategy that fosters stronger customer relationships, better decision-making, and sustainable business growth.

4. Applications of AI/ML in Telecom Networks

4.1. Real-World Examples of Supervised AI/ML Applications

Supervised statistical, data science, AI, ML, and automation applications involve training a model on labeled data (input-output pairs) to make predictions or decisions based on new, unseen data. There are several real-world examples of supervised AI/ML applications:

1. An LMS like SCTE's Absorb supports supervised AI/ML that enables personalized experiences, collaboration through social learning, badging, and content recommendations like Netflix through predictive AI and NLP.
2. Operator email spam filtering uses supervised learning to categorize incoming emails as spam or pass them as a good emails. The spam model is trained on a labeled dataset of emails, with features like words and/or phrases, and their corresponding labels (spam or not spam). The model then predicts the likelihood of new emails being spam based on these features.
3. Comcast Xfinity deploys a supervised cybersecurity system for their Internet gateway devices providing a continuous AI/ML learning system that employs its cumulative knowledge to recognize and predict cyber-attacks.
4. Wi-Fi uses adaptive Wi-Fi services and optimization using an application learning model. In the case of Plume, the model provides application prioritization or fine-tunes application performance. Software automatically detects and optimizes the performance of each application traffic flow, improving latency, and ensuring an optimal customer experience [11].
5. Amazon deploys an image classification and recognition model. The model can identify objects in photos, using supervised learning. The model uses AI/ML to learn from a dataset of images labeled with object categories, enabling Amazon to classify new images into those categories.
6. Speech recognition is offered by voice assistants like Siri, Amazon, and Google Assistants using supervised learning to understand and interpret spoken languages. The NLP model is trained on spoken words or phrases and their corresponding text transcriptions. Another example is the Xfinity voice remote, an NLP speech recognition model.

7. Language translation services employ supervised learning to translate text from one language to another. The model learns from parallel corpora, where sentences in one language are paired with their translations in another language. SCTE's Absorb LMS allows translation to over 30 different languages.
8. Broadband and telecommunications organizations can use supervised learning to predict churn by training ML models in customer or employee behavior.

In conclusion, the application of supervised AI/ML models across broadband and telecommunications is evident in the examples mentioned. From personalized learning experiences in education through SCTE's VirtuLearn™ platform to efficient email filtering systems utilizing predictive algorithms, the power of supervised learning is harnessed to enhance user experiences and streamline processes. In the realm of cybersecurity, organizations like Comcast Xfinity leverage continuous learning systems to fortify their defenses against cyber-attacks. The optimization of Wi-Fi services with adaptive learning models, exemplified by Plume, showcases the adaptability and performance improvements achieved through this approach. Moreover, the wide-ranging applications extend to image recognition, speech interpretation, language translation, and even predictive analysis in telecommunications. The proficiency of supervised learning, drawing insights from labeled datasets, not only facilitates accurate predictions and classifications but also underscores the transformative potential of AI/ML technologies in shaping our modern-day functionalities. As our industry continues to integrate these models, the landscape of possibilities expands, promising further advancements in efficiency, accuracy, and user satisfaction.

4.2. Practical Illustrations of Unsupervised AI/ML in the Telecom Industry

Unsupervised AI/ML techniques are valuable in the broadband telecommunication industry for uncovering patterns, identifying anomalies, and gaining insights from large and complex datasets. There are several real-world examples of unsupervised AI/ML applications:

1. Network anomaly detection by unsupervised learning can help identify abnormal patterns in network traffic, indicating potential security breaches, outside plant issues, irregular CPE operation, or technical issues in the facility. Analyzing our network behavior, a system can detect unusual activities and trigger alerts for further investigation. A good example is DOCSIS proactive network maintenance (PNM) or an optical vendor that uses an optical time domain reflectometer (OTDR) test across a fiber network to detect patterns.
2. Our industry can segment its customer base using unsupervised clustering algorithms. By analyzing customer data such as usage patterns, location, and preferences, the company can identify distinct customer groups, allowing for more targeted marketing and service offerings.
3. Analyzing workforce development in various ways allows an organization to extract valuable insights from employee data, identify patterns, and enhance strategic decision-making. A few examples of unsupervised workforce clustering algorithms include talent segmentation, skill gap analysis, personalized career learning journeys, engagement analysis, ID emerging leaders, team formation, etc.
4. Unsupervised learning can analyze large amounts of network data to optimize resource allocation, improve coverage, and enhance network performance. It can identify areas with high demand and potential congestion, leading to more efficient network planning.
5. Prediction of customer or employee churn by analyzing historical usage data. Unsupervised learning can identify patterns and correlations in behavior, indicating potential churn, and allowing organizations to take proactive measures to retain customers and employees.
6. Optimizing customer support by analyzing customer interactions and feedback, unsupervised learning can identify common issues and trends. This insight can be used to improve self-

install options, and technician interactions during install/repair, refine support processes, and enhance the overall customer experience.
7. Unsupervised NLP techniques can analyze customer call recordings and text interactions to identify sentiment, common concerns, and emerging issues. This information can guide service improvements and strategic decisions.
8. Network traffic analysis can be performed by unsupervised learning helping analyze network traffic patterns to optimize data routing, reduce latency/jitter/packet loss, and ensure efficient Internet data transmission.
9. Unsupervised learning can analyze data on resource usage such as bandwidth and data consumption to optimize resource allocation and pricing plans, ensuring fair distribution and profitability.
10. QoS enhancements by monitoring and analyzing QoS metrics to identify areas of improvement in network performance and service delivery.

In conclusion, the application of unsupervised learning methods across broadband and telecommunications has showcased their transformative potential in uncovering valuable insights, enhancing decision-making, and optimizing various aspects of operations. Whether it's network anomaly detection for technical issue identification, customer segmentation for targeted marketing, or the intricate analysis of workforce development, unsupervised algorithms offer a powerful tool for uncovering hidden patterns and trends within complex datasets. These techniques contribute to efficient resource allocation, personalized services, and improved customer experiences. Furthermore, the ability of unsupervised learning to discern patterns in historical data extends to predicting churn, optimizing customer support, and refining network performance. Leveraging unsupervised natural language processing aids in sentiment analysis and issue identification from customer interactions, while network traffic and quality of service improvements bolster data transmission efficiency. Ultimately, the versatility of unsupervised learning serves as a catalyst for innovation and optimization across a spectrum of industries, underscoring its role in shaping more data-driven, efficient, and customer-centric landscapes.

4.3. Applying Big Data Analytics to Enhance Telecom Networks

Applying big data analytics to enhance our broadband and telecommunication networks holds the potential to usher in a new era of improved network performance, elevated customer experiences, enhanced workforce benefits, and streamlined operations. This innovative approach can be harnessed across various facets of the telecom industry.

Big data analytics offers a range of benefits in network performance optimization. By scrutinizing massive volumes of network traffic data, it identifies congestion points and bottlenecks, enabling targeted optimization efforts. Furthermore, predictive maintenance (e.g., DOCSIS PNM) based on equipment and sensor data foresees potential hardware failures, allowing for proactive intervention. Real-time monitoring of QoS metrics facilitates swift issue detection and resolution, preserving user satisfaction.

Understanding customer and employee behavior through data analysis facilitates personalized services and strategic customer marketing. Historical data assessment and predictive analytics empower companies to anticipate and mitigate customer/employee churn, enhancing retention efforts. Personalized service packages and learning plans aligning with individual preferences contribute to enriched interactions.

All operators are looking to expand and evolve their networks for the next technology, whether it is a DOCSIS, wireless, or fiber roadmap. Anticipating future demand, optimizing network resources, and re-skilling our workforce are central to expansion strategies. Big data analytics offers demand forecasting insights to support capacity planning, labor, and infrastructure growth. Geographic analysis aids in identifying optimal locations for new networks, ensuring efficient coverage in a footprint. Furthermore, optimizing resource allocation, including bandwidth and spectrum, improves efficiency while minimizing waste. Additionally, analyzing energy consumption data supports eco-friendly practices in network operations, infrastructure, and future footprint expansions.

Big data analytics bolsters operator cybersecurity efforts like Rogers' Ignite gateway by detecting anomalous patterns in network traffic indicative of fraudulent activities or security breaches [13]. Timely identification and response mitigate potential risks, safeguarding network integrity. Customers also value such programs and may choose operators because of their cybersecurity value. Couple efforts in cybersecurity with real-time dashboards and visualizations for network health, performance, and usage to empower rapid decision-making by operators. Predictive insights derived from trend and pattern analysis offer early warnings and guide proactive measures.

With the proliferation of IoT and mobile connectivity in our industry, leveraging big data analytics for data processing allows insights into connected device behaviors and facilitates optimization for diverse applications. The management of such devices' intricate connectivity requirements ensures seamless communication and efficient data handling.

Incorporating big data analytics into broadband and telecommunication networks and their workforce not only empowers companies to make informed decisions but also amplifies network efficiency and elevates customer satisfaction. This data-driven approach ushers in proactive responses to challenges, fosters innovation, and cements telecom providers at the forefront of technological advancement.

5. Managing Big Data in Telecom

Data pipelines play a critical role in collecting, organizing, and storing data in the telecom industry. They facilitate the efficient and automated flow of data from various sources to storage systems like a data warehouse, enabling telecommunication organizations to process and analyze data for insights and decision-making. Here's how data pipelines are used in the telecom sector:

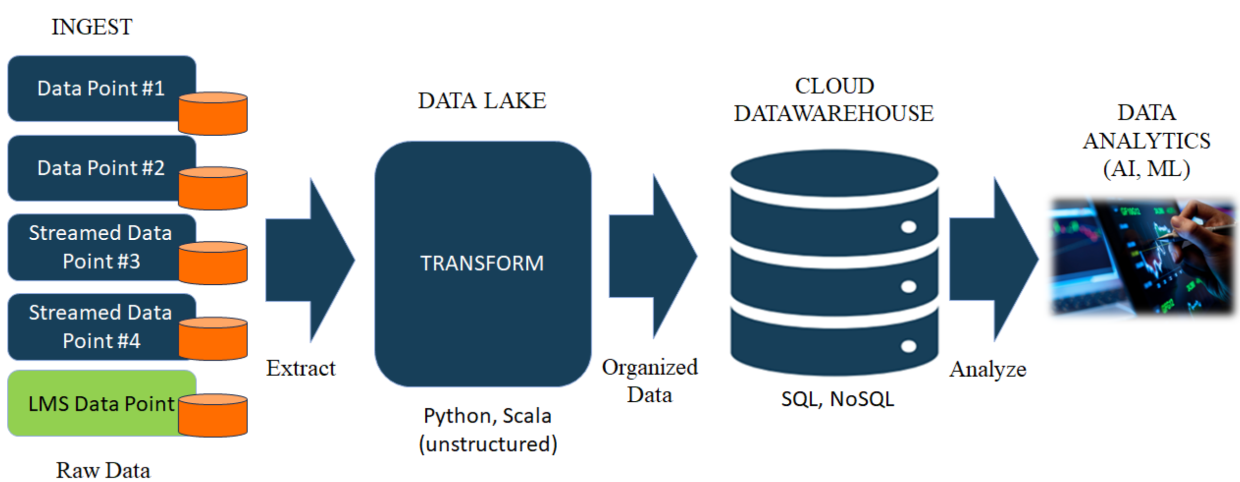

Figure 10 - Data Pipeline

Data pipelines ingest real-time data from network and system equipment, such as routers, gateways, modems, servers, LMSs, call centers, cell towers, etc. This data includes KPIs, traffic patterns, and device status. Data pipelines capture customer interaction data, including call logs, text messages, and app usage. This information helps analyze customer behavior and preferences. Data pipelines can be used to ingest workforce data, including learning asset completion, credentials tracking, learner preferences, predicting learner success/churn, personalization KPIs, optimized learning journeys, etc. [11]. Additional data includes services supporting IoT devices, collecting data generated by sensors and connected devices, and enabling monitoring and control of operator IoT deployments. Ingest billing and financial data to track revenue, expenses, and customer billing information.

After the ingestion phase, data is transformed and organized. The data pipelines clean and transform raw ingested data into usable formats, removing duplicates, correcting errors, and enriching data with additional context. Data from various sources is aggregated to provide a unified view of performance, interactions, and other relevant metrics. Data pipelines ensure that data is consistent and standardized, enabling accurate comparisons and analysis.

The next phase of the data pipeline is storing the transformed and organized data in the data warehouse. The warehouse provides a centralized repository for historical and real-time data. Broadband and telecommunication companies may also use data lakes to store raw and unstructured data, enabling flexible and scalable storage for future analysis. Frequently, storage entails utilizing cloud services, offering a flexible solution for managing vast amounts of data in the order of PBs or EBs. Cloud-based storage solutions offer scalability and cost-effectiveness, allowing telecom companies to store large volumes of data without the need for extensive on-premises infrastructure. Several cloud models exist, allowing flexibility with analysis, public, private, hybrid (public/private), and community.

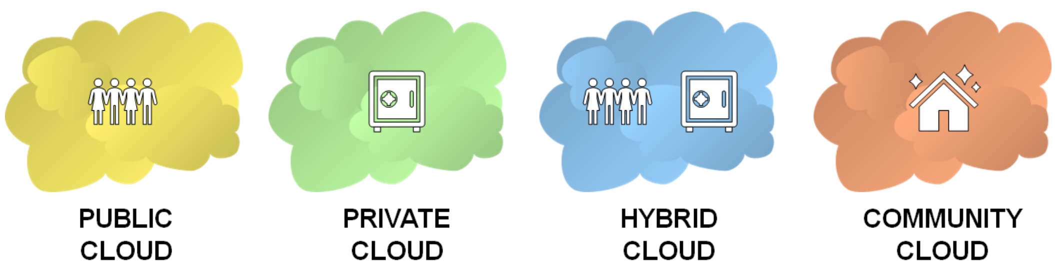

Figure 11 - Cloud Models

The final phase is the processing of the data, as well as data analysis. Processed data is made available for analysis using analytics platforms and tools. This enables operators to gain insights into performance, behavior, and other relevant metrics. Some data pipelines support real-time processing, allowing for immediate analysis and response to events as they occur. For example, operators could be notified when a technician makes a mistake in the network caused by human error factor (HEF). This real-time data event can be analyzed and correlated with network sensors and a technician's learning journey.

Data pipelines are monitored for data quality, performance, and potential issues. Monitoring ensures that data flows smoothly and accurately. Pipelines include error handling mechanisms to address issues such as data source failures or data format inconsistencies. Data pipelines are designed to adhere to regulatory requirements and data governance standards, ensuring data privacy and security. Data is often encrypted during transmission and storage to protect sensitive or private information.

There are a few solutions that operators can leverage in the marketplace. The Hadoop family is an open-source framework used for processing large data sets, as well as updated methods for data ingestion, transformation, storage, and analysis. Hadoop offers a fault-tolerant and parallel processing solution that will be important for our success in the future. Another popular solution is Microsoft Azure is a comprehensive cloud platform that includes a vast range of services, including those for big data processing. this sample conclusion text with your text. Rogers will leverage Azure's enhanced data and AI capabilities across its business to enable use cases and operational improvements through Microsoft's new Data Centre of Excellence in Toronto [13].

In summary, data pipelines are a fundamental component of telecom operations, enabling the seamless collection, organization, and storage of data. They support data-driven decision-making, enhance customer experiences, and drive operational efficiency in the dynamic and data-intensive telecom industry.

5.1. Beyond SQL: Benefits of Unstructured Big Data Databases

Unstructured big data databases offer a range of significant benefits to our industry in the realm of data management and analysis. Unlike traditional structured databases, which rely on rigid schemas (SQL), these databases excel at handling vast amounts of diverse and unstructured data, such as text, images, videos, and social media posts. This adaptability allows broadband and telecommunication organizations to capture and store data in its raw, original form, eliminating the need for predefined structures and enabling flexibility in data ingestion. Unstructured big data databases empower businesses to gain deeper insights by processing and analyzing this heterogeneous data, uncovering hidden patterns and trends that might otherwise be overlooked. Additionally, these databases are well-suited for applications in AL, NLP, and ML, as they provide the foundation for training advanced models on diverse data types. Overall, unstructured big data databases empower enterprises to harness the full potential of their data, fostering innovation, informed decision-making, and enhanced customer experiences. Several examples of big data databases are MongoDB, Apache Cassandra, Amazon simple storage service (S3), Amazon DynamoDB, and Hadoop HDFS. Shaw, now Rogers, chose to use DynamoDB to manage customer data in JSON format [16], while Comcast and Cox are using Hadoop HDFC [5, 14].

5.2. Summarizing and Visualizing Telecom Data for Better Decision-Making

Summarizing and visualizing broadband and telecommunication big data is a powerful approach that enables better decision-making by providing insights, trends, and patterns clearly and intuitively. Summarization and visualization can enhance decision-making in our industry. As shared Cox centralized all their analytics teams to drive best decision making [14].   Summarization includes customer and workforce KPIs, usage patterns, segmentation (demographics, behavior, and preferences), network KPIs, and churn analysis. Visualization includes interactive dashboards, geospatial maps, trend analysis, heatmaps, network topologies (OSP/ISP), customer/employee flowchart journeys, predictive analysis, and social media sentiment.

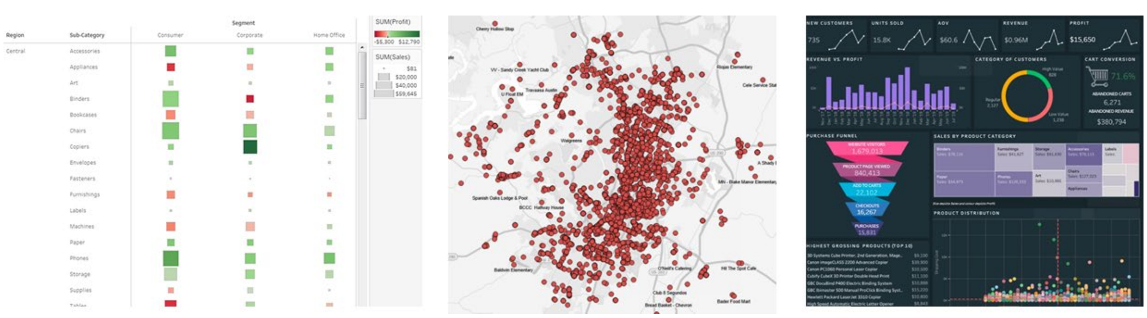

Figure 12 - Visualization Examples: Heatmap, Geospatial and Dashboards

There are a host of benefits of summarization and visualization. One of the big benefits is allowing complex data to be more accessible, enabling decision-makers to quickly grasp the key insights without delving into raw data. Visual representations help identify recurring patterns, correlations, trends, and anomalies that may not be evident in tabular or raw data. Interactive dashboards and visualizations provide real-time information, enabling timely and informed decision-making. For example, summarized data and visualizations facilitate effective communication between technical operation folks and non-technical L&D stakeholders by presenting information in an easily digestible format. Insights gained from summarization and visualization support strategic planning, resource allocation, workforce evolution, and investment decisions for network upgrades and expansion.

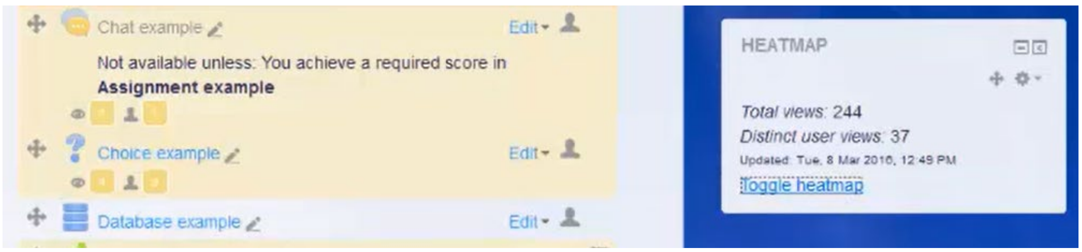

Figure 13 - Moodle LMS Visualization Heatmap, Identification of Popular Learning Assets

Incorporating summarization and visualization techniques into broadband and telecommunication data analysis empowers decision-makers to make data-driven choices, optimize network performance, enhance customer experiences, and drive overall business growth.

6. Key Workforce Skills for Handling Big Data

As we transform our industry into the digital age, our workforce must evolve in the same direction as the business. Proficiency in data science is highly relevant to our industry, as it empowers us to extract valuable insights from the vast amount of data generated by our networks, services, workforce, educational programs, and customer interactions.

Proficiency in data science has many benefits for broadband and telecommunication operators such as network optimization, capacity planning, proactive maintenance, customer/employee insights, customer/employee churn prediction, segmentation, service/product improvement, QoS/QoE, personalization, fraud detection, cybersecurity, anomaly detection, optimizing resources, energy management, resource allocation, footprint expansion, real-time insights, data governance, CPE management, strategic planning, market insights, and data-driven innovation.

The workforce's skill set must align with the entire data pipeline ecosystem. Nurturing a data science mindset is essential, beginning with a comprehensive grasp of data science operations. It is recommended to have a workforce knowledge of data storage frameworks like SQL and NoSQL. Our workforce must be familiar with big data tools like Hadoop with Apache Spark, Azure, and Kafka. Alteryx for the structural data transformations and scripting and programming skills like Python and Scala. As well as cloud-based AWS, Azure, and Google Cloud [14]. MongoDB, Apache Cassandra, Amazon S3, Amazon DynamoDB, and Hadoop HDFS big data databases are growing workforce skills. Soft skills are also required program solving, collaboration, project management, adaptability to learn new tools, and troubleshooting.

In conclusion, proficiency in data science is crucial for telecom operators to effectively manage their networks, improve customer experiences, optimize resources, enhance security, and drive business growth. By leveraging data science techniques, operators can stay competitive, provide high-quality services, and adapt to the evolving demands of the telecommunications industry.

7. SCTE's New Programs for Data Analytics and Python

Through the SCTE education partnership with Cisco Network Academy, we can offer three new courses related to big data and data analytics.

SCTE provides an innovative online course titled "Introduction to Data Science", offering a comprehensive overview of how data analytics adds value to organizations, businesses, governments, and industries. Learners understand how data is being utilized, collected, and visualized, as well as the role of AI. The course helps learners understand the roles and responsibilities of a data analyst and how data scientists and engineers play a role in analytics.

Besides the "Introduction to Data Science", SCTE offers a “Data Analytics Essentials” course teaching learners the fundamental tools of a data analyst. The program was designed for learners to understand how data is collected, transformed, organized, and visualized with spreadsheet tools such as Excel. The program also allows learners to query data from a relational database using SQL and to improve data presentations using powerful business intelligence tools like Tableau. The course allows learners to create an analytics portfolio complete with an analysis of popular datasets. The portfolio will allow learners to showcase their skills in Excel, SQL, and Tableau.

Finally, SCTE offers Python, a general-purpose programming language adopted by the telecommunication industry and used to build just about anything. For this course, no prior programming knowledge is required! Python is key for backend web development, data transformation, data analysis, AI, and scientific computing, all of which are key for expanding a career in broadband and telecommunications. Learners will design, write, debug, and run programs encoded in the Python language using an integrated hands-on Sandbox. The course begins with the very basics guiding learners step by step until they become adept at solving more complex problems. By the end of this course, the learner will have a working knowledge of how computer programs are executed, improve critical thinking, and problem-solving skills; and translate real-world issues into computer-solvable problems.

All three of these courses all learners to earn badges and course certificates, as well as a certification as a Python-certified professional with the Python institute.

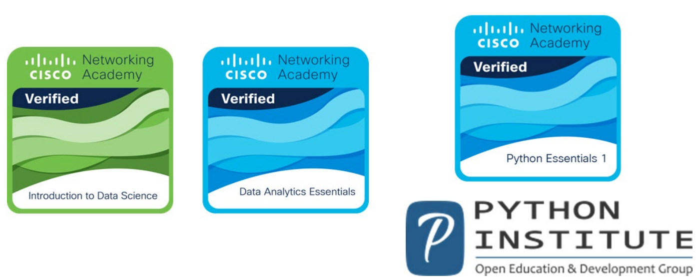

Figure 14 - SCTE Big Data Courses

8. Conclusion

In conclusion, this paper has provided a comprehensive exploration of the significance of big data in the broadband and telecommunications industry, highlighting the essential connection between advanced workforce skills and the successful utilization of this data-driven landscape. L&D teams can start to explore data’s value beyond the LMS, correlating across the business and with educational partners like SCTE. Throughout the paper, we delved into the fundamentals of big data, its relevance to broadband and telecommunications operators, and the pivotal role of data science and engineering in transforming customer experiences, employee experiences, and business strategies. The applications of AI/ML were examined through real-world examples, showcasing both supervised and unsupervised approaches.

Moreover, the paper underscored the importance of managing big data in the broadband and telecommunications sector, emphasizing the benefits of unstructured databases and data visualization techniques in making informed decisions. Recognizing that the potential of big data can only be harnessed by skilled professionals, the discussion of key workforce skills highlighted the need for expertise in handling and interpreting vast datasets.

Additionally, the insight into SCTE's new programs for data analytics and Python underscored the industry's proactive approach to equipping professionals with the necessary skills to navigate the evolving landscape of big data in broadband and telecommunications.

As the broadband and telecommunications industry continues to evolve with its digital transformations, this paper serves as a valuable resource for understanding the interconnectedness of data-driven strategies, workforce readiness, and technological advancements. By embracing the insights and strategies discussed here, stakeholders in the broadband and telecommunications sector can position themselves to thrive in an era where data is the cornerstone of innovation and progress.

Abbreviations

AI	artificial intelligence
CMTS	cable modem termination system
CPE	customer premises equipment
DOCSIS	data over cable service interface specification
EB	exabyte
HDFS	Hadoop distributed file system
HEF	human error factor
IoT	Internet of things
ISP	inside plant
JSON	JavaScript object notation
KPI	key performance indicator
L&D	learning and development
LMS	learning management system
ML	machine learning
MS	Microsoft
NLP	natural language processing
NoSQL	not only structured query language
OSP	outside plant
OTDR	optical time domain reflectometer
PB	petabyte
PDF	portable document format
PNM	proactive network maintenance
PON	passive optical network
QoE	quality of experience
QoS	quality of service
ROI	return on investment
S3	simple storage service
SCTE	Society of Cable Telecommunications Engineers
SQL	structured query language
TB	terabyte
Wi-Fi	wireless fidelity
WWW	world wide web
XML	extensible markup language
ZB	zettabyte

Bibliography & References

1. Johnson, Jeff; Friend, Scott; Lee, Hannah (2017). "Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process: BIG DATA VOLUME, VARIETY, AND VELOCITY." Journal of Product Innovation Management, 34. DOI: 10.1111/jpim.12397.
2. Schomm, F., Stahl, F., Vossen, G. (2013). Marketplaces for data: An initial survey. SIGMOD Record, 42(1), 15-26.
3. Monte Carolo. (2023). Data Engineering Trends & Predictions. Retrieved July 15, 2023, from, https://cdn.pathfactory.com/assets/10854/contents/468540/232de070-4c94-40b5-a1a3-34d885e8313a.pdf
4. KBV Research. (2019). Telecom Analytics Market Size, Share, Trends & Forecast 2019-2025. Retrieved from https://anywhere.epam.com/business/big-data-analytics-for-telecommunications
5. Informatica. (2023). CUSTOMER SUCCESS STORIES COMCAST. Retrieved August 1, 2023, from https://www.informatica.com/about-us/customers/customer-success-stories/comcast.html
6. Happiest Minds Technologies. (2014). Big Data: Why should enterprises adopt it? Retrieved August 15, 2023, from https://www.happiestminds.com/whitepapers/Big-Data-Why-should-enterprises-adopt-it.pdf
7. Hasan Bakhshi, H., Downing, J., Osborne, M., Schneider, P., The Future of Skills Employment 2030. (2017). Retrieved from https://ora.ox.ac.uk/objects/uuid:86577437-1353-4743-8520-401c1f99ad1b/download_file?file_format=application%2Fpdf&safe_filename=Future%2Bof%2Bskills%2BVoR.pdf&type_of_work=Report
8. Rane, Nitin & Achari, Anand & Choudhary, Saurabh. (2023). Enhancing customer loyalty through quality of service: Effective strategies to improve customer satisfaction, experience, relationship, and engagement. International Research Journal of Modernization in Engineering Technology and Science. 5. 427-452. 10.56726/IRJMETS38104.
9. Osman, M. (2023). Wild and Interesting Facebook Statistics and Facts. Retrieved from https://kinsta.com/blog/facebook-statistics/
10. Top Trends in Data and Analytics, 2022
11. Plume leverages AI to introduce 'Full Stack Optimization'--a set of game-changing network features to drive application performance to the next level. (2023). Retrieved from https://www.prnewswire.com/news-releases/plume-leverages-ai-to-introduce-full-stack-optimizationa-set-of-game-changing-network-features-to-drive-application-performance-to-the-next-level-301890384.html
12. Is Big Data Better? LMS Data and Predictive Analytic Performance in Postsecondary Education. https://www.edworkingpapers.com/sites/default/files/ai22-647.pdf
13. Rogers and Microsoft announced a strategic alliance to revolutionize hybrid workplace communications and power 5G innovation across Canada with Azure. Retrieved from https://news.microsoft.com/en-ca/2022/03/16/rogers-and-microsoft-announce-strategic-alliance-to-revolutionize-hybrid-workplace-communications-and-power-5g-innovation-across-canada-with-azure/
14. Kuruppath, R. (2023). Emerging Trends in Continuous Integration and Continuous Deployment.
15. Dr. Hou, J. (2023). Zero Touch Operations: How we're using AI to enable IT systems/process automation. https://careers.vodafone.com/life-at-vodafone/projects-stories/how-were-using-ai-to-enable-it-systems/
16. Bone, G., Agarwal, G. (2022). Translating Customer & Employee Experience with Shaw's Data Journey