

डेटा विश्लेषण में अंतर्दृष्टि

पंकज दास, भारती एवं राहुल बनर्जी

भाकृअनुप—भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

संवादी लेखक का ई-मेल: pankaj.iasri@gmail.com

पिछले कुछ दशकों में प्रौद्योगिकी में बदलाव आने के कारण हर क्षेत्र में बड़ी मात्रा में डेटा एकत्रित हो रहे हैं। यह डेटा उस क्षेत्र के बारे में मूल्यवान अंतर्दृष्टि प्रदान कर सकता है। हालाँकि एकत्रित डेटा का उपयुक्त विश्लेषण करने की आवश्यकता है। व्यवसायों और उद्योगों के एकत्रित डेटा से मिली जानकारी को समझने और उचित निर्णय लेने में डेटा एनालिटिक्स मदद करता है।

डेटा विश्लेषण क्या है?

डेटा विश्लेषण डेटा एकत्र करने, मॉडलिंग और विश्लेषण करने की प्रक्रिया है जो निर्णय लेने के लिए अंतर्दृष्टि प्रदान करता है। विश्लेषण के उद्देश्य के आधार पर विश्लेषण करने के लिए कई तरीके और तकनीकें हैं। डेटा विश्लेषण की विभिन्न विधियाँ मुख्य रूप से दो मुख्य क्षेत्रों पर आधारित हैं: अनुसंधान में मात्रात्मक विधियाँ और गुणात्मक विधियाँ। डेटा विश्लेषण के लिए विभिन्न तकनीकों, और मात्रात्मक अनुसंधान के तरीकों के साथ-साथ गुणात्मक अंतर्दृष्टि की जानकारी के द्वारा विश्लेषण के प्रयासों को अधिक स्पष्ट रूप से परिभाषित दिशा मिलेगी, इसलिए इस विशेषज्ञान के लिए समय निकालना उचित है। डेटा विश्लेषण असंगठित डेटा को सार्थक जानकारी में परिवर्तित करने में मदद करता है। डेटा विश्लेषण प्रक्रिया के दौरान तीन आवश्यक चीजें होती हैं। i) डेटा संगठन, ii) सारांशीकरण और वर्गीकरण जो आँकड़ा लघुकरण के लिए उपयोग की जाने वाली विधि है iii) डेटा विश्लेषण— शोधकर्ता इसे टॉप—डाउन या बॉटम—यू दोनों विधियों से करते हैं।

डेटा विश्लेषण क्यों महत्वपूर्ण है?

डेटा विश्लेषण, डेटा की जाँच, पुनः क्रम और प्रस्तुत करने की प्रक्रिया है जो कि योग्य जानकारी उपलब्ध करवाने के लिए महत्वपूर्ण है। शोधकर्ता को डेटा से सार्थक निष्कर्ष

निकालने के लिए यह आवश्यक है। यह शोध का एक बहुत ही महत्वपूर्ण हिस्सा है ताकि परिणामों को ठीक से इस्तेमाल और लागू किया जा सके। यदि किसी अनुसंधान में उपयुक्त डेटा विश्लेषण नहीं किया जाता है, तो यह शोध कार्य के अनुप्रयोग में त्रुटि उत्पन्न कर सकता है। एक शोध पत्र में डेटा विश्लेषण बहुत महत्वपूर्ण है इसलिए प्रत्येक शोधकर्ता को यह सुनिश्चित करने की पूरी कोशिश करनी चाहिए कि शोध कार्य के दौरान जो भी डेटा एकत्रित हुआ है उसका ठीक से विश्लेषण किया जाये। शोध कार्य में, अनुसंधान के लक्ष्य को ध्यान में रखते हुए विभिन्न माध्यमों का उपयोग करके डेटा एकत्र किया जा सकता है, लेकिन डेटा एकत्र करने के बाद उसकी व्याख्या और अनुप्रयोग के लिए डेटा का सही विश्लेषण करना गंभीर रूप से महत्वपूर्ण है। डेटा संग्रह की उचित विधियों का उपयोग करने के बावजूद, यदि डेटा का सही विश्लेषण नहीं किया जाता है, तो यह शोध कार्य वास्तव में अच्छा नहीं होगा। एक उचित डेटा विश्लेषण विश्वसनीय जानकारी, डेटा जटिलता, मानव पूर्वाग्रह का प्रभाव आदि को सही तरीके से समझने में मदद कर सकता है।

अनुसंधान में डेटा के प्रकार

प्रत्येक प्रकार के डेटा में एक विशिष्ट मूल्य निर्दिष्ट करने के बाद चीजों का वर्णन करने का गुण होता है। विश्लेषण के द्वारा, इसे उपयोगी बनाने के लिए, इन मूल्यों को किसी दिए गए संदर्भ में संसाधित और प्रस्तुत करने की आवश्यकता है। डेटा विभिन्न रूपों में हो सकता है। इस खंड में प्राथमिक डेटा के प्रकारों का वर्णन किया गया है।

गुणात्मक डेटा (Qualitative dat): जब प्रस्तुत डेटा में गुणों का विवरण होता है, तो हम इसे गुणात्मक डेटा कहते हैं। गुणात्मक डेटा में किसी व्यक्ति या व्यक्तियों के समूह के महत्वपूर्ण गुणों की व्याख्या होती है, जिसे मात्रात्मक रूप से





नहीं मापा जा सकता है। उदाहरण: लिंग, स्वाद, अनुभव, बनावट, व्यवहार, इत्यादि। हालांकि इस डेटा का अवलोकन किया जा सकता है लेकिन अनुसंधान में विशेष रूप से तुलना के लिए इस डेटा का विश्लेषण करना कठिन है। इस प्रकार का डेटा आमतौर पर फोकस समूहों, व्यक्तिगत साक्षात्कारों या सर्वेक्षणों में प्रश्नों का उपयोग करके एकत्र किया जाता है।

मात्रात्मक डेटा (Quantitative data): संख्यात्मक आंकड़ों में व्यक्त किए गए किसी भी डेटा को मात्रात्मक डेटा कहा जाता है। इस प्रकार के डेटा को श्रेणियों में वर्गीकृत, समूहीकृत, परिकलित या रैंक किया जा सकता है। उदाहरण: उम्र, रैंक, लागत, लंबाई, वजन, अंक आदि इस प्रकार के डेटा के अंतर्गत आते हैं। आप इस तरह के डेटा को ग्राफिकल प्रारूप, चार्ट में प्रस्तुत कर सकते हैं और इस डेटा पर सांख्यिकीय विश्लेषण विधियों को भी लागू कर सकते हैं।

श्रेणीबद्ध डेटा (Categorical data) : यह समूहों में प्रस्तुत डेटा है। हालांकि, श्रेणीबद्ध डेटा में एक वस्तु का एक से अधिक समूहों से संबंध नहीं हो सकता है। उदाहरण: एक व्यक्ति की जीवन शैली, वैवाहिक स्थिति, धूम्रपान की आदत या शराब पीने की आदत श्रेणीबद्ध डेटा के अंतर्गत आता है।

डेटा विश्लेषण के चरण

- **चरण 1: अध्ययन का उद्देश्य परिभाषित करना:** उद्देश्य को परिभाषित करना अर्थात् एक परिकल्पना बनाना और यह पता लगाना कि इसका परीक्षण कैसे किया जाए। उद्देश्य स्पष्ट और संक्षिप्त होना चाहिए। उदाहरण के लिए, यदि एक किसान अपने उत्पाद की बिक्री को बढ़ाना चाहता है तो वह पूछ सकता है कि क्या ग्राहक उत्पाद की गुणवत्ता से संतुष्ट है?
- **चरण 2: डेटा संग्रहण:** उद्देश्य को स्पष्ट रूप से परिभाषित करने और आपकी माप प्राथमिकताओं को निर्धारित करने के बाद दूसरा चरण डेटा एकत्र करना है। सबसे पहले यह निर्धारित करना आवश्यक है कि डेटा किस स्रोत से प्राप्त होगा। क्या डेटा किसी संगठन या वेबसाइट से प्राप्त होगा या एक शोध प्रयोग/सर्वेक्षण करने की आवश्यकता है। इस चरण का लक्ष्य यह सुनिश्चित करना है कि आप जिस समस्या को हल करना

चाहते हैं उस के लिए आपके पास सम्पूर्ण डेटा हो।

- **चरण 3: डेटा क्लीनिंग, वर्गीकरण और सारणीकरण:** डेटा एकत्र करने के पश्चात् अगला चरण इसे विश्लेषण योग्य बनाना है अर्थात् उच्च गुणवत्ता वाला डेटा जैसे कि त्रुटियों, डुप्लिकेट और आउटलेर्स, अवांछित डेटा बिंदुओं को हटाना आदि। एकत्रित डेटा, हमेशा एक असंगठित रूप में होते हैं और उन्हें संगठित करने की आवश्यकता होती है। इसे सांख्यिकीय विश्लेषण के योग्य बनाने के लिए इसे सार्थक और आसानी से समझने योग्य रूप में प्रस्तुत किया जाता है। कुछ विशेषताओं के आधार पर डेटा को विभिन्न वर्ग या उपवर्ग में समूहबद्ध करने की प्रक्रिया को वर्गीकरण कहा जाता है।
- **चरण 4: डेटा विश्लेषण:** इस चरण का उद्देश्य डेटा से सार्थक अंतर्दृष्टि निकालना है। डेटा विश्लेषण की तकनीकों और विधियों का उपयोग करके, डेटा में छिपे हुए पैटर्न और संबंधों की तलाश, अंतर्दृष्टि और भविष्यवाणियों की जा सकती है। डेटा विश्लेषण की तकनीकों और विधियों काफी हद तक शोध कार्य के उद्देश्य पर निर्भर करती है। अविभाज्य या द्विचर विश्लेषण, समय-श्रृंखला विश्लेषण और प्रतिगमन विश्लेषण आमतौर पर प्रयोग होने वाली डेटा विश्लेषण विधियों हैं।
- **चरण 5: परिणामों की व्याख्या:** डेटा विश्लेषण के बाद अगला चरण परिणामों की व्याख्या करना और उनसे सार्थक अंतर्दृष्टि प्राप्त करना है। यदि आप चाहते हैं कि आपके मूल्यवान शोध कार्य को लागू किया जाए, तो आपको इसे स्टेकहोल्डर्स और निर्णय लेने वालों के सामने इस तरह से प्रस्तुत करना चाहिए जो समझने में आसान हो।

डेटा विश्लेषण की चुनौतियाँ

शोधकर्ताओं और डेटा विश्लेषकों को लगातार और गुणवत्तापूर्ण डेटा एकत्र करने में कई चुनौतियों का सामना करना पड़ता है। डेटा संग्रह प्रथाओं को बेहतर बनाने के तरीकों को विकसित करने के लिए, पहले डेटा संग्रह के लिए बाधाओं की पहचान करना आवश्यक है। सबसे प्रचलित चुनौतियों में से कुछ इस प्रकार हैं—



1. असंगत डेटा संग्रह मानक

डेटामानक इस बात की रूप रेखा तैयार करते हैं कि सामान्य डेटा आइटम और जनसांख्यिकीय जानकारी कैसे एकत्र की जानी चाहिए। स्थापित मानकों में आम तौर पर डेटा परिभाषाएं, मानकीकृत प्रश्न और स्वीकृत प्रतिक्रिया विकल्प होते हैं जो लगातार संग्रह प्रथाओं का मार्गदर्शन करते हैं। वर्तमान में, कई राष्ट्रीय और राज्य-व्यापी डेटामानक हैं जिनका उपयोग प्रशासनिक डेटा एकत्र करने के लिए किया जाता है। ये मानक हमेशा व्यापक रूप से लागू नहीं होते हैं, और स्वयं असंगत हो सकते हैं, और यह डेटा संग्रह की तुलना को प्रभावित कर सकता है।

2. डेटा संग्रह का संदर्भ

डेटा संग्रह विभिन्न परिस्थितियों पर निर्भर करता है और एकत्रित डेटा भिन्न स्थितियों के आधार पर भिन्न हो सकता है। कई परिस्थितियों में पूर्ण व सटीक जानकारी प्राप्त करना मुश्किल हो सकता है जैसे कि आय का डेटा संग्रह, आपराधिक मामलों का डेटा, इत्यादि। ज्यादातर मामलों में, डेटा एकत्र करने वाले व्यक्ति की प्राथमिक भूमिका होती है उदाहरण के लिए, एक पुलिस अधिकारी, सहायक कार्यकर्ता या चिकित्सा व्यवसायी के रूप में। संकट या आपातकालीन आदि स्थितियों में डेटा संग्रह सीमित हो सकते हैं जहाँ कार्यकर्ता किसी व्यक्ति की सुरक्षा को प्राथमिकता दे रहे हैं, या ऐसी परिस्थितियों जहाँ किसी व्यक्ति की गोपनीयता से समझौता नहीं किया जा सकता है।

3. डेटा संग्रह व्यावसायिक कार्य के लिए मूल नहीं है

संगठन के मुख्य कार्य और सेवा वितरण में समय का दबाव संगठन द्वारा एकत्र किए जाने वाले डेटा के प्रकार और गुणवत्ता को प्रभावित कर सकता है। प्रशासनिक डेटा को आम तौर पर परिचालन आवश्यकताओं के उप-उत्पाद के रूप में या आंतरिक व्यावसायिक आवश्यकता को पूरा करने के लिए एकत्र किया जाता है और इसमें केवल ग्राहक के संपर्क विवरण जैसे सेवा करने के लिए आवश्यक मुख्य जानकारी शामिल हो सकती है। ऐसे मामलों में, किसी व्यक्ति की यौन अभिविन्यास, सांस्कृतिक पृष्ठभूमिया विकलांगता की जानकारी को विशेष सेवाओं की पेशकश नहीं करने वाले

संगठनों के लिए एक परिचालन आवश्यकता के रूप में नहीं देखा जा सकता है। नतीजतन, संगठन केवल डेटा आइटम की एक संकीर्ण श्रेणी एकत्र कर सकते हैं, जिसमें व्यापक माध्यमिक उपयोग के उद्देश्यों के लिए पर्याप्त विवरण की कमी होती है, जैसे कि राज्य-व्यापी सेवाविश्लेषण, निगरानी या अनुसंधान करना।

4. डेटा जटिलता

कुछ मामलों में, एक डेटा आइटम के माध्यम से किसी व्यक्ति की पृष्ठ भूमि के बारे में पर्याप्त जानकारी का पता नहीं लगाया जा सकता है, उदाहरण के लिए विकलांग लोगों के लिए। जहाँ इसका प्रयास किया जाता है, यह अक्सर उन लोगों का कम प्रतिनिधित्व करता है जो सेवाओं तक पहुंचने के लिए उच्च जोखिम और बाधाओं का सामना करते हैं। इसमें विभिन्न अवधारणाओं के बारे में भ्रम जोड़ने की भी क्षमता है जो विशिष्ट समुदायों के बाहर के लोगों द्वारा पूरी तरह से समझ में नहीं आती हैं।

5. डेटा संग्रह में प्रशिक्षण का अभाव

चूंकि फ्रंट-लाइन सेवा और नैदानिक कर्मचारियों की प्राथमिक भूमिका आम तौर पर डेटा संग्रह नहीं होती है, उन्हें इस क्षेत्र में प्रशिक्षण प्राप्त नहीं होती है। यदि कर्मचारी में प्रशिक्षण प्राप्त नहीं होता है तो वे संबंधित प्रश्न पूछने के लिए कम आत्मविश्वास महसूस कर सकते हैं। कुछ प्रकार के डेटा को कैसे और क्यों एकत्र किया जाए, इस बारे में प्रशिक्षण की कमी डेटा संग्रह को प्रभावित कर सकती है।

6. गुणवत्ता आश्वासन प्रक्रियाओं का अभाव

शुरु में एकत्र किए गए डेटा को सत्यापित नहीं किया जा सकता है क्योंकि किसी सेवा के संपर्क में रहने वाले व्यक्ति के साथ जानकारी की पुष्टि करने के सीमित अवसर हो सकते हैं। इसके अतिरिक्त, रिकॉर्ड रखरखाव सिस्टम का परिष्कार भिन्न हो सकता है और डेटा की गुणवत्ता अक्सर डेटा को सही ढंग से दर्ज करने वाले व्यक्ति पर निर्भर होती है। किसी भी संगठन में संसाधन की उपलब्धता के आधार पर, कर्मचारियों को पूर्ण जानकारी प्राप्त करने का समय नहीं मिलता है।





7. आर्थिक और आईटी प्रतिबंध

कुछ संगठनों के पास डेटा संग्रह प्रणालियों और प्रक्रियाओं में सुधारों का बुनियादी ढांचा नहीं होता। यह बजट की कमी के कारण हो सकता है। कई आईटी सिस्टम सरकारी विभागों द्वारा प्रदान किए जाते हैं, जो सिस्टम अपडेट और संचालन की जिम्मेदारी लेते हैं। ये अपडेट महंगे हो सकते हैं और इसमें समय लग सकता है। कुछ मामलों में, इन आईटी प्रणालियों में कई प्रतिक्रिया मूल्यों या गतिशील पूछताछ को शामिल करने की सीमित क्षमता हो सकती है, जो परिष्कृत डेटा संग्रह का समर्थन करती है। एकाधिक प्रतिक्रिया विकल्प डेटा विश्लेषण के लिए समस्याएं पैदा कर सकते हैं।

इन चुनौतियों से निपटने के उपाय

उपरोक्त चुनौतियाँ सबसे अधिक प्रचलित हैं। कुछ समस्याएं प्रकृति में अंतर्निहित हैं, उनके लिए 100 प्रतिशत सुधार संभव नहीं हो सकता है। लेकिन कुछ समस्याएं दोषपूर्ण तरीकों या तकनीकों के कारण होते हैं जिन्हें सुधारा जा सकता है। सामान्य तौर पर, समस्याओं को दूर करने के लिए कुछ कदम उठाए जा सकते हैं जैसे कि डेटा संग्रह में सुधार

- डेटा संग्रह में सुधार के लिए संगठन के सभी स्तरों से प्रतिबद्धता
- एक डेटा स्रोत के भीतर जनसांख्यिकीय डेटा की गुणवत्ता में सुधार
- एकाधिक डेटा स्रोतों में डेटा की गुणवत्ता में सुधार
- संबंधित कर्मियों का प्रशिक्षण
- आईटी मुकदमेबाजी के लिए सरकारी नीतियों में सुधार
- नियमित निगरानी और समीक्षा

निष्कर्ष

डेटा विश्लेषण, उपयोगी जानकारी की खोज, निष्कर्ष निकालने और निर्णय लेने में मदद करता है। डेटा के अध्ययन को सरल और सटीक बनाने में डेटा विश्लेषण का महत्वपूर्ण योगदान है। यह शोधकर्ताओं को डेटा की व्याख्या करने और सार्थक जानकारी में परिवर्तित करने में मदद करता है।

अपनी मातृभाषा बंगला में लिखकर मैं बंगबन्धु तो हो गया, किन्तु भारतबन्धु मैं तभी हो सकूँगा जब भारत की राष्ट्रभाषा में लिखूँगा।

– बंकिम चन्द्र चट्टोपायाय

