

संवेदनशील डाटा के संरक्षण में डाटा खनन का महत्व

शालिनी लाम्बा¹, श्वेता सिन्हा², विजय कुमार पाण्डे³^{1,2,3}असिस्टेंट प्रोफेसर, कम्प्यूटर विभाग

नेशनल पीएचडी कॉलेज, लखनऊ-228001, उ०प्र०, भारत

shalinilamba22@gmail.com, singh_shweta2005@yahoo.com, vijay.kumar0384@gmail.com

प्राप्त तिथि 29.07.2015, स्वीकृत तिथि 20.08.2015

गत वर्षों में संचय करने और उपभोक्ताओं के बारे में व्यक्तिगत डाटा रिकॉर्ड करने की क्षमता में वृद्धि करने के लिए हार्डवेयर प्रौद्योगिकी में प्रगति लायी गयी है। इसने इस चिन्ता को जन्म दिया है कि यह व्यक्तिगत डाटा कुछ प्रयोजनों के लिए दुरुपयोग किया जा सकता है। यह विश्लेषण उद्देश्यों के लिए संवेदनशील डाटा के आदान-प्रदान की अनुमति देता है, क्योंकि इन चिन्ताओं को दूर करने के लिए डाटा खनन के संरक्षण की गोपनीयता तेजी से लोकप्रिय हो गयी है। जिसके परिणामस्वरूप, जो अपना डाटा साझा नहीं करना चाहते हैं, उनकी संख्या में वृद्धि होती जा रही है या फिर वह अपना गलत डाटा उपलब्ध करा रहे हैं। डाटा खनन के संरक्षण एवं गोपनीयता का बड़े पैमाने पर अध्ययन किया गया है जिससे इंटरनेट पर संवेदनशील जानकारी का व्यापक प्रचार प्रसार किया जा सके।

डाटा खनन का लक्ष्य डाटा से बड़ी मात्रा में व्यक्तिगत ज्ञान को निकालना है। हालांकि डाटा अक्सर कई अलग-अलग साइटों से एकत्र किया जाता है। सूचना प्रणाली में गोपनीयता सबसे महत्वपूर्ण सम्पत्ति में से एक है। इस कारण से कई प्रयासों के ज्ञान की खोज के दौरान संवेदनशील सूचना के प्रकटीकरण को रोकने के क्रम में डाटा खनन एल्गोरिदम के साथ गोपनीयता के संरक्षण तकनीकों को शामिल करने के लिए समर्पित किया गया है। गोपनीयता, कानूनी और व्यवसायिक चिन्ता इस डाटा के केंद्रीकृत उपयोग को प्रतिबंधित करता है। गोपनीयता के संरक्षण डाटा खनन एल्गोरिदम का उद्देश्य एक ही समय में संवेदनशील जानकारी पर रक्षा करते हुए डाटा की बड़ी राशि से प्रासंगिक ज्ञान को निकालने के लिये है। ऐसे एल्गोरिदम की डिजाइन में एक महत्वपूर्ण पक्ष उपयुक्त मूल्यांकन मानदंडों की पहचान और संबंधित मानक का विकास है।



यह उल्लेख किया जा सकता है कि डाटा माइनिंग एक जटिल और कठिन प्रक्रिया है। डाटा हमेशा शाब्दिक प्रारूप में मौजूद नहीं है। विशिष्ट खनन तकनीक सभी प्रकार के डाटा को संभालने के लिये विकसित किया जाना चाहिए। दूसरी बात, डाटा माइनिंग सिस्टम एल्गोरिदम कुशल और मापक ढंग से डाटा को संभालने के लिए सक्षम होना चाहिये। तीसरा, डाटा माइनिंग एक डाटा सेट के भीतर शोर या लापता डाटा को संभालने और डाटा खनन एल्गोरिदम के एक व्यापक ज्ञान के रूप में डाटा का सही प्रतिनिधित्व उत्पादन करने में सक्षम होना चाहिये। डाटा की गुणवत्ता डाटा माइनिंग का एक महत्वपूर्ण पहलू है। डाटा खनन कार्य के लिए विशेष रूप से तैयार किया जा चुका है कि उच्च गुणवत्ता वाले डाटा, उपयोगी डाटा माइनिंग मॉडल और उत्पादन में उपयोगी परिणाम देता है।

इसके विपरीत, कम गुणवत्ता डाटा, डाटा खनन परिणामों की उपयोगिता पर महत्वपूर्ण नकारात्मक प्रभाव डालता है। यदि इसका हल नहीं है, तो डाटा गुणवत्ता की समस्या भी डाटा गोदाम के कार्यान्वयन और उपयोगिता में डाटा खनन के परिणाम में देरी कर सकता है। संगठन दैनिक आधार पर बहुत अधिक डाटा इकट्ठा करता है तो प्रश्न यह नहीं है, कि इसे रखा कैसे जाये, बल्कि प्रश्न यह है कि इससे उपयोगी जानकारी को हासिल कैसे किया जाय।

यह आभास हो रहा है कि आज कल डेटा खनन विशाल रेंज के संगठन के लिए एक व्यापक रूप से स्वीकृत तकनीक है। संगठन उनके हर दिन की गतिविधियों में डाटा खनन पर अधिकाधिक निर्भर है। प्रायः डाटा खनन में मेडिकल और वित्तीय जानकारी के रूप में व्यक्तिगत जानकारी होती है। डाटा खनन की पूरी प्रक्रिया के दौरान अक्सर कलेक्टर उपयोगकर्ता और मापक सहित कई दलों की जानकारी मिल जाती है। इस तरह के संवेदनशील सूचनाओं का खुलासा व्यक्तिगत जानकारी का उल्लंघन कर सकता है। गोपनीयता तथा संगठन में जनता में विश्वास की कमी के कारण जनता जागरूकता एवं डाटा संग्रहण करने के कार्य में अतिरिक्त जटिलता पेश आ सकती है। सख्त गोपनीयता कानून भविष्य में लाया जायेगा या नहीं, की संभावना नहीं है। दूसरी तरफ इस तरह के कानून में व्यक्ति के बिना व्यक्तिगत विषय में पूरी जानकारी प्राप्त करने में अतिरिक्त कठिनाई हो सकती है, जिसके परिणाम स्वरूप व्यक्ति अपनी निजी जानकारी देने के लिए संकोच कर सकता है।

एक ही समय में संवेदनशील जानकारी परीक्षण, तकनीक की उत्पत्ति एवं गोपनीयता संरक्षण डाटा माइनिंग तकनीक की विशाल बुद्धि कहा जाता है। एक आवश्यक मुद्दा इन गोपनीयता संरक्षण तकनीक के बीच यह है कि एक संवेदनशील सूचना की सुरक्षा के लिए किस प्रकार की व्यवस्था करनी चाहिये।

संदर्भ

1. अग्रवाल, चारु सी0 एवं फिलिप एस0 वाई0 यू0(2008) ए जनरल सर्वे ऑफ प्राइवैसी-प्रिजर्विंग डाटा माइनिंग मॉडल्स एण्ड एल्गोरिदम, अध्याय-प्राइवैसी-प्रिजर्विंग डाटा माइनिंग, खण्ड-34, मु0पृ0 11-52।
2. तोमर, अर्चना; रिचहरीया, विनीतया एवं पाण्डे, आर0 के0(2011) ए कम्प्रेहेंसिव सर्वे ऑफ प्राइवैसी प्रिजर्विंग एल्गोरिदम ऑफ एसोसिएशन रूल माइनिंग इन सेंट्रलाइज्ड डाटाबेस, इण्टरनैशनल ज0 ऑफ कम्प्यूटर एप्लीकेशंस, खण्ड-16, अंक-5, मु0पृ0 23-27।
3. नायक, गायत्री एवं देवी, स्वागतिका(2011) ए सर्वे ऑन प्राइवैसी प्रिजर्विंग डाटा माइनिंग: एप्रोचेज एण्ड टेक्नीक्स, इण्टरनैशनल ज0 ऑफ साइंस एण्ड टेक्नोलॉजी, खण्ड-3, अंक-3, मु0पृ0 2127-2133।