

उच्च शिक्षा में डाटा माइनिंग

श्वेता सिन्हा एवं शालिनी लाम्बा
असिस्टेंट प्रोफेसर, कम्प्यूटर विज्ञान विभाग
नेशनल पी0 जी0 कॉलेज, लखनऊ-226001, उ0प्र0, भारत
singh_shweta2005@yahoo.com, shalinilamba22@gmail.com

प्राप्त तिथि-30.07.2015, स्वीकृत तिथि-10.08.2015

शिक्षा देश की प्रगति के लिये एक अनिवार्य तत्व है। आज की अर्थव्यवस्था की स्थिति में प्रगति के लिये उच्च शिक्षा का महत्व स्पष्ट हो चुका है। उच्च शिक्षण संस्थानों का मुख्य उद्देश्य अपने छात्रों को गुणवत्ता परक शिक्षा प्रदान करना है। उच्च शिक्षा प्रणाली में गुणवत्ता के उच्चतम स्तर को तभी प्राप्त किया जा सकता है, जब विभिन्न स्तरों पर जैसे- एक विशेष कोर्स में छात्रों के नामांकन की भविष्यवाणी के लिये, पारंपरिक कक्षा शिक्षण मॉडल का अलगाव, ऑनलाईन परीक्षा में प्रयोग होने वाले अनुचित साधन का पता लगाने में, परीक्षा परिणाम में असामान्य मूल्यों का पता लगाने में, छात्रों के संभावित निष्पादन के बारे में आदि, की भविष्यवाणी के लिये ज्ञान प्राप्त हो। उच्च शिक्षा प्रणाली में गहरी और पर्याप्त ज्ञान की कमी के कारण सिस्टम प्रबंधन गुणवत्ता के उद्देश्यों को प्राप्त करने में असमर्थ है। यह ज्ञान शैक्षिकी डाटा सेट के बीच छिपा हुआ है और यह खनन(माइनिंग) तकनीकों के माध्यम से निष्कर्षण है। शैक्षिक वातावरण में खनन को शिक्षा डाटा खनन (ईडीएम) के नाम से जाना जाता है। इसमें छात्र की प्रवृत्तियों और शिक्षा के प्रति व्यवहार का विश्लेषण करने के लिये शिक्षा के बड़े खजाने के डाटा से ज्ञान की खोज करने के लिये नए उपकरण तकनीक और अनुसंधान को विकसित किया जाता है।

शैक्षिक डाटा खनन शिक्षा से संबंधित बड़े डाटाबेस से भविष्य में उपयोग में आने वाली छिपी जानकारी की निकासी की एक तकनीक है। यह एक शक्तिशाली तकनीक है जो विश्वविद्यालयों और संस्थानों को उनके डाटा गोदामों में छिपी महत्वपूर्ण सूचना पर ध्यान केंद्रित करने की क्षमता प्रदान करने में मदद करती है। शिक्षण संस्थानों के समक्ष सबसे बड़ी चुनौती यह है कि शैक्षिक डाटा में विस्फोटक वृद्धि को शिक्षा में प्रबंधकीय फैसलों की गुणवत्ता और शैक्षिक स्तर की बढ़ोत्तरी में कैसे उपयोग में लाया जा सकता है। यह डाटा व्यक्तिगत या शैक्षिक हो सकता है जो छात्रों के व्यवहार को समझने के लिए, प्रशिक्षकों द्वारा शिक्षण में सुधार करने में, पाठ्यक्रमों में सुधार लाने, छात्रों को एक बेहतर कैरियर चुनने में मदद करने में, ई-लर्निंग सिस्टम का मूल्यांकन और सुधार करने के लिए सहायता करने में प्रयोग किया जा सकता है। इस तकनीक के तहत संस्थान यह सूचना प्राप्त कर सकते हैं कि छात्रों का किस विषय में विशेष रूप से रुझान है, या किस विषय में अधिक शैक्षिक सहायता की आवश्यकता है। कमजोर छात्र को पहचानने में उपयोगी है जिससे उन्हें अधिक सहायता प्रदान की जा सके। यह सूचना इस बात पर प्रकाश डालती है कि कौन से छात्र असफल हो रहे हैं और किस प्रकार वह और अत्यधिक अध्ययन कर सकते हैं। डाटा माइनिंग छात्रों की झाप आउट संख्या को ज्ञात करके छात्रों को स्कूल में अध्ययन करने के लिए प्रेरित करने में सहायक है और स्कूल छोड़ने वाले छात्रों की संख्या में कमी लाने में अत्यधिक कारगर सिद्ध होता है।

शिक्षा प्रणाली में डाटा खनन को लागू करने के चक्र

शैक्षिक डाटा खनन का मुख्य कार्य विभिन्न तरीकों और एल्गोरिदम की मदद से शिक्षा के क्षेत्र में संग्रहीत डाटा की बड़ी मात्रा में उपयोगी पैटर्न को निकालना है। यह डाटा व्यापक, सूक्ष्म और सटीक होता है और अतिरिक्त विश्लेषण में उपयोग किया जाता है। निष्कासित डाटा में पारस्परिक सम्बन्ध को ज्ञात कर के उसकी मान्यता को तय किया जाता है, फिर शिक्षा में भविष्य की घटनाओं के लिए भविष्यवाणी करने के निर्णय लेने की प्रक्रिया का समर्थन करने के लिए लागू किया जाता है। डाटा माइनिंग के मुख्य चार चरण हैं-

1. **समस्या की परिभाषा-** डाटा खनन विशेषज्ञ और डोमेन विशेषज्ञ के सामूहिक प्रयासों से परियोजना की आवश्यकता और उसके उद्देश्य को परिभाषित करना।
2. **डाटा अन्वेषण-** डाटा का संग्रह कर डाटा का विवरण देना। डाटा की गुणवत्ता की समस्या को पहचानना। इसमें पारंपरिक डाटा विश्लेषण उपकरण, जैसे सांख्यिकी का उपयोग किया जाता है।
3. **डाटा तैयारी-** डाटा खनन परिणामों की गुणवत्ता में सुधार के लिए तथा आगे के विश्लेषण और प्रसंस्करण के लिए आंकड़ों में उपयुक्त हेर-फेर करना। इसमें कई अलग-अलग कार्य शामिल हैं।
4. **डाटा मॉडलिंग-** यह कई प्रक्रियाओं का समूह है, जिसमें डाटा के एकाधिक सेटों को संयुक्त कर विश्लेषण किया जाता है जिससे डाटा में स्थित सम्बन्ध तथा पैटर्न को उजागर किया जा सके।
5. **मूल्यांकन-** प्राप्त किये गये पैटर्न से अपेक्षित ज्ञान प्राप्त हो रहा है या नहीं इस पर निर्भर करता है कि वह मॉडल लागू किया जायेगा अन्यथा मॉडल के मापदंडों को अपेक्षानुसार बदल कर फिर बनाया जाता है।

शैक्षिक डाटा माइनिंग में 'क्लासिफिकेशन', 'कैटेगरीजेशन', 'ऐस्टिमेशन', और 'विज्युलाइजेशन' तकनीकियों का उपयोग कर ज्ञान की खोज करी जाती है। 'क्लासिफिकेशन' संघों और समूह की पहचान कर इस अध्ययन के तहत विषयों को अलग करता है। उच्च शिक्षा संस्थानों में छात्र विशेषताओं के एक व्यापक विश्लेषण के लिए 'क्लासिफिकेशन' का उपयोग कर सकते हैं। 'कैटेगरीजेशन' में इनडक्शन एल्गोरिदम का उपयोग कर अलग अलग कैटेगरी जैसे "ड्रॉप आउट छात्र", "स्थानान्तरण छात्र" आदि बनाये जाते हैं। 'ऐस्टिमेशन' से जी0पी0ए0 और वेतन स्तर के रूप में होने वाले निरंतर परिणाम को ध्यान में रखते हुए भविष्य की संभावनाओं को ज्ञात किया जा सकता है। सफलता के रूप में परिणामों की एक किस्म की संभावना की भविष्यवाणी करने के लिए 'ऐस्टिमेशन' का उपयोग करते हैं। 'विज्युलाइजेशन' गणितीय प्रेरित नियमों और स्कोर को प्रदर्शित करने के लिए इंटेक्टिव रेखांकन का उपयोग करता है और पाई या बार चार्ट की तुलना में कहीं अधिक परिष्कृत है।

शिक्षा प्रणाली में डाटा खनन को लागू करने के चक्र— शैक्षिक डाटा खनन का मुख्य कार्य विभिन्न तरीकों और एल्गोरिदम की मदद से शिक्षा के क्षेत्र में संग्रहीत डाटा की बड़ी मात्रा से उपयोगी पैटर्न को निकालना है। यह डाटा व्यापक, सूक्ष्म और सटीक होता है और अतिरिक्त विश्लेषण में उपयोग किया जाता है। निष्कासित डाटा में पारस्परिक सम्बन्ध को ज्ञात कर के उसकी मान्यता को तय किया जाता है, फिर शिक्षा में भविष्य की घटनाओं के लिए भविष्यवाणी करने के निर्णय लेने की प्रक्रिया का समर्थन करने के लिए लागू किया जाता है। डाटा माइनिंग के मुख्य चार चरण हैं —

1. **समस्या की परिभाषा**— डाटा खनन विशेषज्ञ और डोमेन विशेषज्ञ के सामूहिक प्रयासों से परियोजना की आवश्यकता और उसके उद्देश्य को परिभाषित करना ।
2. **डाटा अन्वेषण**— डाटा का संग्रह कर डाटा का विवरण देना। डाटा की गुणवत्ता की समस्या को पहचानना। इसमें पारंपरिक डाटा विश्लेषण उपकरण, जैसे सांख्यिकी का उपयोग किया जाता है।
3. **डाटा तैयारी**— डाटा खनन परिणामों की गुणवत्ता में सुधार के लिए तथा आगे के विश्लेषण और प्रसंस्करण के लिए आंकड़ों में उपयुक्त हेर-फेर करना। इसमें कई अलग-अलग कार्य शामिल हैं।
4. **डाटा मॉडलिंग**— यह कई प्रक्रियाओं का समूह है, जिसमें डाटा के एकाधिक सेटों को संयुक्त कर उसका विश्लेषण किया जाता है जिससे डाटा में स्थित सम्बन्ध तथा पैटर्न को उजागर किया जा सके।
5. **मूल्यांकन**— प्राप्त किये गये पैटर्न से अपेक्षित ज्ञान प्राप्त हो रहा है या नहीं इस पर निर्भर करता है कि वह मॉडल लागू किया जायेगा अन्यथा मॉडल के मापदंडों को अपेक्षानुसार बदल कर फिर बनाया जाता है।

शैक्षिक आंकड़ों पर यह डाटा माइनिंग तकनीक उच्च शिक्षण संस्थानों के प्रशासन और बुद्धिजीवियों के लिए शैक्षिक प्रक्रियाओं की गुणवत्ता में सुधार के लिए एक उपयोगी सामरिक उपकरण साबित हो सकता है।

संदर्भ

1. रोमेरो, सी0 एवं वेंचुरा, एस0(2007) शैक्षिक डाटा माइनिंग: 1995—2005 एक सर्वेक्षण, विशेषज्ञ प्रणाली आवेदनों के साथ, खण्ड—33, मु0पू0 135—146।
2. आयशा, शीला; मुस्तफा, तस्लीम; सत्तार, एहसान रजा एवं खान, एम0 इनायत(2010) उच्च शिक्षा प्रणाली के लिए डाटा माइनिंग मॉडल, यूरोपियन जर्नल वैज्ञानिक अनुसंधान, खण्ड—43, अंक—1, मु0पू0 24—29।
3. रशान, के0 एच0 एवं पीरिज, अनुष्का(2010) शिक्षा के क्षेत्र में डाटा माइनिंग अनुप्रयोग, एम.एस.आई.टी. कार्नेगी मेलॉन विश्वविद्यालय, रिट्रीव्ड ऑन 28.01.2011

