



भारतीय कृषि सांख्यिकी संस्था की पत्रिका

अंक 73, खंड 2, अगस्त 2019 175-180

हिन्दी परिशिष्ट: इस खंड में प्रकाशित शोधपत्रों के सारांश

अंक 73

अगस्त 2019

खंड 2

अनुक्रमणिका

कृषि सांख्यिकी: सिद्धांत एवं अनुप्रयोग

- निकटतम प्रभाव के साथ लैटिन वर्गीय अभिकल्पना
सोबिता सपम, नृपेश मंडल एवं बिकास सिन्हा
- इकाई स्तर पर समष्टि स्तर की सहायक जानकारी उपलब्ध होने पर दो चरण के प्रतिदर्श की अभिकल्पना के तहत परिमित समष्टि का उच्च क्रम अंशांकन अनुमान
कौस्तव आदित्य, हुकुम चन्द्र, सुशील कुमार एवं श्रीला दास
- विभिन्न अनाजों के भौतिक गुणों व आकार संबंधित निर्धारण के लिए इमेज प्रोसेसिंग आधारित सॉफ्टवेयर
करण सिंह, नचिकेत कोतवालीवाले, अभिमन्यु कल्ले एवं माधवी तिवारी
- के-मीन्स क्लस्टरिंग के साथ फजी तकनीक पर आधारित स्पेशियो-टेम्पोरल मॉडलिंग व पूर्वानुमान के लिए एक नया दृष्टिकोण
अमित साहा, के.एन. सिंह, मृणमय रे, संजय कुमार एवं संतोष राठोड़
- संतुलित द्विपक्षीय उत्पन्न पंक्ति-स्तंभ अभिकल्पना
अनिन्दिता दत्ता, सीमा जग्गी, सिनी वर्गीज एवं एल्डो वर्गीज
- पपीते में बाह्य आकड़ों के होते हुए रोबस्ट एनोवा सिद्धांत का अनुप्रयोग
आर. वेणुगोपालन एवं बी.एल. मंजुनाथ
- द्वि आदेश ऑर्थोगोनल लैटिन हाइपरक्यूब अभिकल्पना पर एक नोट
सुकांत दाश, बैद्यनाथ मंडल, राजेन्द्र प्रसाद एवं सुशील कुमार सरकार
- ओडिशा में ग्रामीण निर्धनता को प्रभावित करने वाले कारकों पर एक अध्ययन
बिज्ञान नंद मोहन्ती, ए.के.पी.सी. स्वेन एवं फिरोज खान
- फसल उपज अनुमान के लिए पदानुक्रमित बेज कुल स्तर पर स्पेशियल प्रतिमान
प्रियंका अंजोय एवं हुकुम चन्द्र
- अपूर्ण बहु-प्रतिक्रिया प्रयोगों में आउटलियर्स मार्किंग की उपस्थिति
राजू कुमार एवं एल.एम. भर
- खाद्य पदार्थों के लिए एसिड बैक्टीरिया-स्टार्टर कल्चर का विकास
सुनीता सिंह, संगीता गुप्ता, सुकांत दाश एवं के.एन. सिंह

संगणक अनुप्रयोग

- C4.5 निर्णय वृक्ष एल्गारिदम का उपयोग करते हुए ऑनलाइन वर्गीकरण व विजुअलाइजेशन
शशि दहिया, सुवजीत दास एवं अंशु भारद्वाज

निकटतम प्रभाव के साथ लैटिन वर्गीय अभिकल्पना

सोबिता सपम¹, नृपेश मंडल² एवं बिकास सिन्हा³

¹मणिपुर विश्वविद्यालय, कांचीपुर, इंफाल

²कलकत्ता विश्वविद्यालय, कोलकाता

³भारतीय सांख्यिकीय संस्थान, कोलकाता

इस लेख में प्रस्तुत शोध कार्य कृषि प्रयोगों के संदर्भ में एक वास्तविक जीवन परिदृश्य से प्रेरित है। यह माना जाता है कि इन भूखंडों में लागू उपचारों के पड़ोसी प्रभावों के माध्यम से एक ब्लॉक डिजाइन या एक लैटिन वर्गीय अभिकल्पना (एल एस डी) में पड़ोसी भूखंड औसत उपज के संदर्भ में एक दूसरे को प्रभावित करते हैं। हम एक रेखीय मॉडल पर विचार करते हैं और इसके विश्लेषण का काफी विवरणों में अध्ययन करते हैं। अनुमाननीय पैरामीट्रिक कार्यों की प्रकृति और विशेष रूप से त्रुटि कार्यों की प्रकृति पर विशेष जोर देने के साथ-साथ हम ऑर्डर के लैटिन वर्गीय अभिकल्पना पर ध्यान केन्द्रित करते हैं।

इकाई स्तर पर समष्टि स्तर की सहायक जानकारी उपलब्ध होने पर दो चरण के प्रतिदर्श की अभिकल्पना के तहत परिमित समष्टि का उच्च क्रम अंशांकन अनुमान

कौस्तव आदित्य¹, हुकुम चन्द्र¹, सुशील कुमार¹
एवं श्रीला दास²

¹भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

²भा.कृ.अ.प.-भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली

सहायक जानकारी का उपयोग प्रायः परिमित समष्टि के अनुमानकों की सटीकता में सुधार करने के लिए किया जाता है। सर्वेक्षण दृष्टिकोण में सहायक जानकारी का कुशल उपयोग करने के लिए अंशांकन दृष्टिकोण (डेविल एवं सरंडल, 1992) का व्यापक रूप से उपयोग किया जाता है। आदित्य व अन्य (2016) ने प्रस्तावित प्रतिगमन प्रकार की अनुमान समष्टि का कुल अंशांकन दृष्टिकोण के तहत अंशांकन दृष्टिकोण का उपयोग करते हुए समष्टि स्तर की सहायक जानकारी द्वि चरण प्रतिदर्श अभिकल्पना

के तहत माध्यमिक चरण इकाई स्तर पर उपलब्ध कराई है। इस लेख में हमने आदित्य व अन्य द्वारा प्रस्तावित प्रतिगमन प्रकार के अनुमानक के एक बेहतर विचरण अनुमानक का प्रस्ताव किया है (सिंह व अन्य 1998) हमने प्रस्तावित अनुमानकों के अनुभव जन्य प्रदर्शन को प्रदर्शित करने के लिए सीमित सिमुलेशन अध्ययन किए हैं। हमारे अनुभवजन्य परिणाम बताते हैं कि प्रस्तावित अनुमानक प्रतिगमन प्रकार के अनुमानक के प्रकारों के सामान्य अनुमानक (आदित्य व अन्य 2016) से बेहतर प्रदर्शन करता है।

विभिन्न अनाजों के भौतिक गुणों व आकार संबंधित निर्धारण के लिए इमेज प्रोसेसिंग आधारित सॉफ्टवेयर

करण सिंह¹, नचिकेत कोतवालीवाले¹,
अभिमन्यु कल्ले² एवं माधवी तिवारी²

¹भा.कृ.अ.प.-केन्द्रीय कृषि अभियांत्रिकी संस्थान, भोपाल

²इंदिरा गांधी कृषि विश्वविद्यालय, रायपुर

विभिन्न कृषि अनुप्रयोगों में कृषि वस्तुओं जैसे आकार, बनावट, रंग आदि के रूपात्मक गुणों की आवश्यकता होती है, जिसमें गुणवत्ता मूल्यांकन, फसल, किस्म की पहचान और मशीनरी का डिजाइन शामिल है। इन विशेषताओं का मैनुअल माप कठिन है, समय भी अधिक लगता है और प्रायः इतना सटीक भी नहीं होता है। छवि प्रसंस्करण पर आधारित एक दृष्टिकोण को कंप्यूटर सॉफ्टवेयर के रूप में विकसित किया गया है। यह दृष्टिकोण एक तेज गैर आक्रामक और मात्रात्मक है। सॉफ्टवेयर फ्लैटबेड स्कैनर का उपयोग करके एक छवि पढ़ता है, अधिमानतः टीआईएफएफ में और तुरंत 0.1 मिमी की सटीकता के लिए भौतिक पैरामीटर प्रदान करता है। वर्तमान कार्य का उद्देश्य वस्तु के ज्ञात आकार के माध्यम से अंशांकन द्वारा विकसित सॉफ्टवेयर की सटीकता का पता लगाना था। सॉफ्टवेयर की सटीकता के अनाजों की कई छवियों पर परिक्षण किया गया है और परिणाम मैनुअल विधियों के अनुरूप पाए जाते हैं। लंबाई व चौड़ाई के संदर्भ में आरएमएसई औसत मान क्रमशः 0.36 और 0.33 पाया गया।

के-मीन्स क्लस्टरिंग के साथ फजी तकनीक पर आधारित स्पेशियो-टेम्पोरल मॉडलिंग व पूर्वानुमान के लिए एक नया दृष्टिकोण

अमित साहा, के.एन. सिंह, मृणमय रे, संजय कुमार एवं संतोष राठोड़

भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

अनुपात लौकिक डेटा की बढ़ती पहुंच के साथ कई क्षेत्रों में दिन-प्रतिदिन अस्थायी मॉडलिंग का महत्व बढ़ रहा है। स्पेशियो-टेम्पोरल टाइम सीरिज के डेटा आपस में स्पेशियो-टेम्पोरल सम्बन्ध प्रदर्शित करते हैं। इसलिए, उन संबंधों को पहचानना और बेहतर पूर्वानुमान प्राप्त करने के लिए इसे मॉडल में सम्मिलित करना आवश्यक है। दूसरी ओर, फजी तकनीकों पर आधारित पूर्वानुमान वर्षा, तापमान आदि जैसे आकस्मिक आंकड़ों के लिए बहुत उपयोगी होते हैं। गत तीन दशकों में, फजी विभिन्न अविभाज्य विधियों का विकास किया गया है। यद्यपि ये विधियाँ केवल एकल समय श्रृंखला के लिए ही मान्य हैं, इसलिए यह आवर्त-लौकिक समय श्रृंखला के लिए उपयोग नहीं किया जा सकता है। इस सीमा को पार करने के लिए मीन्स क्लस्टरिंग के साथ संयोजन में फजी तकनीकों का उपयोग करके एक नई पूर्वानुमान विधि विकसित की गई है। पश्चिम बंगाल के छ जिलों की वार्षिक वर्षा के आँकड़ों का पूर्वानुमान लगाने के लिए प्रस्तावित विधि को आनुभाविक रूप से चित्रित किया गया है। इस अध्ययन के परिणाम, क्लासिकल अविभाज्य चैन विधि की तुलना में प्रस्तावित दृष्टिकोण की श्रेष्ठता को प्रकट करते हैं।

संतुलित द्विपक्षीय उत्पन्न पंक्ति-स्तंभ अभिकल्पना

अनिन्दिता दत्ता¹, सीमा जग्गी¹, सिनी वर्गीज¹
एवं एल्डो वर्गीज²

¹भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

²भा.कृ.अ.प.-केन्द्रीय समुद्री मात्स्यिकी अनुसंधान संस्थान, कोच्चि

यह लेख सामान्यीकृत पंक्ति स्तंभ अभिकल्पनाओं से संबंधित है, जब उपचार के दो सेट होते हैं, एक सेट जिसमें

परीक्षण उपचार सम्मिलित होते हैं और नियंत्रण उपचार के अन्य जिन्हें बारपाटाईट सामान्यीकृत रो-कॉलम अभिकल्पना कहा जाता है। इस मायने में दो सेट असंबद्ध है कि दोनों के बीच आमतौर पर कोई प्रक्रिया नहीं है। यहाँ परीक्षण उपचारों में संबंधित विरोधाभासों का अनुमान लगाना है। जितना संभव हो सके उतना उच्च परिशुद्धता के साथ उपचार को नियंत्रित करना। बाइपाटाईट की श्रृंखला सामान्यीकृत रो-कॉलम अभिकल्पना में परीक्षण उपचार के एक सेट की तुलना नियंत्रण उपचार के एक सेट से की गई है। ये अभिकल्पनाएँ यह सुनिश्चित करते हैं कि परीक्षण बनाम नियंत्रण से संबंधित सभी विरोधाभासों का परीक्षण बनाम परीक्षण उपचार से संबंधित तुलना में कम विचरण के साथ अनुमान लगाया गया है।

पपीते में बाह्य आंकड़ों के होते हुए रोबस्ट एनोवा सिद्धांत का अनुप्रयोग

आर. वेणुगोपालन एवं बी.एल. मंजुनाथ

भा.कृ.अ.प.-भारतीय बागवानी अनुसंधान संस्थान, बैंगलूरु

पॉल एवं भर (2011) द्वारा डिजाइन किए गए प्रायोगिक डेटा में आउटलियर से निपटने के मुद्दे को संबोधित करने के लिए एम-आकलन विधियों के उपयोग की सिफारिश की गई है। वास्तविक समय बागवानी बारहमासी फसल प्रयोगात्मक डेटा का उपयोग करते हुए नियमित एनोवा विधि पर इस पद्धति की प्रभावकारिता को स्पष्ट करने का प्रयास किया गया है। परिणामों ने मजबूत तरीकों की प्रभावकारिता को मजबूत किया है जबकि सभी पात्रों के लिए स्वीकार्य प्रायिकता मूल्यों के साथ युग्मित वर्गों की त्रुटि राशि में तीन से पांच गुणा कमी के रूप में आउटलियर्स के साथ व्यवहार करते हुए एक प्रयास किया गया है। इस प्रकार यह अध्ययन भविष्य में अनुसंधान में बारहमासी बागवानी फसल प्रयोगों में आउटलियर से निपटने के दौरान मजबूत एनोवा दृष्टिकोण को अपनाने के लिए कहता है।

द्वि आदेश ऑर्थोगोनल लैटिन हाइपरक्यूब अभिकल्पना पर एक नोट

सुकांत दाश, बैद्यनाथ मंडल, राजेन्द्र प्रसाद एवं सुशील
कुमार सरकार

भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

लैटिन हाइपरक्यूब अभिकल्पना आमतौर पर कंप्यूटर प्रयोगों को डिजाइन करने में उपयोग किया जाता है। हाल के वर्षों में, साहित्य में ऑर्थोगोनल लैटिन हाइपरक्यूब डिजाइन के निर्माण के कई तरीके प्रस्तावित किए गए हैं। इस लेख में, छः कारकों के लिए ऑर्थोगोनल लैटिन हाइपरक्यूब डिजाइनों की दो नई श्रृंखलाएँ दी गई हैं।

ओडिशा में ग्रामीण निर्धनता को प्रभावित करने वाले कारकों पर एक अध्ययन

बिज्ञान नंद मोहन्ती¹, ए.के.पी.सी. स्वेन¹ एवं फिरोज खान²

¹अर्थ एवं सांख्यिकी निदेशालय, ओडिशा, भुवनेश्वर

²उत्कल विश्वविद्यालय, भुवनेश्वर

इस अध्ययन का उद्देश्य उपभोक्ता व्यय पर राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (एन एस एस ओ) द्वारा एकत्रित घरेलू स्तर के 68वें राउंड के प्राथमिक डेटा (2011-12) का उपयोग करके ग्रामीण ओडिशा में निर्धनता को प्रभावित करने वाले सबसे महत्वपूर्ण कारकों की पहचान करना है। इस अध्ययन में एक लॉजिस्टिक रिग्रेशन एनालिसिस किया गया है, जहाँ डिपेंडेंट वेरिएबल एक डाइकोटोमस वेरिएबल (Y) है, जिसे अधिकारिक गरीबी रेखा से नीचे के घर के लिए Y=1 के रूप में कोडित किया गया है और Y=0 गरीबी रेखा से ऊपर के घरों के लिए और तेरह स्पष्टीकरण चर आयु और उम्र के हैं, घर के मुखिया का लिंग, शिक्षा, घर के मुखिया की वैवाहिक स्थिति, परिवार का आकार, घर में कम से कम एक वेतन कमाने वाला, घर का सामाजिक समूह, घर की कुल भूमि, घर की कुल खेती योग्य भूमि घरों में आवास इकाइयों की संख्या, प्रकाश व्यवस्था और खाना पकाने के लिए ऊर्जा का स्रोत और खाद्य पदार्थों के तहत प्रति व्यक्ति मासिक खर्च का

प्रतिशत। यह पाया गया कि आयु, वैवाहिक स्थिति और घर के मुखिया की सामान्य शिक्षा का स्तर, परिवार का आकार, घर में कम से कम एक वेतन कमाने वाला, घर का सामाजिक समूह प्रकाश व्यवस्था के लिए ऊर्जा का स्रोत और मासिक प्रति व्यक्तित व्यय का प्रतिशत ग्रामीण ओडिशा में गरीबी को प्रभावित करने वाले खाद्य पदार्थ महत्वपूर्ण भविष्यवक्ता हैं।

फसल उपज अनुमान के लिए पदानुक्रमित बेज कुल स्तर पर स्पेशियल प्रतिमान

प्रियंका अंजोय एवं हुकुम चन्द्र

भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

प्रतिदर्श सर्वेक्षणों से स्वीकार्य असमान स्तर के आँकड़ों की मांग पिछले दशकों में सार्वजनिक और निजी क्षेत्रों में उनके व्यापक और विविध उपयोग के कारण काफी बढ़ी है। मूल रूप से यह छोटे प्रतिदर्शों के आकार की समस्या का जवाब देने के लिए छोटे डोमेन के लिए एक लक्ष्य सांख्यिकीय की ध्वनि भविष्यवाणी का उत्पादन करने के लिए लघु क्षेत्र आकलन (एस ए ई) के दृष्टिकोण का मुख्य प्रयास है। प्रतिदर्श आकार की समस्या के कारण असहमति या छोटे डोमेन स्तर अनुमान उत्पन्न करने के लिए पारंपरिक सर्वेक्षण अनुमान दृष्टिकोण पर्याप्त उपयुक्त नहीं हैं। लघु क्षेत्र आकलन तकनीक इसलिए मौजूदा सर्वेक्षण डेटा से अलग स्तर पर विश्वसनीय अनुमान लगाने के लिए एक व्यवहार्य तरीका प्रदान करती है। यह लेख लघु क्षेत्र आकलन के लिए जिले से धान (हरा) की फसल का अनुमान लगाने के लिए एक स्थानिक आश्रित एग््रीगेटड स्तर श्रेणीबद्ध बेज (एच बी) प्रतिमान की खोज करना है। भारत में उत्तर प्रदेश राज्य का स्तर। यह दृष्टिकोण राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय द्वारा एकत्रित फसल आँकड़ों (आई सी एस) योजना में सुधार के सर्वेक्षण डेटा का उपयोग करता है और जनगणना के साथ जुड़ा हुआ है। समय स्तर के छोटे क्षेत्र के प्रतिमान में स्थानिक जानकारी का दोहन करते हुए काफी लाभ प्राप्त किया गया है।

अपूर्ण बहु-प्रतिक्रिया प्रयोगों में आउटलियर्स मार्स्किंग की उपस्थिति

राजू कुमार एवं एल.एम. भर

भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

अपूर्ण बहु-प्रतिक्रिया अभिकल्पना के लिए मार्स्किंग की उपस्थिति में आउटलेयर्स के सब सेट की पहचान करने की एक विधि को विकसित किया गया है। अभिकल्पना प्रयोगात्मक इकाइयों के दो सेट से बना है। इन दो सेटों से विभिन्न प्रकार के प्रतिक्रिया चर देखे जाते हैं। अपूर्ण बहु-प्रतिक्रिया प्रयोगों के लिए ब्लॉक डिजाइन में एक सशर्त कुल के आँकड़े मार्स्किंग की उपस्थिति में आउटलेयर की पहचान के लिए विकसित किए गए हैं। विकसित सांख्यिकी को वास्तविक जीवन डेटा सेट के साथ चित्रित किया गया है। यह पदर्शित किया गया है कि मार्स्किंग की उपस्थिति में आउटलेयर एक प्रयोग से समग्र निष्कर्ष को विकृत कर सकते हैं।

खाद्य पदार्थों के लिए एसिड बैक्टीरिया-स्टार्टर कल्चर का विकास

सुनीता सिंह¹, संगीता गुप्ता¹, सुकांत दाश²
एवं के.एन. सिंह²

¹भा.कृ.अ.प.-भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली

²भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

लैक्टिक एसिड बैक्टीरिया, लैक्टोकोकसैलैक्टिस, खाद्य किण्वन में स्टार्टर कल्चर्स के रूप में उपयोग किया जाता है। इस तरह की शुरुआत के लिए विकास जानकारी का अनुमान लगाया जा सकता है जो कि लैक्टिक बैक्टीरिया के लिए आवश्यक किण्वन के शोषण में व्यावहारिक रूप से मूल्यवान हो सकता है। ये बैक्टीरिया विभिन्न चयापचयों का उत्पादन करने के लिए उत्तरदायी हैं। रूचि की चयापचयों को इनकी शुरुआत से ज्ञात विकास स्थितियों के एक सेट के तहत उत्पादित किया जाता है। उनके विकास को चयनित गणितीय कार्यों का उपयोग करके मॉडलिंग की जा सकती है। इन कार्यों का उपयोग परिभाषित पर्यावर्णीय परिस्थितियों में विशिष्ट विकास दर और जीव के अंतराल समय जैसे मापदंडों को इस अध्ययन निर्धारित करने में किया जा सकता है।

इस अध्ययन में प्रयुक्त किए गए तीन कार्यों (गोम्पर्टज, लॉजिस्टिक और रिचर्ड्स फंक्शन) में से गोम्पर्टज फंक्शन का चयन किया गया था। मॉडल ने स्थिरांक दिए जो कि इसके विकास के लिए जैविक मापदंडों को प्राप्त करने के लिए उपयोग किए गए थे। इस प्रकार मुख्य उद्देश्य एक विशेष कार्य और फिट की अच्छाई (आर 2 मूल्यों से) का पता लगाना था जो कि फिटेड ग्रोथ-डेटा को मापा गया था। डर्बिन-वाट्सन परीक्षण का उपयोग ऑटोकोरिलेशन द्वारा अवशिष्ट निर्भरता का परीक्षण करने के लिए किया गया था। एक बड़ी संख्या के ऑफर्स मान सुझाव देने के मापदंड में या तो लॉजिस्टिक या गोम्पर्टज फंक्शन (3 पैरामीटर फंक्शन) पर प्लॉट किए गए डेटा अधिक सहायक थे गोम्पर्टज फंक्शन को रिचर्ड्स फंक्शन (एक चार पैरामीटर फंक्शन) के विरुद्ध लॉजिस्टिक फंक्शन (दोनों जागृत पैरामीटर फंक्शन) से बेहतर पाया गया जो कि गोम्पर्टज और लॉजिस्टिक फंक्शन के परीक्षण के लिए स्किनेस की चर डिग्री थी। चयनित फंक्शन गोम्पर्टज फंक्शन, तब व्युत्पन्न और जैविक मापदंडों की गणना की गई थी जो कि निरंतर मूल्यों से प्राप्त की गई थी। इस प्रकार समय के संबंध में गतिज डेटा को सरल समीकरणों के संदर्भ में जैविक मापदंडों की गणना करने के लिए एक आसान तरीके से हल किया गया था। इस पद्धति को रूचि के विभिन्न चयापचयों का उत्पादन करने वाले लैक्टिक एसिड बैक्टीरिया के लिए बढ़ाया जा सकता है।

C4.5 निर्णय वृक्ष एल्गोरिदम का उपयोग करते हुए ऑनलाइन वर्गीकरण व विजुअलाइनेशन

शशि दहिया, सुवजीत दास एवं अंशु भारद्वाज

भा.कृ.अ.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

वर्गीकरण कई पूर्व निर्धारित श्रेणियों में से एक को वस्तुओं को निर्दिष्ट करने का डेटा खनन कार्य है। यह एक पूर्वानुमानात्मक मॉडलिंग कार्य है जिसमें एक मॉडल को व्याख्यात्मक चर के एक समारोह के रूप में लक्ष्य चर के लिए बनाया गया है। प्रशिक्षण पर्यवेक्षकों के पास पूर्वनिर्धारित लेबल वाली कक्षाओं के रिकार्ड होने के बाद से इसे पर्यवेक्षित शिक्षण भी कहा जाता है। ये लेबल प्रशिक्षण रिकार्ड वर्गीकरण मॉडल की शिक्षा की निगरानी

करते हैं। विभिन्न श्रेणी के लेवल असतत मूल्यों के द्वारा दर्शाए जा सकते हैं जहाँ मूल्यों के बीच क्रम का कोई अर्थ नहीं है। वर्गीकरण के लिए कई अच्छी तरह से स्थापित तकनीकें हैं, जिनमें से निर्णय वृक्ष तकनीक मशीन लर्निंग डोमेन से एक बहुत ही महत्वपूर्ण और लोकप्रिय तकनीक है। C4.5 एक प्रसिद्ध निर्णय वृक्ष एल्गोरिद्म है जो डेटासेट को वर्गीकृत करने के लिए उपयोग किया जाता है जो सभी डेटा माइनिंग सॉफ्टवेयर में उपलब्ध है। यद्यपि यह निर्णय वृक्षों को प्रेरित करने और डेटासेट से नियमों को सटीक रूप से उत्पन्न करने के लिए एक महत्वपूर्ण एल्गोरिद्म है,

इसलिए इसका उपयोग डेटा माइनिंग और मशीन लर्निंग समुदाय द्वारा किया जाता है। किसी भी डेटा माइनिंग सॉफ्टवेयर को स्थापित किए बिना अपने डेटासेट पर एल्गोरिद्म को लागू करने के लिए उपयोगकर्ताओं को एक ऑनलाइन मंच प्रदान करने के लिए C4.5 एल्गोरिद्म का उपयोग करके नियम निर्माण और निर्णय वृक्ष इंडक्शन के लिए एक वेब आधारित सॉफ्टवेयर विकसित किया गया है। वृक्ष संरचना के रूप में दृश्य उत्पन्न नियमों की समझ को बढ़ाता है।