석 사 학 위 논 문

의사결정나무와 로지스틱 회귀분석을 이용한 태권도 수련생 이탈 예측을 위한 비교 연구

Comparative Analysis of Prediction Taekwondo Trainee`s Defection using Decision Tree and Logistic Regression.

구 유 희

한 양 대 학 교 일 반 대 학 원

2007년 8월

석 사 학 위 논 문

의사결정나무와 로지스틱 회귀분석을 이용한 태권도 수련생 이탈 예측을 위한 비교 연구

Comparative Analysis of Prediction Taekwondo Trainee`s Defection using Decision Tree and Logistic Regression.

지도교수 권 태 원

이 논문을 체육학 석사학위논문으로 제출합니다.

2007년 8월

한양대학교 일반대학원
생활스포츠학과
구 유 희

이 논문을 구유회의 석사학위 논문으로 인준함.

2007년 8월

심사위원장 남 행 웅 (인)

심사위원 최 인 애 (인)

심사위원 권 태 원 (인)

한 양 대 학 교 대 학 원

감사의 글

이제 이 작은 결실을 맺으며 대학원 석사과정을 마감하게 되었습니다. 작고 뜻깊은 결실에 부끄러움이 앞서지만, 지금에 이르기까지 도움을 주신 모든 분들께 면을 통해서나마 감사의 마음을 전하고자 합니다.

그동안 대학원 생활을 함에 있어서 여러 가지 부족함이 많았던 저에게 많은 배려와 자상한 가르침을 주신 권태원 지도교수님께 진심으로 감사드리며, 항상 관심과 격려로 보살펴주시고 이끌어주신 남행웅 교수님과 최인애 교수님께도 깊은 감사의 말씀드립니다. 그리고 어렵게만 느껴졌던 대학원 생활에 아낌없는 격려와 조언으로 힘이 되어주신 한양대학교 생활체육과학대학 교수님들께도 머리 숙여 감사드립니다. 또한 힘들고 어려울 때 항상 곁에서 도와주시고, 위로해주신 박승현 선생님께 깊은 감사드립니다. 그리고 본 논문을 위해 어려움을 접고 여러모로 많은 도움을 준 이호열 선생님, 박호철 선생님, 손준호 선생님, 김태완 선생님, 이병관 선생님, 김동화 선생님, 김인산 선생님, 최성락 선생님께 진심으로 감사드리며, 조교생활 하는데 있어 가장 큰 힘이 되었던 상현이형, 대필이형, 상걸이형, 동석이형, 쌍기형, 승용, 희규, 상훈, 종현, 동진, 동학, 신일, 계환, 상인, 진협, 지원 모두에게 감사드립니다. 그리고 우리 교학과의 박경란 과장님과 김선영 선생님께 감사드립니다. 이분들을 수식하는 몇 마디의 단어보다 감사하다는 말로 대신하고 싶고, 이분들에 대한 저의 사랑과 고마움을 평생 마음속에 간직하고 소중히 생각하겠습니다.

그리고 지금까지 많은 어려움에도 학업에 정진할 수 있도록 격려와 사랑을 주신 아버님, 어머니께 깊은 감사드리며 사랑하는 큰이모, 누나들과 매형들께도 감사드립니다. 그리고 제가 사랑하는 여자친구 정희와 이 기쁨을 오랫동안 함께 나누겠습니다. 마지막으로 이 논문의 자료에 도움을 주신 인천체고동문님들께도 깊은 감사드립니다.

2007. 8

구 유 희

국 문 초 록

의사결정나무와 로지스틱 회귀분석을 이용한 태권도 수련생 이탈 예측을 위한 비교 연구

한양대학교 일반대학원
생활스포츠학과
구 유 회

본 연구는 로지스틱 회귀분석과 데이터마이닝 기법 가운데 의사결정나무 기법을 통한 태권도장의 이탈 수련생 예측에 가장 적합한 예측모형을 제시하는데 목적이 있다. 이와 같은 연구의 목적을 달성하기 위하여 경기도 및 인천광역시 소재의 태권도장 수련생을 모집단으로 설정하고 비확률 표본 추출법 중 편의추출법을 사용하여 수련생에게 1500부를 배포하였다. 회수된 설문지 가운데 신뢰성이 없다고 판단되는 자료를 분석대상에서 제외한 1,149부를 분석자료로 사용하였다. 설문지를 회수한 후 3개월이 경과한 뒤 등록유무를 파악하여 이월수련생 938명과 이탈수련생 211명으로 파악하였다.

자료처리는 SPSSwin 13.0 프로그램을 이용하여 교차분석, 신뢰도검증, 요인분석, 카이제곱 검증, t-검증, 의사결정나무 기법, 로지스틱회귀분석을 사용하였다.

이러한 절차와 방법을 통하여 도출된 본 연구의 결과는 다음과 같다.

1. 의사결정나무 기법과 로지스틱 회귀분석간 특이도의 차이를 알아본 결과 실제 이월자를 이월자로 예측하는 특이도에서는 로지스틱 회귀분석은 92.9%, 의사결정나무 기법이 96.3%로 다소 높게 나타났다.

2. 의사결정나무 기법과 로지스틱 회귀분석간 민감도의 차이를 알아본 결과

실제 이탈자를 이탈자로 예측하는 민감도에서는 로지스틱 회귀분석이 64.0%, 의사결정나무 기법은 44.5%로 나타났다.

3. 의사결정나무 기법과 로지스틱 회귀분석간 정확도의 차이를 알아본 결과 전체적인 분류정확도에서 의사결정나무기법이 86.8%, 로지스틱 회귀분석은 87.6%로 나타났다.

4. 의사결정나무 기법과 로지스틱 회귀분석간의 태권도장 수련생 이탈에 영향을 미치는 변인을 알아본 결과 의사결정나무 기법은 수련기간, 학년, 지도자만족, 추천의사, 시설만족도 변인으로 나타났으며, 로지스틱 회귀분석은 수련기간, 학년, 추천의사, 지도자만족도, 성별변인으로 나타났다.

이상의 결과를 종합해보면 의사결정나무 기법과 로지스틱 회귀분석간 태권도장 수련생 이탈 예측에 대한 비교 분석결과 두 모형 모두 분류정확도에서 높은 예측률을 나타내었고, 두 분석모형간 예측률의 차이는 나타나지 않았다. 이러한 결과는 이탈에 영향을 미치는 변인이 동일하게 사용되었기 때문이라 사료된다.

따라서 태권도장의 일선 지도자는 본 연구에서 밝히고 있는 예측모형과 예측률을 실제 수련생 관리와 경영전략수립에 이용함으로써 수련생 이탈 방지 및 효율적 경영에 도움을 줄 수 있을 것이다.

목 차

표 목차

그림 목차

Ⅰ. 서 론

1. 연구의 필요성

현대인의 삶에 대한 가치의식이 다양화되고 질적인 삶을 추구하려는 욕구가 증대되고 있다. 이러한 시대적 변화 속에서 스포츠의 정치, 사회, 문화적인 역할과 그 중요성은 높아지고 있으며 스포츠산업도 날로 번창하고 있다. 따라서 스포츠는 현대 사회의 핵심적이고 보편적인 사회현상의 하나로 자리 잡아 가고 있다. 또한 스포츠 수요의 급속한 증가는 대중적 여가활동과 대중소비를 향해 가는 사회적 추세를 반영하고 스포츠 산업이 더욱 팽창하리라는 예측을 가능하게 한다. 최근 한국도 스포츠관련 시장의 성장이 급속하게 진행되고 있다. 이는 국내 체육 시설에 관한 법률의 개정과 고도성장에 의해 안정기를 맞고 있는 기업이 늘고 있고 이러한 기업의 경영다각화 전략으로서 그동안 미개척 상태에 머물렀던 스포츠 레져 분야에 진출하기 시작한 것과 더불어 대중의 스포츠와 건강에 대한 관심증대는 스포츠 산업의 현격한 변화와 함께 각종 스포츠 관련 상업시설은 급격한 증가 경향을 나타내고 있다(박진기, 2001). 특히, 올림픽에서의 효자 종목으로써 국민들에게 커다란 관심을 받아온 태권도는 많은 사람들의 관심 증가에 따라 배우는 사람의 수가 늘어나게 되었다. 현재 우리나라 등록 태권도장수가 약 7,000여개에 달하는 등 급격한 증가로 인하여 태권도장 간의 경쟁은 더욱 치열해진 상태이다. 이러한 가운데 준비되지 못한 지도자, 비계획적인 경영, 부실한 교육 프로그램 등으로 인해 교육의 질과 지도자 자질의 향상이 뒤따르지 못함으로써 더 이상 수련생들의 욕구를 충족시킬 수 없는 상황이 되었다(김태형, 이동호, 황관식, 2004; 석종우, 이선장, 2001; 이승재, 조광민, 1998). 이로 인한 태권도장의 일선 지도자들은 수련생들의 이탈에 대한 공통적인 문제점을 갖고 있으며, 이탈방지 및 지속적인 참여에 대한 방안을 강구하고 있다. 즉, 수련생의 일반적인 특성과 환경요인적인 만족도를 파악하여 이탈 가능성이 있는 수련생들을 관리 할 수 있는 방안을 모색함으로써 수련생의 이탈을 방지하여 지속적인 참여를 유지하는 것에 중요

성을 두고 모든 노력을 경주하고 있다고 할 수 있다(조옥성외 2인, 2005). 이에 따라 데이터베이스 (Database)나 고객관계관리(Customer Relationship Management)등의 중요성이 크게 부각되기 시작하면서 고객의 데이터를 이용하여 고객 유형을 분석하고 이를 활용하려는 움직임 또한 활발해지고 있다. 그러나 기업의 데이터베이스 내에 존재하는 관련 데이터는 연속형(Continuous) 데이터보다 범주형(Categorical)데이터가 많기 때문에 이러한 데이터들을 기존의 통계적 기법으로 분류, 해석, 예측하는데 있어 한계가 있기 때문에 최근에는 데이터마이닝(Datamining)에 대한 관심이 증대되고 있다.

데이터마이닝이란 의사결정을 위해 대용량의 데이터로부터 이전에 알려지지 않은 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 함으로써 유용한 지식을 추출하는 반복적인 프로세스이다. 데이터 마이닝은 1983년 IMB Almaden 연구소에서 Rakesh Agrawal 박사를 중심으로 Quest 데이터 마이닝 프로젝터가 시작된 이후로 선진국의 유수 연구소와 대학원을 중심으로 활발하게 연구가 되어왔다(심규석, 2001). 특히 1990년부터 Pos시스템(Point of sales system), CATV(cable TV), 홈쇼핑(home shopping), 인터넷(internet)등이 급속히 확산중인 온라인 시스템(online system) 환경 하에서, 컴퓨터 정보통신의 비약적인 발전으로 고객들에 관한 데이터를 쉽사리 그리고 저비용으로 축적할 수 있게 되고 축적된 고객데이터를 빠르게 처리할 수 있게 됨에 따라서 그 중요성이 역시 급속히 증가중이다. 우리나라 기업들의 경우, 데이터베이스 마케팅의 활용측면에서는 전반적으로 선진국 기업들의 그것에 비해 뒤쳐져있기는 하지만 기존고객과의 지속적인 관계유지를 통해서 기업의 이익을 증가시킬 가능성이 상대적으로 높은 분야들인 은행, 항공회사, 카드회사, 통신회사 등 서비스 기업들을 중심으로 데이터베이스 마케팅을 적극적으로 활용하는 시도를 하기 시작했으며 타 업종으로도 그런 시도가 점차 확산되고 있다.

데이터 마이닝을 활용하기 위해 데이터베이스에 적용하는 데이터마이닝 기법에는 인공신경망(artificial neural networks), 의사결정나무(Decision Tree), 군집분석(cluster analysis)등이 있다. 이러한 데이터마이닝 기법들(data mining techniques) 가운데서 의사결정나무(Decision Tree)는 예측(forecasting)과 분류(classification)에서 다른 데이터마이닝 기법들에 비해 유용성이 많은 것으로 평가되고 있다. 반면 기존의 전통적인 통계기법들 가운데

서 이와 같은 통계학적, 방법론적 문제점들을 덜 내포하고 있다고 평가되는 것은 로지스틱 회귀분석 기법(logistic regression analysis technique)이다.

로지스틱 회귀분석은 독립변인들의 정규분포를 가정하는 회귀분석이나 판별분석과는 달리, 독립변인들에 대해 어떠한 가정을 하지 않음으로써 독립변수들의 분포(distribution)에 상관없이 사용할 수 있다는 장점을 가지고 있다(최대호, 1999).

데이터 마이닝을 활용한 이탈 예측에 대한 선행연구를 살펴보면 데이터마이닝을 이용한 보험회사 고객이탈분석에 관한 연구 이현정(2001), 데이터 마이닝 기법을 활용한 스포츠 센터고객 이탈가능성 예측모형 개발 박진기(2001), 데이터 마이닝을 이용한 인터넷 쇼핑몰의 이탈 고객 분석 모형에 관한 연구 서해옥(2001), 데이터 마이닝 기법을 이용한 전공 이탈학생 분석 및 예측모형 개발 이지영(2002), 데이터 마이닝을 이용한 이탈학생의 예측모형개발 옥기현(2003), 데이터 마이닝 기법을 이용한 고객 이탈 분석 및 예측모형 개발 윤해원(2004), 통계적 기법과 데이터마이닝 기법을 이용한 이동통신 VAS 가망고객 scoring 모형 비교연구 정해원(2004), 데이터 마이닝을 이용한 제약회사 이탈고객 분석 예측에 관한 연구 신선영(2004), 고객이탈 예측 데이터 마이닝 기법 비교연구 배행수(2005)등이 이탈 예측을 위한 연구로 수행되어왔다. 하지만 태권도장에 대한 선행연구를 살펴보면 태권도 체육관에서 수련생 이탈원인에 관한 연구 이종천(2001), 초등학교 태권도 수련생들의 도장 만족도와 중도포기 원인의 연구가 진행되어 왔으나 데이터 마이닝을 활용한 이탈 수련생 예측에 대한 연구는 이루어지지 않고 있다. 데이터 마이닝 분석기법 중에 통계학적, 방법론적 문제점을 덜 내포하고 있는 의사결정나무 기법과 예측력과 분류력이 가장 뛰어난 것으로 밝혀진 로지스틱 회귀분석 기법과의 예측력과 정확도를 비교·분석하고 예측 및 분류도구로서의 유용성을 평가해보는 것도 가치 있는 연구라고 판단된다.

따라서 본 연구는 태권도장의 수련생의 이탈 원인에 대해 알아보고 적합한 예측모형의 기초자료를 태권도장의 일선 지도자들에게 제시함으로써 태권도장 수련생 이탈방지에 유용하게 활용될 것으로 기대된다.

2. 연구의 목적

본 연구의 목적은 아래와 같이 두 가지로 정한다.

첫째, 태권도장의 중요한 당면 문제점 가운데 하나인 수련생이탈에 대한 예측력을 평가하기 위하여 로지스틱 회귀분석과 데이터마이닝 기법 가운데 의사결정나무 기법을 적용하여 가장 우수한 예측력을 가진 기법을 비교·분석하는데 목적을 둔다.

둘째, 로지스틱 회귀분석과 의사결정나무 기법을 적용해서 파악된 이탈성향이 높은 수련생의 특성과 그 원인을 설명해 주는 변인들을 비교·분석함으로서 이탈현상의 원인을 규명하는데 목적을 둔다.

3. 연구 가설

본 연구는 데이터 마이닝의 기법 중 예측모형인 의사결정나무 기법과 로지스틱 회귀분석을 비교분석을 통해 적합한 수련생 이탈 예측모형을 찾는데 있다. 본 연구에서 설정한 관련 변수와 연구 모형을 기초로 하여 다음과 같은 구체적인 영가설을 설정하고 이를 검증하고자 한다.

가설 Ⅰ. 의사결정나무 분석과 로지스틱 회귀분석간의 정분류에는 차이가 없을 것이다.

Ⅰ-1 의사결정나무 분석과 로지스틱 회귀분석간의 특이도에는 차이가 없을 것이다.

Ⅰ-2 의사결정나무 분석과 로지스틱 회귀분석간의 민감도에는 차이가 없을 것이다.

Ⅰ-3 의사결정나무 분석과 로지스틱 회귀분석간의 정확도에는 차이가 없을 것이다.

가설 Ⅱ. 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인에는 차이가 없을 것이다.

Ⅱ-1 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 성별변인은 차이가 없을 것이다.

Ⅱ-2 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 학년변인은 차이가 없을 것이다.

Ⅱ-3 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 수련기간 변인은 차이가 없을 것이다.

Ⅱ-4 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 결석횟수 변인은 차이가 없을 것이다.

Ⅱ-5 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 학원참여 수 변인은 차이가 없을 것이다.

Ⅱ-6 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 추천의사 변인은 차이가 없을 것이다.

Ⅱ-7 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 부모맞벌이 변인은 차이가 없을 것이다.

Ⅱ-8 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 프로그램 만족도 변인은 차이가 없을 것이다.

Ⅱ-9 의사결정나무분석과 로지스틱회귀분석간의 이탈에 영향을 미치는 변인으로 시설 만족도 변인은 차이가 없을 것이다.

Ⅱ-10 의사결정나무 분석과 로지스틱 회귀분석간의 이탈에 영향을 미치는 변인으로 지도자 만족도 변인은 차이가 없을 것이다.

4. 연구의 제한점

본 연구에서는 연구대상과 구성요소에서 나타나는 문제로 다음과 같은 제한점을 둔다.

첫째, 본 연구의 대상은 경기도 및 인천에 소재한 태권도장 초등학생 수련

생으로 제한하였기 때문에 그 결과를 일반화하기에는 한계가 있다.

둘째, 태권도장 수련생 이탈에 영향을 미칠 수 있는 또 다른 요인인(학부모의 의견)이 포함되지 않았으므로 해석으로 확대하기에는 한계가 있다.

5. 용어의 정의

본 연구의 내용 이해를 돕기 위해 다음과 같은 용어들에 대한 정의를 제공하고자 한다.

1) 데이터 마이닝(data mining)

대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 함으로써 유용한 지식을 추출하는 일련의 과정.

2) 이탈고객 분석(churn analysis)

데이터 마이닝 분류 작업에 주로 사용되는 기법으로, 의사결정규칙을 나무구조로 도표화하여 분류와 예측을 수행하는 분석방법이다. 순환적 분할 방식을 이용하여 나무를 구축하며 나무의 가장 상단에 위치하여 뿌리마디, 속성의 분리기분을 포함하는 내부마디, 마디와 마디를 이어주는 가지 그리고 최종분류를 의미하는 잎으로 구성된다. 분류도는 예측의 과정이 나무구조에 의한 추론규칙에 의해서 표현되기 때문에 분석자가 그 과정을 쉽게 이해하고 설명할 수 있다는 장점이 있다.

3) 데이터(data)

정보의 원재료로서 사람이나 기계가 잘 다를 수 있거나 통신, 번역 등을 살 수 있게 규격화하여 표현한 사실이나 개념.

4) 의사결정나무(dicision tree)

데이터 마이닝 분류 작업에 주로 사용되는 기법으로, 의사결정규칙을 나무구조로 도표화하여 분류와 예측을 수행하는 분석방법이다. 순환적 분할 방식을 이용하여 나무를 구축하며, 나무의 가장 상단에 위치하여 뿌리마디, 속성의 분리기준을 포함하는 내부마디, 마디와 마디를 이어주는 가지 그리고 최종분류를 의미하는 잎으로 구성된다. 분류 또는 예측의 과정이 나무구조에 의한 추론규칙에 의해서 표현되기 때문에 분석자가 그 과정을 쉽게 이해하고 설명 할 수 있다는 장점이 있다.

5) 데이터 분할(data split)

데이터 마이닝 과정 중 엄정한 모형평가를 위하여 데이터를 구축용, 검정용, 시험용으로 분리시키는 작업.

6) 스트림(stream)

여러 개의 노드들이 각종 변환과 데이터 탐색 그리고 모형화를 위해 노드들을 연결을 하게 되는 상태의 위상을 가리킨다. 영어의 원 뜻이 물줄기의 흐름인 것처럼 데이터의 흐름이라 볼 수 있다. 이 흐름에 따라 어떤 부분에서는 데이터의 탐색을 다른 부분에서는 변환과 조성을 그리고 뒷부분에서는 데이터의 모형화까지 흐름도에 따라 나타난다.

7) 마디(node)

의사결정나무나 신경망의 구성요소로 부모마디(parent node), 자식마디(child node), 종단마디(terminal node)로 나눌 수 있다. 의사결정나무의 구조에서 하위마디들을 거느리는 상위마디를 부모마디, 나무에서 하나의 부모마디에 종속되어 배치된 한개 이상의 마디들을 자식마디, 의사결정나무 구조의 최종마디로서 최종분류와 사후 확률을 포함한 잎을 종단마디라 한다.

8) 가지치기(pruning)

의사결정나무 모형 구축 시 과잉맞춤으로 인해 불필요하게 복잡해진 나무의 의미 없는 마디(속성)들을 제거하는 작업. 분류오류(classification error)를 크

게 할 위험(risk)이 높거나 부적절할 추론규칙(induction rule)을 가지고 있는 가지(branch)를 제거한다.

9) 다중분리(multiway split)

나무 구축 시 각 마디를 두 개 이상의 하위마디로 분리하는 것. 부모마디에서 자식마디들이 생성될 때 2개 이상의분리가 일어나는 것을 허용함을 의미한다.

10) 로지스틱 회귀(Logistic Regression)

한 개의 종속변인과 여러 개의 독립변수간의 상호관련성에 대해 분석하려 할 때 가장 널리 사용되는 통계적 방법이 회귀분석법이며, 로지스틱 회귀분석법은 범주형 종속변인과 여러 가지 독립변인과의 관계를 알아보고자 할 때 사용되는 분석기법이다.

11) 정분류(Classification)

목표변인이 얼마나 정확하게 분류되었는지 빈도로 나타낸다.

12) 특이도(specificity)

종속변수 y의 값이 “0”인 개체들 중에서 “0”으로 맞게 예측하는 경우의 비율로 정의한다.

13) 민감도(sensitivity)

종속변수 y의 값이 “1”인 개체들 중에서 “1”로 맞게 예측하는 경우의 비율로 정의한다.

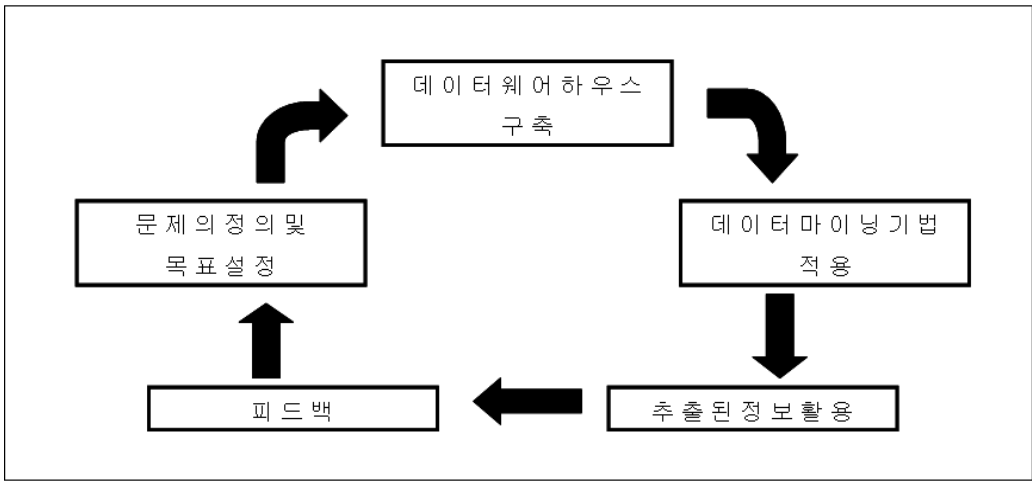
Ⅱ. 이론적 배경

예측 도구로서의 데이터 마이닝 기법과 통계적 기법에 관한 문헌 연구

1. 데이터 마이닝의 정의 및 특징

급속도로 증가하고 있는 데이터로부터 유용한 정보를 찾아내고자 하는 비즈니스 주체들의 니즈(needs)와 방대한 데이터 베이스를 처리할 수 있는 컴퓨터 처리 속도의 향상으로 가능해진 데이터 마이닝(data mining)은 마이닝(mining)이라는 단어의 뜻에서 알 수 있듯이 다양한 자료들의 집합인 데이터 베이스에서 일정한 규칙성이나 패턴(a certain regularities or pattern)등 의사결정에 도움을 줄 수 있는 정보를 밝혀내는 프로세스(process)로 정의되며 데이터로부터 유용한 정보나 지식을 추출하기 때문에 지식개발(knowledge discovery)과도 비슷한 의미로 사용된다.

위에 제시한 정의에서도 알 수 있듯이 데이터 마이닝은 그 대표적인 인공신경망(artificial neural networks)이나 의사결정나무(decision tree)와 같은 특정 기법이 아니라 비즈니스의 문제와 목적에 맞는 데이터를 준비하는 것으로부터 데이터 마이닝 기법가운데 적절한 기법을 사용하여 정보를 추출하고 이를 바탕으로 전략을 수립하거나 의사결정을 내리는 하나의 프로세스(process)를 의미한다. 데이터마이닝의 프로세스를 그림으로 나타내면 [그림 1]과 같다. 즉, 데이터마이닝은 ① 해결하는 비즈니스 문제의 정의 및 달성하고자 하는 목표의 결정, ② 데이터 웨어하우스(data warehouse)의 구축, ③ 데이터 마이닝 기법의 적용, ④ 추출된 정보의 의사결정에 활용, ⑤ 피드백(feedback)에 이르는 하나의 프로세스로 이해해야 하며 데이터 마이닝을 수행하기 위해서는 전 과정에 대한 준비와 계획이 필요하다.



[그림 1] 데이터 마이닝 프로세스

자료원 : Berry,Michael J. A,. and Gorden Linoff(1998), (Data Mining Techniques:For Marketing, Sales, and Customer Support), NY: John Wiley & Sons, p.23

1) 데이터 마이닝과 데이터 웨어하우스

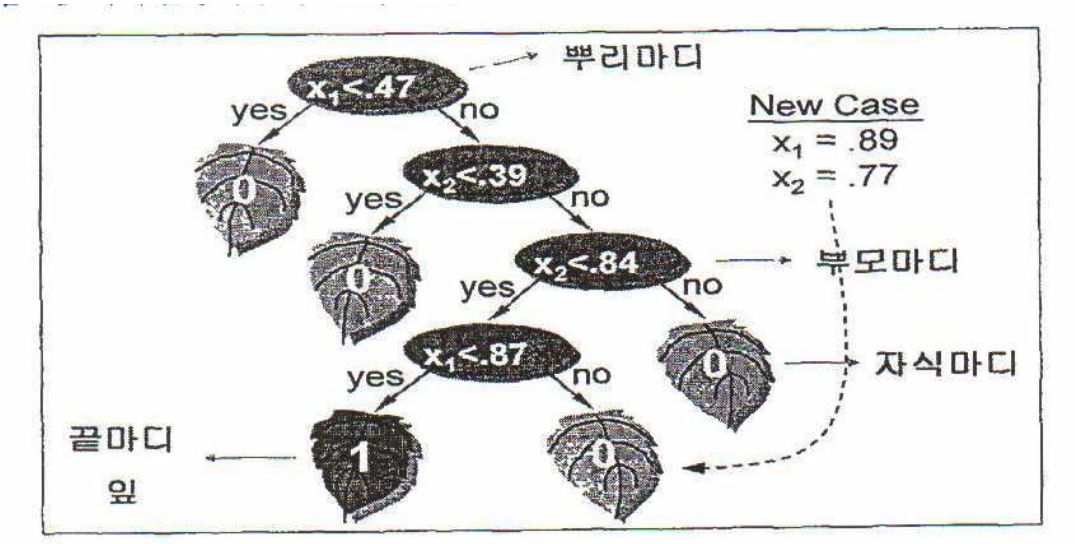
데이터의 단순집합이 아니라 통합(integration)과 공용(sharing)을 특성으로 하는 데이터베이스(database)에 대하여 McFadden and Hoffer(1991)는 조직의 여러 가지 다양한 정보욕구에 대응하기 위해 설계된 서로 연관된 자료의 집합으로 정의하였고 서길수(1995)에 의하면 여러 사용자가 자신의 정보욕구를 충족시키기 위해 사용할 수 있도록 서로 관련 있는 데이터를 최소한의 중복으로 통합해 놓은 데이터의 집합체로 정의하고 있다.

이처럼 데이터베이스는 단순한 자료의 집합이 아니라 서로 유기적으로 관련을 맺고 있는 자료의 집합이다. 또한 데이터베이스는 목적 지향적으로 설계되어서 조직의 요구에 충분히 부응할 수 있어야 하며 한 걸음 더 나아가 필요할 때에 필요한 자료를 추출하고 가공하는 것이 가능할 수 있도록 다양한 방법의 분석 및 응용의 지원이 가능할 수 있도록 설계되어야 한다.

데이터 웨어하우스는 위에서 정의한 데이터베이스 보다 포괄적이고도 다양한 형태의 데이터베이스들로 이루어져 커다란 데이터의 집합이며 의사결정지원에 적극적으로 활용하기 위해서 축적하고 관리하는 하나의 시스템이다. 기업의 운영시스템(operation system)에서 발생된 수년간의 내부데이터(internal data)와 외부데이터(external data)를 주제 중심적으로(subject-oriented) 즉시분석(on-line analysis)을 가능케 하는 통합데이터베이스 시스템으로 정의될 수 있는 데이터웨어하우스는 데이터마이닝을 성공적으로 수행하기 위한 하나의 필요조건이다. 왜냐하면 오류가 있는 데이터에는 아무리 훌륭한 방법을 적용한다 해도 얻어진 결과를 신뢰할 수 없기 때문이다. 마찬가지로 데이터마이닝 기법들의 퍼포먼스가 우수하다고 할지라도 정확하고 올바른 결과를 얻기 위해서는 적용대상이 되는 데이터웨어하우스가 정제되고 체계적인 구조로 이루어져 있어야 한다.

2. 의사결정나무의 개념

표본 집단을 특정 기준 값에 의해 유사한 집단으로 분류하고, 분류된 하위집단을 다시 특정 기준을 찾아 분류하는 과정을 반복함으로써 종속변인과 독립변인들 또는 목표변수와 입력변인들 간의 패턴이나 관계를 찾아내는 분석방법인 의사결정나무기법(decision tree technique)은 종속변인에 가장 큰 영향을 주는 독립변인의 특정 값을 기준으로 표본 집단에 대한 최초 분리가 이루어지며 순차적(sequential)으로 더 이상의 분리가 이루어지지 않을 때까지 분리를 수행한다. 의사결정나무 기법은 다양한 알고리듬에 의해 분리가 이루어지고 [그림 2]와 나무구조로 표현되며 나무구조가 시작되는 뿌리마디(root node), 하나의 마디로부터 분리되어 나간 두 개 이상의 마디들인 자식마디(child node), 자식마디의 상위마디인 부모마디(parent node), 각 나무줄기의 끝에 위치하고 있는 끝마디(terminal node) 등 여러 가지의 마디(node)라고 불리는 구성요소들로 이루어져 있다.



[그림 2] 의사결정나무 분리구조

김은철(1998), "기업도산예측모형에 관한 비교연구," 서강대학교 경영학과 석사학위 논문, p.18.

1) 의사결정나무(decision tree)분석

의사결정나무는 의사결정규칙(decision rule)을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측(prediction)을 수행하는 분석방법이다. 분석과정이 나무구조에 의해서 표현되기 때문에, 신경망(neural networks), 판별분석(discriminant analysis), 회귀분석과 같은 방법들에 비해, 연구자가 분석과정을 쉽게 이해하고 설명할 수 있다는 장점을 가지고 있다. 의사결정나무는 분류 또는 예측을 목적으로 하는 어떤 경우에도 사용될 수 있으나 분석의 정확도보다는 분석과정의 설명이 필요한 경우에 더 유용하게 사용된다. 의사결정나무 분석이 유용하게 활용되는 응용분야는 다음과 같다.

· 세분화(segmentation) : 데이터를 비슷한 특성을 갖는 몇 개의 그룹으로

분할하여 각 그룹별 특성을 발견하는 경우.

· 분류(classification) : 관측개체(observation)를 여러 예측변인들에 근거하여 목표변인(target variable)의 범주를 몇 개의 등급으로 분류하고자 하는 경우.

· 예측(prediction) : 자료로부터 규칙을 찾아내고 이를 이용하여 미래의 사건을 예측하고자 하는 경우.

· 자료축소 및 변인선택(data reduction and variable screening) : 여러 개의 예측변인들이 결합하여 목표변수에 작용하는 교호작용을 파악하고자 하는 경우.

· 교호작용효과와 파악(interaction effect identification) : 여러 개의 예측변인들이 결합하여 목표변인에 작용하는 교호작용을 파악하고자 하는 경우.

· 범주의 병합 또는 연속형 변인의 이산화(category merging and discretizing continuous variable) : 범주형 목표변인의 범주를 소수의 몇 개로 병합하거나 연속형 목표변인을 몇 개의 등급으로 이산화 하고자 하는 경우.

일반적으로 의사결정나무 분석은 다음과 같은 단계를 거친다.

· 의사결정나무의 형성 : 분석의 목적과 자료구조에 따라서 적절한 분리기준(split criterion)과 정지규칙(stopping rule)을 지정하여 의사결정나무를 얻는다.

· 가지치기 : 오분류(classification error)를 크게 할 위험(risk)이 높거나 부적절한 추론규칙을 가지고 있는 가지(branch)를 제거한다.

· 타당성평가 : 이익도표(gains chart)나 위험도표(risk chart)는 검정용 자료(test data)에 의한 교차타당성(cross validation)등을 이용하여 의사결정나무를 평가한다.

· 해석 및 예측 : 의사결정나무를 해석하고 분류 및 예측모형을 설정한다.

이상과 같은 정지기준, 분리기준, 평가기준 등을 어떻게 지정하느냐에 따라서 서로 다른 의사결정나무가 형성된다.

2) 의사결정나무의 구성

의사결정 나무는 하나의 나무구조를 이루고 있으며 마디(node)라고 불리는 구성요소들로 이루어져 있다. 뿌리마디에서 어느 한 끝마디 (leaf node)로 가는 길은 유일하며 그 길은 규칙으로 표현가능하며 서로 다른 끝마디들이 같은 분류(classification)를 만들 수도 있지만 각 끝마디는 서로 다른 근거에 의해 그 분류를 만든다.

<표 1> 의사결정나무 노드 설명

분 류	설 명
뿌리마디 (root node)	나무구조가 시작되는 마디로써 전체 자료로 이루어져 있다.
자식마디 (child node)	하나의 마디로부터 분리되어 나간 두개 이상의 마디들을 의미한다.
부모마디 (parent node)	자식마디의 상위마디를 의미한다.
중간마디 (terminal node)	각 나무줄기의 끝에 위치하고 있는 마디로써 잎(leaf)이라고도 하며, 결국 끝 마디의 개수만큼 분류규칙이 생성되는 것이다.
가지 (branch)	하나의 마디로부터 끝마디까지 연결된 일련의 마디들을 의미하여, 이때 가지를 이루고 있는 마디의 개수를 깊이 (depth)라고 한다. 가지 수와 깊이는 다양하다.

3) 의사결정나무의 분리 기준

분리기준(split criterion)이란 하나의 부모마디로부터 자식마디들이 형성될 때 입력변인의 선택과 기준 값을 의미한다. 즉, 어떤 입력변인을 사용하고 또 그 변인의 어떤 값을 기준으로 분리하는 것이 목표변인의 분포를 가장 잘 구별하는지를 나타내는 기준으로 다음과 같은 몇 가지 알고리듬에 의해서 분리

기준이 정해진다.

(1) CHAID 알고리듬

CHAID(Chi-squared Automatic Interaction Detection) 알고리듬은 목표변인의 비율이 깨지는가를 계산하고 목표변인의 비율이 특정변인 값에 의해서 분리가 이루어지는 경우 어느 정도로 비율이 깨지는가를 계산하고 목표변인의 비율이 유지되는 확률이 가장 적은 변인을 기준으로 분리가 일어나게 하는 알고리듬이다.

(2) CART 알고리듬

마디(node)의 순수함(purity)을 나타내는 지니계수(Gini index)에 의해 분리여부를 결정하는 CART(Classification And Regression Tree) 알고리듬은 특성 변인에 의해 집단을 구분하고 구분된 하나의 집단에서 나머지 집단의 개체가 선택될 확률을 계산하여 집단을 분리하는 알고리듬이다. 즉, 집단이 순수할수록 지니계수의 값은 작아지며 확률 또한 작아지게 되는 것이다. 지니계수의 값에 따라 변인이 선택되는 알고리듬이다.

(3) C4.5 알고리듬

지니계수와 마찬가지로 마디의 순수함을 측정하는 엔트로피 지수(Entropy index)에 의해 분리되는 알고리듬으로 집단이 순수할수록 엔트로피 지수의 값은 작아지게 되고 엔트로피 지수의 값이 가장 작은 변인을 기준으로 분리가 이루어지는 알고리듬이다.

(4) QUEST 알고리듬

QUEST(Quick Unbiased Efficient Statistical Tree)는 C4.5와 마찬가지로 명목형 목표변수에 대해서만 분석을 수행할 수 있다. 분리방법은 예측 변인의 척도에 따라서 서로 다른 분리기준을 사용하여 이진분리를 하는데 분리변인의 선택과 선택된 분리변인에서 분리점의 선택으로 나누어 실행된다. 분리변인의 선택은 예측변인가 순서형 또는 연속형인 경우에는 ANOVA F-검정 또는 Levene 의 검정을 사용하며, 예측변인이 명목형인 경우에는 Pearson의 카이제

곱 검정을 사용하여 가장 작은 유의확률에 대응되는 변수를 분리변인으로 선택한다. 다음의 <표 2>에서는 의사결정나무 분석의 알고리즘을 비교하였다.

<표 2> 의사결정나무분석의 알고리즘의 비교

구 분	CHAID	C5.0	CART	QUEST
입력변수에 의한 분리형식	다중분할	다중분할	이진분할	이진분할
연속형 목표변인	처리불가	처리가능	처리가능	처리가능
나무구조 생성 시 오분류 비용 사용여부	사용 안함	사용	사용	사용 안함
결측치 처리	하나의 범주로 처리	하나의 범주로 처리	결측치 대체	결측치 대체
사전확률 사용여부	사용 안함	사용 안함	사용	사용

4) 의사결정나무 기법의 장점 및 한계점

의사결정나무 기법의 장점들을 다음과 같다. 첫째, 분류 또는 예측의 과정이 나무구조에 의해서 표현되기 때문에 분석자가 그 과정을 쉽게 이해하고 설명할 수 있다는 점이다. 즉, 특정 데이터의 개체가 어떤 집단에 속하게 되는지를 나무구조 만에 의해서도 시각적으로 파악되기 때문에 해석이 편리하다는 장점을 가지고 있다. 둘째, 변인들 간의 상호작용효과(interaction effects)를 파악할 수 있다는 장점을 가지고 있다. 그것은 회귀분석이나 판별분석과 같은 모수적 모형(parametric model) 모형에서 변수들 간의 가능한 모든 상호장용을 고려하기가 아주 어렵다는 점을 감안할 때 매우 큰 장점일 수 있다.

한편 의사결정나무 기법은 단점들도 지니고 있다. 첫째, 의사결정나무 기법은 연속형 변인(continuous variable)를 비연속적인 값으로 취급하기 때문에 분리의 경계점 부근에서는 오류가 발생할 확률이 높다는 단점을 지니고 있다. 예를 들어 월 소득 200만원을 기준으로 분리가 이루어졌다고 하자. 월 소득이 200만원보다 아주 낮거나 높은 경우에는 신뢰할 수 있지만 199만원이나 201만원인 경우에는 오류가 발생할 확률이 아주 높아지게 되는 것이다. 둘째, 주효

과(main effect)를 얻을 수 없다는 점이다. 회귀분석과 판별분석의 경우에는 독립변인의 종속변인에 미치는 효과를 다른 독립변인과 관련 없이 주효과를 밝혀낼 수 있지만 의사결정나무의 경우에는 변인간의 상호작용효과만을 보기 때문에 한계가 있다.

3. 로지스틱 회귀 분석 (Logistic Regression Analysis)

1) 로지스틱회귀분석의 개념

일반적으로 회귀분석(regression analysis)은 아래의 식으로 요약되는데,

$$[식1] \quad Y_i = \beta_0 + \sum \beta_k X_{ik} + \varepsilon_i$$

이때 Y_1 는 종속변수, $X_{1k}(k=1, \ldots, K)$는 독립변수, βk는 회귀계수(β0는 상수항)이며 ε_1는 오차항으로 기본 가정은 $E(\varepsilon 1) = 0$, $Var(\varepsilon 1) = \delta^2$ 그리고 $Cov(\varepsilon 1, \varepsilon_j)=0$이다($i \neq j$). 회귀분석에서는 X_{1k}, βk, ε1가 취할 수 있는 값에 대하여 제한이 없으므로 따라서 종속변수 Y_1의 값도 $-\infty$에서$+\infty$사이의 어떤 값도 자유로이 취할 수 있다. 그러나 만일 Y1가 질적 변수로서 단지 두 개의 값(예컨대 0과 1)만을 취한다고 하고 $E(\varepsilon 1)=0$ 이라고 하면 종속변수 Y_1 의 기대 값 $E(Y_1)$ 는 다음과 같이 표현 될 수 있다.

$$[식2] \quad E(Y_1)=1 \cdot Prob(Y_1=1) +0 \cdot Prob(Y_i=0)=Prob(Y_i=1)$$

$$따라서, \ Prob(Y_1=1) = E(Y_1) = \beta_0 + \sum \beta_K X_{ik}$$

즉, 종속변수 Y의 기재치 $E(Y_1)$는 Y_1가1이 될 확률로 요약되며 [식2]는 이 확률이 독립변수와 선형적인 관계를 가지므로 선형확률모델(linear probabillity model: LPM)이라 하는데 [식2]를 회귀분석기법에 그대로 적용하

는 데에는 문제가 있다. 왜냐하면 $E(Y_1)=P(Y_i=1)$는 확률로 0과1사이의 값을 취하지만 $\beta_0 + \sum\beta_K X_{ik}$의 값은 제한이 없으므로 확률$P(Y_i=1)$의 값이 음수가 되거나 1보다 큰 경우가 발생할 수 있기 때문이다. 따라서 확률 $P(Y_i=1)$의 값이 선형(linear)이 아닌 비선형(nonlinear)의 관계를 가정하는 것이 논리적으로 타당하고 이러한 문제를 해결하기 위해서는 확률 $P(Y_i=1)$의 값이 $-\infty$에서 $+\infty$사이의 값을 갖도록 변환되어야 한다.

두 확률의 비율에 자연로그를 취함으로써 변환하며 변환된 [식3] 을 로짓모델(logit model)또는 로지스틱 회귀분석(logistic regression)이라 한다.

$$[식3] \quad P_1 = \frac{\exp(\beta X_1)}{1+\exp(\beta X_1)}$$

종속 변인이 이분형 척도로 측정된 경우에는 로지스틱 회귀 분석을 사용하여 독립 변인과 종속 변인간의 관계를 분석할 수 있다(이학식, 김영, 2001). 로지스틱 회귀 분석은 어떤 사건이 발생할 확률을 예측한다. 따라서 종속 값은 0과 1사이의 값을 갖는다. 분석 결과 종속 변인 값, 즉 확률이 0.5보다 크면 그 사건이 일어나면 0.5보다 작으면 그 사건이 일어나지 않는 것으로 예측하게 된다. 독립 변인과 종속 변인의 로지스틱 관계는 S자형으로 가정한다. 로지스틱 회귀 분석은 로짓 분석 (Logit Analysis)이라고도 불린다. 종속 변인은 명목척도로 나타내고 독립 변인은 명목 척도, 간격 척도, 비율척도로 측정된다. 독립 변인이 명목 척도로 측정된 경우 일반적 회귀 분석에서처럼 더미 변수로 변경하여 입력한다.

로지스틱 회귀 모형의 가정을 요약하면 다음과 같다(성웅현 , 2001).

첫째, 로지스틱 회귀 모형은 종속 변인과 독립 변인들 사이의 함수 관계를 선형으로 가정하지 않고 비선형 관계로 가정한다.

둘째, 종속 변인의 분포는 정규 분포에 따를 필요가 없다.

셋째, 종속 변인은 주어진 독립 변인의 수준에서 동일한 분산을 가질 필요가 없다.

넷째, 오차항의 정규성이 가정되지 않는다.

다섯째, 로지스틱 회귀 모형에서 독립 변인들의 척도는 연속형, 이산형, 혹은 혼합형으로 구성될 수 있다.

여섯째, 로지스틱 회귀 모형에서는 최우추정법을 사용하기 때문에 추정량의 신뢰성을 확보하기 위해서 표본 크기가 어느 정도 대표본이 되어야 한다.

2) 로지스틱 회귀분석의 장점 및 한계점

로지스틱 회귀분석(logistic regression)은 종속변인과 독립변인들의 관계를 비선형적으로 표현하기 때문에 소비자의 선택(customer choice)과 같이 비선형적인 관계를 보이고 있는 현상들의 예측이나 분석에 접합한 분석방법이다. 또한 로지스틱 회귀분석은 독립변인들의 정규분포를 가정하는 회귀분석이나 판별분석과 달리 독립변인들의 연속성 여부에 상관없이 종속변수가 질적 변수인 경우에 사용된다는 장점을 가지고 있다. 그러나 로지스틱 회귀분석은 일반적인 회귀분석 기법이 가지고 있는 단점인 독립변인들 간의 상호작용효과(interaction effect)와 독립변인의 수에 대한 한계를 극복하지 못하고 있다는 문제점을 지니고 있다. 즉, 변인들 간의 상관관계가 높은 경우, 그 효과를 반영하지 못한다는 단점과 독립변인들의 수가 증가함에 따라 설명력이 계속 증가한다는 점이다.

4. 의사결정나무 기법과 로지스틱 회귀분석 비교

지금까지 살펴본 바와 같이 의사결정나무 기법과 로지스틱 회귀분석 기법들의 주요 특징들을 정리하면 다음과 같다.

의사결정나무 기법은 분석이 이루어지는 과정을 나무구조를 통하여 쉽게 이해할 수 있으며 변인들 간의 상호작용효과를 파악할 수 있다는 장점을 가지고 있다. 그러나 연속형 변인을 비연속적인 값으로 취급하여 분리점 근처의 자료들의 값에 대해서는 잘못 분류할 오류 가능성이 높으며 독립변인들의 종속변인에 대한 주효과(main effect)를 파악할 수 없다는 단점을 가지고 있다. 로지스틱 회귀분석(logistic regression analysis)의 경우는 독립변인과 종속변인의 관계가 비선형인 경우에 적합하며 독립변인들의 연속성 여부에 상관없이 분석

을 할 수 있다는 장점을 가지고 있으나 독립변인들 간의 상관관계가 높은 경우 상호작용 효과(interaction effect)를 보여줄 수 없으며 독립변인의 수가 많아질수록 경우 설명력이 높아진다는 단점을 가지고 있다.

5. 모형의 평가

- 모형평가의 개념 -

모형을 개발하기 위한 방법에는 여러 가지 방법이 사용될 수 있다 .이러한 방법들은 데이터의 성격에 따라 어느 것이 반드시 우수하다고 결론지을 수 없다. 예를 들어 카드회사와 보험회사에서 평점 모형을 개발한다고 할 때 카드회사에서 사용되는 변수의 성격과 보험회사에서 사용되는 변수의 성격이 다르기 때문이다. 따라서 추출된 하나의 고객 데이터로부터 최적의 신용평점 모형을 개발할 때에는 가능한 여러 가지 통계모형을 상정하여 분석하는 것이 바람직하다. 또한 최적의 모형을 얻기 위해서는 구축된 모형들을 비교, 평가해야 하고 이를 통해 선택되어진 모형이 다른 모형에 비해 우수하다는 것을 입증할 수 있어야 한다. 이러한 일련의 과정을 모형평가(model assesment)라고 한다. 즉, 모형평가란 예측을 위해 만든 모형이 임의의 모형보다 과연 우수한지, 고려된 서로 다른 모형들 중 어느 것이 가장 우수한 예측력을 보유하고 있는지를 비교분석하는 과정이라 말할 수 있다(강현철 외, 2001).

6. 정분류(Classification)

정분류란 목표변수의 실제 범주와 모형에 의해 예측된 분류범수 사이의 관계를 나타내는 것이라 할 수 있다. 즉, 목표변수의 범주별로 이를 제대로 분류한 빈도를 나타낸다. 이는 목표변수의 범주가 C 개인 경우 C×C개의 셀로 이루어진 형식을 취한다. 이 때 분류의 대각에 존재하는 셀은 목표변수의 원래 범주가 제대로 분류된 부분이고 비대각에 존재하는 셀은 분류가 잘못된 부분

으로 구분되어 각각의 빈도가 제시되어야한다.

7. 타 분야의 고객이탈에 대한 연구

최근 들어 국내의 기업들은 경쟁우위 확보 전략의 일환으로 데이터 웨어하우스를 구축하였거나 구축 중에 있으며 이로써 고객 이탈 방지를 위하여 각종 데이터를 효율적으로 이용할 수 있게 되었다. 이탈 고객 및 유지 고객의 유형을 세분화하고 이들의 특징을 분석하는 연구는 주로 통신 산업에서 기존의 통신 수단을 해지하고 다른 업체의 통신 수단으로 이동하는 고객들의 특징을 분석하기 위하여 연구되었다. 또한 은행 계좌의 해지 예측, 항공사의 마일리지 부여, 병원의 병실 요금 할인 프로그램 등 다양한 산업에서도 고객 이탈에 대한 연구가 수행되고 있다. 이탈 고객 분석 모형은 고객 반응 정보 중 고객들의 자사 제품에 대한 지속적 사용여부에 초점을 두어 고객의 이탈 확률을 계산하여 고객 세분화를 시도한 것이다. 따라서 고객 이탈의 원인을 파악하고 이를 이용하여 이탈 고객을 최소화시키는 것이 이탈 고객 분석 모형의 의의라고 할 수 있다(이건창 등, 2001). 개별 고객이 이탈할 확률을 추정하여 예측 확률 스코어를 구하고 이탈 확률이 높은 고객을 선별하여 그 고객층을 대상으로 데이터베이스 프로그램을 실행함으로써 고객 이탈률을 줄이는 것이 고객 이탈 분석 모형의 주목적이다(최정환, 최종학, 2001). 이탈 고객을 관리하는 것은 일정 기간 동안의 이탈된 고객의 특성과 기존 고개의 특성을 분석하여 고객 이탈의 주요 변인을 찾아내고 이를 기존 고객에게 적용함으로써 고객별 이탈 가능 확률을 최소화하여 고객의 평생 고객화를 추구하기 위한 것이다(정지택, 1999). 이러한 고객 이탈 분석 모형은 고객 이탈에 영향을 주는 관련 변인 또는 통제 가능한 변인을 밝혀내는 동시에 개인별 이탈 확률을 계산하여 예측 이탈 확률이 높은 고객들을 찾고 그들의 특징을 알아내는 데 매우 유용한 모형이 된다. 일단 개인별 이탈 확률이 구해지면 상대적으로 이탈 확률이 높은 고객들을 군집화 하여 인구 통계학적 변인, 라이프스타일에 관한 변인, 그 외에 개인 특징 변인들로 분석하게 된다. 그래서 이처럼 이탈 확률이 높은 군집과 유사한 특징을 가진 고객들은 경쟁 업체로의 이탈이 생기기 쉬우므로 그러

한 고객들을 잘 선별하여 목표 고객으로 삼아 적절한 방어적 마케팅 프로그램을 개발하고 실행함으로써 이탈을 방지하거나 줄이도록 해야 한다. 이탈 고객 분석 모형은 이탈 자체에만 초점을 맞추는 것이 아니라 이탈 분석 결과를 이용하여 고객의 유치, 유지 탈환이라는 전체적인 관점에서 접근함으로써 가치 높은 고객에 상응하는 판촉을 추진하고, 새로운 고객을 유치하고 이탈률을 감소시키고 유통 채널을 분석할 수 있어야 한다(김영만, 1999). 고객 이탈률은 수익 변동의 선행 지표하고도 할 수 있다. 고객의 이탈 사유를 분석하여 수익이 감소되기 전에 예방함으로써 고객이 계속 거래를 하게 되고 시간이 지날수록 많은 수익을 실현시킬 수 있을 것이다. 그러므로 데이터베이스 마케팅에서는 고객 데이터베이스 안에 내장 되어 있는 구매 데이터와 통계 모형을 이용하여 개인별 이탈률을 예측하는 것뿐만 아니라 고객들의 이탈을 막기 위해 해야 할 것들을 찾으려 하는 것이다. 고객의 이탈을 방지하는 가장 효과적인 방법은 미리 고객의 동향을 분석하여 이탈하려는 고객의 여러 가지 패턴을 찾아내어 고객이 이탈을 고려하는 단계에서 적절한 대응책을 구사하는 것이라 할 수 있다.

8. 수련생 이탈

수련생의 태권도장에 대한 선택 폭이 넓어질수록 각도장들에 있어서 수련생 이탈을 방지하기 위한 중요성은 증가한다고 할 수 있다. 기존 수련생을 유지하는 것은 신규 수련생을 획득하는 것 보다 훨씬 더 중요하며 기존 수련생의 유지가 신규 수련생을 얻는 것에 비해 여러 가지 경제적인 장점이 있기 때문이다. 태권도장은 기존 수련생으로부터 수련생과의 지속적 유지 관계에 따른 미래의 이익을 얻을 수 있을 뿐만 아니라 수련생이 기간에 따른 자격증 및 용품을 구입함으로써 얻게 되는 추가 이익이 있으며 이와 함께 신규 수련생을 유입하는 것도 가능하기 때문에 이익을 향유할 수 있다.

기존에 태권도장들이 이탈 수련생을 분석하기 위해 사용되는 통계적 방법에 따른 문제점은 다음과 같다.

첫째, 수련생 또는 이탈에 대한 정의를 내리는 것이 어렵고 수련생 이탈의

원인을 결정하기는 더욱 어렵다. 수련생 이탈의 원인이 파악되고 수련생 이탈로부터 교훈을 얻었지만 경영자가 실행에 옮기기를 주저한다. 이러한 이유로 이탈 수련생 관리는 방치된다.

둘째, 이미 구축된 데이터베이스에서는 이탈한 수련생의 유형, 사업, 우편번호 등에 따라 단순히 정렬한다. 이러한 통계가 중요한 첫 번째 단계이긴 하지만 수련생이 이탈한 원인을 밝혀주지 못한다.

셋째, 이탈한 수련생과 인터뷰 조사를 위해 전문가를 이용한다. 이것은 여러 가지 이유 때문에 좋은 결과를 가져오지 못한다. 전문적인 인터뷰 조사자는 태권도장과 수련생의 관계를 분석하는데 필요한 친숙함이 부족하다. 외부 조사자에 의해 완성된 보고서에 대해 경영자는 이를 심각하게 받아들이지 않는다.

Ⅲ. 연구방법

1. 연구대상

본 연구는 경기도와 인천광역시 소재 20여 곳의 태권도장 수련생 4학년, 5학년, 6학년을 모집단으로 설정하였고 비확률 표본추출법 중 편의추출법에 의해서 총 1,500명을 연구대상으로 하였다. 연구대상자들로 하여금 자기평가기입법(selp-administration method)으로 설문지를 작성하게 하였다. 총 1,470명의 설문지를 회수하였으며 이 가운데 응답이 부실하거나 누락된 설문지를 제외한 총 1,149명의 유효표본을 얻었다. 또한 이탈고객 확보를 위하여 설문지를 회수한 후 3개월이 경과한 뒤 태권도장의 기록을 참조로 등록유무를 파악하여 이월고객 938명과 이탈고객 211명으로 분류하였다.

본 연구에서 연구대상자의 일반적 특성(general feature of interviewers)은 다음<표 3>와 같다.

<표 3> 연구대상자의 일반적 특성

구분	항 목	빈도(N)	백분율(%)	누적백분율(%)
성 별	남	926	80.6	80.6
	여	223	19.4	100.0
학년	4학년	466	40.6	40.6
	5학년	304	26.5	67.0
	6학년	379	33.0	100.0
수련기간	2년 이하	691	60.1	60.1
	2년 이상	458	39.9	100
이탈유무	이월	938	81.6	81.6
	이탈	211	18.4	100

2. 연구도구

본 연구의 태권도장 수련생 이탈에 미치는 요인을 측정하기 위하여 사용된 연구도구는 기존의 설문지를 바탕으로 설문지 초안을 작성하였다. 작성된 초안은 예비검사를 실시하여 설문내용의 신뢰도와 타당성을 검증한 후 최종도구로 사용하였다.

1) 설문지의 구성

인구통계학적에 관한 4문항, 이용형태에 관한 3문항, 만족도에 관한 20문항 <표 4>와 같이 구성하였다.

<표 4> 설문지의 구성

구성지표	구성내용	문항수
인구통계학적 특성	성별, 학년, 부모맞벌이, 학원참여수	4
이용형태	수련기간, 추천의사, 결석횟수	3
만족도	지도자, 프로그램, 시설	20

환경요인에 대한 만족도의 설문문항은 프로그램요인 8문항, 시설요인 8문항, 지도자요인 4문항 등으로 구성되었다. 문항들의 신뢰도를 검증하기 위한 방법으로 Cronbach's α를 산출하였다. <표 5>에서 나타난 것과 같이 태권도 도장의 환경요인 만족도에 대한 각 문항 간의 신뢰도는 α=.833~.870 사이를 보임으로써 연구도구의 신뢰도는 비교적 높은 것으로 평가할 수 있다.

<표 5> 연구도구의 신뢰도

요인	변인수	Cronbach's Alpha
프로그램	8	.860
시설	8	.870
지도자	4	.833

2) 타당도

태권도장의 수련생들이 환경요인들에 대한 만족도에 관련된 문항의 구성타당도(construct related validity) 검사를 위해 요인분석을 실시하였다<표 6>. 요인 추출 모델은 주성분분석(principal component)을 이용하여 직각회전방법 중의 배리맥스(varimax) 방법을 이용하였으며 요인의 고유치가 1이상인 요인들만 추출하였다. 환경요인에 대한 수련생들의 만족도는 3개 요인으로 추출되었는데 요인 1은 프로그램요인, 요인 2는 시설요인 그리고 요인 3은 지도자요인으로 명명하였다. 이들 요인들은 전체 분산의 55.5%를 설명하고 있다.

<표 6> 요인분석 결과표

요인	요인1	요인2	요인3
p1	.693		
p2	.680		
p3	.678		
p4	.675		
p5	.666		
p6	.659		
p7	.633		
p8	.601		
f1		.755	
f2		.709	
f3		.683	
f4		.675	
f5		.620	
f6		.602	
f7		.590	
f8		.535	
t1			.756
t2			.755
t3			.704
t4			.682
Eigen value	8.250	1.628	1.228
% of Variance	41.250	8.140	6.138
Cumulative %	41.250	49.390	55.528

-통계분석에 사용한 변인선정-

본 연구의 분석에 사용된 변인의 범주는 <표 7>과 같으며 그 구체적인 내용은 다음과 같다.

<표 7> 통계분석에 선정된 변인의 정의

	변수명	범 주
예측변인	성 별	· 남/여
	학 년	· 4학년/5학년/6학년
	수련기간	· 2년 이하/2년 이상
	추천의사	· 있음/없음
	시설 만족도	· 청결, 안전, 운동기구, 바닥, 냉·난방양호, 공기쾌적, 식수관리, 실내조명. · 평균(3.96±1.01)을 기준으로 평균 이하일 경우 0, 평균 이상일 경우 1.
	프로그램 만족도	· 지루하지 않음, 나이별지도, 수준별지도, 적당한 운동, 체력향상, 스트레스해소, 인성발달, 다양성. · 평균(4.41±.86)을 기준으로 평균 이하일 경우 0, 평균 이상일 경우 1.
	지도자 만족도	· 용모단정, 품성, 예절교육, 지도능력우수. · 평균(4.02±1.02)을 기준으로 평균 이하일 경우 0, 평균 이상일 경우 1.

3. 자료처리

본 연구의 목적은 태권도장 수련생들의 인구통계학적 특성 및 태권도장의 환경요인에 대한 만족도 등의 변인을 통하여 이탈 가능성을 예측하기 위한 것이다. 회수된 자료 내용의 신뢰성이 없다고 판단되는 자료를 분석대상에서 제외하고 SPSS version 13.0을 연구목적에 맞게 자료를 처리 하였으며 그 구체적인 분석방법은 다음과 같다.

첫째, 연구대상자의 일반적 특성에 따른 이탈유무를 알아보기 위해 교차분석을 실시하였다.

둘째, 측정 도구의 신뢰도를 알아보기 위하여 Cronbach's α 계수를 산출하였고 타당도를 알아보기 위하여 요인분석을 실시하였다.

셋째, 이탈에 영향을 미치는 변인의 유의성을 평가하기 위하여 종속변인과 독립변인이 모두 명목척도인 변인은 카이제곱(Chi-square)검정을 하였으며 환경요인의 만족도에 따른 집단 간의 평균 차이에 대한 검정은 t-검정을 실시하였다.

넷째, 이탈 가능성을 예측하기 위하여 로지스틱 회귀분석을 사용하였다.

다섯째, 이탈 가능성을 예측하기 위하여 의사결정나무 분석을 사용하였고 알고리즘으로는 CHAID(Chi-squared Automatic interaction Detection)을 이용하였다.

IV. 연구결과

본 연구는 태권도장 기존 수련생들을 대상으로 이탈 가능성을 예측하는데 목적이 있다. 이러한 목적을 달성하기 위하여 수련생의 인구통계학적 특성과 이용행태, 만족도와 관련된 여러 변인들을 독립변인으로 설정하고 이탈유무를 종속변인으로 선정하였다. 본 연구에 사용된 프로그램인 로지스틱 회귀분석과 데이터마이닝 기법 중 의사결정나무 분석을 실시한 결과 다음과 같은 결과를 얻었다.

1. 이탈 유무에 따른 만족도에 대한 *t*-검정 결과

입력변인 중 비율척도를 갖는 각 만족도에 대한 t-검정 결과는 <표 8>과 같다. 시설만족도에서 이월(3.93±1.03)과 이탈(4.08±.94)에서 유의하게 나타났다. 프로그램만족도에서 이월(4.03±1.02)과 이탈(3.95±1.05)로 유의하게 나타났다.

<표 8> 이탈유무에 따른 t-검정 결과표

	이탈유무	평균	표준편차	t	p
시설만족	이월	3.93	1.03	-3.678	.000***
	이탈	4.08	.94		
프로그램	이월	4.03	1.02	2.809	.005**
	이탈	3.95	1.05		
지도자만족	이월	4.40	.87	-.492	.623
	이탈	4.44	.81		

**p<0.01, ***p<0.001

2. 인구통계학적 변인과 이용행태에 대한 *Chi-square* 검정결과

입력변인 중 명목척도를 갖는 인구통계학적 변인과 이용행태에 대한

Chi-square 검정 결과는 <표 9>와 같다. 성별(χ^2=11.96), 학년(χ^2=172.068), 수련기간(χ^2=314.146), 추천의사(χ^2=4.921)에서 유의하게 나타난 네 개의 변인을 입력변인으로 선정하였다.

<표 9> 이탈유무에 따른 Chi-square 검정 결과표

	χ^2	p
성별	11.96	.001***
학년	172.068	.000***
수련기간	314.146	.000***
결석횟수	.112	.738
학원참여수	.000	.989
추천의사	4.921	.027*
부모맞벌이	3.170	.075

*p<0.05, **p<0.01, ***p<0.001,

1) 성별에 따른 이탈유무의 차이

분석에 사용된 총 1149명의 연구대상 가운데 남자는 926명(80.6%), 여자는 223명(19.4%)이며 성별과 이탈여부의 관계를 살펴보면 <표 10>와 같다. 남자는 926명중에 188명(89.1%), 여자는 223명중에 23명(10.9%)이 이탈로 나타났다.

<표 10> 성별에 따른 이탈유무

이탈 / 성별	이탈여부		합 계
	이월	이탈	
남자	738(78.7%)	188(89.1)	926(80.6%)
여자	200(21.3%)	23(10.9%)	223(19.4%)
합계	938(100%)	211(100%)	1149(100%)

2) 학년에 따른 이탈유무의 차이

표본집단은 4학년부터 6학년이며 4학년은 466명(40.6%), 5학년은 304명(26.5%), 6학년은 379명(33.0%)으로 나타났다. 학년에 따른 이탈여부를 살펴보면<표 11>과 같다. 4학년 466명 가운데 21명(10%)이 이탈로 나타났고 5학년은 304명 가운데 42명(19.9%) 그리고 6학년은 379명 가운데 148명(70.1%)가 이탈로 나타났다.

<표 11> 학년에 따른 이탈유무

이탈 / 학년	이탈여부		합 계
	이월	이탈	
4학년	445(47.4%)	21(10.0%)	466(40.6%)
5학년	262(27.9%)	42(19.9%)	304(26.5%)
6학년	231(24.6%)	148(70.1%)	379(33.0%)
합계	938(100%)	211(100%)	1149(100%)

3) 수련기간에 따른 이탈유무의 차이

수련기간은 2년 이하인 수련생 691명(60.1%)과 2년 이상인 수련생 458명(39.9%)으로 수련기간에 따른 이탈여부를 살펴보면 <표 12>과 같다. 2년 이하의 수련기간의 수련생 691명 가운데 13명(6.2%)이 이탈로 나타났고 2년 이상의 수련기간의 수련생 458명 가운데 198명(93.8%)이 이탈로 나타났다.

<표 12> 수련기간에 따른 이탈유무

이탈 / 수련기간	이탈여부		합 계
	이월	이탈	
2년 이하	678(72.3%)	13(6.2%)	691(60.1%)
2년 이상	260(27.7%)	198(93.8%)	458(39.9%)
합계	938(100%)	211(100%)	1149(100%)

4) 추천의사에 따른 이탈유무의 차이

추천의사는 주변사람에게 자신이 다니고 있는 태권도장을 추천할 의사여부를 알아본 것이며 그 특성은 <표 13>과 같다. 추천의사가 있는 경우 629명(54.7%)이며 추천의사가 없는 사람은 520명(45.3%)로 나타났다. 추천의사가 있을 경우 629명 가운데 130명(61.6%), 추천의사가 없을 경우 520명 가운데 81명 38.4%)가 이탈로 나타났다.

<표 13> 추천의사에 따른 이탈유무의 차이

이탈 / 추천의사	이탈여부		합 계
	이월	이탈	
있음	499(53.2%)	130(61.6%)	629(54.7%)
없음	439(46.8%)	81(38.4%)	520(45.3%)
합계	938(100%)	211(100%)	1149(100%)

3. 로지스틱 회귀분석 기법에 의한 예측결과

로지스틱 회귀분석에서는 분석 결과를 살펴보면 성별, 학년, 수련기간, 추천의사, 시설, 프로그램 지도자 만족도 변인이 입력변인으로 선택되었다. <표 14>은 로지스틱 회귀분석의 결과 요약표이다. 분석결과를 살펴보면 통계적으로 의미가 있는 변인들은 성별, 학년, 수련기간, 추천의사, 시설만족도 등 5가지로 나타났다.

성별(B값 :-.663)변인과 이탈과의 관계는 역(negative)의 관계가 있는 것으로 나타났다. 학년(B값:.991)의 변인은 4학년, 5학년, 6학년의 순차적인 단계로 이탈과 정(positive)의 관계를 보였다. 수련기간(B: 3.374)의 변인은 이탈과 정(positive)의 관계를 보였다. 추천의사(B: -.448)의 변인은 이탈과 역(negative)의 관계가 있는 것으로 나타났다. 시설만족도(B: .241)의 변인은 이탈과 정(positive)의 관계를 보였다.

<표 14> 로지스틱 회귀분석 결과 요약표

변수	B 값	S.E	Wald	df	Sig.	Exp(B)
성별	-.663	.284	5.441	1	.020	.515
학년	.991	.127	61.283	1	.000	2.693
수련기간	3.374	.303	124.197	1	.000	29.188
추천의사	-.447	.208	4.649	1	.031	.639
시설만족도	.241	.105	5.280	1	.022	1.273
프로그램 만족도	-.043	.098	.191	1	.662	.958
지도자 만족도	.001	.100	.000	1	.994	1.001
Constant	-5.016	.516	94.464	1	.000	.007

*p<0.05, **p<0.01, ***p<0.001,

1) 로지스틱 회귀분석의 정분류표

<표 15>은 로지스틱 회귀분석의 예측을 나타내는 정분류표이다. 표에서 볼 수 있듯이 로지스틱 회귀분석의 전체 분류정확도는 87.6%, 실제 이월수련생을 이월수련생으로 예측하는 특이도는 92.9%, 실제 이탈수련생을 정확히 이탈 수련생으로 예측하는 민감도는 64.0%를 나타냈다.

<표 15> 로지스틱 회귀분석의 정분류표

		예측값		합계	예측력	
		이월	이탈			
실제값	이월	871	67	938	특이도	92.9 %
	이탈	76	135	211	민감도	64.0 %
합계		947	202	1149	정확도	87.6 %

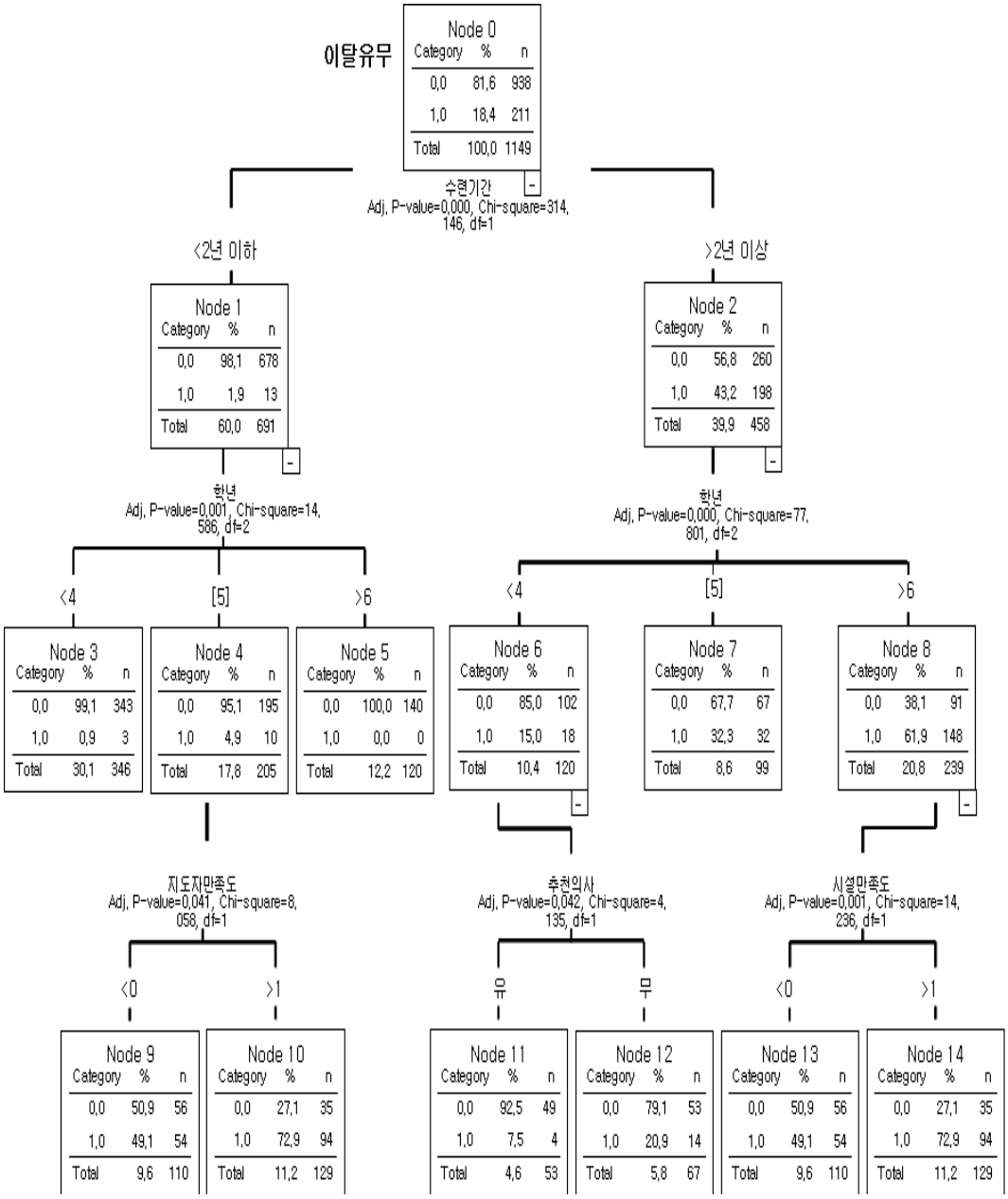
4. 의사결정나무 기법에 의한 예측결과

<표 16> 의사결정나무분석 모델 요약

■ 정지규칙 (stopping Rules) 분리정지(Maximum Tree Depth) : 3 부모마디의 관측수 제한 (Minimum Number of Case for Parent) : 100 자식마디의 관측수 제한 (Minimum Number of Case for Child) :50 분리기준의 유의수준(Significant Level for Splitting) :0.05 ■ 나무구조 결과 (Resulting Tree) 전체마디수 (Total Number of Nodes) : 14 전체수준수 (Total Number of Nodes) : 3 전체종료마디수 (Total Number of Terminal Nodes) : 9

<그림 3>은 의사결정나무 기법에 의해 도출된 분석결과이다. 의사결정나무 분석결과를 살펴보면 여덟 개의 끝마디로 이루어져 있다. 이탈에 가장 큰 영향을 주는 변인은 수련기간이며 그 다음으로 학년, 추천의사, 시설만족도 변인 순으로 나타났다. 분석결과를 구체적으로 살펴보면 다음과 같다. 전체 1149명의 검증집합에서 수련기간이 2년 이하 수련생들은 98.1%인 678명이 지속적인 참여를 하였으며 1.9%인 13명은 이탈한 것으로 나타났다. 수련기간이 2년 이상 수련생들은 56.8%인 260명이 지속적인참여를 하였으며 43.2%인 198명이 이탈한 것으로 나타났다. 수련기간이 2년 이하이고 학년이 4학년일 경우 343명(99.1%)이 이월로 나타났으며 3명(0.9%)이 이탈로 나타났다. 수련기간이 2년 이하이고 학년이 5학년일 경우 195명(95.1%)이 이월로 나타났고 이탈은 10명(4.9%)으로 나타났다. 수련기간이 2년 이하이고 학년이 6학년일 경우 140명(100%)전원이 이월로 나타났다. 수련기간이 2년 이하이고 5학년이며 지도자 만족도가 높을 경우 이월이 86명(90.5%)이 이월로 나타났고 9명(9.5%)이 이탈로 나타났다. 지도자 만족도가 낮을 경우 109명(99.1%)이 이월로 나타났고 1명(0.9%)이 이탈로 나타났다. 수련기간이 2년 이상이고 4학년일 경우 이월이 102명(85.0%)이고 이탈이 18명(15.0%)로 나타났다. 수련기간이 2년 이상이고

5학년일 경우 이월이 67명(67.7%)이고 이탈이 32명(32.3%)로 나타났다. 수련기간이 2년 이상이고 6학년일 경우 이월이 91명(38.1%)이고 이탈이 148명(61.9%)로 나타났다. 수련기간이 2년 이상이고 4학년이며 추천의사가 있을 경우 이월이 49명(92.5%)이고 이탈이 4명(7.5%)로 나타났다. 수련기간이 2년 이상이고 4학년이며 추천의사가 없을 경우 이월이 53명(79.1%)이고 이탈이 14명(20.9%)로 나타났다. 수련기간이 2년 이상이고 6학년이며 시설만족도가 낮은 경우 이월이 56명(50.9%)이고 이탈이 54명(49.1%)로 나타났다. 수련기간이 2년 이상이고 6학년이며 시설만족도가 높은 경우 이월이 35명(27.1%)이고 이탈이 94명(72.9%)로 나타났다.



<그림 3> 의사결정나무 결과 분류 결과표

1) 의사결정나무 기법의 예측을 나타내는 정분류표

<표 17>는 의사결정나무 기법의 예측을 나타내는 정분류표이다. 표에서 볼 수 있듯이 의사결정나무 분석의 전체 분류정확도는 86.8%, 실제 이월수련생을 이월수련생으로 예측하는 특이도는 96.3%, 실제 이탈수련생으로 예측하는 민감도는 44.5%를 나타냈다.

<표 17> 의사결정나무기법의 정분류표

		예측값		합계	예측력	
		이월	이탈			
실제값	이월	903	35	938	특이도	96.3%
	이탈	117	94	211	민감도	44.5%
합계		1020	129	1149	정확도	86.8%

5. 로지스틱 회귀분석, 의사결정나무 기법의 비교 및 평가

<표 18> 로지스틱 회귀분석, 의사결정나무 기법의 예측 비교분석

분석기법 / 예측력	로지스틱 회귀분석	의사결정나무 기법
특이도	92.9%	96.3%
민감도	64.0%	44.5%
정확도	87.6%	86.8%

로지스틱 회귀분석과 데이터마이닝 기법인 의사결정나무분석 기법의 예측 비교・평가해 보면<표 18> 전반적인 분류 정확도는 의사결정나무 기법(86.8%)과 로지스틱 회귀분석(87.6%)은 유사하게 나타났으며 실제 이탈자를 이탈자로 예측하는 정도인 민감도는 로지스틱 회귀분석 기법(64.0%)이 의사결정나무 기법(44.5%)보다 높게 나타났다 또한 실제 이월자를 이월자로 예측하는 정도인 특이도는 의사결정나무 기법(96.3%)이 로지스틱 회귀분석(92.9%)보다 약간의 높은 차이가 있는 것으로 나타났다. 전반적으로 의사결정나무 기법과 전통적인 예측기법으로 사용되었던 로지스틱 회귀분석 기법은 예측에 유의

한 차이가 없는 것으로 나타났다.

<표 19> 로지스틱 회귀분석, 의사결정나무 기법의 이탈에 영향을 미치는 변인의 비교분석

분석기법 / 변인	로지스틱 회귀분석	의사결정나무 기법
예측변인	수련기간, 학년, 추천의사, 시설 만족도, 성별	수련기간, 학년, 추천의사, 시설만족도, 지도자 만족도
변인 수	5	5

<표19>는 로지스틱 회귀분석에서 예측된 변인으로는 수련기간, 학년, 추천의사, 시설만족도, 성별 등 5가지 변인이 이탈에 영향을 미치는 것으로 나타났으며, 의사결정나무 기법에서 예측된 변인으로는 수련기간, 학년, 추천의사, 시설만족도, 지도자 만족도 등 5가지 변인이 이탈에 영향을 미치는 것으로 나타났다.

Ⅴ. 논 의

본 연구의 목적은 태권도장의 중요한 당면 문제점 가운데 하나인 수련생 이탈에 대한 예측력을 평가하기 위하여 로지스틱 회귀분석과 데이터마이닝 기법 가운데 의사결정나무 기법을 적용하여 가장 우수한 예측력을 가진 기법을 비교·분석하였다. 또한 로지스틱 회귀분석과 의사결정나무 기법을 적용해서 파악된 이탈성향이 높은 수련생의 특성과 그 원인을 설명해 주는 변인들을 비교·분석함으로서 이탈현상의 원인을 규명하는데 목적이 있다.
본 장에서는 앞에서 얻어진 연구결과를 기초로 선행연구를 이용하여 논의를 하고자 한다.

1. 의사결정나무분석과 로지스틱회귀분석간의 정분류 차이

‘가설 Ⅰ. 의사결정나무분석과 로지스틱회귀분석간의 정분류에는 차이가 없을 것이다.’를 정분류를 통해 검증한 결과 전반적인 분류 정확도는 의사결정나무 기법과 로지스틱 회귀분석에서 유사하게 나타났으며 실제 이탈자를 이탈자로 예측하는 민감도는 로지스틱 회귀분석 기법이 의사결정나무 기법보다 예측력이 높게 나타났다. 또한 실제 이월자를 이월자로 예측하는 특이도는 의사결정나무 기법이 로지스틱 회귀분석보다 약간 높은 예측력이 있는 것으로 나타났다. 전반적으로 의사결정나무 기법과 전통적인 예측기법으로 사용되었던 로지스틱 회귀분석 기법간에 태권도장 수련생 이탈 예측에 차이가 없는 것으로 나타났다. 또한 민감도에서는 로지스틱 회귀분석이 특이도에서는 의사결정나무 기법이 태권도장 수련생 이탈 예측에 효과적이며, 이러한 결과를 토대로 태권도장 수련생 이탈 예측모형을 구축하는 것이 효율적이라 사료된다. 이는 배행수(2005), 서해옥(2002), 이지영(2002)의 데이터 마이닝을 이용한 고객 이탈 예측 연구결과에서 의사결정나무 기법과 로지스틱 회귀분석 모형 간 전반적인 분류 정확도는 두 모형 간에 일치하는 결과로 나타나 본 연구의 결과를

뒷받침 해주고 있다. 신선영(2005), 윤해원(2004), 정해원(2004), 김혜영(2005), 이현정(2001)의 데이터 마이닝을 이용한 이탈고객 분석예측에 관한 연구에서 의사결정나무 기법과 로지스틱 회귀분석 모형 간 민감도, 특이도, 정확도에서 부분적으로 일치하는 결과를 나타내어 본 연구를 뒷받침해주고 있다.

2. 의사결정나무분석과 로지스틱회귀분석간의 이탈에 미치는 변인의 차이

'가설Ⅱ. 의사결정나무분석과 로지스틱회귀분석간의 이탈에 미치는 변인의 차이가 없을 것이다.'를 검증한 결과 로지스틱 회귀분석에서 가장 이탈에 영향을 미치는 변인으로는 수련기간, 학년, 추천의사, 시설 만족도, 성별 등 5가지로 나타났으며, 의사결정나무 분석 결과로는 수련기간, 학년, 추천의사, 지도자 만족, 시설만족도 등 5가지로 나타났다.

첫째, 의사결정나무분석과 로지스틱회귀분석을 통해 이탈 요인을 분석한 결과 가장 영향력이 높은 변인은 수련기간이었다. 즉, 수련기간이 길었을 경우 수련생들의 이탈율은 높은 것으로 나타났다. 이러한 결과는 수련기간이 길어질수록 수련생들이 느끼는 환경요인적 만족도에 대한 욕구를 충족시켜 주지 못했기 때문이라 판단된다. 이와 같은 결과는 김주훈(2001)의 태권도장 운영의 활성화 전략에 관한 연구에서 수련기간이 길수록 환경요인적 만족도가 낮은 것으로 나타났고 박진기(2001)의 스포츠센터 고객 이탈가능성 예측에 관한 연구에서 이용빈도가 많아질수록 이탈에 중요한 영향을 미치는 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

둘째, 의사결정나무분석과 로지스틱회귀분석을 통해 이탈 요인을 분석한 결과 두 번째로 영향력이 높은 변인은 학년으로 나타났다. 즉, 고학년일수록 수련생들의 이탈율은 높은 것으로 나타났다. 이러한 결과는 저학년에서 고학년으로 올라갈수록 운동보다는 학업에 치중하는 경향이 있기 때문이라 판단된다. 이와 같은 결과는 김주훈(2001), 이웅현(2001)의 태권도 체육관 수련생 만

족도에 관한 연구에서 학년이 높을수록 지도자 만족도가 낮은 것으로 나타났고 옥기현(2003)의 데이터 마이닝을 이용한 이탈학생의 예측 연구에서 학년이 이탈에 중요한 영향을 미치는 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

셋째, 의사결정나무분석과 로지스틱회귀분석을 통해 이탈 요인을 분석한 결과 세 번째로 영향력이 높은 변인은 추천의사로 나타났다. 즉, 추천의사가 낮을수록 이탈율이 높은 것으로 나타났다. 이러한 결과는 연구대상자가 태권도장에 대한 환경요인적 요인 만족도가 낮았기 때문에 타인들에게 추천하고 싶지 않았을 것이라 판단된다. 이와 같은 결과는 정억순(2003)의 태권도 서비스 만족도가 차후 태권도수련 추천의도에 미치는 영향에 관한 연구에서 환경요인적 만족도가 추천의도에 영향을 미치지 않는다는 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

넷째, 의사결정나무분석과 로지스틱회귀분석을 통해 이탈 요인을 분석한 결과 네 번째로 영향력이 높은 변인은 지도자만족도로 나타났다. 즉, 지도자에 대한 만족도가 낮을수록 이탈율이 높은 것으로 나타났다. 이러한 결과는 지도자의 지도방법이나 품성 그리고 자질에 대한 평가가 낮았기 때문이라 판단된다. 이와 같은 결과는 김백윤(2005)의 태권도 수련생의 지도자만족도에 관한 연구에서 수련생이 지도자에 대한 외모, 언행, 관원존중 등의 지도자요인에서 낮은 만족도를 나타난다는 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

다섯째, 로지스틱회귀분석을 통해 이탈 요인을 분석한 결과 다섯 번째로 영향력이 높은 변인은 성별이다. 즉, 성별이 여자일 경우 이탈율이 높은 것으로 나타났다. 이러한 결과는 태권도장의 수련생 대부분이 남성으로 이루어져 있고, 지도자 또한 남성으로 구성되었으며, 프로그램도 남성 수련생 위주로 짜여져 있다. 따라서 여자 수련생들의 이탈율이 높은 것은 환경요인 만족도에 대해 불만족하기 때문이라 판단된다. 이와 같은 결과는 최원보(2002), 이장원(2001)의 태권도장 경영전략연구의 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

여섯째, 의사결정나무 분석을 통해 이탈 요인을 분석한 결과 여섯 번째로 영향력이 높은 변인은 시설만족도이다. 즉, 시설에 대한 만족도가 낮을수록 이탈율이 높은 것으로 나타났다. 이러한 결과는 노후된 시설이나 태권도장 용품 등에 대한 평가가 낮았기 때문이라 판단된다. 이와 같은 결과는 이송학(2005)의 태권도장 서비스품질 향상을 위한 남·여 고객 만족도에 관한 연구에서 수련생이 식수, 탈의실, 차량운행 등의 시설요인에서 낮은 만족도를 나타낸다는 결과와 일치하여 본 연구의 결과를 뒷받침 해주고 있다.

V. 결론 및 제언

1. 결 론

본 연구는 로지스틱 회귀분석과 데이터마이닝 기법 가운데 의사결정나무 기법의 비교 분석을 통한 태권도장의 바람직한 이탈 수련생 관리를 위한 기초적 자료를 제공하기 위하여 태권도장의 수련생들을 대상으로 이탈유무에 따른 세분화와 이탈가능성을 예측하고자 실시하였다. 이와 같은 연구의 목적을 달성하기 위하여 경기도 및 인천광역시 소재의 태권도장 수련생을 모집단으로 설정하고 비확률 표본 추출법 중 편의추출법을 사용하여 경기도 및 인천소재의 태권도장 수련생에게 1500부를 배포 및 수집하여 1,149명의 표본을 추출하였다. 설문지를 회수한 후 3개월이 경과한 뒤 등록유무를 파악하여 이월수련생 938명과 이탈수련생 211명으로 파악하였다. 자료처리는 SPSSwin 13.0 프로그램을 이용하여 교차분석, 신뢰도검증, 요인분석, 카이제곱 검증, t-검증, 의사결정나무 기법, 로지스틱회귀분석을 사용하였다.

이러한 절차와 방법을 통하여 도출된 본 연구의 결과는 다음과 같다.

첫째, 의사결정나무 기법과 로지스틱 회귀분석간 특이도의 차이를 알아본 결과 실제 이월자를 이월자로 예측하는 특이도에서는 로지스틱 회귀분석은 92.9%, 의사결정나무 기법이 96.3%로 다소 높게 나타났다.

둘째, 의사결정나무 기법과 로지스틱 회귀분석간 민감도의 차이를 알아본 결과 실제 이탈자를 이탈자로 예측하는 민감도에서는 로지스틱 회귀분석이 64.0%, 의사결정나무 기법은 44.5%로 나타났다.

셋째, 의사결정나무 기법과 로지스틱 회귀분석간 정확도의 차이를 알아본 결과 전체적인 분류정확도에서 의사결정나무기법이 86.8%, 로지스틱 회귀분석은 87.6%로 나타났다.

넷째, 의사결정나무 기법과 로지스틱 회귀분석간의 태권도장 수련생 이탈에 영향을 미치는 변인의 차이를 알아본 결과 의사결정나무 기법은 수련기간, 학년, 지도자만족, 추천의사, 시설만족도 변인이며, 로지스틱 회귀분석은 수련기간, 학년, 추천의사, 지도자만족도, 성별 변인으로 나타났다.

이상의 결과를 종합해보면 의사결정나무 기법과 로지스틱 회귀분석간 태권도장 수련생 이탈 예측에 대한 비교 분석결과 두 모형 모두 분류정확도에서 높은 예측률을 나타내었고, 두 분석모형간 예측률의 차이는 나타나지 않았다. 이러한 결과는 이탈에 영향을 미치는 변인이 동일하게 사용되었기 때문이라 사료된다. 특히 의사결정나무의 분류규칙에 의한 태권도장 수련생의 세분화를 통해 태권도장 수련생 이탈 가능성 예측 나무구조모형을 통해 제시할 수 있다. 따라서 태권도장의 일선 지도자나 실제 의사결정권자인 수련생은 본 연구에서 밝히고 있는 예측모형과 예측률을 실제 수련생 관리와 경영전략수립에 이용함으로써 수련생 이탈 방지 및 효율적 경영에 도움을 줄 수 있을 것이다.

2. 제 언

이상과 같은 본 연구의 비교분석과정을 통하여 얻어진 본 연구의 결과는 실제 태권도장에서 수련생관리에 매우 중요한 기초적 자료를 제공할 수 있을 것으로 여겨진다. 그러나 연구의 과정에서 발견된 문제점을 기초로 하여 후속연구를 위하여 다음과 같은 제언을 두고자 한다.

첫째, 본 연구는 지역을 수도권으로 한정하였기 때문에 지역을 보다 광범위하게 설정하여 보다 다양한 후속연구가 진행되어야 할 것이다.

둘째, 본 연구에서는 수련생 이탈 가능성을 예측하기 위한 설명변수를 설문지를 통한 자료수집방법을 이용하였지만, 후속 연구에서는 실제 태권도장의 전산화된 수련생 관리 프로그램을 통하여 자료를 얻는 것이 바람직할 것으로 여겨진다.

셋째, 본 연구에서는 수련생 이탈 가능성을 예측하기 위한 설명 변수로 선행연구와 예비조사를 거쳐 가능한 유의한 변수를 선정하고자 노력하였으나, 후속연구에서는 보다 더 많은 반복분석을 통하여 이탈유무에 영향을 미치는 변수를 선정할 것이 요구되어진다.

넷째, 본 연구에서는 수련생 이탈가능성을 예측하기 위해 의사결정나무 기법과 로지스틱 회귀분석을 사용하였지만, 후속연구에서는 신경망분석, 판별분석 등 다양한 예측 통계모형과 비교분석하여 이탈 예측에 유용한 모형을 구축할 수 있는 연구가 필요하겠다.

참 고 문 헌

[국내문헌]

강성구 (2004). 데이터 마이닝에서 상호정보를 이용한 변수선택방법에 대한 연구. 석사학위논문. 성균관대학교 대학원.

강한구 (2003). 이탈 고객 분류를 위한 데이터마이닝 방법의 비교 연구 .석사학위논문. 동의대학교 대학원.

강병길 (2003). 태권도 도장 시설 및 운영에 관한 학부모의 만족도 분석. 한국스포츠산업・경영학회지, 7(2), 101-112.

기현옥 (2003). 데이터마이닝을 이용한 이탈학생의 예측모형개발. 석사학위논문. 한림대학교 대학원.

김백윤 (2005). 태권도 수련생의 인구통계학적 특성에 따른 체육관 프로그램 및 지도자 만족도 분석. 한국스포츠산업・ 경영학회지 10(3), 71-81.

김갑수 (2003). 태권도 도장 수련생의 교육 및 시설만족도. 스포츠리서치, 15(2), 911-918.

권혜숙 (2002). 데이터 마이닝 패키지에서 분류나무 알고리즘의 비교 연구. 이학석사학위논문. 서울대학교 대학원.

김수정 (2005). 초등학교 태권도 수련생들의 도장 만족도와 중도포기 원인. 석사학위논문. 세종대학교 대학원.

고봉수 (2005). 태권도장 경영프로그램이 만족과 재구매에 미치는 영향. 이학박사 학위논문. 원광대학교 대학원.

김영갑, 이정식 (2005). 태권도의 수련정도와 참여 동기 및 지도자 이미지의 관계. 한국체육학회지, 44(1), 523-533.

김학덕, 임재구 (2005). 태권도 도장 관원들의 참여 동기에 따른 지속성여부에 관한 연구. 한국스포츠리서치, 16(4), 897-906.

김기덕 (2002). 태권도장의 서비스 품질이 수련 지속 행동에 미치는 영향. 석사 학위논문. 연세대학교 교육대학원.

김주안 (2002). 로지스틱 회귀분석을 이용한 제품구매의도에 영향을 미치는 온자인 커뮤니티 요인에 관한 연구.

배행수 (2005). 고객 이탈 예측 데이터 마이닝 기법 비교 연구. 석사 학위논문. 연세대학교 정보대학원.

박진기 (2001). 데이터 마이닝 기법을 활용한 스포츠센터 고객 이탈가능성 예측 모형 개발. 박사학위 논문. 계명대학교 대학원.

서해옥 (2001). 데이터마이닝을 이용한 인터넷 쇼핑몰의 이탈 고객 분석 모형에 관한 연구. 석사학위 논문. 이화여자대학교 경영대학원.

신선영 (2004). 데이터 마이닝을 이용한 제약회사 이탈고객 분석예측에 관한 연구. 석사학위 논문. 이화여자대학교 경영대학원.

이송학 (2005). 태권도장 서비스 품질 향상을 위한 남·여 고객 서비스 만족도 및 불평행동 분석. 스포츠리서치, 16(4), 955-965.

이지영 (2002). 데이터 마이닝 기법을 이용한 전공 이탈학생 분석 및 예측모형 개발. 석사학위논문. 계명대학교 대학원.

이종천 (2001). 태권도 체육관에서 수련생 이탈 원인에 관한 연구. 석사학위논문. 경희대학교 대학원.

이현정 (2001). 데이터마이닝을 이용한 보험회사 고객이탈분석에 관한 연구. 석사학위논문. 중앙대학교 대학원.

윤혜원 (2004). 데이터 마이닝 기법을 이용한 고객이탈 분석 및 예측모형 개발. 석사학위논문. 고려대학교 대학원.

장해진 (2005). 태권도 체육관의 효율적 경영을 위한 마케팅 요인 분석. 석사학위논문. 건국대학교 교육대학원.

정혜원 (2004). 통계적 기법과 데이터마이닝 기법을 이용한 이동통긴 VAS 가망고객 scoring 모형 비교 연구. 석사학위논문. 연세대학교 대학원.

정억순 (2003). 태권도장 서비스에 대한 만족도가 차후 태권도수련 추천의도에 미치는 영향. 한국체육학회지, 42(5), 589-596.

최상진. 조옥성 (2003). 태권도 체육관의 프로그램 및 시설에 따른 수련생 만족도 분석. 한국스포츠리서치, 14(5), 313-326.

최대호 (1999). 데이터마이닝 기법과 로지스틱회귀분석 기법의 예측 퍼포먼스 비교분석. 석사학위논문. 서강대학교 대학원.

[국외문헌]

Anwar, T M, H W beck, and S, B Navathe(1992), “Knowledge Mining by imprecise Querying A Classification-Based Approach,” IEEE 8th International Conference on Data Engineering, Phoenix, Arizona, February, 35-48.

Becher, J. Berkhin, P. & Freeman, E. (2000). Automating exploratory data analysis for efficient date mining. Proceedings 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA USA. 424-429.

Berry, M. J. A. & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. Canada: John Wiley & Sons, Inc.

Brachman, R J, T Khabaza, W. Kioesgen, G Piatetsky-shapiro, and E Simoudis(1996) "Mining Business Databases," Communications of the ACM. 39(11), 35-48 .

Breiman, L, J H Freidman R, A. Olshen, and C. J Stone(1984), Classification and Regression Tree, Belmont: Wordsworth

Courtheoux, R J (1989), "Database Techniques How th Key Company Resource,"Brown, Herbert E, and B Buskirk (eds.), Redings & Cases in Direct Marketing, Chicago NTC Business Books, 93-102.

Dallal, G. E. (2001). Logistic regression. Retrieved April 17, 2001, from htpp://www.tufts.edu/~gdallal/logistic.htm

Edelstein. (1988) Introduction to Data Mining & Knowledge Discovery. Two Crows Crp, 7-8.

Fayyad, U M, G Piatetsky-Shapiro, and P Smith (1996),"From Data Mining to Knowledge Discovert," In Knowledge Discovery and Data Mining, (eds) Fayyas, U M, G Piatetsky-Shapiro, P Smith, and R Uthurusamy, CA The MIT Press, 1-34.

Friedman, J. H. (1997). data mining and statistics: What's the connection? Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics, Houston, Texas. 1-7

Goodwin L, Van Dyne M, Lin S, Tallbert S. (2003) Data mining issues and opportunities for building nursing knowledge. Journal of biomedical informatics, 36, (4~5), 229-231.

Han, J. & Kamder, M. (2001). Data mining: concepts and techniques. San Diego, CA: Academic Press.

Han, J. (1996). Data mining techniques. Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada. 545.

Hirji, K. K. (2001). Exploring data mining implementation. Communications of the ACM, 44, 87-93.

Inmon, W. H(1996), Building the Data Warehouse, NY. John Wiley & Sons, 유영일 역(1997) [데이터웨어하우스 구축방법론], 제2판, 서울: 홍릉과학출판사.

Mannila, H.(1996), "Data Mining: Machine learning, statistics, and databases,"Eight International Conference on Scientific and Database Management, p 1-8.

Quinlan, J R.(1986), "Introduction th decision tree,"Machine Learning, Vol. 1, 81-106.

Saarenvirta, G(1998), "Data Mining th improve profitabihty," CMA Magazine, Vol 72, No. 2, 8-12.

ABSTRACT

Comparative Analysis of Prediction Taekwondo Trainee`s Defection using Decision Tree and Logistic Regression.

Koo, You-Hoe
Department of Sports and Well-Being
General Graduate School
Hanyang University

The purpose of this study is to suggest the most appropriate prediction model for prediction of defection of trainees of Taekwondo gymnasium through decision-making tree technique and logistics regression analysis. In order to accomplish the purpose of this study, I have distributed 1,500 questionnaires sheets to the trainees using by convenience sampling method among non probability sampling extraction methods by setting trainees of Taekwondo gymnasium located in Gyeonggi-Do and Incheon Metropolitan City. Among retuned questionnaires, I have used 1,149 sheets of questionnaires as analysis data by excluding data that are judged as data without reliability from the object of analysis. After the lapse of three months after collecting the questionnaires, I have found out 938 trainees of carry forward and 211 trainees of defection by examining the registration or not at that time.

For data processing, I have used cross analysis, reliability verification, factors analysis, chi-square verification, t-test, decision-making tree technique, and logistics regression analysis using SPSS win 13.0 program.

Results of this study derived from this procedure and method are as follows.

1. As the result of examining the difference of level of specialty between decision-making tree technique & logistics regression analysis, in the level of specialty predicting carry forward number of people with actual number of carry-over, logistics regression analysis was 92.9% and decision-making tree technique showed a little higher at 96.3%.

2. As the result of examining the difference of level of sensitivity between decision-making tree technique & logistics regression analysis, in the level of sensitivity predicting carry forward number of people with actual number of carry-over, logistics regression analysis was 64.0% and decision-making tree technique showed 44.5%.

3. As the result of examining the difference of level of accuracy between decision-making tree technique & logistics regression analysis, in the level of accuracy predicting carry forward number of people with actual number of carry-over, logistics regression analysis was 86.8% and logistics regression analysis showed 87.6%.

4. As the results of examining variables affecting defection of trainees of Taekwondo gymnasium between decision-making tree technique and logistics regression analysis, in case of decision-making tree technique, variables were training period, grade, satisfaction of instructor, recommendation intention, and satisfaction of facility and in case of logistics regression analysis, variables were training period, grade,

recommendation intention, satisfaction of instructor, and sex.

Summarizing the above result, as the result of comparison analysis of defection prediction of trainees of Taekwondo gymnasium between decision-making tree technique and logistics regression analysis, the two models showed all high prediction rate in the level of accuracy and there was no difference of prediction rate between the two analysis models. This result is believed to be because the same variables affecting defection were used in the same.

Therefore, a front line leaders of Taekwondo gymnasium trainees can prevent trainees defection and help efficient management by using the prediction models and prediction rates in actual management of trainees as well as in the establishment of management strategy.

태권도 도장 수련생의 환경요인별 만족도에 관한 설문지

안녕하십니까?

먼저 바쁘신 학업에도 불구하고 연구에 협조하여 주심에 감사드립니다. 이 설문지는 학문적 연구 이외에는 어떠한 목적에도 사용하지 않을 뿐 아니라, 통계적인 자료처리에만 이용될 것입니다.

귀하의 좋은 의견이 본 연구에 활용될 수 있도록 도와주신 점 다시 한번 깊은 감사를 드리며 귀하의 앞날에 무궁한 발전이 있으시길 바랍니다.

2006년 12월

한양대학교 일반대학원

생활스포츠학 전공 구 유 회

연락처 : 031) 400-5730

010-4784-2540

1. **일반사항**

※ 다음은 귀하에 대한 일반적인 사항에 대한 물음입니다. 해당되는 곳에 "✔"표를 하거나 자세히 기입하여 주시기 바랍니다.

1. 귀하의 성별은?
① 남 ( ) ② 여 ( )

2. 귀하의 학년은?
① 4학년 ② 5학년 ③ 6학년

3. 귀하는 태권도를 배운지 얼마나 되었습니까?
① 6개월 이하 ② 6개월-1년 이하 ③ 1년-2년 이하 ④ 2년-3년 이상

4. 귀하는 한달에 몇 번 정도 수업에 결석합니까?
① 1번 ② 2번 ③ 3번 ④ 4번 ⑤ 5번 이상

5. 귀하는 주변사람에게 현재 다니고 있는 태권도 도장에 대해 추천할 의사가 있습니까?
① 매우 그렇다 ② 그렇다 ③ 보통이다 ④ 아니다 ⑤ 전혀 아니다

6. 귀하는 앞으로 지속적으로 태권도 도장에 다닐 것입니까?
① 그렇다 ② 아니다

7. 귀하는 현재 다니고 있는 태권도장이 처음입니까?
① 그렇다 ② 아니다

7-1. 위의 아니다 라고 응답한 사람만 작성해 주시기 바랍니다.
도장 옮긴 횟수 / ① 2회 ② 3회 ③ 4회 ④ 5회 이상

7-2. 옮긴 이유는 무엇 때문입니까?
① 체육관 부적응 ② 프로그램의 다양성 부족 ③ 지도자의 능력부족
④ 거리가 멀어서 ⑤ 재미없어서 ⑥ 부모님이 다니지 말라고 해서
⑦ 시설이 좋지 않아서 ⑧ 기타

8. 귀하는 태권도장 외에 몇 개의 학원을 다니고 있습니까?
① 1개 ② 2개 ③ 3개 ④ 4개 이상

9. 귀하의 부모님은 직장에 다니고 있습니까?
① 아버지 ② 어머니 ③ 두분 모두

다음의 만족도에 관한 각 문항을 읽으신 후, 학생이 생각하는 것과 일치하는 정도에 따라 해당 번호에 “○” 또는 “✓”를 해주시기 바랍니다.

***1.* 다음은 시설에 대한 관련 문항들입니다.**

번호	항 목	전혀 그렇지 않다.	그렇지 않다	보통 이다	그렇다	매우 그렇다
1	운동하기에 체육관은 충분히 넓었다.					
2	체육관은 항상 깨끗했다.					
3	체육관은 늘 안전했다.					
4	운동기구는 상태가 좋았다.					
5	체육관 바닥은 좋았다.					
6	냉방상태는 양호했다.					
7	난방상태는 양호했다.					
8	체육관 공기는 항상 쾌적했다.					
9	마시는 물은 깨끗하게 관리되었다.					
10	실내는 충분히 밝았다.					
11	체육관과의 거리는 적절하였다.					
12	차량운행은 만족스러웠다.					

***2.* 다음은 지도자에 대한 관련 문항들입니다.**

번호	항 목	전혀 그렇지 않다.	그렇지 않다	보통 이다	그렇다	매우 그렇다
1	관장님과 사범님은 항상 단정했다.					
2	관장님과 사범님은 친절하였다.					
3	관장님과 사범님의 인격/품성은 훌륭하였다.					
4	관장님과 사범님은 운동을 지도하는 능력은 좋았다					
5	관장님과 사범님은 예절교육을 강조 하였다.					

3. **다음은 프로그램에 대한 관련 문항들입니다.**

번호	항 목	전혀 그렇지 않다.	그렇지 않다	보통 이다	그렇다	매우 그렇다
1	운동프로그램은 지루하지 않았다.					
2	운동프로그램은 학년(나이)에 맞게 지도되었다.					
3	운동프로그램은 띠별로 수준에 맞게 지도되었다.					
4	운동프로그램은 운동량은 적당했다.					
5	운동프로그램은 체력향상에 도움이 되었다.					
6	운동프로그램은 스트레스해소에 도움이 되었다.					
7	운동프로그램은 좋은 인성발달에 도움이 되었다.					
8	운동프로그램은 다양하게 짜여져 있다.					

※ 설문에 응답해 주셔서 대단히 감사합니다.