

의사결정모형을 이용한 과원조성 결정요인

소남호*, 정재원*, 이성빈*, 오상현*

*농촌진흥청 농산업경영과

e-mail:soxxxx007@korea.kr

Decision Factors of Orchard Establishment with Decision-making Tree Model

Namho So*, Jaewon Jeong*, Seongbin Lee*, Sangheon Oh*

*Dept. of Agribusiness Farm Management, RDA

요약

본 논문에서는 최근 재배면적과 생산량이 급증하고 있는 샤인머스켓 과원조성비 및 과원진입에 영향을 미치는 변수를 의사결정나무 모형과 랜덤포레스트 모형을 통해 실증분석을 수행했다. 분석결과 의사결정나무 모형은 과원조성비를 중심으로 농가수취가격과 수량으로 나타났고, 랜덤포레스트 모형은 샤인머스켓 재배면적을 중심으로 노동시간과 경영비에 의해서 결정됨을 보였다. 모형평가 결과 랜덤포레스트 모형이 모든 지표에서 우월함을 보였으며, 주요 결정변수는 재배면적과 노동시간이 중요함을 보였다. 본 논문은 148 포도재배농가를 샤인머스켓과 그 외 농가로 구분하였으나 실제 농가는 조생종, 중생종, 만생종 등으로 작물포트폴리오를 구성하여 투입 가능한 노동력(자가, 고용)을 고려한 최적 재배면적을 경작하고 있기 때문에, 이를 고려하지 않고 단일품목 전업농으로 분석한 결과는 한계가 있을 수 있다.

* 본 논문은 농촌진흥청에서 수행 중인 '주요 과수의 조성비 분석' 과제(PJ01391601)의 일환으로 진행된 것임.

1. 서론

농가소득의 정체 또는 완만한 증가에도 불구하고 그 중 농업소득이 차지하는 비중이 최근 10년간 정체되고, 농외소득(겸업소득, 사업소득)과 이전소득(공적 부조, 사적부조) 비중은 지속적으로 증가하고 있다. 농업소득은 1년간 작물을 재배하고 시장에 판매하여 얻는 총수입에서 재배를 위해 투입요소들의 화폐가치인 경영비를 빼서 산출한다. 따라서 소득에 영향을 미치는 요인은 크게 세 가지로 나뉜다. 첫째 단위면적당 수량, 둘째는 단위수량 당 농가수취가격, 셋째는 경영비이다[1]. 기후변화로 인한 수량변화와 시장개방과 폭넓은 소비자의 선호로 결정되는 가격에 의해서 결정되는 총수입과 달리 상대적으로 경영비는 농가 의사결정과 경영전략에 따른 경영개선으로 비용절감을 할 수 있는 분야이다. 따라서 외부 환경에 영향을 받는 수량과 가격에 비해서, 비용절감을 통한 소득증대는 단기적으로 농업인이 선택할 수 있는 방법 중 하나다. 또한 본 논문에서 분석대상으로 삼는 과수는 과중한 해에 수확하여 소득을 얻는 식량작물이나 노지채소와 달리 한번 심으면 성목수령에 이르기 전까지 지속적인 비용투입(음의 소득)과 초기 기반투자비용(토지개량, 묘목재식, 지주설치,

관수시설, 배수시설 등)으로 작물선택에 매우 신중한 접근을 필요로 한다. 최근 국내외 소비자 선호증가와 지자체 지원에 힘입어 샤인머스켓의 재배면적과 생산량이 급증하고 있는 상황에서 기존 과수농가 작물전환을 하거나 신규 농가가 진입하기 전 농가의 현금흐름, 기술수준, 출하처 등에 대한 면밀한 분석과 향후 전망에 대한 대비를 충분히 해야 할 필요가 있다.

농촌진흥청에서는 매년 작물별 소득을 조사·분석하여 농산물소득조사자료집을 발간하여 농가의 경영진단 연구·지도 자료 및 농업정책 의사결정 등의 기초자료로 활용하고 있다. 이중 과수의 경영비 중 과원조성비¹⁾가 있으며, 조성비는 농가에서 과수묘목을 심은 후 총수입과 생산비가 같아지는 시기²⁾의 총투입비용에서 그해까지의 수입을 뺀 금액을 해당과수의 내용연수로 나누어서 산출하고 있다.

과원조성비 관련 선행연구로 한국농촌경제연구원 과수(사과, 배, 포도) 농가의 고정자산투자결정모형을 제시하여,

- 1) 조성비는 조성기간 중 투입된 묘목, 비료, 농약 등과 울타리, 관배수시설 등에 투입된 비용을 현재가격으로 환산한 금액에서 같은 기간에 과일을 수확 판매하여 얻은 총수입을 현재가격으로 환산한 금액을 뺀 차액을 과수의 내용연수로 나누어 연간 과원조성비로 계산한다.
- 2) 성목수령은 당해연도 과일을 수확 판매하여 얻은 총수입이 당해연도 투입비용을 보상하고 순수익이 발생하는 수령(손익분기점)을 의미하며 성목수령까지 기간을 조성기간(2020년 농산물소득조사분석방법)

과원조성에 필요한 투자결정에 필요한 순자산에 대한 수익발생 여부를 기준으로 모형 설계하여 농가가 과수선정, 생산기간과 경영규모를 입력하여 평가결과를 도출하게 하였다[2]. 우수곤 등(2006)은 6개 과수의 재식형태별 신규 과원의 기반정비, 재식비용, 지주시설 등 시설비를 조사하고 연차별 경영성과분석을 통한 과원조성비를 산출하여 제시하였다[3]. 최재혁 등(2009)은 단감의 수령별 생산량 실태조사 및 유형별 수익성 분석을 통하여 과실 품질과 수량을 조사하여 생산량 및 조수입 분석을 위한 기초자료로 제공하였다[4]. 김승희(2012)는 농가재해 보험금 산정을 위한 기준자료로 활용하기 위하여 사과, 배 등 7개 과종에 대한 현지조사를 통해서 품종별 재식주수별 표준 수확량을 산출하였다[5]. 하두중 등(2016)은 합리적인 생산계획 수립과 오래된 과원의 갱신을 위한 의사결정에 합리적인 보상기준 설정을 위한 수령별 경제적 가치를 분석하여 제공하였다[6]. 그러나, 선행연구는 대부분 과수의 합리적 생산계획 수립과 과원 갱신 의사결정을 위한 수령별, 재식형태별 생산량과 수익성 분석이며, 과종도 한정되어 있고, 무엇보다 분석시기 경과로 현재 적용하기에는 한계가 있다. 따라서 분석시기를 현재와 가까운 시기로 바꾸고 과종을 확대하여 수령별, 재식형태별 생산량과 수익성 분석을 제공하는 것도 중요하며, 더욱이 과원조성비에 영향을 미치는 요인들에 대한 연구가 부족했다는 점에서 본 논문은 과원조성비에 영향을 미치는 요인(범위형, 수치형)을 밝힘으로써 농업인이 보유한 역량 및 현황에 맞는 작물을 선택할 수 있게 정보를 제공함을 목적으로 한다. 이를 위해 범위형 자료와 수치형 자료를 동시에 분석할 수 있으며, 농업인들도 결과를 직관적이고 도표화된 형태로 이해할 수 있는 의사결정나무모형(Decision Tree Analysis)을 적용하고자 한다.

2. 본론

2.1 의사결정나무 모형

의사결정나무 모형은 자료 내 존재하는 ‘관계, 패턴, 규칙 등을 파악하고 발견하여 모형화하는 데이터마이닝 기법 중 하나이며, 스무고개와 같이 단계별 특정 요인에 대한 참과 거짓을 통해서 가지처럼 뻗어 나가는 형태를 나뭇가지로 표현하는 것으로 나무의 모형은 마디(node)들로 구성된 하향식 나무구조를 가진다. 나무 마디는 기능적 요소에 따라 뿌리마디(Root node), 자식가지(Child node), 중간마디(Internal node), 끝마디(Terminal node)와 마디를 이어주는 가지(Branch)로 구별할 수 있다[7].

의사결정나무 모형의 분석과정은 다음과 같다. 첫째, 나무 모형의 최상위에 존재하는 뿌리마디에서 시작하여 링크를 통해 의사결정에 따라 결정되는 경로(path)를 거쳐 끝마디까지

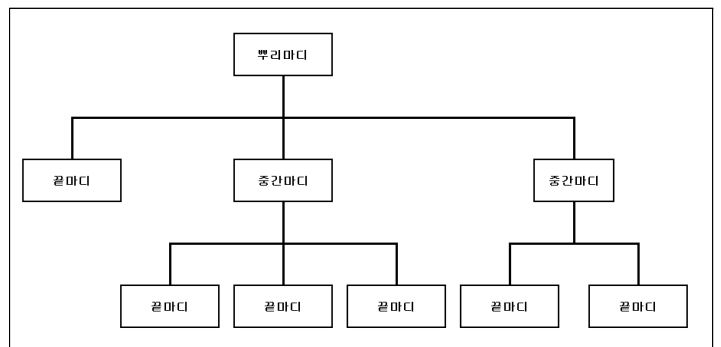
자식마디를 만들어서 특정 그룹으로 분류되거나 예측된다. 여기서 뿌리마디는 최상위 마디를 의미하며, 분류·예측 대상이 되는 모든 모형자료를 포함한다. 부모마디는 분류기준에 따라 링크로 연결된 자식마디를 형성하고, 더 이상 가지로 분기되지 않는 최종마디를 끝마디라고 한다[8]. 이 같은 구성요소를 정리하면 아래와 같다.

[표 1] 의사결정나무의 구성요소

구분	설명
뿌리마디 (Root node)	나무구조가 시작되는 마디(전체자료로 구성)
자식마디 (Child node)	하나의 마디에서 분기되어 나간 2개 이상 마디
부모마디 (Parent node)	자식마디의 상위마디
끝마디 (Terminal node)	각 나무줄기의 끝에 위치한 마디로 잎(leaf)라고도 하며, 끝마디 개수만큼 분류규칙 발생
중간마디 (Internal node)	나무구조의 중간에 있는 끝마디가 아닌 마디
가지 (Branch)	뿌리마디에서 끝마디까지 연결된 일련의 마디
깊이 (Depth)	뿌리마디부터 끝마디까지의 중간마디 개수

출처: 최중후·서두성(1999)

의사결정나무 모형은 상위마디(뿌리마디, 중간마디 등)에서 하위마디(자식마디, 끝마디 등)로 가지(Branch)가 뻗어나가는 과정을 가지분할(Split)이라고 하며, 가지분할 할 경우 ‘분류변수’와 ‘분류기준’값의 선택이 중요하다. 분류기준은 부모마디에서 자식마디로 뻗어갈 때 예측변수의 선택과 범주의 병합이 이루어질 기준을 의미한다. 선택기준은 상위마디에서 하위마디로 갈 때 마디 내에서는 동질성, 마디 간에는 이질성이 가장 커지는 방향을 선택한다.



[그림 1] 의사결정나무 모형의 기본구조

2.1.1 의사결정나무 모형(Decision Tree Model)

의사결정나무는 목표변수가 이산형 자료인 경우의 분류나무(Classification tree)와 목표변수가 연속형 자료인 경우의 회귀나무(Regression tree)로 구분된다. 이산형인 분류나무는 상위마디에서 가지분할 할 때, 분류변수와 분류기준값의 선택을 위해 카이제곱통계량(Chi-square statistic)의 p값, 지니

계수(Gini index), 엔트로피 지수(Entropy index) 등이 사용된다. 카이제곱통계량의 p값이 작을수록, 지니계수와 엔트로피 지수가 클수록 마디 간 이질성이 크며, 순수도가 낮아지므로, 마디 간에는 이질성이 크게 가지분할을 수행한다.

불확실성 측정값(Uncertainty measure)인 지니계수와 엔트로피 지수는 다음과 같이 정의한다.

$$G = 1 - \sum_i p_i^2 \quad (0 \leq G \leq 0.5)$$

$$E = - \sum_i p_i \log_2 p_i \quad (0 \leq E \leq 1)$$

의사결정나무 분석과정은 1단계 목표변수와 관계가 있는 설명변수를 선택, 2단계 분석목적과 자료의 구조에 따라 적절한 분류기준과 규칙을 정하여 의사결정나무 생성, 3단계 부적절한 나무는 제거(가지치기 Pruning), 4단계 이익, 위험, 비용 등을 고려하여 모형평가 5단계 분류 및 예측 수행으로 구성된다. 해석이 용이하고, 두 개 이상의 변수결합으로 목표변수에 미치는 영향(교호효과) 해석이 쉬우며, 비모수적 방법으로 선형성, 정규성, 등분산성 등의 제약을 받지 않으며 순위로 분석하므로 이상치에 민감하지 않은 장점이 있으나, 연속형 자료를 비연속형으로 취급하여 분류기준값 경계에서 오류발생 문제가 있으며 경로 의존적이므로 새로운 자료예측에서 불안정성이 높을 수 있다[9].

2.1.2 랜덤포레스트 모형(Random Forest Model)

의사결정나무 모형이 갖고 있는 문제점을 해결하기 위해 제시된 대안이 랜덤포레스트 모형이다. 의사결정나무 모형에서 만든 나무(Tree)들을 엮어서 Support Vector Machine, Boosting 등 기계학습을 통해 숲(Forest)을 만듦으로써 더 좋은 예측을 가능하게 한다. 랜덤포레스트 분석과정은 1단계는 50개에서 1,000개 의사결정나무를 만들고, 2단계는 나무에 임의성³⁾을 부여하고, 3단계는 앞 단계에서 도출한 의사결정나무의 결과들을 합하여 숲을 만들고 각각의 의사결정나무에 자료를 입력하여 더 많은 표를 받는 나무를 랜덤포레스트 결과값으로 채택한다. 과추정(Overfitting)되지 않고, 새로운 자료적용에 따른 예측불안정성이 낮아진다[10][11].

2.2 연구자료와 분석결과

포도 과원 조성비에 영향을 미치는 요인을 찾기 위해 2019년 농산물소득조사(농촌진흥청)에서 총 148개 포도재배농가의 소득원자료를 사용하였다. 또한 샤인머스켓과 그 외 품종(캠벨엘리, 거봉, MBA 등)으로 나뉜 품종 결정에 주요결정

변수를 찾기 위해 의사결정모형과 랜덤포레스트모형을 적용하였고, 모형결과를 평가하였다.

2.2.1 품종별 현황비교

최근 들어 재배면적이 크게 늘고 있는 샤인머스켓은 2006년 일본에서 등록된 품종으로 국내에 도입되어 재배면적이 1,867ha(2019년)으로 증가하였다. 조사 농가 중 샤인머스켓은 40농가(28%)가 재배하고, 총 영농경력, 작물재배경력 및 재배면적이 상대적으로 적으나, 지속적으로 재배농가 및 재배면적이 확대되고, 높은 교육시간(현장애로기술 해결)과 수출 증대를 위한 고품질 재배기술 확립이 나타나고 있다.

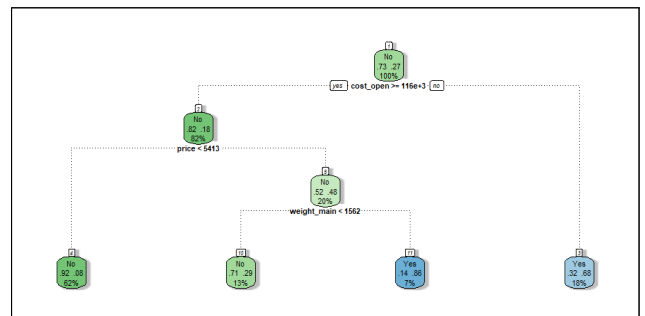
[표 2] 품종별 조사농가 현황비교

구분	샤인머스켓	그 외 품종
조사 농가수	40농가	108농가
총 영농경력	15.8년	23.7년
작물재배경력	3.5년	18.0년
작물재배면적	3.2a	59a
교육시간	41.8시간	25.3시간

출처: 2019년 농산물소득조사

2.2.2 분석결과

범위형 변수(지역, 성별)과 연속형 변수(수량, 총수입, 소득, 경영비) 등이 샤인머스켓 과원조성에 영향을 미치는 결과를 의사결정나무 모형을 통해 도출했고, 최적변수를 구하여 가지치기를 통해 얻은 결과는 그림2와 같다. 샤인머스켓 품종선택에 영향을 미치는 주요 변수는 과원조성비를 중심으로 농가수취가격과 수량으로 나타났다.



[그림 2] 의사결정나무 분석결과

의사결정나무 모형평가 결과는 표3과 같다. 정확도가 70.5%, 민감도 68.8%, 명확도 75.0%, 우위성 72.7%로 나타났다.

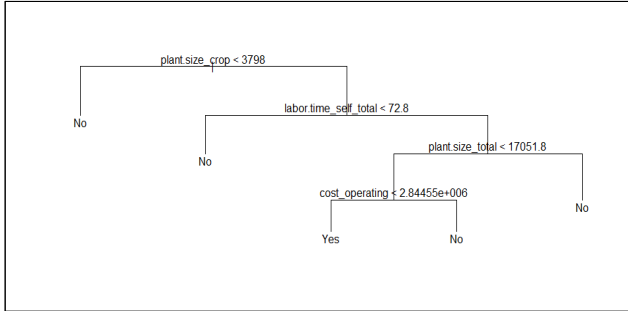
[표 3] 의사결정나무 모형평가

항목	정확도	민감도	명확도	우위성	수정 정확도
값	70.5%	68.8%	75.0%	72.7%	71.9%

랜덤포레스트 모형을 통해 도출했고, 최적변수를 구하여

3) 임의성을 주는 방법은 Bootstrap sampling process, Random attribute selection을 이용하여 의사결정나무를 다르게 만든다(의사결정나무와 Bagging 혼합모형).

가지치기를 통해 얻은 결과는 그림3와 같다. 샤인머스켓 품종 선택에 영향을 미치는 주요 변수는 샤인머스켓 재배면적을 중심으로 노동시간과 경영비에 의해서 결정됨을 보였다. 즉 샤인머스켓 재배면적은 0.4ha 이하, 자가노동시간은 72.8시간 이하, 총재배면적 1.7ha 이하이며 경영비가 280만원 이하일 경우 샤인머스켓 과원조성에 진입하게 된다.



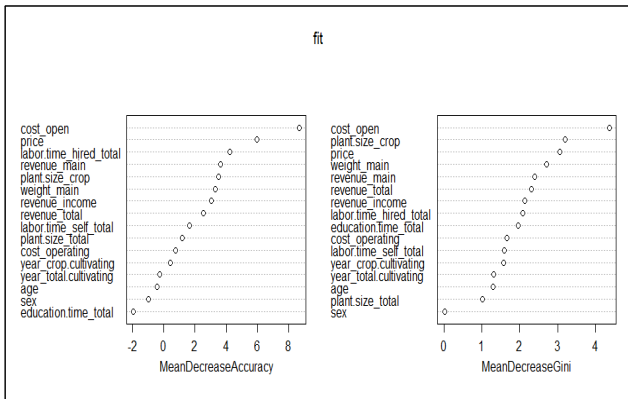
[그림 3] 랜덤포레스트 분석결과

표3과 표4로 두 모형을 비교하면 정확도 등 모든 비교지수에서 랜덤포레스트 모형이 우월함을 보인다.

[표 4] 랜덤포레스트 모형평가

항목	정확도	민감도	명확도	우위성	수정 정확도
값	72.7%	80.2%	77.0%	74.7%	78.6%

그림4를 통해 과원조성비 주요 결정변수는 샤인머스켓 재배면적, 노동시간이 가장 중요함을 보인다.



[그림 4] 과원조성비 주요 결정변수

3. 결론

최근 재배면적과 생산량이 급증하고 있는 샤인머스켓 과원 조성비 및 과원진입에 영향을 미치는 변수를 의사결정나무 모형과 랜덤포레스트 모형을 통해 실증분석을 수행했다. 분석결과 의사결정나무 모형은 과원조성비를 중심으로 농가수취가격과 수량으로 나타났고, 랜덤포레스트 모형은 샤인머스켓 재배면적을 중심으로 노동시간과 경영비에 의해서 결정됨을 보였다. 모형평가 결과 랜덤포레스트 모형이 모든 지표에서 우

월함을 보였으며, 주요 결정변수는 재배면적과 노동시간이 중요함을 보였다. 본 논문은 148 포도재배농가를 샤인머스켓과 그 외 농가로 구분하였으나 실제 농가는 조생종, 중생종, 만생종 등으로 작물포트폴리오를 구성하여 투입가능한 노동력(자가, 고용)을 고려한 최적 재배면적을 경작하고 있기 때문에, 이를 고려하지 않고 단일품목 전업농으로 분석한 결과는 한계가 있을 수 있다.

참고문헌

- [1] 농촌진흥청, “2020년 농산물소득조사분석방법”, 농촌진흥청, 2019년
- [2] 한국농촌경제연구원, “과수농가의 투자결정모형과 응용 프로그램 개발”, 연구보고서 2002년
- [3] 우수곤, “농축산물 소득조사 분석”, 농촌진흥청 연구과제(농축산물 소득조사 분석) 연구보고서 2006년
- [4] 최재혁, 신현열, 최성태, 하두중, 정정석, 오호상, “단감의 수령별 생산량 실태조사 및 유형별 수익성 분석”, 농촌진흥청 연구과제 보고서, 2009년
- [5] 김승희, “농작물 재해보험 과수품목 표준수확량 연구”, 농촌진흥청 연구과제, 2012년
- [6] 하두중, 박원흠, 최재혁, 고상환, “주요 과수의 수령별 경제적 가치 분석”, 농촌진흥청 연구과제(사과, 배 수령별 경제적 가치 분석) 보고서, 2016년
- [7] 임은정·정순희, “의사결정나무 분석을 통한 한국 중고령자의 점진적 은퇴의사결정에 관한 연구”, Financial Planning Review, 제8권 제3호, pp. 167-195, 2015년
- [8] 최종후, 서두성, “데이터마이닝 의사결정나무의 응용”, 통계분석연구, 제 4권 1호, pp. 61-83, 봄, 1999년
- [9] 이민아, “의사결정나무분석을 이용한 공무원 이직의도의 영향요인 예측모형”, 서울대학교 석사논문 2020년
- [10] Breiman, L., “Random Forests”, Machine Learning 45(1), pp. 5-32, 2001
- [11] Breiman, L., “Manual On Setting Up, Using, and Understanding Random Forests v3.1”, https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf