

# 로그 회귀분석 및 CART를 활용한 수력사업의 CDM 승인여부 예측 모델에 관한 연구

박종호<sup>1</sup> · 구본상\*

<sup>1</sup>서울과학기술대학교 건설시스템공학과

## Predicting the success of CDM Registration for Hydropower Projects using Logistic Regression and CART

Park, Jong-Ho<sup>1</sup>, Koo, Bonsang\*

<sup>1</sup>Department of Civil Engineering, Seoul National University of Science and Technology

**Abstract :** The Clean Development Mechanism (CDM) is the multi-lateral 'cap and trade' system endorsed by the Kyoto Protocol. CDM allows developed (Annex I) countries to buy CER credits from New and Renewable (NE) projects of non-Annex countries, to meet their carbon reduction requirements. This in effect subsidizes and promotes NE projects in developing countries, ultimately reducing global greenhouse gases (GHG). To be registered as a CDM project, the project must prove 'additionality,' which depends on numerous factors including the adopted technology, baseline methodology, emission reductions, and the project's internal rate of return. This makes it difficult to determine ex ante a project's acceptance as a CDM approved project, and entails sunk costs and even project cancellation to its project stakeholders. Focusing on hydro power projects and employing UNFCCC public data, this research developed a prediction model using logistic regression and CART to determine the likelihood of approval as a CDM project. The AUC for the logistic regression and CART model was 0.7674 and 0.7231 respectively, which proves the model's prediction accuracy. More importantly, results indicate that the emission reduction amount, MW per hour, investment/Emission as crucial variables, whereas the baseline methodology and technology types were insignificant. This demonstrates that at least for hydro power projects, the specific technology is not as important as the amount of emission reductions and relatively small scale projects and investment to carbon reduction ratios.

**Keywords :** CDM, Hydro Power, Data mining, Logistic Regression, CART

## 1. 서론

### 1.1 연구의 배경 및 목적

건설 경기의 연이은 침체로 국내 건설기업은 신규 시장의 창출의 일환으로 신재생에너지사업에 적극 진출하려는 움직임을 보이고 있다. 신재생에너지사업은 기존 화석 에너지에 기반한 전력 생산 방식에서 탈피하여 태양광, 풍력, 수력, 매립가스 등과 같은 저탄소 및 친환경 에너지 사업을 일컬으며, 전 세계적으로 확대되고 있는 추세이다.

그러나 신재생에너지 사업은 재래식 발전소에 비해 전력 단가가 높아 수익성이 떨어지고 적용 기술의 출력 효율 및 내구성이 검증되지 않은 문제도 있다. 이에 따라 에너지 수요가 급증하는 개발도상국에서는 활발히 적용되지 않는 한계가 있다(Koo 2013).

청정개발체제(Clean Development Mechanism, 이하 CDM)는 이런 문제를 해결하기 위해 유엔기후변화협약(United Nations Framework Convention on Climate Change, 이하 UNFCCC)에서 고안해 낸 제도 중 하나이다(KEMCO 2007). CDM제도에서는 신재생에너지 사업을 통해 감축된 온실가스의 실적을 배출권의 형태로 국제 탄소 시장에서 거래할 수 있게 해준다. 사업의 투자는 감축의무가 있는 선진국(부속서 I 국가)에서 하지만 실제 사업의 물리적 위치는 개발도상국(비부속서 I 국가)에서 수행하게 된다(Koo et al, 2014). 즉, 선진국이 개발도상국에 온실가스를 감축할 수

\* Corresponding author: Koo, Bonsang, Department of Civil Engineering, Seoul National University of Science and Technology, Seoul 139-743, Korea  
E-mail: bonsang@seoultech.ac.kr  
Received November 26, 2014; revised January 6, 2015  
accepted February 13, 2015

있는 사업을 찾아 수행하면 그 결과 발생한 온실가스 감축실적을 자국의 실적으로 인정받아 자국에 할당된 감축목표에 활용할 수 있는 것이다.

CDM 사업으로 등록되면 감축한 온실가스에 해당하는 탄소배출권(Certificated Emission Reduction, 이하 CER)을 받을 수 있으며 이의 판매를 통해 추가 수익원을 얻게 된다. 이에 따라 해당 국가 및 기업은 개발도상국에서 신재생에너지 사업을 추진하는 인센티브가 생기게 되는 것이다. CDM 사업은 개별 국가의 온실 가스 저감 정책이 아닌 국제적으로 합의된 몇 안 되는 거래제도로서 그 의미가 중요하다(Yim and Yun 2009).

CDM 사업의 보조 없이는 사업성이 떨어져 신재생에너지 사업을 포기할 수 있기 때문에, 개발국에서 CDM 사업의 등록과 이에 따른 배출권의 획득은 사업의 성패에 매우 중요한 요인이 될 수 있다.

그러나 모든 신재생에너지 사업이 CDM 사업으로 자동적으로 승인되는 것은 아니다. CDM 사업으로 승인되기 위해서는 사업의 경제적, 기술적, 환경적 '추가성'(Additionality)을 입증해야 한다. 즉, 온실가스의 실질적 감축과 이를 친환경 기술로 적용하였는지의 여부, 그리고 경제적으로는 보조를 통해 사업성이 생기는 것을 입증해야 승인이 가능하다(Chiba and Kwak 2006).

실제로는 적용 기술, 베이스라인 측정 방법론, 온실 가스 감축량, 사업 내부 수익률 등의 여러 변수에 따라 승인이 판가름 나기 때문에 특정 사업의 승인여부를 파악하기가 어렵다.

본 연구에서는 신재생에너지 사업 중 수력 사업의 CDM 사업으로의 승인여부를 예측할 수 있는 모델을 개발하는 것을 목표로 하였다. 구체적으로 UNFCCC에서 공개적으로 제공하는 CDM 사업의 데이터를 활용하여 데이터 마이닝 분석을 실시하여 CDM 사업으로의 등록 여부를 예측할 수 있는 방법론을 제시하였다. 이와 더불어, 구축된 모델의 평가를 통해 등록 여부에 미치는 핵심 인자를 파악할 수 있었다.

기존 연구에서는 CDM 사업의 성공을 위한 요인 분석(Song 2010; Castro and Michaelowa 2008) 및 전략(Lee 2010)을 제시하고 있으나 예측모델을 구축한 연구는 없다.

본 분석모델을 통해, 신재생에너지 사업자는 CDM 사업의 등록을 위해 사업의 기술적, 환경적 및 경제적 조건을 어떻게 설정할지를 사전에 파악할 수 있고 해당 사업의 승인 여부를 사전에 파악해 볼 수 있게 된다.

## 1.2 연구의 범위 및 방법

UNFCCC에서는 CDM 사업의 개시 해인 1997년부터 현재까지의 총 11089개의 전 세계 CDM 사업에 대한 정보를 공개하고 있으며, 제반 관련 지표를 제공하고 있다.

본 연구에서는 이 중에서 수력사업을 대상으로 분석하는

것으로 국한하였다. CDM으로 신청된 수력 사업은 총 2717개로 전체 사업의 24.5%를 차지하고 있어, 신재생에너지 사업의 주력 사업 중 하나인 것을 알 수 있다. 또한 국내외에서 수력 사업을 CDM 사업으로 한 다수의 사례가 있으며 해외 수출 품목으로서 잠재력이 커 이를 선택하였다.

UNFCCC의 공개 자료에는 개별 사업에 대한 환경적, 기술적, 경제적 데이터와 더불어 등록 및 탈락 여부를 제공하고 있다. 이 데이터를 기반으로 데이터 마이닝 기법인 로지스틱 회귀 분석(Logistic Regression)과 Classification and Regression Tree(이하 CART)분석을 실시하여 개별 영향 인자 중 유의성이 있는 인자를 구분하고 최종적으로 기계 학습을 통한 예측 모델을 구축하였다.

## 2. CDM 사업 개요

### 2.1. CDM 사업의 메커니즘

1997년 교토의정서가 채택되고 2005년에 발효되면서 기후 변화협약의 목적을 달성하기 위하여 어떤 주체에서, 어떤 방식을 통해, 얼마만큼의 온실가스 발생량을 줄이는데에 대한 절차와 온실가스 감축에 대한 법적 구속력을 갖게 되었다. 이러한 온실가스를 감축하기 위한 방안 중 하나가 배출권 거래제도(Cap and Trade)이다. 배출권 거래제도는 온실가스 배출행위에 권리를 부여하여 해당국가의 감축목표치 보다 많은 배출량이 발생했다면 그에 상응하는 배출권을 시장에서 매매할 수 있도록 한 제도이다(KEMCO 2007).

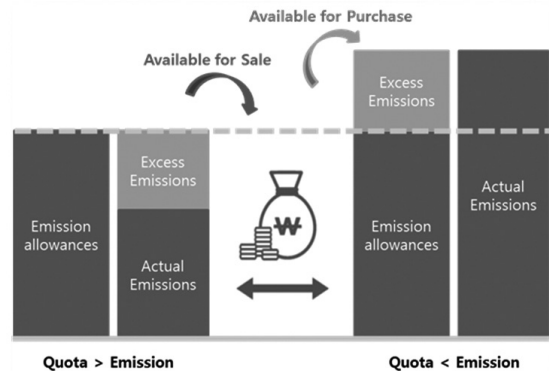


Fig. 1. Emission Trading

(Ref: Greenhouse Gas Inventory & Research Center of Korea)

Fig. 1의 경우처럼 허용량보다 적게 온실가스를 배출 하였을 때 잉여배출량 만큼의 배출권을 판매하는 방식으로, CDM 사업의 경우가 이에 해당한다.

CDM 사업은 온실가스 감축 실적을 인정해주는 교토메커니즘에 속한 감축 수단으로, CDM 사업을 통하여 배출권을 인정받을 수 있다. 하지만 CDM 사업을 실시함에 있어서 검

증 절차를 수행해야 하고 만약 사업의 타당성을 인정받지 못하였을 경우 CDM 사업으로 승인을 받을 수 없게 된다. 이는 곧 탄소배출권에 대한 획득 실패를 의미한다. 따라서 사업 타당성을 인정받는 것은 매우 중요하다.

### 2.2 CDM 사업의 신청 비용 및 승인을

실제로 검증(Validation) 단계에 있는 사업 1,135개를 제외한 현재(2014년 8월1일)까지의 CDM 사업 승인 여부를 살펴보면 승인을 68%(총 11,089개중 승인 7,546개, 탈락 3,543개)를 보여주고 있다(Table 1). CDM 사업은 그 자체로는 수익성이 낮고 탄소배출권을 받았을 때 사업의 수익을 내는 것을 목표로 하기 때문에 68%의 승인율은 결코 높은 숫자로 볼 수 없다.

Table 1. Share of CDM Projects

CDM Projects	# of projects	percentage	Approved	Approval rate
Wind	2785	25.0%	2403	86.2%
Hydro	2717	24.5%	2037	74.9%
Biomass energy	1333	12.0%	640	48.0%
Methane avoidance	1026	9.2%	634	61.7%
Landfill gas	560	5.1%	361	64.4%
EE own generation	745	6.7%	310	41.6%
Solar	398	3.6%	367	92.2%
Others	1525	13.7%	794	52.0%
Total	11089	100%	7546	68.0%

또한 사업 참여자는 등록 시 발생하는 사업 등록비를 UNFCCC에 납입하여야 하고, 향후 모니터링 기간 동안 발급되는 CER 양에 따라 진행수수료를 준비하여야 한다. 수수료는 승인 유무에 관계없이 지출되어야 하는 매몰비용이기 때문에 승인에 실패 하였을 경우 재정적인 손실 또한 피할 수 없다.

그러므로 CDM 사업에 등록이 되려면 타당성 검증에 철저한 준비가 필요하다. 그러나 사업 등록은 여러 기준에 의해서 결정되기 때문에 사전에 승인 여부를 파악하기는 쉽지 않다.

### 2.3 CDM 사업의 추가성 기준

신재생에너지 사업을 CDM으로 승인받기 위한 핵심 사항은 해당 사업의 추가성(Additionality) 입증 여부이다. 추가성은 등록된 CDM 사업 활동이 부재할 경우 발생하는 온실가스 배출량이 해당 프로젝트로 인해 일정 수준 이하로 감축될 경우를 의미한다(KEMCO 2007). 다시 말해, 해당 사업을 수행했을 경우와 재래식 사업의 배출량을 비교해 인위적인 온실가스 배출 감축이 있을 때 추가성이 있는 것으로 보고 있다. 구체적으로 추가성은 다음과 같이 총 3가지로 세분화 된다(Fig. 2).

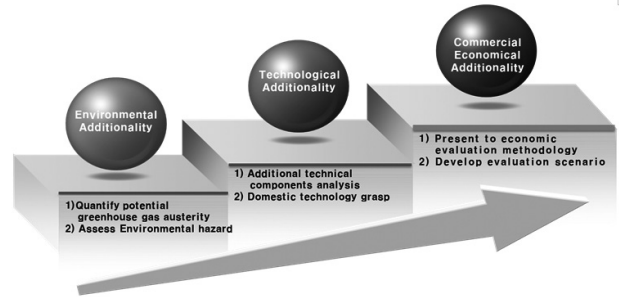


Fig. 2. UNFCCC CDM Regulation Guideline (Ref. KEMCO 2007)

1) 환경적 추가성(Environmental Additionality)  
 해당사업의 온실가스 배출량이 베이스라인(baseline) 배출량보다 적게 배출할 경우, 대상 사업은 환경적 추가성이 있음. 여기서 베이스라인(Baseline)이란 전력을 기존의 재래식 방식(예, 화력 발전소)으로 생산했을 경우 발생할 온실가스의 양을 일컬으며, CDM 사업을 통해 감축되는 온실가스 배출량의 척도가 된다. 만약, 풍력이나 태양광 사업과 같이 탄소 배출량이 전혀 없다면, 베이스라인이 온실 가스 감축량이 된다. 환경보호단체 'Atmosfair' 는 베이스라인에 대한 정의를 Fig. 3과 같이 설명한다.

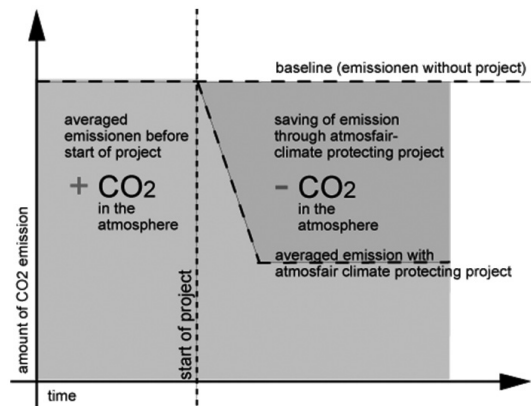


Fig. 3. Quantification of the saved emissions

2) 기술적 추가성(Technological Additionality)  
 CDM 사업에 활용되는 기술은 현재 유치국에 존재하지 않거나 개발되었지만 여러 가지 장애요인으로 인해 활용도가 낮은 선진화된(more advanced) 기술이어야 함. 신재생에너지 기술은 검증되지 않은 신기술이 대부분이기에 사업의 리스크를 증가시켜 일반 사업에서는 활용을 기피하기 마련이다. 따라서, 신기술을 사용한 것을 입증하면 그에 대한 보조를 해준다는 취지이다.  
 3) 경제적 추가성(Commercial/Economical Additionality)  
 기술의 낮은 경제성, 기술에 대한 이해부족 등의 여러 장애

요인으로 인해 현재 투자가 이루어지지 않는 사업을 대상으로 하여야함.

경제적 추가성의 경우, 해당 사업에서 CDM 사업을 통한 CER과 관련한 수입을 제외하였을 때, 대안이 되는 재래식 사업들보다 경제적으로 이익이 없다는 것을 증명하여야 하는 것이 중요하다. 즉, 이 사업은 경제적으로 매력적이지 못하기 때문에 경제적 보조의 필요성을 입증해야 하는 것이다.

상기 기준들을 볼 때, CDM 사업의 등록 여부는 ① 배출저감량, ② 신기술의 도입 여부, 그리고 ③ CER의 판매로 인한 수익성 개선 수준에 따라 달라진다는 것을 추정할 수 있다.

특히 경제적 추가성은 CER 판매 이전과 그 이후의 수익성의 차이가 클수록 CDM 사업으로서 적합하다는 것을 추정해 볼 수 있다. 이는 CDM 사업의 핵심 영향 요인들을 사전에 구분해 내는데 중요한 잣대가 된다.

이는 역으로 CDM 사업에 신청된 사업들은 승인되는 것이 사업에 중요하다는 것을 보여준다. 즉, 신재생에너지 사업 중 수익성이 좋은 사업은 굳이 등록비용을 감당하면서까지 신청할 이유가 없다. 그러나 수익성이 낮은 사업은 그 만큼 CDM 으로부터 얻는 수익이 중요한 보조수단이 될 수 있을 것이다.

### 3. 수력 CDM 사업 개요

#### 3.1 수력 CDM 사업 현황

본 연구에서는 수력(Hydro power) 사업에 국한하였다. 수력 사업은 국내에서도 꾸준히 진행되고 있다. 2014년 9월30일 까지 온실가스검증원에 의해 조사된 국내 CDM사업 현황은 한국수자원공사의 경기, 경북, 전남 소수력발전과, 남동발전의 삼척, 영흥 소수력발전 사업, 중부발전의 보령 소수력발전 등 여러 기관들이 참여하여 성과를 내고 있다.

또한 최근에는 수자원공사가 파키스탄 파트린드 수력발전 사업을 CDM 으로 등록하는 등 해외 수출 품목으로서 그 잠재력이 크다.<sup>3)</sup> 따라서 수력사업의 승인여부에 대한 국내의 관심이 높다.

Table 1과 같이 UNFCCC에 CDM으로 승인되었거나 탈락된 수력 사업은 총 2717개로 전체사업의 24.5%를 차지하고 있으며, 이중 승인된 사업은 2037개 탈락된 사업은 680개로 승인을 74.9%를 보여주고 있다.

#### 3.2 수력 CDM 사업 데이터

수력 사업과 관련된 구체적 데이터는 UNFCCC에서 제공하는 'CDM pipeline'자료<sup>4)</sup>에서 구할 수 있다. 본 자료에는 모든 CDM 사업의 환경적, 기술적, 경제적 데이터를 제공하며

지속적으로 사업의 등록, 검증 및 승인여부에 대한 업데이트를 해주고 있다.

본 자료 중에서 승인 일자(Credit Start) 및 검증기관(Validator)과 같은 행정 정보는 제외시키고 추가성과 관련된 변수들은 모두 데이터 마이닝의 분석 대상으로 선정하였고 Table 2에 정리되어 있다.

Table 2. Characteristic of Variables

Variables	Definition
<b>Environmental Additionality</b>	
Methodology	Using Methodology
Annual Emission	Amount of annual Greenhouse gas reductions
Expected 2012	Amount of reductions expected to be accumulated by 2012
Expected 2020	Amount of reductions expected to be accumulated by 2012~2020
<b>Technological Additionality</b>	
Small_hydro	Using small hydropower or not
Sub_type	Types of generation
MWh	Power plant capacity
<b>Commercial/Economical Additionality</b>	
Investment	Investment cost
Investment/Emission	Emission per Investment
Investment/Generation	Electricity capacity per Investment
Emission/Generation	Electricity capacity per Emission
IRR_excl_CER	IRR without CER
IRR_incl_CER	IRR with CER
<b>Normal variable</b>	
Region	Location
Sub_region	Sub Location
<b>Input variable</b>	
Status	Approval or Failure

#### 3.2.1 환경적 추가성과 관련된 데이터

1) Methodology: 해당 사업의 온실가스 감축량을 산정하기 위한 방법론이다.

베이스라인 배출량을 높게 설정할수록 감축량에 대하여 얻는 크레딧(CER)의 양이 많아지게 되기 때문에 사업자는 베이스라인 배출량을 높게 잡으려는 성향을 보인다. 그러므로 베이스라인을 객관적으로 설정할 수 있는 방법론이 필요하며 UNFCCC에서는 사업유형에 따라 검증된 다수의 방법론을 정의해 놓고 있다. 구체적으로 사업 규모에 따라 대규모(AM<sup>5)</sup>, ACM<sup>6)</sup> 소규모(AMS<sup>7)</sup>) 및 로 구분되고 기술의 적용 방식에 따라서도 나누어진다. 이처럼 탄소감축량의 산정 방법

3) "K-water, 파키스탄 파트린드 수력발전 CDM UN등록" (2013.04.13.)

4) 자료는 CDM pipeline(<http://www.cdmpipeline.org/>)에서 지속적으로 업데이트가되며 엑셀 형태로 다운이 가능하다.

5) Approved Methodology

6) Approved Consolidated Methodology

7) Approved Methodology, Small Scale

론이 규정되어 있으며 사업자는 진행하고자 하는 사업에 적용 가능하다고 판단되는 방법론을 선택하면 된다.

분석의 용이함을 위하여 수력사업에 사용되는 방법론을 Table 3처럼 종합하여 종류별로 6가지로 분류하였다.

Table 3. Types of Methodologies

Methodology	Number	Share	Approval	Approval rate
ACM2	1339	49.2%	1055	79%
AM_others	21	0.7%	8	38%
AMS-I.D.	1331	48.3%	959	72%
AMS_others	8	0.2%	6	75%
ACM+AMS	16	0.5%	7	44%
AMS+AMS	2	0.1%	2	100%
Total	2717	100%	2037	

선택한 방법론에 따라 온실가스 감축량이 산정되므로 방법론은 '환경적 추가성'을 입증하는데 사용된다. 더불어 Table 3에 제시된 바와 같이 적용된 방법론에 따라 승인 여부가 달라지는 것을 알 수 있다. 이에 따라 방법론은 승인 여부에 중요한 영향 요인 중 하나인 것으로 선정 하였다. 실제로(Han 2006)에 의하면 방법론이 승인 여부에 가장 큰 요인 중에 하나로 조사되었고, 신규 방법론보다는 이미 검증된 방법론의 승인율이 높은 것으로 조사되었다.

2) Annual Emission: CDM 사업은 CER을 부여받는 연수(Credit Period)를 7년 또는 10년으로 구분하는데, 이 기간 동안 매년 평균적으로 감축할 것으로 예상되는 배출량이다.

3) Expected 2012: UNFCCC에 의하여 지정된 1차 감축기간(2008~2012)에 해당하는 2012년까지의 누적 감축량

4) Expected 2020: UNFCCC에 의하여 지정된 2차 감축기간(2013~2020)에 해당하는 2020년까지의 누적 감축량

일반적으로 온실가스 감축량이 클수록 승인율이 높을 것으로 판단되어 평균 및 누적 감축량 모두 중요한 영향 요인으로 선정하였다.

### 3.2.2 기술적 추가성과 관련된 데이터

1) Small\_hydro: 발전하는 시설용량이 10,000kW 이하의 설비를 소수력발전이라고 한다(Lee and Lee 2005). 이러한 소수력발전은 519개로, 전체의 19%를 차지하고 있고 승인율(승인 : 336개, 탈락 : 183개)은 65%에 해당하고 소수력발전이 아닌 수력발전은 2189개 중 승인율(승인 : 1701개, 탈락 : 497개)은 77%에 해당한다.

2) Sub\_type: 수력사업은 발전 방식에 따라 Existing Dam, Higher efficiency hydro power, New dam, Run of river 등의 4가지로 구분된다. CDM pipeline에는 'Run of river', 즉, 흐르는 물을 이용한 발전방식이 71%로 가장 많은 비중을 차지하고 있다. Table 4에서와 같이 4가지 발전 방식에 따라서

도 승인율의 차이가 나는 것을 볼 수 있다. 따라서, Sub\_type도 영향인자 중 하나로 추정되어 '기술적 추가성'을 입증하는 요인으로 선정하였다.

Table 4. Hydro Project's Type

Sub_type	Number	Share	Approval	Approval rate
Existing Dam	133	4.8%	91	68%
Higher efficiency hydro power	5	0.2%	1	33%
New dam	645	24%	486	75%
Run of river	1934	71%	1459	75%
Total	2717	100%	2037	

3) MWh: CDM 사업으로 승인받으려는 발전소를 운영하였을 때 시간당 생산할 수 있는 전력량을 의미한다.

### 3.2.3 경제적 추가성과 관련된 데이터

1) Investment: 해당 발전소를 건설하는데 투자한 비용(Millon US 달러)이다.

2) Investment/Emission: 투자대비 온실가스 감축량을 의미한다. 온실가스를 감축시키기 위하여 어느 정도의 투자를 하였는지 알 수 있다.

3) Investment/Generation: 발전량 대비 투자량을 의미한다. 투자한 금액에 대하여 발전량이 얼마나 되는지 판단할 수 있는 지표로서 활용할 수 있다.

4) Emission/Generation: 발전량 대비 온실가스 감축량이다. 발전효율과 온실가스 감축량의 상관관계를 고려할 수 있는 지표로서 활용할 수 있다.

5) IRR\_excl\_CER: CER 수익을 고려하지 않았을 때의 IRR(Internal Rate of Return, 내부수익률)이다.

6) IRR\_incl\_CER: CER수익을 고려 한 내부수익률이다. CER 수익을 고려하지 않는 'IRR\_excl\_CER'보다 항상 높은 값을 갖는다.

이상의 데이터는 사업의 경제적 추가성을 파악하는데 유효할 것으로 추정되어 선정하였다.

### 3.2.4 기타 일반 데이터

1) Region: 수력 CDM 사업을 진행하고 곳의 지리적 특성을 반영해 주기 위하여 대륙별로 분류를 실시하였다.

2) Sub\_Region: 대륙별 분류를 한 단계 더 세분화 하여 사업 참여율이 높은 지역이 어느 곳인지 알기위하여 지역별 구분을 실시하였다. 참여율에 대한 수치는 다음 Table 5에서 확인할 수 있다.

3) Status: CDM 사업의 승인여부를 구분한다. 승인은 'Registered'로, 탈락은 'Rejected', 'Validation terminated', 'Withdrawn'등으로 구분되어 있다. 데이터 마이닝을 통해 향후 CDM 사업에 대해 예측하고자 하는 출력변수가 된다.

Table 5. Regional Hydro Power Application

Region	Sub_Region	Number	Approval	Approval rate
Africa	Africa	21	15	71%
Asia	East Asia	1716	1332	78%
	Southern Asia	273	171	88%
	Southeast Asia	285	250	63%
Europe & Central Asia	Europe	10	8	80%
	Central Asia	22	7	31%
Latin America	Central America	71	49	69%
	North America	11	6	55%
	South America	308	185	60%
total		2717	2023	74%

## 4. 데이터 마이닝 개요

### 4.1 데이터 마이닝 개요

데이터 마이닝(Data Mining)이란 분석하고자 하는 자료의 경향과 규칙을 발견하기 위하여 데이터를 대량으로 수집하여 그 데이터로부터 자동화 혹은 반자동화 도구를 활용하여 탐색하고 분석하는 과정이다(Park and Koh 2006).

데이터는 입력변수와 출력변수로 구분할 수 있는데 입력 변수란 예측을 위한 근거가 되는 변수들을 의미하고 출력 변수란 입력변수를 통하여 알고자 하는 결과를 의미한다. 출력 변수의 유무에 따라 데이터 마이닝을 지도학습(Supervised Learning)과 자율학습(Unsupervised Learning)으로 구분할 수 있다.

본 연구에서 사용하는 지도학습은 회귀분석(Regression)과 분류(Classification)로 나누어진다. 회귀란 입력들로부터 연속형(Continuous) 출력을 예측하는 방식이고, 분류란 입력들로부터 범주형 출력을 예측하는 방식이다. 지도학습에 의한 분석을 진행하는 경우 Training Data와 Test Data가 존재 하여야 한다.

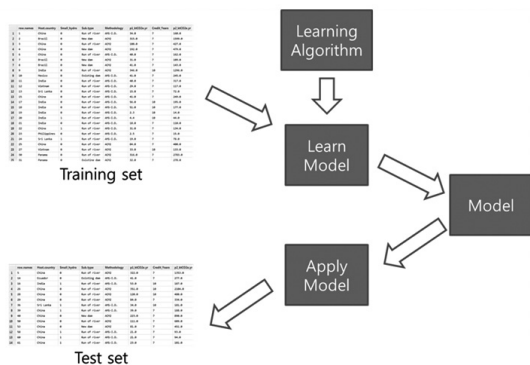


Fig. 4. Supervised Learning

Fig. 4와 같이 구축된 데이터(Training set)를 통하여 입력 변수와 출력변수간의 관계를 학습하고 검증용 데이터(Test set)를 이용하여 출력값이 알려지지 않은 새로운 데이터에 대

한 값을 예측하는 것으로 지도학습을 위한 알고리즘에는 회귀분석(Regression analysis), 의사결정나무(CART), 판별분석(Discriminant analysis), 신경망분석(Neural network) 등이 있다.

데이터가 축적될수록 추출할 수 있는 정보와 그 가치의 질은 증가한다. 데이터로부터 알고자 하는 내용을 출력변수로 지정하고 상관성이 예측되는 입력변수를 이용한다면 분석기법의 특징에 따라 변수의 유의성을 확인할 수 있다. 이에 따라 유의미한 변수의 조합으로 실제 제시된 데이터와 비교하여 정확성의 정도를 확인할 수 있다.

## 4.2 Logistic Regression과 CART분석

### 4.2.1 로그 회귀분석(Logistic Regression Analysis)

로그 회귀분석은 데이터의 분석 결과가 두 집단(Binary Outcome)으로 나누어지는 경우에 각각의 관측되는 데이터들이 어떤 집단으로 분류될 수 있는가를 분석하고 이를 예측하는 모형을 개발하는데 사용되는 알고리즘 방법이다. 즉, 어떤 사건이 발생하는지 안하는지에 대하여 발생확률을 예측하여 확률이 0.5보다 크면 사건이 일어나고 0.5보다 작으면 그 사건이 일어나지 않는 것으로 예측하게 되는 방식이다.

이 분석은 입력변수로 범주형 및 연속형 변수 모두 사용할 수 있다는 장점이 있다.

로그 회귀모형은 다음 식(1)과 같이 정의된다.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

좌측항을 로짓(logit)이라고 부르는데, 로짓을 출력변수로 정의하고 로짓과 n개의 입력변수와의 관계를 선형함수로 모형화한다.

p는 하나의 데이터에서 X에 해당하는 변수들을 위의 식에 삽입했을 때, 그에 따른 결과가 예측하고자 하는 집단에 속할 확률이고 반대로 1-p는 집단에 속하지 않을 확률이다. 이 값을 n개의 입력변수를 이용하여 예측하고자 하는 것이 로그 회귀분석인 것이다.

회귀계수가 양의 값을 가질 때에 대응하는 변수의 값이 높아질수록 특정집단에 속할 확률이 높아지고 음의 값을 가질 때 특정집단에 속할 확률은 낮아진다. 결국 이 계수들을 해석하면 어떤 변수가 특정집단에 속할 확률이 높은지 혹은 낮은지를 알 수 있다.

로그 회귀분석은 대상들이 두 집단 혹은 그 이상의 집단으로 나누어진 경우, 개별 관측 값들이 어느 집단으로, 어떻게 분류될 수 있는가를 분석하고 이를 예측하는 모델을 모형화한다. 입력변수의 숫자형, 범주형 여부에 상관없이 분석이 가능하며 입력변수가 출력변수에 미치는 영향을 수치화할 수 있는 장점이 있다.

### 4.2.2 CART분석(Classification and Regression Trees : CART)

CART분석의 목적은 새로운 데이터에 대하여 분류를 실시하거나 해당 범주의 값을 예측 하는 것이다. CART 역시 입력 변수로 범주형 및 연속형 변수를 모두 사용할 수 있다는 장점을 가지고 있다.

알고리즘의 분류방법은 우선 루트노드를 만들고 데이터 중에서 최적의 분할기준을 선정해 그 기준을 기반으로 하여 하위에 두 개의 노드로 데이터들을 분류한다. 그리고 하위노드들을 같은 방식으로 확장시켜 나가는 것이다. 이 때 데이터들 간의 특성 및 형태가 유사한 동질성(Homogeneity)이 높은 그룹들을 판별하기 위하여 데이터의 분할 기준을 불순도(Impurity)가 낮은(동질성이 낮은 그룹과 혼합정도가 가장 적은) 때로 하는 것이 중요하다.

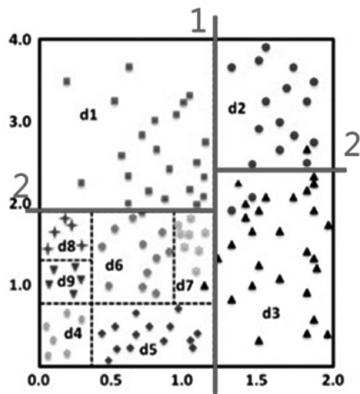


Fig. 5. CART Model

Fig. 5와 같이 전체적인 데이터들의 집합에서 최초로 불순도가 가장 낮게 나누어줄 수 있는(기준 1)을 적용하여 d2, d3과 d1들을 분류하고 또다시 같은 방식으로 하위노드를 확장(기준 2,3)시킨다. 이러한 방식으로 개별적인 특징을 가진 분류를 찾아낼 수 있다.

CART분석은 의사결정나무(Decision Tree) 알고리즘으로써 변수의 특성에 따라 데이터를 구분하고, 구분된 집단의 확률을 계산하여 집단을 분리하는 방법으로, 분류와 예측의 과정이 Tree구조에 의하여 도표화된다. 때문에 변수의 중요도가 쉽게 파악된다는 점이 장점이다. 또한 Tree구조로부터 변수의 값에 대하여 분리되어 표현되기 때문에 정량적인 값을 제시해 줄 수 있다. 하지만 연속적인 변수를 비연속적인 값으로 취급하기 때문에 분리의 경계점 부근에서 오류가 발생할 확률이 높고 입력변수가 출력변수에 미치는 효과에 대해서는 알 수 없다는 단점이 존재한다.

### 4.2.3 상호 결과비교의 필요성

본 연구에서 로그 회귀분석(Logistic Regression)과 CART

분석(Classification And Regression Trees)분석을 동시에 실시하는 이유는 각 방식의 결과를 비교 분석함으로써 결과에 대한 해석력을 높일 수 있기 때문이다.

로그 회귀분석의 경우 개별 입력변수의 유의성을 분석할 수 있고 결과가 두 가지로만 나누어지는 이항 분석이 가능하다는 장점이 있지만 입출력 변수간의 선형관계를 가정하는 단점이 있다. 반면에 CART분석의 경우 변수를 비선형 형태로 가정하며 결과에 대한 값을 정량적으로 구분해줄 수 있다.

이처럼 두 가지 분석을 통해 상호 보완적으로 결과에 대한 정확성과 신뢰성을 향상시킬 수 있다.

### 4.3 정확성 검증

구축된 모델을 사용하여 유의성 있는 변수를 분별하였더라도 모델자체의 신뢰도가 떨어진다면 실패한 분석이 될 수 있다. 본 연구에서는 Accuracy와 'Area Under the Curve(AUC)'를 산정하여 모델의 정확성 및 예측성을 검증하였다.

		Predict	
		Positive	Negative
Actual	True	TP	TN
	False	FP	FN

Fig. 6. Confusion Matrix

Accuracy 및 AUC를 산정하기 위해서는 Fig. 6과 같이 Confusion Matrix가 사용된다. Confusion Matrix는 실제와 예측 간의 관계를 나타내며 주로 알고리즘의 성능을 평가하는 지표로 많이 사용된다. True/False는 실제의 값의 예측이 일치하는 경우와 일치하지 않는 경우를 뜻하고, Positive/Negative는 예측한 값이 일치하는 경우와 일치하지 않는 경우를 뜻한다.

이때 Accuracy는 구축모델을 통하여 실제와 예측이 모두 정확하게 예측된 데이터(True Positive Rate + False Negative Rate)의 수를 모든 데이터의 수로 나누어준 것  $[(TP+FN)/(TP+TN+FP+FN)]$ 으로 예측의 정확성을 알 수 있는 지표로 사용된다.

AUC는 실제로 승인된 데이터가 승인될 것으로 예측하는 민감도(Sensitivity=True Positive Rate,  $TP/(TP+FP)$ )와 탈락된 데이터를 탈락할 것으로 예측하는 특이도(Specificity=False Positive Rate,  $1-TN/(TN+FP)$ )를 계산한 후 Fig. 7과 같은 그래프를 통하여 산정된다. 이때의 그래프를 ROC Curve라고 한다.



Fig. 7. ROC Curve  
(Ref. Jun 2012)

그래프의 형태가 왼쪽 꼭대기에 가깝게 그려질수록 예측력이 우수하다. 이것은 ROC Curve 아래의 면적(AUC)이 1에 가까우면 모델의 예측력이 우수하다는 뜻이다. 예측력의 정확도는 Swets(1988)에 의하여 AUC 수치에 따라 다음 Table 6과 같이 구분된다.

Table 6. Guideline of AUC

Range	Accuracy
AUC=0.5	Non-informative
0.5<AUC<0.7	Less accurate
0.7<AUC<0.9	Moderately accurate
0.9<AUC<1	Highly accurate
AUC=1	Perfect tests

## 5. 수력 CDM 사업의 데이터 마이닝

### 5.1 데이터 전처리

#### 5.1.1 입력 및 출력변수 선정 및 정리

본 연구의 목적은 수력 CDM 사업에 대한 승인여부를 로그 회귀 분석 및 CART 분석 기법을 통해 예측 모델을 구축하는 것이다.

본 분석을 위해 Table 7에 제시된 바와 같이 Status(승인 여부)를 이항 출력변수로 하고 나머지 변수들을 입력변수로 사용하였으며, 개별 변수들을 범주형(categorical) 및 숫자형(numerical) 변수로 구분하였다.

#### 5.1.2 누락 정보 보정을 위한 Imputation

수집된 변수들의 실제 값들은 CDM pipeline 및 사업승인을 위하여 제출된 보고서(PDD)로부터 얻은 정보들이다. 대부분 데이터는 두 자료 중에서 취합이 가능하였다. 그러나 IRR\_incl\_CER의 경우 누락된 값이 다수 존재하여, 이 경우 R program의 mice package에 있는 'imputation' 명령어를 사용하여 보정을 하였다. MICE(Multivariate Imputation by Chained Equations)는 누락된 데이터에 대하여 조건부 분포

Table 7. Classification of variables

categorical data		
	variable	type
input variable	Methodology	6 Methodologies
	Small_hydro	Yes=1, No=0
	Sub_type	4 Sub_types
	Region	4 Regions
	Sub_region	9 Sub_regions
output variable	Approval	Approval=1, No=0
numerical data		
	variable	unit
input variable	Annual Emission	ktCO2e
	Expected 2012	ktCO2e
	Expected 2020	ktCO2e
	MWh	MW/h
	Investment	millionUSD
	Investment/Emission	kUSD/tCO2e
	Investment/Generation	USD/kW
	Emission/Generation	ktCO2e/MWh
	IRR_excl_CER	%
	IRR_incl_CER	%

가 존재한다면, 그 분포에 해당하는 특정한 변수를 만들어낼 수 있는 방법으로서, imputation 시 대표적으로 활용되는 기법이다(Van Buuren 1999).

### 5.1.3 Training 및 Test set 분류

앞서 설명한 바와 같이 수력 CDM 사업은 총 2717개이며 이중 승인된 사업은 2023개, 탈락된 사업은 694개로 승인을 74.4%를 보여주고 있다. 이를 균등하게 Training 및 Test set로 나누기 위해 R의 caTools packages의 'split' 명령어를 사용하였다. 이 명령어는 전체 데이터를 무작위로 나눠주되 승인, 탈락 사업들을 균등하게 분배해 준다.

데이터의 70%에 해당하는 1785개의 데이터를 Training Set로 무작위로 추출하여 모델구성에 이용하고 나머지 30%의 데이터는 Test Set로 완성된 모델을 검증하는 방식으로 분석을 진행하였다.

## 5.2 Logistic Regression 분석

### 5.2.1 Logistic Regression 분석 절차

구축된 Training set를 기반으로 로그 회귀분석을 실시하였다. 로그 회귀분석은 출력변수에 영향을 주는 유의한 입력 변수들을 구분해 준다. 이에 따라 Table 7에 제시된 입력 변수(input variable)들을 첫 회에는 모두 넣고 회귀분석 실시 후, 매회마다 유의하지 않은 변수들을 제외하면서 최종적으로 유의한 변수들만 남을 때까지 반복 수행하였다. 이때 유의수준(p value)은 0.05로 수행하였으며 R 프로그램의 glm package를 이용하였다.



### 5.2.2 Logistic Regression 분석 결과

로그 회귀분석 결과 최종 모델의 유의한 변수들은 Table 8 과 같다. 여기서 'Estimate'가 공식 (1)의 회계 계수에 해당되며 'Pr(>|z|)'는 p-value에 해당된다.

앞서 언급한 대로 로그 회귀분석에서 입력변수의 계수 크기 및 부호는 출력변수와의 단조(monotonic) 관계를 정량화한 것이다. 그러므로 Table 8에서 Estimate가 양인 변수들, 즉, Expected 2020 및 Investment/Emission은 승인 확률과 양의 선형 관계에 있는 인자들이다. 그 외의 변수들은 승인 확률과 음의 선형 관계에 있는 인자들이다.

각각에 대한 해석은 다음과 같다.

1) Expected 2012: 이는 2012년까지 누적된 온실가스 예상 감축량으로서 값이 클수록 승인 여부 확률이 높아야 논리적으로 타당하다. 그러나 실제로 대부분의 사업들이 운영단계에 들어가서 실제로 탄소저감을 하는 시점은 2012년 이후이다. 따라서, 2012년 이전에는 탄소 저감량이 작고 2020년 이후에 저감량이 큰 사업들이 승인 확률이 높게 나왔으며 이에 따라 Expected 2012와 승인 여부(Status)의 관계는 유의하나 음의 관계로 드러났다.

2) Expected 2020: 본 변수는 2020년까지 누적된 온실가스 예상 감축량으로서 그 값이 클수록 승인 확률이 높은 것으로 드러났다. 상기 언급한대로 대부분의 사업들이 2012년부터 2020년 사이에 본격적인 탄소저감이 계획되어 있어 본 변수가 클수록 승인 확률 역시 큰 것으로 드러났으며 이는 저감량이 큰 사업일수록 승인 확률도 높아질 것이라는 연구의 전제와 상통한다.

3) MWh: 시간당 전력생산량이 작을수록 승인가능성이 높은 것으로 분석되었다. 이는 규모가 상대적으로 작은 사업일수록 보조를 필요로 하고 이에 따라 승인이 잘되는 것을 반증하는 것으로 해석된다.

4) Investment/Emission: 본 변수는 온실가스 감축량 대비 투자금액을 나타내는 것으로 감축의 효율성을 금액으로 표시한 것이다. 이 역시 승인여부와 음의 상관관계를 보였으며 이는 탄소를 저감하는데 투자비용이 큰 사업들, 즉, 수익성이 낮은 사업들이 승인 여부가 높은 것으로 드러났다.

5) IRR\_excl\_CER: IRR값이 작을수록 사업승인에 유리하다는 분석결과를 확인할 수 있었다. 즉, 사업성이 낮은 사업들이 CER의 판매를 통한 추가 수익의 필요성이 가장 크므로 승인여부가 큰 것으로 해석된다.

이외의 변수들, 특히 방법론(Methodology), 권역별 국가 (Region, Sub-region), 적용 기술(Sub-type, Small hydro)에 해당하는 변수들은 유의하지 않는 것으로 드러났다.

로그 회귀분석은 이처럼 유의한 변수들을 선별해 주는 장점은 있으나 각 변수의 경계 값(boundary value)을 파악하기

는 쉽지 않다. 이의 보완을 위해 다음과 같이 CART 분석을 실시하였다.

Table 8. Logistic Regression Analysis

	Estimate	Std. Error	z value	Pr(> z )
Intercept	1.448 × 10	2.155 × 10 <sup>-1</sup>	6.718	1.84 × 10 <sup>-11</sup>
Expected 2012	-3.651 × 10 <sup>-3</sup>	4.358 × 10 <sup>-4</sup>	-8.378	2.00 × 10 <sup>-16</sup>
Expected 2020	1.172 × 10 <sup>-3</sup>	1.690 × 10 <sup>-4</sup>	6.934	4.08 × 10 <sup>-12</sup>
MWh	-1.987 × 10 <sup>-6</sup>	3.077 × 10 <sup>-7</sup>	-6.459	1.05 × 10 <sup>-4</sup>
Investment/Emission	3.740 × 10 <sup>-4</sup>	1.790 × 10 <sup>-4</sup>	2.089	0.0367
IRR_excl_CER	-4.54 × 10 <sup>-2</sup>	2.285 × 10 <sup>-2</sup>	-1.987	0.0469

### 5.3 CART 분석

#### 5.3.1 CART 분석 절차

구축된 Training set와 로그 회귀분석에서 활용된 동일한 변수(Table 7)를 가지고 CART 분석을 실시하였다. CART는 개별 변수들을 유의성 기준으로 구분하지 않고 유사한 데이터끼리 군집화 및 세분화하는데 변수들을 사용하며, 이때 군집 간의 구분에 활용된 변수의 경계 값을 제공해 준다. CART에서는 이에 따라 군집(bin)의 크기를 설정할 수 있다. 본 분석에서는 R 프로그램의 rpart, rpart.plot package에서 default로 주어진 bin 크기로 분석을 실시하였다.

#### 5.3.2 CART 분석 결과

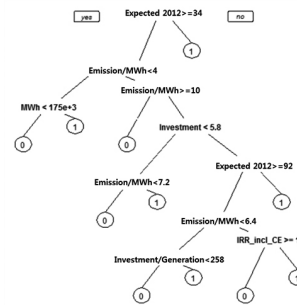


Fig. 8. CART Analysis

Fig. 8에서와 같이 가장 상위 기준으로 나누어지는 변수는 2012년까지의 누적 감축량(Expected 2012)이며 이 값이 34ktCO<sub>2</sub>e 보다 작을 경우 승인 여부가 높은 것으로 드러났다. 이는 상기 로그 회귀분석과 같은 결과를 보여주고 있다. 즉, 2012년까지는 감축량이 작거나 없는 사업이 많고 2012년부터 2020년까지의 누적 감축량이 큰 사업이 많다는 것을 시사한다.

이와 같은 결과는 'Expected 2012' 변수의 데이터 재검토를 통해서 확인이 된다. Table 9에서와 같이 누적 감축량이

34ktCO<sub>2</sub>e 보다 작은 사업들이 90%로 높은 승인율을 보이고 있다.

Table 9. Classification of Expected 2012

Total 2717 Projects				
	Over 34 ktCO <sub>2</sub>		Lower 34 ktCO <sub>2</sub>	
2717	1690		1027	
	Approval	Fail	Approval	Fail
	1114	576	923	104
Approval rate	66%		90%	

두 번째 변수는 Emission/Generation으로서 Table 10에서와 같이 4ktCO<sub>2</sub>e/MWh 이상이면 승인 여부가 높은 것으로 드러났다.

Table 10. Classification of Emission/Generation

Total 2717 Projects				
	Over 4 ktCO <sub>2</sub> /MWh		Lower 4 ktCO <sub>2</sub> /MWh	
1690	1518		172	
	Approval	Fail	Approval	Fail
	1054	464	112	60
Approval rate	70%		34%	

이 변수의 값이 4ktCO<sub>2</sub>e/MWh 보다 크기 위해서는 로그 회귀분석의 결과와 동일하게 탄소 저감량이 크거나, 시간당 전력생산량이 작은 사업일수록 승인율이 높을 것으로 해석 가능하다. 이는 상대적으로 소규모 사업이 보조를 필요로 하기 때문에 승인율이 높다고 예측되며, CART를 통해 구체적인 값의 추정이 가능해졌다.

## 6. 소결

### 6.1 유의성을 갖는 인자 비교

수집한 데이터를 두 가지 데이터 마이닝 기법을 이용하여 분석하고 유의성을 갖는 인자만을 추출하여 확인해 보았다. 그 결과 승인여부에 영향을 미치는 인자는 Table 11과 같이 대부분 공통적으로 해당하는 변수였고 이는 두 모델의 분석에 대하여 신뢰도를 갖게 해 주었다. 두 분석을 통하여 추출된 특정 변수들을 분석해 보면, 2012년 까지 탄소 감축량이 작거나 없는 사업(0~34ktCO<sub>2</sub>e)은 승인율이 높았으며 2020년 까지 감축 대상에 속해있는 사업은 누적 감축량이 클수록 승인율이 높다. 즉 연간 감축량이 크면 클수록 승인에 유리하다는 의미이다. MWh의 경우 연간감축량 대비 4분의 1 수준 보다 작을수록 승인에 유리할 것이다. 이것은 상대적으로 소규모의 사업에 대하여 승인에 대한 가능성이 높다는 것을 의미한다.

또한 수력사업의 경우에는 방법론(Methodology)이나 지역적 특성(Region, Sub\_Region), 또는 적용기술(Sub\_type, Small\_hydro)보다는 연간 탄소저감량(ktCO<sub>2</sub>)과 발전용량(MWh)이 승인에 더욱 중요하게 작용하는 것을 알 수 있다.

Table 11. Variables of Having Significance of results

Environmental Additionality	
Logistic Regression	CART
Expected 2012	Expected 2012
Expected 2020	
Technological Additionality	
MWh	MWh
Commercial/Economical Additionality	
	Investment
Investment/Emission	Investment/Emission
IRR_excl_CER	
	IRR_incl_CER
	Emission/MWh

### 6.2 Accuracy, AUC 비교

분석모델의 신뢰도를 판단하는 Accuracy와 AUC의 결과는 Table 12에 정리되어 있다.

두 분석 모두 Accuracy와 AUC값이 비슷한 범위에 속하였다. 분석결과가 0.7<AUC<0.9 사이에 존재하기 때문에 Table 6의 기준에 따라 중등도의 신뢰도(moderately accurate)를 구축했다고 할 수 있다. 더 높은 신뢰도의 구축을 위해서는 향후 진행될 수력 CDM 사업 데이터의 추가 분석과 'randomForest'와 같은 심층적인 분석이 필요할 것으로 판단된다.

Table 12. Model's Accuracy and AUC

CDM Projects		Logistic Regression		CART	
		Predict			
		False	True	False	True
Actual	False	30	174	63	209
	True	27	584	48	767
Accuracy		0.7533		0.7635	
AUC		0.7674		0.7231	

### 6.3 예측 모델의 활용 및 적용성

구축된 모델은 CDM으로 등록을 요하는 신규 수력 사업의 승인여부를 예측하는데 활용될 수 있다. 만약 모델에서 승인이 불가한 것으로 판별됐다면, 그 원인을 더욱 쉽고 신속하게 파악할 수 있다. 즉, Table 11의 유의한 인자들 중에서 해당 사업이 결여되는 항목을 파악하고 이의 개선을 통해 CDM 사업으로의 등록 확률을 높이는데 활용될 수 있게 된다.

## 7. 결론

본 연구에서는 CDM 으로 등록된 수력사업에 대해서 데이터 마이닝 기법을 적용하여 승인여부 판별 모델을 구축하고 동시에 주요 영향인자를 파악하였다.

승인 판별모델을 구축하기 위해 로그 회귀분석과 CART분석을 사용하였으며, 두 분석은 상호 보완적으로 작용하여 유의한 변수들을 선정해주고, 승인에 대한 구체적인 경계치를 제시하였다.

분석 결과 수력사업은 CDM 사업 승인에 있어서 기존에 유리하게 작용한다고 알려진 특정 방법론(Methodology)사용이나, 적용 기술(Sub-type, Small\_hydro) 보다 높은 연간 탄소 감축량(Annual Emission) 및 낮은 설비용량(MWh)등이 사업 승인에 유의하게 영향을 미치는 것으로 나타났다. 이는 수력사업의 경우 상대적으로 소규모인 사업들이 승인에 유용하게 작용하는 것을 시사한다.

구축된 모델의 정확성은 중등도의 신뢰도를 보였다. 향후 수력 사업으로 등록된 CDM 사업이 증가되고 데이터의 양이 축적되면, 정확성 및 신뢰도가 향상된 모델구축이 가능할 것으로 판단된다.

재래식 발전에 비하여 사업의 수익성이 떨어지는 신재생에너지 사업은 CDM 사업 승인을 통하여 CER을 발급받아 수익성을 향상 시키는 것이 중요하다. 본 연구에서 구축된 승인 판별 모델을 통해 사전에 승인 여부를 예측하고 승인 가능성을 높여줄 수 있도록 사업을 조정하는데 활용될 수 있을 것으로 기대한다.

## 감사의 글

이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다.

## References

Castro, P., and Michaelowa, A. (2008). "Empirical analysis of performance of CDM projects: FINAL REPORT." Climate Strategies Report.

Chiba, M., and Kwak, S. H. (2006). "The Meaning of Additionality in CDM Project and the Potential Impacts of Additionality on the CDM Project". *Journal of Energy & Climate Change*, 1(2), pp. 92-98.

Han, S., H. (2006). "Application of Approved Baseline Methodologies for CDM Projects in Korea(Case Study: Landfill Gas-to-Electricity Projects)", *Korea*

*Energy Management Corporation*.

Jun, C., H. (2012). "Data Mining Techniques." Seoul, Korea: Han Na Rae.

KEMCO (2007). "Clean Development Mechanism." *Korea Energy Management Corporation*, KOTRA.

Koo, B. (2013). "Evaluating the Economic Feasibility of Green Construction Projects using FiT and CDM Support Mechanisms." *Korean Journal of Construction Engineering and Management*, KICEM, 14(3), pp. 123-133.

Koo, B., Park, J. H., and Kim, C. W. (2014). "Using the Binomial Option Pricing Model for Strategic Sales of CER's to Improve the Economic Feasibility of CDM projects." *Korean Journal of Construction Engineering and Management*, KICEM, 15(1), pp. 111-121.

Lee, G., B. and Lee, E., W. (2005). "Overview and Trend of Small Hydropower Development in Korea." *Korean Society for Fluid Machinery*, pp. 735-741.

Lee, J. H. (2010). "Status and Strategy of CDM projects." *Journal of the Electrical World*, 398, pp. 33-36.

Park, J., H. and Koh, C. (2006) "An Efficient Data Mining Algorithm based on the Database Characteristics," *Journal of the Korean Society for Industrial and Applied Mathematics*, 10(1), pp. 107-119.

Song, J. (2010). "The road to the successful clean development mechanism: lessons from the past." Doctoral dissertation, Massachusetts Institute of Technology.

Swets, J. A., (1988). "Measuring the accuracy of diagnostic systems." *Science*, 240, pp. 1285-1293.

Van Buuren, S. (1999). "Flexible multivariate imputation by MICE, TNO *Prevention and Health*," report PG/VGZ/99,054.

Yim, H. S., and Yun, S. J. (2009). "An Evaluation of Clean Development mechanism(CDM) From a Perspective of Sustainable Development," *ECO*, 13(2), pp. 141-174.

---

**요약 :** 청정개발체제(CDM) 사업은 신재생에너지사업의 보조를 통해 지구온난화 가스의 감축을 꾀하는 대표적인 국가 및 기업 간 배출권 거래(cap and trade)제도이다. 재래식 발전 방식에 비해 수익성이 낮은 태양광, 풍력, 수력 등의 사업이 CDM 사업으로 승인을 받으면 매년 탄소배출권(CER)을 제공받고, 이의 판매를 통해 발생한 추가 수익으로 인해 사업 타당성이 향상될 수 있다. 그러나 CDM 사업으로 인정받기 위해서는 환경적, 기술적, 경제적 추가성(Additionality)를 입증해야 하는데, 해당 적용 기술, 베이스라인 측정 방법론, 온실 가스 감축량, 사업 내부 수익률(IRR) 등 다수의 변수에 따라 결과가 달라지기 때문에 사전적으로 승인 여부를 파악하기가 어렵다. 본 연구에서는 신재생에너지로 분류되는 수력 사업의 CDM 승인 여부를 예측할 수 있는 모델을 개발하는 것을 목표로 하였다. 구체적으로 UNFCCC에서 제공하는 수력 사업 데이터를 활용하여 로그 회귀분석 및 CART 분석을 실시하여 예측 모델을 개발하였으며 이와 함께 승인 여부에 유의하게 영향을 미치는 핵심 인자들을 파악하였다. 구축된 로그 회귀 및 CART 예측 모델은 AUC가 각각 0.7674 및 0.7231로 예측 정확성이 비교적 높게 나왔다. 또한 수력 사업에서는 온실가스 저감량 대비 투자액, 시간당 발전량 및 내부수익률이 승인여부에 유의한 변수들로 파악되었고, 이에 비해 특정 기술이나 측정 방법론은 영향이 없는 것으로 드러났다. 즉, 특정 기술을 불문하고 온실가스를 투자 대비 가장 효율적으로 저감하는 사업과 수력사업들 중 상대적으로 소규모로 진행되는 사업이 CDM 사업으로 승인될 가능성이 높다는 것으로 해석된다.

**키워드 :** CDM, 수력사업, 데이터 마이닝, 로그 회귀분석, CART

---