

캡스톤디자인 중간보고서

제 목	국문	디지털 트윈을 활용한 스마트 팩토리 에너지 효율화 모델링 및 플랫폼 개발		
	영문	Development of Energy Efficiency Modeling and Platform for Smart Factories Using Digital Twin		
진 행 상 황	중요마일스톤	1) 스팀 사용량 예측 모델 구현 - LSTM 기반 스팀 예측 모델 구현 2) 펄프 배합비 클러스터 기반 공정 분류 및 영향도 분석 - 펄프 배합 비율 정규화 및 클러스터링 - 펄프 배합비 클러스터 기반 변수 영향도 분석 및 도출 3) 다목적 함수 기반 공정 최적화 시나리오 도출 - 생산량/품질 효율을 고려한 다목적 함수 정의 - 상관계수 기반 변수 가중치 설정 4) 강화학습 기반 시나리오 생성 및 평가 - 공정 시간 구간(초반/중반/후반)별 전략 학습 - 기존 시나리오와 성능 비교 및 시각화 5) 디지털 트윈 시뮬레이터 프로토타입 구성 - 시나리오 기반 시뮬레이션 시각화 UI 구성 - 실공정과 유사한 변수 궤적 확인 가능 화면 개발 6) 시스템 기능별 테스트 계획 및 실행 - 각 기능(예측/최적화/시뮬레이션)의 처리 속도, 정합성 테스트 계획 수립 및 수행		
	진행상황	LSTM 기반 스팀 예측 모델 구현 완료 펄프 태그 정규화 및 클러스터링 완료 생산량 효율/품질을 고려한 다목적 함수 정의 완료 상관계수 기반 변수 가중치 계산 로직 적용 완료 구간별 전략 최적화 에이전트 학습 완료 강화학습 기반 최적 시나리오 추출 완료 시나리오 기반 변수 궤적 시각화 완료		
산출물	요구사항 정의서(별첨 1), 중간보고서(별첨 2)			
팀 구성원	학년	학 번	이 름	연락처(전화번호/이메일)
	4	20222024	박선아	010-5177-9950 / 20222024@edu.hanbat.ac.kr
	4	20221991	서지윤	010-5039-1073 / 20221991@edu.hanbat.ac.kr
<p>컴퓨터공학과의 프로젝트 관리규정에 따라 다음과 같이 요구사항 정의서와 중간보고서를 제출합니다</p> <p style="text-align: center;">2025 년 05 월 02 일</p> <p style="text-align: right;">책임자 : 박선아 (인) 지도교수 : 이상금 (인)</p>				

[별첨1]

프로젝트명 : 디지털 트윈을 활용한 스마트 팩토리 에너지
효율화 모델링 및 플랫폼 개발

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):박선아
서지윤

대표 연락처: 010-5177-9950

e-mail: 20222024@edu.hanbat.ac.kr

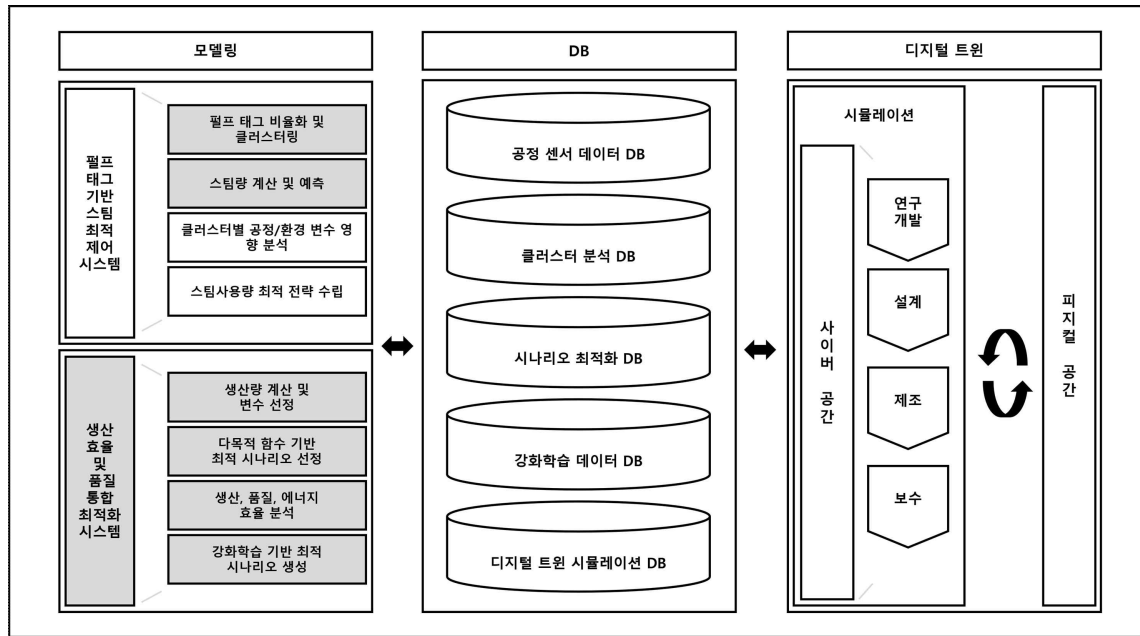
목차

1. 개요
2. 시스템 장비 구성요구사항
3. 기능 요구사항
4. 성능 요구사항
5. 인터페이스 요구사항
6. 데이터 요구사항
7. 테스트 요구사항
8. 보안 요구사항
9. 품질 요구사항
10. 제약 사항
11. 프로젝트 관리 요구사항

1. 시스템 개요

본 시스템은 디지털 트윈 기술을 기반으로 제지 공정의 실증 데이터를 활용하며, LSTM 기반 예측 모델을 활용하여 스팀 사용량을 예측하고, 이를 펄프 태그의 배합 비율로 정규화하여 스팀과 관련된 공정 변수들의 패턴을 분석함으로써, 스팀 사용 효율을 최적화한다. 동시에 다목적 함수와 강화학습 기법을 적용하여 생산성과 품질을 향상시킬 수 있는 최적 운전 시나리오를 제안함으로써, 스마트 공정 최적화 시스템을 구축하는 것을 목표로 한다.

디지털 트윈 기반 스마트 공정 최적화 시스템 구성도
(추진범위는 그레이색상으로 표기함)



2. 시스템 장비 구성요구사항

요구사항 고유번호		ECR-001
요구사항 명칭		모델링 서버
요구사항 분류		시스템 장비 구성 요구사항
요구사항 상세 설명	정의	AI 모델 학습 서버
	세부 내용	<ul style="list-style-type: none"> - 장비 품목 : 연산서버(Jupyter) - 장비 수량 : 1식 - 장비 기능 : AI 모델 학습 및 데이터 분석 - 장비 성능 및 특징 : AI 프레임워크 호환 필요 및 GPU 내장

요구사항 고유번호		ECR-002
요구사항 명칭		DB관리 소프트웨어
요구사항 분류		시스템 장비 구성 요구사항
요구사항 상세 설명	정의	DB 관리
	세부 내용	<ul style="list-style-type: none"> - 장비 품목 : DBMS(InfluxDB) - 장비 수량 : 1식 - 장비 기능 : 실시간 데이터 수집/저장, 시나리오 결과 관리 - 장비 성능 및 특징 : Open API 지원, 시계열 기반의 데이터 베이스 처리 능력 지원 및 최신 표준 SQL 규격 지원

요구사항 고유번호		ECR-003
요구사항 명칭		디지털 트윈 시뮬레이션 소프트웨어
요구사항 분류		시스템 장비 구성 요구사항
요구사항 상세 설명	정의	공정 시나리오 재현 및 시뮬레이션 소프트웨어
	세부 내용	<ul style="list-style-type: none"> - 장비 품목 : 디지털 트윈 플랫폼 SW(UNREAL) - 장비 수량 : 1식 - 장비 기능 : 운전 시나리오 실행 및 시각화, 물리적 공정 매핑 - 장비 성능 및 특징 : Web 기반 시각화 제공, 시나리오별 파라미터 설정 기능

3. 기능 요구사항

요구사항 고유번호		SFR-001
요구사항 명칭		스팀 사용량 최적화 기능
요구사항 분류		기능
요구사항 상세 설명	정의	<p>펄프 기반 스팀 최적 제어 시스템</p> <p>LSTM 모델을 활용하여 스팀 사용량을 예측하고 펄프 배합비 기반으로 영향 요인을 분석하여 제어 가능한 최적 조건을 도출하는 기능</p>
	세부 내용	<ul style="list-style-type: none"> - 공정 변수(속도, 압력 등)와 환경 변수(기압, 일조량 등)를 입력받아 스팀 사용량을 분 단위 예측 - 펄프 배합비 기반 클러스터별 스팀 사용량 및 예측 변수 영향 분석 기능 - 알고리즘 기반 제어 가능 변수 조건 도출을 통한 스팀 사용량 최적화 기능

요구사항 고유번호		SFR-002
요구사항 명칭		품질을 고려한 생산 효율 최적화 기능
요구사항 분류		기능
요구사항 상세 설명	정의	생산 효율 및 품질 통합 최적화 시스템 다목적 함수와 강화학습을 통해 생산 효율과 품질을 고려한 공정 변수 최적 운전 조건 시나리오를 도출하는 기능
	세부 내용	<ul style="list-style-type: none"> - 공정 변수(평량, 속도, 스팀량 등)을 기반으로 다목적 함수 계산 - 다목적 함수 : $y = \sum_{j=1}^3 a_j \cdot \frac{1}{ T_j } \sum_{t \in T_j} (\sum_{i=1}^n w_i \cdot x_i(t))$ - 제약조건 설정과 품질 변수 가중치, 구간 가중치 계산을 통해 품질 유지 기능 - 시나리오별 결과 비교, 시각화 기능

요구사항 고유번호		SFR-003
요구사항 명칭		디지털 트윈 기반 공정 시뮬레이션 기능
요구사항 분류		기능
요구사항 상세 설명	정의	디지털 트윈 기반 공정 시뮬레이션 기능 최적화된 공정 조건을 실제 공장에 적용하기 전 가상 환경에서 시뮬레이션을 수행하여 결과를 예측하고 검증하는 기능
	세부 내용	<ul style="list-style-type: none"> - 최적 시나리오를 사이버 공간에서 재형하여 변수 궤적 확인 기능 - 실제 공정과 시뮬레이션 결과 비교 기능 - 다양한 조건(계절, 품질 편차 등)에 따른 전략 효과 검토 기능

4. 성능 요구사항

요구사항 고유번호		PER-001
요구사항 명칭		스팀 사용량 예측 응답 시간
요구사항 분류		성능
요구사항 상세 설명	정의	스팀 예측 모델이 입력값(공정변수, 환경변수)를 받아 예측값을 반환하는 데 걸리는 평균 시간
	세부 내용	<ul style="list-style-type: none"> - 시스템이 정상 상태일 경우, 사용자가 특정 시점의 공정 조건을 입력했을 때, 1분 이내에 예측 결과를 반환해야 함 - 시뮬레이션 모드에서도 실시간 1분 단위 예측이 누락되지 않도록 처리 속도 보장 필요

요구사항 고유번호		PER-002
요구사항 명칭		최적 시나리오 도출 시간
요구사항 분류		성능
요구사항 상세 설명	정의	다목적 최적화 모델이 입력값을 기반으로 각 시나리오별 y 점수를 계산하고 최적 시나리오를 도출하는 데 소요되는 시간
	세부 내용	<ul style="list-style-type: none"> - 목적함수 y 계산시, 총 처리 시간이 5분을 초과하지 않아야 함 - 강화학습 기반 시나리오 생성 모델도 동일한 수준의 처리 시간 내에 수행되어야 함

5. 인터페이스 요구사항

요구사항 고유번호		SIR-001
요구사항 명칭		스팀 사용량 예측 결과 시각화 대시보드
요구사항 분류		인터페이스
요구사항 상세 설명	정의	스팀 사용량 예측 결과와 계산식 기반 스팀 사용량을 비교할 수 있는 시각화 기능 제공
	세부 내용	<ul style="list-style-type: none"> - 스팀 사용량 계산값 및 예측값을 확인하기 쉽게 1분 단위 그래프로 시각화 기능이 제공되어야 함 - 현 시점의 스팀 사용량을 직관적으로 파악할 수 있게 수치가 제공되어야 함
주석		공정 운영자는 실시간 모니터링과 이상 징후 탐지 등을 위해 직관적인 그래프와 수치표시 기능이 필요함
요구사항 출처		ETRI

요구사항 고유번호		SIR-002
요구사항 명칭		최적 시나리오 비교/선택 시각화 대시보드
요구사항 분류		인터페이스
요구사항 상세 설명	정의	지중, 평량별 다목적 함수로 도출된 기존 최적 시나리오와 강화학습을 통해 생성된 최적 시나리오를 비교하고 선택할 수 있는 기능 제공
	세부 내용	<ul style="list-style-type: none"> - 최적 시나리오 후보를 공정 변수별로 비교가 용이 하게 시각화 기능이 제공되어야 함 - 생산량 효율, 품질, 에너지 효율 관련 공정 변수의 증가/감소 폭이 수치적으로 제공 되어야 함
주석		공정 운영자의 신속하고 정확한 의사결정을 지원할 수 있도록, 정보는 이해하기 쉬운 표 형식으로 제공되어야 하며, 주요 지표나 상태는 직관적인 색상 구분 등을 통해 시각적으로 구분 가능해야 함
요구사항 출처		ETRI

6. 데이터 요구사항

요구사항 고유번호	DAR-001
요구사항 명칭	공정 데이터 수집 및 구축
요구사항 분류	데이터
요구사항 세부내용	<ul style="list-style-type: none"> - 수집된 데이터는 시간 기반 인덱스를 기준으로 저장되고 조회 가능해야 함 - 데이터 저장 주기는 1분 단위로 설정되며, 대용량 처리를 위한 시계열 DB를 사용해야 함

요구사항 고유번호	DAR-002
요구사항 명칭	공정 데이터 정합성 검증
요구사항 분류	데이터
요구사항 세부내용	<ul style="list-style-type: none"> - 외부 데이터의 연계 시, 반드시 데이터의 정합성을 체크하고 로그를 유지하여야 함 - 데이터 타입, 단위, 해상도(초/분 단위 등)에 대한 표준 정의가 필요함

7. 테스트 요구사항

요구사항 고유번호	TER-001
요구사항 명칭	테스트 방안
요구사항 분류	테스트
요구사항 세부내용	<ul style="list-style-type: none"> - 시스템이 공정 데이터를 받아 예측, 최적화, 시뮬레이션 등 전 과정을 통합적으로 수행하는지 검증 - 각 모듈(LSTM 예측, 다목적 최적화, 디지털 트윈 시뮬레이션, UI 대시보드)이 정상적으로 연동되는지 확인 및 그래프 UI 반응 테스트 포함 - 결과의 정확성 및 누락 여부 확인, 에러 로그 자동 기록 기능 확인

8. 보안 요구사항

요구사항 고유번호	SER-001
요구사항 명칭	보안지침 준수
요구사항 분류	보안
요구사항 세부내용	<ul style="list-style-type: none"> - ETRI, 국가정보원 보안 규정 및 지침을 준수하여 개발이 수행되어야 함

요구사항 고유번호	SER-002
요구사항 명칭	응용 및 DB보안
요구사항 분류	보안
요구사항 세부내용	<ul style="list-style-type: none"> - 강화학습 결과, 예측 결과, 공정 데이터 등 주요 정보가 저장되는 DB는 암호화를 통해 보호되어야 함 - 외부 시스템(MES/ERP 등) 연동 시, 데이터 요청 및 응답은 SSL 또는 VPN 기반으로 암호화하여 통신되어야 함 - 데이터 조회·수정·삭제 시 모든 행위에 대해 로그가 자동으로 기록되며, 관리자 이외에는 직접 접근이 제한되어야 함

9. 품질 요구사항

요구사항 고유번호		QUR-001
요구사항 명칭		시스템 신뢰성
요구사항 분류		품질
요구사항 상세 설명	정의	시스템이 지정된 조건에서 일정 시간 동안 고장 없이 연속적으로 작동하고, 고장 발생 시 신속하게 복구될 수 있어야 함
	세부 내용	<div>- 시스템 주요 기능(LSTM 예측, 시뮬레이션, 시나리오 선택 등) 중단 시 알람이 발생하며 관리자에게 자동 통보되어야 함</div> <div>- 고장이 발생한 경우, 시스템은 평균 복구 시간(MTTR)이 30분 이내로 제한되어야 함</div>

요구사항 고유번호		QUR-002
요구사항 명칭		사용자 운용 및 학습
요구사항 분류		품질
요구사항 상세 설명	정의	사용자가 시스템 기능을 빠르게 이해하고 쉽게 운용할 수 있도록 직관적인 UI와 학습 가이드를 제공해야 함
	세부 내용	<ul style="list-style-type: none">- 초기 사용자라도 30분 이내 튜토리얼을 통해 기본 기능(예측 결과 조회, 시뮬레이션 실행, 최적 시나리오 확인 등)을 습득할 수 있어야 함- 기능 사용 중 발생 가능한 오류에 대해 명확한 에러 메시지와 해결 안내가 제공되어야 함- 화면 구성은 표, 그래프, 색상 등 시각적 요소 중심으로 구성되어야 하며, 복잡한 기능은 단계적으로 안내되도록 설계해야 함

10. 제약 사항

요구사항 고유번호	COR-001
요구사항 명칭	개발 및 구현 환경 제약
요구사항 분류	제약사항
요구사항 세부내용	<ul style="list-style-type: none">- 모델링 및 예측 기능은 Python 기반 프레임워크(PyTorch, TensorFlow 등)를 사용해야 함- 웹 인터페이스는 사내 서버 환경에 배포 가능한 구조여야 함

11. 프로젝트 관리 요구사항

요구사항 고유번호	PMR-001
요구사항 명칭	수행 조직
요구사항 분류	프로젝트 관리
요구사항 세부내용	<ul style="list-style-type: none">- 프로젝트 수행을 위해 다음과 같은 팀을 구성하고 역할을 명확히 해야 함- AI 모델링팀: 데이터 분석, LSTM/강화학습 모델 설계 및 검증- 시각화/시뮬레이션팀: 디지털 트윈 시뮬레이션 UI/UX 및 통합 운영 화면 구현

요구사항 고유번호	PMR-002
요구사항 명칭	일정 계획
요구사항 분류	프로젝트 관리
요구사항 세부내용	<ul style="list-style-type: none">- 계획된 프로젝트의 일정 및 단계별 산출물을 명확히 정의해야 함- 1단계 - 요구사항 분석: 요구사항 정의서, 초기 데이터 진단 보고서- 2단계 - 모델 개발 및 최적화: 예측/최적화 모델 코드, 평가 보고서- 3단계 - 시뮬레이션 구현 및 UI 연동: 디지털 트윈 시각화 화면, 테스트 시나리오 결과

[별첨2]

중간보고서

1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 분석, 설계, 구현(소스코드 작성) 및 테스트한 내용을 기술하시오.

1. 다목적 함수를 활용한 생산량/품질 최적 시나리오 도출

1-1. 데이터 전처리 및 변수 선정

제지 공정의 생산량 및 품질 최적화를 위해 다목적 함수 기반 분석을 수행하였다. 사용된 데이터는 실제 제지공정에서 수집된 2000여개의 열을 가진 데이터로 기간은 2022년 1월 1일~11월 30일의 1분 단위 시계열 데이터이다. 분석에 앞서, 다음과 같은 기준으로 데이터를 전처리하였다.

- 공정 단위 필터링 : 30분 이상 60분 이하의 공정 시간 데이터를 대상으로 필터링.
- 결측치 처리: 결측 비율이 높은 변수 제거 후, 시간 순서를 기반으로 Forward Fill과 Backward Fill 방식의 보간 수행.
- 이상치 제거: 공정 데이터의 시간적 연속성과 국소적 이상값 제거를 위해 Rolling Z-score 방식으로 이상치 탐지 및 제거.

또한 생산량/품질 최적 시나리오 도출에 사용할 주요 변수로는 아래 변수를 사용하였다.

- 생산량 관련 변수: 속도, 지로 유량
- 품질 관련 변수: 평량(종이의 무게), 수분
- 에너지 관련 변수 : 후건조기 압력

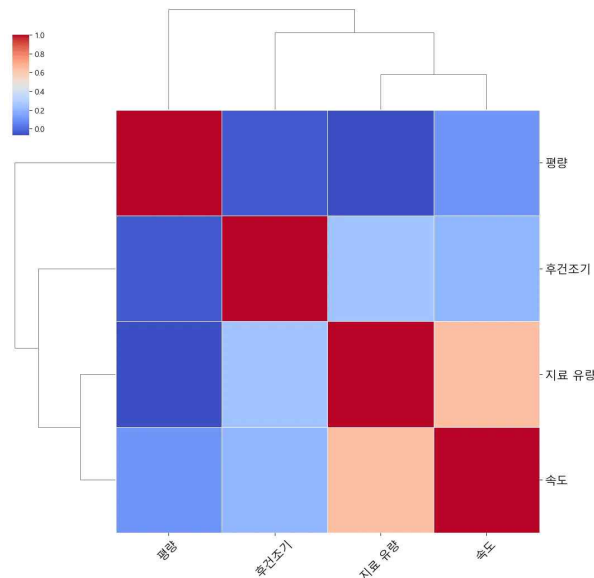


그림 1. 사용된 주요 변수의 상관관계 히트맵

1-2. 다목적 함수 정의

생산량과 품질을 종합적으로 고려하기 위해 다음과 같은 다목적함수를 구성하였다.

$$y = \sum_{j=1}^3 a_j \cdot \frac{1}{|T_j|} \sum_{t \in T_j} \left(\sum_{i=1}^n w_i \cdot x_i(t) \right)$$

수식 1. 다목적 함수

$x_i(t)$	시간 t에서의 공정 변수 i 값
w_i	변수 i의 중요도 (가중치)
T_j	전체 공정 시간의 구간 j (예: 앞, 중간, 뒤 15%/70%/15%)
a_j	구간 j의 품질 민감도 또는 분산 계수
y	시나리오의 생산성과 품질 종합 지표

해당 다목적함수는 시간 흐름에 따른 공정의 세부적 특성과 생산 품질 간의 상관성을 반영하여 설계된 구조이다. 다음과 같은 세 가지 주요 논리를 바탕으로 구성되었다.

- 공정 구간의 분할 및 구간 가중치(a_j)

제지 공정은 하나의 공정 단위 내에서도 초반, 중반, 후반 구간에 따라 운전 조건과 품질의 안정성이 상이하다. 특히 초반부와 후반부는 다른 공정 단위(이전/이후 공정 단위)와 연속적으로 연결되는 구간이므로, 설비 상태 변화나 조건 전이로 인해 흔들림이 발생하기 쉽고, 이로 인해 품질 저하의 가능성이 크다. 반면, 중반부는 상대적으로 안정적인 운전이 유지되는 구간으로, 실제 상품으로 출하되는 종이의 품질은 중반부 상태에 가장 큰 영향을 받는다. 따라서 공정 단위를 시간 세 구간(초반 15%, 중반 70%, 후반 15%)으로 나누고, 이 구간별로 품질 민감도를 반영한 구간 가중치(a_j)를 주었다. 이를 통해 생산량 자체뿐만 아니라 품질이 안정된 구간에서의 생산성이 높은 시나리오가 더 높은 평가를 받을 수 있도록 다목적함수를 설계하였다.

- 공정 변수별 가중치(w_i)

다목적함수 계산에 포함되는 공정 변수는 다수 존재하나, 모든 변수를 동일한 비중으로 취급할 경우, 실제로 품질이나 생산량에 큰 영향을 주지 않는 값들까지 분석에 반영되어 왜곡될 가능성이 있다. 이를 방지하고 실질적으로 영향력이 큰 변수 중심의 평가를 하기 위해, 각 변수와 아래 방정식을 통해 계산된 1분 단위의 계산 생산량 간의 상관계수를 기반으로 변수별 가중치(w_i)를 산정하였다.

$$\text{생산량} = \text{지폭} \times \text{평량} \times \text{속도} \times \text{시간}$$

수식 2. 계산 생산량 방정식

따라서 전체 데이터에서 각 변수(x_i)와 계산 생산량 간 상관계수를 계산하고 절댓값 기준으로 정규화 후 가중치(w_i)로 변환 하였다. 이러한 방식을 통해 다목적함수는 단순 평균이 아닌, 생산성과 품질에 실질적으로 영향을 주는 변수 중심의 가중 합산 구조로 구성된다.

- 공정 구간별 시간 정규화($\frac{1}{|T_j|}$)

각 공정 단위는 시간 길이(구간별 시점 개수)와 총 생산량이 다르다. 이를 그대로 다목적함수에 적용할 경우 긴 LOT일수록 높은 점수를 받는 구조적 편향이 발생할 수 있다. 따라서 각 구간 T_j 에 대해 시점 수 $|T_j|$ 로 나누어 평균화하는 정규화를 수행하였다. 이 과정을 통해서 다른 길이의 공정 단위간에도 공정하고 일관된 비교가 가능하게 하며, 최종적으로 구간 가중 평균을 종합하여 하나의 점수 y 로 수치화하였다.

또한 실제 공정에서는 각 운전 변수들이 갖는 물리적 또는 운전상의 한계값이 존재한다. 이를 고려하지 않고 최적화 시나리오를 도출할 경우, 이론적으로는 다목적함수 y 가 높더라도 현실에서 적용 불가능한 결과가 도출될 가능성이 있다. 이를 방지하기 위해, 본 연구에서는 다음과 같은 변수 제약조건을 정의하여 최적화 과정에 적용하였다.

$$x_{i,\min} \leq x_i(t) \leq x_{i,\max} \quad \forall i, t$$

수식 3. 제약 조건

1-3. 최적 시나리오 도출 및 평가

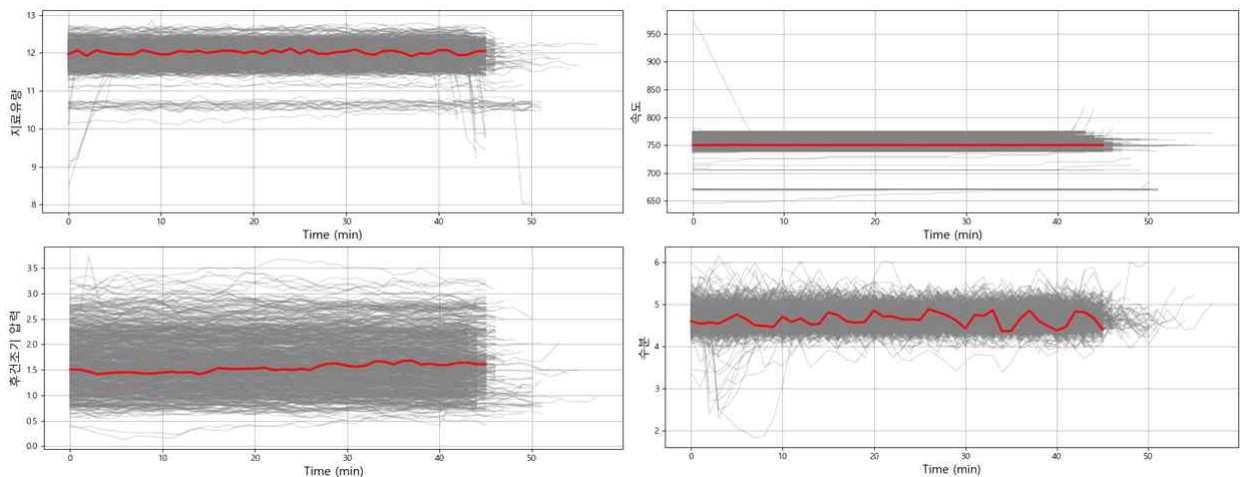


그림 2. 다목적 함수로 선정된 최적 시나리오

다목적 함수 기반 최적화 분석을 통해 기존 공정 단위 데이터 중에서 생산성과 품질 지표를 동시에 고려한 이상적인 운전 시나리오(최적 공정 단위)를 도출하였다. 해당 시나리오는 공정의 핵심 변수들을 종합적으로 평가한 결과 가장 우수한 다목적함수 점수 y 를 기록하였으며, 기존 운전 평균과 비교하여 다음과 같은 개선 효과를 확인하였다.

- 분당 생산량 : 기존 평균 448.78 g/min에서 최적 시나리오에서는 461.01 g/min으로 12.23 g/min 증가하였다. 이는 약 2.7%의 생산성 향상에 해당한다.
- 수분 : 4.66%에서 최적 시나리오에서는 4.63%로 약 0.03% 감소하여 품질의 안정성은 거의 유지되었다.

- 평량 : 해당 지종의 평량은 172로 기존 평균 161.65g에서 166.93g으로 5.28g 증가하여 품질 및 상품성이 향상되었다.
- 후건조기 압력 : 에너지 효율과 관련된 후건조기는 1.65kg/cm²에서 1.54kg/cm²으로 0.12kg/cm² 감소하여 에너지 효율성이 향상될 수 있는 운영 조건이 선택되었음을 확인 할 수 있었다.

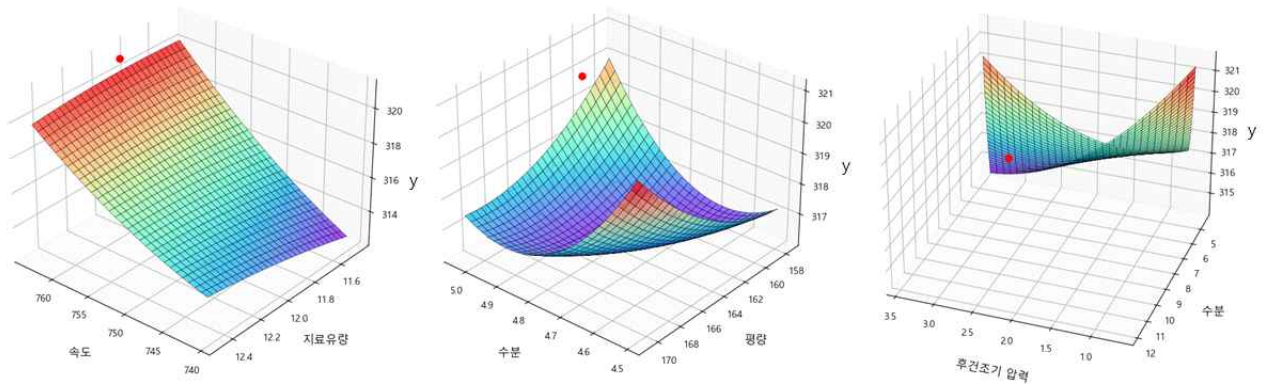


그림 3. 다목적 함수 기반 3D 회귀 곡면

도출된 최적 시나리오의 변수 조합은 전체 운전 데이터 기반의 다항 회귀 곡면 위에서도 높은 목적함수 값을 가지는 영역에 위치하고 있음을 확인하였다. 위 그림 3에서는 주요 변수 조합별로 계산된 다목적 함수 곡면과, 그 위에 표시된 최적 공정 구간(최적 시나리오)를 함께 시각화하였다. 이를 통해 본 최적 시나리오가 데이터 전체 분포 구조 안에서 실현 가능한 범위 내에 위치하며, 동시에 높은 평가 점수를 획득하고 있음을 확인할 수 있다.

2. 강화학습 기반 최적 시나리오 생성

2-1. 강화학습 기반 구간별 제어 구조 및 보상 계산

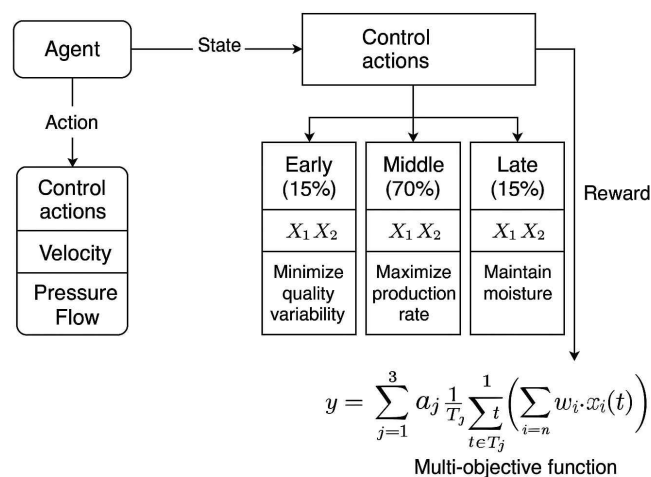


그림 4. 강화학습 구조

기존에 정의한 다목적 함수 구조를 강화학습의 보상 함수로 활용하여, 공정 구간별 최적 운

전 전략을 학습할 수 있는 강화학습 환경을 설계하였다. 전체 공정 단위 데이터를 기반으로 각 시점별 공정 상태를 상태(state)로 정의하고, 이를 바탕으로 속도, 지료유량, 압력 등의 공정 변수를 조절하는 행동(action)을 선택하도록 하였다. 특히 공정은 시간 순서를 기준으로 초기(15%) / 중간(70%) / 후반(15%)의 세 구간으로 나누어 각 구간별로 전략 목표를 다르게 설정하였다. 초기에는 품질 변동성 최소화, 중간에는 생산량 극대화, 후반에는 품질 유지를 목표로 하며, 이에 따라 각 구간의 중요도 가중치(a_j)를 반영한 다목적 함수 형태의 보상 y 를 계산하였다. 그림 4는 강화학습 환경 구조를 시각적으로 나타낸 것으로, 에이전트가 상태를 기반으로 행동을 선택하고, 구간별 전략 목적을 달성했는지에 따라 보상 y 를 받는 과정이다.

2-2. 강화학습 기반 최적 시나리오 생성 및 평가

앞서 학습된 강화학습 에이전트(PPO 기반)를 활용하여 공정 데이터를 기반으로 시뮬레이션을 수행하고, 최적 운전 시나리오를 생성하였다. 에이전트는 각 시간 구간(초기, 중간, 후반)에 따라 설정된 전략 목표와 다목적 함수 기반 보상을 고려하여, 실시간으로 속도, 압력, 유량 등 주요 공정 변수의 운전 값을 선택하였다. 강화학습 기반으로 도출된 최적 시나리오와 기존 LOT 전체 평균을 비교한 결과는 다음과 같다.

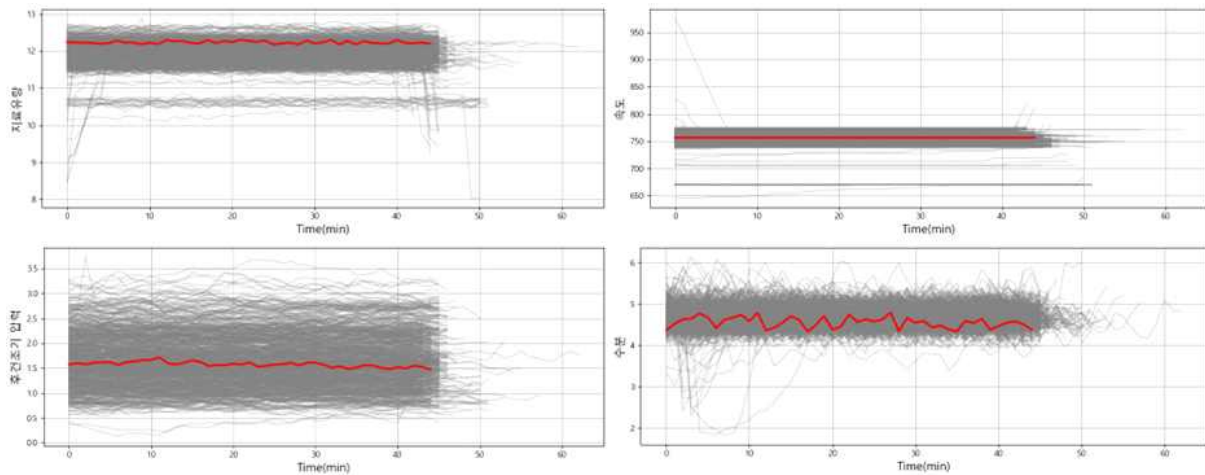


그림 5. 강화학습으로 선정된 최적 시나리오

- 분당 생산량 : 기존 평균 438.71 g/min에서 에이전트 시나리오에서는 433.81 g/min으로 4.90g/min 감소하였다. 이는 강화학습 과정에서 일부 변수에 대해 품질 혹은 안정성 위주 전략이 선택된 결과로 해석된다.
- 수분 : 44.66% → 4.56%로 0.11% 감소하여 품질의 안정적이다.
- 평량 : 해당 지종도 마찬가지로 평량은 172이다. 하지만 결과는 161.66g → 159.23g으로 2.42g 감소하여 품질 측면에서 다소 부족한 결과를 보였다.
- 후건조기 압력 : 에너지 효율과 관련된 후건조기는 1.65kg/cm² → 1.57kg/cm²로 0.08kg/cm² 감소하여, 에너지 효율 향상을 고려한 운전 조건이 선택되었음을 알 수 있다.

결론적으로 생산량/품질 최적화 모델링 및 분석에서는 동일한 다목적함수 구조를 기반으로 두 가지 방식의 최적 시나리오를 비교하였다. 하나는 실제 데이터를 활용해 도출한 다목적 함수 기반 최적 시나리오, 다른 하나는 강화학습을 통해 새롭게 생성된 전략 기반 시나리오이다. 다목적 함수 기반 시나리오는 실제 공정에서 관측된 공정 단위 데이터를 바탕으로 하여, 현실성이 높고 신뢰할 수 있는 운전 조건을 제시한다. 이 시나리오에서는 분당 생산량이 약 12.23g/min(2.7%) 증가하였고, 수분 및 평량 등 품질 지표 또한 양호하게 유지 또는 향상되었으며, 후건조기 압력은 감소하여 에너지 효율 향상 가능성까지 나타났다.

반면, 강화학습 기반 시나리오는 과거 데이터로부터 학습한 정책에 따라 새로운 운전 시나리오를 생성한 것으로, 실제 공정 단위 데이터에서 관측되지 않은 전략을 탐색한 점에서 의미가 있다. 생산량은 소폭 감소하였으나, 수분은 더 안정적인 수준으로 유지되었고, 후건조기 압력도 절감되는 방향을 보였다. 그러나 평량은 기대와 달리 감소하는 결과를 보였으며, 이는 reward 설계 및 학습 방향에 따라 조정이 필요한 부분이다. 이와 같이, 강화학습 기반 전략은 절대적인 성능에서는 다목적 함수 기반 시나리오에 미치지 못했지만, 기존 데이터로부터 파생되지 않았던 새로운 운전 전략을 제시했다는 점에서 의미가 있다. 특히, 강화학습은 reward 구조와 데이터 구성에 따라 전략적 방향을 자유롭게 탐색할 수 있으므로, 향후 reward 설계 개선, 에피소드 수 확장, 제약조건 반영 방식 조정 등을 통해 성능 향상이 충분히 가능하다고 생각한다.

최종적으로, 본 연구는 현실성과 신뢰성이 높은 다목적 기반 최적화 시나리오와, 확장성과 탐색성이 뛰어난 강화학습 기반 전략을 병렬적으로 분석함으로써, 제지 공정 운전 최적화에 있어 다양한 접근 방식을 실험하고 가능성을 확인하였다.

3. 계절성을 반영한 LSTM 기반 스팀 사용량 예측 모델 구현

3-1. 데이터 전처리 및 변수 선정

제지 공정의 스팀 사용량 파악 및 계절성을 반영하기 위해 스팀 사용량 예측 모델링을 진행하였다. 사용된 데이터는 실제 제지공정에서 수집된 2000여개의 열을 가진 데이터로 기간은 2022년 1월 1일~ 11월 30일의 1분 단위 시계열 데이터이다. 또한 계절성 반영을 위해 기상청 데이터를 병합하여 함께 사용하였다. 분석에 앞서, 다음과 같은 기준으로 데이터를 전처리하였다.

- 상위 지종 선정 : 전체 데이터 중 샘플 수가 많은 상위 5개 지종을 선정하여 분석에 활용
- 결측치 처리 : 결측치의 비율이 높은 변수는 제거하고, 나머지는 시간 순서를 기반으로 최근 값 대체 방식으로 보간.
- 이상치 제거 : 시계열의 특성을 반영하여 이상치를 제거하고자 Rolling window 방식으로 IQR 기반 이상치 검출 및 제거.

스팀 사용량 예측 모델 구현에 사용할 주요 변수는 다음과 같은 방법으로 선정하였다.

- 스팀 사용량 : 아래 방정식을 통해 계산된 1분 단위 계산 스팀 사용량 사용

$$\begin{aligned} \text{계산 스팀 사용량} = & 0.5234 \times \text{Reel 평량} - 1.5031 \times \text{Size 평량} \\ & - 66.4690 \times \text{Reel 수분량} + 55.3265 \times \text{지료유량} + 582.0949 \end{aligned}$$

수식 4. 계산 스팀 사용량 방정식

- 환경 변수 : MI(Mutual Information) score를 이용해 환경 요인과 스팀 사용량의 상호 의존성 분석 → MI score가 높은 일조, 일사, 현지 기압 선정
- 공정 변수 : 스피어만 상관계수를 통해 에너지 및 스팀 관련 변수와 스팀 사용량 분석 → 상관계수의 절댓값이 0.4 이상인 지료유량(0.73), 전건조기 압력(0.43)을 주요 변수로 선정
- 공정 영향 요소 : 평량(종이의 무게)도 변수로 사용

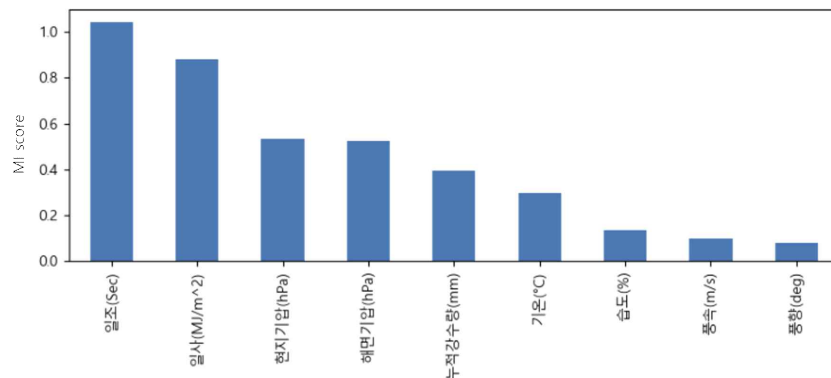


그림 6. MI score를 통한 환경변수 선정

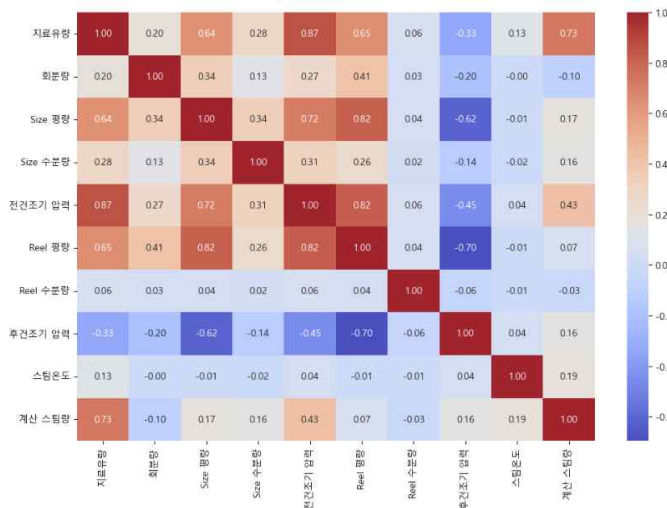


그림 7. 공정 변수 선정을 위한 상관관계 히트맵

3-2. LSTM 기반 스팀 사용량 예측 모델링

입력변수 X는 앞서 말한 평량, 지료유량, 전건조기 압력, 일조량, 일사량, 현지기압이며, 출력변수 Y는 계산된 스팀 사용량이다. 모든 변수들은 MinMaxScaler를 통해 정규화를 진행하였다. 모델은 시계열 형태로 되어있는 공정 데이터의 특성을 반영하기 위해 LSTM(Long Short-Term Memory) 모델을 선정하였으며, 하이퍼파라미터는 다음과 같이 설정하였다.

- Optimizer: Adam
- Loss function: MSE

- Epochs: 50

3-3. 모델 성능 평가

모델은 R^2 와 오류 평가 지표인 MAE, RMSE로 평가하였으며 각각 0.9187, 8.20, 11.75로 높은 성능을 보여주었다. 또한 하나의 공정 사이클에 해당하는 구간에 적용했을 때 아래 그림 8과 같이 예측이 진행되었으며, 예측 정확도 약 99.51% 평균 오차 3.43kg/m², 평균 절대 오차를 0.49%를 보였다. 여러 샘플 구간에 대해 동일 모델을 적용했을 때도 전체 평균 오차는 7.92kg/m², 전체 평균 절대 오차율은 1.10% 수준으로, 본 모델이 다양한 조건에서도 높은 예측 정확도를 유지함을 확인할 수 있었다. 이러한 결과는 해당 LSTM 기반 모델이 시간 흐름에 따라 변화하는 공정 데이터의 특성을 잘 반영하고 있으며, 정확하고 신뢰성 있는 스팀 사용량 예측이 가능함을 보여준다.

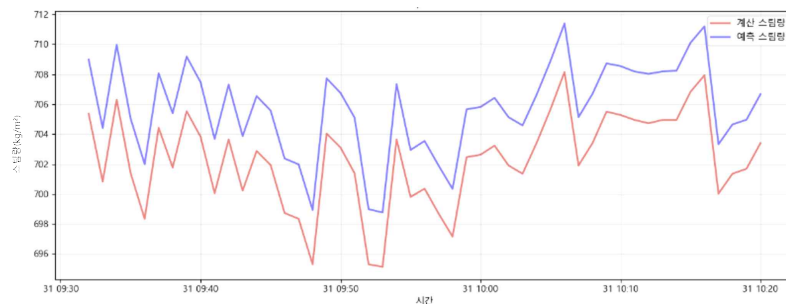


그림 8. 공정 사이클별 스팀 사용량 예측 시각화

4. 펄프 배합 정규화 및 클러스터링

4-1. 펄프 배합 비율 전환

제지 공정에서 스팀은 원지를 건조하는데 사용되며, 이 스팀 사용량은 펄프의 종류와 배합 비율에 영향을 받는다. 기존에 구축한 계절성을 반영한 스팀 사용량 예측 모델의 결과를 기반으로, 스팀 사용량과 펄프 배합 간의 관계를 분석하고 최적화 방향을 도출하기 위해 펄프 데이터를 비율로 전환하였다. 총 5종류의 펄프가 혼합되며, 각 1분 단위(row)를 기준으로 이들 펄프의 상대적인 비율을 정규화하여 활용하였다.

4-2. 펄프 배합비 클러스터링

정규화된 펄프 배합 비율 데이터를 기반으로 클러스터링을 수행하여, 배합 유형의 특성을 파악하고 분석을 용이하게 하였다. 클러스터링에는 K-means 알고리즘을 사용하였으며, 클러스터 수(k)는 2로 설정하였다. 이는 작업자 인터뷰를 통해 실제 작업이 잘 안될 시 조정하는 두 종류의 펄프를 중심 변수로 삼아 군집화를 설계한 것이다. 클러스터링 결과로 도출된 라벨은 기존 데이터에 병합하였으며, 각 클러스터에 속한 배합 유형이 공정 성능(스팀 사용량, 품질 등)에 어떤 영향을 미치는지 추가적으로 분석할 예정이다.

한편, 계절성을 반영한 스팀 예측 모델은 환경 변수, 공정 변수, 계산식 기반 스팀 사용량을 통합하여 1분 단위의 스팀 사용량을 예측하였다. 이 모델은 계절적 요인에 따른 공정 변동성을 반영하면서도 비교적 낮은 오차와 높은 예측 성능을 보여 신뢰성을 입증하였다.

하지만 클러스터링을 통해 구분된 펄프 배합비 클러스터와 스팀 사용량 간의 변수 영향도 등의 구체적인 분석은 진행하지 못해 추가 분석이 필요한 상황이다. 다음 단계에서는 클러스터 유형 간 공정 성능 차이와 품질 영향도에 대한 정량적 비교 분석을 수행할 예정이다. 에너지 효율 측면에서는 각 클러스터 별 평균 스팀 사용량을 비교해 동일한 생산 조건에서 배합 차이에 따른 에너지 소비량의 차이를 평가해 볼 계획이다. 또한, 환경변수에 따른 스팀 사용량의 민감도를 분석하고, 품질 지표를 조합하여 펄프 배합비율에 따라 품질을 유지하되 스팀 사용량이 낮은 공정 최적 조건을 모델을 통해 도출하고자 한다.

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.

본 프로젝트는 당초, 공정 센서 데이터를 기반으로 생산량 예측 모델(LSTM 기반)을 구축하고, 이를 활용하여 제지 공정의 생산 효율을 향상시키는 것이 목표였다. 특히, 지폭, 평량, 속도, 시간을 활용하여 생산량을 계산한 후, 공정 변수들과의 상관관계를 분석하여 예측에 적합한 입력 변수를 선정하고자 하였다. 그러나 초기 실험 결과, LSTM 모델은 전체 시점 중 공정 시작 구간(1~5분)에서 예측 성능이 매우 불안정하게 나타났다. 이 현상에 대해 분석한 결과, 공정의 초기 구간은 앞단 공정과 연결되는 전환 시점이기 때문에 센서 값이 많이 흔들리고 품질 변동성이 크다. 이로 인해 예측 모델이 안정적인 패턴을 학습하지 못하고 있음을 발견하였다.

이를 통해 단순히 정량적 생산량 예측만으로는 실제 공정의 안정성과 품질 특성을 충분히 반영하기 어렵다는 문제를 인식하게 되었고, 이에 따라 방향을 전환하였다. 단일 목표(생산량 예측)에서 다목적 최적화로의 전환을 결정하였으며, 생산량뿐 아니라 공정 안정성(변동성), 품질(평량/수분), 에너지 효율(건조기 압력)까지 종합적으로 고려한 목적함수를 새롭게 설계하였다. 새롭게 정의된 목적함수는 전체 공정을 시간 기준으로 세 구간(초기/중간/후반)으로 나누고, 구간별로 품질 안정성, 생산성, 수분 유지 등의 전략 목표를 반영하여 가중치를 부여하였다. 이를 통해 특정 시점의 오차가 아닌, 운전 시나리오 전반의 전략적 최적화를 추구하는 구조로 다시 설계하였다.

이러한 과정을 통해 기존 예측 모델만으로는 파악하기 어려웠던 운전 전략의 흐름과 특성을 시나리오 단위로 더 명확하게 분석할 수 있게 되었으며 다목적 함수 기반의 구조 덕분에 강화학습을 활용한 새로운 운전 전략 탐색도 가능해졌다. 이런 식으로 문제를 발견하고 방향을 바꾼 경험은, 단순히 데이터를 예측하는 것만으로는 부족하고, 실제 공정에서 일어나는 상황과 운전 방식까지 함께 생각해야 한다는 걸 알게 해주었다.

덕분에 이 프로젝트는 처음보다 더 현장에 맞고, 앞으로도 다양하게 활용할 수 있는 방식으로 발전할 수 있었다.

본 프로젝트에서 스팀 사용량은 예측 기반으로 한 에너지 효율화, 공정 데이터와 환경 변수를 활용해 계절성을 반영한 예측 모델을 구축하는 것을 1차 목표로 설정하였다. 초기에는 공정 변수와 함께 환경 변수를 기온, 습도로 설정해 LSTM 스팀량 예측 모델을 학습했을 때, 성능 평가 지표가 이론적으로는 높게 나왔으나, 실제 예측 오차가 더 커지거나, 변수의 조합에 따라 오히려 성능이 저하되는 경우도 종종 발생했다. 기온과 습도의 경우, 건조해지는 등

계절의 변화가 공정과의 연관성은 높지만, 실제 공정이 폐쇄 공정의 형태로 진행되어 해당 변수가 일관되게 영향을 미치지 않거나, 영향력이 크지 않다고 판단하였다.

이에 따라, 일조량, 일사량, 기압 등 다른 환경 변수의 조합을 실험하였고, 상관관계를 기반으로 주요 입력 변수를 재구성한 결과, 더 안정적인 예측 성능을 얻을 수 있었다. 이런 경우 공정에 실질적인 영향을 미치는 핵심 변수를 잘 파악하고, 선택하여 입력하는 방식이 예측 정확도와 해석 모두에서 효과적이라는 결론을 얻을 수 있었다. 이를 통해 연관된 변수의 수보다는, 목적에 맞는 최소한의 정보가 더 높은 예측력을 가질 수 있다는 것을 다시금 느낄 수 있었다.

그리고 해당 예측 모델을 완성한 이후, 펄프 배합 비율, 수분량 등 공정 내 물성 특성에 따라 스팀량이 영향을 받는다는 점을 고려하게 되면서, 분석 범위를 단순 스팀량 예측에서 펄프 배합비에 따른 공정 효율화로 확대하게 되었다. 그러나 분석 범위가 넓어짐에 따라 고려해야 할 변수와 상호작용이 복잡해졌고, 이로 인해 전체적인 프로젝트 진행 방향을 잡고 흐름을 해석하는 데 어려움이 생겼다. 다양한 데이터와 변수의 연관성을 생각해야하는데 기존에 진행했던 스팀량 예측 모델에만 생각을 고정하고, 펄프 배합비의 영향을 한 번에 파악하려고 해 분석 방향을 설정하는 데 많은 시간을 소요하게 되었다.

이 문제를 해결하기 위해서 범위가 너무 넓을 때는 비슷한 특성이나 역할을 가진 변수끼리 그룹을 묶어서 작고 명확한 단위로 분석하는 방안으로 접근하라는 조언을 듣고 클러스터링을 시작해 방향성을 조금 잡을 수 있었다. 또한 기존에 진행했던 모델에 얽매이지 않고, 아예 처음부터 다시 문제를 재정의해 분석을 진행하는 것도 체계적인 하나의 방법이 될 수 있다는 걸 느끼고 있다.

이런 과정들을 통해 단순히 해당 단계에서만 쓰이는게 아니라 멈추지 않고 연관 되어있는 변수를 생각해보면서, 실무 공정에서 각각의 요소들이 분리되어 보이지만 실제로는 유기적으로 연결되어 있다는 점을 인지하고, 데이터를 분석하고 적용할 때 시야를 넓게 바라볼 수 있게 해주는 기회가 되었던 것 같다.